



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ & ΡΟΜΠΟΤΙΚΗΣ

Ανίχνευση και Ανάλυση Υφής και
Μικροδομών με Μη-γραμμικά και
Στατιστικά Μοντέλα

ΕΚΘΕΣΗ ΕΝΔΙΑΜΕΣΗΣ ΚΡΙΣΗΣ

του

ΓΕΩΡΓΙΟΥ ΕΥΑΓΓΕΛΟΠΟΥΛΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π. (2001)

Αθήνα, Σεπτέμβριος 2005



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

Ανίχνευση και Ανάλυση Υφής και Μικροδομών με Μη-γραμμικά και Στατιστικά Μοντέλα

ΕΚΘΕΣΗ ΕΝΔΙΑΜΕΣΗΣ ΚΡΙΣΗΣ

του

ΓΕΩΡΓΙΟΥ ΕΥΑΓΓΕΛΟΠΟΥΛΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π. (2001)

Συμβουλευτική Επιτροπή: καθ. Πέτρος Μαραγκός
καθ. Σπύρος Τζαφέστας
καθ. Γεώργιος Καραγιάννης

Εγκρίθηκε από την πενταμελή εξεταστική επιτροπή την 2005.

...
Π. Μαραγκός	Σ. Τζαφέστας	Γ. Καραγιάννης
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

...	...
Γ. Παπαβασιλόπουλος	Σ. Κόλλιας
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2005

...

ΓΕΩΡΓΙΟΣ ΕΥΑΓΓΕΛΟΠΟΥΛΟΣ

Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

© 2005 - All rights reserved

Περιεχόμενα

1	Εισαγωγή	1
1.1	Στόχοι	1
1.2	Κίνητρα και σημασία	2
1.2.1	Εργαλεία	4
1.3	Υφής προβλήματα	7
1.4	Συνεισφορά και πρόοδος	8
1.5	Διάρθρωση κειμένου	9
2	Οπτική Υφή και Μικροδομές Διαμόρφωσης	11
2.1	Ορισμοί και ιδιότητες	11
2.1.1	Μοντέλα υφής	13
2.2	Χωρικές συχνότητες και φίλτρα	14
2.2.1	Συστοιχίες Gabor φίλτρων	17
2.3	Μοντέλα διαμόρφωσης, ενεργειακοί τελεστές και αποδιαμόρφωση	18
2.3.1	1Δ ενεργειακός τελεστής	19
2.3.2	Ενεργειακός τελεστής για 2Δ σήματα	20
2.3.3	AM-FM σήματα και διαχωρισμός ενέργειας	21
2.3.4	Διακριτοί τελεστές και αποδιαμόρφωση εικόνων	22
2.3.5	Ανάλυση του 2Δ τελεστή σε θόρυβο	23
2.4	Τεχνικές ανάλυσης υφής	27
2.4.1	Στοχαστικά φράκταλ	27
2.4.2	Τυχαία πεδία	28
2.4.3	Διαχωρισμός υφής	32
3	Ενέργεια και Ανάλυση Υφής	35
3.1	Διαμορφώσεις υφής	35
3.1.1	Ενέργεια συνιστωσών υφής	36
3.2	Πολυκαναλική ανάλυση	36
3.2.1	Ζωνοπερατή αποδιαμόρφωση	37
3.2.2	Ανάλυση συνιστωσών κατά κανάλια	38
3.3	Ανάλυση κυρίαρχων συνιστωσών	42
3.4	Ενέργεια διαμόρφωσης και χαρακτηριστικά υφής	44
4	Κατάτμηση Υφής	47
4.1	Κατάτμηση με εξέλιξη καμπυλών και κυρίαρχες διαμόρφώσεις	47
4.2	Χαρακτηριστικά διαμόρφωσης	48
4.2.1	Η επιλογή καναλιού ως στατιστική απόφαση	48
4.3	Μεταβολική κατάτμηση χωρίς επισκόπηση	53
4.4	AM-FM ανταγωνισμός περιοχών	54

5	Ανάλυση Εδαφοτομών	57
5.1	Περιγραφή εδαφοτομών και ανάλυση εδάφους	57
5.2	Υφή εδαφοτομών, ανάλυση και κατάτμηση	58
5.2.1	Μορφολογική κατάτμηση	58
5.2.2	Διαχωρισμός γεωμετρίας και υφής με levelings	59
5.2.3	Συζευγμένη, πολυκριτριακή κατάτμηση εδαφοτομών	60
5.3	Αποτελέσματα κατάτμησης και συγκρίσεις	62
5.3.1	Βελτίωση με συγχώνευση περιοχών	64
5.3.2	Συγκρίσεις	65
5.4	Κατηγοριοποίηση υφής	67
5.5	Συμπεράσματα και αποτιμήσεις	71
6	Ανίχνευση Μικροδομών Φωνής	73
6.1	Φωνή σε θόρυβο	73
6.2	Ενεργειακοί τελεστές και πολυζωνικές διαμορφώσεις	74
6.3	Κίνητρα και ανίχνευση μέσω στατιστικών υποθέσεων	75
6.4	Χαρακτηριστικά διαμόρφωσης	77
6.5	Εντοπισμός συνόρων φωνής σε θόρυβο	78
6.6	Ανίχνευση φωνητικής δραστηριότητας	81
6.6.1	Ανίχνευση ενέργειας διαμόρφωσης	81
6.6.2	Αποτελέσματα και συγκρίσεις	82
6.7	Συμπεράσματα	85
7	Μελλοντική Έρευνα και Επεκτάσεις	87
7.1	Δεδομένα, Μερική πρόοδος	87
7.2	Ζητούμενα	89
7.3	Προς το τέλος	90
7.3.1	Υφή.....	90
7.3.2	Φωνή.....	90
7.4	Επίλογος	91
	Ορολογία	93
	Βιβλιογραφία	95
	Κατάλογος Δημοσιεύσεων του συγγραφέα	111

Κατάλογος Σχημάτων

2.1	Παραδείγματα υφής.....	13
2.2	Ζωνοπερατό Gabor φιλτράρισμα.....	17
2.3	Συστοιχία φίλτρων Gabor	19
2.4	Συνθετική υφή από στοχαστικά φράκταλ	28
2.5	Σύνθεση υφής με το multi-level logistic μοντέλο	31
2.6	Ιεραρχικά Μαρκοβιανά Τυχαία Πεδία (MRF)	31
2.7	Διαχωρισμός υφής $u + v$	33
3.1	Αποδιαμόρφωση και ανάλυση υφής ανά κανάλι	39
3.2	Πολυκαναλική ανάλυση και συνιστώσες διαμόρφωσης υφής	40
3.3	Διάγραμμα ανάλυσης κυρίαρχων συνιστωσών (DCA).....	42
3.4	DCA ανάλυση συνθετικού AM-FM	44
3.5	Σύγκριση DCA πλάτους και ενέργειας.....	45
3.6	Ενέργεια διαμόρφωσης υφής	46
4.1	Χαρακτηριστικά διαμόρφωσης (κυρίαρχων συνιστωσών) για κατάτμηση .	49
4.2	Πληροφορία υφής (κλίμακα, μέγεθος, προσανατολισμός) μέσω DCA....	49
4.3	Πιθανοτική ανίχνευση υφής ανά κλίμακα	51
4.4	Χαρακτηριστικά διαμόρφωσης βελτιωμένα με «μάσκα» υφής (I)	52
4.5	Χαρακτηριστικά διαμόρφωσης βελτιωμένα με «μάσκα» υφής (II).....	52
4.6	Κατάτμηση υφής ('cone' & 'scale')	54
4.7	Κατάτμηση υφής ('zebra')	55
4.8	Κατάτμηση υφής ('house')	55
4.9	Κατάτμηση υφής ('rice')	55
4.10	Κατάτμηση υφής ('rocks')	56
5.1	Πολυκλιμακωτά leveling «καρτούν» και $u + v$ διαχωρισμός.....	61
5.2	Ανάλυση υφής εδαφοτομών	63
5.3	Κατάτμηση εδαφοτομών με συνδυασμό κριτηρίων.....	64
5.4	Ομαδοποίηση με ένωση περιοχών.....	65
5.5	Σύγκριση μεθόδων κατάτμησης (I)	66
5.6	Κατάτμηση με συνδυασμό κριτηρίων (διαφορετικοί σημαδευτές)	67
5.7	Κατάτμηση και σύγκριση μεθόδων (II)	68
5.8	Λεπτομέρειες από τυπικές μικροϋφές χρώματος.....	70
6.1	Διάγραμμα εξαγωγής χαρακτηριστικών διαμόρφωσης	77
6.2	Ανάλυση και αποδιαμόρφωση φωνής ανά κανάλι	78
6.3	Χαρακτηριστικά ανάλυσης και ανίχνευσης φωνής	79
6.4	Παραδείγματα εντοπισμού συνόρων φωνής	80

6.5	Καμπύλες ανίχνευσης - ανοχής λάθους	81
6.6	Παραδείγματα ανίχνευσης φωνητικής δραστηριότητας	83
6.7	Καμπύλες ROC για αποτίμηση απόδοσης της ανίχνευσης φωνής	84
6.8	Νόρμες ολικών σφαλμάτων ανίχνευσης φωνητικών γεγονότων	85

Κατάλογος Πινάκων

5.1	Ποσοτική σύγκριση μεθόδων κατάτμησης (I)	67
5.2	Ποσοτική σύγκριση μεθόδων κατάτμησης (II)	68
5.3	Διαχωριστικότητα τάξεων υφής για Gabor ενεργειακά χαρακτηριστικά ..	70
6.1	Ορθή ανίχνευση ορίων φωνής (%) στη βάση NTIMIT	79
6.2	Ακρίβεια ανίχνευσης φωνητικής δραστηριότητας, Aurora 2	84
6.3	Ακρίβεια ανίχνευσης φωνητικής δραστηριότητας, Aurora 3	84

Περίληψη Προόδου Διατριβής

Στα κατώτερα επίπεδα συστημάτων όρασης η επεξεργασία και ανάλυση της οπτικής πληροφορίας γίνεται με την αναζήτηση χαρακτηριστικών και μικροδομών, σε πολλαπλές κλίμακες. Οι επιφάνειες σε φυσικές εικόνες είναι συνήθως ανομοιόμορφες και παρουσιάζουν μεταβολές έντασης φωτεινότητας, σε διαφορετικές κλίμακες, σχηματίζοντας επαναλαμβανόμενα πρότυπα. Η υφή είναι αποτέλεσμα των φυσικών ιδιοτήτων μιας επιφάνειας όπως είναι η τραχύτητα, ο προσανατολισμός, το βάθος, το χρώμα, η ανάκλαση, εξελικτικών διαδικασιών ή μεταβολής συνθηκών περιβάλλοντος. Η παρούσα έρευνα στοχεύει στην ανάδειξη και ανάλυση μικροδομών της υφής, για προβλήματα όπως η εξαγωγή χαρακτηριστικών, η κατάτμηση, η κατηγοριοποίηση και η σύνθεση. Ερευνώνται και προτείνονται προηγμένες τεχνικές μοντελοποίησης και ανάλυσης σε παράλληλες κατευθύνσεις: i) Μέθοδοι βασισμένες σε μοντέλα χωρικών φίλτρων και τοπικών διαμορφώσεων, ii) Μοντέλα στοχαστικών φράκταλ και κλασματικοί θόρυβοι, και iii) Μεταβολικές και μη-γραμμικές μέθοδοι διαχωρισμού εικόνων σε απλοποιημένες εκδοχές τους και συνιστώσες υφής. Κοινός παρονομαστής είναι η χρήση μοντέλων διαμορφώσεων και οι μη-γραμμικές τοπικές ενεργειακές μετρήσεις. Τέτοιες τεχνικές εφαρμόζονται ειδικότερα για την ανάλυση και κατάτμηση εικόνων εδαφοτομών, που χαρακτηρίζονται από πολύπλοκη δομή. Η επέκταση και η αναζήτηση κοινών μοντέλων για υφή και παραπλήσια φαινόμενα σε ακουστικά σήματα, οδηγεί σε εκδοχές των μεθόδων για ανίχνευση και ανάλυση μικροδομών φωνής.

Λέξεις Κλειδιά: οπτική υφή, μικροδομές, διαμορφώσεις, πολυκαναλική ανάλυση, συστοιχίες φίλτρων, ενεργειακοί τελεστές, ανίχνευση, ανάλυση, κατάτμηση, διαχωρισμός εικόνων, αναπαράσταση εικόνων, μη-γραμμικά μοντέλα, στοχαστικά μοντέλα, μεταβολικές τεχνικές, ανάλυση εδαφοτομών, ανίχνευση φωνής.

Abstract

At the higher levels of vision systems, the processing and analysis of visual information is accomplished by seeking features and structures in multiple scales. Surfaces in natural scenes are usually non-uniform, characterized by intensity variations in various scales that create repetitive patterns. Texture, is a result of natural surface properties, like coarseness, orientation, depth, color, reflection, evolutionary processes or change in environmental conditions, like shading or lighting. This research aims at detecting and analyzing texture microstructures, for problems like texture feature extraction, segmentation, classification and synthesis. Advanced modeling and analysis methods are examined and proposed in parallel directions: i) Methods based in spatial filtering and local modulation detection, ii) Stochastic models and fractional noises and iii) Variational and non-linear methods of image decomposition in simplified versions and texture components. The common denominator is the use of modulation modeling and non-linear local energy tracking. Such techniques are applied for the specific task of analysis and segmentation of soil-section images, that posses a complex structure. The extension of texture models for analogous phenomena in acoustic signals, has led to versions of the developed methods appropriate for detection of speech microstructures.

Keywords: visual texture, microstructures, modulations, multiband analysis, filter-banks, energy operators, feature extraction, detection, segmentation, image decomposition, image representation, non-linear models, variational techniques, soilsection analysis, speech event detection.

Κατάλογος Συντμήσεων

MΔE	:	Μερική Διαφορική Εξίσωση
AM	:	Amplitude Modulation (Διαμόρφωση Πλάτους)
ADCA	:	Amplitude-based Dominant Components Analysis
CCA	:	Channelized Components Analysis (Ανάλυση Συνιστωσών κατά κανάλια)
DCA	:	Dominant Components Analysis (Ανάλυση Κυρίαρχων Συνιστωσών)
DESA	:	Discrete Energy Separation Algorithm (Διακριτός Αλγόριθμος Διαχωρισμού Ενέργειας)
EDCA	:	(Teager) Energy-based Dominant Components Analysis
ESA	:	Energy Separation Algorithm (Αλγόριθμος Διαχωρισμού Ενέργειας)
fBm	:	Fractional Brownian Motion (Κλασματική Κίνηση Brown)
FM	:	Frequency Modulation (Διαμόρφωση Συχνότητας)
GAC	:	Geodesic Active Contours (Γεωδαιτικά Ενεργά Περιγράμματα)
GAR	:	Geodesic Active Regions (Γεωδαιτικές Ενεργές Περιοχές)
GRF	:	Gibbs Random Fields (Gibbs Τυχαία Πεδία)
MCMC	:	Markov Chain Monte Carlo
MDA	:	Multiband Demodulation Analysis (Πολυκαναλική Ανάλυση Αποδιαμόρφωσης)
MDL	:	Minimum Description Length
MLE	:	Maximum Likelihood Estimates (Εκτιμήσεις Μέγιστης Πιθανοφάνειας)
MRF	:	Markov Random Fields (Μαρκοβιανά Τυχαία Πεδία)
PDEs	:	Partial Differential Equations (Μερικές Διαφορικές Εξισώσεις)
PDF	:	Probability Density Function (Συνάρτηση Πυκνότητας Πιθανότητας)
RC	:	Region Competition
RSST	:	Recursive Shortest Spanning Tree (split and merge)
TV	:	Total Variation
WSS	:	Wide Sense Stationary (Στατική με την ευρεία έννοια)

ZR	:	average Zero-crossings Rate (Μέσος ρυθμός μεταβάσεων από το μηδέν)
LTED	:	Long-term Multiband Teager Energy Divergence (Μακροπρόθεσμη Απόκλιση Πολυκαναλικής Teager Ενέργειας)
LTSD	:	Long Term Spectral Divergence (Μακροπρόθεσμη Φασματική Απόκλιση)
mAA	:	mean Average Amplitude (Μέσο Απόλυτο Πλάτος)
mSA	:	mean Squared Amplitude (Μέσο Τετραγωνικό Πλάτος)
MTE	:	mean Multiband Teager Energy (Πολυκαναλική μέση Teager Ενέργεια)
MTED	:	Multiband Teager Energy Divergence (Απόκλιση Πολυκαναλικής Teager Ενέργειας)
MIA	:	mean Multiband Instant Amplitude (Πολυκαναλικό μέσο Στιγμιαίο Πλάτος)
MIF	:	mean Multiband Instant Frequency (Πολυκαναλική μέση Στιγμιαία Συχνότητα)
PTE	:	Power-spectrum based Teager Energy
ROC	:	Receiver Operating Characteristics
TKEO	:	Teager-Kaiser Energy Operator
SNR	:	Signal to Noise Ratio (Σηματοθορυβικός λόγος)
STE	:	Short-time Teager Energy
VAD	:	Voice Activity Detection (Ανίχνευση Δραστηριότητας Φωνής)

Κεφάλαιο 1

Εισαγωγή

Η αντίληψη του φυσικού κόσμου στηρίζεται στην ικανότητα του ανθρώπινου, υποκειμενικά ιδανικού, συστήματος αισθήσεων να ερμηνεύει πληροφορία κλιμακωτά. Η παρούσα δουλειά έχει να κάνει με αναζήτηση δομής και εξαγωγή πληροφορίας σε συστήματα τεχνητής αντίληψης, κυρίως όρασης, όπου η έννοια της δομής σε μικρές κλίμακες απαντάται ως *οπτική υφή* (visual texture) στις δισδιάστατες ψηφιακές αναπαραστάσεις φυσικών σχημάτων, αλλά και επικοινωνίας λόγου, όπου μονοδιάστατες μη-γραμμικές χρονικές μεταβολές χαρακτηρίζουν τη *μικροδομή φωνής* (speech microstructure). Μια ολοκληρωμένη περιγραφή ή ερμηνεία που ανακύπτει από τέτοια συστήματα οφείλει να συνυπολογίζει μικρομεταβολές και χαρακτηριστικά σε κλίμακες και διαστάσεις πέρα από τις φαινομενικά επαρκείς.

1.1 Στόχοι

Αυτό που γίνεται άμεσα αντιληπτό με την αφή σε αντικείμενα και στοιχεία του κόσμου προσλαμβάνεται διαφορετικά μέσω της όρασης. Και ενώ μπορούν να γίνουν υποθέσεις ως προς το είδος ή την επιφάνεια και τα χαρακτηριστικά του ορατού αντικειμένου, το μεγαλύτερο μέρος της πληροφορίας υψής γίνεται εκμεταλλεύσιμο υποσυνείδητα, σε ένα ευρύτερο πλαίσιο ανάλυσης σύνθετης οπτικής πληροφορίας (χρώμα, σχήμα, σχετική θέση, φωτεινότητα κ.λ.π.). Το κύριο τμήμα της παρούσας έρευνας έχει να κάνει με την ανίχνευση και ανάλυση υψής σε τεχνητά συστήματα όρασης (artificial vision). Σε ένα δεύτερο μικρότερο σε έκταση στάδιο ιδέες και μέθοδοι γενικεύονται σε αισθητήρια σήματα όπως είναι η φωνή ή βιολογικές μετρήσεις, υπό ένα κοινό πρίσμα αναζήτησης μικροδομών και κρυφής γεωμετρίας σε στοχαστικά σήματα.

Εξ ορισμού η οπτική υφή εικόνων αποτελεί δομή με χαρακτηριστικά πέραν των άμεσα προσλαμβανόμενων. Στην ουσία περιλαμβάνει μικρομεταβολές πάνω από τις μακρομεταβολές μεγεθών όπως η φωτεινότητα, το χρώμα ή το σχήμα. Πέρα από την υφή, και ανεξάρτητα από τον ασαφή ορισμό της, ενδιαφέρον υπάρχει για μικρομεταβολές όπως διαμορφώσεις, ταλαντώσεις, φράκταλ δομές, στατιστική αυτοομοιότητα σε πολλαπλές κλίμακες κ.α.

Με γνώμονα την κατανόηση και εφαρμογή της μελέτης των μικρομεταβολών πολυδιάστατων σημάτων σε συστήματα όρασης, ενδιαφέρει ο ορισμός της υψής. Να οριστεί η οπτική υφή για ψηφιοποιημένες αναπαραστάσεις του φυσικού κόσμου αλλά και να γενικευθεί η έννοια της για μονοδιάστατα σήματα. Θα μπορούσε κανείς διαισθητικά να μιλήσει για "άκουστική" υφή σημάτων ήχου ή ομιλίας όπως χρησιμοποιείται περιφραστικά στη μουσική η υφή ή το ηχόχρωμα;

Παράλληλα αλλά και συμπληρωματικά στόχος είναι η αναζήτηση και χρήση εύρωστων μοντέλων για την εύρεση και την περιγραφή υφής. Το θέμα της επιλογής του κατάλληλου μοντέλου έχει απασχολήσει εκτενώς τη σχεική βιβλιογραφία με επιμέρους ζητήματα την επιλογή κατάλληλης βάσης περιγραφής, τον ορισμό των σχετικών παραμέτρων, την εκτίμηση τους καθώς και τρόπους σύζευξης ιδεών από διαφορετικές προσεγγίσεις. Η εφαρμογή ενός κατάλληλου μοντέλου μπορεί να γίνει μόνο εφόσον υπάρχει υφή και μπορεί να ανιχνευθεί. Μια δεύτερη επομένως κατεύθυνση είναι η *ανίχνευση* τέτοιων δομών, που θα μπορούσαν να χαρακτηριστούν υφή, με κατάλληλα εργαλεία που ταιριάζουν στις ιδιότητες τους και ο καθορισμός των συνθηκών ώστε κάτι τέτοιο να είναι εφικτό και να έχει νόημα. Τέλος στόχος είναι η *ανάλυση* και η *εξαγωγή* χαρακτηριστικών για περιγραφή, μέτρηση, κατηγοριοποίηση, διάκριση, αναγνώριση, κατάτμηση υφής κ.α. μέσα από ανεπτυγμένα εργαλεία και καινοτόμες μεθόδους, κατάλληλες για τη φύση της πληροφορίας υφής.

Στην προσπάθεια να γίνει περισσότερο απτή αυτή η πληροφορία, αξιοποιείται ένα σύνολο μη-γραμμικά και στατιστικά εργαλεία, τα οποία διαισθητικά και μαθηματικά ταιριάζουν σε τέτοιες δομές. Το ερευνητικό πλαίσιο περιλαμβάνει γενικευμένους μη-γραμμικούς, πολυδιάστατους ενεργειακούς τελεστές, μοντελοποίηση σύνθετων σημάτων ως υπέρθεση διαμορφωμένων ημιτονοειδών, πολυκαναλική ανάλυση μέσω ζωνοπερατού φίλτραρίσματος σε πολλαπλές ζώνες ξεχωριστού συχνοτικού ενδιαφέροντος και εκτίμηση των επιμέρους ισχυρών συντελεστών που απαρτίζουν την υφή. Στο περιθώριο εξετάζονται μέθοδοι σχετικές με την ύπαρξη στατιστικής αυτοομοιότητας και ποσοτικοποίησης φράκταλ δομής, στατιστική μοντελοποίηση και εκτίμηση με μαρκοβιανές διαδικασίες, λογισμό μεταβολών και ελαχιστοποίηση ενεργειακών.

Η εφαρμογή των προτεινόμενων μεθόδων για ανάλυση υφής και μικροδομών επιδεικνύεται σε διαφορετικές κατηγορίες εικόνων και «αισθητήριων» σημάτων. Πέρα από φυσικές εικόνες, η ανάλυση εφαρμόζεται κύρια σε δεδομένα από μια βάση εικόνων λεπτών εδαφοτομών με απώτερο σκοπό την κατηγοριοποίηση και εκτίμηση της βιολογικής ποιότητας εδαφών. Παράλληλα εξετάζεται η εφαρμογή σε περιβαλλοντολογικές και γεωλογικές εικόνες, δεδομένα τηλεπισκόπησης (remote sensing), βιολογικές μετρήσεις και 2Δ απεικονίσεις. Από τη σκοπιά της γενίκευσης ή μάλλον της εξειδίκευσης σε 1Δ δεδομένα εξετάζεται η εφαρμογή των μεθόδων για ανίχνευση και ανάλυση φωνής καθώς και χρονο-συχνοτική ανάλυση σημάτων.

1.2 Κίνητρα και σημασία

Η ανάλυση υφής ως εξειδικευμένο κομμάτι ενός συστήματος τεχνητής, υπολογιστικής όρασης αποτελεί ενεργό ερευνητικό πεδίο στο οποίο συναντιόνται ετερογενή γνωστικά αντικείμενα και διαφορετικές επιστημονικές περιοχές, όπως η ψυχολογία, η νευροφυσιολογία, οι γνωσιακές επιστήμες, η επεξεργασία εικόνων και μηχανική μάθηση, η στατιστική φυσική κ.α. Η σημασία της αυτόματης αντίληψης της υφής έγκειται αφενός στις ψυχοφυσικές ενδείξεις γύρω από την ανθρώπινη αντίληψη και αφετέρου στις άμεσες και έμμεσες εφαρμογές σε θέματα μηχανικής όρασης [189].

Η επεξεργασία της υφής είναι άρρηκτα συνδεδεμένη με την οπτική αντίληψη (visual perception). Έτσι για παράδειγμα, σχηματισμοί ή πρότυπα υφής (textural patterns) είναι τα κεντρικά ζητήματα στην ψυχολογία Gestalt [203], στη θεωρία *αρχικής όρασης* (early vision theory) και στο *πρωταρχικό σκίτσο* (primal sketch) του Marr [129]. Από μια τελείως φιλοσοφική πλευρά έχει σημασία η κατανόηση των επιμέρους διαδικασιών που περιλαμβάνει η αντίληψη της υφής, για την κατανόηση των διεργασιών όρασης,

αντίληψης και συνείδησης. Έννοιες και συζητήσεις γύρω από την υφή ξεκινάνε ακόμη και πριν από την εισαγωγή της υπολογιστικής συλλογιστικής [Gibson 1950] [63], ενώ με την ανάπτυξη των πρώτων μηχανιστικών συστημάτων όρασης η υφή ορίζεται ευριστικά ανάλογα με τις ανάγκες [68]. Ακριβής και αποδεκτός ορισμός της φυσικής υφής επιφανειών και της οπτικής υφής δεν υπάρχει, με τις περισσότερες προσπάθειες να επικεντρώνονται στον καθορισμό ενός ενιαίου πλαισίου επεξεργασίας και ανάλυσης [216].

Ένα σύνολο ψυχοφυσικών (psychophysical) μελετών (ανθρώπινης συμπεριφοράς) προέκυψαν με κίνητρο την ανάγκη να ερμηνευθούν και να αναγνωριστούν εκείνες οι οπτικές διεργασίες που ευθύνονται για την αναγνώριση αντικειμένων και σχημάτων με βάση την υφή και μόνο. Ο Beck [8] θεωρεί ότι το ανθρώπινο σύστημα όρασης δεν αναζητά πρότυπα και σύνθετες φόρμες υφής αλλά στατιστικά όπως το μήκος, ο προσανατολισμός κ.α. Ο Julesz χώρισε τον τρόπο λειτουργίας της οπτικής αντίληψης σε δύο διαφορετικά μοτίβα (preattentive & attentive vision), βασιζόμενος κύρια σε πειράματα και παρατηρήσεις σχετικά με αναγνώριση υφής από παρατηρητές [85,86,89]. Προχώρησε στη μελέτη συνθετικών υφών μικροπροτύπων [29,90] και τη δημιουργία υποθέσεων σχετικά με τα απαραίτητα χαρακτηριστικά για την αντίληψη υφής, στη διατύπωση ερωτήσεων σχετικά με τη σημασία της σε εικόνες με κοινά τα υπόλοιπα χαρακτηριστικά (χρώμα, αντίθεση, φωτεινότητα) και στη διαμόρφωση της θεωρίας των τεξτονίων (textures) [86] ως εξήγηση της αυθόρμητης διάκρισης υφής, χωρίς σχολαστική επισκόπηση. Τα *τεξτόνια* (blobs, terminators, crossings, color) θεωρήθηκαν οι βασικές μονάδες της preattentive αντίληψης υφής [89]. Ανάλογα πειράματα πραγματοποιήθηκαν εκτός από την πυκνότητα ή τη συγκέντρωση, σχετικά με το μέγεθος [10] ή την αναγνώριση συνόρων υφής [199,200] παράλληλα με διαπιστώσεις όπως ότι η αντίληψη συγκεκριμένων προτύπων εξαρτάται πρωτίστως από ανάλυση χωρικών συχνοτήτων και ότι ο αυθόρμητος διαχωρισμός περιοχών και συχετίζεται με το συχνοτικό τους περιεχόμενο [9].

Παράλληλες ενδείξεις για την αντίληψη υφής προέρχονται και από πρώιμες νευροφυσιολογικές μελέτες [89]. Προτάθηκε η ύπαρξη αναλυτών στο οπτικό σύστημα για την ανίχνευση τεξτονίων [28], ενώ διαπιστώθηκε ότι νευρικά κύτταρα στον οπτικό φλοιό [79] συμπεριφέρονται δεκτικά σε οπτικά χαρακτηριστικά που κατά σύμβαση σημαίνουν υφή. Ο ρόλος τέτοιων ενδείξεων ως προς τη σημασία της υφής στην όραση τονίζεται επίσης από την χρήση ανιχνευτών στοιχειωδών χαρακτηριστικών μιας εικόνας (elongated blobs, bars, edges, terminators) σε ένα από τα πρώτα ολοκληρωμένα μοντέλα αρχικής όρασης. Το primal sketch του Marr [129] θεωρητικά προτάθηκε ως ένα πρώτο επίπεδο εσωτερικής αναπαράστασης της οπτικής πληροφορίας με όρους τέτοιων στοιχειωδών χαρακτηριστικών (image primitives) και φέρει πολλούς σχηματισμούς υφής να περιέχουν ιεραρχικά οργανωμένες σημασιολογικές δομές.

Ως διεργασία χαμηλού επιπέδου (low-level processing) του συστήματος όρασης, η ανάλυση υφής συμμετέχει και συμπεριλαμβάνεται σε πολλά από τα πεδία εφαρμογών της υπολογιστικής όρασης, είτε άμεσα, σε εφαρμογές προσανατολισμένες στην υφή είτε έμμεσα ως τμήμα ενός πολυ-συμβολικού συστήματος ανάλυσης οπτικών αναπαραστάσεων. Η παρουσία υφής σε μια φυσική σχηνή μπορεί να εκμεταλλευθεί εναλλακτικά: (1) Τα χαρακτηριστικά μιας μοναδικής υφής μπορεί να χρησιμοποιηθούν για να αναγνωριστεί το υλικό ή η δομή μιας επιφάνειας, όπως π.χ. το να ξεχωρίζει κανείς το ξύλο από την άμμο (texture appearance). (2) Οι ιδιαιτερότητες στα χαρακτηριστικά ενός συνόλου εικόνων υφής ή επιφανειών, επιτρέπουν την ταξινόμηση τους και την κατηγοριοποίηση με βάση ένα σύνολο κανόνων (texture classification). (3) Ανίχνευση και εντοπισμός ακμών, βασικό στάδιο του προβλήματος *πρώτου πλάνου-φόντου* (figure-ground), σε ακμές που χαρακτηρίζονται από αλλαγές στην υφή (σε συνδυασμό

με μεταβολές σε άλλες ιδιότητες μιας επιφάνειας όπως η ένταση φωτεινότητας και το χρώμα [131]). (4) Η διάκριση και ο διαχωρισμός υφής (texture segregation) μπορεί να αποτελεί σημαντική συνιστώσα για την αναγνώριση αντικειμένων. (5) Ιδιότητες της υφής σε μια εικόνα μπορεί να χρησιμοποιηθούν για την εξαγωγή συμπερασμάτων της τρισδιάστατης διάταξης αντικειμένων, του βάθους, του σχήματος [2, 6, 184] (shape from texture [63]) ή του προσανατολισμού [111, 206]. (6) Κατάτμηση με βάση την υφή (texture segmentation) [165] αλλά και στα πλαίσια σύνθετης κατάτμησης εικόνων με πολλαπλά κριτήρια (multi-cue image segmentation) [112, 113].

Η ανάλυση και η σύνθεση υφής απαντάται σε σχεδιασμό και γραφικά με υπολογιστές [77, 205], ψηφιακές τέχνες και εφαρμογές εικονικής πραγματικότητας για τη σύνθεση ρεαλιστικών επιφανειών. Η επιδιόρθωση μερικώς κατεστραμμένων εικόνων και η «συμπλήρωση» ή αντικατάσταση χαμένης πληροφορίας στατιστικά (texture inpainting, completion, filling-ing) [1, 12, 14, 53] αναπτύχθηκε πρόσφατα στα πλαίσια μοντέλων διαχωρισμού υφής από γεωμετρία και θόρυβο. Παράλληλα η περιγραφή και η εξαγωγή χαρακτηριστικών υφής, ως στοιχείο φυσικών σκηνών, περιλαμβάνεται σε ένα πλήθος σύνθετων εφαρμογών όρασης όπως η αναζήτηση εικόνων με ταίριασμα χαρακτηριστικών, η τηλεπισκόπηση (remote-sensing), οι βιοϊατρικές και οικολογικές - περιβαλλοντολογικές εφαρμογές, η αυτόματη επίβλεψη διεργασιών, η αναγνώριση κίνησης και προσώπων κ.α.

1.2.1 Εργαλεία

Η αναπαράσταση υφής μέσω επιλογής και κατασκευής κατάλληλων μαθηματικών μοντέλων περιγραφής έχει να κάνει με τις ιδιότητες και τις μεταβολές μιας εικόνας που γίνονται αντιληπτές ή και μεταφράζονται ως διαφορές στην φυσική υφή των αντίστοιχων επιφανειών. Κάποιες γενικώς παραδεκτές αρχές σχετικά με την υφή συνοψίζονται στα εξής [106, 189]: α) Η υφή είναι ιδιότητα περιοχών σε μια εικόνα και ορίζεται τοπικά σε χωρικές γειτονιές, β) ορίζεται στατιστικά και περιλαμβάνει κατανομές γκριζών ή έγχρωμων επιπέδων φωτεινότητας, γ) σε διαφορετικές κλίμακες ή επίπεδα ευκρίνειας (resolution) και δ) ο αριθμός και το μέγεθος των επιμέρους στοιχειωδών χαρακτηριστικών καθορίζει το αν οι περιοχές γίνονται αντιληπτές σαν υφή, γεωμετρικές δομές ή αντικείμενα. Σαν αποτέλεσμα τα μοντέλα και εργαλεία για την αναπαράσταση οπτικής υφής προσανατολίζονται, χωρίς όμως να περιορίζονται, ανάλογα.

Η επιλογή στατιστικών μεθόδων και εργαλείων ανέκυψε ως φυσική και διαισθητική επιλογή από πολύ νωρίς. Το πρώτο γενικό μοντέλο υφής προτάθηκε από τον Julesz [85] και στηρίχθηκε σε στατιστικά τάξης k , δηλαδή στατιστικά τιμών k -ομάδων, ενώ πολλές από τις πρώτες προσεγγίσεις βασίστηκαν σε μετρήσεις και στατιστικά, αρκετά διαδεδομένα αργότερα, όπως οι ροπές (moments), περιθώρια ιστογράμματα (marginal histograms) και από κοινού ιστογράμματα (joint histograms) των εξόδων γραμμικών φίλτρων [211]. Εξάλλου το ανθρώπινο σύστημα πιστεύεται ότι εξάγει πληροφορία από στατιστικά με $k \geq 2$ και γι'αυτό το σκοπό χρησιμοποιήθηκε μια σειρά χαρακτηριστικών πινάκων «συνύπαρξης» (co-occurrence matrices) (εντροπία, ενέργεια, συσχέτιση κ.α.) [68]. Στο στατιστικό πλαίσιο εντάχθηκαν μεταξύ άλλων διάφοροι τύποι, μεγέθη και προσανατολισμοί τεξτονίων [112, 215] και κλίκες (cliques) τυχαίων πεδίων.

Ο Besag [15, 16] θεμελίωσε τη θεωρία των Μαρκοβιανών τυχαίων πεδίων (Markov random fields), με ιδέες από τη στατιστική μηχανική. Τέτοια πεδία χαρακτηρίζουν πρότυπα υφής ως χωρικά εξαρτώμενα σε ένα πεπερασμένο πλέγμα με στατιστικό τρόπο και χρησιμοποιήθηκαν συστηματικά για την κατασκευή διακριτών μοντέλων υφής

εικόνων [41, 48, 62, 120, 204]. Η στατιστικά θεώρηση επιτρέπει την εκτίμηση του σχετικά μικρού αριθμού παραμέτρων τέτοιων πιθανοτικών μοντέλων από παρατηρήσεις υψής αλλά και τη σύνθεση εικόνων με δειγματοληψία από τις εν λόγω κατανομές. Διάφορες παραλλαγές των αρχικών μοντέλων προτάθηκαν εισάγοντας περιορισμούς γεωμετρικής δομής [176] ή σχήματος [213]. Στο πνεύμα μιας ενοποιημένης και μαθηματικά βάσιμης θεώρησης της υψής, ερευνήθηκαν συνδυασμοί μοντέλων τυχαίων πεδίων με χαρακτηριστικά στις εξόδους γραμμικών φίλτρων [216] και ορισμοί της υψής με βάση τα στατιστικά του Julesz [208] και ισοδυναμίες με κατανομές Gibbs.

Τεχνικές που βασίζονται σε διανύσματα χαρακτηριστικών στις εξόδους γραμμικών φίλτρων, υποστηρίζονται από μελέτες φυσιολογίας και νευροφυσιολογίας βιολογικών συστημάτων όρασης, από ψυχοφυσικές ενδείξεις για συχνοτική ανάλυση κατά την οπτική αντίληψη και από την ευελιξία και περιγραφικότητα τέτοιων μοντέλων. Η γενική αρχή πολυκαναλικής ή πολυζωνικής ανάλυσης μέσω φίλτρων με επιλεκτικότητα στον προσανατολισμό και στην κλίμακα χωρικών συχνοτήτων ακολουθεί την παράλληλη επεξεργασία της οπτικής πληροφορίας στα πρώτα στάδια όρασης [202]. Η αποτύπωση στον αμφιβληστροειδή (retinal image) αποσυνθέτει μια εικόνα σε πολλαπλές φιλτραρισμένες εκδοχές της, η κάθε μια από τις οποίες περιέχει μεταβολές έντασης σε ένα περιορισμένο, στενό εύρος συχνοτήτων [49, 114]. Επιπλέον τα δεκτικά πεδία (receptive fields), η χωρική απόκριση δηλαδή, των κυττάρων-νευρώνων (simple cells) στον οπτικό φλοιό (primary visual cortex), μοντελοποιούνται από ζωνοπερατά φίλτρα βέλτιστης χωρικής και συχνοτικής τοποθέτησης [45, 128, 192]. Το σύνολο των αποτελεσμάτων πειραματικής φυσιολογίας των Hubel και Wiesel [79, 81] περιγράφει τα προφίλ και τις ιδιότητες τέτοιων φίλτρων, διακρίνοντας τα σε απλά και σύνθετα, μελετώντας τις αποκρίσεις νευρώνων σε στοιχειώδη ερεθίσματα. Επιπλέον ψυχοφυσιολογικά πειράματα απόκρισης σε ερεθίσματα ανέδειξαν μηχανισμούς κατανομής της ενέργειας σε πολλαπλά ιστροπικά κανάλια [10] και τοπικούς ανιχνευτές διαφορών σε στοιχειώδη χαρακτηριστικά [79, 87, 129, 130]. Οι Julesz, Bergen [89] παρατήρησαν την ισοδυναμία συνδυασμού φίλτρων με ανιχνευτές τεξτονίων και επεσήμαναν την αναγκαιότητα πολλαπλών φίλτρων για την καλύτερη διάκριση υψής [88].

Το πολυκαναλικό μοντέλο είναι παράλληλα ελκυστικό και ευέλικτο καθώς επιτρέπει την αναζήτηση διαφορετικών μεγεθών και προσανατολισμών που κυριαρχούν δομικά σε μια υφή, ενώ καλύπτει και την ανάγκη για κλιμακωτή, πολλαπλής ευκρίνειας ανάλυση. Τα χαρακτηριστικά υψής που εξάγονται από τις αποκρίσεις των διαφόρων φίλτρων είναι συνήθως στατιστικά γκριζών τιμών από μη-γραμμικούς τελεστές ή κατανομές πιθανότητας και έχουν μεγαλύτερη περιγραφική ικανότητα από στατιστικά μοντέλα και μετρήσεις. Το μοντέλο χρησιμοποιήθηκε σε διάφορες εκδοχές από το 1980 και μετά, με την ίδια πάντα φιλοσοφία των χωρικά και συχνοτικά επιλεκτικών φίλτρων για την αναπαράσταση και την ανάλυση υψής εικόνων π.χ. πυραμίδες Laplacian of Gaussians, συναρτήσεις Gabor, ορθοκανονικές βάσεις, wavelets, steerable pyramids κ.α. [9, 20, 27, 35, 52, 58, 66, 84, 114, 116, 156, 157, 163, 166, 192, 216].

Η μέτρηση χαρακτηριστικών υψής γίνεται στις εξόδους μιας συστοιχίας φίλτρων (filterbank) και αποσκοπεί στην κατασκευή διανυσμάτων περιγραφής και ανάλυσης περιοχών με υφή, με όσο το δυνατόν πιο εύρωστο τρόπο και μικρό αριθμό παραμέτρων ή διάστασης. Οι διαφορετικές προσεγγίσεις για την εξαγωγή πολυκαναλικών χαρακτηριστικών περιλαμβάνουν είτε υπολογισμό γενικευμένων ενεργειών, είτε στατιστικά κατανομών γκριζών τιμών. Στην πρώτη περίπτωση, ακολουθώντας πάντα υπολογιστικά μοντέλα βιολογικής όρασης, η γραμμικότητα του μοντέλου που επιβάλλεται από τις αποκρίσεις των φίλτρων είναι ανεπαρκής για διάκριση και αντίθετη με ψυχοφυσικές

ενδείξεις [114]. Αυτή η ανάγκη για μη-γραμμική επεξεργασία μετά το φιλτράρισμα οδήγησε σε τελεστές ανόρθωσης μισού [114] και πλήρους κύματος [10, 18] (τετραγωνισμός και απόλυτη τιμή), ενέργειες ζευγαριών φίλτρων με κάθετη διαφορά φάσης (quadrature pair) όπως π.χ. ημιτονικά και συνημιτονικά Gabor [58, 192], περιβάλλουσα και φάση της απόκρισης μιγαδικών φίλτρων [20], τελεστές με βιολογικές ερμηνείες όπως ανίχνευση blobs [84], bars [155] ή περιοδικότητας (σε κατευθυντικές δομές) [100] και πολυφασματικά opponent χαρακτηριστικά [83]. Στη δεύτερη περίπτωση περιλαμβάνεται ο χαρακτηρισμός περιοχών υψής με τη βοήθεια ιστογραμμάτων των εξόδων των φίλτρων [77, 216], με τα κέντρα των τάξεων που προκύπτουν από clustering των διανυσμάτων [113] και με στατιστικούς περιορισμούς των συντελεστών της πολυκαναλικής αναπαράστασης [157].

Η εμφάνιση της υψής ως μακροσκοπικές, τοπολογικές μεταβολές της επιφάνειας, έχει επαληθευθεί και πειραματικά [129, 199]. Τόσο η πυκνότητα, όσο και ο προσανατολισμός των διαφορετικών δομών-προτύπων παίζουν θεμελιώδη ρόλο στην αντίληψη υψής. Το γεγονός αυτό υπονοεί μια αποκλειστική σχέση ανάμεσα στην υφή και σε τοπικές χωρικές συχνότητες. Επιπλέον είναι γνωστό ότι οι ηλεκτροχημικές αποκρίσεις των κυττάρων του οπτικού φλοιού μεταδίδουν πληροφορία αντίθεσης φωτεινότητας και χωρικής φάσης. Αυτό προκαλεί την ερμηνεία της υψής με όρους ομαλά μεταβαλλόμενων *διαμορφώσεων* σε επιλεκτικά ως προς τη συχνότητα και τον προσανατολισμό κανάλια [72].

Ένα πρόσφατο μοντέλο αναπαράστασης υψής με πολλαπλές συνιστώσες ανάγει τις μεταβολές μιας εικόνας σε ένα σύνολο τοπικών διαμορφώσεων και σε τοπικά αναδυόμενες συχνότητες. Οι περισσότερες οπτικές υφές φυσικών επιφανειών, χαρακτηρίζονται από τέτοιες τοπικές συχνότητες [21] και μπορούν να περιγραφούν ως υπέρθεση συνιστωσών διαμορφωμένων κατά πλάτος και φάση [123]. Πρόκειται για τοπικά μη-στατικά ημίτονα, των οποίων το πλάτος καθορίζει τις μεταβολές της αντίθεσης, ενώ το διάνυσμα της στιγμιαίας μεταβαλλόμενης φάσης τις τοπικά αναδυόμενες ταλαντώσεις υψής. Η μοντελοποίηση φυσικών και συνθετικών εικόνων υψής με τέτοια AM-FM μοντέλα οδήγησε σχεδόν φυσικά στην αναζήτηση και δημιουργία εύρωστων αλγορίθμων για την εξαγωγή των στιγμιαίων σημάτων διαμόρφωσης από το σύνολο των συνιστωσών [73, 123]. Ένας τέτοιος απλός, αποτελεσματικός και ακριβής αλγόριθμος, βασίστηκε σε ένα μη-γραμμικό πολυδιάστατο ενεργειακό τελεστή [123], γενίκευση του 1Δ ενεργειακού τελεστή Teager-Kaiser [91] που έχει εφαρμοστεί με επιτυχία για ανάλυση σημάτων φωνής [125]. Για την απομόνωση των επιμέρους ζωνοπερατών συνιστωσών η ανάλυση και αποδιαμόρφωση εφαρμόζεται στην έξοδο συστοιχίας Gabor φίλτρων [74], που επιβάλλουν μια επιπλέον ομαλότητα και ανοχή στο θόρυβο στα ενεργειακά και στιγμιαία σήματα. Έτσι η εξαγωγή διανυσμάτων χαρακτηριστικών υψής ανάγεται σε ανάλυση διαμορφώσεων εικόνων, σε πολλαπλά κανάλια, κατά την οποία ο μη-γραμμικός τελεστής εντοπίζει την ενέργεια του συστήματος που θα μπορούσε να παράγει χωρικές 2Δ ταλαντώσεις και οι παράμετροι της υψής είναι τα χωρικά διαμορφωμένα στιγμιαία μεταβαλλόμενα σήματα πλάτους και φάσης.

Μια κατηγορία μοντέλων υψής, η οποία εντάσσεται τόσο στο στατιστικό όσο και στο μη-γραμμικό πλαίσιο είναι τα μοντέλα φράκταλ (fractal). Σε εφαρμογές υψής, ανάλυσης μικροδομών και σχήματος τέτοια μοντέλα ποσοτικοποιούν και περιγράφουν με μικρό αριθμό παραμέτρων την τραχύτητα μιας επιφάνειας, την πολυπλοκότητα μιας καμπύλης ή τη γεωμετρική πολυπλοκότητα και στατιστική κατανομή ενός συνόλου. Ντετερμινιστικά, η φράκταλ διάσταση (fractal dimension) [7, 118, 151] χρησιμοποιείται κατά κόρον σε μετρήσεις υψής όπου οι επιφάνειες ή τα περιγράμματα των αντικειμένων

είναι φράκταλ σε μια περιορισμένη περιοχή κλιμάκων. Μια ενδιαφέρουσα κατηγορία στατιστικών μοντέλων, είναι αυτή των στοχαστικών φράκταλ και ιδιαίτερα μια οικογένεια $1/f$ θορύβων (1-over-f noises), οι κλασματικές κινήσεις Brown (fractional Brownian motion ή fBm) [119]. Οι $1/f$ θόρυβοι είναι στοχαστικές διαδικασίες με φάσμα ισχύος αντιστρόφως ανάλογο μιας δύναμης της συχνότητας. Πολύ συνηθισμένες διεργασίες στη φύση, χρησιμοποιούνται για την περιγραφή πληθώρας φυσικών φαινομένων. Οι fBm, έχουν χρησιμοποιηθεί για την περιγραφή σε πολλαπλές κλίμακες φράκταλ αναπαραστάσεων, φυσικών σκηνών και υψής επιφανειών [93, 152]. Εξάλλου, τα μοντέλα φράκταλ επιφανειών συμφωνούν με τα πολυκαναλικά μοντέλα χωρικών φίλτρων και τις ψυχοφυσικές μελέτες για τη σχέση τραχύτητας επιφάνειας, διάστασης και αντίληψης υψής [102, 152].

1.3 Υψής προβλήματα

Τα βασικά προβλήματα σε θέματα όρασης, βιολογικής και υπολογιστικής, που περιλαμβάνουν αντίληψη, αναπαράσταση και ερμηνεία υψής πηγάζουν από την αρχική δυσκολία ορισμού. Παρά τις προσπάθειες για ένα κοινό εννοιολογικό, μαθηματικό και πληροφοριακό πλαίσιο, εκκρεμεί ακόμη η διατύπωση ενός ευρέως αποδεκτού ορισμού που θα περιλαμβάνει τόσο τα αίτια όσο και τις εμφανίσεις της υψής στη φύση και κατ'επέκταση σε ψηφιακές εικόνες.

Τι είναι η φυσική υψή επιφανειών, ποιες οι ιδιότητες της, ποιες οι αιτίες της και ποιες οι σχέσεις μεταξύ αιτιών και αποτελέσματος είναι ορισμένα αρχικά ερωτήματα σε επίπεδο φυσικής και οπτικής. Από πλευράς ψυχοφυσικής, πώς γίνεται αντιληπτή η υψή, πώς διακρίνεται και πώς γίνεται η εσωτερική αναπαράσταση της όποιας πληροφορίας. Από την πλευρά της ανάλυσης εικόνων, πώς εμφανίζεται η οπτική υψή σε εικόνες, πώς οι μεταβολές της έντασης του γκριζου ή του έγχρωμου φάσματος μεταφράζονται σε υψή και ποια η σχέση της με τη φυσική υψή επιφανειών. Τέλος από πλευράς πληροφοριακών συστημάτων και υπολογιστικής όρασης, ποια η επεξεργασία που πρέπει να προσομοιώσουν τέτοια συστήματα και ποιος ο ρόλος στην οικουμενική απεικόνιση και ερμηνεία φυσικών σκηνών.

Διαισθητικά η υψή περιλαμβάνει, όπως προαναφέρθηκε, τις έννοιες περιοχής, κατανομής τιμών, κλίμακας και θεμελιωδών δομικών στοιχείων. Για κάθε μια από αυτές ανακύπτουν ερευνητικές προκλήσεις. Η ύπαρξη υψής κατά περιοχές σχετίζεται με τα προβλήματα και τις εφαρμογές της, όπως ανίχνευση, ανάλυση, διάκριση, κατηγοριοποίηση, κατάτμηση κ.α. Η κατανομή και η χωρική εξάρτηση των γκριζων τιμών υψής, παρακινούν την αναζήτηση κατάλληλων στατιστικών μοντέλων για περιγραφή και σύνθεση. Ο ρόλος της κλίμακας είναι θεμελιώδης καθώς σε διαφορετικές κλίμακες, αντικείμενα, σύνολα ή λεπτομέρειες αντικειμένων γίνονται αντιληπτά ως υψή ή μικροδομή. Επομένως η σύνδεση κλίμακας, δομής και αυτοομοιότητας βρίσκονται στον πυρήνα οποιασδήποτε μελέτης της υψής. Τέλος, σχετικά με το τι δομεί τις διαφορετικές υψές, ενδιαφέρουν τα θεμελιώδη στοιχεία ή χαρακτηριστικά της (texture primitives), είτε αυτά είναι τεχτόνια, είτε πρωτογενή γεωμετρικά στοιχεία. Συγκεκριμένα πέραν του να ορίσει και να ανιχνεύσει κανείς τέτοια δομικά συστατικά, χρειάζεται να καθορίσει τη σχέση συγκέντρωσης και πλήθους με την αντίληψη και την κλίμακα της υψής και να αναπτύξει αποτελεσματικούς τρόπους μέτρησης και εξαγωγής τους.

Η υψή είναι σύμφωνα με το Gibson [63] ένα από τα επίπεδα δομής της οπτικής πληροφορίας. Είναι οι μικροδομές ή fine structure της ύλης και συνυπάρχει μαζί με τις μακροδομές ή coarse structures και τα επαναλαμβανόμενα πρότυπα. Δομές που

καθορίζονται από κλειστές καμπύλες, περιγράφονται με όρους όπως *φόρμες*, *φιγούρες* ή *σχήματα* ενώ σχηματισμοί μεγάλης κλίμακας από τέτοιες δομές αποτελούν την *οργάνωση*. Αλλάζοντας κλίμακες η μετάβαση από τη μια δομή στην άλλη είναι σύμφωνη με τους περιορισμούς και την ανάλυση του συστήματος οπτικής αντίληψης, βιολογικού ή τεχνητού. Η πολυεπίπεδη περιγραφή εικόνων είναι βασική σε θέματα όρασης και πηγή προβλημάτων σχετικών με την εξαγωγή του ενός ή περισσότερων επιπέδων της αναπαράστασης και την διαφορετική και εξειδικευμένη επεξεργασία τους [129]. Ο διαχωρισμός της υψής και η ανίχνευση της, με όρους κατάλληλων μικροδομών, από τις κυρίαρχες δομές μιας εικόνας αλλά και από τον στοχαστικό θόρυβο είναι πολλά υποσχόμενες και σημαντικές κατευθύνσεις.

Τέλος η οπτική υφή είναι κάτι άμεσα ή έμμεσα αντιληπτό και καθορίζει τη λεπτομέρεια που αποδίδεται σε μια σκηνή ή επιφάνεια. Ανάλογη λεπτομέρεια προσδίδεται σε φωνητικά, ηχητικά ή άλλα αισθητήρια σήματα, περιγράφεται όμως με διαφορετικούς όρους και έχει διαφορετική φυσική σημασία. Αλληλεπιδράσεις των περιγραφών και των μοντέλων και μεταβάσεις από δύο διαστάσεις σε μία ή και αντίστροφα έχουν αποδειχθεί χρήσιμες στο παρελθόν, με εργαλείο πάντα γενικευμένη ν-διάστατη επεξεργασία σημάτων. Τέτοιες διαδραστικές προσεγγίσεις για την υφή έχουν ενδιαφέρον τόσο από άποψη διευκόλυνσης και ιδεών, όσο και από άποψη μιας ολοκληρωμένης θεώρησης της αντίληψης.

1.4 Συνεισφορά και πρόοδος

Η συγκεκριμένη δουλειά, μια σύνοψη της οποίας περιγράφεται εδώ, είναι μια προσπάθεια ολοκληρωμένης προσέγγισης της υψής με καινοτόμα εργαλεία. Η συνεισφορά έγκειται σε μια προσπάθεια ορισμού και περιγραφής της υψής στα πλαίσια στοιχειωδών μικροδομών. Τέτοιες μικροδομές είναι οι χωρικές διαμορφώσεις και οι ταλαντώσεις κυρίως αλλά και τα φράκταλ ή οι χωρικές πλεγματικές εξαρτήσεις. Μια πρώτη σύνδεση αιτίας και αποτελέσματος γίνεται και προτείνεται ανίχνευση και ποσοτικοποίηση της δραστηριότητας της υψής με όρους ενέργειας μικροδιαμορφώσεων. Οι κλίμακες, ο ορισμός τοπικά ανά περιοχές και η ύπαρξη στοιχειωδών συνιστωσών συμπεριλαμβάνονται στη συγκεκριμένη θεώρηση.

Η υφή εικόνων αναλύεται πολυκαναλικά και κατασκευάζονται διανύσματα χαρακτηριστικών με εξαγωγή των επιμέρους συνιστωσών της και εξετάζονται τα θέματα της ανίχνευσης, ανάλυσης, κατηγοριοποίησης και κατάτμησης υψής. Ερευνήθηκαν και προτείνονται δυο διαφορετικές μέθοδοι κατάτμησης. Η μία εξειδικεύεται σε εικόνες υψής, με κλίμακα, προσανατολισμό και περιοδικότητα χρησιμοποιώντας τα προτεινόμενα διανύσματα χαρακτηριστικών [55,98,99]. Η δεύτερη αποτελεί μια προσπάθεια κατάτμησης με πολλαπλά χαρακτηριστικά όπου οι γεωμετρικές δομές και η υφή επεξεργάζονται διαφορετικά και οδηγούν την τελική ομαδοποίηση της εικόνας σε ομοιόμορφες, και με τις δυο έννοιες, περιοχές [177,178].

Αρκετά θέματα εξετάστηκαν αλλά βρίσκονται ακόμη σε πρωταρχικό στάδιο, των οποίων οι ιδέες και κάποια αρχικά αποτελέσματα θα παρατεθούν για λόγους πληρότητας και κατευθύνσεις μελλοντικής έρευνας. Έτσι η μοντελοποίηση εικόνων υψής, και συγκεκριμένα της φάσης των συνιστωσών της, με τη βοήθεια στοχαστικών θορύβων και παραμέτρους αυτοομοιότητας εξετάστηκε σε θέματα ανάλυσης και σύνθεσης. Τα μοντέλα MRF ειδικά με ενεργειακούς περιορισμούς, ως προς την ενέργεια των διαμορφώσεων των μικροδομών της υψής, ερευνήθηκαν ως στατιστικό μοντέλο υψής. Ενδιαφέρουσες κατευθύνσεις ορίστηκαν επίσης στο θέμα του διαχωρισμού της υψής

από τη γεωμετρία των δομών της εικόνας ενώ ερευνάται το μοντέλο και με επιπλέον διαχωρισμό συνιστωσών θορύβου.

Τέλος αρκετά ολοκληρωμένα αποτελέσματα προκύψανε από την παράλληλη εφαρμογή των μεθόδων και ιδεών σε θέματα ανάλυσης φωνής και ανίχνευσης των μικροδομών της. Τα μοντέλα ενέργειας των μικροδομών και ανάλυσης σε πολλές συνιστώσες εφαρμόστηκαν για θέματα ανίχνευσης φωνητικών γεγονότων σε περιβάλλον θορύβου με συστηματική βελτίωση απέναντι σε συμβατικούς ή πολύ πετυχημένους αλγόριθμους. [56,57].

1.5 Διάρθρωση κειμένου

Στο Κεφάλαιο 2 γίνεται μια εισαγωγή στην έννοια της υφής, βιβλιογραφικά και διαισθητικά και ορίζεται το γενικό θεωρητικό υπόβαθρο και τα εργαλεία που θα χρησιμοποιηθούν στη συνέχεια. Περιγράφεται το μοντέλο και οι προτεινόμενες μέθοδοι ανάλυσης υφής και αναλύονται στατιστικά οι ενεργειακοί τελεστές ως βασικά εργαλεία. Στο Κεφάλαιο 3 ορίζεται η υφή στο πλαίσιο ενέργειας διαμορφώσεων και περιγράφονται νέες μέθοδοι περιγραφής και εξαγωγής χαρακτηριστικών. Στο Κεφάλαιο 4 παρουσιάζεται μια μέθοδος κατάτμησης προσανατολισμένη στην υφή (texture-oriented) με πληθώρα αποτελεσμάτων σε φυσικές και συνθετικές εικόνες. Στο Κεφάλαιο 5 παρουσιάζεται η γενικότερη εφαρμογή της ανάλυσης ψηφιοποιημένων εικόνων εδαφοτομών, με έντονη υφή και μικροδομές, τόσο σε θέματα ανάλυσης όσο και σε θέματα ανίχνευσης και κατάτμησης. Παρουσιάζεται η πρόοδος σε θέματα διαχωρισμού της εικόνας σε δύο συνιστώσες, υφής και γεωμετρίας και η σύζευξη τους σε ένα κοινό πλαίσιο κατάτμησης (coupled segmentation). Το Κεφάλαιο 6 είναι αφιερωμένο σε 1Δ σήματα και συγκεκριμένα σε σήματα φωνής όπου οι προτεινόμενες μέθοδοι εξειδικεύονται σε εφαρμογές ανίχνευσης γεγονότων. Η αναφορά καταλήγει στο Κεφάλαιο 7 με τα βήματα για να ολοκληρωθεί η τρέχουσα έρευνα, ανοιχτά ερευνητικά ζητήματα και μελλοντικές κατευθύνσεις.

□

Κεφάλαιο 2

Οπτική Υφή και Μικροδομές Διαμόρφωσης

Η οπτική υφή (visual texture) είναι η αναπαράσταση της φυσικής υφής επιφανειών που αποτυπώνεται σε ψηφιακές εικόνες μεταξύ άλλων χαρακτηριστικών όπως είναι το χρώμα, το σχήμα, η θέση, η διάταξη, ο προσανατολισμός, το βάθος, η κίνηση κ.α. Ένα ενιαίο πλαίσιο ιδιοτήτων και αρχών διέπουν τους περισσότερους ορισμούς και χρήσεις της υφής, κάποια από τα οποία θα αναπτυχθούν εδώ. Μια ερμηνεία της υφής δίνεται μέσω τοπικών διαμορφώσεων σε εικόνες, με συγκεκριμένη δομή και ανάλογες μεθόδους χειρισμού. Αναλύεται το μοντέλο διαμορφώσεων για την υφή και περιγράφονται οι ενεργειακοί τελεστές αποδιαμόρφωσης μαζί με μια σύντομη θεωρητική ανάλυση της αναμενόμενης συμπεριφοράς τους. Τέλος γίνεται μια σύντομη περιγραφή τριών άλλων καθιερωμένων μοντέλων περιγραφής μικροδομών υφής που θα ενταχθούν στη συνέχεια στο πλαίσιο των περιγραφών με όρους χωρικών διαμορφώσεων.

2.1 Ορισμοί και ιδιότητες

Οι αναπαραστάσεις των επιφανειών φυσικών αντικειμένων σε συστήματα εικόνων και υπολογιστικής όρασης δεν είναι πάντα ομοιόμορφες αλλά παρουσιάζουν μεταβολές της έντασης φωτεινότητας (intensity). Αυτές οι μεταβολές καταλήγουν σε σχηματισμούς και πρότυπα, κάποιες φορές επαναλαμβανόμενα και σε διάφορες κλίμακες, που απαρτίζουν την οπτική υφή. Τέτοιοι σχηματισμοί μπορεί να είναι αποτέλεσμα των φυσικών ιδιοτήτων της επιφάνειας όπως είναι η τραχύτητα, το βάθος, το χρώμα, η διαφορά ανακλαστικότητας κ.α. [189]. Λόγω της απουσίας μιας ενιαίας, φορμαλιστικής περιγραφής για την υφή παραθέτονται στη συνέχεια ιδέες, ορισμοί και σχόλια από τη βιβλιογραφία σε ελεύθερη μετάφραση:

“Οι οπτικές υφές ορίζονται ως σύνολα πολλών μικρών στοιχείων, που μπορεί να είναι κουκίδες διαφόρων χρωμάτων ή απλοί σχηματισμοί.” [Julesz & Bergen, 1983] [89]

“Η ερμηνεία μέσω πεπλεγμένων σωματιδίων είναι άμεσα συνδεδεμένη με την ιδέα της ευκρίνειας της υφής, την οποία μπορεί να σκεφτεί κάποιος ως το μέσο αριθμό pixel για κάθε ευδιάκριτο στοιχείο υφής (texture element).” [Ballard & Brown, 1982] [5]

“Η υφή σχετίζεται με τη χωρική κατανομή των τιμών έντασης φωτεινότητας της εικόνας και τα διακριτά χαρακτηριστικά τονικότητας ... τρία χαρακτηριστικά είναι κρίσιμα σ’ αυτή τη διάκριση (περιοχών υφής από ομοιόμορφες), το μέγεθος των μικρών περιοχών, τα σχετικά μεγέθη και ο αριθμός ξεχωριστών διακριτών χαρακτηριστικών τονικότητας.” [Haralick & Shapiro, 1992] [67]

“Το αν είναι υφή ή όχι ένα φαινόμενο, τυπικά καθορίζεται από την κλίμακα παρατήρησης. Η υφή προκύπτει από ένα αριθμό διαφορετικών πηγών. Πρώτον, τις συγκεντρώσεις πολλών μικρών αντικειμένων μπορούμε να τις σκεφτούμε καλύτερα σαν υφή...Δεύτερον, πολλές επιφάνειες χαρακτηρίζονται από οργανωμένους σχηματισμούς που μοιάζουν με πλήθος από μικρά αντικείμενα...” [Forsyth & Ponce, 2003] [59]

“Μέσα σε μια μοναδική επιφάνεια μπορεί να υπάρχει μεταβολή στην ανακλαστικότητα, στο χρώμα ή στον τρισδιάστατο προσανατολισμό (τραχύτητα επιφάνειας). Αυτές οι μεταβολές καταλήγουν στον αμφιβληστροειδή σε μια εικόνα υφής. Τέτοιες μεταβολές υφής μπορεί να είναι κανονικές (...), τυχαίες (...) ή ενδιάμεσες (...).” [Landy, 2004] [105]

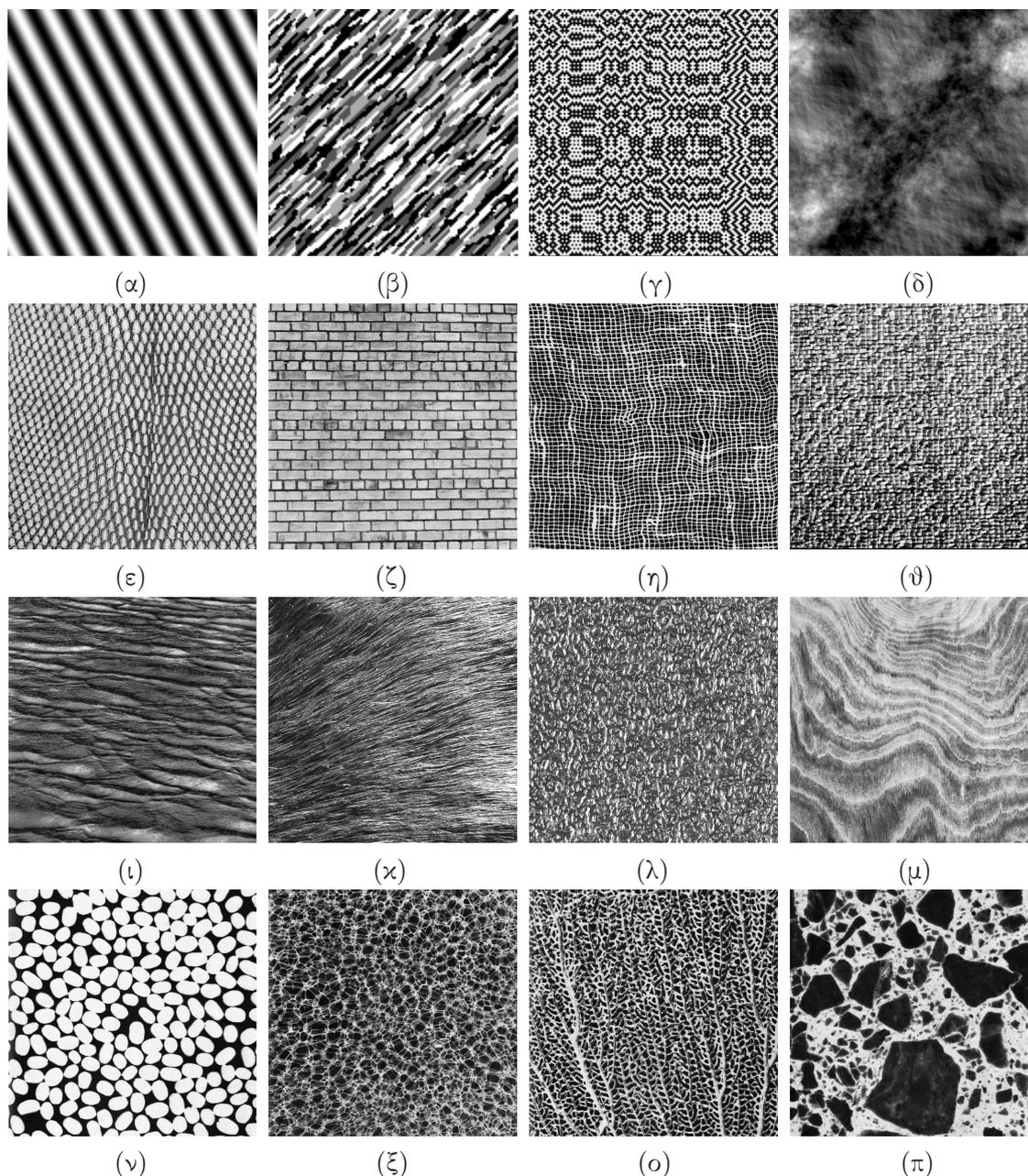
Οι ιδιότητες της υφής που αποτυπώνονται σε μια εικόνα μπορούν να ερμηνευθούν ως: 1) φυσικές μεταβολές των ιδιοτήτων της επιφάνειας (τραχύτητα, στατικότητα, γεωμετρία, ανακλαστικότητα, φυσική υφή, θέση) 2) ως μεταβολές των συνθηκών φωτισμού, απεικόνισης και περιβαλλοντικών παραγόντων 3) χημικές - βιολογικές, εξελικτικές διαδικασίες [191](σχηματισμοί και πρότυπα σε οργανισμούς, ζώα και φυτά, χημική υφή). Η οπτική υφή, στον αμφιβληστροειδή ή σε κάποιο σύστημα ψηφιακής αναπαραστάσης, δεν είναι μία ένα προς ένα απεικόνιση της φυσικής υφής μιας επιφάνειας, λόγω της συνύπαρξης των ανωτέρων ερμηνειών και των περιορισμών της απεικόνισης.

Υπάρχει ένα σύνολο διαισθητικών ιδιοτήτων της υφής, όπως αυτή γίνεται αντιληπτή, που συμπεριλαμβάνονται στα περισσότερα μοντέλα και ορισμούς και θεωρούνται πραγματικές [189]:

- Η υφή είναι ιδιότητα περιοχών, ορίζεται σε χωρικά πλέγματα, των οποίων το μέγεθος και το είδος εξαρτάται από τον τύπο της υφής ή το μέγεθος των primitives.
- Προϋποθέτει και συμπεριλαμβάνει τη χωρική κατανομή γκριζών ή έγχρωμων τιμών, οπότε μπορεί να οριστεί και να αναλυθεί στατιστικά.
- Η υφή γίνεται αντιληπτή σε διαφορετικές κλίμακες ή ευκρίνεια, και ο αριθμός, το μέγεθος και το είδος των primitives μπορούν να οριστούν διαφορετικά για κάθε κλίμακα.
- Μια περιοχή είναι ή διαθέτει υφή αν χαρακτηρίζεται από μεγάλο αριθμό primitives και από την απουσία σημαντικά κυρίαρχων μακροδομών ή μετρήσιμων συνόλων.

Η οπτική υφή διαθέτει ένα σύνολο περιγραφικών γνωρισμάτων όχι απαραίτητα ανεξάρτητα μεταξύ τους. Τέτοια μπορεί να είναι η συγκέντρωση, η ομοιομορφία, η πυκνότητα, η τραχύτητα, η σκληρότητα, η κανονικότητα, η γραμμικότητα, η ένταση, η κατευθυντικότητα, η κατεύθυνση, η συχνότητα, η φάση, η περιοδικότητα, η τυχαιότητα, η δομή,

η φυσικότητα. Στο σχήμα 2.1 φαίνονται τεχνητές και φωτογραφικές εικόνες υφής, από φυσικές επιφάνειες με περισσότερο ή λιγότερο αντιληπτές ιδιότητες. Στο σύνολο τους πάντως είναι χωρικά ομογενείς, με επαναλαμβανόμενους σχηματισμούς και συχνά κάποια τυχαία μεταβολή, στη θέση, στο μέγεθος, στον προσανατολισμό ή στις στάθμες του γκριζου.



Σχήμα 2.1: Παραδείγματα εικόνων υφής. (α)-(β): Συνηθισμένες περιοδικές και (γ)-(δ) συνθετικές τυχαίες. (ε)-(θ): Φωτογραφικές ψευδοπεριοδικές (reptile, bricks, burlap, raffia), (ι)-(μ): Φωτογραφικές τυχαίες (water, fabric, metal, wood), (ν)-(π): Φωτογραφικές δομημένες. Οι φωτογραφικές είναι από τη βάση VisTex, MIT Media Lab και τη συλλογή Brodatz [23].

2.1.1 Μοντέλα υφής

Οι μέθοδοι και τα μοντέλα περιγραφής υφής στη βιβλιογραφία μπορούν να χωριστούν σε τρεις μεγάλες κατηγορίες, ανάλογα με τις γενικές αρχές τους:

- I. *Στατιστικές*: Χαρακτηριστικά και μοντέλα από την κατανομή των γκριζών τιμών σε ένα διδιάστατο πλέγμα.
- Μοντέλα γραμμικής πρόβλεψης (2D Autoregressive model, Autocorrelation function) [33, 188].
 - Χαρακτηριστικά - στατιστικές μετρήσεις πινάκων «συνύπαρξης» (co-occurrence matrices) (εντροπία, ενέργεια, συσχέτιση, αντίθεση, ομοιογένεια), 2nd-order στατιστικά [68].
 - Gibbs (Markov) Τυχαία Πεδία [41, 48, 120, 216].
 - Στοχαστικά φράκταλ, στατιστικά αυτο-όμοιες διαδικασίες, fractional Brownian motion [93, 101, 110, 152, 154].
 - Granulometries, τοπικές κατανομές μεγέθους και σχήματος [181, 197].
- II. *Γεωμετρικές και Δομικές*: Η υφή ορίζεται από τα θεμελιώδη δομικά συστατικά της.
- Τεξτόνια (textons). [89, 108, 112, 214].
 - Ντετερμινιστικά φράκταλ, μέτρηση γενικευμένων φράκταλ διαστάσεων [31, 34, 96].
 - Διαγράμματα Voronoi [190].
 - Γραμματικά Μοντέλα (shape, tree, array grammars) [5].
- III. *Ενεργειακές - Πολυκαναλικές*: Συχνοτική ανάλυση και αναπαράσταση της υφής με πολλαπλές φασματικές συνιστώσες.
- Χαρακτηριστικά στην έξοδο γραμμικών φίλτρων (κατανομές, ενέργεια, μη γραμμικότητες) [20, 114, 163, 211].
 - Αναπαραστάσεις space-frequency, scale-space (wavelets, pyramids, filter-banks). [66, 77, 117, 156, 157].
 - Χωρικά AM-FM μοντέλα, Τοπικές διαμορφώσεις [21, 72, 123].
 - Gabor ενεργειακά χαρακτηριστικά [65].

Οι σύγχρονες τάσεις θέλουν τη χρήση ενοποιημένων μοντέλων, και την εναλλαγή ή την αναζήτηση ισοδυναμίας ανάμεσα στις περιγραφές έτσι ώστε να αφομοιώνονται τα πλεονεκτήματα των διαφορετικών προσεγγίσεων και να αντιμετωπίζεται μεθοδικά η ανάλυση της υφής.

2.2 Χωρικές συχνότητες και φίλτρα

Η αναπαράσταση εικόνων ως πίνακες τιμών φωτεινότητας ορίζει την περιγραφή της στο χώρο (spatial domain) και τις μεταβολές των τιμών έντασης από pixel σε pixel σε ένα διακριτό πλέγμα σημείων. Η μετάβαση μέσω Fourier μετασχηματισμού στο πεδίο συχνότητων (frequency domain) δίνει το συχνοτικό περιεχόμενο της εικόνας μέσω της υπέρθεσης μιγαδικών εκθετικών συντελεστών. Με τον όρο χωρική συχνότητα (spatial frequency) μιας εικόνας, ορίζεται ο ρυθμός επανάληψης ή εναλλαγής (cycle rate) των μεταβάσεων ανάμεσα στις διάφορες στάθμες φωτεινότητας. Έκφράζεται

συνήθως σε κύκλους/βαθμίδα (cycles/degree ή cycles/pixel) ή κανονικοποιημένα, στο πεπερασμένο διακριτό πλέγμα της εικόνας, σε κύκλους/εικόνα (cycles/image).

Ένα ημίτονο της μορφής $\sin[2\pi(u_c x + v_c y)]$ έχει δύο χωρικές συχνότητες, u_c και v_c που εκφράζουν αντίστοιχα το ρυθμό ταλάντωσης στην οριζόντια και κάθετη διεύθυνση, με ακτινική συχνότητα (radial frequency), $w = \sqrt{u_c^2 + v_c^2}$ και γωνία μεταβολής $\theta = \arctan(v_c/u_c)$. Η εικόνα στο σχήμα 2.1 (α) είναι ένα τέτοιο χωρικό ημίτονο με σταθερό διάνυσμα χωρικών συχνοτήτων (u_c, v_c) . Οι περιορισμοί δειγματοληψίας Nyquist-Shannon καθορίζουν τη μέγιστη φυσική συχνότητα που μπορεί να διαθέτει μια εικόνα πεπερασμένων διαστάσεων. Στο βιολογικό σύστημα όρασης η ανώτερη ανιχνεύσιμη χωρική συχνότητα, σε ψηλά επίπεδα φωτισμού είναι 50–60 cycles per degree ενώ σε χαμηλά επίπεδα 20–30 cpd.

Οι περισσότερες φυσικές εικόνες υψής είναι ευρείας ζώνης (wideband) υπό την έννοια ότι το εύρος των συχνοτήτων, που κυριαρχούν στο φάσμα τους είναι μεγάλο, παρουσιάζοντας συχνοτικό περιεχόμενο σε περισσότερες από μια ζώνες του πεδίου συχνοτήτων. Η περιγραφή της υψής σε διαφορετικές περιοχές χωρικών συχνοτήτων και οποιαδήποτε χωρο-συχνοτική (space-frequency) ανάλυση γίνεται με τη βοήθεια γραμμικών ζωνοπερατών φίλτρων [163]. Μια οικογένεια φίλτρων με πολλά ελκυστικά χαρακτηριστικά είναι τα φίλτρα Gabor, που αποτελούν χωρικά διαμορφωμένα ημίτονα από συναρτήσεις Gauss [19, 43, 45, 60]. Η απόκριση ενός μιγαδικού Gabor φίλτρου είναι:

$$h(x, y) = \exp \left[-\frac{(x - x_c)^2}{2a^2} - \frac{(y - y_c)^2}{2b^2} \right] \cdot \exp [ju_c(x - x_c) + jv_c(y - y_c)] \quad (2.1)$$

και η απόκριση συχνότητας, οπότε και η επιλεκτικότητα του στις συχνότητες:

$$H(u, v) = 2\pi ab \exp \left[-\frac{(u - u_c)^2 a^2}{2} - \frac{(v - v_c)^2 b^2}{2} \right] \cdot \exp [-ju_c x_c - jv_c y_c] \quad (2.2)$$

όπου το (x_c, y_c) προσδιορίζει την τοποθέτηση του φίλτρου χωρικά, (a, b) οι τυπικές αποκλίσεις της 2Δ Gaussian, καθορίζουν το ενεργό εύρος του φίλτρου προς τις δύο κατευθύνσεις και το διάνυσμα (u_c, v_c) καθορίζει τη διαμόρφωση συχνότητας ή την συχνοτική τοποθέτηση του φίλτρου. Ένα τέτοιο φίλτρο επιτρέπει συχνότητες σε εύρος $\pm 1/a$ και $\pm 1/b$ σε κάθε κατεύθυνση ενώ ο λόγος $\lambda = a/b$ καθορίζει τη μορφή του, με $a = b$ για ισοτροπική συμπεριφορά.

Οι ομοιότητες στην απόκριση στο χώρο και στο πεδίο συχνοτήτων μεταφράζονται σε ταυτόχρονη επεξεργασία πληροφορίας και στα δύο πεδία. Για ένα φίλτρο τοποθετημένο στην αρχή, ο Fourier μετασχηματισμός του είναι μια γκαουσιανή τοποθετημένη στις συχνότητες (u_c, v_c) , δηλαδή ένα ζωνοπερατό φίλτρο που επιτρέπει τη διέλευση περιορισμένου εύρους συχνοτήτων στις δύο διευθύνσεις. Τα Gabor αποτελούν ομαλά και εύχρηστα ζωνοπερατά φίλτρα, εύκολα ελέγξιμα και ικανά να παράγουν μια ολόκληρη οικογένεια με παρόμοια χαρακτηριστικά μεταβάλλοντας ένα μικρό σύνολο παραμέτρων.

Επιπλέον είναι τα μόνα που ικανοποιούν το κατώτερο όριο στη σχέση αβεβαιότητας που συνδέει χωρική και φασματική πληροφορία. Από τον D. Gabor [60] αποδείχθηκε ότι μονοδιάστατα τέτοια φίλτρα ελαχιστοποιούν τη σχέση αβεβαιότητας $\Delta x \cdot \Delta u \geq 1/4$, όπου το Δ εκφράζει ενεργό rms εύρος. Επεκτείνοντας τον ορισμό τους στις δύο διαστάσεις ο Daugman [45] απέδειξε ότι οι 2Δ Gabor συναρτήσεις της μορφής 2.1, 2.2

είναι οι μόνες που ταυτόχρονα ελαχιστοποιούν την αβεβαιότητα στις δύο διαστάσεις:

$$\begin{aligned} (\Delta x)(\Delta y)(\Delta u)(\Delta v) &\geq 1/16 & (2.3) \\ \Delta x = \frac{a\sqrt{2}}{2}, \Delta u = \frac{1}{2\sqrt{2}a} &\Rightarrow \Delta x \cdot \Delta u = 1/4 \\ \Delta y = \frac{b\sqrt{2}}{2}, \Delta v = \frac{1}{2\sqrt{2}b} &\Rightarrow \Delta y \cdot \Delta v = 1/4 \end{aligned}$$

Η σημασία των παραπάνω σχέσεων είναι ότι τα Gabor καταφέρνουν ταυτόχρονα να δώσουν μέγιστη πληροφορία στο χωρικό και στο φασματικό πεδίο με την ελάχιστη αβεβαιότητα. Έτσι εξάγουν με βέλτιστο τρόπο πληροφορία τόσο προς τη δομή («τι») όσο και προς τη θέση («που»).

Ένα επιπλέον κίνητρο για την χρήση τέτοιων φίλτρων για εξαγωγή πληροφορίας υφής είναι οι βιολογικές και φυσιολογικές ενδείξεις γύρω από τη συμπεριφορά των απλών κυττάρων (simple cells) [79, 80], μια κατηγορία νευρώνων, στον οπτικό φλοιό των θηλαστικών. Εμπειρικές μελέτες στα προφίλ των χωρικών δεκτικών πεδίων (spatial receptive fields) τέτοιων γενικευμένων απλών κυττάρων συνηγορούν στη μοντελοποίηση από Gabor συναρτήσεις [40, 44, 45, 128].

Με κατάλληλους μετασχηματισμούς βάσης, οποιαδήποτε εικόνα μπορεί να περιγραφεί από ένα πεπερασμένο σύνολο στοιχειωδών 2Δ Gabor συναρτήσεων [43, 46, 156]. Μια τέτοια αναπαράσταση αποτελεί ένα κοινό χωρικό-συχνοτικό πλαίσιο ανάλυσης οπτικής πληροφορίας και μπορεί να ερμηνευθεί ως μια συνεχής γέφυρα ανάμεσα στα πεδία χώρου και της χωρικής συχνότητας. Μεταβάλλοντας τις παραμέτρους κλίμακας a , b της Gaussian επιτυγχάνεται μετάβαση από το ένα πεδίο περιγραφής στο άλλο με ενδιάμεση την Gabor αναπαράσταση. Για μεγάλες τιμές ο Gabor μετασχηματισμός τείνει οριακά στο Fourier και στο πεδίο συχνοτήτων, ενώ για πολύ μικρές, όπου ο όρος της συνάρτησης Gauss εκφυλίζεται σε μια 2Δ συνάρτηση δέλτα, ο μετασχηματισμός είναι στην ουσία δειγματοληψία στο χωρικό πεδίο:

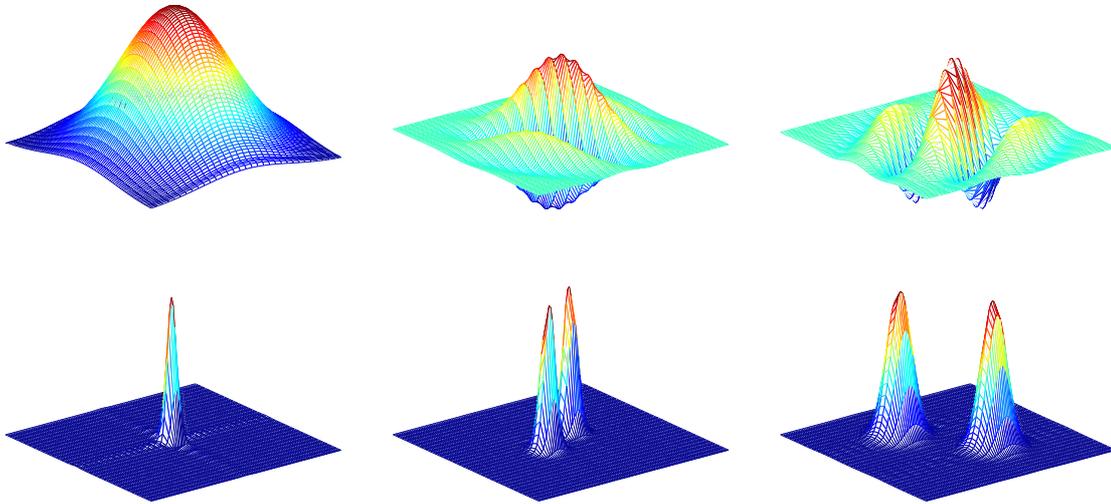
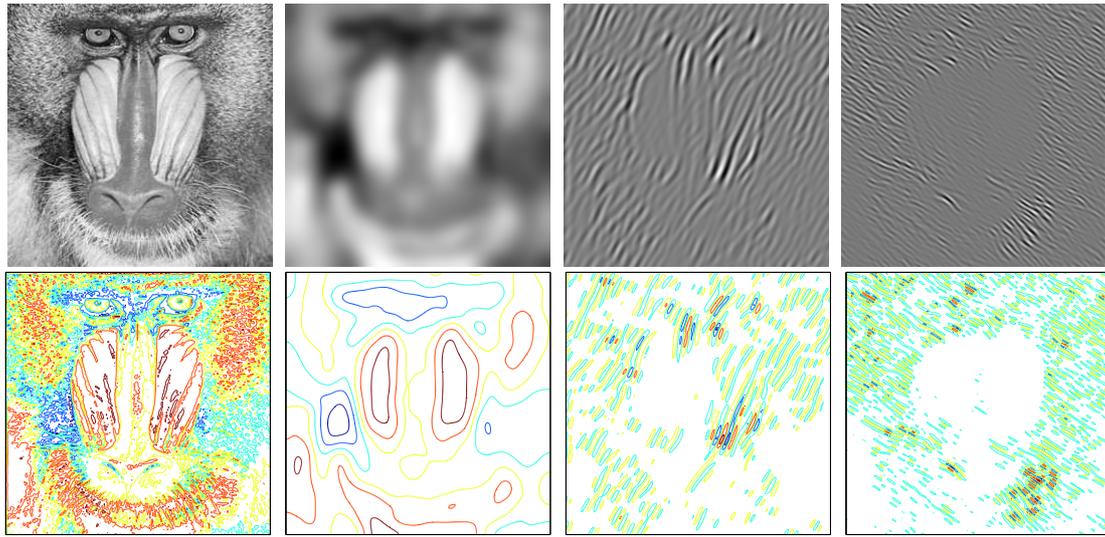
$$\lim_{a,b \rightarrow \infty} [h(x, y|a, b) * I(x, y)] = F\{I\}(u, v) \quad (2.4)$$

$$\lim_{a,b \rightarrow 0} [h(x, y|a, b) * I(x, y)] = \delta(x, y) * I(x, y) \quad (2.5)$$

όπου $I(x, y)$ η εικόνα, $F\{ \cdot \}$ ο Fourier μετασχηματισμός της και $h(x, y|a, b)$ η απόκριση του φίλτρου (2.1). Με χρήση μιας τέτοιας Gabor πυραμίδας μπορεί να επιτευχθεί πλήρης αναπαράσταση, περιγραφή, αναγνώριση, συμπίεση και ανακατασκευή οπτικής πληροφορίας [46].

Η εφαρμογή του μιγαδικού φίλτρου Gabor της (2.1), ισοδυναμεί με φιλτράρισμα με το πραγματικό και το φανταστικό μέρος του $h(x, y) = h_c(x, y) + jh_s(x, y)$ [18]. Τα δύο φίλτρα αποτελούν ένα ζεύγος με κάθετη διαφορά φάσης (quadrature pair), όπου το πραγματικό έχει άρτια και το φανταστικό περιττή συμμετρία ως προς τον προσανατολισμό τους θ . Οι δύο εκδοχές της φιλτραρισμένης εικόνας έχουν διαφορά φάσης $\pi/2$ κατά μήκος της θ .

Στο σχήμα 2.2 φαίνεται η εφαρμογή σε μια εικόνα με έντονη υφή, τριών Gabor σε διαφορετικές χωρικές συχνότητες (ή κλίμακες και προσανατολισμούς). Το βαθυπερατό Gauss φίλτρο ομαλοποίησης στο κέντρο του πεδίου συχνοτήτων αποτελεί ένα γενικευμένο Gabor και η έξοδος του είναι απαραίτητη για ανακατασκευή της αρχικής εικόνας από τις επιμέρους ζωνοπερατές συνιστώσες της. Η συμμετρία στο Fourier πεδίο των αποκρίσεων συχνότητας των φίλτρων οφείλεται στη χρήση του πραγματικού μέρους των μιγαδικών συναρτήσεων.



Σχήμα 2.2: Ζωνοπερατό Gabor φιλτράρισμα. Πάνω σειρά: Αρχική εικόνα 256×256 και έξοδοι τριών Gabor φίλτρων, σε διαφορετικές κλίμακες και προσανατολισμούς. Δεύτερη σειρά: Επιπεδοσύνολα (level curves) σε 6 επίπεδα. Τρίτη & τέταρτη σειρά: Αποκρίσεις κρουστικές και συχνότητας των Gabor φίλτρων σε προοπτικές απεικονίσεις. Χαρακτηριστικά των τριών φίλτρων (κάθετη συχνότητα u_c , οριζόντια συχνότητα v_c , εύρος ζώνης): $\{[0, 0, 27], [13, 7, 20], [-15, 21, 11]\}$.

2.2.1 Συστοιχίες Gabor φίλτρων

Τα φίλτρα Gabor, απομονωμένα ή σε συστοιχίες χρησιμοποιήθηκαν συστηματικά για κατάτμηση, διάκριση και κατηγοριοποίηση υφής, ερευνήθηκαν και συστηματοποιήθηκαν οι παράμετροι τους (κλίμακα, προσανατολισμός, εύρος, χωρική τοποθέτηση) [19, 51] και παρουσιάστηκε η δυνατότητα αναζήτησης πεπερασμένων συνιστωσών υφής σε απλές εικόνες με χρήση ανάλογου αριθμού φίλτρων [18, 20]. Για κάλυψη όλου του εύρους των δυνατών χωρικών συχνοτήτων, χρησιμοποιούνται συστοιχίες φίλτρων (filterbanks) ακτινικής διάταξης και αυξανόμενου ανά οκτάβες εύρους, με φίλτρα των οποίων η κλίμακα (scale) αλλά και ο προσανατολισμός (orientation) δίνονται από αναλυτικές σχέσεις [21, 70, 72, 73, 163].

Για την ανάλυση πληροφορίας ευρείας ζώνης της υφής, η εικόνα χωρίζεται σε απο-

μονωμένες συνιστώσες στενού εύρους (narrowband components), των οποίων το φασματικό περιεχόμενο είναι περιορισμένο σε μια μικρή περιοχή συχνοτήτων, μέσω πολλαπλών καναλιών χωρικών συχνοτήτων τύπου wavelet. Ο διαχωρισμός γίνεται από ένα σύνολο φίλτρων, κατάλληλης δομής και διάταξης έτσι ώστε να καλύπτεται επαρκώς το διάστημα των χωρικών συχνοτήτων. Ένα επιπλέον πλεονέκτημα είναι ο περιορισμός του θορύβου ανά ζώνη και η ομαλοποίηση των όποιων ενεργειακών μετρήσεων. Το πλαίσιο πολλαπλών φίλτρων, παράλληλης επεξεργασίας και εξαγωγής χαρακτηριστικών ανά ζώνη συχνοτήτων είναι γνωστό ως *πολυζωνική ανάλυση* (multiband analysis).

Τα συνήθη κριτήρια για την επιλογή των φίλτρων της συστοιχίας είναι να είναι καλά εντοπισμένα και τοποθετημένα στο χωρικό και στο συχνοτικό πεδίο, να είναι όσο το δυνατόν στενής ζώνης και στο σύνολο να καλύπτουν πυκνά τις σημαντικές συχνότητες της εικόνας. Για το σκοπό αυτό επιλέγονται συνήθως 2Δ Gabor φίλτρα λόγω της βέλτιστης απόκρισης τους στα πλαίσια αβεβαιότητας χώρου-συχνότητας (2.4). Χρησιμοποιούνται ιστροπικά φίλτρα, σε διαφορετικές κλίμακες και προσανατολισμούς, με κάποιο κανόνα σύνδεσης του εύρους ζώνης με τις κεντρική συχνότητα του καιθενός. Κατά κανόνα σε μεγαλύτερες κεντρικές συχνότητες αντιστοιχίζεται μεγαλύτερο εύρος ζώνης.

Η συνηθέστερη διάταξη θέλει το πεδίο των συχνοτήτων να χωρίζεται ακτινικά με φίλτρα αυξανόμενης κλίμακας (radial symmetry). Για να είναι δυνατή μια πλήρης ανακατασκευή της εικόνας από το άθροισμα των επιμέρους συνιστωσών της, διατηρείται σταθερός ο λόγος της κεντρικής συχνότητας προς το εύρος ζώνης του φίλτρου i (constant-Q filterbanks):

$$Q = \frac{w_i}{\sqrt{a_i^{-2} + b_i^{-2}}} \quad , \quad w_i = \sqrt{u_{ci}^2 + v_{ci}^2} \quad (2.6)$$

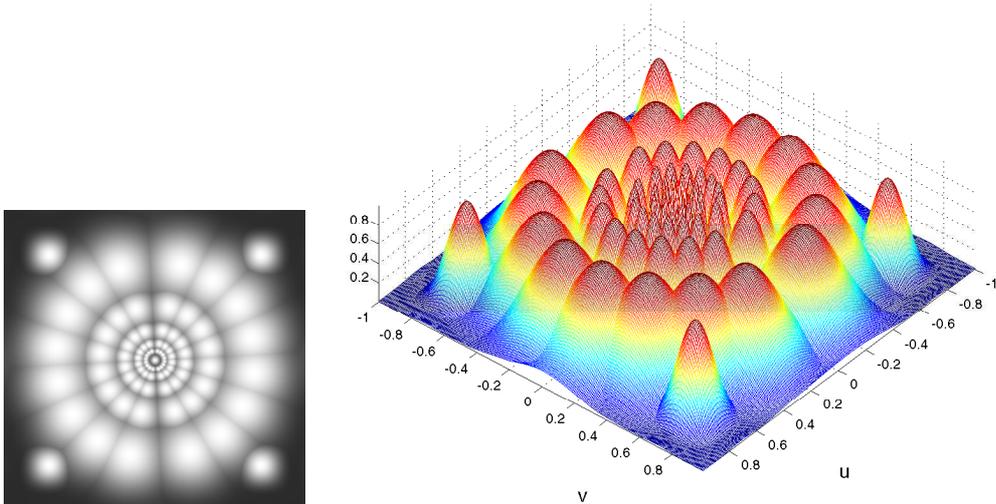
πράγμα που εξασφαλίζει σταθερή, λογαριθμική απόσταση ανάμεσα σε διαδοχικά φίλτρα της ίδιας κατεύθυνσης $\theta_i = \arctan(v_{ci}/u_{ci})$. Η απόσταση αυτή εκφράζεται σε οκτάβες έτσι ώστε φίλτρα που απέχουν μία οκτάβα να τέμνονται ακριβώς σε ύψος μισής κορυφής. Ενδεικτικά, για κλίμακες 2, 4, 8, 16 και 32 pixel, σε 8 διευθύνσεις $0 \leq \theta \leq \pi$, με βήμα $\pi/8$ και κεντρικές συχνότητες ανά κατεύθυνση $w_i = \log_2 Q - 0.5 \log_2(a_i^{-2} + b_i^{-2})$, για απόσταση 2 οκτάβες ανάμεσα σε γειτονικά φίλτρα ($Q = 2$), προκύπτει μια συστοιχία 40 φίλτρων, ακτινικής συμμετρίας με 5 φίλτρα ανά ακτίνα.

Ένα τέτοιο σύνολο, στο πεδίο των συχνοτήτων, φαίνεται στο σχήμα 2.3 ως γκριζα εικόνα και ως προοπτική 3Δ απεικόνιση. Για λόγους αναπαράστασης κάθε φίλτρο έχει κανονικοποιηθεί στο μέγιστο εύρος γκριζων τιμών. Ανάλογα ορίζονται και συστοιχίες ιστροπικών φίλτρων, με ίδιο εύρος ζώνης και γραμμικό διαχωρισμό ανά κατεύθυνση (linear filterbanks). Μια τέτοια διάταξη συμπεριφέρεται με τον ίδιο τρόπο στις ψηλές και χαμηλές ζώνες συχνοτήτων αλλά δεν εξασφαλίζει απαραίτητα την ανακατασκευή της αρχικής εικόνας από τις επιμέρους ζωνοπερατές εκδοχές της.

2.3 Μοντέλα διαμόρφωσης, ενεργειακοί τελεστές και αποδιαμόρφωση

Εικόνες υφής, τοπικά στενής ζώνης κατά την έννοια [21] μπορούν να μοντελοποιηθούν ως χωρικά AM-FM σήματα

$$f(x, y) = a(x, y) \cos[\phi(x, y)] \quad , \quad \vec{\omega}(x, y) = \nabla \phi(x, y) \quad (2.7)$$



Σχήμα 2.3: Αποκρίσεις συχνότητας (σε γκριζα κλίμακα φωτεινότητας και σε 3Δ προοπτική απεικόνιση) συστοιχίας 2Δ Gabor φίλτρων. Υπάρχουν 40 φίλτρα σε ακτινική διάταξη τύπου wavelet, σε 8 ακτίνες με 5 φίλτρα ανά ακτίνα, ένα επιπλέον στο $(0, 0)$ και δύο στο άνω και κάτω επίπεδο συχνοτήτων. Η συμμετρία ως προς το αριστερό ημιεπίπεδο οφείλεται στη χρήση πραγματικών φίλτρων και στη συμμετρία του 2Δ Fourier.

δηλαδή μη-στατικά (non-stationary) ημίτονα με διαμόρφωση πλάτους (AM) και διαμόρφωση συχνότητας (FM). Το χωρικά μεταβαλλόμενο πλάτος $a(x, y)$ μοντελοποιεί την τοπική αντίθεση (contrast) και το χωρικά μεταβαλλόμενο διάνυσμα στιγμιαίων συχνοτήτων $\vec{\omega}(x, y) = (\omega_1, \omega_2)(x, y)$ περιέχει πληροφορία για τις τοπικά αναδυόμενες συχνοτήτες. Το πλάτος και οι συχνοτήτες είναι ομαλά, τοπικά στενής ζώνης, σήματα. Τέτοια μοντέλα διαμόρφωσης προτάθηκαν [21, 70, 73, 74, 123] και εφαρμόστηκαν σε μια πληθώρα προβλημάτων όρασης και ανάλυσης εικόνων. Δεν εφαρμόζονται απευθείας στην εικόνα ευρείας ζώνης αλλά στις ζωνοπερατές συνιστώσες της στην έξοδο μιας συστοιχίας Gabor φίλτρων, που εκτός από τον διαχωρισμό, αυξάνουν την ανοχή σε θόρυβο και επιβάλλουν κάποια ομαλότητα στα σήματα πλάτους και συχνότητας.

2.3.1 1Δ ενεργειακός τελεστής

Ο μη-γραμμικός διαφορικός τελεστής Ψ που αναπτύχθηκε από τον Teager [186] και προτάθηκε συστηματικά από τον Kaiser [91, 92], για 1Δ σήματα συνεχούς χρόνου $f(t)$ ορίζεται ως:

$$\Psi(f)(t) \triangleq [f'(t)]^2 - f(t)f''(t) \quad (2.8)$$

με $f' = df/dt$ και $f'' = d^2f/dt^2$. Το διακριτό αντίστοιχο του Ψ είναι ο τελεστής

$$\Psi_d(f)(n) \triangleq f^2(n) - f(n-1)f(n+1) \quad (2.9)$$

για σήματα διακριτού χρόνου $f(n)$. Οι Ψ και Ψ_d αναφέρονται ως *ενεργειακοί τελεστές* υπό την έννοια ότι εντοπίζουν την ενέργεια απλών αρμονικών ταλαντώσεων ημιτονικών σημάτων. Αυτή η ενέργεια είναι ανάλογη τόσο του τετραγώνου του πλάτους όσο και του τετραγώνου της συχνότητας της ταλάντωσης.

Οι ενεργειακοί τελεστές μπορούν να εκτιμήσουν στιγμιαία την περιβάλλουσα διαμορφωμένων σημάτων κατά πλάτος ή συχνότητα [126] ενώ για γενικά AM-FM σήματα

της μορφής $f(t) = a(t) \cos[\phi(t)]$, εντοπίζουν το ενεργειακό γινόμενο

$$\Psi[a(t) \cos(\phi(t))] \approx a^2(t)\omega^2(t) \quad (2.10)$$

όπου $\omega(t) = d\phi(t)/dt$ η στιγμιαία γωνιακή συχνότητα. Η προσέγγιση ισχύει με μικρό σφάλμα όταν τόσο το πλάτος όσο και η συχνότητα δεν μεταβάλλονται πολύ γρήγορα ή σε μεγάλο εύρος σε σχέση με το φέρον. Εφαρμόζοντας επιπλέον τον Ψ στην παράγωγο $f'(t)$ αναπτύχθηκε στο [125] ένας 1Δ αλγόριθμος αποδιαμόρφωσης που διαχωρίζει την ενέργεια (2.10) σε συντελεστές πλάτους και συχνότητας και αποδιαμορφώνει AM-FM σήματα εκτιμώντας την περιβάλλουσα $|a(t)|$ και τη στιγμιαία συχνότητα $\omega(t)$ [22, 125].

2.3.2 Ενεργειακός τελεστής για 2Δ σήματα

Μια επέκταση του 1Δ ενεργειακού τελεστή προτάθηκε στο [123] σε ένα γενικότερο πλαίσιο πολυδιάστατων ενεργειακών τελεστών και 2Δ τελεστών για εικόνες. Αν $f(\vec{x})$ ένα διδιάστατο πραγματικό σήμα με $\vec{x} = (x, y) \in \mathbf{R}^2$ τότε ο 2Δ ενεργειακός τελεστής ορίζεται ως:

$$\Psi(f)(\vec{x}) \triangleq \|\nabla f(\vec{x})\|^2 - f(\vec{x})\nabla^2 f(\vec{x}), \quad \vec{x} = (x, y) \in \mathbf{R}^2 \quad (2.11)$$

όπου ∇f η κλίση (gradient) της f , $\|\cdot\|$ η Ευκλείδεια νόρμα, και $\nabla^2 f$ η Laplacian του f . Ο $\Psi(f)$ μπορεί εξ ορισμού να εκφραστεί ως το άθροισμα δύο μονοδιάστατων τελεστών στις δύο διευθύνσεις δηλ.:

$$\Psi(f) = \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 - f\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) = \sum_{k \in \{x, y\}} \Psi_k(f) \quad (2.12)$$

με $\Psi_k(f) = f_k^2 - f f_{kk}$. Έτσι η έξοδος του Ψ είναι ένα άθροισμα των ενεργειακών συνιστωσών του 1Δ τελεστή στην οριζόντια και κάθετη διεύθυνση. Ένα σύνολο ιδιοτήτων του 2Δ Ψ δίνονται στο [123].

2.3.2.1 Μιγαδικά σήματα

Η επέκταση του 2Δ τελεστή για μιγαδικά σήματα δόθηκε στο [123] με βάση ένα γενικευμένο τελεστή για διανυσματικά σήματα. Ο τελεστής C για ένα μιγαδικό σήμα $f(\vec{x}) = (f_r + j f_i)(\vec{x})$ ορίστηκε ως

$$C(f)(\vec{x}) \triangleq \|\nabla f(\vec{x})\|^2 - \text{Real}[f^*(\vec{x})\nabla^2 f(\vec{x})] \quad (2.13)$$

όπου $()^*$ ορίζει το μιγαδικό συμπλήρωμα και $f_r = \text{Real}(f)$ και $f_i = \text{Imag}(f)$. Εύκολα διαπιστώνουμε ότι προκύπτει

$$C(f) = \Psi(f_r) + \Psi(f_i) = \Psi[\text{Real}(f)] + \Psi[\text{Imag}(f)] \quad (2.14)$$

Έτσι ανάλυση εφαρμογής του ενεργειακού τελεστή σε μιγαδικά σήματα καταλήγει σε ξεχωριστή ανάλυση του πραγματικού και του φανταστικού τους μέρους.

Ενδιαφέρον παρουσιάζει η σύγκριση του συμπεράσματος της 2.14 με την απευθείας εφαρμογή του Ψ της σχέσης 2.11 στο μιγαδικό σήμα $f(\vec{x})$. Με απλούς υπολογισμούς

$$\Psi(f) = \Psi(f_r + j f_i) = \Psi(f_r) - \Psi(f_i) + j \nabla^2 [f_r f_i] \quad (2.15)$$

Συγκρίνοντας την (2.14) με την (2.15) φαίνεται ότι $C(f) \neq \text{Real}[\Psi(f)]$, οπότε ο ο ενεργειακός τελεστής για μιγαδικά σήματα, μια πραγματική ποσότητα, δεν είναι ισοδύναμος με το πραγματικό μέρος της εφαρμογής του ενεργειακού τελεστή στο σήμα.

2.3.3 AM-FM σήματα και διαχωρισμός ενέργειας

Έστω ένα πραγματικό 2Δ ημιτονοειδές σήμα, διαμορφωμένο κατά πλάτος και συχνότητα, της μορφής $f(x, y) = a(x, y) \cos[\phi(x, y)]$, όπου $a(x, y)$ το χωρικά μεταβαλλόμενο πλάτος, $\phi(x, y)$ το σήμα φάσης και

$$\vec{\omega}(x, y) \triangleq \nabla\phi(x, y) = [\omega_1(x, y), \omega_2(x, y)] \quad (2.16)$$

το διάνυσμα των στιγμιαίων χωρικά μεταβαλλόμενων γωνιακών συχνοτήτων

$$\omega_1(x, y) \triangleq \frac{\partial\phi}{\partial x}(x, y) \quad , \quad \omega_2(x, y) \triangleq \frac{\partial\phi}{\partial y}(x, y). \quad (2.17)$$

Το κάθε σήμα στιγμιαίας συχνότητας μπορεί να γραφεί στη μορφή $\omega_k(x, y) = \omega_{c,k} + \omega_{m,k}q_k(x, y)$, $k = \{1, 2\}$, όπου $\omega_{c,k}$ η σταθερή συχνότητα φέροντος, $q_k(\vec{x}) \in [-1, 1]$ το σήμα διαμόρφωσης συχνότητας με $\omega_{m,k}$ τη μέγιστη απόκλιση του ω_k από την κεντρική του τιμή.

Εφαρμογή του Ψ σε ένα AM-FM σήμα f της μορφής (2.7) καταλήγει προσεγγιστικά:

$$\Psi[a \cos(\phi)] \approx a^2 \|\vec{\omega}\|^2 \quad (2.18)$$

Το σφάλμα προσέγγισης είναι αμελητέο κάτω από ρεαλιστικές συνθήκες [123], που θέλουν τις μεταβολές πλάτους a και στιγμιαίων συχνοτήτων $\vec{\omega}$ να μην είναι πολύ γρήγορες στο χώρο ούτε πολύ μεγάλες σε τιμές σε σχέση με τις φέρουσες ($\omega_{c,1}, \omega_{c,2}$). Το (2.18) είναι ένα ενεργειακό γινόμενο των τετραγώνων του πλάτους και του μέτρου της συχνότητας. Διαχωρισμός αυτής της ενέργειας στα δύο σήματα διαμόρφωσης είναι εφικτός αν εφαρμόσουμε επιπλέον τον Ψ στα μερικά διαφορικά του f στις δύο διευθύνσεις $\partial f/\partial x$ και $\partial f/\partial y$.

$$\Psi\left(\frac{\partial f}{\partial x}\right) \approx a^2 \omega_1^2 \|\vec{\omega}\|^2 \quad , \quad \Psi\left(\frac{\partial f}{\partial y}\right) \approx a^2 \omega_2^2 \|\vec{\omega}\|^2 \quad (2.19)$$

Συνδυάζοντας τις (2.18),(2.19) προκύπτει ο μη γραμμικός Συνεχής Αλγόριθμος Διαχωρισμού Ενέργειας (Continuous Energy Separation Algorithm ή CESA) :

$$\sqrt{\frac{\Psi(\partial f/\partial x)}{\Psi(f)}} \approx |\omega_1(x, y)| \quad , \quad \sqrt{\frac{\Psi(\partial f/\partial y)}{\Psi(f)}} \approx |\omega_2(x, y)| \quad (2.20)$$

$$\frac{\Psi(f)}{\sqrt{\Psi(\partial f/\partial x) + \Psi(\partial f/\partial y)}} \approx |a(x, y)| \quad (2.21)$$

Ο CESA είναι επέκταση του 1Δ αλγορίθμου για αποδιαμόρφωση χρονικά μεταβαλλόμενων σημάτων [22, 125, 126] και μπορεί να εκτιμήσει σε κάθε θέση (x, y) την περιβάλλουσα του πλάτους και το μέτρο της στιγμιαίας συχνότητας ενός χωρικά μεταβαλλόμενου AM-FM σήματος. Για το πρόσημο των συχνοτήτων χρησιμοποιούνται τα πρόσημα των φερουσών, δηλ. της μέσης τιμής του σήματος σε κάθε διεύθυνση. Για εικόνες στην έξοδο συμμετρικού ζωνοπερατού φίλτρου, το διάνυσμα των φερουσών προσεγγίζεται από το διάνυσμα των κεντρικών συχνοτήτων του φίλτρου.

2.3.4 Διακριτοί τελεστές και αποδιαμόρφωση εικόνων

Για την ενεργειακή ανάλυση και την αποδοδιαμόρφωση διδιάστατων σημάτων και συγκεκριμένα εικόνων, που ορίζονται σε ένα 2Δ πεπερασμένο χωρικό πλέγμα, χρησιμοποιούνται οι διακριτές εκδοχές των προηγούμενων σχέσεων. Η ερμηνεία (2.12) του Ψ επιτρέπει την επέκταση του σε σήματα διακριτού χώρου (discrete-space) $f(m, n)$. Αντικαθιστώντας τις ενεργειακές συνιστώσες με τις εξόδους του 1Δ διακριτού τελεστή Ψ_d προκύπτει ο 2Δ διακριτός ενεργειακός τελεστής

$$\begin{aligned}\Psi_d(f)(m, n) &\triangleq \Psi_{d,1}(f)(m, n) + \Psi_{d,2}(f)(m, n) \\ &= 2f^2(m, n) - f(m-1, n)f(m+1, n) - f(m, n-1)f(m, n+1)\end{aligned}\quad (2.22)$$

όπου ο $\Psi_{d,1}$ ενεργεί κάθετα σε όλες τις στήλες $0 \leq m \leq M$ ενώ ο $\Psi_{d,2}$ οριζόντια σε όλες τις γραμμές $0 \leq n \leq N$ του f . Η έκφραση (2.22) είναι όμοια με τον διακριτό τελεστή στο [138] που αναπτύχθηκε για βελτίωση εικόνων (image enhancement) και ανίχνευση ακμών (edge detection). Αντικαθιστώντας τα διαφορικά στο Ψ με εναλλακτικά 2Δ σχήματα χωρικών διαφορών μπορούν να προκύψουν διαφορετικά σχήματα 2Δ διακριτών ενεργειακών τελεστών.

Έστω ένα διακριτό AM-FM σήμα

$$f(m, n) = a(m, n) \cos[\phi(m, n)] \quad (2.23)$$

με κάθετη στιγμιαία συχνότητα $\Omega_1(m, n) \triangleq \partial\phi/\partial m$ σε ακτίνια/δείγμα (radians/sample) και οριζόντια $\Omega_2(m, n) \triangleq \partial\phi/\partial n$ με εύρος τιμών $[-\pi, \pi]$ και σταθερό πρόσημο για κάθε (m, n) . Κάτω από την προϋπόθεση ότι τόσο το πλάτος a όσο και τα σήματα στιγμιαίων συχνοτήτων είναι ζωνοπεριορισμένα και ισχύουν κάποιες ρεαλιστικές συνθήκες σχετικά με το εύρος της μεταβολής τους [123, 126] τότε

$$\Psi_d[a(m, n) \cos(\phi(m, n))] \approx a^2(m, n)(\sin^2[\Omega_1(m, n)] + \sin^2[\Omega_2(m, n)]). \quad (2.24)$$

Προσεγγίζοντας τα διαφορικά του f στις δύο κατευθύνσεις με συμμετρικές διαφορές τριών δειγμάτων προκύπτει ο Διακριτός Αλγόριθμος Διαχωρισμού Ενέργειας (Discrete Energy Separation Algorithm ή DESA) :

$$\arcsin\left(\sqrt{\frac{\Psi_d[f(m+1, n) - f(m-1, n)]}{4\Psi_d[f(m, n)]}}\right) \approx |\Omega_1(m, n)| \quad (2.25)$$

$$\arcsin\left(\sqrt{\frac{\Psi_d[f(m, n+1) - f(m, n-1)]}{4\Psi_d[f(m, n)]}}\right) \approx |\Omega_2(m, n)| \quad (2.26)$$

$$\frac{2\Psi_d[f(m, n)]}{\sqrt{\Psi_d[f(m+1, n) - f(m-1, n)] + \Psi_d[f(m, n+1) - f(m, n-1)]}} \approx |a(m, n)| \quad (2.27)$$

Ο DESA μπορεί να εκτιμήσει σε κάθε σημείο την περιβάλλουσα και το μέτρο των δύο στιγμιαίων συχνοτήτων του χωρικού AM-FM σήματος. Ο περιορισμός του είναι ότι $0 \leq |\Omega_1|, |\Omega_2| \leq \pi/2$, δηλ. εκτιμάει συχνότητες μέχρι το ένα τέταρτο της συχνότητας δειγματοληψίας. Τα πρόσημα τους μπορούν να βρεθούν από τα πρόσημα

των αντίστοιχων συχνοτήτων φέροντος. Αν το AM-FM σήμα έχει σταθερό πλάτος A και σταθερές συχνότητες $\Omega_{c,1}$ και $\Omega_{c,2}$, τότε ο DESA παρέχει ακριβείς εκτιμήσεις του απόλυτου πλάτους $|a(m, n)| = |A|$ και συχνοτήτων $|\Omega_1(m, n)| = |\Omega_{c,1}|$ και $|\Omega_2(m, n)| = |\Omega_{c,2}|$.

Τα πλεονεκτήματα του DESA για αποδιαμόρφωση των στιγμιαίων συνιστωσών AM-FM σημάτων είναι η απλότητα, η αποδοτικότητα, η μικρή υπολογιστική πολυπλοκότητα, η εξαιρετική, σχεδόν στιγμιαία, χωρική ανάλυση σε παράθυρα 3×3 και η ικανότητα του να εντοπίζει χωρικές δομές διαμόρφωσης. Διαφορετικές προσεγγίσεις για αποδιαμόρφωση χωρικά διαμορφωμένων σημάτων είναι με χρήση του μετασχηματισμού Hilbert, που για 1Δ σήματα και σήματα φωνής δίνονται συγκρίσεις στο [158], και του πολυδιάστατου κατευθυντικού Hilbert στην αναλυτική εικόνα [39, 75].

2.3.5 Ανάλυση του 2Δ τελεστή σε θόρυβο

Στην ενότητα αυτή παρουσιάζονται κάποια πρώιμα θεωρητικά αποτελέσματα σχετικά με την αναμενόμενη συμπεριφορά των ενεργειακών τελεστών σε 2Δ σήματα και εικόνες. Αναπτύσσονται οι βασικές στατιστικές ιδιότητες (ροπές κατώτερης τάξης) της εφαρμογής του (2.11) σε ένα τυχαίο 2Δ σήμα $N(x, y)$ και δίνεται μια σχέση ανάμεσα στο εύρος ζωνοπερατού φιλτραρίσματος και μέσης τιμής του τελεστή. Μια τέτοια ανάλυση είναι σημαντική για τη συμπεριφορά του τελεστή σε συστήματα εικόνων με θόρυβο και ανάλογη ανάλυση έχει γίνει για τον 1Δ Teager-Kaiser τελεστή στο [22].

2.3.5.1 Στατιστικά

Το $N(x, y)$ υποτίθεται να είναι στατική με την ευρεία έννοια (wide-sense stationary), μηδενικής μέσης τιμής (zero-mean), Gaussian τυχαία διαδικασία, με συνάρτηση αυτοσυσχέτισης $R(\tau_1, \tau_2)$ και φασματική πυκνότητα ισχύος (power spectral density)

$$\Phi(\vec{\omega}) = \int_{\mathbb{R}^2} R(\vec{\tau}) e^{-\vec{\tau} \cdot \vec{\omega}} d\vec{\tau}, \quad \vec{\tau} = (\tau_1, \tau_2), \quad \vec{\omega} = (\omega_x, \omega_y) \quad (2.28)$$

$$\Phi(\omega_x, \omega_y) = \int_{\mathbb{R}^2} R(\tau_1, \tau_2) e^{(-\omega_x \tau_1 - \omega_y \tau_2)} d\tau_1 d\tau_2 \quad (2.29)$$

Επειδή η $N(x, y)$ είναι wss Gaussian διαδικασία, οι μερικές παράγωγοι $N_x, N_{xx}, N_y, N_{yy}, N_{xy}$, όπου οι υποδείκτες δηλώνουν μερική παραγωγή προς την αντίστοιχη μεταβλητή, είναι επίσης wss Gaussian. Επιπλέον, η N_x είναι στατιστικά ανεξάρτητη τόσο της N όσο και της N_{xx} και ομοίως η N_y είναι στατιστικά ανεξάρτητη της και N_{yy} [145, σελ. 133]. Έτσι η εφαρμογή του τελεστή (2.12) στην N

$$\Psi(N) = (N_x)^2 - NN_{xx} + (N_y)^2 - NN_{yy} \quad (2.30)$$

είναι το άθροισμα τεσσάρων ανεξάρτητων διαδικασιών. Έστω συμβολικά

$$R^{(m,n)}(\tau_1, \tau_2) = \frac{\partial^k R(\tau_1, \tau_2)}{\partial \tau_1^m \partial \tau_2^n}, \quad k = m + n \quad (2.31)$$

και τα $E[\cdot]$ και $\text{Var}[\cdot]$ δηλώνουν στατιστική αναμενόμενη τιμή (expectation) και μεταβλητότητα (variance). Ισχύουν τα εξής [145, σελ. 319]

$$\text{Var}[N] = R(0, 0) = \gamma_0^2 \quad (2.32\alpha')$$

$$\text{Var}[N_x] = -R^{(2,0)}(0, 0) = \alpha_1^2 \quad (2.32\beta')$$

$$\text{Var}[N_y] = -R^{(0,2)}(0, 0) = \beta_1^2 \quad (2.32\gamma')$$

$$\text{Var}[N_{xx}] = R^{(4,0)}(0, 0) = \alpha_2^2 \quad (2.32\delta')$$

$$\text{Var}[N_{yy}] = R^{(0,4)}(0, 0) = \beta_2^2 \quad (2.32\epsilon')$$

$$E[N N_{xx}] = R^{(2,0)}(0, 0) = -\alpha_1^2 \quad (2.32\zeta')$$

$$E[N N_{yy}] = R^{(0,2)}(0, 0) = -\beta_1^2 \quad (2.32\zeta')$$

$$E[N_{xx} N_{yy}] = R^{(2,2)}(0, 0) = \gamma_1^2 \quad (2.32\eta')$$

Λόγω των μηδενικών μέσων τιμών $\text{Var}[\cdot] = E[(\cdot)^2]$, ενώ επειδή η N είναι wss ισχύει $E[N_{x^m} N_{x^n}] = (-1)^n R^{(m,n)}(0, \tau_2)$, $n > m$.

Με βάση τα παραπάνω για τις ροπές του Ψ έχουμε

$$\begin{aligned} E[\Psi(N)] &= E[\Psi_x(N)] + E[\Psi_y(N)] \\ &= \text{Var}[N_x] - E[N N_{xx}] + \text{Var}[N_y] - E[N N_{yy}] \\ &= 2\alpha_1^2 + 2\beta_1^2 \end{aligned} \quad (2.33)$$

με Ψ_x, Ψ_y τους κατευθυντικούς ενεργειακούς τελεστές ως προς x και y αντίστοιχα. Έτσι η αναμενόμενη τιμή $\Psi(N)$ είναι δύο φορές το άθροισμα μεταβλητότητας των $\partial N/\partial x$ και $\partial N/\partial y$. Από την παραπάνω σχέση προκύπτει επίσης η εξάρτηση της εξόδου του τελεστή από το φάσμα ισχύος (2.29)

$$\begin{aligned} E[\Psi(N)] &= -R^{(2,0)}(0, \tau_2) - R^{(0,2)}(\tau_1, 0) \\ &= \frac{1}{\pi} \int_{\mathbb{R}^2} \omega_y^2 \Phi(\omega_x, \omega_y) d\omega_x d\omega_y + \frac{1}{\pi} \int_{\mathbb{R}^2} \omega_x^2 \Phi(\omega_x, \omega_y) d\omega_x d\omega_y \\ &= \frac{1}{\pi} \int_{\mathbb{R}^2} |\vec{\omega}|^2 \Phi(\omega_x, \omega_y) d\omega_x d\omega_y \end{aligned} \quad (2.34)$$

λόγω της σχέσης Fourier αυτοσυσχέτισης και φάσματος ισχύος και τις ιδιότητες παραγωγίσης του μετασχηματισμού. Επομένως η μέση τιμή του $\Psi(N)$ αυξάνεται γραμμικά με την φασματική ενεργειακή μεταβλητότητα του N . Αυτό μπορεί να αποτελεί πρόβλημα για πολύ θορυβώδη σήματα.

Η αυτοσυσχέτιση $R_{\Psi(N)}(\tau_1, \tau_2)$ μπορεί να βρεθεί χρησιμοποιώντας τη φόρμουλα του Isserli για την αναμενόμενη τιμή γινομένων τεσσάρων τυχαίων μεταβλητών μηδενικής μέσης τιμής [82]

$$E[Y_1 Y_2 Y_3 Y_4] = E[Y_1 Y_4] E[Y_1 Y_3] + E[Y_1 Y_3] E[Y_2 Y_4] + E[Y_1 Y_2] E[Y_3 Y_4] \quad (2.35)$$

όπως και στο [22], με σκοπό τον περιορισμό των ροπών πολλαπλών συντελεστών.

$$\begin{aligned} R_{\Psi(N)}(\tau_1, \tau_2) &= R_{\Psi_x(N)}(\tau_1, \tau_2) + R_{\Psi_y(N)}(\tau_1, \tau_2) \\ &\quad + \underbrace{E[\Psi_x(N(x + \tau_1, y)) \Psi_y(N(x, y))]}_{E_1} + \underbrace{E[\Psi_x(N(x, y + \tau_2)) \Psi_y(N(x, y))]}_{E_2} \end{aligned} \quad (2.36)$$

Με λογισμούς στις αναμενόμενες τιμές των τελεστών, το άθροισμα στο δεύτερο μέλος, με το συμβολισμό της (2.31) καταλήγει

$$\begin{aligned} E_1 + E_2 = & 8\alpha_1^2\beta_1^2 + 2[R^{(1,1)} + 2R^{(0,1)}R^{(2,1)} \\ & + 2R^{(1,0)}R^{(1,2)} + R^{(2,0)}R^{(0,2)} + RR^{(2,2)}](\tau_1, \tau_2) \end{aligned} \quad (2.37)$$

ενώ για τους όρους αυτοσυσχέτισης

$$R_{\Psi_x(N)}(\tau_1, \tau_2) = 4\alpha_1^4 + [3(R^{(0,2)})^2 + 4R^{(0,1)}R^{(0,3)} + RR^{(0,4)}](\tau_1, \tau_2) \quad (2.38)$$

$$R_{\Psi_y(N)}(\tau_1, \tau_2) = 4\beta_1^4 + [3(R^{(2,0)})^2 + 4R^{(1,0)}R^{(3,0)} + RR^{(4,0)}](\tau_1, \tau_2) \quad (2.39)$$

Το σύνολο των (2.36)-(2.39) δίνει τη συνάρτηση αυτοσυσχέτισης $R_{\Psi(N)}$. Με χρήση της ανισότητας Cauchy-Schwarz [145]

$$|R(\tau_1, \tau_2)| \leq \frac{1}{\pi} \int_{\mathbb{R}^2} \Phi(\omega_x, \omega_y) d\omega_x d\omega_y = R(0, 0) \quad (2.40)$$

Επομένως η R έχει μέγιστο στο 0 οπότε $R^{(1,0)}(0, \tau_2) = 0, \forall \tau_2, R^{(0,1)}(\tau_1, 0) = 0, \forall \tau_1$ και ισχύουν $R^{(1,0)}(0, 0) = R^{(0,1)}(0, 0) = 0$ αλλά και $R^{(1,1)}(0, 0) = 0$. Με χρήση και των (2.32), η μεταβλητότητα του $\Psi(N)$ θα είναι

$$\begin{aligned} \text{Var}[\Psi(N)] = & R_{\Psi(N)}(0, 0) - E^2[\Psi(N)] \\ = & [2R^{(2,0)}R^{(0,2)} + 3((R^{(0,2)})^2 + (R^{(2,0)})^2) + R(R^{(4,0)} + R^{(0,4)} + 2R^{(2,2)})](0, 0) \\ = & 2\alpha_1^2\beta_1^2 + 3(\alpha_1^4 + \beta_1^4) + \gamma_0^2(2\gamma_1^2 + \alpha_2^2 + \beta_2^2) \end{aligned} \quad (2.41)$$

και αυξάνεται, όπως και η αναμενόμενη τιμή με την παρουσία μεγαλύτερων συχνοτήτων στο N . Για 1Δ σήμα N , $\beta_1 = \beta_2 = \gamma_1 = 0$ οπότε οι σχέσεις (2.33),(2.41) ανάγονται απευθείας στα 1Δ αντίστοιχα τους που αναφέρονται στο [22].

Παρόμοια τα βασικά στατιστικά της εφαρμογής του τελεστή στα διαφορικά N_x και N_y έχουν ενδιαφέρον για την εφαρμογή του ESA. Έτσι οι αναμενόμενες τιμές δίνονται από

$$\begin{aligned} E[\Psi(N_x)] = & 2R^{(4,0)}(0, 0) - R^{(1,1)}(0, 0) + R^{(2,2)}(0, 0) = 2\alpha_2^2 + \gamma_1^2 \\ E[\Psi(N_y)] = & 2R^{(0,4)}(0, 0) - R^{(1,1)}(0, 0) + R^{(2,2)}(0, 0) = 2\beta_2^2 + \gamma_1^2 \end{aligned} \quad (2.42)$$

και οι μεταβλητότητες

$$\begin{aligned} \text{Var}[\Psi(N_x)] = & 2\alpha_2^2\gamma_1^2 + 3(\alpha_2^4 + \gamma_1^4) - \alpha_1^2(2R^{(4,2)} + R^{(2,4)} + R^{(6,0)}) \\ \text{Var}[\Psi(N_y)] = & 2\beta_2^2\gamma_1^2 + 3(\beta_2^4 + \gamma_1^4) - \beta_1^2(2R^{(2,4)} + R^{(4,2)} + R^{(0,6)}) \end{aligned} \quad (2.43)$$

2.3.5.2 Ζωνοπερατή ενέργεια

Έστω ότι θεωρούμε τη ίδια τυχαία διαδικασία N η οποία φιλτράρεται από ένα γραμμικό ζωνοπερατό φίλτρο της μορφής

$$g_\sigma(x, y) = 2h_\sigma(x, y) \sin(\omega_{c1}x + \omega_{c2}y) \quad (2.44)$$

με κεντρική $\omega_c = (\omega_{c1}, \omega_{c2})$ και απόκριση συχνότητας

$$G_\sigma(\omega_1, \omega_2) = \frac{1}{j} [H_\sigma(\omega_1 - \omega_{c1}, \omega_2 - \omega_{c2}) + H_\sigma(\omega_1 + \omega_{c1}, \omega_2 + \omega_{c2})]. \quad (2.45)$$

Θεωρώντας $H(\vec{\omega})$ την απόκριση συχνότητας του αντίστοιχου βαθυπερατού φίλτρου και $\sigma = (\sigma_1, \sigma_2)$ μια παράμετρο κλίμακας που καθορίζει το εύρος και την κεντρική συχνότητα, τότε

$$H_\sigma(\vec{\omega}) = \frac{1}{\sqrt{|\sigma|}} H(\vec{\omega}/\sigma) \quad (2.46)$$

Για λόγους απλότητας έστω ότι η κρουστική h είναι άρτια συμμετρική έτσι ώστε $G_\sigma(0, 0) = 0$, οι θετικές και αρνητικές συνιστώσες του $G_\sigma(\omega)$ να μην επικαλύπτονται και το φίλτρο να είναι ισοτροπικό, δηλ. $\sigma_1 = \sigma_2$. Τότε

$$|G_\sigma(\vec{\omega})|^2 = |H_\sigma(\vec{\omega} - \vec{\omega}_c)|^2 + |H_\sigma(\vec{\omega} + \vec{\omega}_c)|^2 \quad (2.47)$$

Το φιλτραρισμένο σήμα θα είναι $N_\sigma(x, y) = (N * g_\sigma)(x, y)$, όπου $*$ δηλώνει συνέλιξη. Από τη σχέση (2.34) που δίνει την αναμενόμενη τιμή του τελεστή στη διαδικασία N προκύπτει η γενίκευση για την ζωνοπερατά φιλτραρισμένη εκδοχή του $\Psi(N_\sigma(x, y))$

$$\begin{aligned} \mathbb{E}[\Psi(N_\sigma)] &= \frac{1}{\pi} \int |\vec{\omega}|^2 |G_\sigma(\vec{\omega})|^2 \Phi(\vec{\omega}) d\vec{\omega} \\ &= \frac{1}{\pi} \int_{\mathbb{R}^2} |\vec{\omega}|^2 (|H_\sigma(\vec{\omega} - \vec{\omega}_c)|^2 + |H_\sigma(\vec{\omega} + \vec{\omega}_c)|^2) \Phi(\vec{\omega}) d\vec{\omega} \\ &= \frac{1}{\pi\sigma} \int_{\mathbb{R}^2} |\vec{\omega}|^2 (|H((\vec{\omega} - \vec{\omega}_c)/\sigma)|^2 + |H((\vec{\omega} + \vec{\omega}_c)/\sigma)|^2) \Phi(\vec{\omega}) d\vec{\omega} \end{aligned} \quad (2.48)$$

Στην περίπτωση ιδανικού ζωνοπερατού τετραγωνικού παλμού με $H(\vec{\omega}) = 1, |\vec{\omega}| \leq |\vec{\omega}_c|$ τότε προσεγγιστικά η μέση τιμή του φιλτραρισμένου σήματος είναι

$$\mathbb{E}[\Psi(N_\sigma)] \cong \frac{1}{\pi\sigma} \int_{\mathbb{R}^2} |\vec{\omega}|^2 \Phi(\vec{\omega}) d\vec{\omega} = \frac{1}{\sigma} \mathbb{E}[\Psi(N)]. \quad (2.49)$$

Γενικά, και για μη ισοτροπικά φίλτρα με $\sigma_1/\sigma_2 = \lambda$,

$$\mathbb{E}[\Psi(N * g_\sigma)] \cong \frac{\lambda}{\sigma} \mathbb{E}[\Psi(N)] \quad (2.50)$$

η αναμενόμενη μέση τιμή μεταβάλλεται γραμμικά, αντιστρόφως ανάλογα με τη θετική σταθερά που καθορίζει το εύρος του ζωνοπερατού φιλτραρίσματος. Μείωση του σ , δηλ. περιορισμός της ενεργής ζώνης του φίλτρου αυξάνει τη μέση τιμή του ενεργειακού τελεστή. Κάτι τέτοιο οφείλεται διαισθητικά αφενός στον περιορισμό των εντός ζώνης ταλαντώσεων και αφετέρου, εφόσον η διαδικασία είναι λευκός θόρυβος, στην εξ' ορισμού ιδιότητα του ζωνοπερατού φιλτραρίσματος να περιορίζει τον ευρυζωνικό θόρυβο.

2.4 Τεχνικές ανάλυσης υφής

Οι τεχνικές που περιγράφονται σύντομα στη συνέχεια σχετίζονται άμεσα με την αναζήτηση μικροδομών και την περιγραφή της υφής σε ένα τέτοιο επίπεδο. Καθιερωμένα μοντέλα υφής που βασίζονται σε στοχαστικά φράκταλ, τυχαία πεδία και διαχωρισμό της λεπτομέρειας από την εικόνα θα ενταχθούν στο πλαίσιο της αναζήτησης και της ανάλυσης των μικροδιαμορφώσεων της υφής.

2.4.1 Στοχαστικά φράκταλ

Τα μοντέλα φράκταλ (fractal) είναι δημοφιλή για ανάλυση υφής γιατί προσφέρουν τρόπους ποσοτικοποίησης και μοντελοποίησης, με μικρό αριθμό παραμέτρων, της τραχύτητας μιας επιφάνειας. Ένα μέγεθος που χρησιμοποιείται κατά κόρον σε μετρήσεις υφής είναι η φράκταλ διάσταση (fractal dimension) [7, 118, 150, 151]. Η μέτρηση της γίνεται μαθηματικά σε διάφορες κλίμακες και ποσοτικοποιεί την αυτοομοιότητα στο εύρος των κλιμάκων [93, 96, 185]. Σε πολλές κατηγορίες κυρίως φυσικών υφών, περιορίζοντας το εύρος των υπαρκτών και μετρήσιμων κλιμάκων, οι επιφάνειες, τα περιγράμματα τους αλλά και τυχαίοι σχηματισμοί υφής είναι φράκταλ. Η φράκταλ διάσταση καμπυλών παίρνει τιμές $1 \leq D \leq 2$, ενώ για επιφάνειες που η τοπολογική τους διάσταση είναι 2, είναι $2 \leq D \leq 3$. Εκτός από τη φράκταλ διάσταση κι άλλα μεγέθη έχουν χρησιμοποιηθεί τη διερεύνηση των ιδιοτήτων της υφής όπως η «κενότητα» (lacunarity) [34, 96, 118], που ανιχνεύει και ποσοτικοποιεί στατιστικά δεύτερης τάξης φράκταλ επιφανειών.

Οι περισσότερες φυσικές επιφάνειες με υφή ή σχηματισμούς και μικροδομές δεν είναι ντετερμινιστικές αλλά διαθέτουν μια στατιστική μεταβολή [101, 152]. Επομένως η μοντελοποίηση τους ως ντετερμινιστικά φράκταλ και η μέτρηση της διάστασης τους είναι δύσκολη. Εναλλακτικά παρουσιάζεται η μοντελοποίηση υφής με στοχαστικά φράκταλ και ειδικά μια κατηγορία $1/f$ θόρυβων (1-over-f noises), τις κλασματικές κινήσεις Brown (fractional Brownian motion) [119]. Οι $1/f$ θόρυβοι είναι στοχαστικές διαδικασίες με φάσμα ισχύος αντιστρόφως ανάλογο μιας δύναμης της συχνότητας. Για τα fBm ο εκθέτης παίρνει τιμές από 1 μέχρι 3, και αποτελούν γενικευμένες επεκτάσεις της κίνησης Brown.

Οι 2Δ επεκτάσεις των fBm [93, 151, 152, 154] είναι διαδικασίες $B_H(x, y)$ που ορίζονται σε ένα διδιάστατο χωρικό πλέγμα και παρουσιάζουν στατιστική αυτοομοιότητα σε πολλαπλές κλίμακες, δηλαδή στατιστικά παρόμοια συμπεριφορά ανεξάρτητα από το διάστημα στο οποίο αυτή εξετάζεται. Η συμπεριφορά σε διαφορετικές κλίμακες, εξαρτάται από την παράμετρο H (Hurst exponent), που καθορίζει πόσο λεπτομερές ή τραχύ είναι το αποτέλεσμα της διαδικασίας σε μικρές περιοχές. Οι αλλαγές τιμών μετά από χωρική μετατόπιση κατά $(\Delta x, \Delta y)$ καθορίζονται τόσο από το H όσο και από την ίδια τη μετατόπιση και η μεταβλητότητα της διαδικασίας είναι

$$\text{Var}[B_H] = \Delta B_H = B_H(x + \Delta x, y + \Delta y) - B_H(x, y) \propto (\Delta x^2 + \Delta y^2)^H \quad (2.51)$$

Το 2Δ φάσμα ισχύος μιας τέτοιας διαδικασίας εξαρτάται αντίστροφα από μια δύναμη του μέτρου $|\vec{\omega}|$ του διανύσματος χωρικών συχνοτήτων $\vec{\omega} = (u, v)$

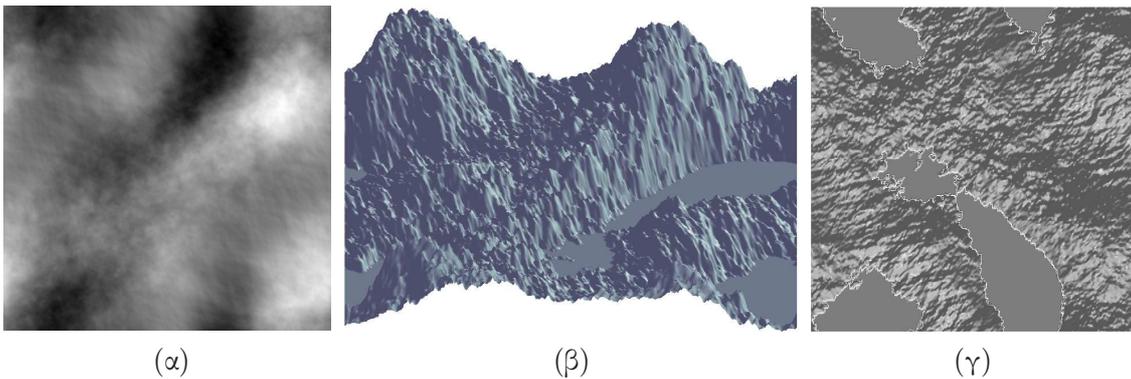
$$S_H(u, v) \propto 1/(u^2 + v^2)^{(H+1)/2} \quad (2.52)$$

και η φράκταλ διάσταση της διαδικασίας εξαρτάται από το H

$$D = 3 - H, \quad 0 \leq H \leq 1 \quad (2.53)$$

Η χρήση και ο έλεγχος ενός τέτοιου μοντέλου υφής γίνεται με δύο τρόπους. Ανάλυση οπτικής υφής επιφανειών και μέτρηση του συντελεστή H , και δημιουργία συνθετικών εικόνων υφής από τυχαία παραγόμενα fBm. Η εκτίμηση των παραμέτρων του μοντέλου και η μέτρηση της πολυπλοκότητας της υφής γίνεται με γεωμετρικές μεθόδους [96], στατιστικές μεθόδους γραμμικής παλινδρόμησης [93] και εκτιμητές μέγιστης πιθανοφάνειας [110] και μεθόδους μαθηματικής μορφολογίας [121]. Για σύνθεση διδιάστατων διαδικασιών fBm (και των παραγώγων τους fractional Gaussian noises ή fGn) χρησιμοποιείται μια πληθώρα μεθόδων από τις οποίες αναφέρονται ενδεικτικά οι random midpoint displacement, random cuts, Fourier spectrum synthesis [201]. Η τρίτη μέθοδος περιλαμβάνει δημιουργία του 2Δ φάσματος ισχύος μιας υποθετικής διαδικασίας που υπακούει στην 2.52 και αντίστροφο FFT μετασχηματισμό για να προκύψει η χωρικά μεταβαλλόμενη αναπαράσταση. Όσο μεγαλύτερη επιλέγεται η διάσταση (μικρό H) της διαδικασίας, τόσο περισσότερο «θρυμματισμένη» φαίνεται η συνθετική επιφάνεια.

Στο σχήμα 2.4 φαίνεται ένα παράδειγμα σύνθεσης υφής φράκταλ διάστασης 2.4, στο Fourier πεδίο, με επιλογή της παραμέτρου $H = 0.6$. Η διαδικασία απεικονίζεται ως (α) αναπαράσταση γκριζας εικόνας, (β) 3Δ προοπτική απεικόνιση σαν φυσική σκηνή και (γ) επιφάνεια που φαίνεται κατακόρυφα.



Σχήμα 2.4: Σύνθεση (α) υφής, (β) φυσικών τοπίων και (γ) επιφανειών με προσομοίωση 2Δ στοχαστικών φράκταλ (fractional Brownian motion). Η σύνθεση έγινε με δημιουργία των συντελεστών του 2Δ Fourier σε χωρικό πλέγμα 256×256 και η φράκταλ διάσταση των επιφανειών επιλέχθηκε $D = 2.4$ ($H = 0.6$).

2.4.2 Τυχαία πεδία

Τα Μαρκοβιανά Τυχαία Πεδία (Markov Random Fields) και οι ισοδύναμες κατανομές Gibbs μοντελοποιούν στατιστικά φυσικές και τεχνητές εικόνες υφής. Βασίζονται στην υπόθεση ότι η στάθμη του γκριζου σε κάθε σημείο του χωρικού πλέγματος της εικόνας, η οποία αντιμετωπίζεται σαν στοχαστικό διδιάστατο πεδίο, εξαρτάται μόνο από τις τιμές στα γειτονικά του pixel. Οι παράμετροι αυτού του στοχαστικού, χωρικά ομογενούς πεδίου, όπως το σύστημα της γειτονίας, οι δυνατές στάθμες φωτεινότητας και τα βάρη στην αλληλεπίδραση των σωματιδίων-pixel καθορίζουν την ακρίβεια του μοντέλου [64, 204].

Η θεωρία των MRF θεμελιώθηκε από τον Besag [15, 16] με ιδέες από τη χωρική στατιστική και τη στατιστική μηχανική. Η ισοδυναμία της περιγραφής με τα Gibbs Τυχαία Πεδία ή GRF, μέσω της ένα προς ένα αντιστοιχίας που θεμελιώνει το θεώρημα

Hammersley-Clifford, επιτρέπει την περιγραφή σχηματισμών υφής με όρους συντελεστών χωρικής εξάρτησης σε ένα πεπερασμένο πλέγμα [41, 48, 54, 62, 120].

2.4.2.1 Ορισμοί

Για ένα τυχαίο πεδίο X σε ένα τετραγωνικό πλέγμα $E = \{s = (i, j) \in \mathbb{Z}^2 : 1 \leq i \leq M_1, 1 \leq j \leq M_2\}$ η τιμή X_s σε κάθε θέση (site) s είναι μια τυχαία μεταβλητή¹. Ορίζεται μια γειτονιά \mathcal{N}_s με κέντρο στο s με τις ακόλουθες ιδιότητες

$$s \neq \mathcal{N}_s \quad \text{and} \quad s \in \mathcal{N}_q \iff q \in \mathcal{N}_s$$

Η 4-pixel γειτονιά $\mathcal{N}_s^{(1)}$ περιλαμβάνει τους οριζόντιους και κάθετους γείτονες γύρω από το s , ενώ η $\mathcal{N}_s^{(2)}$ 8-pixel καλύπτει και τις διαγώνιες. Τα υποσύνολα μιας r -τάξης γειτονιάς $\mathcal{N}^{(r)}$ ονομάζονται κλίκες (cliques).

Ένα τυχαίο πεδίο X καλείται Μαρκοβιανό Τυχαίο Πεδίο (MRF), τάξης r , αν 1) $P(X = x) > 0$ για κάθε x , 2) Για κάθε θέση s , η υπό συνθήκη πιθανότητα του X_s δεδομένων των τιμών του X σε κάθε θέση $E \setminus \{s\}$ εξαρτάται μόνο από τις τιμές του X στη γειτονιά $\mathcal{N}_s^{(r)}$:

$$P(x_s | x_q, q \in E \setminus \{s\}) = P(x_s | x_q, q \in \mathcal{N}_s^{(r)}) \quad (2.54)$$

Από τη στατιστική μηχανική ένα τυχαίο πεδίο X καλείται Gibbs Τυχαίο Πεδίο (GRF), αν η κατανομή του ακολουθεί την Gibbs-Boltzmann εκθετική κατανομή

$$P(X = x) = \frac{1}{Z} \exp[-U(x)], \quad \forall x, \quad (2.55)$$

όπου $U(x)$ είναι η ενέργεια της πραγματοποίησης x του πεδίου και η σταθερά κανονικοποίησης Z ονομάζεται συνάρτηση διαχωρισμού (partition function)

$$Z = \sum_x \exp[-U(x)]$$

Η ενεργειακή ποσότητα μπορεί να γραφεί και σαν άθροισμα συναρτήσεων V_c σχετιζόμενων με διαφορετικές κλίκες c , που καλούνται συναρτήσεις δυναμικού (potential functions), στη μορφή

$$U(x) = \frac{1}{T} \sum_{c \in \mathcal{C}} V_c(x)$$

Η παράμετρος T είναι ανάλογη της απόλυτης θερμοκρασίας του φυσικού συστήματος του οποίου η κατάσταση ακολουθεί την ανωτέρω διασπορά. Η φυσική διαίσθηση είναι ότι για μεγάλα T η κατανομή είναι περισσότερο ομοιόμορφη και πολλές πραγματοποιήσεις είναι πιθανές.

Με κατάλληλη επιλογή των συναρτήσεων δυναμικού, το βασικό μοντέλο μπορεί να δώσει ως ειδικές περιπτώσεις μια ποικιλία κλάσεων τυχαίων πεδίων, όπως τα μοντέλα διακριτών τιμών δυαδικό (binary) autologistic, πλειότιμο (multi-valued) autologistic, διωνυμικό (binomial), μοντέλο Poisson και τα συνεχή μοντέλα όπως τα Gaussian τυχαία πεδία.

¹Οι ορισμοί εδώ «μεταφέρονται» από το [122].

2.4.2.2 Μοντέλα υφής

Οι παράμετροι ενός επιλεγμένου MRF μοντέλου για ανάλυση και σύνθεση υφής είναι το σύστημα γειτονιάς, ο αριθμός των διακριτών γκριζών τιμών και τα βάρη εξάρτησης στα υποσύνολα του \mathcal{C} . Έστω μια γειτονιά 2ης τάξης (8-σημεία) του x_s στις θέσεις x_i , $i = 1 \dots 8$

$$\mathcal{N}_s = \begin{array}{|c|c|c|} \hline x_5 & x_1 & x_6 \\ \hline x_4 & x_s & x_2 \\ \hline x_8 & x_3 & x_7 \\ \hline \end{array}$$

και $U(x_s)$ η συνεισφορά στη συνάρτηση δυναμικού της θέσης x_s . Κάποιες ειδικές περιπτώσεις μοντέλων διακριτών τιμών, δίνονται εν συντομία με τη μορφή της συνάρτησης ενέργειας τους. Η επιλογή των παραμέτρων α_s, β_i καθορίζει την ομοιογένεια (homogeneity) και την κατευθυντικότητα (directionality) του πεδίου.

Auto-logistic [15]: Δυαδικές τιμές $X_s \in \{0, 1\}$, κλίκες σε όλες τις διευθύνσεις

$$U(x_s) = \alpha_s x_s + \sum_{i=1}^9 \beta_i x_s x_i \quad (2.56)$$

Ising: Δυαδικές τιμές $X_s \in \{-1, 1\}$, κάθετες και οριζόντιες κλίκες

$$U(x_s) = \alpha_s x_s + \sum_{i=1}^4 \beta_i x_s x_i \quad (2.57)$$

Auto-binomial [15, 41]: Διακριτές τιμές $X_s \in \{0 \dots M\}$, κλίκες σε όλες τις διευθύνσεις

$$U(x_s) = -\ln \binom{M-1}{x_s} - \alpha_s x_s - \sum_{i=1}^9 \beta_i x_s x_i \quad (2.58)$$

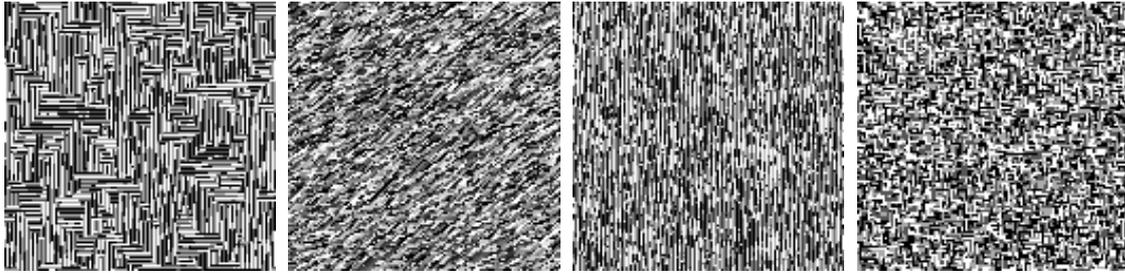
Multi-level logistic [47, 48]: Διακριτές τιμές $X_s \in \{0 \dots M\}$, κλίκες σε όλες τις διευθύνσεις

$$U(x_s) = \alpha_s x_s + \sum_{i=1}^9 \beta_i [2\delta(x_s - x_i) - 1] \quad (2.59)$$

Auto-normal και άλλα συνεχή μοντέλα (*continuous-valued models*) (π.χ. Gaussian MRFs [32]): $X_s = \mathbb{R}$

$$U(x_s) = \frac{(x_s - \mu_s)^2}{2\sigma^2} + \sum_{i=1}^9 \beta_i \frac{(x_s - \mu_s)(x_i - \mu_i)}{2\sigma^2} \quad (2.60)$$

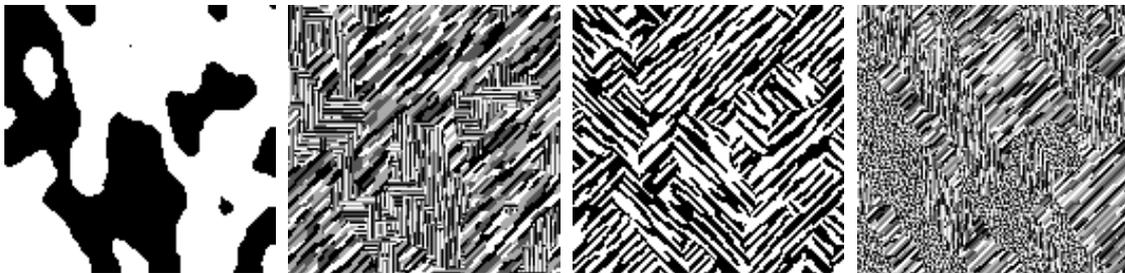
Για τη σύνθεση MRF διαδικασιών που προσομοιώνουν φυσικές υφές έχουν αναπτυχθεί αλγόριθμοι δειγματοληψίας κατανομών Gibbs (2.55). Τέτοιες διαδικασίες βασίζονται σε δειγματοληψία Markov Chain Monte Carlo όπως οι αλγόριθμοι Metropolis [134] και Hastings [69] ή ο Gibbs sampler [62], συχνά και με την προσθήκη τεχνικών βελτιστοποίησης όπως η μέθοδος simulated annealing [62]. Οι συνθετικές υφές μπορεί να παρουσιάζουν φαινόμενα συγχέντρωσης (clustering), ανισοτροπίας (anisotropy), φαινόμενα έλξης - απώθησης (attraction-repulsion effects) ή διατεταγμένα πρότυπα (ordered patterns) [41, 48]. Μερικές συνθετικές υφές χρησιμοποιώντας



Σχήμα 2.5: Εικόνες συνθετικής υφής για διαφορετικά βάρη κλικών και γκρίζες στάθμες με το μοντέλο multi-level logistic. Εκτός από την πρώτη, οι υπόλοιπες έχουν 4 γκρίζες στάθμες και συνδυασμούς συντελεστών ώστε να προσδίδουν συγκεκριμένο προσανατολισμό.

το multi-level logistic μοντέλο φαίνονται στο σχήμα 2.5. Μέθοδοι κατηγοριοποίησης ή κατάτμησης υφής με μοντέλα MRF βασίζονται στην εκτίμηση των παραμέτρων της αντίστοιχης GRF κατανομής [32, 36], με τεχνικές όπως Bayesian estimation, maximum likelihood [37], coding-maximum pseudolikelihood [16, 104], least squares [48] κ.α.

Αξίζει τέλος να αναφερθούν δύο παραλλαγές των βασικών μοντέλων. Τα ιεραρχικά πεδία (hierarchical MRFs) [48] αποτελούν διαδικασίες σε πολλαπλά επίπεδα που ελέγχουν 1) Μεγάλες περιοχές της εικόνας, 2) Υφή ή θόρυβο εντός των περιοχών. Δύο ταυτόχρονες διαδικασίες εκτιμώνται και δύο ή περισσότερα σύνολα παραμέτρων υπολογίζονται κατά τη διάρκεια της μάθησης. Το κάτω επίπεδο, δίνει ένα πρόχειρο χωρισμό της εικόνας σε περιοχές ομοιόμορφης υφής και το πάνω επίπεδο δίνει τη λεπτομέρεια ανά περιοχή. Σύνθεση τέτοιων εικόνων με υπολογιστικές προσομοιώσεις φαίνεται στο σχήμα 2.6 με κάτωτερο το δυαδικό πεδίο της αριστερής εικόνας. Τα GRF



Σχήμα 2.6: Εικόνες συνθετικής υφής με πολυεπίπεδα τυχαία πεδία και ιεραρχική δομή. Το πρώτο δυαδικό πεδίο είναι το επίπεδο πάνω στο οποίο τα διαφορετικά MRF προσθέτουν λεπτομέρεια και υφή. Δύο τελικά αποτελέσματα φαίνονται στις μεσαίες εικόνες, ενώ η τελευταία δεξιά αντιστοιχεί σε σύνθεση με τριαδικό κατώτερο επίπεδο.

με ενεργειακούς περιορισμούς (constrained MRFs) [176] είναι τυχαία πεδία των οποίων οι κατανομές είναι της μορφής

$$P(X = x) = \frac{1}{Z} \exp[-(1 - \lambda)U(x) - \lambda V(x)], \quad \forall x, \quad (2.61)$$

όπου η $V(x)$ μια συνάρτηση που ελέγχει γεωμετρικούς περιορισμούς στην εικόνα. Ελαχιστοποίηση της ενέργειας σημαίνει ταυτόχρονα και ικανοποίηση των περιορισμών που έχουν να κάνουν με τη δομή και τους σχηματισμούς του πεδίου. Τέτοια πεδία έχουν εφαρμοστεί με μορφολογικούς περιορισμούς για μοντελοποίηση εικόνων μέσω της κατανομής μεγεθών των δομών της.

2.4.3 Διαχωρισμός υφής

Μια πρόσφατη έννοια και πεδίο έρευνας σε θέματα ανάλυσης εικόνων είναι ο διαχωρισμός (decomposition) μια δοσμένης εικόνας $f : \Omega \rightarrow \mathbb{R}$ σε δυο συνιστώσες $u + v$ όπου το u καλείται «καρτούν» συνιστώσα (cartoon component), μια απλοποιημένη τμηματικά ομαλή προσέγγιση και αποτελείται από σχετικά επίπεδες περιοχές της εικόνας, σταθερής φωτεινότητας ², με απότομες ακμές ενώ το v κομμάτι, οι συναρτήσεις ταλάντωσης ή «ταλαντώσεις υφής», αποτελείται από υφή και θόρυβο. Σκοπός είναι ο διαχωρισμός της υφής από την αρχική εικόνα για περαιτέρω επεξεργασία, ανάλυση αλλά και μια πρώτη κατάτμηση. Το μοντέλο έχει βρει πολλούς υποστηρικτές και διάφορες μεθοδολογίες προτείνονται για την εξαγωγή των u και v συνιστωσών [4, 144, 182, 195].

Πολλές από τις μη-γραμμικές, τεχνικές ομαλοποίησης εικόνων που διατηρούν τις ακμές μπορούν να προσεγγίσουν την u συνιστώσα. Αυτά τα μοντέλα αναπαριστούν τη u σαν μια συνάρτηση φραγμένης μεταβολής (bounded variation) σε ένα χώρο $BV(\Omega)$ επιτρέποντας ασυνέχειες κατά μήκος καμπυλών και επομένως ακμές και περιγράμματα. Η διαφορά $v := f - u$ μοντελοποιείται από μια συνάρτηση στο $L^2(\Omega)$. Μια σύνοψη τέτοιων μοντέλων μπορεί να βρεθεί στο [139] και αποδεικνύεται ότι είναι ειδικές περιπτώσεις ή στενά συνδεδεμένες με το συναρτησιακό Mumford-Shah [140]

$$E_{MS}(u, C) = \int_{\Omega \setminus C} (|\nabla u|^2 + \lambda(u - f)^2) dx dy + \mu \text{Len}(C) \quad (2.62)$$

Μια τέτοια προσέγγιση είναι η τεχνική ελαχιστοποίησης της ολικής μεταβολής (total variation (TV) minimization) των Rudin-Osher-Fatemi [168] που υπολογίζει το «καρτούν» με επίλυση του ακόλουθου προβλήματος ελαχιστοποίησης

$$E_{ROF}(u) = \int_{\Omega} |\nabla u| dx dy + \lambda \int_{\Omega} |f - u|^2 dx dy \quad (2.63)$$

όπου $BV(\Omega)$ ο χώρος συναρτήσεων φραγμένης μεταβλητότητας, λ μια θετική παράμετρος και

$$|u|_{TV} = \int_{\Omega} |\nabla u| dx dy \quad (2.64)$$

είναι η ολική μεταβολή του u .

Η μέθοδος TV δουλεύει καλά για αποθορυβοποίηση και ομαλοποίηση εικόνων αλλά δε διατηρεί την υφή για μικρά λ . Ο Meyer [137], πρότεινε τον επαναπροσδιορισμό του TV προβλήματος με χρήση διαφορετικών νορμών και εισήγαγε το διαχωρισμό σε u και v , που το v κομμάτι μοντελοποιεί ταλαντώσεις στις δύο χωρικές διευθύνσεις (και πιθανώς θόρυβο) και μπορεί να εκφραστεί ως $v = \text{div} \vec{g}$, όπου $\vec{g} = (g_1, g_2)$ ένα διανυσματικό πεδίο

$$\vec{g} = -\frac{1}{2\lambda} \left(\frac{\nabla u}{|\nabla u|} \right), \quad v = \frac{\partial g_1}{\partial x} + \frac{\partial g_2}{\partial y} \quad (2.65)$$

Οι Vese, Osher [195, 196] προτείνανε ένα νέο μοντέλο για διαχωρισμό εικόνων σε

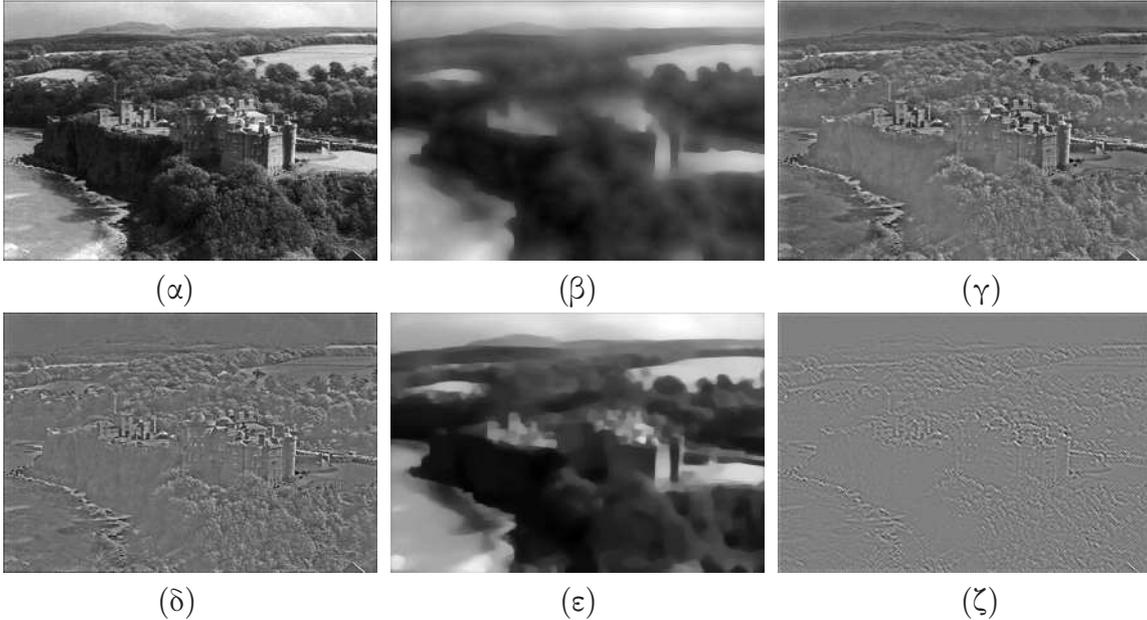
²Στα κόμικς, την *αόρατη τέχνη*, “το «καρτούν» είναι ένας τρόπος ενίσχυσης μέσω απλοποίησης (amplification through simplification). Μετατρέποντας σε «καρτούν» μια εικόνα, δεν αφαιρούμε λεπτομέρεια αλλά επικεντρωνόμαστε σε συγκεκριμένες λεπτομέρειες.” [133].

ομαλή συνιστώσα και υφή, ελαχιστοποιώντας ως προς (u, g_1, g_2) το

$$E_{VO}(u, \vec{g}) = \int_{\Omega} |\nabla u| dx dy + \lambda \int_{\Omega} |f - (u + \partial_x g_1 + \partial_x g_2)|^2 dx dy + \mu \int_{\Omega} (\sqrt{g_1^2 + g_2^2})^p dx dy$$

(2.66)

όπου $u \in BV(\Omega)$ και $v = f - u \approx \operatorname{div} \vec{g}$ ενώ οι παράμετροι $\lambda, \mu > 0$ ελέγχουν τη σύγκλιση και το $p \geq 1$ καθορίζει τη νόρμα του \vec{g} . Για $p = 2$ προκύπτει το μοντέλο [144]. Στο [195] παρουσιάζονται αριθμητικά, επαναληπτικά σχήματα, που προσεγγίζουν τις ασθενέστερες νόρμες του Meyer, για την εκτίμηση των Euler-Lagrange εξισώσεων-λύσεων της (2.66).



Σχήμα 2.7: Διαχωρισμός υφής με $u + v$ μοντέλα. Πάνω σειρά: (α) Αρχική εικόνα f , (β) u_{TV} συνιστώσα με ελαχιστοποίηση ολικής μεταβολής (total variation minimization) των Rudin-Osher-Fatemi [168], για 100 επαναλήψεις με δυναμικό λ , (γ) v_{TV} συνιστώσα όπου $v_{TV} = f - u_{TV}$ συν μια σταθερά $c = 100$ για λόγους απεικόνισης. Κάτω σειρά: $u + v$ αποσύνθεση με $\lambda = \mu = 0.01$ και 100 επαναλήψεις. (ε),(ζ) u_{VO} και v_{VO} συνιστώσες με το μοντέλο Vese-Osher [195], (δ) $f - u_{VO}$ συν μια σταθερά $c = 100$.

Μοντέλα διαχωρισμού σε δύο συνιστώσες προτάθηκαν επίσης στο [182, 183] με χρήση λεξικών για την περιγραφή της υφής και στο [3] εμπνευσμένα από τα [137, 195]. Το ιδανικό μοντέλο θα διαχώριζε μια εικόνα σε τρεις συντελεστές $u + v + w$, όπου η συνιστώσα u συνιστώσα κρατάει τη δομή και τη γεωμετρία, η v την υφή της εικόνας και η w το θόρυβο. Πρόσφατα μεταβολικά μοντέλα αναπτύχθηκαν για τέτοιο τριπλό διαχωρισμό [4] όπου η αποσύνθεση γίνεται ελαχιστοποιώντας ενεργειακά με διαφορετικές νόρμες για κάθε συνιστώσα, επεκτείνοντας στην ουσία τα προηγούμενα μοντέλα.

Το εύρος των εφαρμογών του διαχωρισμού της γεωμετρικής δομής μιας εικόνας από την υφή της και ιδανικά το θόρυβο, εκτείνεται σε θέματα όπως αποθρομβοποίηση, ομαλοποίηση και κανονικοποίηση εικόνων, κωδικοποίηση, αποσύνθεση μιας εικόνας σε αναπαραστάσεις διαφορετικής σημασιολογίας, διάκριση και κατάτμηση υφής μέσω της απομονωμένης συνιστώσας της και συμπλήρωση χαμένης υφής ή δομής (texture/structure in-painting) σε σύνθετες εικόνες [12–14]. Στο σχήμα 2.7 φαίνεται η

αποσύνθεση της εικόνας (α), που διαθέτει δομή και υφή, στις δύο συνιστώσες με την TV [168] και την $u + v$ [195, 196] προσέγγιση. Η πρώτη, μετά από 100 επαναλήψεις, θολώνει πολύ τα περιγράμματα των αντικειμένων και ομαλοποιεί υπερβολικά την u συνιστώσα στο (β), ενώ η v συνιστώσα στο (γ) λαμβάνεται ως διαφορά από την αρχική. Η μέθοδος Meyer/Vese-Osher κρατάει καλύτερα τη γεωμετρία στο u κομμάτι, που φαίνεται στο (ϵ), και σχεδόν αφήνει μόνο τη λεπτομέρεια της υφής στο v που φαίνεται στην εικόνα (ζ). Συγκρινόμενη με τη διαφορά $f - u$ στο (δ), φαίνεται ότι από το v , που έχει ληφθεί ταυτόχρονα από τη διαδικασία ελαχιστοποίησης, απουσιάζουν χαρακτηριστικές δομές ή περιγράμματα αντικειμένων.

□

Κεφάλαιο 3

Ενέργεια και Ανάλυση Υφής

Η υφή επιφανειών αναλύεται μοντελοποιώντας τις μεταβολές μιας εικόνας ως τοπικές συνιστώσες διαμόρφωσης, χρησιμοποιώντας φιλτράρισμα σε πολλαπλές συχνότητες και μη-γραμμικούς τελεστές στιγμιαίου εντοπισμού ενέργειας. Η δραστηριότητα υφής εκφράζεται με εκτίμηση της ενέργειας χωρικών διαμορφώσεων στις κυρίαρχες ζώνες συχνότητων. Παρουσιάζονται οι μέθοδοι πολυκαναλικής ανάλυσης και εξαγωγής χαρακτηριστικών διαμόρφωσης και ανάλυσης κυριαρχών συνιστωσών. Το πρόβλημα της ανάλυσης υφής ανάγεται σε ανίχνευση ενεργειακών όρων και περιγραφή με έντονα σε πληροφορία διανύσματα χαρακτηριστικών χαμηλής διάστασης.

3.1 Διαμορφώσεις υφής

Στοιχειώδεις συνιστώσες οπτικής υφής μπορούν να ερμηνευθούν σαν τοπικά ομαλές διαμορφώσεις, και επομένως μη-στατικά σήματα, καλά εντοπισμένα σε μια στενή ζώνη χωρικών συχνότητων, της μορφής (2.7) που επαναλαμβάνεται εδώ για λόγους αναφοράς

$$f(x, y) = a(x, y) \cos[\phi(x, y)] , \quad \vec{\omega}(x, y) = \nabla\phi(x, y) \quad (3.1)$$

όπου $a(x, y)$ το σήμα διαμόρφωσης πλάτους και $\phi(x, y)$ το σήμα διαμόρφωσης φάσης. Δεδομένου ενός σήματος f υπάρχουν άπειροι τρόποι να ορίσει κανείς τα σήματα a και ϕ . Παρ' όλα αυτά μόνο συγκεκριμένες ερμηνείες τους έχουν νόημα στη μοντελοποίηση εικόνων τοπικά στενής ζώνης (locally narrowband). Το πλάτος $a(x, y)$ μοντελοποιεί την αντίθεση φωτεινότητας (contrast) και το χωρικά μεταβαλλόμενο διάνυσμα $\vec{\omega}(x, y) = (\omega_1, \omega_2)(x, y)$ περιέχει πληροφορία για τις τοπικά αναδυόμενες χωρικές συχνότητες [21]. Τέτοιες συχνότητες μεταβάλλονται ομαλά σε μια στενή φασματική συχνότητων.

Κατά συνέπεια, περιοχές μιας εικόνας που χαρακτηρίζονται από υφή μοντελοποιούνται ως υπέρθεση χωρικών σημάτων AM-FM διαμόρφωσης

$$I(x, y) = \sum_{k=1}^K a_k(x, y) \cos[\phi_k(x, y)] , \quad \vec{\omega}_k(x, y) = \nabla\phi_k(x, y) \quad (3.2)$$

όπου ο κάθε ένας από τους K συντελεστές είναι ένα 2Δ μη-στατικό ημίτονο, με χωρικά μεταβαλλόμενο πλάτος $a_k(x, y)$ και χωρικά μεταβαλλόμενο διάνυσμα στιγμιαίων συχνότητων $\vec{\omega}_k(x, y)$, το οποίο μπορεί να αναλυθεί στα δύο επιμέρους σήματα, στην οριζόντια και κάθετη διεύθυνση της εικόνας

$$(\omega_{1k}(x, y), \omega_{2k}(x, y)) = \left(\frac{\partial\phi_k}{\partial x}, \frac{\partial\phi_k}{\partial y} \right) (x, y) \quad (3.3)$$

Οι στιγμιαίες χωρικές συχνότητες σχετίζονται με την κλίμακα, τον προσανατολισμό (orientation) και την τραχύτητα (coarseness) της προκύπτουσας υφής. Τέτοια μοντέλα διαμόρφωσης για την υφή προτάθηκαν στα [21, 73, 74] με φραγμένα σφάλματα προσέγγισης γύρω από την ομαλότητα των διαμορφώσεων και εφαρμόστηκαν σε μια πληθώρα προβλημάτων ανάλυσης εικόνων και όρασης υπολογιστών.

Οι τοπικές μεταβολές της υφής σε μια εικόνα μοντελοποιούνται με πεπερασμένο αριθμό χωρικά διαμορφωμένων σημάτων τα οποία διαχωρίζονται μέσα από ζωνοπερατό φίλτράρισμα σε πολλαπλές ζώνες συχνοτήτων. Οι διαμορφώσεις αυτές, σε πολλαπλές κλίμακες, αντιστοιχούν σε ξεχωριστά ποσά φυσικής ενέργειας και κατ' επέκταση συμβάλλουν ανομοιόμορφα στις υπαρκτές, ξεχωριστές υφές στην εικόνα. Για την ανάλυση της κάθε συνιστώσας σε συγκεκριμένο ζεύγος AM και FM συναρτήσεων [193] ξεχωρίζουν δύο προσεγγίσεις. Η πρώτη, ενεργεί απευθείας στις τιμές της εικόνας $f_k(x, y)$ και σχετίζεται με τον 2Δ ενεργειακό τελεστή 2.11 και τον αλγόριθμο διαχωρισμού ενέργειας ESA (2.20, 2.21) [123] που περιγράφονται στην ενότητα (2.3). Η δεύτερη περιλαμβάνει τη δημιουργία της αναλυτικής εικόνα από κάθε συνιστώσα μέσω του πολυδιάστατου κατευθυντικού Hilbert μετασχηματισμού [75]. Οι δύο προσεγγίσεις είναι σχεδόν ισοδύναμες. Περισσότερες λεπτομέρειες για αποδιαμόρφωση μέσω της 2Δ αναλυτικής εικόνας μπορούν να βρεθούν στα [72, 74, 76].

Οι συναρτήσεις ταλάντωσης που ορίστηκαν από το Meyer [137] για μοντελοποίηση και εξαγωγή της υφής από μια εικόνα, μπορούν να αναζητηθούν στο πλαίσιο της AM-FM μοντελοποίησης, μέσω διαμορφώσεων και την ύπαρξη ταλαντώσεων έντασης και χωρικών συχνοτήτων. Το κίνητρο εδώ είναι οι αναλογίες ανάμεσα στα μοντέλα διαμόρφωσης και τις μαθηματικές θεωρήσεις για τις νόρμες μεταβολής της υφής [137]. Μια τέτοια μοντελοποίηση παραμένει ανοιχτό ερευνητικό ζήτημα και πεδίο μελλοντικής έρευνας.

3.1.1 Ενέργεια συνιστωσών υφής

Εφαρμόζοντας τον 2Δ ενεργειακό τελεστή σε ένα σήμα $f_k(x, y) = a_k(x, y) \cos[\phi_k(x, y)]$ που μοντελοποιεί μια συνιστώσα υφής προκύπτει το ενεργειακό γινόμενο

$$\Psi[a_k \cos(\phi_k)] \approx a_k^2 \|\nabla \phi_k\|^2 \quad (3.4)$$

Ορισμός 3.1 Το γινόμενο (3.4) είναι η σύζευξη των τετραγώνων του στιγμιαίου πλάτους και του μέτρου της συχνότητας και μπορεί να ονομαστεί η ενέργεια διαμόρφωσης συνιστώσας υφής (*texture component modulation energy*).

Θεωρώντας ότι οι μεταβολές του πλάτους και των στιγμιαίων συχνοτήτων δεν είναι πολύ γρήγορες στο χώρο ούτε πολύ μεγάλες σε τιμές σε σχέση με τις φέρουσες, το παραπάνω σφάλμα προσέγγισης γίνεται αμελητέο [123, 125, 126]. Ο διαχωρισμός αυτής της ενέργειας στα δύο σήματα διαμόρφωσης είναι εφικτός εφαρμόζοντας τον ESA (Εξ. 2.20, 2.20). Επομένως μια μεμονωμένη, χωρικά μεταβαλλόμενη, AM-FM συνιστώσα υφής χαρακτηρίζεται σε κάθε θέση από την εκτιμώμενη περιβάλλουσα πλάτους και το μέγεθος της μεταβολής των στιγμιαίων συχνοτήτων.

3.2 Πολυκαναλική ανάλυση

Οι εικόνες υφής είναι συνήθως σήματα ευρείας ζώνης, δηλαδή χαρακτηρίζονται από ενέργεια σε διαφορετικές φασματικές ζώνες. Ως συνέπεια τα AM-FM μοντέλα δεν

εφαρμόζονται απευθείας σε μια εικόνα αλλά στις ζωνοπερατά φιλτραρισμένες εκδοχές της [21]. διαχωρισμό των σημαντικών διαμορφώσεων υφής ή την ανάδειξη των κυρίαρχων τοπικών συχνοτήτων [21], το ζωνοπερατό φιλτράρισμα αυξάνει την ευρωστία και ομαλοποιεί τα σήματα διαμόρφωσης.

Για την απομόνωση των τοπικών συνιστωσών διαμόρφωσης, χρησιμοποιούνται μηχανισμοί φιλτραρίσματος με επαρκή χωρική και φασματική τοποθέτηση. Μια συστοιχία από γραμμικά, μεταφορικά αμετάβλητα ζωνοπερατά φίλτρα καλύπτουν πυκνά το πεδίο συχνοτήτων με μόνη υπόθεση ότι σε κάθε pixel η απόκριση του κάθε φίλτρου κυριαρχείται από το πολύ μία AM-FM συνιστώσα. Το κάθε φίλτρο αποτελεί ένα κανάλι ενώ ο διαχωρισμός σε συνιστώσες διαμόρφωσης της (3.2) δεν είναι μοναδικός και εξαρτάται από τη δομή και το σχεδιασμό της συστοιχίας. Για την επιλογή των κεντρικών πυρήνων που δημιουργούν τη συστοιχία μεταβάλλοντας ένα σύνολο παραμέτρων, η ανάγκη για από κοινού καλή φασματική και χωρική τοποθέτηση επιβάλλει τη χρήση Gabor φίλτρων (2.1, 2.2) ως βέλτιστη επιλογή [21, 45, 74], για τα οποία αναλυτική αναφορά γίνεται στην ενότητα 2.2. Ο σχεδιασμός της συστοιχίας Gabor περιγράφεται αναλυτικά στην ενότητα 2.2.1.

Η έννοια της χρήσης πολλαπλών ζωνών συχνότητας για ανάλυση ή κατάτμηση υφής έχει εφαρμοσθεί προηγούμενα με επιτυχία π.χ. [19–21, 35, 74, 84, 163, 192]. Εδώ, εικόνες με πολλαπλούς σχηματισμούς υφής φιλτράρονται από μια συστοιχία K πραγματικών, ισοτροπικά, 2Δ Gabor φίλτρων, της μορφής

$$h_k(x, y) = \exp \left[-\frac{x^2}{2a_k^2} - \frac{y^2}{2b_k^2} \right] \cos(u_k x + v_k y) \quad , \quad 1 \leq k \leq K$$

όπου $(2\pi a_k)^{-1} = (2\pi b_k)^{-1}$ είναι το ενεργό rms εύρος ζώνης [45] αντίστοιχα στην οριζόντια x και στην κατακόρυφη y διάσταση του φίλτρου k και (u_k, v_k) είναι οι συντεταγμένες κεντρικής συχνότητας στο πεδίο συχνοτήτων. Οι αντίστοιχες αποκρίσεις συχνότητας H_k είναι διατεταγμένες ακτινικά σε διάφορους προσανατολισμούς και κλίμακες. Μια συνήθης επιλογή είναι η χρήση $K = 40$ φίλτρων σε 5 κλίμακες και 8 διευθύνσεις. Το όλο σχήμα αναφέρεται βιβλιογραφικά ως πολυκαναλική ή πολυζωνική ανάλυση (multiband analysis).

3.2.1 Ζωνοπερατή αποδιαμόρφωση

Έστω ότι σε ένα συγκεκριμένο χωρικό σημείο η απόκριση του φίλτρου $h_k(x, y)$, $H_k(u, v)$ στην εικόνα $I(x, y)$ (3.2), κυριαρχείται από την k συνιστώσα υφής f_k , έτσι ώστε σε μια γειτονιά γύρω από το σημείο

$$g_k(x, y) = I(x, y) * h_k(x, y) \approx f_k(x, y) * h_k(x, y)$$

Μπορεί να αποδειχθεί ότι κάτω από ρεαλιστικές και ασθενείς συνθήκες, όπως στο [22], ότι σ' αυτή τη γειτονιά [76]

$$\begin{aligned} \Psi[g_k(x, y)] &\approx a_k^2(x, y) |\nabla \phi_k(x, y)|^2 |H_k[\nabla \phi_k(x, y)]|^2 \\ &\approx \Psi[f_k(x, y)] |H_k[\nabla \phi_k(x, y)]|^2. \end{aligned} \quad (3.5)$$

Για την αποδιαμόρφωση μέσω ESA στα σήματα χωρικής μεταβολής, ο ενεργειακός τελεστής εφαρμόζεται στις απομονωμένες συνιστώσες και στις μερικές παραγωγούς τους. Για τα σήματα συχνοτήτων στις σχέσεις (2.20) ο Ψ εμφανίζεται τόσο στον αριθμητή όσο και στον παρονομαστή, οπότε ο όρος $|H_k[\nabla \phi_k(x, y)]|^2$ ακυρώνεται λόγω

διαίρεσης. Έτσι η εφαρμογή των (2.20) για την αποδιαμόρφωση μιας συνιστώσας γίνεται απευθείας στη ζωνοπερατή συνιστώσα στην έξοδο του φίλτρου

$$\sqrt{\frac{\Psi(\partial g_k/\partial x)}{\Psi(g_k)}} \approx |\omega_{k1}(x, y)|, \quad \sqrt{\frac{\Psi(\partial g_k/\partial y)}{\Psi(g_k)}} \approx |\omega_{k2}(x, y)| \quad (3.6)$$

Παρόμοια θεώρηση επεκτείνεται και στα διακριτά ισοδύναμα τους (2.26, 2.27). Αντίθετα για την ανάκτηση του σήματος πλάτους από τα ζωνοπερατά σήματα, λόγω του ότι οι συνιστώσες f_k υπόκεινται μεμονωμένα σε διαφορετική κλίμακα κατά την εφαρμογή της συστοιχίας, η (2.21) τροποποιείται με κανονικοποίηση ως εξής

$$|H_k[(\omega_{k1}, \omega_{k2})(x, y)]|^{-1} \frac{\Psi(g_k)}{\sqrt{\Psi(\partial g_k/\partial x) + \Psi(\partial g_k/\partial y)}} \approx |a_k(x, y)| \quad (3.7)$$

όπου οι ω_{k1}, ω_{k2} έχουν υπολογιστεί ήδη από την (3.6). Πρακτικά αυτή η τροποποίηση λαμβάνεται υπόψιν με κατάλληλη κανονικοποίηση των αποκρίσεων των φίλτρων κατά το σχεδιασμό της συστοιχίας. Τα πρόσημα των στιγμιαίων συχνοτήτων (3.6) προσεγγίζονται από τα πρόσημα των κεντρικών συχνοτήτων των φίλτρων. Συγκεκριμένα, για πραγματικά φίλτρα με συζυγή Fourier συμμετρία, αν οι κεντρικές των φίλτρων έχουν το ίδιο πρόσημο τότε $\text{sgn}(\omega_{k1}) = \text{sgn}(\omega_{k2}) = +1$, διαφορετικά $\text{sgn}(\omega_{k1}) = +1, \text{sgn}(\omega_{k2}) = -1$.

Επομένως, εφαρμόζοντας τον DESA, δηλ. τη διακριτοποίηση των σχέσεων (3.6, 3.7) στις K ζωνοπερατές συνιστώσες, στην έξοδο της συστοιχίας φίλτρων, προκύπτουν οι εκτιμήσεις για τις τοπικά και φασματικά κυρίαρχες συνιστώσες υφής. Η κάθε μία χαρακτηρίζεται από χωρικά μεταβαλλόμενα σήματα πλάτους $a_k(x, y)$ και στιγμιαίων συχνοτήτων $\omega_{k1}(x, y), \omega_{k2}(x, y)$.

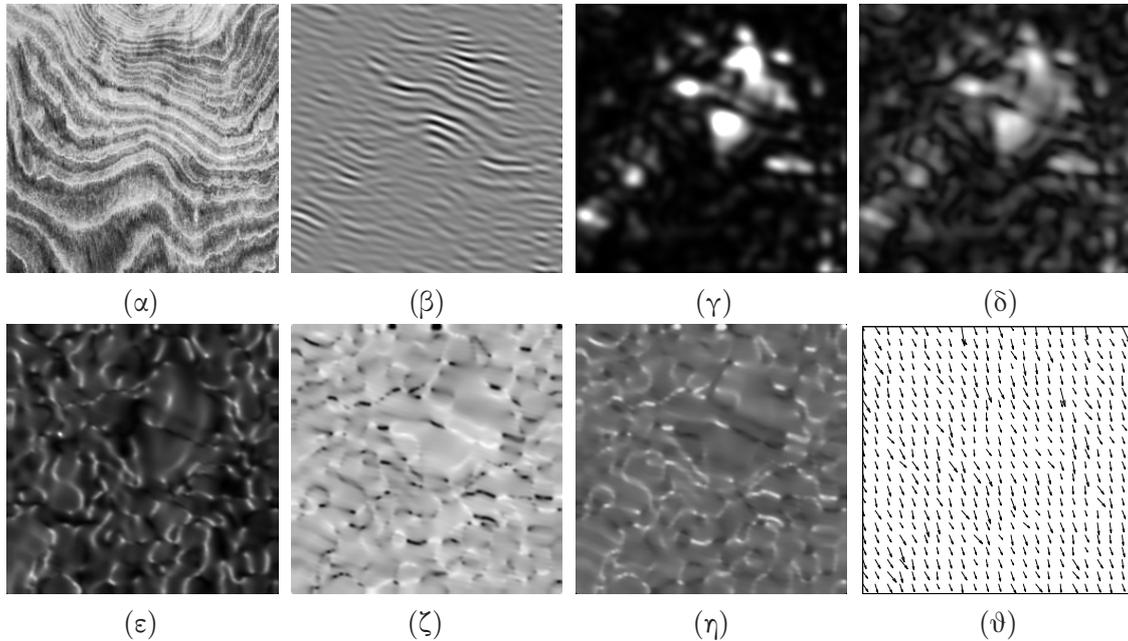
3.2.2 Ανάλυση συνιστωσών κατά κανάλια

Οι μεθοδολογίες ανάλυσης υφής φυσικών εικόνων που σχετίζονται με ανάλυση σε πολλαπλές ζώνες συχνοτήτων, ανακτούνε την πληροφορία στην έξοδο κάθε φίλτρου, δημιουργώντας διανύσματα χαρακτηριστικών ή εναλλακτικά συνδυάζουν τις εξόδους σε μια σύνθετη μέτρηση, για όλο το φάσμα συχνοτήτων, με κάποιο κριτήριο απόφασης. Οι Havlicek, Bovik κ.α. [72, 74] αναφέρονται στις δύο ξεχωριστές προσεγγίσεις ως *ανάλυση συνιστωσών κατά κανάλια* (Channelized Component Analysis, CCA) και *ανάλυση κυρίαρχων συνιστωσών* (Dominant Component Analysis, DCA) αντίστοιχα.

Θεωρώντας δεδομένο εκ των προτέρων ότι τα κανάλια της συστοιχίας χωρίζουν την εικόνα σε συνιστώσες με τρόπο χωρικά οικουμενικό (spatially global), κάθε συνιστώσα διαμόρφωσης ανήκει σε ένα κανάλι για όλο το πεδίο της εικόνας. Στο σχήμα CCA ο αριθμός των συνιστωσών στο προκύπτον μοντέλο είναι αναγκαστικά ίσος με τον αριθμό των καναλιών της συστοιχίας. Αποδιαμορφώνοντας τις προκύπτουσες συνιστώσες για την ολική εικόνα, με διαχωρισμό ενέργειας (3.6, 3.7), προκύπτει μια πυκνή αναπαράσταση, που χαρακτηρίζει τις ισχυρές δομές της εικόνας και τα τοπικά αναδύμενα χαρακτηριστικά της υφής. Για κάθε pixel κατασκευάζεται ένα διάνυσμα χαρακτηριστικών, διάστασης ίσης με τον αριθμό των συνιστωσών επί τον αριθμό των σημάτων διαμόρφωσης.

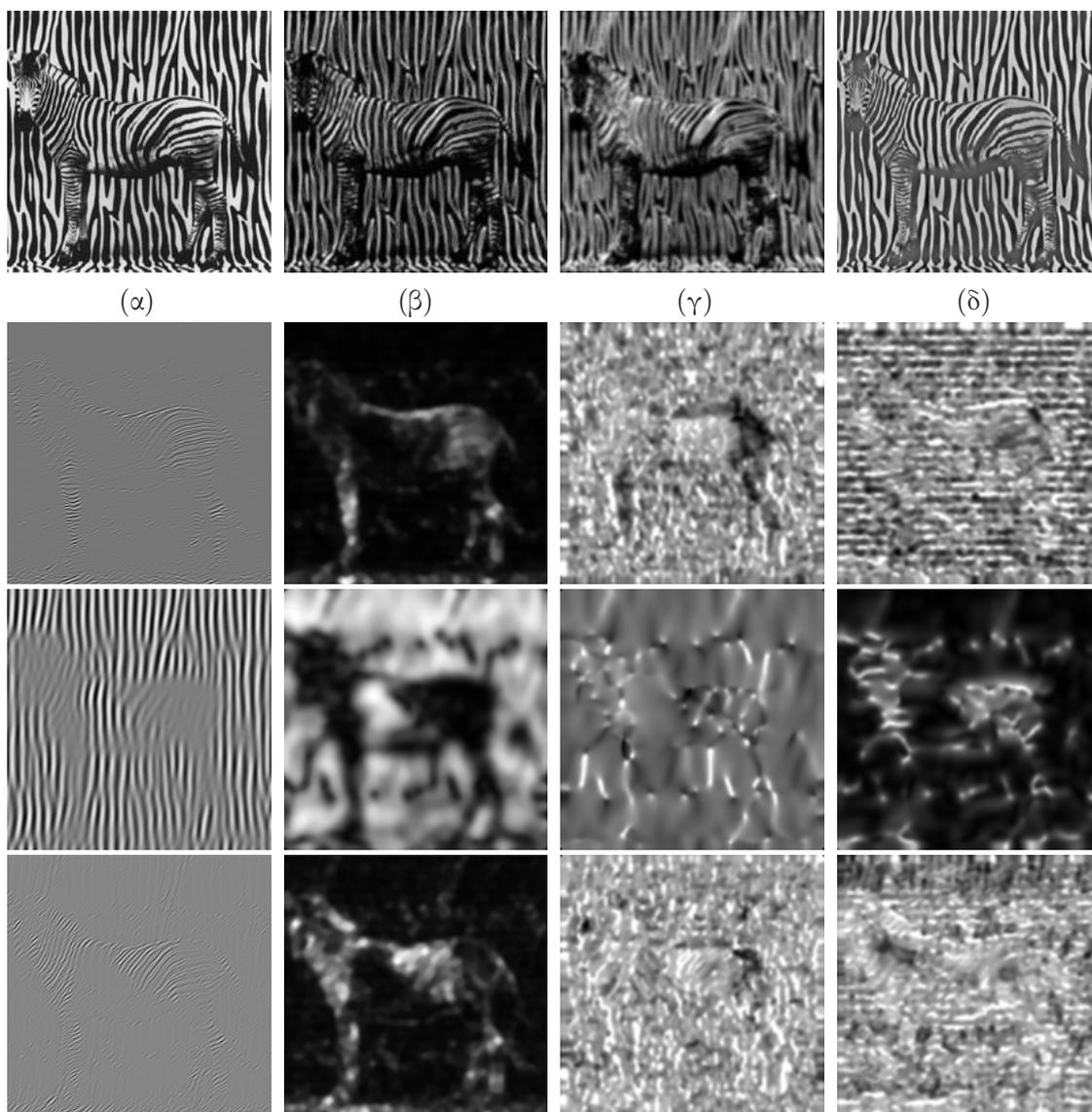
Στο σχήμα 3.1 φαίνεται ένα παράδειγμα ανάλυσης μιας εικόνας φυσικής υφής ξύλου [23] σε ένα κανάλι της συστοιχίας. Η εξαγώμενη ζωνοπερατή συνιστώσα στο (β) εμφανίζει διαμορφώσεις σε μια συγκεκριμένη κλίμακα και προσανατολισμό, οι οποίες

μοντελοποιούνται από ένα μη-στατικό ημιτονοειδές ενώ η έξοδος του 2Δ ενεργειακού τελεστή στο (γ) εκφράζει την ενέργεια διαμόρφωσης της συνιστώσας. Με εφαρμογή του DESA στο κανάλι προκύπτουν τα εκτιμώμενα σήματα πλάτους (δ) , οριζόντιας (ϵ) και κατακόρυφης (ζ) στιγμιαίας χωρικής συχνότητας. Τέλος στο (η) φαίνεται το μέτρο του διανύσματος συχνότητας, μέτρηση της κλίμακας της υφής και στο (θ) ο προσανατολισμός του ως needle διάγραμμα. Η ενέργεια διαμόρφωσης ποσοτικοποιεί την ύπαρξη και την μεταβολή της υφής στη συγκεκριμένη ζώνη συχνότητας.



Σχήμα 3.1: Αποδιαμόρφωση και ανάλυση υφής σε ζωνοπερατό κανάλι. (α) Υφή ξύλου ($D71$) 256×256 , (β) Ζωνοπερατή συνιστώσα από φίλτρο [$u = 3, v = 14, B = 20$], (γ) Ενέργεια διαμόρφωσης συνιστώσας υφής (έξοδος 2Δ ενεργειακού τελεστή), (δ) Περιβάλλουσα πλάτους, (ε) Οριζόντια στιγμιαία συχνότητα, (ζ) Κάτακόρυφη στιγμιαία συχνότητα, (η) Μέτρο του διανύσματος συχνότητας, (θ) Προσανατολισμός διανύσματος συχνότητας.

Αποτελέσματα ανάλυσης κατά κανάλια παρουσιάζονται και στο σχήμα 3.2 για μια εικόνα, με έντονες διαμορφώσεις σε πολλαπλές κλίμακες και προσανατολισμούς. Η σχηματισμοί υφής του δέρματος της ζέβρας, κυριαρχούν και στο φόντο της εικόνας με αποτέλεσμα τη δυσκολία εντοπισμού ακμών και διαχωρισμού της φιγούρας χωρίς εντατική επισκόπηση. Στην πρώτη σειρά παρουσιάζεται η εφαρμογή του 2Δ ενεργειακού τελεστή στη συνολική εικόνα ευρείας ζώνης (β) , που λειτουργεί σαν φίλτρο ενίσχυσης του της αντίθεσης της εικόνας [123], αλλά και η εφαρμογή σε μια εκδοχή μηδενικής μέσης τιμής της αρχικής εικόνας (γ) , όπου και εμφανίζει σχεδόν βαθυπερατές ιδιότητες. Στο (δ) παρουσιάζεται η ανακατασκευασμένη εικόνα από την εισαγωγή των 35 AM-FM συνιστωσών στο μοντέλο (3.2), που δεν είναι τέλεια αλλά πλησιάζει σε ικανοποιητικό βαθμό την αρχική. Οι τρεις σειρές που ακολουθούν δείχνουν την ανάλυση και τα σήματα διαμόρφωσης σε τρία διαφορετικά κανάλια της συστοιχίας, αντιπροσωπευτικά των δομών υφής της εικόνας. Η πρώτη και τρίτη συνιστώσα έχουν την ίδια κλίμακα αλλά διαφορετικό προσανατολισμό και εντοπίζουν πτυχώσεις διαφορετικής κατεύθυνσης στο σώμα της ζέβρας. Η δεύτερη έχει διαφορετική κλίμακα και προσανατολισμό και αναδεικνύει τις σχεδόν κάθετες και πιο αργές μεταβολές του φόντου. Όπως φαίνεται και από τα σήματα πλάτους και συχνότητας οι τρεις συνιστώσες είναι σχεδόν συμπληρωματικές, οριοθετώντας πλήρως τις διαφορετικές περιοχές υφής στην εικόνα.



Σχήμα 3.2: Συνιστώσες AM-FM υφής και αποδιαμόρφωση. Πάνω σειρά: (α) Τυφή κυματισμών ζέβρας I (200×200), (β) Ενεργειακός τελεστής στην αρχική εικόνα $\Psi(I)$, (γ) Ενεργειακός τελεστής στην αρχική με μηδενική μέση τιμή $\Psi(I) - \bar{I}$, (δ) Ανακατασκευασμένη εικόνα από 35 AM-FM συνιστώσες. Αριστερά προς δεξιά για τις υπόλοιπες: Ζωνοπερατή συνιστώσα, περιβάλλουσα πλάτους $a(x, y)$, οριζόντια $\omega_1(x, y)$ και κατακόρυφη $\omega_2(x, y)$ στιγμιαία συχνότητα. Δεύτερη σειρά: συνιστώσα φίλτρου [$u = 16, v = 44, B = 5$], Τρίτη σειρά: συνιστώσα φίλτρου [$u = 14, v = 2, B = 16$], Τέταρτη σειρά: συνιστώσα φίλτρου [$u = 31, v = 35, B = 5$].

3.2.2.1 Επεξεργασία και ομαλοποίηση

Για τον περιορισμό των σφαλμάτων εκτίμησης της ενέργειας και της διαδικασίας αποδιαμόρφωσης, καθώς και για την επιβολή κάποιας ομαλότητας σήματα πλάτους και συχνοτήτων συνίσταται η εφαρμογή τεχνικών μετα-επεξεργασίας (post-processing). Γενικώς οι εικόνες υψής περιέχουν ασυνέχειες, σκιάσεις και «σπασίματα», με αποτέλεσμα να απέχουν από το ιδανικό μοντέλο των τοπικά ομοιόμορφων διαμορφώσεων. Τέτοιες ιδιομορφίες εμφανίζονται ως μη ομαλές μεταβολές στις περιβάλλουσες και στις φάσεις των αποκρίσεων φίλτρων της συστοιχίας. Στα [19, 20] όπου η απόκριση πλάτους και φάσης χρησιμοποιήθηκαν για κατάτμηση εικόνων, βαθυπερατά φίλτρα Gauss τυπικής απόκλισης περίπου ίσης με αυτή των Gabor καναλιών, χρησιμοποιήθηκε για ομαλοποίηση των αποκρίσεων.

Στην περίπτωση αποδιαμόρφωσης των καναλιών είναι προτιμότερο η ομαλοποίηση και η διαχείριση των σφαλμάτων να εφαρμόζεται απευθείας στις εκτιμώμενες AM και FM συναρτήσεις. Τα σφάλματα αποδιαμόρφωσης οφείλονται στα σφάλματα προσεγγίσεων του ενεργειακού τελεστή, στους λόγους με διακριτές, μη-γραμμικές ποσότητες και στις προσεγγίσεις παραγώγων μέσω διαφορών. Για την πρώτη περίπτωση, τα τυχόν σφάλματα προσέγγισης του 2Δ ενεργειακού τελεστή έχουν μια υψιπερατή και μια βαθυπερατή συνιστώσα [126, 158]. Η υψιπερατή αφαιρείται με εφαρμογή στην έξοδο του τελεστή ενός βαθυπερατού φίλτρου π.χ. διωνυμικά ή Gauss φίλτρα. Επιπλέον τυχόν αρνητικές τιμές στην έξοδο του τελεστή, που θεωρητικά δεν ανταποκρίνονται στον ορισμό της ενεργειακής ποσότητας και πρακτικά δεν μπορούν να εξισωθούν με την κατά προσέγγιση τετραγωνική ποσότητα (3.4), αντικαθίστανται με κάποιο τοπικό μέσο.

Για τον περιορισμό των σφαλμάτων αποδιαμόρφωσης (3.6, 3.7), χρησιμοποιείται ένας συνδυασμός από βαθυπερατά φίλτρα ομαλοποίησης των εκτιμήσεων και στατιστικών - median φίλτρων σε τοπικά παράθυρα. Το βαθυπερατό φιλτράρισμα αντισταθμίζει τα σφάλματα προσεγγίσης του αλγορίθμου ενώ το median ενδείκνυται για την απομάκρυνση μεμονωμένων ακραίων μεταβολών (spikes, singularities) λόγω των διαιρέσεων του DESA. Στα παραδείγματα των σχημάτων 3.1 και 3.2, χρησιμοποιήθηκε ενδεικτικά διωνυμικό φιλτράρισμα στον ενεργειακό τελεστή, Gauss φίλτρα τυπικής απόκλισης 2 και μεγέθους 3 και median σε παράθυρα 3×3 .

3.2.2.2 Ανακατασκευή

Η ανάκτηση της αρχικής εικόνας από τις παραμέτρους του μοντέλου σε κάθε θέση, περιλαμβάνει ανακατασκευή των συνιστωσών (2.7) από τα χωρικά μεταβαλλόμενα σήματα διαμόρφωσης και σύνθεση της (3.2) με υπέρθεση των επιμέρους AM-FM συνιστωσών. Η δυσκολία έγκειται στο μη καλά ορισμένο ανάστροφο πρόβλημα της ανάκτησης της φάσης $\phi_k(x, y)$ από το διάλυμα στιγμιαίων συχνοτήτων (ω_{k1}, ω_{k2}) για κάθε συνιστώσα. Η ανακατασκευή φάσης είναι ένα γενικότερα δύσκολο πρόβλημα λογισμού διανυσματικών πεδίων και για την μη αναλυτική επίλυση του έχουν προταθεί αλγεβρικές, μεταβολικές και αριθμητικές μέθοδοι. Για τα AM-FM μοντέλα μη-βέλτιστος αλλά πρακτικός αλγόριθμος προτάθηκε στο [73].

Οι K συνιστώσες φάσης ανακτώνται χρησιμοποιώντας πρώτης τάξης Taylor σειρές για τα διαφορικά στις δύο κατευθύνσεις, καταλήγοντας στη διακριτή σχέση

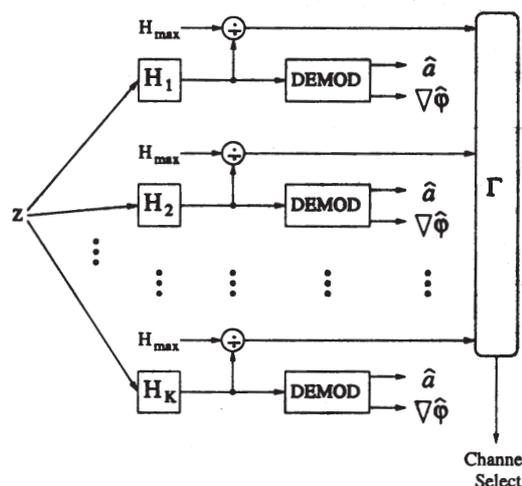
$$\phi_k(m, n) = \frac{1}{2}[\phi_k(m-1, n) + \omega_{k1}(m-1, n) + \phi_k(m, n-1) + \omega_{k2}(m, n-1)] \quad (3.8)$$

Η αρχικοποίηση της (3.8) γίνεται με τη βοήθεια δειγμάτων της φάσης της απόκρισης του καναλιού k που κατά προσέγγιση είναι ισοδύναμη με τη φάση της συνιστώσας που κυριαρχεί στο κανάλι [74]. Τα δείγματα της φάσης αποθηκεύονται κατά τη διαδικασία ανάλυσης και για την αποφυγή διάδοσης σφαλμάτων εκτίμησης που υπεισέρχονται στα στιγμιαία σήματα συχνοτήτων, η διαδικασία ανακατασκευής της φάσης μπορεί να γίνεται ανεξάρτητα σε περιοδικά block από pixels. Η φάση αρχικοποιείται σε κάθε τετράγωνο, εκτιμάται ανεξάρτητα στην πρώτη στήλη και πρώτη γραμμή και από την (3.8) ανακατασκευάζεται στο εσωτερικό του.

3.3 Ανάλυση κυρίαρχων συνιστωσών

Η πληροφορία των πολυδιάστατων χαρακτηριστικών της ανάλυσης κατά κανάλια, παρά το ότι ερμηνεύει πιστά το μοντέλο διαμορφώσεων, μπορεί να καθίσταται περιττή και να αυξάνει την πολυπλοκότητα της περιγραφής. Η ανάλυση κυρίαρχων συνιστωσών (dominant components analysis ή DCA) [72, 74] αποτελεί μια εναλλακτική περιγραφή με διανύσματα χαρακτηριστικών χαμηλής διάστασης. Η μεθοδολογία για μια εικόνα υφής που μοντελοποιείται κατά την (3.2), εκτιμά τις συναρτήσεις διαμόρφωσης των τοπικά ισχυρών συνιστωσών ανά θέση. Το εκτιμώμενο πλάτος κυρίαρχης συνιστώσας $a_D(x, y)$ ή κυρίαρχο πλάτος αντιστοιχεί στην αντίθεση της επικρατέστερης δομής υφής, ενώ οι συχνότητες κυρίαρχης συνιστώσας $\nabla\phi_D(x, y)$ ή κυρίαρχες συχνότητες είναι εκτιμήσεις των αναδυόμενων τοπικά συχνοτήτων.

Οι μη-στατικές, στιγμιαίες συχνότητες που κυριαρχούν σε χωρικά τοπική βάση στο φάσμα των περισσότερων φυσικών εικόνων υφής ονομάζονται αναδυόμενες (emergent frequencies) [21] και δίνονται από το διάνυσμα φάσης της συνιστώσας που είναι κυρίαρχη σε κάθε σημείο. Τέτοιες συχνότητες χαρακτηρίζουν την λεπτομέρεια, τη ροή και τον προσανατολισμό των περισσότερων φυσικών υφών (τοπικά συνεκτικές, κανονικής δομής, αποτελέσματα φυσικών διαδικασιών εξέλιξης [191, 205], κοκκώδεις, προσανατολισμένες υφές). Ως συνέπεια οι όποιες εκτιμήσεις αυτών των συχνοτήτων φέρουν πληροφορία σχετική με την τοπική δομή της υφής [21].



Σχήμα 3.3: Διάγραμμα ανάλυσης κυρίαρχων συνιστωσών από το [71].

Το σχήμα DCA εξάγει χαρακτηριστικά υφής [71, 74] καταλαμβάνοντας τις τοπικά ισχυρές ταλαντώσεις φωτεινότητας όπως περιγράφονται από το AM-FM μοντέλο και

επιλέγοντας ανάμεσα στις διαφορετικές ζώνες συχνότητων. Τα στάδια φαίνονται στο διάγραμμα του σχήματος 3.3 (από το [74]). Το σύνθετο διδιάστατο σήμα αναλύεται από μια συστοιχία ζωνοπερατών φίλτρων για την απομόνωση των επιμέρους συνιστωσών σε χωρό-φασματική με βάση τις αρχές της ενότητας 3.2. Οι διαφορετικές εκδοχές στενής ζώνης, που ιδανικά αντιστοιχούν σε μεμονωμένες συνιστώσες του μοντέλου υφής, αποδιαμορφώνονται με κάποιον αλγόριθμο αποδιαμόρφωσης, όπως οι εξισώσεις διαχωρισμού (3.6, 3.7). Ανάλογα με την ανάλυση κατά κανάλια, η κάθε συνιστώσα δίνει ξεχωριστές εκτιμήσεις πλάτους και στιγμιαίων συχνότητων.

Η επιλογή του καναλιού, που δίνει την καλύτερη εκτίμηση για την κυρίαρχη τοπικά συνιστώσα, βασίζεται στον υπολογισμό ενός μετρικού, του κριτηρίου επιλογής καναλιού Γ σε κάθε σημείο στο πεδίο της εικόνας. Εκτιμήσεις των συναρτήσεων διαμόρφωσης των κυρίαρχων συνιστωσών δίνονται για κάθε σημείο από το κανάλι της συστοιχίας που μεγιστοποιεί το κριτήριο

$$i = \arg \max_{1 \leq k \leq K} \{\Gamma_k\} \quad (3.9)$$

όπου το $k = 1 \dots K$ αντιστοιχεί σε ένα από τα K φίλτρα της διάταξης, ενώ το αντίστοιχο κυρίαρχο πλάτος και κυρίαρχο μέτρο συχνότητας είναι για όλο το πεδίο

$$a_D(x, y) = a_i(x, y) \quad , \quad |\vec{\omega}_D(x, y)| = |\vec{\omega}_i(x, y)| \quad (3.10)$$

Ανάλογα με την επιλογή του κριτηρίου Γ υπάρχουν δύο εναλλακτικά DCA σχήματα. Το DCA πλάτους ή ADCA, που αποτελεί την αρχική προσέγγιση [72], χρησιμοποιεί ένα κριτήριο κανονικοποιημένου πλάτους

$$\Gamma_k(x, y) = \frac{|a_k(x, y)|}{\max_{\vec{w}} |H_k(\vec{w})|} \quad (3.11)$$

όπου $H_k(\vec{w})$, $w = (u, v)$ η απόκριση χωρικών συχνότητων του φίλτρου k , που ισοδύναμα αποκτάται χρησιμοποιώντας τόσο το πραγματικό h_{ck} όσο και το φανταστικό h_{sk} μέρος της απόκρισης μιγαδικού φίλτρου (ζεύγος με κάθετη διαφορά φάσης)

$$\Gamma_k(x, y) = \frac{\sqrt{(I * h_{ck})^2(x, y) + (I * h_{sk})^2(x, y)}}{\max_{\vec{w}} |H_k(\vec{w})|} \quad (3.12)$$

Τα δύο κριτήρια αποδεικνύονται να είναι θεωρητικά ισοδύναμα και η περιβάλλουσα της μιγαδικής απόκρισης $|(I * h_k)(x, y)|$ προσεγγιστικά είναι μια ομαλοποιημένη έκδοση του πλάτους $|a_k(x, y)|$ [19, 20]. Η ομαλοποίηση επιβάλλεται από μια συνάρτηση Gauss τυπικής απόκλισης $\sigma_k / \sqrt{2}$, όπου σ_k η απόκλιση για το συγκεκριμένο κανάλι. Συγκεκριμένα μπορεί να δειχθεί ότι

$$|(I * h_k)(x, y)| = c g_k(x/\sqrt{2}, y/\sqrt{2}) |a_k(x, y)| \approx |a_k(x, y)|$$

όπου c μια σταθερά και g_k ο παλμός διαμόρφωσης Gauss του φίλτρου k .

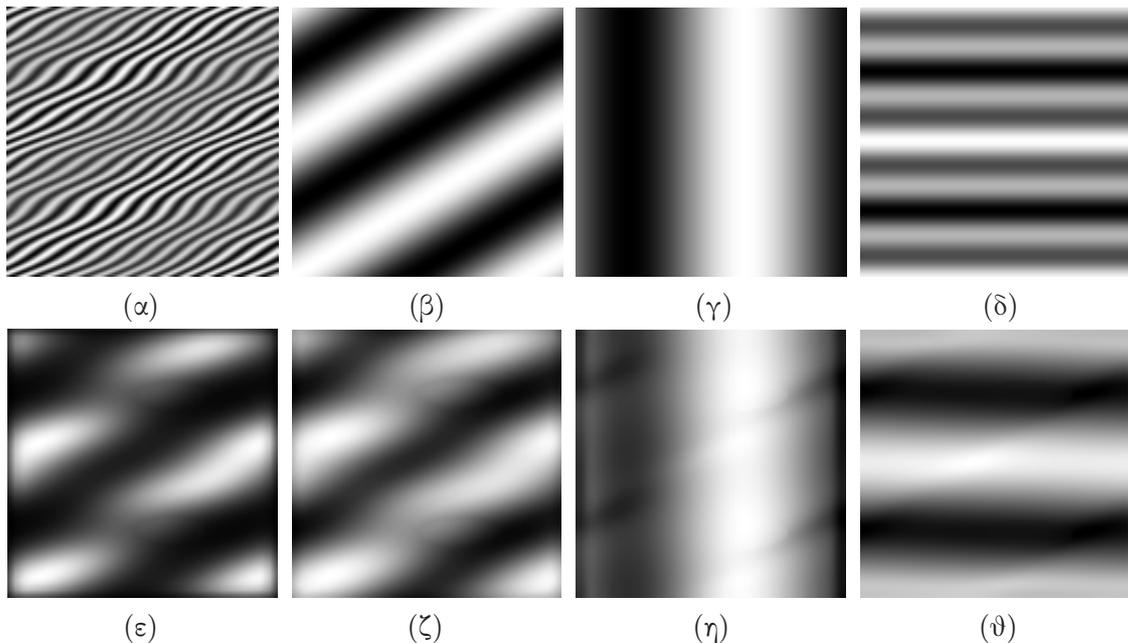
Εναλλακτικά και στα πλαίσια της παρούσας δουλειάς [98, 99] προτάθηκε ένα σχήμα DCA βασισμένο στην Teager ενέργεια των κυρίαρχων συνιστωσών το οποίο και καλείται DCA ενέργειας ή EDCA. Συγκεκριμένα το κυρίαρχο τοπικά κανάλι επιλέγεται με βάση την μέση έξοδο του ενεργειακού τελεστή στην πραγματική απόκριση του φίλτρου

$$\Gamma_k(x, y) = \overline{\Psi[(I * h_{ck})(x, y)]} \quad (3.13)$$

Και στις δύο προσεγγίσεις οι εκτιμήσεις στις παραμέτρους διαμόρφωσης των κυρίαρχων συνιστωσών, υποβάλλονται σε πρόσθετη μετα-επεξεργασία και φιλτράρισμα για

τον περιορισμό σφαλμάτων αλλά και την επιβολή ομαλότητας στα σημεία μετάβασης από μια κυρίαρχη συνιστώσα σε άλλη. Η ενεργειακή ανάλυση κυρίαρχων συνιστωσών παρουσιάζει πλεονεκτήματα αλλά και φορμαλισμό σε σχέση με την ανάλυση πλάτους. Συγκεκριμένα ισοσταθμίζει το ρόλο του πλάτους και το ρόλο του εύρους συχνοτήτων στην επιλογή των τοπικά κυρίαρχων συνιστωσών κάτι που διαισθητικά φαίνεται πιο φυσικό. Εξάλλου η ενέργεια διαμόρφωσης υψής, όπως ορίστηκε προηγουμένα, περιλαμβάνει ταυτόχρονα πληροφορία πλάτους και συχνοτήτων. Επομένως μια ερμηνεία του EDCA είναι ότι εξάγει τις κυρίαρχες συνιστώσες της υψής με βάση την ενέργεια των διαμορφώσεων τους. Μια συνιστώσα χαρακτηρίζεται κυρίαρχη όχι μόνο λόγω μεγάλων τιμών διαμόρφωσης περιβάλλουσας, που ερμηνεύεται μέσω υψηλού contrast τοπικά αλλά και λόγω των αναδυόμενων συχνοτήτων που κυριαρχούν στο φάσμα της εικόνας. Επιπλέον οι εκτιμήσεις του EDCA, λόγω του συχνοτικού όρου, συλλαμβάνουν οξείες μεταβολές της υψής ανάμεσα σε διαφορετικές περιοχές της εικόνας και ακμές υψής.

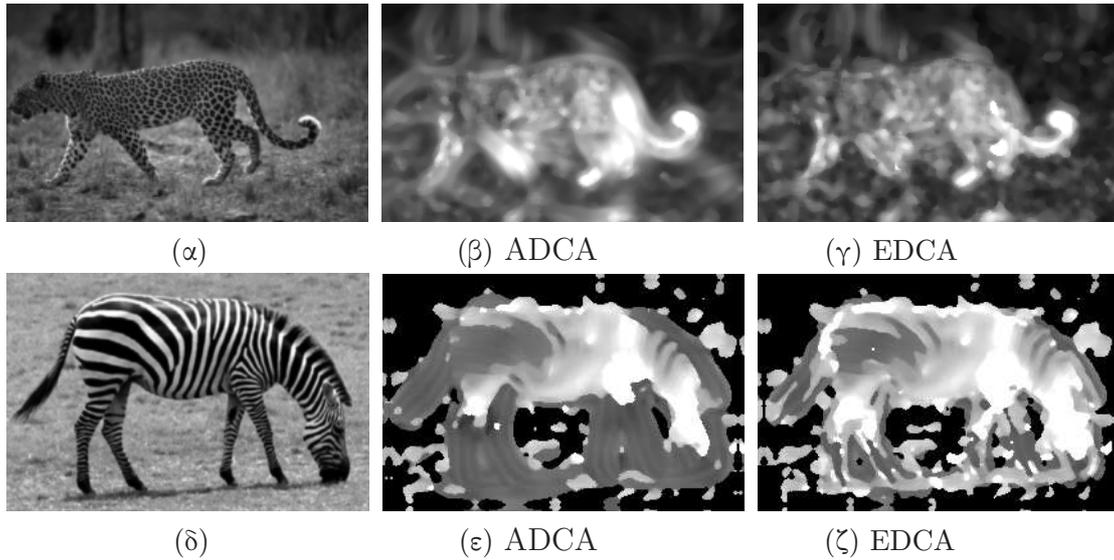
Στο [98] αποδόθηκε στο DCA σχήμα για την επιλογή του κυρίαρχου καναλιού μια θεωρητική στατιστική ανάλυση ανίχνευσης σημάτων μαζί με απόφαση υψής \ μη υψής σε κάθε σημείο μιας εικόνας, σε διαφορετικές κλίμακες. Στο [99] χαρακτηριστικά κυρίαρχων συνιστωσών διαμόρφωσης προτάθηκαν και χρησιμοποιήθηκαν για μεταβολική κατάτμηση εικόνων υψής.



Σχήμα 3.4: Ανάλυση Κυρίαρχών Συνιστωσών Ενέργειας (EDCA) σε ένα (α) συνθετικό 2D AM-FM σήμα, (β) - (δ) Πραγματική περιβάλλουσα πλάτους, Οριζόντια και κατακόρυφη στιγμιαία συχνότητα αντίστοιχα, (ε) Κυρίαρχη ενέργεια Teager (ζ) - (θ) Κυρίαρχες συνιστώσες διαμόρφωσης πλάτους και συχνοτήτων μέσω DCA.

3.4 Ενέργεια διαμόρφωσης και χαρακτηριστικά υψής

Ο μηχανισμός εντοπισμού της ενέργειας των συνιστωσών της υψής που πραγματοποιεί η ανάλυση κυρίαρχων συνιστωσών λειτουργεί στον πολυδιάστατο χώρο χαρα-



Σχήμα 3.5: EDCA και ADCA. (β) - (γ) Κυρίαρχες συνιστώσες πλάτους και (ε) - (ζ) Κυρίαρχες συνιστώσες μέτρου στιγμιαίων συχνοτήτων [98].

κτηριστικών που αποτελείται από τις αποκρίσεις των φίλτρων h_k . Οι φιλτραρισμένες, στενής ζώνης συνιστώσες υφής υποβάλλονται στη μη-γραμμικότητα του 2Δ ενεργειακού τελεστή Ψ και αποδιαμορφώνονται μέσω του ESA (2.20, 2.21). Οι μετρήσεις ενέργειας ομαλοποιούνται από ένα τοπικό παράθυρο-φίλτρο μέσου όρου h_a και συγκρίνονται σημειακά. Το φίλτρο i με τη μέγιστη μέση Teager ενέργεια (Maximum Average Teager Energy)

$$\Psi_{\text{mat}}[I(x, y)] = \max_k \{(\Psi(I * h_k) * h_a)(x, y)\}, \quad (3.14)$$

καθορίζει την επικρατέστερη κυρίαρχη συνιστώσα υφής.

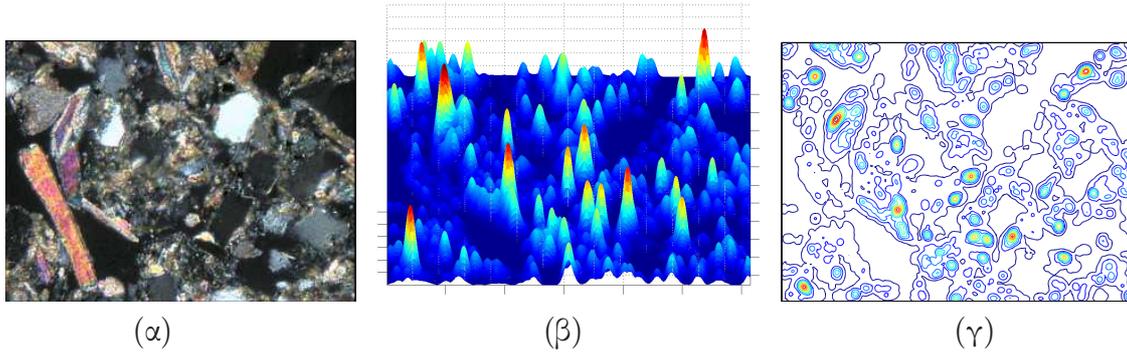
Ορισμός 3.2 Η μέγιστη μέση ενέργεια διαμόρφωσης των τοπικών συνιστωσών της υφής (3.14) είναι μια αργά μεταβαλλόμενη ένδειξη της ενέργειας διαμόρφωσης υφής (*texture modulation energy*).

Η ενέργεια διαμόρφωσης παρέχει πληροφορία για την ύπαρξη ή μη, δομής υφής σε περιοχές της εικόνας, υπό την έννοια και την περιγραφή του μοντέλου AM-FM διαμορφώσεων. Μπορεί επομένως να χρησιμοποιηθεί για ανίχνευση και διάκριση υφής και να κατηγοριοποιήσει διάφορα ενεργειακά επίπεδα και επομένως υφές διαφορετικής φύσης. Παράλληλα ένα σύνολο χαρακτηριστικών χαμηλής διάστασης, το πλάτος κυρίαρχης διαμόρφωσης (*dominant modulation amplitude*) $a_D(x, y)$ και το διάνυσμα συχνοτήτων κυρίαρχης διαμόρφωσης (*dominant modulation frequency vector*) $\vec{\omega}_D(x, y)$, σχηματίζονται από τις αποδιαμορφωμένες τιμές του κυρίαρχου καναλιού σε κάθε θέση του χωρικού πεδίου.

Τα χαρακτηριστικά διαμόρφωσης (*modulation features*) που εξάγονται αναζητώντας τις κυρίαρχες συνιστώσες υφής σε πολλαπλές ζώνες του φάσματος, παρέχουν τοπική (περιοχές, ακμές, σύνορα) όσο και ολική πληροφορία (ομοιογένεια, φασματικές ζώνες) δομής της υφής και μπορούν να περιγράψουν ένα εύρος υφών. Η ενεργειακή αυτή προσέγγιση προτάθηκε στα [177, 178] για ανίχνευση, περιγραφή, ανάλυση της υφής και πολυκριτηριακή κατάτμηση εικόνων, όπου η Ψ_{mat} αποτέλεσε ένα εύχρηστο, αποτελεσματικό και καλά εντοπισμένο χαρακτηριστικό ανίχνευσης υφής. Όπως αναφέρθηκε

για το EDCA προηγούμενα, η Ψ_{mat} προσφέρει βελτιωμένη τοποθέτηση σημαντικών γεγονότων και συνυπολογίζει τη συχνότητα και το πλάτος διαμόρφωσης της επικρατούσας συνιστώσας.

Στο σχήμα 3.6 παρουσιάζεται η ενέργεια διαμόρφωσης υφής της σχέσης (3.14) για μια πλούσια σε διαφορετικές υφές εικόνα λεπτής εδαφοτομής. Η μέτρηση εξάχθηκε από μια συστοιχία 40 φιλτρων, με μέσους όρους σε γειτονιά 11×11 pixel και γραμμική και μη-γραμμική ομαλοποίηση 11×11 , των κυρίαρχων χαρακτηριστικών. Η ενέργεια παρουσιάζεται ως προοπτική 3Δ απεικόνιση, αλλά και με τη μορφή 10 επιπεδοκαμπυλών της (level curves). Η ύπαρξη αλλά και τα επίπεδα έντασης της υφής ανά περιοχές στην εικόνα, χαρακτηρίζονται από την ενέργεια διαμόρφωσης.



Σχήμα 3.6: Ενέργεια διαμόρφωσης και ανίχνευση υφής. (α) Εικόνα εδαφοτομής (768×567 pixels), (β) Μέγιστη μέση Teager ενέργεια Ψ_{mat} , σε 3Δ προοπτική προβολή και (γ) ως υπέρθεση 10 επιπεδοκαμπυλών της.

□

Κεφάλαιο 4

Κατάτμηση Υφής

Ο διαχωρισμός μιας εικόνας σε περιοχές με ομοιογενή χαρακτηριστικά, όπως η γεωμετρία, η φωτεινότητα, το χρώμα, η υφή, η κίνηση κ.α. αποτελεί μια από τις σημαντικότερες διεργασίες όρασης χαμηλού επιπέδου. Η φιλοσοφική - γνωσιακή του αντιστοιχία που είναι γνωστή ως διαχωρισμός *πρώτου πλάνου-φόντου* (figure-ground). Η κατάτμηση υφής περιλαμβάνει τα ανεξάρτητα στάδια της εξαγωγής χαρακτηριστικών υφής, της επιλογής ή μείωσης της διάστασης τους και ενός αλγορίθμου κατάτμησης. Η εξαγωγή κυρίαρχων συνιστωσών μέσω πολυκαναλικής ανάλυσης και εντοπισμού της ενέργειας καταλήγει σε διανύσματα χαρακτηριστικών χαμηλής διαστασης (πλάτος, συχνότητες, προσανατολισμός). Τέτοια χαρακτηριστικά εμπλέκονται και λειτουργούν ως κινητήριες δυνάμεις σε μεταβολικά (variational) σχήματα (α) κατάτμησης υφής χωρίς επισκόπηση, (β) πολυκριτηριακής κατάτμησης εικόνων με ανίχνευση υφής. Περιγράφεται εδώ το πρώτο σχήμα, βασισμένο στα εσωτερικά των περιοχών, που είναι περισσότερο προσανατολισμένο στην κατάτμηση υφής.¹

4.1 Κατάτμηση με εξέλιξη καμπυλών και κυρίαρχες διαμόρφώσεις

Η κατάτμηση εικόνων υφής είναι ένα από τα σημαντικότερα και πιο διάσημα προβλήματα αρχικής όρασης, που έχει αντιμετωπιστεί από διαφορετικές προοπτικές όπως στατιστικά μοντέλα [61, 120] γεωμετρικές και μορφολογικές μεθόδους [17, 176, 181] και μεταβολικές τεχνικές βασισμένες στα εσωτερικά των περιοχών (region-based) [107, 147, 209, 212] ή βασισμένες στα όρια (boundary/edge-based) (π.χ. [114, 153], snakes, deformable models [38, 94], geodesic active contours [30, 115, 146]).

Στο region-based σενάριο κατάτμησης, τα pixel της εικόνας ομαδοποιούνται σε περιοχές ανάλογα με τη συγγένεια των χαρακτηριστικών τους. Ένα σύνολο αρχικών καμπυλών εξελίσσεται στο πεδίο της εικόνας με οδηγό ένα διάνυσμα περιγραφικών χαρακτηριστικών, συνήθως περισσότερο εκλεπτυσμένων από απλές τιμές φωτεινότητας. Οι εικόνες υφής χαρακτηρίζονται από τυχαία φύση, αβεβαιότητα στην τοποθέτηση συνόρων και ακμών, συχνοτικό περιεχόμενο ευρείας ζώνης και τοπική χωροφασματική αβεβαιότητα. Η πολυκαναλική ανάλυση και εξαγωγή χαρακτηριστικών είναι συνηθισμένη διαδικασία για την ανίχνευση ευρυζωνικής πληροφορίας υφής σε διαφορετικά κανάλια, όπως περιγράφηκε σε προηγούμενη ενότητα. Παρά τον πληροφοριακό πλου-

¹Η έρευνα και τα αποτελέσματα αυτής της ενότητας ανήκουν στο πλαίσιο των [55, 98, 99] σε συνεργασία με τον Ιάσωνα (jason) Κόκκινο.

ραλισμό της, η εξαγωγή χαρακτηριστικών στην έξοδο μιας συστοιχίας φίλτρων οδηγεί σε διανύσματα μεγάλης διάστασης που καταλήγουν σε υποβέλτιστες κατατμήσεις.

Στο σχήμα κατάτμησης με επισκόπηση (supervised texture segmentation [147]), η βέλτιστη κατάτμηση επιτυγχάνεται με την επιλογή των καναλιών που διαχωρίζουν καλύτερα τις διαφορετικές υφές. Στο σενάριο αυτόματης κατάτμησης χωρίς επισκόπηση (unsupervised), η επιλογή των καναλιών των χαρακτηριστικών και η μείωση της διάστασης του διανύσματος πραγματοποιείται συνήθως με ευριστικά κριτήρια [169,170]. Πρόσφατα μια διανυσματική διαδικασία διάχυσης (vector-valued diffusion) [167] χρησιμοποιήθηκε για την απλοποίηση του διανύσματος χαρακτηριστικών.

Τα χαρακτηριστικά διαμόρφωσης, μέσω του AM-FM μοντέλου, ερευνήθηκαν σε ένα σχήμα κατάτμησης χωρίς επισκόπηση στα [98,99]. Η ανάλυση κυρίαρχων συνιστωσών (3.3) χρησιμοποιείται για την εξαγωγή διανυσμάτων χαρακτηριστικών χαμηλής διάστασης που περιλαμβάνουν πληροφορία μεγέθους, κλίμακας και προσανατολισμού, ενώ μπορεί να ερμηνευθεί ως περιγραφή των τοπικών ημιτονοειδών σημάτων που μοντελοποιούν καλύτερα τις συνιστώσες της υφής. Η κατάτμηση γίνεται με εξέλιξη καμπυλών (curve evolution) και χρήση μεθόδων επιπεδοσυνόλων (level sets) όπου οι κινητήριες δυνάμεις των καμπυλών καθορίζονται από ένα πιθανοτικό κριτήριο, με βάση τα εσωτερικά των περιοχών, όπως στα [148,212].

Μια πρόσφατη προσπάθεια σύζευξης μεταβολικών μεθόδων με χαρακτηριστικά διαμόρφωσης για κατάτμηση υφής [164] περιλαμβάνει ένα σχήμα γεωδαιτικών ενεργών περιγραμμάτων (geodesic active contours), παρόλο που το μεταβολικό κομμάτι χρησιμεύει ως βελτίωση σε ένα καθαρά στατιστικό αλγόριθμο.

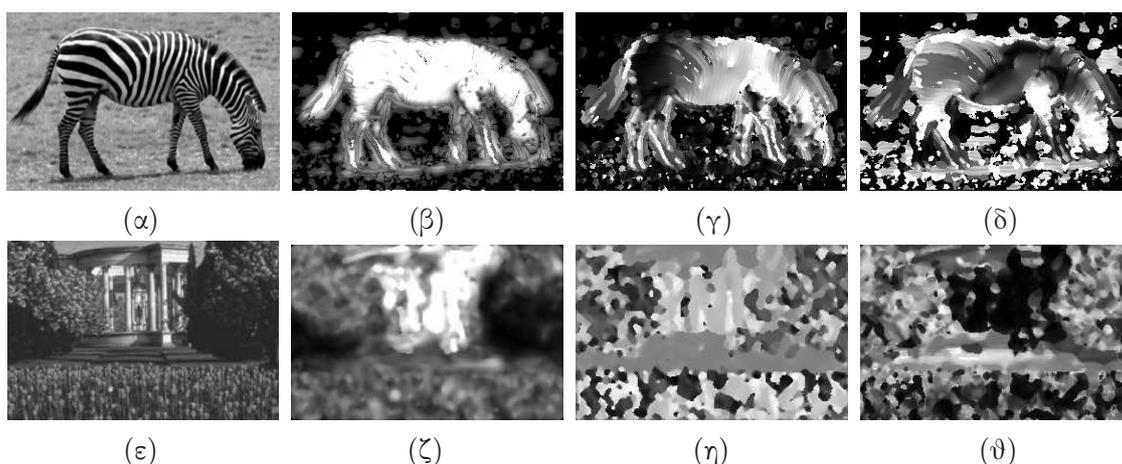
4.2 Χαρακτηριστικά διαμόρφωσης

Η υφή περιγράφεται μέσω δομών διαμόρφωσης, θεωρώντας το σύνολο των χαρακτηριστικών που εξάγεται μέσω της ανάλυσης κυρίαρχων συνιστωσών DCA. Εξετάστηκε τόσο η πρωταρχική μέθοδος με βάση το πλάτος (εξ. 3.11), όσο και η εναλλακτική πρόταση βασισμένη στην ενέργεια που εντοπίζει ο 2Δ ενεργειακός τελεστής (εξ. 3.13). Στα σχήματα 4.1, φαίνεται το σύνολο των χαρακτηριστικών υφής για δύο φυσικές εικόνες με περισσότερες από μία περιοχές διαφορετικής υφής (πέντε και δύο αντίστοιχα). Το πλάτος αντιστοιχεί σε πληροφορία μεγέθους και έντασης ενώ το διάνυσμα κυρίαρχων συχνοτήτων αναδεικνύει κλίμακα και προσανατολισμό, όπως εύκολα παρατηρείται στο σχήμα 4.2.

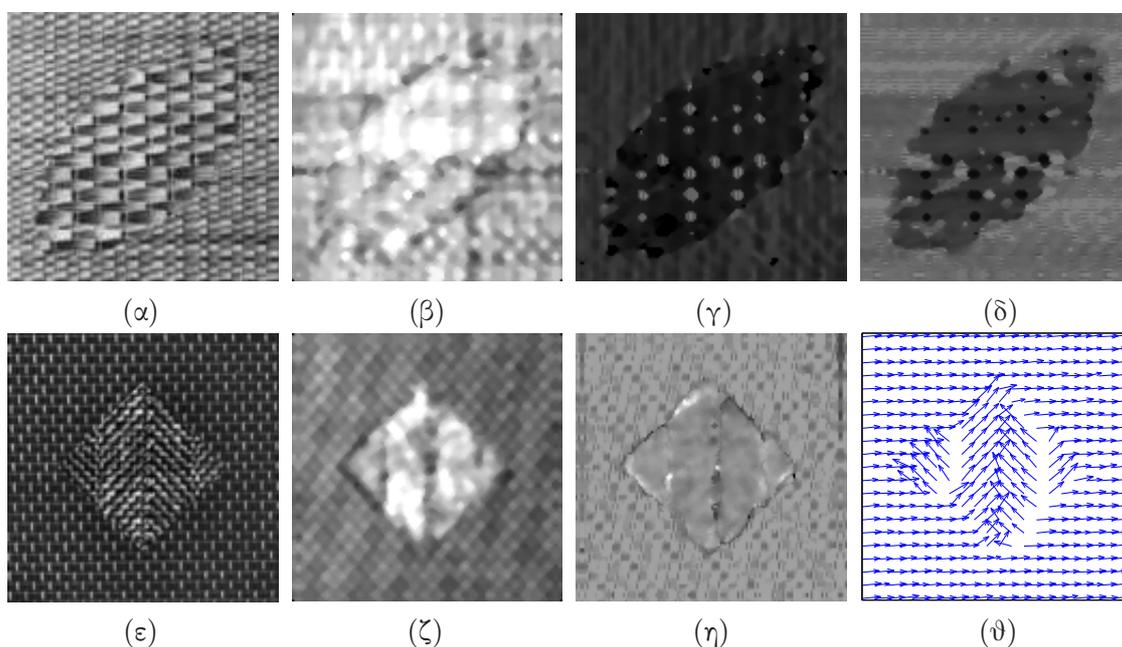
4.2.1 Η επιλογή καναλιού ως στατιστική απόφαση

Το πρόβλημα της επιλογής του καναλιού για την ανάλυση κυρίαρχων συνιστωσών διατυπώνεται ως τεστ πολλαπλών υποθέσεων (multiple hypothesis testing problem) [95]. Για απλότητα εξετάζεται το 1Δ σενάριο, αν και η γενίκευση σε δύο διαστάσεις είναι απλή. Η επιλογή του καναλιού ανάγεται σε ανίχνευση σε θόρυβο ενός ημίτονου της συχνότητας του φίλτρου και άγνωστης φάσης και πλάτους. Τα διαφορετικά κανάλια αντιμετωπίζονται ως ξεχωριστές υποθέσεις και η υπόθεση που εξηγεί καλύτερα τα δεδομένα αντιστοιχεί στο κυρίαρχο κανάλι.

Ένας πρόσθετος παράγοντας είναι η χωρική τοποθέτηση της διαδικασίας απόφασης. Υποθέτοντας ότι σε ένα σημείο απαιτείται η απόφαση για το αν ένα ημίτονο συγκεκριμένης συχνότητας υπάρχει στη γειτονιά του, ελέγχεται πόσο καλά τα σημεία



Σχήμα 4.1: Χαρακτηριστικά διαμόρφωσης κυρίαρχων συνιστωσών με βάση την ενέργεια (EDCA) στις εικόνες 'zebra' και 'house'. Δεξιά προς αριστερά και για τις δύο εικόνες: Γκρίζα εικόνα, κυρίαρχο πλάτος, οριζόντια και κατακόρυφη στιγμιαία συχνότητα.



Σχήμα 4.2: Πληροφορία υφής (μέγεθος, κλίμακα, προσανατολισμός) μέσω DCA. Πάνω σειρά: (α) υφές με διαφορετική κλίμακα, (β) - (δ) Κυρίαρχο πλάτος και στιγμιαίες συχνότητες. Κάτω σειρά: (ε) υφές με διαφορετικό προσανατολισμό, (ζ)-(θ) Κυρίαρχο πλάτος, μέτρο και προσανατολισμός του διανύσματος κυρίαρχης συχνότητας.

της γειτονιάς ικανοποιούν την υπόθεση. Μακρινά σημεία μοντελοποιούνται από ένα λιγότερο περιγραφικό μοντέλο, συγκεκριμένα μια διαδικασία λευκού θορύβου.

Έστω η υπόθεση \mathcal{H}_Ω ότι γύρω από ένα σημείο $x = 0$, το σήμα είναι ημιτονοειδές συχνότητας Ω , άγνωστης φάσης ϕ , πλάτους a και DC συνιστώσας b . Η πιθανοφάνεια (likelihood) του σήματος I στο σημείο x εκφράζεται ως

$$P(I(x); a, b, \phi | \mathcal{H}_\Omega) = \frac{G(x)}{\sqrt{2\pi}\sigma} \exp -\frac{[I(x) - (a \cos(\Omega x + \phi) + b)]^2}{2\sigma^2} + \frac{1 - G(x)}{\sqrt{2\pi}\sigma_n} \exp -\frac{[I(x) - C]^2}{2\sigma_n^2}$$

όπου $G(x) = e^{-x^2/\sigma_G^2}$ ένας παράγοντας υπολογισμού της τοπικότητας της διαδικασίας, σ^2 η μεταβλητότητα λευκού Gaussian θορύβου (WGN) που παρεμβαίνει στο σήμα και το σ_G μοντελοποιεί την χωρική εξασθένιση της ικανότητας της υπόθεσης να εξηγήει τα δεδομένα. Ο δεύτερος όρος είναι το background μοντέλο θορύβου, μέσης τιμής C και μεταβλητότητας σ_n^2 .

Θέτοντας το σ_G ίσο με την τυπική απόκλιση του Gabor φίλτρου που χρησιμοποιείται στο αντίστοιχο κανάλι, προκύπτει μια σχέση ανάμεσα στην επιλογή καναλιού με βάση το πλάτος και στην επιλογή με βάση την πιθανοφάνεια (likelihood). Η πιθανοφάνεια δομής σήματος γύρω από 0 κάτω από την \mathcal{H}_Ω είναι:

$$P(I; a, b, \phi | \mathcal{H}_\Omega) = \prod_{x=-1,0,1,\dots} P(I(x); a, b, \phi | \mathcal{H}_\Omega) \quad (4.1)$$

Λογαριθμίζοντας και με βάση το γεγονός ότι η \log είναι κυρτή συνάρτηση προκύπτει το ακόλουθο κάτω φράγμα στην πιθανοφάνεια $\ln[P(I)] = \sum_x \ln[P(I(x))]$ των δεδομένων:

$$\ln[P(I)] \geq \sum_x G(x) \left\{ -\frac{(I(x) - (a \cos(\Omega x + \phi) + b))^2}{2\sigma^2} - \frac{\ln(2\pi\sigma^2)}{2} \right\} + \sum_x (1 - G(x)) \left\{ -\frac{(I(x) - C)^2}{2\sigma_n^2} - \frac{\ln(2\pi\sigma_n^2)}{2} \right\}$$

Παρότι η τελευταία εξίσωση είναι ένα κάτω όριο μπορεί να χρησιμοποιηθεί για την επιλογή ανάμεσα στις δύο ανταγωνιστικές υποθέσεις. Ένα τεστ γενικευμένου λόγου πιθανοφάνειας (Generalized Likelihood Ratio Test ή GLRT) χρησιμοποιείται για τον υπολογισμό *εκτιμήσεων μέγιστης πιθανοφάνειας* (Maximum Likelihood Estimates ή MLE) των παραμέτρων a, b, ϕ . Μεγιστοποιώντας το (4.1) ως προς τα a, b, ϕ έχουμε

$$\hat{a} = \sqrt{\frac{(h_c * I)^2 + (h_s * I)^2}{(\sum_x G(x))^2}}, \quad \hat{\phi} = \tan^{-1} \left(-\frac{h_s * I}{h_c * I} \right), \quad \hat{b} = G * I \quad (4.2)$$

όπου h_c, h_s ένα ζευγάρι αρτίου-περιττού φίλτρου (quadrature) $h_c(x) = G(x) \cos(\Omega x)$, $h_s(x) = G(x) \sin(\Omega x)$. Με χρήση για την G συνάρτησης διασποράς ίδιας με αυτή του αντίστοιχου Gabor καναλιού, το \hat{a} είναι το εκτιμώμενο πλάτος απόκρισης ενώ το \hat{b} θα μπορούσε να ερμηνευθεί ως ο μέσος των δεδομένων στο σημείο. Ο πρώτος όρος στο κάτω φράγμα αποδεικνύεται ότι είναι ίσος με $\hat{a}^2 / (2\sigma^2 / \sum_x G(x))$, που εκτός από μια σταθερά κλίμακας, είναι το κριτήριο (3.12) για την επιλογή ενός καναλιού. Εφόσον οι υπόλοιποι όροι είναι σταθεροί για όλα τα κανάλια σε μια συγκεκριμένη κλίμακα μπορούν να αγνοηθούν.

Χρησιμοποιώντας ανάλογα επιχειρήματα και λογική, μπορεί να υπολογιστεί η πιθανοφάνεια της υπόθεσης ότι η εικόνα είναι τμηματικά σταθερή γύρω από το 0, απορρίπτοντας την ημιτονική συνιστώσα. Για να συγκριθεί όμως με την υπόθεση ημιτονικών δεδομένων, πρέπει να ληφθεί υπόψιν ότι λιγότερες παράμετροι χρησιμοποιούνται για την κατασκευή του μοντέλου. Διαφορετικά, ένα ημίτονο μηδενικού πλάτους θα μπορούσε το ίδιο καλά να εξηγήσει τα δεδομένα. Για το λόγο αυτό, χρησιμοποιείται το κριτήριο Minimum Description Length (MDL) [95], τροποποιημένο για την περίπτωση

$$\text{MDL}(\mathcal{H}_i) = -\ln(P(I|\mathcal{H}_i)) + \frac{n_i}{2} \ln\left(\sum_x G(x)\right) \quad (4.3)$$

όπου n_i ο αριθμός των παραμέτρων κάθε υπόθεσης.

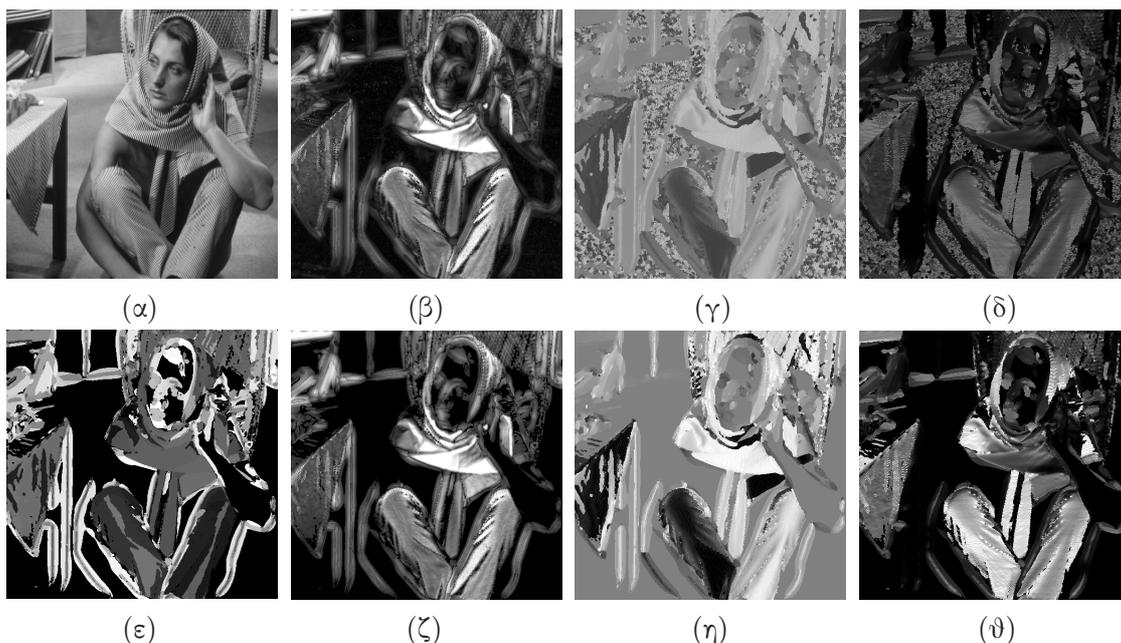
Η εφαρμογή αυτής της στατιστικής απόφασης για την υφή παρουσιάζεται στο σχήμα 4.3, όπου οι δείκτες της υφής παρουσιάζονται ως γκριζες εικόνες. Οι λευκές περιοχές αντιστοιχούν σε περιοχές όπου η υπόθεση μη ύπαρξης υφής επικρατεί, ενώ οι γκριζες τιμές αντιστοιχούν στον προσανατολισμό του Gabor φίλτρου συγκεκριμένης κλίμακας για κάθε εικόνα. Στις μικρές κλίμακες (β), (γ) οι μικροδομές και η λεπτομέρεια ερμηνεύονται ως υφή. Στις μεγάλες κλίμακες, (δ), μεγαλύτερες περιοχές χαρακτηρίζονται από υφή (π.χ. οι πτυχές της μύτης), ενώ ταυτόχρονα περιοχές μικρότερης κλίμακας ερμηνεύονται ως σταθερά σήματα με θόρυβο.



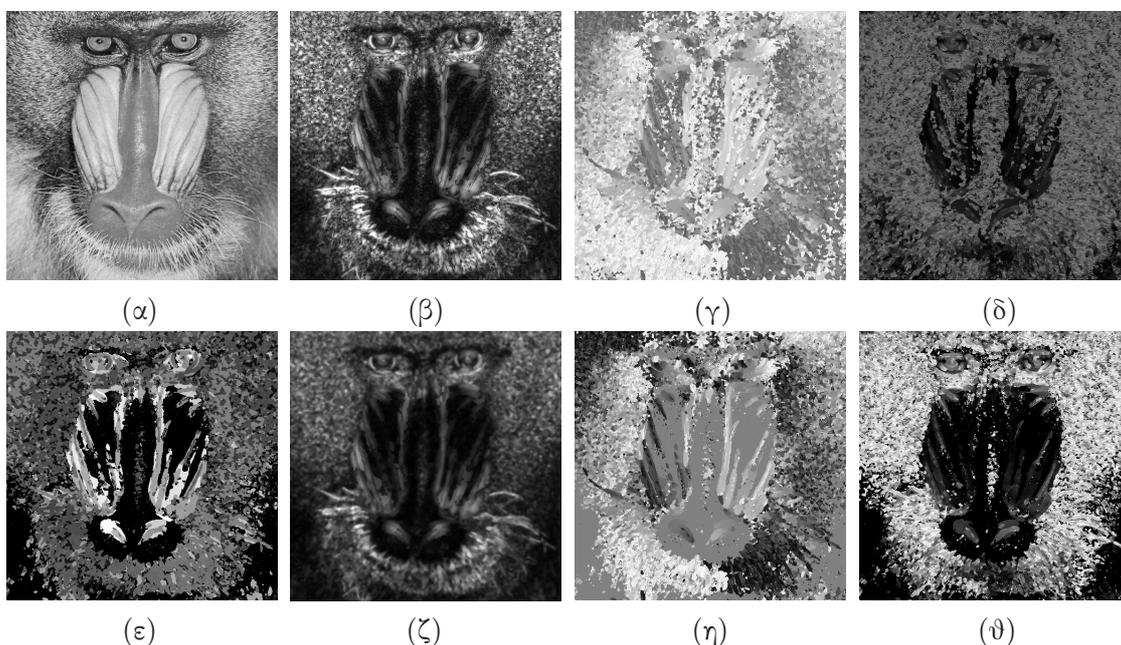
Σχήμα 4.3: Υπόθεση για ύπαρξη υφής, σε κλίμακα του γκριζου, σε τέσσερις κλίμακες για την εικόνα 'mandrill'. Οι κλίμακες, σε άξουσα σειρά προς τα δεξιά, συμπίπτουν με αυτές της συστοιχίας των Gabor φίλτρων ενώ οι γκριζες στάθμες εκφράζουν την πιθανότητα ύπαρξης υφής. Οι λευκές τιμές αντιστοιχούν σε ομαλές περιοχές, όπου τα δεδομένα δεν ερμηνεύονται ως υφή.

Μια πρώτη εφαρμογή της στατιστικής ανίχνευσης υφής είναι η βελτίωση των αρχικών AM-FM χαρακτηριστικών με τον υπολογισμό μιας «μάσκας» υφής/μη-υφής. Για κάθε pixel, μετά την επιλογή του καναλιού, υπολογίζεται ο λόγος πιθανοφάνειας της επικρατούσας υπόθεσης και της υπόθεσης του μοντέλου μη-υφής να εξηγήει στην ίδια κλίμακα τα κοντινά δεδομένα. Σημεία (pixels) όπου η δεύτερη υπόθεση επικρατεί δέχονται τιμές μηδέν για τα χαρακτηριστικά πλάτους και συχνότητων τους. Τα βελτιωμένα χαρακτηριστικά μέσω DCA με βάση την ενέργεια, φαίνονται στο σχήμα 4.4. Όπως φαίνεται στη δεύτερη σειρά (εικόνα 'barb'), η υπόθεση μη-ύπαρξης υφής κυριαρχεί στην περίπτωση του σταθερού δαπέδου, περιοχές οι οποίες αφαιρέθηκαν από τις αρχικές εκτιμήσεις των κυρίαρχων AM-FM συνιστωσών. Αυτό δηλώνουν και τα μαύρα pixel στο 4.4(θ), που κωδικοποιεί τους δείκτες των ενεργών φίλτρων ανά σημείο, και που αντιστοιχούν σε περιοχές που κατηγοριοποιούνται ως μη-υφή.

Ουσιαστικά η εφαρμογή μιας τέτοιας μάσκας απόφασης μπορεί να θεωρηθεί ένα πρώτο στάδιο στατιστικής κατάτμησης και κατηγοριοποίησης σε περιοχές υφής και



Σχήμα 4.4: Χαρακτηριστικά διαμόρφωσης για την εικόνα 'barb' μέσω EDCA. (α)-(δ) Γκριζα εικόνα, κυρίαρχο πλάτος, οριζόντια και κάθετη συχνότητα, (ε) δείκτης κυρίαρχου καναλιού ανά σημείο (ως γκριζα εικόνα), (ζ)-(θ) βελτιωμένα χαρακτηριστικά από την απόρριψη σημείων που χαρακτηρίζονται χωρίς υφή.



Σχήμα 4.5: Χαρακτηριστικά διαμόρφωσης για την εικόνα 'mandrill' μέσω EDCA. (α)-(δ) Γκριζα εικόνα, κυρίαρχο πλάτος, οριζόντια και κάθετη συχνότητα, (ε) δείκτης κυρίαρχου καναλιού ανά σημείο (ως γκριζα εικόνα), (ζ)-(θ) βελτιωμένα χαρακτηριστικά από την απόρριψη σημείων που χαρακτηρίζονται χωρίς υφή. Το πλάτος έχει υποστεί επιπλέον medianφιλτράρισμα.

μη-υφής. Ανάλογα με τη εφαρμογή μπορεί να επιλεχθεί η βελτίωση ή όχι των χαρακτηριστικών, καθώς μια τέτοια απόφαση μπορεί να ενσωματωθεί στον ίδιο τον αλγόριθμο κατάτμησης. Στο [55] πραγματοποιείται μια τέτοια ενσωμάτωση αλλά και μια μίξη κριτηρίων υφής και ακμών, ανάλογα με στατιστική ισχύ τοπικά των διαφορετικών μοντέλων.

4.3 Μεταβολική κατάτμηση χωρίς επισκόπηση

Οι πρώτες region-based τεχνικές κατάτμησης υφής [107, 212] βασίστηκαν σε εκδοχές του συναρτησιακού Mumford-Shah [140] κατάλληλα τροποποιημένες για την ενσωμάτωση πολυδιάστατων χαρακτηριστικών περιγραφής της υφής. Τρέχουσες τεχνικές κατάτμησης χωρίς επισκόπηση [24, 78, 148, 167, 194, 209] βασίζονται στη μεθοδολογία επιπεδοσυνόλων (level sets) [115, 143, 149, 172], που αποτελεί ένα κομψό και αποτελεσματικό μέσο επίλυσης προβλημάτων εξελισσόμενων καμπυλών.

Η μέθοδος ανταγωνισμού περιοχών (Region Competition ή RC) [212], ενοποιεί διαφορετικά κριτήρια και εισάγει μια πιθανοτική θεώρηση στη σχετική βιβλιογραφία. Βασίζεται στη μεγιστοποίηση της πιθανότητας των παρατηρήσεων της εικόνας I , χρησιμοποιώντας ένα σύνολο M περιοχών (regions) R_i , εντός των οποίων οι παρατηρήσεις ακολουθούν υποθετικά μια παραμετρική κατανομή $P(\cdot; a_i)$. Ένας επιπλέον όρος για τα μήκη των συνόρων των περιοχών Γ_i χρησιμοποιείται για να προκύψει το συναρτησιακό

$$J(\Gamma, \{a_i\}) = \sum_{i=1}^M \frac{\mu}{2} \int_{\Gamma_i} ds - \iint_{R_i} \log P(I; a_i) \quad (4.4)$$

Τεχνικές λογισμού μεταβολών καταλήγουν στην εξέλιξη των συνόρων των περιοχών, με κίνηση στην κατεύθυνση που κατατάσσει παρατηρήσεις pixel στην περιοχή που τα μοντελοποιεί καλύτερα, διατηρώντας ομαλά σύνορα

$$\frac{\partial \Gamma_i}{\partial t} = -\mu \kappa \vec{N} + \log \frac{P(I; a_i)}{P(I; a_j)} \vec{N} \quad (4.5)$$

όπου κ η καμπυλότητα (curvature) και \vec{N} το κάθετο εξωτερικό διάνυσμα της καμπύλης (front) Γ_i και j η γειτονική περιοχή που ανταγωνίζεται την i για τις παρατηρήσεις.

Η κατανομές εντός των περιοχών μπορούν να θεωρηθούν πολυμεταβλητές Gaussian, όπου για d πλήθος χαρακτηριστικών

$$P(I; \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} e^{-\frac{1}{2}(I-\mu_i)^T \Sigma_i^{-1} (I-\mu_i)} \quad (4.6)$$

Η εκτίμηση των παραμέτρων $a_i = \mu_i, \Sigma_i$ των κατανομών πραγματοποιείται σε εναλλαγή με την εξέλιξη των καμπυλών, με αποτέλεσμα ένα προσαρμοστικό, χωρίς επισκόπηση, σχήμα κατάτμησης εικόνων.

Στο [148] η εξέλιξη των καμπυλών συνδυάστηκε με όρους ακμών (edge-based) και μεθόδους level sets, στον αλγόριθμο γεωδαιτικών ενεργών περιγραμμάτων (Geodesic Active Regions ή GAR)

$$\frac{\partial \Gamma_i}{\partial t} = \lambda \log \frac{P(I; a_i)}{P(I; a_j)} - (1 - \lambda) \left[g(P_c) \kappa \vec{N} + (\nabla g(P_c) \cdot \vec{N}) \vec{N} \right] \quad (4.7)$$

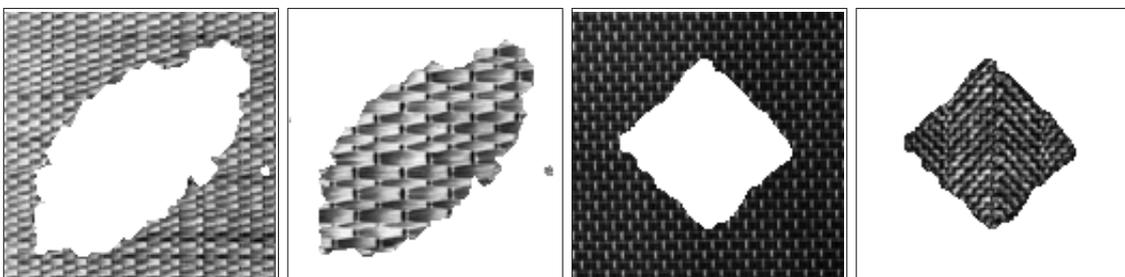
όπου $g(\cdot)$ είναι μια μονοτονικά αύξουσα συνάρτηση, P_c η πιθανότητα ενός pixel να ανήκει σε ένα σύνορο και το λ καθορίζει τα σχετικά βάρη των όρων που βασίζονται στις περιοχές και των όρων ακμών. Ο τελευταίος όρος είναι εμπνευσμένος από το μοντέλο Geodesic Active Contours (GAC) [30, 97] και ωθεί τα σύνορα κοντά σε περιοχές έντονων ακμών. Άλλοι αλγόριθμοι όπως οι [194, 209] μπορούν να θεωρηθούν της ίδιας φύσης με τον αρχικό RC.

4.4 AM-FM ανταγωνισμός περιοχών

Το σχήμα κατάτμησης Region Competition προσαρμόζεται και η αναζήτηση της βέλτιστης κατανομής των σημείων μιας εικόνας υφής, γίνεται με βάση τα κυρίαρχα χαρακτηριστικά διαμόρφωσης. Η εξαγωγή γίνεται με την EDCA μεθοδολογία και τον εντοπισμό της κυρίαρχης ενέργειας διαμόρφωσης. Το τελικό διάνυσμα χαρακτηριστικών αποτελείται από την ένταση φωτεινότητας της εικόνας, το κυρίαρχο πλάτος διαμόρφωσης, μέτρο και τον προσανατολισμό του διανύσματος στιγμιαίων συχνοτήτων. Η ένταση I της εικόνας συμπεριλαμβάνεται καθώς εξακολουθεί να παραμένει σημαντικό κριτήριο για τις περιοχές χωρίς υφή [167].

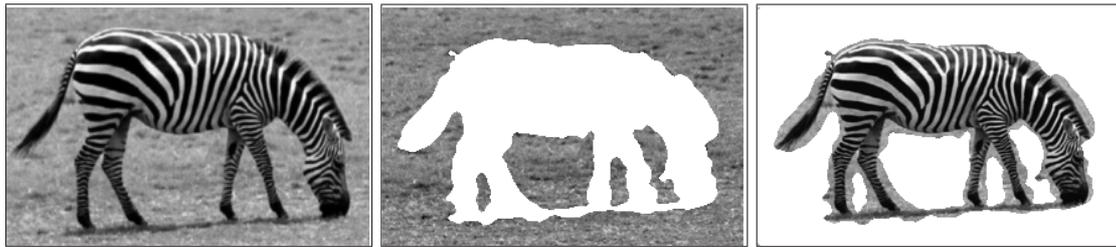
Η εξέλιξη καμπυλών όπως καθορίζεται από την εξ. (4.4) υλοποιείται με μεθόδους level sets, παρόμοιας αρχιτεκτονικής με το [148]. Η κατανομή του διανύσματος χαρακτηριστικών σε κάθε περιοχή μοντελοποιείται ως γινόμενο πολυμεταβλητών Gauss κατανομών της μορφής (4.6), για τις τρεις πρώτες διαστάσεις, και von-Mises κατανομής για τον προσανατολισμό. Παραδείγματα μη-επιβλεπόμενης κατάτμησης, με τυχαία τοποθέτηση και προεπιλογή του αριθμού των αρχικών καμπυλών (fronts) φαίνεται στα σχήματα 4.6-4.10.

Στο 4.6 φαίνεται το αποτέλεσμα της διαδικασίας σε δύο συνθετικές εικόνες υφής, όπου τα κεντρικά τμήματα διαφέρουν ως προς την κλίμακα στη μία και ως προς τον προσανατολισμό στην άλλη περίπτωση. Στη φυσική εικόνα 4.7, η έντονη υφή-πρότυπο στο δέρμα της ζέβρας διαχωρίζεται από το σχεδόν ομοιόμορφο φόντο. Στην 4.8 οι στήλες, λόγω του σχεδόν σταθερού πλάτους (σχήμα 4.1(ζ)) ανιχνεύονται ως ξεχωριστή περιοχή ενώ τα σκαλιά λόγω της έντονης κατακόρυφης κυρίαρχης συχνότητας (σχήμα 4.1(θ)). Παρόμοια συμπεράσματα και παρατηρήσεις εξάγονται και από τα υπόλοιπα σύνολα αποτελεσμάτων (4.9 & 4.10). Σημειώνεται ότι σημαντικό ρόλο παίζει η επιλογή του αρχικού αριθμού των fronts. Η αυτόματη επιλογή του αποτελεί δύσκολο πρόβλημα και σε κάθε περίπτωση διαφορετική επιλογή οδηγεί σε διαφορετική ομαδοποίηση των pixel της εικόνας.

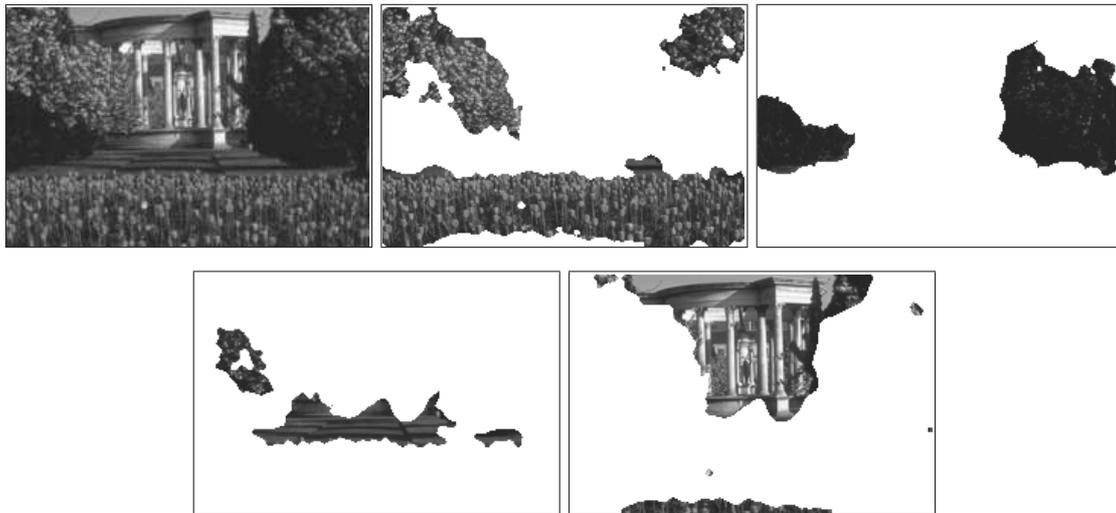


Σχήμα 4.6: Κατάτμηση για τις εικόνες του Σχ. 4.2 (α) και (ε), σε περιοχές διαφορετικής κλίμακας και προσανατολισμού.

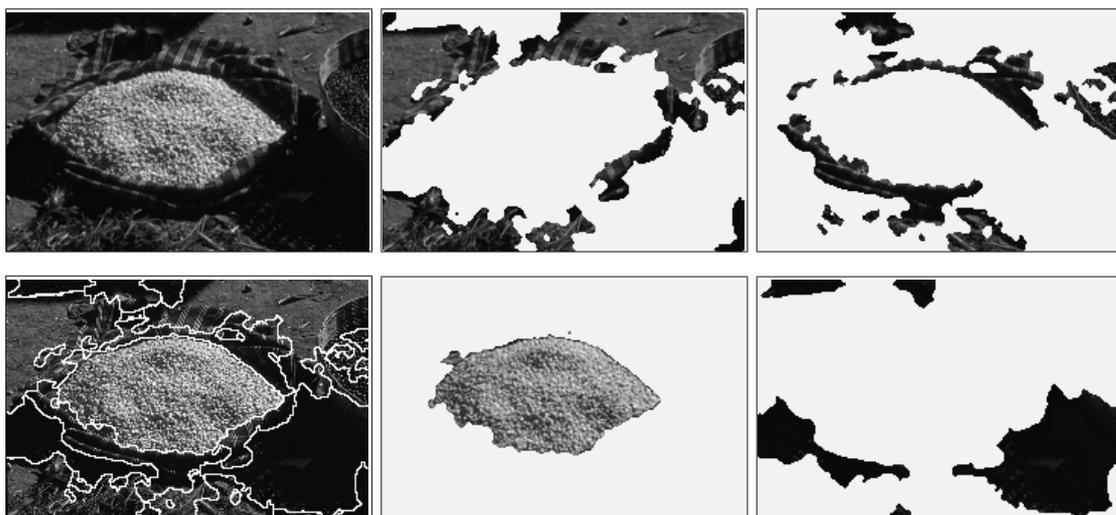
□



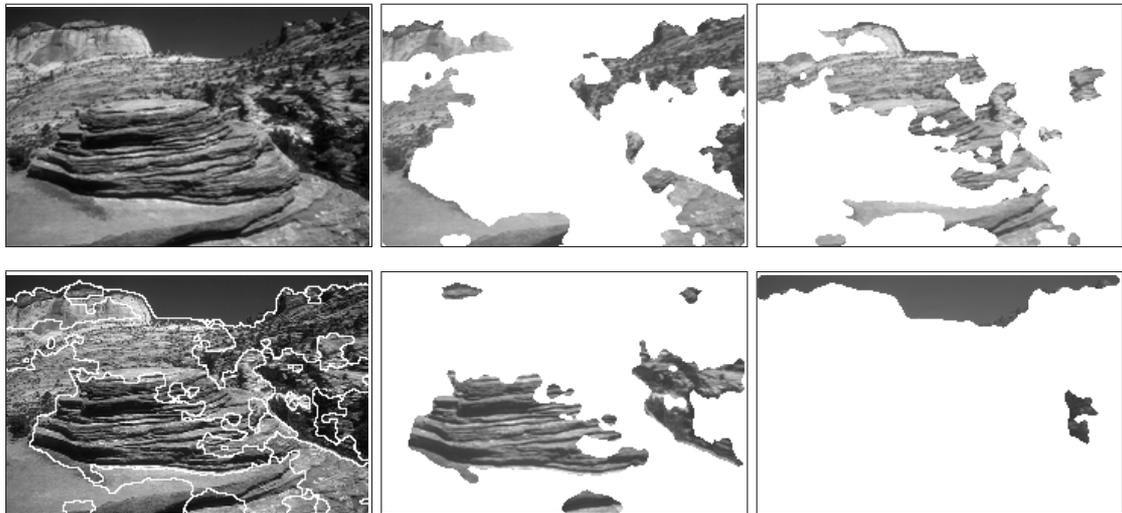
Σχήμα 4.7: Κατάτμηση υφής και οι δύο περιοχές ομοιόμορφων AM-FM κυρίαρχων χαρακτηριστικών διαμόρφωσης.



Σχήμα 4.8: Κατάτμηση υφής για την εικόνα του Σχ. 4.1(ε) σε 4 περιοχές ομοιόμορφων AM-FM κυρίαρχων χαρακτηριστικών διαμόρφωσης.



Σχήμα 4.9: Κατάτμηση υφής σε 4 περιοχές ομοιόμορφων AM-FM κυρίαρχων χαρακτηριστικών διαμόρφωσης.



Σχήμα 4.10: Κατάτμηση υφής σε 4 περιοχές ομοιόμορφων AM-FM κυρίαρχων χαρακτηριστικών διαμόρφωσης.

Κεφάλαιο 5

Ανάλυση Εδαφοτομών

Μια κατηγορία εικόνων με έντονη υφή και παρουσία διαφορετικών δομών είναι οι ψηφιακές εικόνες λεπτών τομών εδάφους ή εδαφοτομών (soilsections). Η αυτόματη επεξεργασία τέτοιων εικόνων μπορεί να αποκαλύπτει στοιχεία της δομής του εδάφους και να εξάγει εκτιμήσεις βιολογικής σημασίας, όπως η γονιμότητα και οι αλλαγές σε οικοσυστήματα του εδάφους. Παρουσιάζεται εδώ η εξαγωγή χαρακτηριστικών, η ανάλυση υφής και η κατάτμηση τέτοιων εικόνων σε ομογενείς περιοχές, σχετικές με την μικρομορφολογία του εδάφους. Με παράλληλη, χωριστή εκμετάλλευση πληροφορίας γεωμετρίας ή αντίθεσης και υφής, παρουσιάζεται ένα σχήμα συζευγμένης (coupled) ή από κοινού κατάτμησης (joint segmentation) για περαιτέρω ερμηνεία εικόνων του εδάφους και μετρήσεις χαρακτηριστικών. Τέλος ελέγχεται η κατηγοριοποίηση των διαφορετικών περιοχών και δομών τέτοιων εικόνων με βάση την υφή τους.¹

5.1 Περιγραφή εδαφοτομών και ανάλυση εδάφους

Ο σκοπός της μικρομορφολογικής ανάλυσης (soil micromorphology), ως κλάδος της επιστήμης του εδάφους (soil science) είναι η περιγραφή, ερμηνεία και μέτρηση των συστατικών και των χαρακτηριστικών του χώματος σε μικροσκοπικό επίπεδο. Βασικά συστατικά του χώματος είναι τα μεμονωμένα σωματίδια (π.χ. κόκκοι κρυστάλλων, οργανικά άλατα, τμήματα φυτών) που μπορούν να διακριθούν με το οπτικό μικροσκόπιο (coarse structures), μαζί με τη μικρομάζα ή μικροδομή (fine material) που δε διαχωρίζεται σε διακριτά άτομα [26]. Η μελέτη της δομής του εδάφους ασχολείται με το μέγεθος, το σχήμα, την οξύτητα, την αντίθεση, τη συχνότητα εμφάνισης και τη χωρική τοποθέτηση των πρωτευόντων σωματιδίων και των κενών (voids). Πολλά από αυτά τα χαρακτηριστικά είναι συνάρτηση του προσανατολισμού των συστατικών και της κατεύθυνσης στην οποία γίνονται οι τομές, καθώς και της μεγέθυνσης που χρησιμοποιείται.

Οι εικόνες εδαφοτομών, που παράγονται μέσω διαδικασιών διακριτοποίησης, χρησιμοποιώντας συμβατικούς σαρωτές, κάμερες ή μικροσκόπια πολωμένου φωτός, παρουσιάζουν μια μεγάλη διαφοροποίηση γεωμετρικών χαρακτηριστικών, προτύπων και σχηματισμών υφής. Σημαντικά χαρακτηριστικά εικόνων που παρέχουν πληροφορία για την εκτίμηση της ποιότητας της δομής του εδάφους, περιλαμβάνουν το σχήμα σωματιδίων και συσωματωμάτων, είτε 1Δ όπως είναι οι ακμές ή οι καμπύλες, είτε 2Δ όπως

¹Η έρευνα και τα αποτελέσματα αυτής της ενότητας ανήκουν στο πλαίσιο του έργου ΠΕΝΕΔ-01, σε συνεργασία με την Νατάσα (nat) Σοφού [177, 178].

φωτεινά ή σκοτεινά blobs (μικρές ομογενείς περιοχές τυχαίου σχήματος), την χωρική τοποθέτηση και την απόσταση συστατικών και την υφή τους.

Η ως τώρα ανάλυση εικόνων εδαφοτομών περιορίζεται σε κατωφλιοποιήσεις, ευριστικά, μετρήσεις ιστογραμμάτων, απλές τεχνικές ψηφιακής επεξεργασίας σημάτων και εκτιμήσεις φράκταλ διάστασης [25,42,159,187]. Ο ευρύτερος, μακροπρόθεσμος στόχος αυτής της έρευνας είναι να ξεπεραστούν αυτές οι τεχνικές από σύγχρονες μεθόδους υπολογιστικής όρασης όπως η ανάλυση υφής και η κατάτμηση εικόνων, με βλέψη ένα σύστημα υψηλού επιπέδου ικανό να αναλύει και να εκτιμά την γονιμότητα και την βιολογική ποιότητα του εδάφους μέσω των εικόνων τομών [127]. Έτσι η ακριβή και χρονοβόρα διαδικασία της παραδοσιακής βιοχημικής ανάλυσης να αυτοματοποιηθεί, να ενισχυθεί και πιθανόν να συνδυαστεί με τεχνικές όρασης.

Η κατάτμηση είναι ο ακρογωνιαίος λίθος στην ανάλυση των εδαφοτομών, με διαδικασίες επεξεργασίας και μετρήσεων να ακολουθούν το στάδιο του διαχωρισμού των περιοχών. Μια τέτοια ανάλυση περιλαμβάνει εκτίμηση της ομοιογένειας των περιοχών, ανάλυση υφής και κατηγοριοποίηση με τα χαρακτηριστικά που έχουν εξαχθεί, και στατιστική ανάλυση μεγεθών όπως η περίμετρος, το εμβαδό, ο προσανατολισμός, η συγκέντρωση, η επιμήκυνση, το μέγεθος, η καμπυλότητα κ.α.

5.2 Υφή εδαφοτομών, ανάλυση και κατάτμηση

5.2.1 Μορφολογική κατάτμηση

Η κατάτμηση είναι ένα σημαντικό αλλά εν γένει δύσκολο πρόβλημα της υπολογιστικής όρασης καθώς απαιτεί σε κάποιο ποσοστό την σημασιολογική κατανόηση της εικόνας. Γενικά είναι η διαδικασία του χωρισμού μιας εικόνας σε ασύνδετες περιοχές, έτσι ώστε κάθε μια από τις οποίες να είναι ομογενής και συνεκτική ως προς κάποια ιδιότητα, όπως η αντίθεση, το χρώμα, η υφή, η κίνηση, το βάθος. Μπορούν να διακριθούν τρία διαφορετικά στάδια δηλ. (1) προεπεξεργασία, (2) εξαγωγή χαρακτηριστικών περιοχών, (3) εφαρμογή του αλγορίθμου κατάτμησης, κάθε ένα από τα οποία είναι σημαντικό για το τελικό αποτέλεσμα της κατάτμησης.

Από το πλήθος μεθόδων κατάτμησης ο *μορφολογικός μετασχηματισμός πλημμυρισμού* (watershed transform) έχει αποδειχθεί πολύ ισχυρός και αποτελεσματικός ιδιαίτερα μέσω σύζευξης με μη-γραμμικούς μορφολογικούς τελεστές [17, 198]. Η θεμελιώδης ιδέα είναι η ακόλουθη: (α) δημιουργείται η εικόνα κλίσης (gradient) της σκηνής· (β) για κάθε αντικείμενο ενδιαφέροντος ή ομογενή περιοχή, ανιχνεύεται ένας «σημαδευτής» ή marker (σύνολο σημείων χαρακτηριστικών μέσα στην επιθυμητή περιοχή), (γ) κατασκευάζονται οι υδατογραμμές (watershed lines) που σχετίζονται με τους σημαδευτές.

Ο μετασχηματισμός watershed μπορεί να περιγραφεί τοπογραφικά ως μια διαδικασία «πλημμυρισμού», όπου η συνάρτηση της εικόνας θεωρείται μια τοπογραφική επιφάνεια βυθισμένη στο νερό. Οι σημαδευτές παίζουν το ρόλο των πηγών πλημμυρισμού, απ' όπου τα διάφορα κύματα αρχίζουν να σχηματίζουν «λίμνες». Στα σημεία όπου διαφορετικά κύματα συναντιούνται, ένα «φράγμα» ανεγείρεται για την αποφυγή της ένωσης των λιμνών. Το σύνολο των φραγμάτων είναι η γραμμή watershed που χωρίζει την εικόνα σε διαφορετικές περιοχές. Στη μαθηματική μορφολογία το watershed έχει υλοποιηθεί μέσω προσομοιώσεων εμβύθισης (immersion simulations) [198] και ιεραρχικές ουρές προτεραιότητας (hierarchical queues) [17].

Εκτός από τη μορφολογική διαδικασία πλημμυρισμού, το watershed μοντελοποιήθηκε επίσης με συνεχή τρόπο μέσω της eikonal Μερικής Διαφορικής Εξίσωσης (ΜΔΕ) [135,142], χρησιμοποιώντας ιδέες από το πεδίο της εξέλιξης καμπυλών [124]. Τα κίνητρα για τη χρήση ΜΔΕ περιλαμβάνουν καλύτερη και περισσότερο διαισθητική μαθηματική μοντελοποίηση, σύνδεση με τη φυσική και καλύτερη προσέγγιση της συνεχούς γεωμετρίας του προβλήματος. Χρησιμοποιώντας ΜΔΕ μοντελοποίηση στη διαδικασία πλημμυρισμού του μετασχηματισμού το σύνορο κάθε αναδιδόμενου κύματος αντιμετωπίζεται ως καμπύλη η οποία εξελίσσεται με προκαθορισμένη ταχύτητα. Συγκεκριμένα το σύνορο του σηματοδευτή θεωρείται ως μια ομαλή, κλειστή καμπύλη $\vec{C}(p, t)$ όπου $p \in [0, 1]$ παραμετροποιεί την καμπύλη και t μια τεχνητή παράμετρος εξέλιξης. Η ΜΔΕ που υλοποιεί τον watershed πλημμυρισμό είναι

$$\frac{\partial \vec{C}}{\partial t} = \frac{c}{A(t)\|\nabla I\|} \cdot \vec{N} \quad (5.1)$$

όπου το $A(t)$ είναι 1 για κατάτμηση βασισμένη μόνο στη χρωματική αντίθεση (contrast-based), που αναφέρεται ως πλημμυρισμός ύψους (height flooding) ή $A(t) = \text{Area}(\vec{C})$ στην περίπτωση κατάτμησης με βάση αντίθεση και μέγεθος (contrast and size), οπότε και αναφέρεται ως πλημμυρισμός όγκου ή ύψους (volume flooding) [179].

Σύμφωνα με την προσέγγιση μέσω επιπεδοσυνόλων [143], όπου η εξελισσόμενη καμπύλη εμβυθίζεται ως το μηδενικό επιπεδοσύνολο $\Gamma(t) = \{(x, y) : \Phi(x, y, t) = 0\}$ μιας χωρο-χρονικής συνάρτησης μεγαλύτερης διάστασης $\Phi(x, y, t)$, η ΜΔΕ που διέπει την εξέλιξη είναι

$$\frac{\partial \Phi}{\partial t} = \frac{c}{A(t)\|\nabla I(x, y)\|} \|\nabla \Phi\| \quad (5.2)$$

Αποτελεσματικοί αλγόριθμοι για την επίλυση χρονικά εξαρτώμενων eikonal ΜΔΕ, είναι η μέθοδος επιπεδοσυνόλων στενής ζώνης (narrow-band level sets) και η μέθοδος ταχείας προέλασης (fast marching) [173, 174].

5.2.2 Διαχωρισμός γεωμετρίας και υψής με levelings

Στην ενότητα 2.4.3 έγινε εκτενής αναφορά στη $U + V$ μεθοδολογία διαχωρισμού εικόνων [195], όπου μια εικόνα I διαχωρίζεται σε δύο συνιστώσες. Η μια διατηρεί πληροφορία γεωμετρίας και ομαλών μεταβολών όπως γεωμετρικής δομής, τοπικής αντίθεσης, ακμών, σχήματος ή κυρίαρχων αντικειμένων και η άλλη πληροφορία υψής. Διάφορες μη-γραμμικές τεχνικές ομαλοποίησης εικόνων δημιουργούν «καρτούν» προσεγγίσεις, όπως το σχήμα ελαχιστοποίησης ολικής μεταβολής [168] ή η ανισοτροπική διάχυση [153].

Ως προσέγγιση της U συνιστώσας, προτείνονται τα levelings [136], μη-γραμμικά, αντικειμενοστραφή φίλτρα, που διατηρούν τα περιγράμματα. Η απλοποίηση μιας εικόνας I γίνεται επεκτείνοντας και συστέλλοντας τοπικά μια αρχική εικόνα, τον marker M , και περιορίζοντας συνολικά την εξέλιξη του από την εικόνα αναφοράς. Πιο συγκεκριμένα, επαναλήψεις του τελεστή εικόνων $\lambda(F|I) = (\delta(F) \wedge I) \vee \varepsilon(F)$, όπου $\delta(F)$ και $\varepsilon(F)$ είναι διαδικασίες dilation και erosion αντίστοιχα του F με ένα μικρό δίσκο [171], δίνουν στο όριο το leveling της I σε σχέση με τον M , που συμβολίζεται ως

$$\Lambda(M|I) = \lim_{k \rightarrow \infty} F_k, \quad F_k = \lambda(F_{k-1}|I), \quad F_0 = M \quad (5.3)$$

Τα levelings έχουν πολλές ενδιαφέρουσες ιδιότητες όπως ότι διατηρούν τις ακμές, δεν εισάγουν νέα ελάχιστα ή μέγιστα περιοχής και αποτελούν αύξοντα και ταυτοδύναμα

(idempotent) φίλτρα. Πρακτικά μπορούν να ανακατασκευάσουν ολόκληρα αντικείμενα της εικόνας διατηρώντας τις ακμές και τα σύνορα τους. Κατά την ανακατασκευή απλοποιούν την αρχική εικόνα εξαλείφοντας αντικείμενα μικρότερα από τον σηματοδευτή με την εικόνα αναφοράς να παίζει το ρόλο ενός ολικού περιορισμού.

Με γνώμονα τις παραπάνω ιδιότητες, μια εναλλακτική μέθοδος για U, V διαχωρισμό μιας εικόνας είναι υιοθετώντας ως «καρτούν» συνιστώσα το leveling της εικόνας και ως συνιστώσα υψής τη διαφορά του από αυτή

$$U \triangleq \Lambda(M|I), \quad V \triangleq I - U \quad (5.4)$$

Για το U η επιλογή του σηματοδευτή M προσφέρει ευελιξία για το τελικό leveling και μπορεί να οριστεί μέσω ανάλυσης σε πολλαπλές κλίμακες. Συγκεκριμένα κατασκευάζονται πολυκλιμακωτά leveling «καρτούν» $U_i = \Lambda(M_i|U_{i-1})$, $i = 1, 2, 3, \dots$, με $U_0 = I$ από μια ακολουθία σηματοδευτών M_i , που λαμβάνονται δειγματοληπτώντας μια Gaussian πυραμίδα (scale-space). Τα πολυκλιμακωτά «καρτούν» διαθέτουν την ιδιότητα της αιτιατότητας, καθώς το U_j είναι leveling του U_i για $j > i$. Τα αντίστοιχα υπόλοιπα $V_i = I - U_i$ συνιστούν μια ιεραρχία από συνιστώσες υψής σε πολλαπλές κλίμακες.

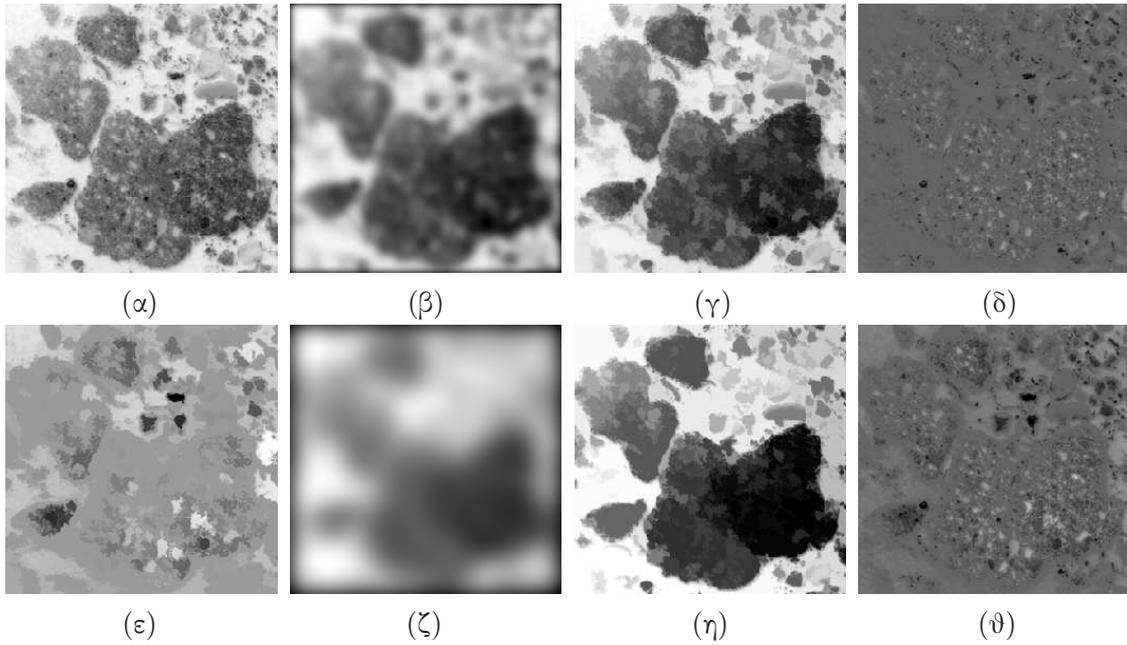
Εναλλακτικά για την επιλογή των σηματοδευτών εξετάστηκε και η χρήση ανισοτροπικής διάχυσης [153], όπου σε κάθε βήμα της ακολουθίας ο leveling marker λαμβάνεται από μια εκδοχή της εικόνας με θολωμένες περιοχές αλλά αρκετά διατηρημένα σύνορα, λόγω της διαδικασίας περιορισμένης διάχυσης. Στο σχήμα 5.1, φαίνονται δύο από τα στάδια της leveling πυραμίδας, όπου οι ακμές και τα σύνορα των κυρίαρχων γεωμετρικών δομών της εικόνας, διατηρούνται στις «καρτούν» συνιστώσες σε διαφορετικές κλίμακες.

Συγκρίνοντας τα leveling «καρτούν» με τα αυτά που λαμβάνονται με την επίλυση του προβλήματος TV ελαχιστοποίησης (2.63): (α) Τα levelings διατηρούν τα ακρότατα των περιοχών και δεν δημιουργούν νέα ενώ τα TV «καρτούν» διατηρούν την ολική μέση τιμή. (β) Τα levelings διατηρούν τη σχέση μεταβολής μεταξύ γειτονικών σημείων ενώ τα TV διατηρούν την ολική μεταβλητότητα. (γ) Τα levelings σχετίζονται με μια ελαχιστοποίηση ολικής μεταβολής, καθώς για $U = \Lambda(M|I)$, είναι $\iint \|\nabla U\| < \iint \|\nabla I\|$. (δ) Η παρουσία του marker M δίνει στο leveling «καρτούν» ευελιξία και πολυκλιμακωτές δυνατότητες σε σχέση με τις απλές σταθερές κανονικοποίησης που ελέγχουν την κλίμακα του TV «καρτούν».

5.2.3 Συζευγμένη, πολυκριτηριακή κατάτμηση εδαφοτομών

Με διαχωρισμό της εικόνας στις δύο συνιστώσες, U, V και χωριστή, εξειδικευμένη επεξεργασία τους προκύπτει ένα ισχυρό σχήμα πολυκριτηριακής κατάτμησης, με σύζευξη πληροφορίας γεωμετρίας και υψής. Οι μεταβολές της αντίθεσης (ή του μεγέθους) λαμβάνονται υπόψιν μέσω του U κομματιού και οι ταλαντώσεις υψής περιγράφονται από τη V συνιστώσα. Η συμβολή και η ποσοτικοποίηση της υψής γίνεται μέσω της διαδικασίας ανίχνευσης και μέτρησης της μέγιστης μέσης Teager ενέργειας ή ενέργειας ταλαντώσεων υψής Ψ_{mat} που ορίστηκε στην ενότητα 3.4 και δίνεται από την εξ.(3.14). Με τον τρόπο αυτό συμπεριλαμβάνεται ένα μετρικό στη διαδικασία της μεταβολικής κατάτμησης που σχετίζεται με το είδος και την ένταση της υψής στις περιοχές της εικόνας.

Η προτεινόμενη διαδικασία κατάτμησης βασίζεται στην εξέλιξη καμπύλης που κα-



Σχήμα 5.1: Πολυκλιμακωτά leveling «καρτούν» και $U + V$ διαχωρισμός. Πάνω σειρά: (α) Λεπτομέρεια εδαφοτομής (αναφορά), (β) Gauss σηματοδευτής 1 ($\sigma = 4$), (γ) Leveling 1 (U_1), (δ) Διαφορά 1 ($V_1 = I - U_1 + 100$). Κάτω σειρά: (ε) Διαφορά $U_1 - U_2$, (ζ) Gauss σηματοδευτής 2 ($\sigma = 12$), (η) Leveling 2 (U_2), (θ) Διαφορά 2 ($V_2 = I - U_2 + 100$).

θορίζει η ακόλουθη ΜΔΕ :

$$\frac{\partial \vec{C}}{\partial t} = \left(\frac{\lambda_1}{A(t) \|\nabla f_1\|} + \lambda_2 \Psi_{\text{mat}}(f_2) - \mu \kappa \right) \vec{N} \quad (5.5)$$

όπου f_1 και f_2 είναι μετασχηματισμοί της εικόνας σχετιζόμενοι με την αρχική I αλλά όχι απαραίτητα ίδιοι. Η ταχύτητα της καμπύλης εξαρτάται από τρεις όρους: οι δύο πρώτοι είναι εικοναλ ενώ ο τρίτος, η ταχύτητα καμπυλότητας, είναι διαχυτικός. Όλοι οι όροι σχετίζονται με κάποιο κριτήριο βελτιστοποίησης. Ο πρώτος όρος οδηγεί την καμπύλη με ταχύτητα που μεγιστοποιεί τον πλημμυρισμό της εικόνας f_1 ως προς το watershed της. Έστω $\vec{C}(t)$ και $\vec{C}(t + \Delta t)$ δύο στιγμιότυπα της εξελισσόμενης καμπύλης που συμπίπτουν με τα επιπεδοσύνολα της f_1 σε ύψη h και $h + \Delta h$ αντίστοιχα. Τότε για τον πρώτο όρο $\Delta h = f_1(x + \Delta x, y + \Delta y) - f_1(x, y) = (\nabla f_1 \cdot \vec{C}_t) \Delta t$. Διατηρώντας την κάθετη ταχύτητα πλημμυρισμού σταθερή $\lambda = \Delta h / \Delta t$, ο όρος μεγιστοποιείται όταν

$$\frac{\partial \vec{C}}{\partial t} = \frac{\lambda}{\|\nabla f_1\|}$$

το οποίο θυμίζει την συνάρτηση τερματισμού ακμών $g(\cdot)$ των GAC με διαφορετική ρίζα. Ο δεύτερος όρος μπορεί ναδειχθεί ότι αντιστοιχεί σε ροή που μεγιστοποιεί την ενέργεια διαμόρφωσης υψής

$$\max \iint_{R(C)} \Psi_{\text{mat}}(f) \implies \frac{\partial \vec{C}}{\partial t} = \Psi_{\text{mat}}(f) \vec{N}$$

με αποτέλεσμα να ωθεί την καμπύλη σε περιοχές με έντονη «δραστηριότητα» υψής.

Ακολουθώντας την θεώρηση επιπεδοσυνόλων [143], η συνάρτηση εμβυθίζεται ως το μηδενικό επιπεδοσύνολο μιας χώρο-χρονικής συνάρτησης $\Phi(x, y, t)$. Η ΜΔΕ που

καθορίζει την εξέλιξη της επίπεδο-συνάρτησης είναι:

$$\frac{\partial \Phi}{\partial t} = \left(\frac{\lambda_1}{A(t) \|\nabla f_1\|} + \lambda_2 \Psi_{\text{mat}}(f_2) - \mu \text{curv}(\Phi) \right) \|\nabla \Phi\| \quad (5.6)$$

όπου $\text{curv}(\Phi)$ η καμπυλότητα των επιπεδοσυνόλων της Φ .

Με βάση τη ΜΔΕ (5.5) μπορούν να προκύψουν διαφορετικά σενάρια μεταβάλλοντας τα σήματα f_1 και f_2 . Η πιο προφανής επιλογή είναι $f_1 = I$, $f_2 = I$ που αντιστοιχεί σε εξέλιξη με ταχύτητα αντιστρόφως ανάλογη της αντίθεσης φωτεινότητας (ή μεγέθους) της εικόνας και ανάλογη της Ψ_{mat} ενέργειας της εικόνας. Προτείνεται επίσης και το πρωτότυπο, εναλλακτικό σενάριο $f_1 = U$, $f_2 = V$ όπου η εξέλιξη καμπύλης είναι αντιστρόφως ανάλογη της αντίθεσης φωτεινότητας (ή μεγέθους) της «καρτούν» συνιστώσας και ανάλογη της Ψ_{mat} ενέργειας της συνιστώσας υψής της εικόνας. Αυτό το δεύτερο σενάριο παρουσιάζει ιδιαίτερο ενδιαφέρον και εξετάζεται περισσότερο, καθώς ενοποιεί πληροφορία ακμών και υψής, συνδυάζοντας τα διαφορετικά σήματα που προκύπτουν από τον $U + V$ διαχωρισμό της εικόνας.

Μια τέτοια επιλογή αιτιολογείται καλύτερα από διάφορες οπτικές: (α) *δαισθητικά*, για μια από κοινού κατάτμηση, αναμένεται μια συνέργια διαφορετικών, παράλληλων διαδικασιών στις μακροδομές (γεωμετρία) και στις μικροδομές (υψή) της εικόνας, (β) *υπολογιστικά* όσο καλύτερη είναι η απομάκρυνση από το προς ανάλυση κομμάτι της υψής, των συνόρων, περιγραμμάτων και επίπεδων επιφανειών, τόσο περιορίζονται οι τιμές της Ψ_{mat} που δεν αντιστοιχούν σε μεταβολές υψής, αλλά σε ακμές και όρια περιοχών.

Στο προτεινόμενο σχήμα υπάρχουν πολλαπλές καμπύλες που πρόκειται να εξελιχθούν ταυτόχρονα και αρχικοποιούνται σαν τα περιγράμματα ενός συνόλου σηματοδευτών, ενδεικτικών των σημαντικών ή(και) ομογενών περιοχών της εικόνας. Γενικά η επιλογή των σηματοδευτών εξαρτάται από την εφαρμογή και μια ποικιλία αλγορίθμων έχει αναπτυχθεί για την εξαγωγή τους. Έτσι οι markers μπορεί να είναι: (i) βασισμένοι στην χρωματική αντίθεση, ανταποκρινόμενοι σε κορυφές ή κοιλάδες συγκεκριμένου ύψους (που προκύπτουν μέσω φίλτρων ανακατασκευής της εικόνας): (ii) κορυφές του τελεστή Ψ_{mat} , εντοπίζοντας περιοχές υψηλής ενέργειας υψής: (iii) συνδυασμός κριτηρίων αντίθεσης και υψής και (iv) απευθείας (manually) τοποθετημένοι με επίβλεψη (supervised segmentation) σε περιοχές ενδιαφέροντος.

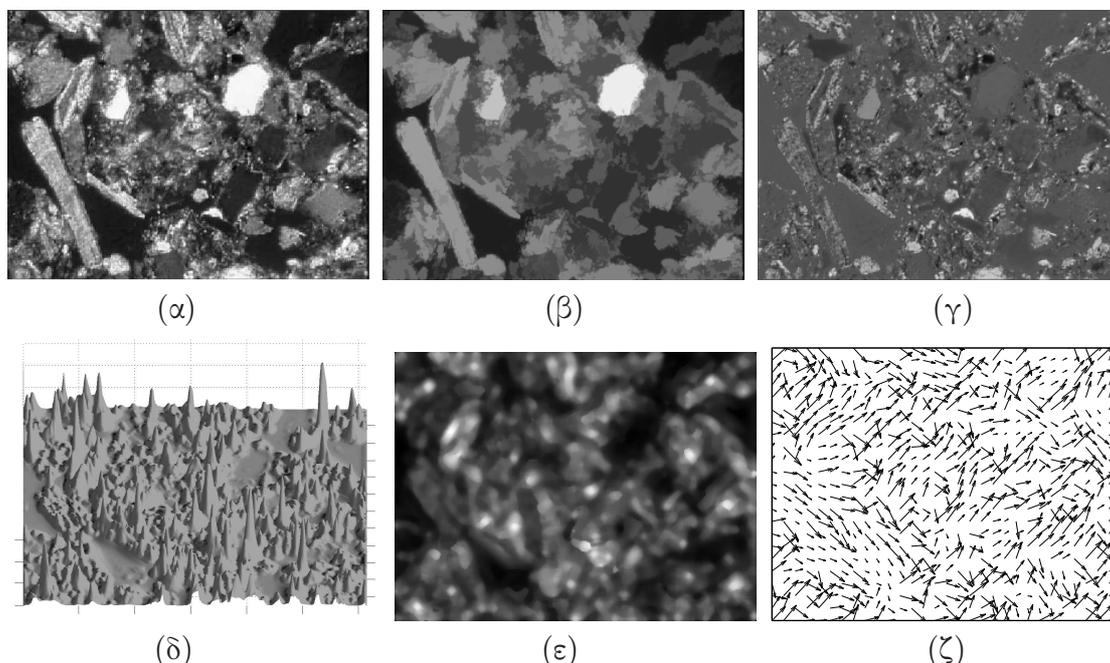
Η υλοποίηση του νόμου διάδοσης των καμπυλών (5.6) γίνεται με καθιερωμένες τεχνικές από τη θεωρία επιπεδοσυνόλων. Αν $\mu = 0$, δηλαδή παραλείποντας τον όρο καμπυλότητας, η ΜΔΕ είναι καθαρού *eikonal* τύπου και η υλοποίηση βασίζεται στη μέθοδο *fast marching* [174], που εξασφαλίζει υπολογιστική ταχύτητα. Αν $\mu \neq 0$, η ΜΔΕ υλοποιείται με χρήση της *narrowband* μεθόδου και τα σύνορα των περιοχών ομαλοποιούνται. Η χρήση του κ -όρου είναι προαιρετική.

5.3 Αποτελέσματα κατάτμησης και συγκρίσεις

Οι μέθοδοι για διαχωρισμό, ανίχνευση και ανάλυση υψής μαζί το κοινό σχήμα κατάτμησης εφαρμόστηκαν σε δύο κατηγορίες εικόνων εδαφοτομών: μεγάλης κλίμακας από σάρωση, ανάλυσης 47 pixel/μm, όπως η λεπτομέρεια του σχήματος 5.1 και μικρής κλίμακας, από μικροσκόπιο πολωμένου φωτός σε ανάλυση 1 pixel/μm. Στο σχήμα 5.2 παρουσιάζεται η διαδικασία διαχωρισμού, ανάλυσης της υψής και εξαγωγής κυρίαρχων χαρακτηριστικών της μέσω πολυκαναλικού εντοπισμού της ενέργειας διαμόρφωσης.

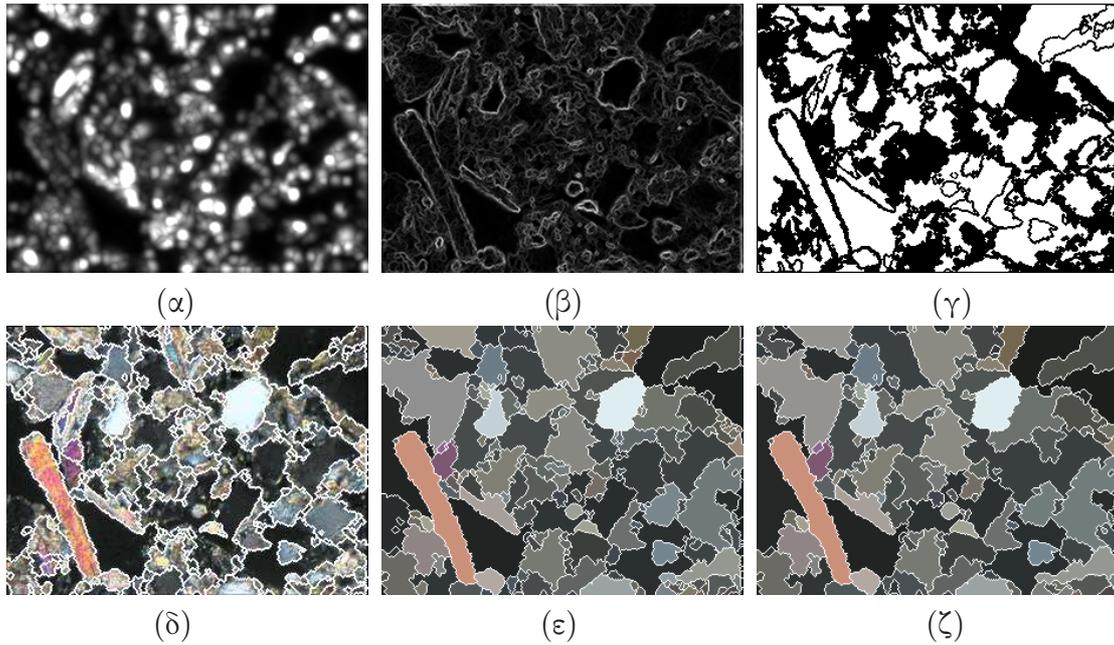
Στο 5.2(α) διακρίνονται οι κυρίαρχες δομές του εδάφους, οργανικό και ανόργανο υλικό (μεγάλα γκρίζα και λευκά αντικείμενα), κενά (σκούρες περιοχές) και συσσωματώματα μικρομάζας. Ο διαχωρισμός σε «καρτούν» συνιστώσα U και συνιστώσα υψής V , στα (β) και (γ) αντίστοιχα, γίνεται με χρήση του δευτέρου επιπέδου της πυραμίδας leveling από Gauss σηματοδευτές ($\sigma_1 = 4$, $\sigma_2 = 8$).

Η καλά εντοπισμένη μέτρηση ενέργειας υψής στο (γ) είναι ενδεικτική των ισχυρών συγκεντρώσεων και μεταβολών της υψής και είναι μια από τις δυνάμεις που οδηγούν τις καμπύλες της κατάτμησης σε περιοχές όπου αυτή η ενέργεια μεγιστοποιείται τοπικά. Η χωρική ανάλυση της μέτρησης είναι άριστη ενεργώντας σε ένα τοπικό παράθυρο 3×3 , ενώ η φασματική εξαρτάται από το εύρος των φίλτρων. Τα χαρακτηριστικά κυρίαρχων διαμορφώσεων της υψής, εξάγονται μέσω EDCA. Το κυρίαρχο πλάτος στο (δ), μετά από median φιλτράρισμα, είναι ομοιόμορφο σε περιοχές ανάλογης έντασης και αντίθεσης. Στο (ε) φαίνεται ο προσανατολισμός του διάνυσματος κυρίαρχων ή τοπικά αναδυόμενων συχνοτήτων, που αναδεικνύει την κατεύθυνση μεταβολής της φάσης των ταλαντώσεων υψής στην εικόνα. Όλες οι μετρήσεις έχουν γίνει το V κομμάτι της εικόνας (γ) μετά τον leveling διαχωρισμό.



Σχήμα 5.2: Ανάλυση και χαρακτηριστικά υψής εδαφοτομών: (α) Εικόνα εδαφοτομής (768×567 pixels) ψηφιοποιημένη σε $1\text{pixel}/\mu\text{m}$ (γκρίζα κλίμακα), (β) «καρτούν» συνιστώσα U , (γ) συνιστώσα υψής V , (δ) Προοπτική 3D απεικόνιση της ενέργειας διαμόρφωσης υψής Ψ_{mat} , (ε) Κυρίαρχο πλάτος, (ζ) Διανύσματα προσανατολισμού κυρίαρχων συχνοτήτων. Τα (δ)-(ζ) έχουν εξαχθεί από τη V συνιστώσα.

Μετά τον μονοχρωματικό $U + V$ διαχωρισμό και την εξαγωγή των χαρακτηριστικών της υψής, στο σχήμα 5.2(α)-(δ), ένα κατάλληλο σύνολο από αυτόματα αποκτημένους σηματοδευτές, που φαίνεται στο σχήμα 5.3(γ), εξελίσσεται με βάση τη ΜΔΕ (5.6), όπου $f_1 = U$ και $f_2 = V$. Τα αρχικά αποτελέσματα κατάτμησης, σχήμα 5.3 (δ)-(ε), λόγω της πολυ-δομικής φύσης αυτών των εικόνων, μπορούν να βελτιωθούν από μια διαδικασία ένωσης περιοχών (region merging) με βάση ένα μετρικό ομοιότητας και τη μεταβλητότητα ενός περιορισμού Fisher απόστασης. Μια βελτιωμένη κατάτμηση φαίνεται στο 5.3(ζ) με καλύτερα ομαδοποιημένες δομές (ειδικά μικρομάζας) και μικρότερο αριθμό τελικών περιοχών.



Σχήμα 5.3: Κατάτμηση της εδαφοτομής του 5.2(α) με συνδυασμό κριτηρίων: (α) Ενέργεια διαμορφώσεων υψής $\Psi_{\text{mat}}(V)$, (β) Μέτρο gradient του U (σχήμα 5.2(β)), (γ) Σημαδευτές, (δ) Συζευγμένη κατάτμηση ($A(t) = 1$), (ε) Περιοχές κατάτμησης, (ζ) Βελτιωμένες περιοχές μετά από ένωση (region post-merging). Από το [178].

5.3.1 Βελτίωση με συγχώνευση περιοχών

Για τη βελτίωση των αποτελεσμάτων κατάτμησης του συζευγμένου, πολυκριτηριακού σχήματος αναπτύχθηκε μια διαδικασία εκ των υστέρων ένωσης περιοχών, η οποία μπορεί να εφαρμοστεί επιλεκτικά και προαιρετικά, ειδικά για την κατηγορία εικόνων εδαφοτομών, για καλύτερη ομαδοποίηση και περισσότερο ερμηνεύσιμα αποτελέσματα. Χρησιμοποιεί διανύσματα χαρακτηριστικών που έχουν εξαχθεί ήδη κατά τη φάση της ανάλυσης υψής και θα μπορούσε να γενικευθεί με στόχο κάποιον αλγόριθμο κατάτμησης εικόνων τύπου split & merge.

Η διαδικασία βασίζεται στον υπολογισμό της απόστασης των μέσων των χαρακτηριστικών ανάμεσα σε γειτονικές περιοχές. Τέτοια χαρακτηριστικά μπορεί να είναι το χρώμα, η ένταση, η αντίθεση των συνιστωσών «καρτούν» και υψής, και τα χαρακτηριστικά κυρίαρχων διαμορφώσεων (ενέργεια, πλάτος, συχνότητες). Το μετρικό απόστασης ανάμεσα σε δύο γειτονικές περιοχές R_1, R_2 , ορίζεται με βάση τη σταθμισμένη Ευκλείδεια απόσταση:

$$D_{R_1 R_2} = \left(\sum_j w_j |\mu_{1j} - \mu_{2j}|^2 \right)^{1/2} \quad (5.7)$$

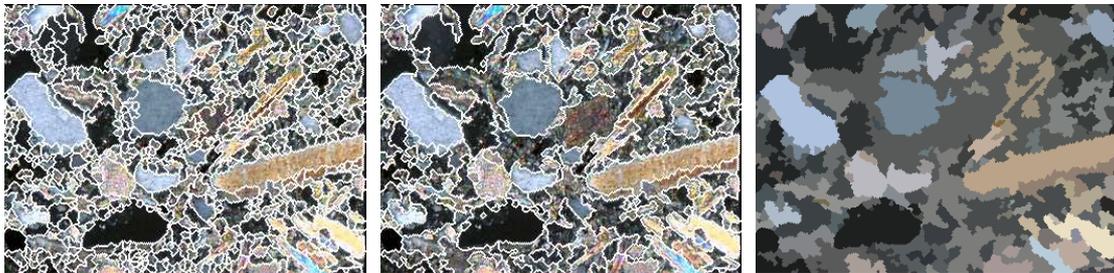
όπου μ_{ij} η μέση τιμή του χαρακτηριστικού j για κάθε περιοχή i . Μια διαφορετική επιλογή θα μπορούσε να είναι η κανονικοποιημένη απόσταση με $w_j = \sigma_{1j}^{-1}$.

Σε κάθε πέρασμα του αλγορίθμου η απόσταση υπολογίζεται για κάθε ζεύγος γειτονικών περιοχών μαζί με τη μέση τιμή μ_{ij} και η μεταβλητότητα σ_{ij}^2 κάθε περιοχής i και για κάθε χαρακτηριστικό j . Αναζητείται το ζεύγος ελάχιστης απόστασης με περιορισμό να ικανοποιείται το τροποποιημένο Fisher κριτήριο

$$F = \text{rank}_k[F_j] = \text{rank}_k \left[\frac{(N_1 + N_2)\sigma_{0j}^2}{N_1\sigma_{1j}^2 + N_2\sigma_{2j}^2} - 1 \right] > c \quad (5.8)$$

όπου το $\text{rank}_k[\cdot]$ συμβολίζει την k -οστή διάταξη ενός συνόλου, N_i είναι ο αριθμός των pixel της περιοχής $i = \{1, 2\}$, F_j είναι ο περιορισμός για τη μεταβλητότητα σ_{0j}^2 του χαρακτηριστικού j στην ενωμένη περιοχή και c ένα κατώφλι-σταθερά (συνήθως τίθεται 0). Αν το κριτήριο (5.8) ικανοποιείται για το ζεύγος ελάχιστης απόστασης, οι περιοχές ενώνονται δίνοντας μια νέα μικρής μέσης απόστασης χαρακτηριστικών και μικρής μεταβλητότητας συγκρινόμενης με τις αρχικές. Σε αντίθετη περίπτωση το ζεύγος της επόμενης μικρότερης απόστασης ελέγχεται και ούτω καθεξής.

Μια σταθερή ελάχιστη επιτρεπόμενη απόσταση ένωσης επιλέγεται κατά την αρχικοποίηση, χρησιμοποιώντας ένα από τα χαμηλότερα κέντρα του ιστογράμματος, για να εγγυάται σύγκλιση της διαδικασίας και να αποτρέπει την ένωση «ανομοιόμορφων» περιοχών. Στο σχήμα (5.4) φαίνεται ένα παράδειγμα της διαδικασίας ένωσης για μια κατάτμηση εδαφοτομής. Τα χαρακτηριστικά που χρησιμοποιήθηκαν είναι το χρώμα, και η ένταση των U και V συνιστωσών ξεχωριστά, με το κατώφλι του κριτηρίου (5.8) $c = 0.5$ και αυτόματη σύγκλιση.



Σχήμα 5.4: Βελτίωση των αποτελεσμάτων κατάτμησης με ένωση περιοχών. Το αρχικό αποτέλεσμα, (αριστερή εικόνα) του συζευγμένου σχήματος, βελτιώνεται με ομαδοποίηση με βάση το χρώμα και τις συνιστώσες γεωμετρίας και υψή(μεσαία εικόνα) Οι τελικές ομαδοποιημένες περιοχές κατάτμησης φαίνονται δεξιά.

5.3.2 Συγκρίσεις

Το προτεινόμενο συζευγμένο σχήμα κατάτμησης δοκιμάστηκε απέναντι σε απλούστερες αλλά καθιερωμένες μεθόδους, με σκοπό να επιδειχθεί η ικανότητα βελτίωσης της κατάτμησης του χώρου της εικόνας. Τέτοιες συγκρίσεις δεν είναι απλές καθώς το πρόβλημα της κατάτμησης εξαρτάται από την εφαρμογή και δεν υπάρχει ένας οικουμενικά βέλτιστος αλγόριθμος που να μπορεί να χρησιμοποιηθεί ως αναφορά σε κάθε περίπτωση. Η έλλειψη ενός ολικού κριτηρίου ποιότητας για την αποτίμηση των διαφορετικών μεθόδων, παράλληλα με την απουσία δεδομένων από ανθρώπινη κατάτμηση εικόνων εδαφοτομών, κάνει δύσκολη την ποσοτική αποτίμηση των αποτελεσμάτων.

Ως μέσο ποσοτικοποίησης των αποτελεσμάτων και συγκρίσεων, χρησιμοποιήθηκαν σ' αυτή την περίπτωση κάποια εμπειρικά κριτήρια ποιότητας, μαζί με απλή παρατήρηση του τελικού χωρισμού των περιοχών. Επιλέχθηκαν τα ακόλουθα μετρικά ελαχιστοποίησης, για τα οποία μικρότερες τιμές υποδηλώνουν καλύτερο αποτέλεσμα κατάτμησης:

- A. Συνάρτηση ολικού κόστους (*Liu Yang Global Cost*) (LYGC) [109], ένας συγχερασμός (tradeoff) ανάμεσα στον περιορισμό της ετερογένειας και στη διατήρηση λεπτομερειών

$$F(I) = \sqrt{N} \times \sum_{i=1}^N \frac{e_i^2}{\sqrt{A_{R_i}}}$$

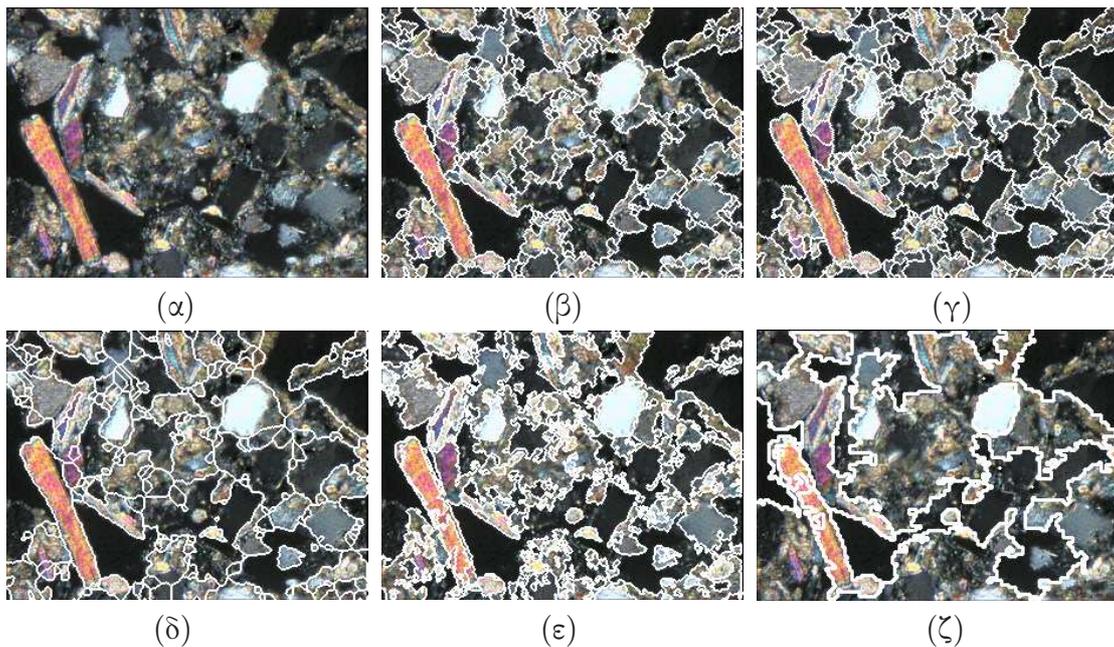
όπου N είναι ο συνολικός αριθμός περιοχών, e_i^2 το σφάλμα χρώματος στην περιοχή i , που ορίζεται ως η Ευκλείδεια απόσταση ανάμεσα στην αρχική και στην καταταμημένη για κάθε pixel ανά περιοχή, και A_{R_i} το εμβαδόν (area) της i (σε pixels).

- B. Συναρτησιακό ενέργειας (*Mumford Shah Energy functional*) (MSF) [140], που πραγματοποιεί κατάτμηση ελαχιστοποιώντας το ολοκληρωτικό μέσο χρωματικό σφάλμα, gradient και μήκος περιγράμματος καμπύλης.

$$E(u, \Gamma) = \mu \int \int_R (u - I)^2 dx dy + \int \int_{R-\Gamma} \|\nabla u\| dx dy + \nu |\Gamma|$$

όπου u η καταταμημένη εικόνα-μωσαϊκό και Γ τα σύνορα κατάτμησης. Το Γ πρέπει να είναι όσο το δυνατόν μικρότερο ενώ τα μ, ν είναι σταθερές.

Το συζευγμένο σχήμα, που θα αναφέρεται ως CTS (contrast-texture) ή VTS (volume-texture) συγκρίθηκε με τις εξής μεθόδους: (α) Μετασχηματισμός *watershed* (WT), καθώς η μέθοδος επεκτείνει την κλασική προσέγγιση περιλαμβάνοντας πληροφορία υψής, (β) Μέθοδος διαχωρισμού και ένωσης *Recursive Shortest Spanning Tree* (RSST) [103], καθιερωμένη region-based τεχνική κατάτμησης, (γ) Ομαδοποίηση *k-means clustering* (KM) [50].



Σχήμα 5.5: Διαφορετικές μέθοδοι κατάτμησης: (α) Εδαφοτομή (έγχρωμη) του 5.2(α), (β) Συζευγμένη κατάτμηση (contrast-based), (γ) Συζευγμένη κατάτμηση (volume-based), (δ) Κατάτμηση watershed, (ε) Κατάτμηση διαχωρισμού και ένωσης (split and merge) βασισμένη στο RSST, (ζ) Ομαδοποίηση κ-μέσων (k-means clustering). Από το [178].

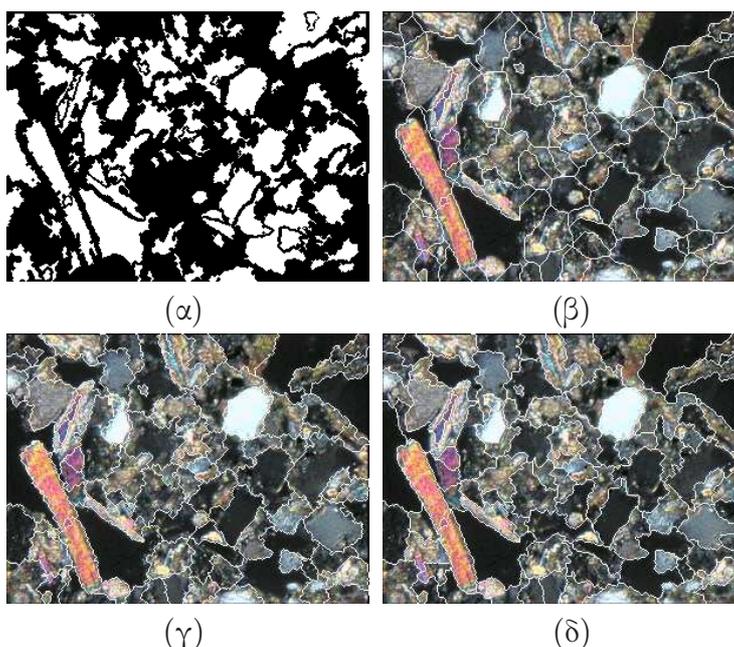
Αποτελέσματα συγκριτικά φαίνονται στο σχήμα 5.5, ενώ τα μετρικά ποιότητας φαίνονται στον πίνακα 5.1. Είναι φανερό ότι το συζευγμένο σχήμα ξεπερνάει σε απόδοση το κλασικό watershed, σχήμα 5.5(δ), γεγονός αναμενόμενο λόγω της διπλής πληροφορίας αντίθεσης φωτεινότητας και υψής. Τα αποτελέσματα του watershed βελτιώνονται ως προς την καλύτερη τοποθέτηση ακμών και τον περιορισμό των συνόρων σε λάθος ακμές (λόγω της παρουσίας υψής). Αν αντί του $A(t) = 1$ στη ΜΔΕ (5.5)

Κριτήριο Ποιότητας	Μέθοδος Κατάτμησης				
	CTS	VTS	WT	KM	RSST
YLGC	2.52	2.24	4.04	4.50	35.57
MSF	1.47	1.45	1.48	2.10	22.31

Πίνακας 5.1: Ποσοτική αποτίμηση μεθόδων κατάτμησης για την εδαφοτομή 5.5(α).

χρησιμοποιηθεί $A(t) \neq 1$ και πλημμυρισμός όγκου [179], η τοποθέτηση των ακμών είναι περαιτέρω βελτιωμένη, σχήμα 5.5(β). Τα αποτελέσματα των RSST και k-means (ακόμη και μετά από επεξεργασία και βελτίωση των αποτελεσμάτων της ομαδοποίησης για τη δημιουργία συνεκτικών περιοχών) στα (ε) και (ζ) αντίστοιχα, φαίνονται υποκατατμημένα (undersegmented) με μέτρια τοποθέτηση ακμών, οι οποίες στην περίπτωση του RSST είναι και ισχυρά κβαντισμένες.

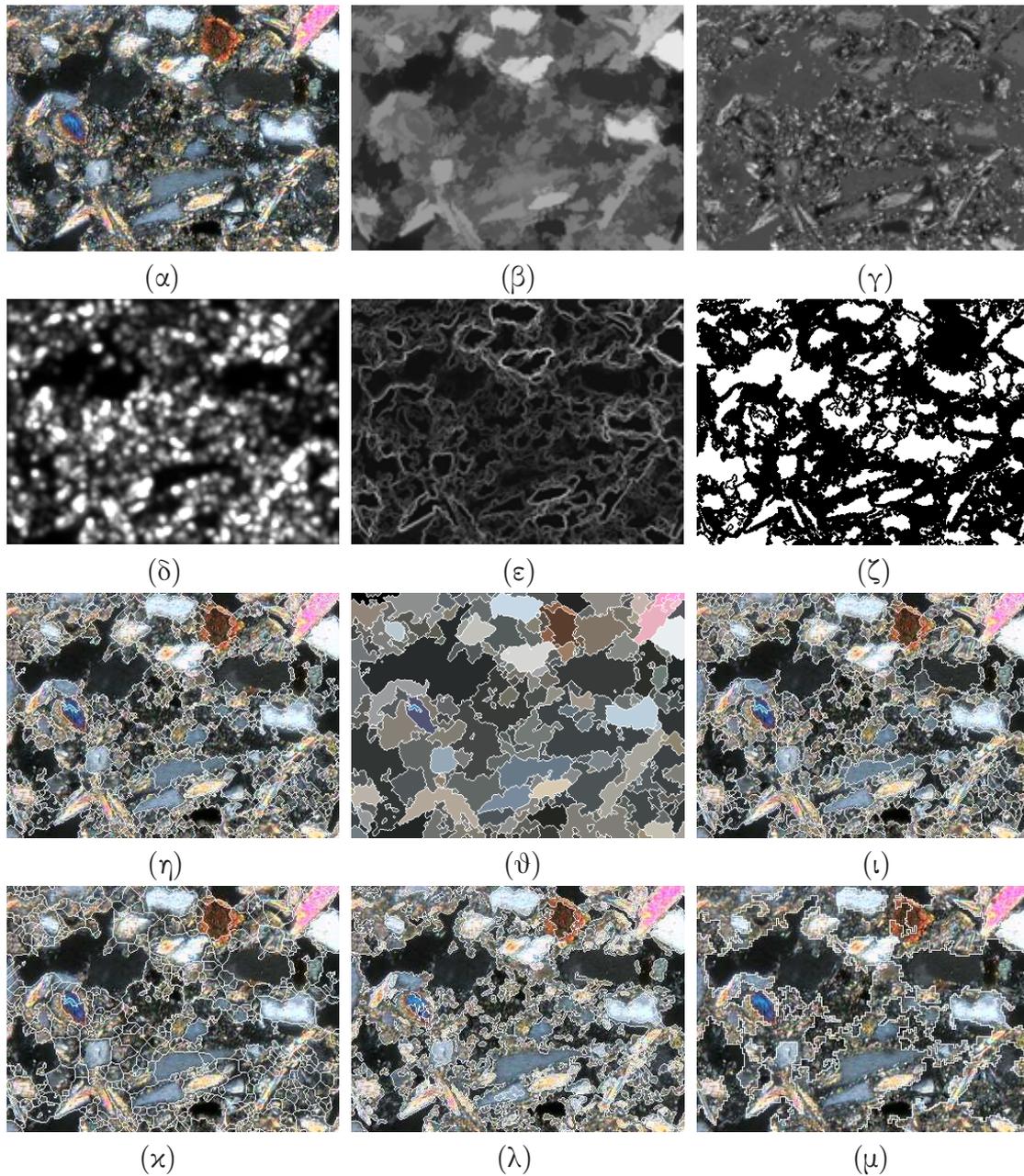
Από τον πίνακα 5.1, τα καλύτερα αποτελέσματα με ποσοτικούς όρους επιτεύχθηκαν από το volume-based συζευγμένο σχήμα πλημμυρισμού, με δεύτερο καλύτερο το εναλλακτικό contrast-based. Ένα δεύτερο σύνολο συγκρίσεων για την ίδια εικόνα, ανάμεσα στα δύο σχήματα και το απλό watershed παρουσιάζεται στο 5.6 με ένα διαφορετικό αρχικό σύνολο σημαδευτών η επιλογή του οποίου φαίνεται στο (α). Τέλος η όλη διαδικασία παρουσιάζεται ξανά για μια διαφορετική εικόνα εδαφοτομής, στο σχήμα 5.7, με όλα τα ενδιάμεσα στάδια και με το σύνολο των σχετικών συγκρίσεων που ποσοτικοποιούνται στον πίνακα 5.2.



Σχήμα 5.6: Κατάτμηση όπως στο σχήμα 5.3 με συνδυασμό κριτηρίων αλλά διαφορετικό σετ σημαδευτών: (α) Σημαδευτές, (β) Κατάτμηση watershed, (γ) Συζευγμένη κατάτμηση (contrast-based watershed), (δ) Συζευγμένη κατάτμηση (volume-based watershed μεγέθους).

5.4 Κατηγοριοποίηση υφής

Η κατηγοριοποίηση υφής (texture classification) βασίζεται στην κατασκευή διανυσμάτων χαρακτηριστικών, που αντιστοιχούν σε κάθε θέση των διαφορετικών περιοχών



Σχήμα 5.7: Κατάτμηση και συγκρίσεις: (α) Εδαφοτομή (έγχρωμη), (β) Συνιστώσα «καρτούν» U , (γ) Συνιστώσα υψής V , (δ) Ενέργεια διαμορφώσεων υψής $\Psi_{\text{mat}}(V)$, (ε) Μέτρο gradient του U , (ζ) Σημαδευτές, (η) Συζευγμένη κατάτμηση (contrast-based), (θ) Περιοχές κατάτμησης βελτιωμένες με συγχωνεύσεις, (ι) Συζευγμένη κατάτμηση (volume-based), (κ) Κατάτμηση watershed, (λ) Κατάτμηση διαχωρισμού και ένωσης (split and merge) βασισμένη στο RSST, (μ) Ομαδοποίηση κ-μέσων (k-means clustering). Από το [177].

Κριτήριο Ποιότητας	Μέθοδος Κατάτμησης			
	CTS	VTS	WT	KM
YLGC	0.25	0.09	0.6	7.7
MSF	3.25	3.30	3.40	3.69

Πίνακας 5.2: Ποσοτική αποτίμηση μεθόδων κατάτμησης για την εδαφοτομή 5.7(α).

της εικόνας και στην επιλογή ενός κατάλληλου κατηγοριοποιητή (classifier), όπως k-means, Bayes κ.α. [50]. Για να εκτιμηθεί η αποτελεσματικότητα των εξαγόμενων χαρακτηριστικών για ανάλυση υφής, με όρους κατηγοριοποίησης των επιμέρους περιούχων, ανεξάρτητα από την επιλογή του classifier, εξετάζεται η διαχωριστικότητα των τάξεων σε ένα σύνολο από τυπικές εικόνες μικροϋφών εδάφους. Η μέθοδος βασίζεται σε ένα στατιστικό μετρικό της ικανότητας τελεστών ή διαδικασιών να διακρίνουν δύο τάξεις, ποσοτικοποιώντας τη διαχωριστικότητα των σημείων στο χώρο χαρακτηριστικών, σύμφωνα με το κριτήριο Fisher [65, 100]. Το κριτήριο Fisher είναι ένα μέτρο της απόστασης ανάμεσα στις διαφορετικές τάξεις που κατασκευάζει ένας classifier [50] και του βαθμού στον οποίο αυτές είναι συμπαγείς.

Οποιαδήποτε δύο πολυδιάστατα διανύσματα χαρακτηριστικών, που αντιστοιχούν σε δύο διαφορετικές τάξεις, προβάλλονται γραμμικά στον 1Δ χώρο για να επιτευχθεί μέγιστη διαχωριστικότητα, σύμφωνα με το Fisher Linear Discriminant

$$y = (\vec{\mu}_1 - \vec{\mu}_2)^T S^{-1} \vec{x}$$

όπου S ο πίνακας συμμεταβλητότητας (pooled covariance matrix), μ οι μέσοι των δύο τάξεων και \vec{x} ένα διάνυσμα χαρακτηριστικών. Η προβολή μεγιστοποιεί το κριτήριο Fisher γραμμικής διάκρισης

$$f = \frac{|n_1 - n_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (5.9)$$

όπου σ_i η τυπική απόκλιση των προβολών των διασπορών και n_i η προβολή του μέσου $\vec{\mu}_i$, για τις δύο τάξεις $i = \{1, 2\}$. Το f εκφράζει σε μια ποσότητα την απόσταση ανάμεσα σε δύο ομαδοποιήσεις σχετικά με το μέγεθος τους. Μεγαλύτερες τιμές του κριτηρίου, εκφράζουν καλύτερη διαχωριστικότητα για τις υπό εξέταση τάξεις.

Για σκοπούς σύγκρισης χρησιμοποιούνται τρία διανύσματα χαρακτηριστικών και η αποτελεσματικότητά τους ποσοτικοποιείται από την τιμή του f . Τα χαρακτηριστικά αυτά βασίζονται στην πολυκαναλική Gabor ανάλυση και στην περιγραφή της υφής στον χώρο των αποκρίσεων των φίλτρων. Σε κάθε pixel αποδίδεται ένα σύνολο τιμών, αποτέλεσμα ενός τελεστή ή μιας διαδικασίας στην έξοδο της συστοιχίας.

- *Ενέργεια* (Gabor energy) [20, 65, 163, 175, 192], δηλ. το μέτρο της απόκρισης των quadrature φίλτρων για κάθε ζώνη συχνοτήτων

$$\vec{F}_G = [(h_{ck} + jh_{sk}) * I]_{k=1 \dots K}$$

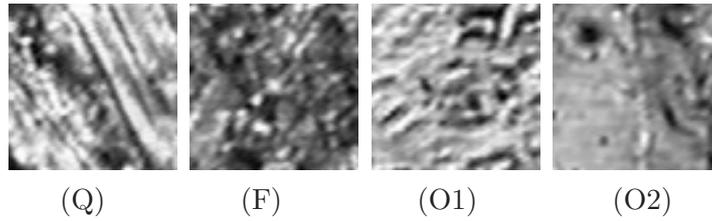
- *Teager ενέργεια* των αποκρίσεων των Gabor φίλτρων (Gabor Teager energy).

$$\vec{F}_T = [\Psi(h_{ck} * I)]_{k=1 \dots K}$$

- *Ενέργεια διαμόρφωσης* Ψ_{mat} , κυρίαρχος προσανατολισμός συχνότητας $\angle \vec{\omega}_D$ και ένταση φωτεινότητας I (Ενότητα 3.4).

$$\vec{F}_D = [\Psi_{\text{mat}}(I), \angle \vec{\omega}_D, I]$$

Τα δύο πρώτα διανύσματα, \vec{F}_G , \vec{F}_T είναι διάστασης, ανά pixel ίσης με τον αριθμό των φίλτρων, τυπικά $K = 40$, ενώ το διάνυσμα «κυρίαρχων» χαρακτηριστικών, μειώνει την διάσταση κατά $3/K$ φορές. Η σύγκριση \vec{F}_G , \vec{F}_T εξυπηρετεί τη σύγκριση του διδιάστατου τελεστή Ψ με τον κλασικό τελεστή για διάκριση ενέργειας. Το \vec{F}_D



Σχήμα 5.8: Τυπικές μικροϋφές χώματος: Κρύσταλλοι (quartz grains - crystals) (Q), Μικρομάζα (fine matter) (F) και οργανικό υλικό (organic matter) (O1, O2).

αποτελεί το εναλλακτικό διάλυμα χαμηλής διάστασης, που έχει εξαχθεί μέσω, εντοπισμού ενέργειας διαμόρφωσης, ανάλυσης κυρίαρχων συνιστωσών και περιλαμβάνει και πληροφορία έντασης φωτεινότητας.

Οι τέσσερις τυπικές τάξεις υφής που φαίνονται στο σχήμα 5.8, έχουν ληφθεί από τις λεπτομέρειες των κατατμημένων εδαφοτομών, και αντιπροσωπεύουν κρύσταλλους/ανόργανο υλικό (Q), μικρομάζα ή συσσωματώματα αδιάκριτου υλικού (F), και οργανικό υλικό (O). Οι υφές επιλέχθηκαν ως αντιπροσωπευτικές των αντίστοιχων τάξεων τυπικών δομών του εδάφους. Για όλα τα ζεύγη, υπολογίσθηκαν οι αποστάσεις Fisher που επιτυγχάνονται από τα τρία διανύσματα, f_g , f_t , f_d . Στον πίνακα 5.3 παρουσιάζονται οι τιμές των αποστάσεων f .

	Ενέργεια Gabor \vec{F}_G				Ενέργεια Teager, \vec{F}_T				Κυρίαρχα, \vec{F}_D			
	Q	F	O1	O2	Q	F	O1	O2	Q	F	O1	O2
Q	-	1.03	1.41	1.16	-	1.10	1.56	1.21	-	1.38	1.03	1.06
F		-	1.08	1.33		-	1.17	1.45		-	1.65	1.01
O1			-	1.37			-	1.48			-	1.52
O2				-				-				-

Πίνακας 5.3: Αποστάσεις Fisher για χαρακτηριστικά υφής μετά από Gabor πολυκαναλική ανάλυση, για τα δείγματα τυπικής υφής στο σχήμα 5.8.

Οι μέγιστες και ελάχιστες αποστάσεις με τα αντίστοιχα ζεύγη τάξεων ανά μέθοδο είναι: $\mathbf{f}_g=[1.41 \text{ (Q-O1)}, 1.03 \text{ (Q-F)}]$, $\mathbf{f}_t=[1.56 \text{ (Q-O1)}, 1.10 \text{ (Q-F)}]$, $\mathbf{f}_d=[1.65 \text{ (F-O1)}, 1.01 \text{ (F-O2)}]$. Το διάλυμα Teager ενεργειών των Gabor αποκρίσεων πετυχαίνει μεγαλύτερη διαχωριστικότητα για κάθε ζεύγος με μέση απόσταση 1.33, συγκρινόμενο με την τυπική Gabor ενέργεια μέσου 1.23. Το χαμηλής διάστασης διάλυμα κυρίαρχων χαρακτηριστικών, που φέρει πληροφορία πλάτους και συχνότητας διαμόρφωσης, δίνει τη μέγιστη απόσταση, (F-O1) και μια μέση 1.27 για όλα τα ζεύγη.

Συμπερασματικά, ο διδιάστατος τελεστής Teager, ξεπερνάει σε απόδοση τη συμβατική Gabor ενέργεια, σε επίπεδο διαχωριστικότητας, για κατηγοριοποίηση και μπορεί να χρησιμοποιηθεί εναλλακτικά σαν ενεργειακός τελεστής ή μη-γραμμικότητα μετά από οποιοδήποτε πολυκαναλικό φιλτράρισμα. Η ενέργεια διαμορφώσεων υφής, ισορροπεί το trade-off ανάμεσα σε διάσταση και αποτελεσματικότητα, πράγμα που αναδεικνύει ότι τα αναδυόμενα κυρίαρχα χαρακτηριστικά διαμόρφωσης μπορούν να χρησιμοποιηθούν για επιπρόσθετη ανάλυση των κατατμημένων εικόνων, κατηγοριοποίηση υφής ή εκτίμηση κατανομών σχετικά με τον τύπο ή την υφή των διαφορετικών περιοχών.

5.5 Συμπεράσματα και αποτιμήσεις

Το πλαίσιο ανάλυσης εδαφοτομών, που έχει αναπτυχθεί και στηρίζεται στην ανίχνευση και στην ανάλυση προτύπων υψής, περιλαμβάνει διαχωρισμό κάθε εικόνας σε δύο συνιστώσες, ανάλυση υψής και συζευγμένη κατάτμηση με πληροφορία υψής και γεωμετρίας. Το κοινό σχήμα κατάτμησης δίνει περισσότερο ερμηνεύσιμα και λογικά αποτελέσματα σε σύγκριση με καθιερωμένες, συμβατικές μεθόδους. Η αποτίμηση πραγματοποιήθηκε τόσο μέσω οπτικής επισκόπησης όσο και με μετρικά ποιότητας κατάτμησης για εικόνες εδαφοτομών πολύπλοκης δομής. Η σημασία της ανάλυσης υψής αναδεικνύεται από την απόδοση και τα αποτελέσματα της κατάτμησης μέσα από το σενάριο διαχωρισμού της από τη «γεωμετρία», δηλ. την αντίθεση ή το μέγεθος των μακροδομών, της εικόνας. Η ενέργεια διαμόρφωσης υψής βελτιώνει επίσης τη διαχωριστικότητα των τάξεων για κατηγοριοποίηση, γεγονός που μπορεί να χρησιμοποιηθεί για πρόσθετη ανάλυση των περιοχών κατάτμησης με προϋπολογισμένα χαρακτηριστικά. Ο συνδυασμός γεωμετρικών, στατιστικών και μετρήσεων υψής στις διαφορετικές περιοχές θα αποδώσει τις χαρακτηριστικές ιδιότητες του εδάφους για την εκτίμηση της βιολογικής του ποιότητας. Ανάλογες τεχνικές μπορούν να εφαρμοστούν και σε άλλες πτυχές της επιφανειακής ή υποεδαφικής ανάλυσης, όπως επίσης και σε δεδομένα τηλεπισκόπησης.

□

Κεφάλαιο 6

Ανίχνευση Μικροδομών Φωνής

Ο λόγος αποτελεί ίσως την πιο πρώιμη μορφή ανθρώπινης επικοινωνίας μαζί με την εικόνα και τη μουσική. Από τους πρώτους κωδικοποιητές φωνής και την εισαγωγή αναλογικών κυματομορφών σε μηχανές μέσω διάτρητων καρτών του Flanagan, προσεγγίζεται η εποχή της πολυμορφικής, αμφίδρομης επικοινωνίας ανθρώπου και μηχανών. Η ανάγκη για τεχνικές και γενικευμένα μοντέλα για την επεξεργασία οπτικής και ηχητικής πληροφορίας προκύπτει από την ύπαρξη κοινών εφαρμογών αλλά και από την αναζήτηση ερμηνείας σε παραπλήσια φαινόμενα. Οι έρευνες του Teager [186] γύρω από τα μη-γραμμικά φαινόμενα κατά την παραγωγή της φωνής και η ακόλουθη μη-γραμμική μοντελοποίηση της, ώθησαν την αναζήτηση μικροδομών διαμόρφωσης σε σήματα φωνής για θέματα επεξεργασίας και αναγνώρισης ομιλίας.

Παρουσιάζεται εδώ η έρευνα σχετικά με την ανίχνευση τέτοιων μικροδομών με σκοπό την αναγνώριση και την ανίχνευση φωνητικών γεγονότων (speech event detection). Εξειδικευμένα και αποτελεσματικά χαρακτηριστικά για ανίχνευση και κατάτμηση φωνής είναι απαραίτητα σε σύγχρονα συστήματα επικοινωνίας λόγου. Χρησιμοποιώντας ιδέες από τη μη-γραμμική ανάλυση της φωνής, όπως η παρουσία διαμορφώσεων και η πολυκαναλική ανάλυση, προτείνονται νέες περιγραφές ενέργειας και φασματικού περιεχομένου. Τα νέα χαρακτηριστικά λαμβάνονται μέσω διαχωρισμού σε πολλαπλές ζώνες συχνοτήτων και εντοπισμού της κυρίαρχης ενέργειας διαμόρφωσης, σε επίπεδο της Teager-Kaiser ενέργειας ξεχωριστών AM-FM συνιστωσών. Η αποτελεσματικότητά τους επιδεικνύεται μέσα από θεωρητική, στατιστική ανάλυση αλλά και συστηματική αποτίμηση σε δύο σενάρια εντοπισμού φωνής, και συγκεκριμένα ανίχνευση συνόρων φωνής (endpoint detection) και ανίχνευση φωνητικής δραστηριότητας (voice activity detection) [56, 57].

6.1 Φωνή σε θόρυβο

Η ανίχνευση ύπαρξης φωνής μπορεί να αντιμετωπισθεί είτε άμεσα σαν αναγνώριση και ονοματισμό γεγονότων (event labeling) σε περιβάλλον σιωπής ή(και) θορύβου, είτε έμμεσα ως αναγνώριση φωνητικής δραστηριότητας (voice activity detection). Ο διαχωρισμός φωνής από περιβάλλοντα θόρυβο είναι ειδική περίπτωση του γενικότερου προβλήματος της κατάτμησης φωνής και είναι σημαντικός για την αναγνώριση, κωδικοποίηση, επεξεργασία και μετάδοση φωνής. Σημαντική μείωση επεξεργασίας επιτυγχάνεται επιλέγοντας μόνο τα χρήσιμα τμήματα ενός ηχογραφημένου σήματος, ενώ τα συστήματα αναγνώρισης ομιλίας απαιτούν υψηλή ακρίβεια κατά τη φάση της δημιουργίας των φωνητικών προτύπων. Η ανίχνευση εμπλέκεται στο χαρακτηρισμό, διαχωρισμό

και προσπέλαση μεγάλων βάσεων και σε τεχνικές ενίσχυσης όπως η εκτίμηση φάσματος θορύβου (noise spectrum estimation) [132], απόρριψη «παραθύρων» (frame dropping) [161], μείωση θορύβου, ακύρωση ηχούς (echo cancelation), κανονικοποίηση ενέργειας και συμπίεση σιωπής. Στις τηλεπικοινωνίες εφαρμόζεται για επιλεκτική, πραγματικού χρόνου, μετάδοση φωνής πάνω από δίκτυα.

Οι σύγχρονες προσεγγίσεις επικεντρώνονται είτε στην ανάπτυξη εξελιγμένων χαρακτηριστικών είτε στην πολυπλοκότητα της λογικής της απόφασης. Πρωτότυπα χαρακτηριστικά για labeling φωνής εμπνέονται από εναλλακτικές ιδιότητες του σήματος. Εκτός από την ενέργεια και των ρυθμό μεταβάσεων από το μηδέν (zero-crossings rate), η βιβλιογραφία περιλαμβάνει “περιοδικότητα” και jitter, ευστάθεια του pitch, χωρική συσχέτιση, φασματική εντροπία, cepstral χαρακτηριστικά, υπόλοιπο LPC, εναλλακτικές ενεργειακές μετρήσεις [210], χρονική περιβάλλουσα ισχύος [132], φασματική απόκλιση [161, 162] και χρονο-συχνοτικές παραμέτρους μέσω πολυκαναλικής ανάλυσης [207]. Πρόσφατα το στατιστικό πλαίσιο κερδίζει ενδιαφέρον με ιδιότητες των στατιστικών της φωνής να χρησιμοποιούνται καθώς και βέλτιστους κανόνες λόγω πιθανοφάνειας [180].

6.2 Ενεργειακοί τελεστές και πολυζωνικές διαμορφώσεις

Πειραματικές και θεωρητικές ενδείξεις για την ύπαρξη διαμορφώσεων σε πολλαπλές κλίμακες κατά τη διάρκεια παραγωγής, οδήγησαν στην πρόταση του AM-FM μοντέλου διαμορφώσεων για τη φωνή [125]. Αποδιαμόρφωση ενός πραγματικού AM-FM σήματος

$$x(t) = a(t) \cos \left(\int_0^t \omega(\tau) d\tau \right) \quad (6.1)$$

με χρονικά μεταβαλλόμενη περιβάλλουσα πλάτους $a(t)$ και σήματα στιγμιαίων συχνοτήτων $\omega(t)$, μπορεί να προσεγγιστεί με χρήση του μη-γραμμικού διαφορικού ενεργειακού τελεστή Teager-Kaiser [91, 186]. Για σήματα συνεχούς χρόνου αυτός ο τελεστής είναι

$$\Psi[x(t)] \equiv [\dot{x}(t)]^2 - x(t)\ddot{x}(t) \quad (6.2)$$

όπου $\dot{x}(t) = dx(t)/dt$. Ο Teager-Kaiser ενεργειακός τελεστής Ψ εντοπίζει τη στιγμιαία ενέργεια πηγής που παράγει μια ταλάντωση. Εφαρμοζόμενος σε ένα AM-FM σήμα της μορφής (6.1), ο Ψ δίνει τη στιγμιαία ενέργεια της πηγής δηλ. $\Psi[x(t)] \approx a^2(t)\omega^2(t)$, όπου το σφάλμα προσέγγισης είναι αμελητέο [125], αν το στιγμιαίο πλάτος $a(t)$ και η στιγμιαία συχνότητα $\omega(t)$ δε μεταβάλλονται πολύ ή γρήγορα σε σχέση με τη μέση τιμή του $\omega(t)$.

Ένα σχήμα AM-FM αποδιαμόρφωσης [125] (*Energy separation algorithm* ή ESA), βασίζεται στον Ψ και στον διαχωρισμό της στιγμιαίας ενέργειας στα σήματα πλάτους και συχνότητας

$$\sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}} \approx \omega(t) \quad , \quad \frac{\Psi[x(t)]}{\sqrt{\Psi[\dot{x}(t)]}} \approx |a(t)| \quad (6.3)$$

Η μεθοδολογία στιγμιαίου διαχωρισμού της ενέργειας οδήγησε σε διάφορες κατηγορίες αλγορίθμων για την αποδιαμόρφωση διακριτών AM-FM σημάτων

$$x[n] = x(nT) = A[n] \cos \left(\int_0^n \Omega[k] dk \right) \quad (6.4)$$

όπου οι ακέραιοι k και n χρησιμοποιούνται συμβολικά από την ολοκλήρωση ως συνεχείς μεταβλητές, $A[n] = a(nT)$ και $\Omega[n] = T\omega(nT)$. Μια άμεση προσέγγιση είναι η εφαρμογή του διακριτού Teager-Kaiser τελεστή $\Psi_d[x_n] \equiv x_n^2 - x_{n-1}x_{n+1}$, όπου $x_n = x[n]$, στο διακριτό (6.4) και η απόκτηση διακριτών ενεργειακών εξισώσεων της μορφής $\Psi_d[x_n] \approx A^2[n] \sin^2(\Omega[n])$. Αυτή είναι η βάση του διακριτού αλγορίθμου διαχωρισμού ενέργειας (*Discrete ESA*) [125]:

$$\arccos \left(1 - \frac{\Psi_d[x_n - x_{n-1}] + \Psi_d[x_{n+1} - x_n]}{4\Psi_d[x_n]} \right) \approx \Omega[n] \quad (6.5)$$

$$\sqrt{\frac{\Psi_d[x_n]}{\sin^2(\Omega[n])}} \approx |A[n]| \quad (6.6)$$

Ο DESA είναι απλός, υπολογιστικά απλός και έχει άριστη, σχεδόν στιγμιαία χρονική ανάλυση.

Για να εφαρμοστεί αποδιαμόρφωση μέσω του ESA σε ένα AM-FM μοντελοποιημένο σήμα, χρειάζονται διαδικασίες φιλτραρίσματος και απομόνωσης συγκεκριμένων ζωνών συχνοτήτων. Το σχήμα πολυκαναλικής ανάλυσης αποδιαμόρφωσης (*multiband demodulation analysis* ή MDA) προτάθηκε στο [22] ως τρόπος εξαγωγής διαμορφώσεων παρουσία θορύβου. Το MDA εφαρμόζεται στην έξοδο μιας συστοιχίας γραμμικά τοποθετημένων Gabor φίλτρων [60], των οποίων η 1Δ απόκριση $h(t)$ και απόκριση συχνότητας $H(\omega)$ είναι

$$h(t) = \exp(-\alpha^2 t^2) \cos(\omega_c t) \quad (6.7)$$

$$H(\omega) = \frac{\sqrt{\pi}}{\alpha} \left[\exp\left(-\frac{(\omega - \omega_c)^2}{4\alpha^2}\right) + \exp\left(-\frac{(\omega + \omega_c)^2}{4\alpha^2}\right) \right], \quad (6.8)$$

όπου ω_c η κεντρική συχνότητα του φίλτρου και α το rms εύρος του.

6.3 Κίνητρα και ανίχνευση μέσω στατιστικών υποθέσεων

Αποτελεσματική ανίχνευση δραστηριότητας φωνής προϋποθέτει την εξέταση τόσο του ενεργειακού επιπέδου της διέγερσης όσο και του συχνοτικού του περιεχομένου. Ο ορισμός της ενέργειας ενός σήματος κατά τον Teager, ως η ενέργεια που παράγεται από την γενετήσια πηγή του, ταιριάζει σε ένα τέτοιο πλαίσιο, με την ταυτόχρονη, στιγμιαία μέτρηση του μεγέθους δύο ποσοτήτων που εκφράζουν τη διαφορετική πληροφορία.

Θεωρώντας το άθροισμα διαμορφωμένων ημιτόνων ως το μοντέλο φωνής [125], η περιγραφή διακριτού χρόνου είναι

$$s[n] = \sum_{k=1}^K A_k[n] \cos(\Omega_{ck} \cdot n + \Phi_k[n]) \quad (6.9)$$

όπου k ο δείκτης συντονισμού (resonance) και K ο αριθμός των θεμελιωδών συντονισμών (formants). Τα στιγμιαία μεταβαλλόμενα σήματα πλάτους A_k και φάσης Φ_k (ή συχνότητας $\Omega_k = \Omega_{ck} + \partial\Phi_k/\partial n$) πρόκειται να εκτιμηθούν και να ανιχνευθούν.

Έστω ότι ένα μοναδικό AM-FM σήμα είναι παρόν, δηλ. $K = 1$, με την απομόνωση μιας ζώνης διαμόρφωσης μέσω ενός αρκετά στενού Gabor φίλτρου. Η φέρουσα Ω_{ck}

μπορεί να θεωρηθεί γνωστή και να προσεγγιστεί μέσω της κεντρικής συχνότητας του φίλτρου. Διατυπώνοντας το πρόβλημα για απλότητα ως την ανίχνευση ενός ημιτονοειδούς με άγνωστες, μη-τυχαίες παραμέτρους και σχεδόν στατικό πλάτος και φάση, ανά παράθυρο ανάλυσης, σε λευκό Gauss θόρυβο άγνωστης μεταβλητότητας, οι δύο στατιστικές υποθέσεις είναι [95]:

$$\begin{aligned} H_0 : X[n] &= W[n] \\ H_1 : X[n] &= W[n] + A \cos(\Omega_c n + \Phi) + B \end{aligned} \quad (6.10)$$

για κάθε παράθυρο μήκους N . Χρησιμοποιώντας εκτιμήσεις μέγιστης πιθανοφάνειας $[\hat{A}, \hat{\Phi}, \hat{B}, \hat{\sigma}_1^2, \hat{\sigma}_0^2]$ για τις άγνωστες παραμέτρους αποδεικνύεται εύκολα [57] ότι η συνάρτηση λογαριθμικής πιθανοφάνειας για τη δεσμευμένη στο H_1 κατανομή Gauss $p(X|H_1)$ είναι:

$$\ln p(X|H_1) \approx N \frac{\hat{A}^2}{4\hat{\sigma}_1^2} + \frac{N}{2\hat{\sigma}_1^2} (\hat{B}^2 - \hat{\sigma}_0^2) - \frac{N}{2} \ln 2\pi\hat{\sigma}_1^2 \quad (6.11)$$

Με στατιστική δοκιμή, αποφασίζεται υπέρ της υπόθεσης H_i που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας. Για να ληφθεί υπόψιν ο διαφορετικός αριθμός εκτιμώμενων παραμέτρων για τις δύο υποθέσεις, χρησιμοποιείται το κριτήριο *Minimum Description Length* ή MDL. Συγκεκριμένα επιλέγεται η H_i που μεγιστοποιεί το $\text{MDL}(i) = \ln p(X|H_i) - \frac{n_i}{2} \ln N$, όπου $n_i = [1, 4]$ η διάσταση του διανύσματος παραμέτρων για $i = 1, 2$.

Η εκτίμηση του Ω_c χρησιμοποιώντας ένα Gaussian παράθυρο εύρους συχνότητας σ_g χαρακτηρίζεται από κάποιες σχέσεις αβεβαιότητας των μέσων διάρκειας και συχνότητας [39], που καταλήγουν στο $\Omega_c^2 + \sigma_g^2 \geq 1/4N^2$ [57]. Χρησιμοποιώντας το κατώτερο όριο αβεβαιότητας με $n_1 = 4$ για το MDL(1), μπορεί να κατασκευαστεί ένας κανόνας για την ανίχνευση ημιτονοειδών συνιστωσών φωνής

$$N \frac{\hat{A}^2}{4\hat{\sigma}_1^2} + \ln(\Omega_c^2 + \sigma_g^2) \underset{H_0}{\overset{H_1}{\gtrless}} \mathcal{O}(\hat{B}, \hat{\sigma}_1^2, \hat{\sigma}_0^2, N) \quad (6.12)$$

όπου \mathcal{O} μια συνάρτηση στατιστικών στο παράθυρο ανάλυσης, το σ_g είναι ανάλογο του εύρους του φίλτρου και $\hat{A}^2/2\hat{\sigma}_1^2$ ο σηματοθορυβικός λόγος (SNR). Παρόμοια λογική εφαρμόζεται επίσης και για την ανίχνευση ενός από K ημιτονοειδή με διαφορετικές φέρουσες, που αντιστοιχούν στα K Gabor φίλτρα. Σε αυτή την περίπτωση δοκιμάζονται $K + 1$ υποθέσεις μεγιστοποιώντας το MDL κριτήριο ενώ ένα παράθυρο ανάλυσης χαρακτηρίζεται ως θόρυβος αν $\text{MDL}(0) > \text{MDL}(i), \forall i \neq 0$.

Από το [22] η αναμενόμενη τιμή του ενεργειακού τελεστή σε ένα φιλτραρισμένο AM-FM σήμα σε θόρυβο προσεγγίζεται από τη σχέση $\Psi(X[n]) \approx A[n]^2 |H(\Omega_c)|^2 [(\Omega_c + \partial\Phi[n]/\partial n)^2 + \Gamma_c]$, όπου Ω_c και H είναι αντίστοιχα η κεντρική και η απόκριση συχνότητας του φίλτρου και Γ_c μια σταθερά ίση με τη μέση φιλτραρισμένη ισχύ θορύβου. Εδώ η προσέγγιση δίνει $\Psi(X[n]) \approx A^2(\Omega_c^2 + \Gamma_c) \cdot |H(\Omega_c)|^2$ και λογαριθμίζοντας

$$\ln \Psi(X[n]) \approx \ln A^2 + \ln(\Omega_c^2 + \Gamma_c) + \text{const.} \quad (6.13)$$

Συγκρίνοντας τις εξισώσεις (6.12) και (6.13), παρατηρείται η αναλογία στους όρους πλάτους και συχνότητας και στις σταθερές που βασίζονται στο μέσο ζωνοπερατό θόρυβο μέσα στους λογαρίθμους. Αυτές οι ομοιότητες, παρά τις προσεγγίσεις που τέθηκαν στο πρόβλημα, δίνουν μια αίσθηση του ρόλου του ενεργειακού τελεστή και των εκτιμήσεων του ESA για την επιλογή του καναλιού και τη διαδικασία ανίχνευσης φωνής.

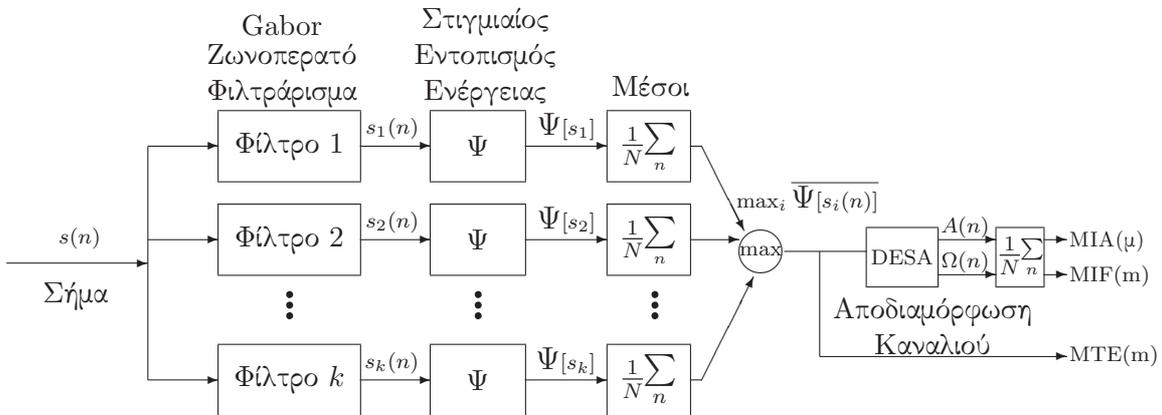
Έτσι τα κίνητρα για χαρακτηριστικά ανίχνευσης φωνής, βασισμένα σε διαμόρφώσεις, είναι: (α) η στατιστική ερμηνεία ενός ανιχνευτή κατωφλίων (6.12), που συγκρίνει ταυτόχρονα μέσες εκτιμήσεις πλάτους και συχνότητας διαταραχών, (β) η δυϊκή πληροφορία που λαμβάνεται από τον Teager-Kaiser ενεργειακό τελεστή, (γ) η ικανότητα του ESA να περιγράφει διαμορφώσεις ακόμη και σε πολύ μικρές κλίμακες, (δ) ο περιορισμός του θορύβου που συνεπάγεται μιας πολυκαναλικής ζωνοπερατής διαδικασίας.

6.4 Χαρακτηριστικά διαμόρφωσης

Οι ζώνες διαμόρφωσης εξάγονται μέσω συστοιχίας K Gabor φίλτρων και ο διακριτός ενεργειακός τελεστής Ψ_d εφαρμόζεται ως μη-γραμμική ενεργειακή μέτρηση. Η διαδικασία αποτυπώνεται γραφικά στο σχήμα 6.1. Για κάθε παράθυρο ανάλυσης m , λαμβάνονται αναπαραστάσεις βραχέως χρόνου των κυρίαρχων συνιστωσών διαμόρφωσης, εντοπίζοντας στον πολυ-διάστατο χώρο των αποκρίσεων των φίλτρων στο σήμα s , τη μέγιστη μέση Teager ενέργεια (*mean Multiband Teager Energy*) ή MTE :

$$\text{MTE}(m) = \max_{1 \leq k \leq K} \frac{1}{N} \sum_{n=1}^N \Psi_d[(s * h_k)(n)] \quad (6.14)$$

όπου το $*$ συμβολίζει συνέλιξη, n είναι ο δείκτης δείγματος με $(m-1)N+1 \leq n \leq mN$ και h_k η απόκριση του k -οστού φίλτρου.



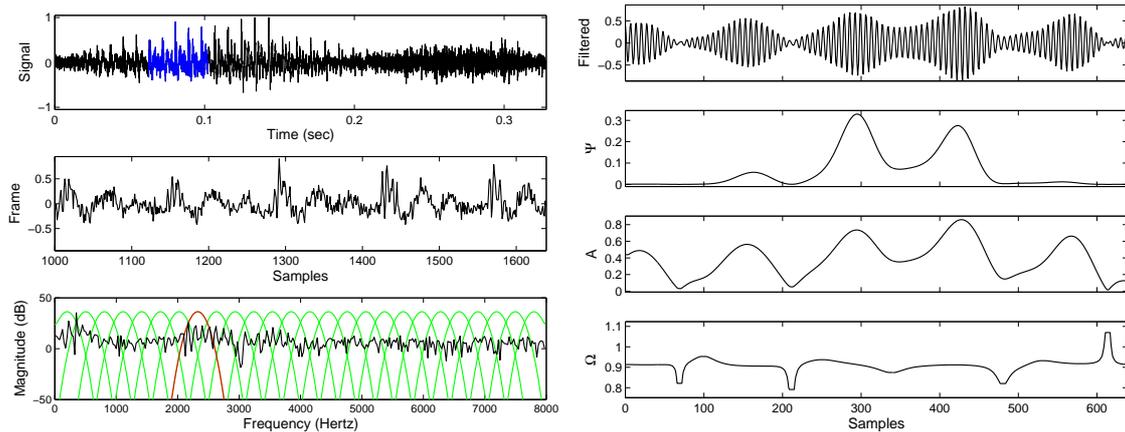
Σχήμα 6.1: Εξαγωγή χαρακτηριστικών διαμόρφωσης: Πολυκαναλικό φιλτράρισμα, εντοπισμός ενέργειας διαμόρφωσης και αποδιαμόρφωση του φίλτρου με την μέγιστη μέση Teager ενέργεια του τελεστή Ψ . Οι μέσες τιμές της αποδιαμορφωμένης εξόδου του επιλεγμένου φίλτρου δίνει τα χαρακτηριστικά μέσο πολυκαναλικό στιγμιαίο πλάτος και συχνότητα. Τα παράθυρα ανάλυσης είναι μήκους N δειγμάτων.

Το φίλτρο $i = \arg \max_k (\text{MTE})$ υποβάλλεται σε αποδιαμόρφωση μέσω του DESA και οι μέσες τιμές των εκτιμήσεων δίνουν, ανά παράθυρο, το κυρίαρχο μέσο πολυκαναλικό στιγμιαίο πλάτος (*mean multiband Instant Amplitude* ή MIA) και στιγμιαία συχνότητα (*mean Multiband Instant Frequency* ή MIF)

$$\text{MIA}(m) = \frac{1}{N} \sum_{n=1}^N |A_i(n)| \quad , \quad \text{MIF}(m) = \frac{1}{N} \sum_{n=1}^N \Omega_i(n) \quad (6.15)$$

Το MTE ερμηνεύεται ως η κυρίαρχη ενέργεια διαμόρφωσης (dominant modulation energy), που εντοπίζει από κοινού πληροφορία πλάτους και συχνότητας, οι οποίες υπάρχουν έμφυτες σε σήματα φωνής.

Στο σχήμα 6.2 φαίνεται για ένα παράθυρο ανάλυσης ενός σήματος, φωνής στην προκειμένη περίπτωση, η διαδικασία σάρωσης του φάσματος από τη συστοιχία (25 Gabor φίλτρα, rms ενεργό εύρος 160 Hz) και η επιλογή και αποδιαμόρφωση του κυρίαρχου. Τα νέα χαρακτηριστικά βραχέως χρόνου για ολόκληρο το σήμα, συγκρινόμενα με κλασσικές αναπαραστάσεις ενέργειας (μέσο απόλυτο πλάτος ή mAA) και συχνοτικής μεταβολής (μέσος ρυθμός μεταβάσεων από το μηδέν ή ZR) φαίνονται στο σχήμα 6.3.



Σχήμα 6.2: Ανάλυση και αποδιαμόρφωση φωνής ανά κανάλι. Για κάθε παράθυρο ανάλυσης, επιλέγεται και αποδιαμορφώνεται το κανάλι με τη μέγιστη μέση Teager ενέργεια ή κυρίαρχη ενέργεια διαμόρφωσης.

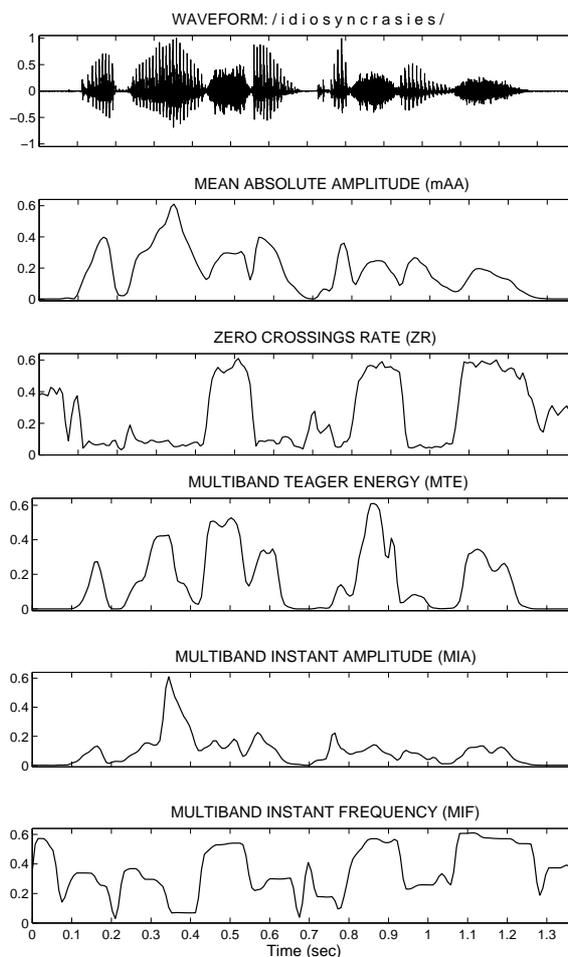
6.5 Εντοπισμός συνόρων φωνής σε θόρυβο

Τα χαρακτηριστικά διαμόρφωσης (MTE, MIA, MIF) χρησιμοποιήθηκαν αντί των συμβατικών mean absolute amplitude (mAA) και zero-crossings rate (ZR) σε μια καθιερωμένη μέθοδο ανίχνευσης ορίων φωνής (endpoint detection) με κατώφλια [160]. Από τα πρώτα 100 ms, που θεωρούνται εκ των προτέρων σιωπή ή θόρυβος περιβάλλοντος, η μέση τιμή μ_{sif} και η τυπική απόκλιση σ_{sif} της MIF υπολογίζονται μαζί με μέγιστες τιμές για την «ενέργεια» MTE σιωπής, S_{max} , αλλά και για ολόκληρο το σήμα P_{max} . Κανόνες κατωφλιών κατασκευάζονται χρησιμοποιώντας εμπειρικά βάρη κ , λ , σύμφωνα με τα

$$\gamma_f = \mu_{\text{sif}} + \kappa \sigma_{\text{sif}}, \quad \gamma_d = \min(T_1, T_2), \quad \gamma_u = 5 \cdot \gamma_d \quad (6.16)$$

$$T_1 = \lambda P_{\text{max}} + (1 - \lambda) S_{\text{max}}, \quad T_2 = 3 \cdot S_{\text{max}} \quad (6.17)$$

Ένας διπλός έλεγχος της «ενεργειακής» μέτρησης του σήματος, αναζητώντας τα ακραία σημεία όπου ξεπερνιέται ένα ψηλό κατώφλι γ_u , αμέσως μετά από ένα χαμηλότερο γ_d , ανιχνεύει το κύριο, συνήθως έμφωνο τμήμα της διαταραχής. Τα αρχικά εκτιμώμενα σύνορα μετατοπίζονται έτσι ώστε να περικλείσουν τυχόν ισχυρή φασματική δραστηριότητα, χρησιμοποιώντας το γ_f κατώφλι μετρήσεων «συχνοτικού» περιεχομένου. Για τις εδώ δοκιμές δόθηκαν στα βάρη οι τιμές $\lambda = 0.02$, $\kappa = 1$. Στο σχήμα 6.4 δίνονται παραδείγματα του αλγορίθμου με τα κλασσικά και τα νέα χαρακτηριστικά διαμόρφωσης.



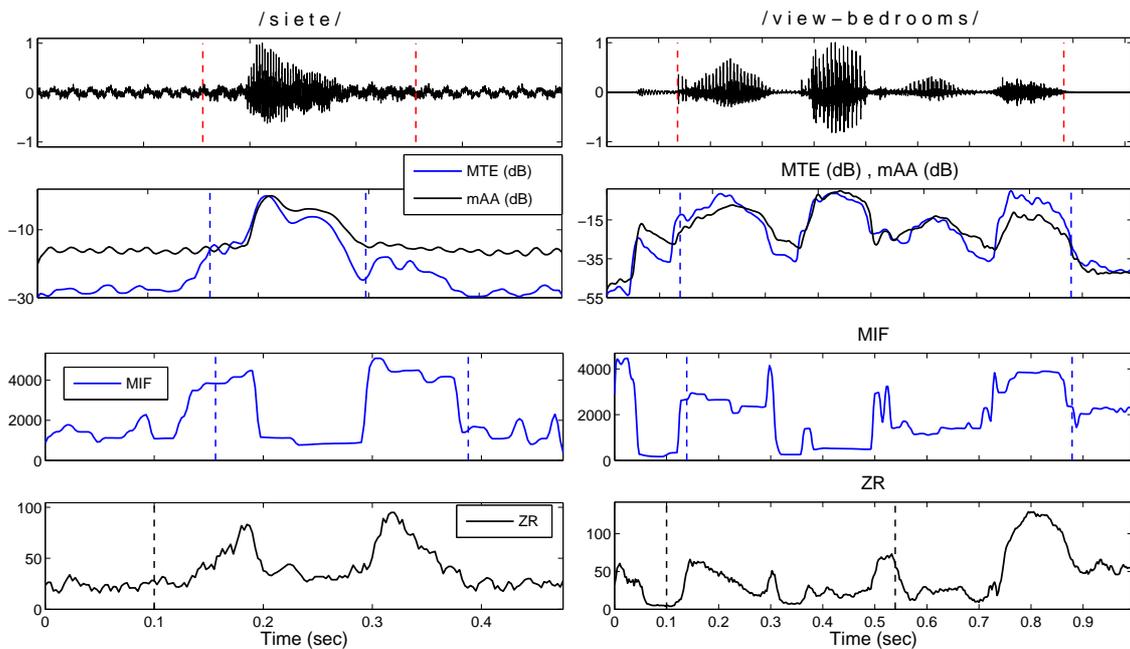
Σχήμα 6.3: Χαρακτηριστικά ανάλυσης και ανίχνευσης φωνής σε 15 ms par’ajura Hamming, με 1/2 της διάρκειας επικάλυψη, στα 16 kHz. Κυματομορφή, κλασικά χαρακτηριστικά (mAA, ZR) και τα νέα χαρακτηριστικά διαμόρφωσης MTE, MIA & MIF. Όλες οι παραστάσεις είναι κανονικοποιημένες και ομαλοποιημένες με συνδυασμό γραμμικών και μη-γραμμικών φίλτρων (median 3 σημείων, Hanning 3 σημείων).

Τα χαρακτηριστικά για ανίχνευση συνόρων δοκιμάστηκαν κάτω από πραγματικές συνθήκες τηλεφωνικού θορύβου στη βάση NTIMIT. Το εγχείρημα ήταν η ανίχνευση ορίων φράσεων, ανεξάρτητα από το είδος της ενδιάμεσης δραστηριότητας, συγκρινόμενα με τα όρια που έχουν προκύψει από ανθρώπινη επισκόπηση της βάσης. Ορίστηκε ως σφάλμα ανίχνευσης οποιαδήποτε λανθασμένη τοποθέτηση των συνόρων άνω των 60 ms. Ποσοστά ορθής ανίχνευσης για ολόκληρη τη βάση (1570 φράσεις), με διαφορετικούς συνδυασμούς χαρακτηριστικών φαίνονται στον πίνακα 6.5. Κάθε συνδυασμός περιλαμβάνει μια «ενεργειακή» και μια «συχνοτική» μέτρηση. Στήλες με ένα μόνο χαρακτηριστικό αντιστοιχούν σε αποτελέσματα από ένα μόνο έλεγχο με κατώφλια, χωρίς βελτίωση των αρχικών εκτιμήσεων.

Το mSA αναφέρεται στο μέσο τετραγωνικό πλάτος, ενώ τα MAA, MZR είναι πο-

Κλασικά		Πολυκαναλικά	Διαμόρφωσης	Teager	
mAA, ZR	mSA, ZR	MAA, MZR	MTE, MIF	MTE	STE PTE
56.1	66.6	51.5	73.5	73.1	71.6 49.5

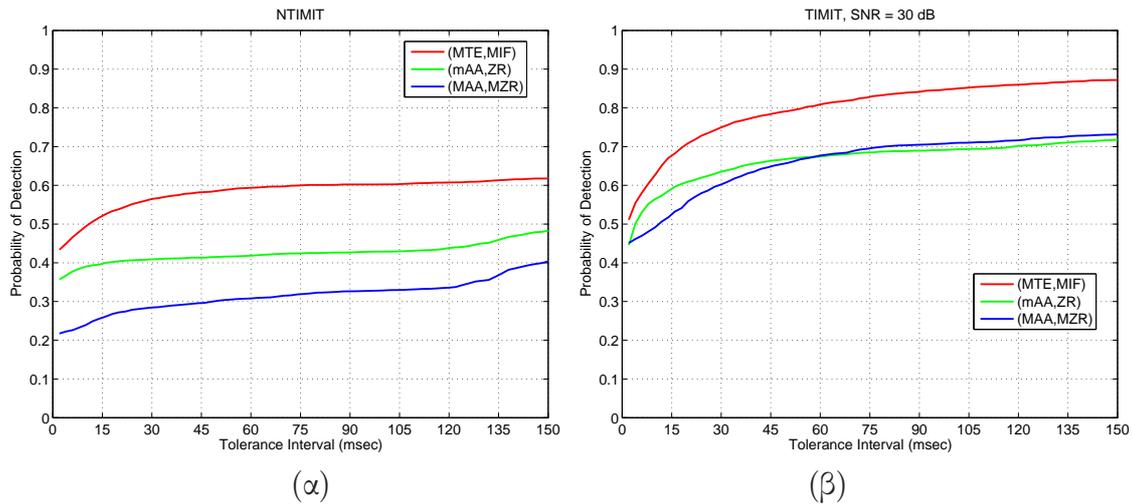
Πίνακας 6.1: Ορθή ανίχνευση ορίων φωνής (%) στη βάση NTIMIT



Σχήμα 6.4: Παραδείγματα εντοπισμού συνόρων φωνής με κλασσικά και νέα χαρακτηριστικά σε δύο λέξεις (πάνω σειρά) με τις κάθετες γραμμές παριστάνουν τα πραγματικά σύνορα. Ακολούθως παρουσιάζονται τα νέα και τα συμβατικά χαρακτηριστικά και τα αποτελέσματά τους. Οι «ενεργειακές» μετρήσεις (mAA, MTE) παρουσιάζονται σε dB σε κοινή παράσταση για λόγους σύγκρισης.

λυκαναλικές παραλλαγές των κλασσικών και αποτελούν αντίστοιχα το μέγιστο από όλες τις ζώνες μέσο πλάτος και ρυθμό zero-crossings στη ζώνη με τη μέγιστη μέση φιλτραρισμένη περιβάλλουσα. Τέλος οι δύο τελευταίες στήλες αντιστοιχούν σε εναλλακτικές μετρήσεις με την Teager ενέργεια. Η PTE περιγράφεται στο [210], ενώ η STE είναι η μέση έξοδος του ενεργειακού τελεστή ανά παράθυρο, χωρίς πολυκαναλική ανάλυση. Συμπερασματικά, η αποτελεσματικότητα των νέων χαρακτηριστικών οφείλεται ισάξια στα πλεονεκτήματα τόσο της πολυκαναλικής ανάλυσης όσο και των μετρήσεων ενέργειας διαμορφώσεων. Η μείωση του μέσου σφάλματος ανίχνευσης, σε σχέση με την κλασσική προσέγγιση είναι 38.7% χρησιμοποιώντας μόνο το MTE και 40.1% βελτιώνοντας τα αποτελέσματα με το MIF.

Για την εκτίμηση των αποτελεσμάτων ανεξάρτητα από τα εμπειρικά ορισμένα διαστήματα σφάλματος, χρησιμοποιήθηκε μια απλή σύμβαση για την παραγωγή καμπυλών που προσεγγίζουν τις καμπύλες ROC (Receiver Operating Characteristics). Για τα δύο είδη σφαλμάτων τοποθέτησης, τίθεται αυστηρά στα 30 ms η ανοχή σε χαμένη διάρκεια φωνημάτων (lost-phoneme error) και αφήνεται η ανοχή πρώιμης αρχής ή καθυστέρησης τέλους (spoke-too-soon error) να μεταβάλλεται στο διάστημα 2 με 150 ms. Αυτό το διάστημα, κανονικοποιημένο στο μέγιστο του, δίνει ένα μέτρο που σχετίζεται με την πιθανότητα εσφαλμένου συναγερού. Οι δύο ποσότητες δεν είναι ίσες αλλά συνδέονται, με μια ένα προς ένα, μονοτονικά αύξουσα, άγνωστη συνάρτηση [57], καθώς αύξηση στην ανοχή, αυξάνει την πιθανότητα κατά κάποιο ποσοστό. Στο σχήμα 6.5 φαίνονται τέτοιες καμπύλες ανίχνευσης - ανοχής για τα κλασσικά χαρακτηριστικά, τις πολυκαναλικές εκδοχές τους και τα χαρακτηριστικά διαμόρφωσης.



Σχήμα 6.5: Καμπύλες ανίχνευσης - ανοχής λάθους, (α) στην NTIMIT και (β) στην TIMIT με προσθετικό θόρυβο (SNR=30 dB). Το χρονικό διάστημα στον οριζόντιο άξονα είναι το διάστημα ανοχής σφάλματος ορίων των φράσεων και σχετίζεται με την πιθανότητα λανθασμένου συναγερμού.

6.6 Ανίχνευση φωνητικής δραστηριότητας

Τα συστήματα ανίχνευσης φωνητικής δραστηριότητας (*Voice Activity Detection* ή VAD), κατηγοριοποιούν εισερχόμενα παράθυρα ενός σήματος, με βάση εξαγόμενα χαρακτηριστικά, χονδρικά σε δύο τάξεις: φωνητικά και μη-φωνητικά γεγονότα (παύσεις, σιωπή, θόρυβος, αλλά και ψίθυροι, ασθενή άφωνα κ.α.). Ένα τέτοιο σύστημα, σχεδιασμένο για ακρίβεια και ανοχή σε θόρυβο, προτάθηκε στα [161,162], και περιλαμβάνει τη χρήση της μακροπρόθεσμης φασματικής απόκλισης (*Long Term Spectral Divergence* ή LTSD). Ο αλγόριθμος βασίζεται σε προσαρμοζόμενα κατώφλια και ενημέρωση των παραμέτρων του θορύβου. Το εκτιμώμενο χαρακτηριστικό ποσοτικοποιεί την απόκλιση της φωνής από τον θόρυβο περιβάλλοντος και είναι στην ουσία μια «ενεργειακή» μέτρηση που διατηρεί και φασματική πληροφορία, αυτή των ισχυρών φασματικών συνιστωσών σε γειτονικά παράθυρα ανάλυσης. Το VAD βασισμένο στο LTSD δοκιμάστηκε εξαντλητικά και επιτυχημένα κάτω από διάφορες, πραγματικές συνθήκες θορύβου και συγκρίθηκε με πρότυπα VAD [11], χρησιμοποιώντας κοινές μεθόδους αποτίμησης αλλά και ακρίβεια αναγνώρισης.

6.6.1 Ανίχνευση ενέργειας διαμόρφωσης

Η αποτελεσματικότητα του MTE ως ανιχνευτή γεγονότων φωνής σε ένα VAD σύστημα γίνεται προσαρμόζοντας τον βασισμένο στο LTSD αλγόριθμο και αλλάζοντας το βασικό χαρακτηριστικό του με την κυρίαρχη ενέργεια διαμόρφωσης MTE σε δύο εναλλακτικές εκφράσεις απόκλισης: μακράς και βραχείας. Κατά τη διάρκεια μιας περιόδου αρχικοποίησης μαθαίνονται τα χαρακτηριστικά του θορύβου. Μετά την εξαγωγή των χαρακτηριστικών, ενότητα 6.4, η διαφορά επιπέδου σε dB σε σχέση με την αντίστοιχη μέτρηση θορύβου, συγκρίνεται με ένα προσαρμοζόμενο κατώφλι $\gamma \in [\gamma_0, \gamma_1]$:

$$\gamma = \gamma_0 + (\gamma_1 - \gamma_0)(E - E_0)/(E_1 - E_0) \quad (6.18)$$

όπου E η ενέργεια θορύβου περιβάλλοντος, ενώ τα όρια του διαστήματος εξαρτώνται από τις ενέργειες E_0, E_1 στις πιο «καθαρές» και πιο θορυβώδεις συνθήκες της υπό

εξέτασης βάσης, υπολογισμένες κατά την αρχικοποίηση. Η μέτρηση του χαρακτηριστικού του θορύβου προσαρμόζεται κάθε φορά που σιωπή ή παύση ανιχνεύεται, με τοπικούς μέσους σε μια μικρή γειτονιά παραθύρων. Με τον τρόπο αυτό λαμβάνονται υπόψη μη-στατικές συνθήκες θορύβου.

Για να μετρηθούν «αποκλίσεις» ενέργειας διαμόρφωσης, στο πνεύμα του LTSD, χρησιμοποιούνται χαρακτηριστικά βασισμένα στο μέγεθος MTE:

1. Απόκλιση Πολυκαναλικής Teager Ενέργειας (*Multiband Teager Energy Divergence* ή MTED), η μέγιστη ανά κανάλι μέση ενέργεια MTE, όπως ορίστηκε προηγούμενα συγκρινόμενη με το αντίστοιχο μέγεθος MTEW για το θόρυβο

$$\text{MTED}(m) = 10 \log_{10} (\text{MTE}(m)/\text{MTEW}) \quad (6.19)$$

Το MTED μετράει την απόκλιση της ενέργειας διαμόρφωσης ενός frame από την ενέργεια αναφοράς του θορύβου. Ουσιαστικά η διαδικασία είναι ίδια με τον αλγόριθμο ανίχνευσης ορίων της ενότητας 6.5, που συγκρίνει τη διαφορά μεγέθους του MTE.

2. Μακροπρόθεσμη Απόκλιση Πολυκαναλικής Teager Ενέργειας (*Long-term Multiband Teager Energy Divergence* ή LTED), όπου το MTE μεγιστοποιείται σε μια γειτονιά $2L$ παραθύρων, καταλήγοντας σε μια διεσταλμένη και κανονικοποιημένη ως προς τον θόρυβο εκδοχή του:

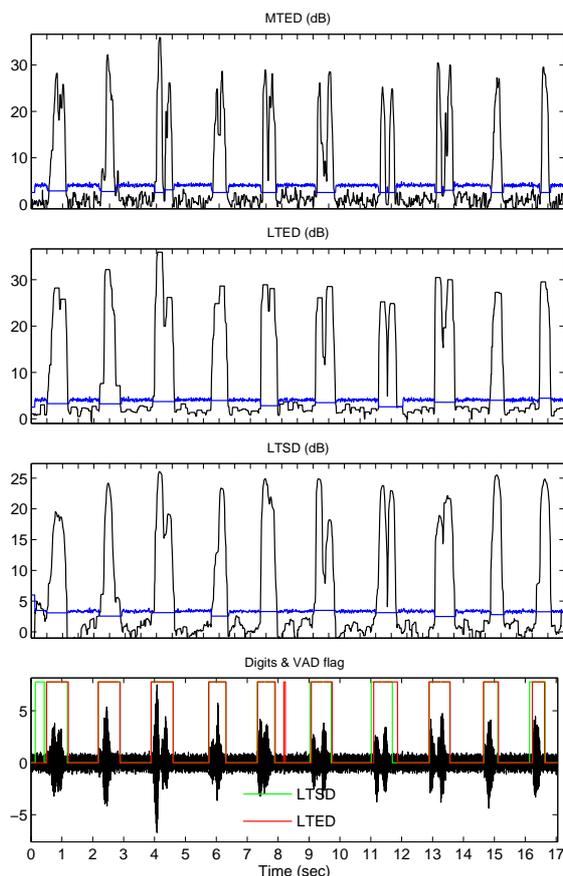
$$\text{LTED}(m) = 10 \log_{10} \left(\max_l \{ \text{MTE}(m+l) \} / \text{MTEW} \right) \quad (6.20)$$

με το $-L \leq l \leq L$ να καθορίζει την τάξη της εξάρτησης από γειτονικά frame.

Στο σχήμα 6.6 παρουσιάζεται ένα παράδειγμα των προτεινόμενων χαρακτηριστικών για VAD σε ένα παράδειγμα ακολουθίας ψηφίων από τη βάση Aurora 3 (quiet, hands-free mic.), για παράθυρα 25 ms με 10 ms μετατοπίσεις και μια συστοιχία 25 Gabor. Μαζί φαίνεται και το προσαρμοζόμενο κατώφλι που σηματοδοτεί τη φωνητική δραστηριότητα. Διαισθητικά τα MTE αναμένονται να δουλεύουν καλύτερα, καθώς μεγιστοποιούν τη μέτρηση ανά frame τόσο συχνοτικά (μέσω της πολυκαναλικής ανάλυσης), όσο και χρονικά (μέσω της τοπικής γειτονιάς). Έτσι εξασφαλίζεται η κυρίαρχη τοπικά χρονο-φασματική περιγραφή του σήματος.

6.6.2 Αποτελέσματα και συγκρίσεις

Το πειραματικό πλαίσιο περιλαμβάνει συγκρίσεις απόδοσης ανίχνευσης του LTSD αναφοράς και των βασισμένων στο MTE VADs, (6.19, 6.20), κάτω από διαφορετικές συνθήκες θορύβου στις βάσεις Aurora 2 (70070 φράσεις) και Aurora 3 (4914 φράσεις). Η αποτίμηση βασίζεται σε λάθη κατηγοριοποίησης σε διαφορετικά SNR [11, 132, 161] χρησιμοποιώντας κάποιο χαρακτηρισμό ή ονοματισμό αναφοράς του καθαρού συνόλου δεδομένων. Εδώ χρησιμοποιήθηκε αυτόματη αναγνώριση φωνής για κατάτμηση και ονοματισμό των γεγονότων φωνής και σιωπής στις βάσεις. Αποτελέσματα υψηλού ποσοστού αναγνώρισης στις καθαρές ακολουθίες ψηφίων καθορίζει το ground truth. Εν συντομία, για το σύνολο Aurora 2, η εκπαίδευση του συστήματος γίνεται χρησιμοποιώντας 32 mixtures, 18 καταστάσεις και το μήκος 39 διάνυσμα χαρακτηριστικών $[\text{MFCC}, \log E, \Delta, \Delta\Delta]^T$ στο σενάριο “clean-train” με τις δοκιμές στα καθαρά δεδομένα να πετυχαίνουν ένα ποσοστό 99.6% ακρίβειας αναγνώρισης λέξεων. Για την Aurora 3,



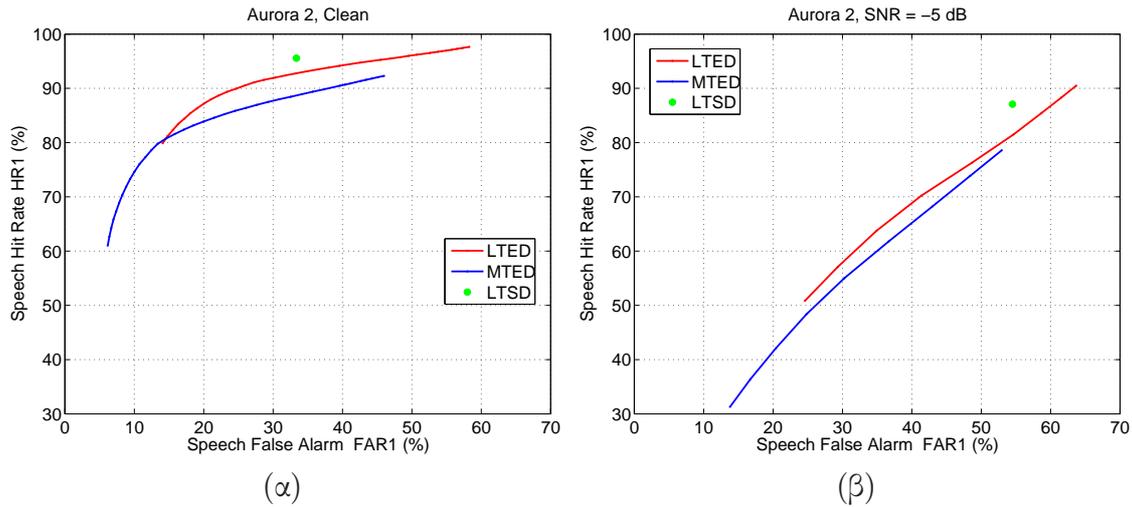
Σχήμα 6.6: Παραδείγματα ανίχνευσης κυρίαρχης ενέργεια διαμόρφωσης MTE για VAD και το χαρακτηριστικό αναφοράς LTSD [161]. Τα διαστήματα δραστηριότητας ανιχνεύθηκαν παρόμοια και με τα δύο χαρακτηριστικά, LTED και MTED.

η εκπαίδευση γίνεται με 16 mixtures, 16 καταστάσεις και το ίδιο χαρακτηριστικά. Το υποσύνολο των 1522 για το “well matched” σενάριο δοκιμών χρησιμοποιήθηκε με 93.7% ακρίβεια αναγνώρισης.

Για το VAD αναφοράς LTSD χρησιμοποιήθηκαν οι προδιαγραφές που αναφέρονται στο [161], ενώ για τα προτεινόμενα VAD τα βέλτιστα κατώφλια (6.18), εκτιμήθηκαν μέσω καμπυλών ROC. Στο σχήμα 6.7 παρουσιάζονται αυτές οι καμπύλες στα πιο θορυβώδη και πιο καθαρά σύνολα για τα VAD με χαρακτηριστικά τα MTED και LTED. Επιλέχθηκαν τα κατώφλια που αντιστοιχούν στα σημεία των καμπυλών ελάχιστης απόστασης από την άνω αριστερή γωνία του ιδανικού σημείου λειτουργίας. Αυτό οδήγησε στα $\gamma_0 = 24$ dB, $\gamma_1 = 0.5$ dB για το MTED VAD και $\gamma_0 = 32$ dB, $\gamma_1 = 2$ dB για το LTED στην Aurora 2. Οι δοκιμές στην Aurora 3 έγιναν με τα ίδια ζεύγη κατωφλιών ($\gamma_0 = 6$ dB, $\gamma_1 = 2.5$ dB) και για τα τρία χαρακτηριστικά.

Η απόδοση των VAD εκτιμήθηκε σχετικά με το ποσοστό ορθών ανιχνεύσεων φωνής (speech Hit Rate ή HR1), που ορίζεται ως ο λόγος των παραθύρων φωνής που έχουν ανιχνευθεί ως προς το συνολικό αριθμό παραθύρων φωνής και το ποσοστό ορθών ανιχνεύσεων σιωπής (non-speech Hit Rate ή HR0), ορισμένο ανάλογα για τη σιωπή. Συμπληρωματικά με αυτές τις ποσότητες ορίζονται τα ποσοστά λανθασμένου συναγερμού (False Alarm Rates, FAR1 και FAR0) των αποφάσεων για φωνή και θόρυβο.

Στους πίνακες 6.2 και 6.3 παρουσιάζεται η απόδοση για τα τρία VAD, στις βάσεις αναφοράς. Τα αποτελέσματα αντιστοιχούν και στα δύο ποσοστά, σε μέσες τιμές από όλα τα δεδομένα, ανά επίπεδο θορύβου. Τα λάθη κατηγοριοποίησης συμβαίνουν



Σχήμα 6.7: Καμπύλες ROC για αποτίμηση απόδοσης της ανίχνευσης φωνής στο (α) για το «καθαρό» σύνολο και (β) για το πιο θορυβώδες (-5 dB) σύνολο της Aurora 2 για τα VAD βασισμένα στο MTE. Παρουσιάζεται επίσης το σημείο λειτουργίας του LTSD-VAD.

Πίνακας 6.2: Ακρίβεια Ανίχνευσης, Aurora 2

Θόρυβος (dB)	Χαρακτηριστικά VAD					
	LTSD		LTED		MTED	
	HR1	HR0	HR1	HR0	HR1	HR0
clean	95.6	66.6	89.7	72.7	88.1	67.3
20	89.9	80.4	83.2	83.6	79.4	78.4
15	90.1	75.8	80.2	83.8	77.1	78.2
10	90.3	70.3	77.2	83.9	73.9	78.4
5	89.8	63.4	71.9	79.7	75.6	68.8
0	89.9	67.4	74.3	81.4	75.1	72.6
-5	87.1	45.5	51.0	75.0	70.7	54.4
Μέσοι (%)	90.4	67.1	75.4	80.0	77.1	71.1
Νόρμα λάθους (%)	34.3		31.7		36.8	

HR1: speech Hit Rate, HR0: non-speech Hit Rate

Πίνακας 6.3: Ακρίβεια Ανίχνευσης, Aurora 3

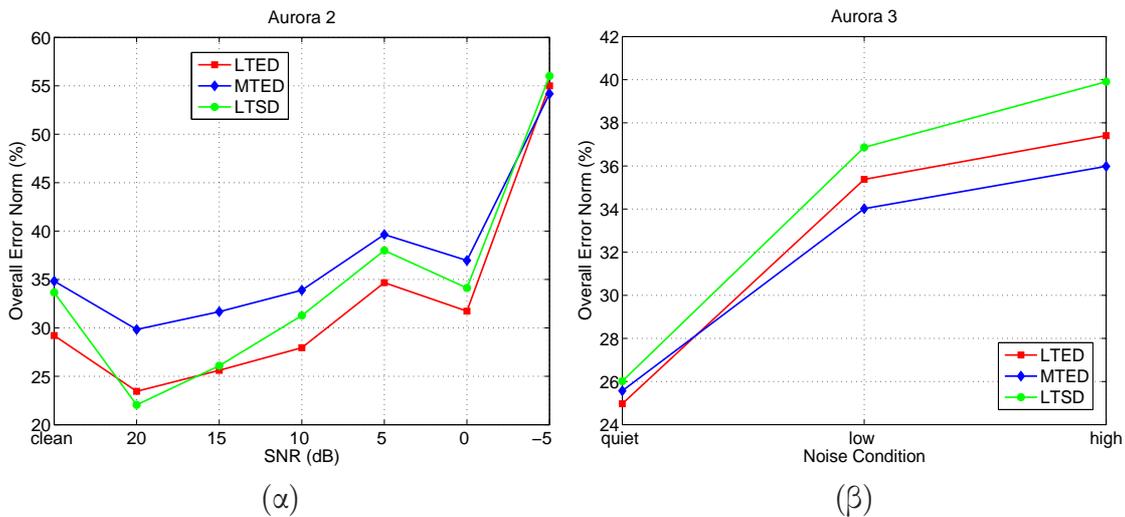
Συνθήκες Θορύβου	Χαρακτηριστικά VAD					
	LTSD		LTED		MTED	
	HR1	HR0	HR1	HR0	HR1	HR0
quiet	85.2	66.4	87.4	70.7	73.2	81.1
low	85.3	66.2	67.8	66.8	74.1	78
high	86.3	62.5	87.3	64.8	71.0	78.7
Μέσοι (%)	85.6	65.1	87.5	67.4	79.3	72.8
Νόρμα λάθους (%)	37.8		34.9		34.2	

HR1: speech Hit Rate, HR0: non-speech Hit Rate

και στις δύο κατηγορίες γεγονότων (φωνής και σιωπής/θορύβου), με αποτέλεσμα τα ποσοστά HR1, HR0 να θεωρούνται ίσης βαρύτητας και σημασίας. Αυτό μπορεί να ποσοτικοποιηθεί από την L_2 νόρμα των λανθασμένων συναγερωμών. Επιδιώκεται, για καλύτερη λειτουργία, η ελαχιστοποίηση της νόρμας λάθους (*overall false alarm error norm*):

$$\|(FAR0, FAR1)\| = [(1 - HR0)^2 + (1 - HR1)^2]^{1/2} \quad (6.21)$$

Στατιστικά αυτό το μετρικό εκφράζει την μέση απόδοση του ανιχνευτή καθώς είναι $\sqrt{2}$ φορές η rms νόρμα των FAR, ενώ γεωμετρικά είναι η μικρότερη Ευκλείδεια απόσταση από το ιδανικό σημείο λειτουργίας στην ROC καμπύλη (HR=100, FAR=0) (σχήμα 6.7). Στο σχήμα 6.8 οι νόρμες λάθους παρουσιάζονται για τα δύο σύνολα δεδομένων και τα τρία διαφορετικά VAD ως συνάρτηση της μείωσης του SNR.



Σχήμα 6.8: Ολικά σφάλματα (overall false alarm error norms) ανίχνευσης φωνητικών γεγονότων κάτω από διαφορετικά SNR στις βάσεις (α) Aurora 2 και (β) Aurora 3.

Στις δοκιμές στην Aurora 2, όπου τα κατώφλια επιλέχθηκαν κατά βέλτιστο τρόπο, οι MTE αλγόριθμοι ισοσταθμίζουν τα δύο ποσοστά δίνοντας μέσα hit rates πάνω από 70% τόσο για παρουσία όσο και για απουσία φωνής. Το LTED ελαχιστοποιεί τη νόρμα λάθους (6.21), με 7.6% σχετική μείωση του ολικού λάθους έναντι στο LTSD. Από το σχήμα 6.8(α) φαίνεται ότι ο LTED ανιχνευτής ελαχιστοποιεί το σφάλμα, εκτός από τα 20 dB SNR, όπου και τα τρία χαρακτηριστικά ακολουθούν ανάλογη πτώση απόδοσης κάτω από αυξανόμενο SNR. Στο σύνολο Aurora 3, όπου τα κατώφλια ανίχνευσης είναι τα ίδια, το LTED πετυχαίνει υψηλότερη απόδοση σε μεμονωμένα hit rate απ' ότι το LTSD και μια συνολική σχετική μείωση σφάλματος του 7.7%. Η ελάχιστη νόρμα λάθους δίνεται από το MTED χαρακτηριστικό με σχετική μείωση κατά 9.5%, ενώ και οι δύο αλγόριθμοι που βασίζονται στα χαρακτηριστικά διαμόρφωσης ξεπερνάνε το LTSD στο συνολικό σφάλμα κάτω από όλες τις συνθήκες θορύβου, όπως φαίνεται και στο σχήμα 6.8(β). Σημειώνεται ότι το LTED είναι με συνέπεια καλύτερο και στα δύο σύνολα δοκιμών.

6.7 Συμπεράσματα

Η ανίχνευση φωνητικών γεγονότων προσεγγίζεται μέσω της ανίχνευσης μικροδομών διαμόρφωσης, εντοπίζοντας και μετρώντας κατάλληλα και σε πολλαπλές ζώνες

την κυρίαρχη ενέργεια τους και εξάγωντας αργά μεταβαλλόμενη πληροφορία κυρίαρχου πλάτους και συχνότητας. Τα χαρακτηριστικά διαμόρφωσης διαπιστώθηκαν ότι βελτιώνουν συστηματικά την ανίχνευση των ορίων της φωνής σε μεμονωμένες λέξεις ή φράσεις και μειώνουν το μέσο ολικό σφάλμα ανίχνευσης ενός σθεναρού και μεγάλης ακρίβειας ανιχνευτή φωνητικής δραστηριότητας, σε μεγάλες βάσεις αναφοράς. Η επέκταση αυτών των ιδεών για γενικευμένη ανίχνευση και ανάλυση γεγονότων μπορεί να γίνει μέσω μιας συστηματικής περιγραφής με όρους μικροδιαμορφώσεων των κατηγοριών ακουστικών σημάτων.

□

Κεφάλαιο 7

Μελλοντική Έρευνα και Επεκτάσεις

Η μελέτη της υψής μέσω των μη-γραμμικών μοντέλων διαμορφώσεων έχει αναδείξει την ικανότητα τέτοιων μοντέλων να εξάγουν και να περιγράφουν δομές σε πολλαπλές κλίμακες, με εφαρμογές σε ζητήματα αναπαράστασης εικόνων και υπολογιστικής όρασης. Η παράλληλη εφαρμογή προς την αναζήτηση παρόμοιων δομών φωνής, με δεδομένο ότι θεματικά οι διαμορφώσεις σε σήματα φωνής έχουν εξετασθεί αρκετά νωρίτερα και εκτενέστερα, οδήγησε στην ανάπτυξη χαρακτηριστικών ικανών για διαχωρισμό φωνητικών γεγονότων. Κλείνοντας επιχειρείται εδώ μια σύνοψη της προόδου, και μια οριοθέτηση των μελλοντικών στόχων και προσεγγίσεων στα ως τώρα ανοιχτά προβλήματα, κυρίως γύρω από την οπτική υφή αλλά και τη μικροδομή της φωνής.

7.1 Δεδομένα, Μερική πρόοδος

Το μοντέλο πολυζωνικών διαμορφώσεων χρησιμοποιείται για την αναπαράσταση πληροφορίας οπτικής υψής. Η ισχύς του έγκειται στην περιγραφή με ένα σύνολο παραμέτρων διαμόρφωσης που περιλαμβάνουν την τοπική χρωματική αντίθεση και τις αναδυόμενες χωρικές συχνότητες οι οποίες ευθύνονται για την κλίμακα, τον προσανατολισμό και την τραχύτητα της υψής. Εντοπίζοντας και εξάγοντας την ενέργεια των κυρίαρχων τοπικών διαμορφώσεων, με τη βοήθεια μη-γραμμικών ενεργειακών τελεστών και πολυκαναλική ανάλυση και διαχωρίζοντας τη στιγμιαία, προκύπτουν πλούσιες σε πληροφορία και μικρές σε διάσταση περιγραφές της υψής μιας εικόνας. Τα χαρακτηριστικά κυρίαρχων συνιστωσών υψής επιλέχθηκαν και χρησιμοποιήθηκαν για θέματα ανάλυσης (εξαγωγή χαρακτηριστικών [98]), κατάτμησης (μεταβολικά σχήματα με βάση τα χαρακτηριστικά διαμόρφωσης [99]), ανίχνευσης (κινητήριες δυνάμεις σε πολυκριτηριακό σχήμα [177, 178]) και κατηγοριοποίησης (διαχωρισμός τάξεων με μικρής διάστασης διανύσματα και ενεργειακούς τελεστές [177]).

Πρόσφατες θεωρίες και μοντέλα για διαχωρισμό υψής, προβλέπουν ορισμούς των διαφορετικών συνιστωσών σε συμπληρωματικούς χώρους συναρτήσεων και μεταβολικές υπολογιστικές τεχνικές για την ανάκτηση τους. Μια πρώτη προσέγγιση της συνιστώσας των ταλαντώσεων της υψής έγινε με τη βοήθεια μορφολογικών τελεστών συναρτήσεων και πυραμίδες από απλουστευμένες εκδοχές της εικόνας σε πολλαπλές κλίμακες. Η προσέγγιση μέσω levelings προσφέρει πλεονεκτήματα αλλά δεν έχει διερευνηθεί ακόμη το αν ικανοποιεί τις παραδοχές του αρχικού $u + v$ μοντέλου ή αν η συνιστώσα υψής, που λαμβάνεται συμπληρωματικά από τη διαδικασία, ανήκει στο χώρο

συναρτήσεων του Meyer [137].

Αν και δεν έχει δοθεί έκταση στην παρούσα αναφορά, μερική πρόοδος έχει επιτευχθεί και στη μοντελοποίηση, ανάλυση και σύνθεση στοχαστικής υψής με χρήση του μοντέλου AM-FM συνιστωσών με στοχαστική (ή και φράκταλ) διαμόρφωση φάσης. Τα σήματα στιγμιαίων συχνοτήτων που προκύπτουν από αποδιαμόρφωση των μη-στατικών συνιστωσών παρουσιάζουν αρκετά τυχαία δομή. Με κίνητρο ανάλογα φαινόμενα σε σήματα φωνής (π.χ. συριστικούς ήχους), η τυχαία φάση των αποδιαμορφωμένων σημάτων μοντελοποιείται ως στοχαστικός, $1/f$ θόρυβος. Η στιγμιαία χωρική μεταβολή της φάσης λαμβάνεται με ολοκλήρωση στο πεδίο της εικόνας, η οποία παραμετροποιεί και το μοντέλο ως προς τον εκθέτη της μεταβολής του φάσματος ισχύος. Η μέτρηση του εκθέτη γίνεται είτε απευθείας με εκτίμηση της κλίσης του λογαριθμικού φάσματος είτε έμμεσα από τον υπολογισμό της φράκταλ διάστασης της επιφάνειας. Η εκτίμηση και ο έλεγχος του μοντέλου γίνεται με ανάλυση συνθετικής υψής γνωστών παραμέτρων και σύνθεση εικόνων με τυχαία παραγόμενη φάση μέσω $1/f$ διαδικασιών, κάποιες από τις οποίες είναι fBm.

Προς μια ενιαία αντιμετώπιση της υψής, με βάση τις απαιτήσεις και τις διαισθητικές παραδοχές του Κεφαλαίου 2, έγινε μια πρώτη προσέγγιση στο πλαίσιο των χωρικών διαμορφώσεων. Τα θεμελιώδη στοιχεία ή primitives της υψής είναι, σε ένα αφηρημένο για την ώρα βαθμό, οι χωρικά τοπικά εμφανιζόμενες συνιστώσες διαμόρφωσης σε στενές ζώνες συχνοτήτων. Τέτοιες συνιστώσες εντοπίζονται και οριοθετούν περιοχές, καλύπτουν διαφορετικές κλίμακες, και μπορούν να ορισθούν με το σύνολο των παραμέτρων διαμόρφωσης σε ένα χωρικό πλέγμα. Μέσω της κυρίαρχης ενέργειας διαμορφώσεων υψής, ποσοτικοποιείται η συγκέντρωση και η παρουσία των κυρίαρχων primitives διαμόρφωσης σε κάθε θέση και καθορίζεται αν μια περιοχή είναι ή διαθέτει κατ' αυτόν τον τρόπο υφή.

Η άμεση εφαρμογή έχει τεθεί σε θέματα ανάλυσης εικόνων εδαφοτομών για ζητήματα βιολογικού ελέγχου. Η υφή τέτοιων εικόνων, που ακόμη και κάτω από ανθρώπινη παρατήρηση είναι δύσκολο να χαρακτηριστεί και να κατηγοριοποιηθεί, εμπλέκεται σε όλα τα στάδια της ανάλυσης. Ως κυρίαρχο και σημαντικό χαρακτηριστικό, περιγράφεται μέσω πολυκαναλικής αποδιαμόρφωσης και κυρίαρχων χαρακτηριστικών, οδηγεί τη διαδικασία κατάτμησης και συμμετέχει ως κριτήριο για βελτίωση του τελικού διαχωρισμού. Επιπλέον βρέθηκε ότι μια πρώιμη κατηγοριοποίηση των ομογενών περιοχών με βάση την υφή τους επιτελείται καλύτερα με χρήση των μεθόδων και των χαρακτηριστικών που έχουν αναπτυχθεί.

Περνώντας στη μία διάσταση, με παρόμοιες τεχνικές εντοπίζονται και αναλύονται οι μικροδομές διαμόρφωσης σε σήματα φωνής και τα εξαγόμενα χαρακτηριστικά χρησιμοποιούνται με επιτυχία για ανίχνευση φωνητικής δραστηριότητας. Αν και δεν υπάρχει η άμεση ερμηνεία τέτοιων δομών, όπως γίνεται για την οπτική υφή, εμφανίζονται σε διαφορετικές κλίμακες και συνδέονται με την ταχέως μεταβαλλόμενη συνιστώσα της ακουστικής πληροφορίας. Η ανίχνευση τους ταυτίστηκε συστηματικά και σε μεγάλες βάσεις αναφοράς με το διαχωρισμό φωνής και θορύβου. Εναλλακτικά εργαλεία χρονοφασματικής ανάλυσης και μέτρησης ενεργειακών φασματικών συγκεντρώσεων έχουν αναπτυχθεί σε πρώιμο στάδιο που δεν έχουν αναφερθεί εδώ, για την επεξεργασία και την ανάλυση ακόμη και σε πολύ μικρές κλίμακες σημάτων που μοντελοποιούνται μέσω AM-FM διαμορφώσεων. Ένα παράδειγμα εφαρμογής είναι η ανάλυση βιολογικών σημάτων υπερηχογραφήματων, με σχεδόν άριστα αποτελέσματα χρονικής ευκρίνειας ανάλυσης και περιγραφής.

7.2 Ζητούμενα

Σε σχέση με την περιγραφή και ανάλυση της υψής, ένα ανοιχτό ζήτημα είναι η ολοκλήρωση του μοντέλου και ο καθορισμός του συνόλου των σημαντικών παραμέτρων του. Τέτοιες παράμετροι μπορεί είναι ο αριθμός των συνιστωσών, ο σχεδιασμός της συστοιχίας των φίλτρων και τα περισσότερο περιγραφικά χαρακτηριστικά για θέματα κατάτμησης και κατηγοριοποίησης.

Η μαθηματική και φυσική σύνδεση των χωρικών διαμορφώσεων με τις συναρτήσεις ταλάντωσης που ορίστηκαν από το Meyer [137] για μοντελοποίηση και διαχωρισμό της υψής, φαντάζει ενδιαφέρουσα καθώς οι αναλογίες ανάμεσα στα δύο μοντέλα είναι χαρακτηριστικές. Οι ταλαντώσεις υψής, που περιγράφονται σε ένα συγκεκριμένο χώρο συναρτήσεων, θα μπορούσαν να ταυτίζονται με τις μη στατικές ταλαντώσεις της έντασης και των τοπικά αναδυόμενων συχνοτήτων. Στο ίδιο πνεύμα σημαντικό ζήτημα παραμένει ο χωρισμός εικόνων σε διαφορετικές συνιστώσες, που για την ώρα περιλαμβάνουν την αφηρημένη γεωμετρική, την υψή και το θόρυβο.

Για την στοχαστική μοντελοποίηση της φάσης ανακύπτουν ζητήματα που ευνοούν την περαιτέρω έρευνα στο πεδίο. Εξ' αρχής το όλο πρόβλημα είναι ασθενώς καθορισμένο καθώς η μοντελοποίηση ανάγεται σε εκτίμηση μίας μόνο παραμέτρου. Για την ανακατασκευή του σήματος φάσης από τα σήματα συχνοτήτων δεν υπάρχει για την ώρα βέλτιστη διαδικασία. Επιπλέον για την εκτίμηση του συντελεστή, υπεισέρχεται αβεβαιότητα στη μέτρηση του. Αν αυτό γίνεται με εκτίμηση κλίσης, πως θα μετρηθεί, ως προς πια διεύθυνση, ποιο είναι το διάστημα εμπιστοσύνης των δεδομένων που θα χρησιμοποιηθούν κ.α. Αν γίνεται με μέτρηση της φράκταλ επιφάνειας το ζήτημα είναι η περιοχή κλιμάκων που θα χρησιμοποιηθεί, έτσι ώστε η επιφάνεια να είναι φράκταλ, όπως και τι γίνεται στην περίπτωση που το τυχαίο σήμα φάσης είναι μεν $1/f$ θόρυβος αλλά όχι φράκταλ.

Για την εφαρμογή σε δεδομένα εδαφοτομών είναι αναγκαία η ενσωμάτωση όλων των επιμέρους σταδίων σε ένα ολοκληρωμένο σύστημα ανάλυσης και η εύρεση τρόπων συσχετισμού της υπολογιστικής, αυτόματης ανάλυσης με γεωλογικές παραμέτρους, βιολογικές παρατηρήσεις και ανθρώπινες εκτιμήσεις. Διάφορα ζητήματα ανάμεσα στα πεδία της όρασης και της γεωλογίας χρειάζεται να διερευνηθούν όπως ποιες ιδιότητες (τραχύτητα, προσανατολισμός, κλίμακα) της υψής χαρακτηρίζουν τις ξεχωριστές δομές του εδάφους όπως απεικονίζονται στις τομές, ποια εξαγώμενα χαρακτηριστικά τις περιγράφουν καλύτερα, ποια η σημασία διαφορετικής «περιεκτικότητας» και παρουσίας υψής κ.α. Τέλος είναι σημαντική η συστηματική αποτίμηση των μεθόδων στο σύνολο των δεδομένων και η εξαγωγή συγκεντρωτικών αποτελεσμάτων.

Στο θέμα της ανίχνευσης των δομών της φωνής ανοιχτά ερευνητικά ζητήματα υπάρχουν σχετικά με την επέκταση των αντίστοιχων μοντέλων εικόνων. Θα μπορούσε κανείς να θεωρήσει για παράδειγμα το διαχωρισμό της φωνής σε διαφορετικές συνιστώσες, μια από τις οποίες θα περιλαμβάνει μόνο τις μικροδομές διαμόρφωσης και θα αποτελεί το αντίστοιχο της υψής, με εφαρμογές σε θέματα αποθρομβοποίησης, αποκατάσταση χαμένης πληροφορίας, κατάτμησης φωνής και αναγνώρισης ομιλίας. Πάντως, λείπει μέχρι τώρα η φυσική ερμηνεία ενός ακουστικού φαινομένου αντίστοιχου της οπτικής υψής.

7.3 Προς το τέλος

Η προσπάθεια που θα γίνει από δω και στο εξής θα επικεντρωθεί κύρια στο κομμάτι της υψής, με άμεση εφαρμογή την ολοκλήρωση του συστήματος ανάλυσης εδαφοτομών, αλλά παραπλευρικά θα αγγίξει και τις επεκτάσεις στην ανάλυση της φωνής. Συγκεκριμένα θα αναλυθεί διεξοδικότερα και θα αναζητηθούν λύσεις σε ένα από τα τρέχοντα σημαντικά προβλήματα της υψής, αυτό του διαχωρισμού από ένα πολυεπίπεδο μοντέλο υπέρθεσης συνιστωσών, με άμεσες και ποικίλες εφαρμογές σε θέματα αναπαράστασης πληροφορίας, μοντελοποίησης, αποκατάστασης, επεξεργασίας και μαθηματικής περιγραφής. Η εφαρμογή των αναπτυσσόμενων τεχνικών στα δεδομένα εδάφους θα ολοκληρωθεί με στόχο ένα ανωτέρου επιπέδου, ολοκληρωμένο σύστημα όρασης (ανάλυση χαμηλού επιπέδου, εξαγωγή σημαντικών χαρακτηριστικών και απόφαση). Τέλος η έρευνα σε θέματα φωνής και η χρήση κοινών μοντέλων θα συνεχιστεί και θα επεκταθεί, με στόχο την αντιμετώπιση γενικότερων θεμάτων ανίχνευσης και την συμβολή σε συστήματα βελτίωσης και αναγνώρισης φωνής. *Επιγραμματικά:*

7.3.1 Υψή

- Προσέγγιση του διαχωρισμού $u + v$ με μοντέλα διαμορφώσεων, αναλυτικά και υπολογιστικά. Αναζήτηση κοινών αρχών και αντιμετώπιση του προβλήματος με σκοπό την εξαγωγή και την ανάλυση της συνιστώσας v των ταλαντώσεων της υψής, της ομαλής συνιστώσας, και πιθανόν μιας συνιστώσας θορύβου. Μελέτη, περιγραφή και διάκριση των επιμέρους συνιστωσών και συγκρίσεις με γνωστά μεταβολικά σχήματα ελαχιστοποίησης ενεργειακών.
- Συνδυασμός $1/f$ διαδικασιών (και fBm) με μοντέλα $AM-FM$ για μη-κανονικές υψές, με στοχαστική μοντελοποίηση της φάσης. Πιθανή επέκταση του μοντέλου υψής με περισσότερες παραμέτρους, που να ανταποκρίνονται στη φράκταλ φάση. Αναζήτηση λύσεων στα προβλήματα ανακατασκευής, εύρους κλίμακας, αβεβαιότητας στην εκτίμηση κ.α. Εφαρμοσιμότητα σε κατηγορίες υψής.
- Εφαρμογή στην ανάλυση εδαφοτομών με βάση την υψή
 - Επιλογή παραμέτρων και σύνοψη του μοντέλου.
 - Ολοκλήρωση των επιμέρους σταδίων ανάλυσης και καθορισμός του συνόλου των σημαντικών μετρήσεων.
 - Ερμηνεία σε γεωλογικό και βιολογικό επίπεδο και ολοκλήρωση.
 - Επιλογή ή(και) ανάπτυξη κατηγοριοποιητή (classifier) για συνδυασμό και εξαγωγή συμπερασμάτων.

7.3.2 Φωνή

- Επέκταση της ανάλυσης σε γενικευμένα ακουστικά γεγονότα.
- Εισαγωγή των τεχνικών ανίχνευσης και δοκιμές σε συστήματα αναγνώρισης.
- Πιθανές εφαρμογές για αποθορυβοποίηση, ενίσχυση ή και κατάτμηση φωνής.
- Γενίκευση και περαιτέρω αναζήτηση κοινών μοντέλων υψής και φωνής.

7.4 Επίλογος

Η υφή επιφανειών ανήκει στα οπτικά χαρακτηριστικά του φυσικού κόσμου που υπάγονται σε συμβολική αναπαράσταση στα βιολογικά συστήματα όρασης και συμμετέχουν στις περισσότερες αναπαραστάσεις χαμηλού επιπέδου τεχνητών συστημάτων. Κρύβει πληροφορία σχετικά με το είδος, τη σύσταση, τη χωρική τοποθέτηση αλλά και την προέλευση ή εξέλιξη των σωμάτων και των προτύπων που χαρακτηρίζει. Και ενώ έχει μελετηθεί από τη βιολογική, την ψυχοφυσική, τη φιλοσοφική, τη γνωσιακή και τη μαθηματική του πλευρά, εξακολουθεί να υπάρχει αβεβαιότητα γύρω από τον ορισμό και την περιγραφή του οπτικού φαινομένου. Όλα αυτά προσθέτουν βαρύτητα στην έρευνα σχετικά με την επεξεργασία πληροφορίας υφής κάτι που δικαιολογείται και από το σύνολο των εφαρμογών της. Η όποια υπολογιστική προσέγγιση της υφής μπορεί να συμβάλλει σε θέματα κατανόησης και αντίληψης του φυσικού κόσμου.

Ο ήχος και η φωνή ανάγονται επίσης σε ζητήματα αντίληψης και ερμηνείας του φυσικού κόσμου. Ο εντοπισμός ομιλίας ή ήχων με νόημα προηγείται της αναγνώρισης του λόγου, όπως και η απομάκρυνση τυχαίων ή ανεπιθύμητων σημάτων. Θα ήταν εξαιρετική απλούστευση να μιλήσει κανείς για ενιαία θεώρηση μηχανιστικής αντίληψης. Αυτό όμως δεν εμποδίζει την αλληλεπίδραση και την έμπνευση ιδεών και μοντέλων που μπορεί να λειτουργήσουν για παραπλήσια φαινόμενα και στα δύο πεδία.

Και μια διαφορετική αντίληψη του φυσικού κόσμου: Η συνείδηση του περιβάλλοντος για τις νυχτερίδες γίνεται μέσω “συνελιζων εκπεμπόμενων παλμών με την ηχώ τους, και η λαμβανόμενη πληροφορία επιτρέπει ακριβείς διακρίσεις απόστασης, μεγέθους, σχήματος, κίνησης και υφής ανάλογες με αυτές που κάνουμε εμείς με την όραση” [141]. Οι εκπεμπόμενοι παλμοί είναι κατά ένα ποσοστό διαμορφωμένοι στη συχνότητα και αυτό παρέχει πληροφορία υφής και θέσης, ενώ μια σταθερή συνιστώσα κωδικοποιεί την ταχύτητα. Ο ανεπιθύμητος θόρυβος φιλτράρεται κατάλληλα έξω από την ηχώ ενώ ένα σύνολο ακουστικών χαρακτηριστικών χρησιμοποιείται για διάκριση υφής. Ενώ οι άνθρωποι βασιζόμαστε στην εικόνα για να δημιουργήσουμε το όραμα του κόσμου, οι νυχτερίδες βασίζονται στον ήχο.

□

Ορολογία

Attentive (vision, perception): Προσεχτική (όραση, αντίληψη), με σχολαστική επισκόπηση

Complex (cortical) Cells: Σύνθετα Κύτταρα (Νευρώνες)

Contrast (color, intensity): Αντίθεση (χρώματος, έντασης φωτεινότητας)

Curve Evolution: Εξέλιξη καμπυλών

Dominant Component(s): Κυρίαρχη(ες) συνιστώσα(ες)

Early Vision: Αρχική Όραση

Fronts (evolving): Καμπύλες (εξελισσόμενες)

Emergent Frequencies: Εμφανιζόμενες ή αναδυόμενες συχνότητες

Filterbank: Συστοιχία Φίλτρων

Geodesic Active Contours (Regions): Γεωδαιτικά Ενεργά Περιγράμματα (Περιοχές)

Level Sets: Επιπεδοσύνολα

Modulations: Διαμορφώσεις

Multiband analysis: Πολυκαναλική ή Πολυζωνική ανάλυση

Quadrature pair (filters): Ζεύγος (φίλτρων) με κάθετη διαφορά φάσης

Preattentive (vision, perception): Απρόσεκτη (όραση, αντίληψη), χωρίς σχολαστική επισκόπηση

Primal Sketch: Πρωταρχικό Σκίτσο

Primary Visual Cortex: Πρωτεύων Οπτικός Φλοιός

Primitives (image, texture): Στοιχειώδη χαρακτηριστικά (εικόνας, υφής)

(Spatial) Receptive Fields: (Χωρικά) Δεκτικά Πεδία

Region: Περιοχή

Region Competition (segmentation): Κατάτμηση με Ανταγωνισμό Περιοχών

Retina: Αμφιβληστροειδής (Χιτώνας)

Simple (cortical) Cells: Απλά Κύτταρα (Νευρώνες)

Textons: Τεξτόνια, βασικές μονάδες αυθόρμητης αντίληψης υφής

Texture (visual, natural): Υφή (οπτική, φυσική)

Event detection: Ανίχνευση γεγονότων ενδιαφέροντος

Formants: Θεμελιώδεις συντονισμοί (φωνής)

(Speech) Endpoint detection: Ανίχνευση συνόρων (φωνής, λέξεων, φράσεων)

Resonance: Συντονισμός (φωνής)

Voice Activity detection: Ανίχνευση φωνητικής δραστηριότητας

Βιβλιογραφία

- [1] ACTON, S., MUKHERJEE, D., HAVLICEK, J., AND BOVIK, A. Oriented texture completion by AM-FM reaction-diffusion. *IEEE Transactions on Image Processing* 10, 6 (June 2001), 885–896.
- [2] ALOIMONOS, Y. Shape from texture. *Biological Cybernetics* 58, 5 (Feb. 1988), 345–360.
- [3] AUJOL, J.-F., AUBERT, G., BLANC-FIRAUD, L., AND CHAMBOLLE, A. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision* 22, 1 (Jan. 2005), 71–88.
- [4] AUJOL, J.-F., AND CHAMBOLLE, A. Dual norms and image decomposition models. *Intl. Journal of Computer Vision* 63, 1 (June 2005), 85–104.
- [5] BALLARD, D. H., AND BROWN, C. M. *Computer Vision*. Prentice-Hall, 1982.
- [6] BALLESTER, C., AND GONZALEZ, M. Affine invariant texture segmentation and shape from texture by variational methods. *Journal of Mathematical Imaging and Vision* 9, 2 (Sept. 1998), 141–171.
- [7] BARNSLEY, M. F. *Fractals Everywhere*. Academic Press, 1988.
- [8] BECK, J. Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology* 85 (1972), 1–19.
- [9] BECK, J., SUTTER, A., AND IVRY, R. Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing* 37, 2 (1987), 299–325.
- [10] BERGEN, J. R., AND ADELSON, E. H. Early vision and texture perception. *Nature* 333 (May 1988), 363–364.
- [11] BERITELLI, F., CASALE, S., RUGGERI, G., AND SERRANO, S. Performance evaluation and comparison of G.729/AMR/Fuzzy voice activity detectors. *IEEE Signal Processing Letters* 9, 3 (Mar. 2002), 85–88.
- [12] BERTALMIO, M., SAPIRO, G., CASELLES, V., AND BALLESTER, C. Image inpainting. In *SIGGRAPH 00* (2000), ACM Press, pp. 417–424.
- [13] BERTALMIO, M., VESE, L., SAPIRO, G., AND OSHER, S. Image filling-in in a decomposition space. In *Proc. Intl. Conf. on Image Processing* (Barcelona, Spain, Nov. 2003).

- [14] BERTALMIO, M., VESE, L., SAPIRO, G., AND OSHER, S. Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing* 12, 8 (Aug. 2003), 882–889.
- [15] BESAG, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society B* 36 (1974), 192–236.
- [16] BESAG, J. On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society B* 48 (1986), 259–302.
- [17] BEUCHER, S., AND MEYER, F. The morphological approach to segmentation: The watershed transformation. In *Mathematical Morphology in Image Processing*, E.R.Dougherty, Ed. Marcel Dekker, New York, 1993.
- [18] BOVIK, A., CLARK, M., AND GEISLER, W. Texture segmentation using a class of narrowband filters. In *Intl. Conf. on Acoustics Speech and Signal Processing* (Dallas, TX, Apr. 1987).
- [19] BOVIK, A. C. Analysis of multichannel narrow-band filters for image texture segmentation. *IEEE Transactions on Signal Processing* 39, 9 (Sept. 1991), 2025–2043.
- [20] BOVIK, A. C., CLARK, M., AND GEISLER, W. S. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 1 (Jan. 1990), 55–73.
- [21] BOVIK, A. C., GOPAL, N., EMMOTH, T., AND RESTREPO, A. Localized measurement of emergent image frequencies by Gabor wavelets. *IEEE Transactions on Information Theory* 38, 3 (Mar. 1992), 691–712.
- [22] BOVIK, A. C., MARAGOS, P., AND QUATIERI, T. F. AM-FM energy detection and separation in noise using multiband energy operators. *IEEE Transactions on Signal Processing* 41, 12 (Dec. 1993), 3245–3265.
- [23] BRODATZ, P. *Textures: A Photographic Album for Artists and Designers*. Dover, New York, 1966.
- [24] BROX, T., AND WEICKERT, J. A TV flow based local scale measure for texture discrimination. In *European Conf. on Computer Vision* (Prague, Czech Republic, 2004).
- [25] BRUNEAU, P. M. C., DAVIDSON, D. A., AND GRIEVE, I. C. An evaluation of image analysis for measuring changes in void space and excremental features on soil thin sections in an upland grassland soil. *Geoderma* 120 (2004), 165–175.
- [26] BULLOCK, P., FEDOROFF, N., JONGERIUS, A., STOOPS, G., AND TURSI-NA, T. *Handbook for Soil Thin Section Description*. Waine Research Publications, Albrighton, Woverhampton, England, 1985.
- [27] BURT, P. J., AND ADELSON, E. H. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications COM-31*, 4 (1983), 532–540.

- [28] CAELLI, T. *Visual Perception*. Pergamon Press, 1981.
- [29] CAELLI, T., AND JULESZ, B. On perceptual analyzers underlying visual texture discrimination: Part I. *Biological Cybernetics* 28, 3 (Sept. 1978), 167 – 175.
- [30] CASELLES, V., KIMMEL, R., AND SAPIRO, G. Geodesic Active Contours. *Intl. Journal of Computer Vision* 22, 1 (Feb. 1997), 61–79.
- [31] CHAUDHURI, B. B., AND SARKAR, N. Texture segmentation using fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 1 (Jan. 1995), 72–77.
- [32] CHELLAPPA, R., AND CHATTERJEE, S. Classification of textures using gaussian markov random fields. *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-33*, 4 (1985), 959–963.
- [33] CHELLAPPA, R., AND KASHYAP, R. Texture synthesis using 2-D noncausal autoregressive models. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33, 1 (Feb. 1985), 194–203.
- [34] CHEN, S., KELLER, J. M., AND CROWNOVER, R. M. On the calculation of fractal features from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 10 (Oct. 1992), 1087–1090.
- [35] COGGINS, J., AND JAIN, A. K. A spatial filtering approach to texture analysis. *Pattern Recognition Letters* 3 (1985), 195–203.
- [36] COHEN, F. S., AND COOPER, D. B. Simple parallel hierarchical and relaxation algorithms for segmenting noncausal markovian random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9, 2 (Mar. 1987), 195–219.
- [37] COHEN, F. S., FAN, Z., AND PATEL, M. A. Classification of rotated and scaled textured images using Gaussian Markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13, 2 (Feb. 1991), 192–202.
- [38] COHEN, L. On active contour models and balloons. *CVGIP: Image Understanding* 53, 2 (May 1991), 211–218.
- [39] COHEN, L. *Time-Frequency Analysis*. Prentice Hall, 1995.
- [40] CORMACK, L. K. Computational models of early human vision. In *Handbook of Image and Video Processing*, A. Bovik, Ed., 2 ed. 2005, pp. 325–345.
- [41] CROSS, G. C., AND JAIN, A. K. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5, 1 (Jan. 1983), 25–39.
- [42] DATHE, A., EINS, S., NIEMEYER, J., AND GEROLD, G. The surface fractal dimension of the soil-pore interface as measured by image analysis. *Geoderma* 103 (2001), 203–229.

- [43] DAUGMAN, J. E. Complete discrete 2-D Gabor transform by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36, 7 (July 1988), 1169–1179.
- [44] DAUGMAN, J. G. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research* (1980), 847–856.
- [45] DAUGMAN, J. G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A* 2, 7 (July 1985), 1160–1169.
- [46] DAUGMAN, J. G. Image analysis and compact coding by oriented 2-D Gabor primitives. In *SPIE Proc. of Vis. Comm. Image Proc.* (1987), vol. 758, pp. 19–30.
- [47] DERIN, H., AND COLE, W. S. Segmentation of textured images using Gibbs random fields. *Computer Vision, Graphics, and Image Processing* 35, 1 (July 1986), 72–98.
- [48] DERIN, H., AND ELIOT, H. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9, 1 (Jan. 1987), 39–55.
- [49] DEVALOIS, R. L., AND DEVALOIS, K. K. *Spatial Vision*. Oxford University Press, New York, 1988.
- [50] DUDA, R. O., HART, P. E., AND STORK, D. G. *Pattern Classification*, 2 ed. Wiley-Interscience, 2001.
- [51] DUNN, D., AND HIGGINS, W. E. Optimal Gabor filters for texture segmentation. *IEEE Transactions on Image Processing* 4, 7 (July 1995), 947–964.
- [52] DUNN, D., HIGGINS, W. E., AND WAKELEY, J. Texture segmentation using 2-D Gabor elementary functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, 2 (Feb. 1994), 130–149.
- [53] EFROS, A. A., AND LEUNG, T. K. Texture synthesis by non-parametric sampling. In *Intl. Conf. on Computer Vision* (Corfu, Greece, Sept. 1999), pp. 1033–1038.
- [54] ELFADEL, I. M., AND PICARD, R. W. Gibbs random fields, coocurences and texture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, 1 (Jan. 1994), 24–37.
- [55] EVANGELOPOULOS, G., KOKKINOS, I., AND MARAGOS, P. Advances in variational image segmentation using AM-FM models: Regularized demodulation and probabilistic cue integration. In *Proc. 3rd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision* (Beijing, China, 2005).
- [56] EVANGELOPOULOS, G., AND MARAGOS, P. Speech event detection using multiband modulation energy. In *Proc. Interspeech '05 - Eurospeech* (Lisbon, Portugal, 2005).

- [57] EVANGELOPOULOS, G., AND MARAGOS, P. Multiband modulation energy tracking for noisy speech detection. *IEEE Transactions on Speech and Audio Processing* (2006), to appear.
- [58] FOGEL, I., AND SAGI, D. Gabor filters as texture discriminator. *Biological Cybernetics* 61 (1989), 103–113.
- [59] FORSYTH, D. A., AND PONCE, J. *Computer Vision: A Modern Approach*. Prentice-Hall, 2003.
- [60] GABOR, D. Theory of communication. *Journal Inst. of Elec. Eng. London* 93, III (1946), 429–457.
- [61] GEMAN, D., GEMAN, S., GRAFFIGNE, C., AND DONG, P. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 7 (July 1990), 609–628.
- [62] GEMAN, S., AND GEMAN, D. Stochastic relaxation, Gibbs distributions and the Bayesian treatment of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1984), 721–741.
- [63] GIBSON, J. J. *The perception of the visual world*. Houghton Mifflin, Boston, MA, 1950.
- [64] GOUTSIAS, J. Markov random fields: Interacting particle systems for statistical image modeling and analysis. Tech. Rep. 96-01, John Hopkins University, ECE, 1996.
- [65] GRIGORESCU, S. E., PETKOV, N., AND KRUIZINGA, P. Comparison of texture features based on Gabor filters. *IEEE Transactions on Image Processing* 11, 10 (Oct. 2002), 1160–1167.
- [66] HALEY, G. M., AND MANJUNATH, B. S. Rotation-invariant texture classification using a complete space-frequency model. *IEEE Transactions on Image Processing* 8, 2 (Feb. 1999), 255–269.
- [67] HARALICK, R., AND SHAPIRO, L. *Computer and Robot Vision*. Addison Wesley, 1992.
- [68] HARALICK, R. M. Statistical and structural approaches to texture. *Proc. IEEE* 67 (1979), 786–804.
- [69] HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 1 (Apr. 1970), 97–109.
- [70] HAVLICEK, J., AND BOVIK, A. C. Multi-component AM-FM image models and wavelet-based demodulation with component tracking. In *Proc. Intl. Conf. on Image Processing* (Austin, TX, Nov. 1994), pp. I41–I45.
- [71] HAVLICEK, J. P., AND BOVIK, A. C. Image modulation models. In *Handbook of Image and Video Proc.*, A. Bovik, Ed. Academic Press, 2000, pp. 305–316.

- [72] HAVLICEK, J. P., BOVIK, A. C., AND CHEN, D. AM-FM image modeling and Gabor analysis. In *Visual Communication & Image Processing*. Marcel Dekker, New York, 1999.
- [73] HAVLICEK, J. P., HARDING, D. S., AND BOVIK, A. C. The multicomponent AM-FM image representation. *IEEE Transactions on Image Processing* 5, 6 (June 1996), 1094–1100.
- [74] HAVLICEK, J. P., HARDING, D. S., AND BOVIK, A. C. Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models. *IEEE Transactions on Image Processing* 9, 2 (Feb. 2000), 227–242.
- [75] HAVLICEK, J. P., HAVLICEK, J. W., AND BOVIK, A. C. The analytic image. In *Proc. Intl. Conf. on Image Processing* (Santa Barbara, CA, Oct. 1997), pp. 446–449.
- [76] HAVLICEK, J. P., TAY, P. C., AND BOVIK, A. C. AM-FM image models: Fundamental techniques and emerging trends. In *Handbook of Image and Video Processing*, A. Bovik, Ed., 2 ed. 2005, pp. 377–395.
- [77] HEEGER, D. J., AND BERGEN, J. R. Pyramid-based texture analysis/synthesis. In *SIGGRAPH 95* (1995), ACM Press, pp. 229–238.
- [78] HEILER, M., AND SCHNORR, C. Natural image statistics for natural image segmentation. *Intl. Journal of Computer Vision* 63, 1 (June 2005), 5–19.
- [79] HUBEL, D. *Eye, Brain and Vision*. Scientific American Library, 1988.
- [80] HUBEL, D. H., AND WIESEL, T. N. Receptive fields, binocular interaction, and functional architecture in the cat’s visual cortex. *Journal of Physiology (London)* 160, 1 (1962), 106–154.
- [81] HUBEL, D. H., AND WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)* 195 (1968), 215–243.
- [82] ISSERLIS, L. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* 12, 1/2 (Nov. 1918), 134–139.
- [83] JAIN, A., AND HEALEY, G. A multiscale representation including opponent color features for texture recognition. *IEEE Transactions on Image Processing* 7, 1 (Jan. 1998), 124–128.
- [84] JAIN, A. K., AND FARROKHANIA, F. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition* 24, 12 (1991).
- [85] JULESZ, B. Visual pattern discrimination. *IRE Trans. on Information Theory IT-8* (Feb. 1962), 84–92.
- [86] JULESZ, B. Textons, the elements of texture perception and their interactions. *Nature* 290 (Mar. 1981), 91–97.

- [87] JULESZ, B. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics* 41 (1981), 131–181.
- [88] JULESZ, B. *Dialogues on Perception*. MIT Press, 1995.
- [89] JULESZ, B., AND BERGEN, J. R. Textons, the fundamental elements in preattentive vision and perception of textures. *The Bell Syst. Tech Journal* 62, 6 (1983), 1619–1645.
- [90] JULESZ, B., GILBERT, E. N., AND VICTOR, J. D. Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics* 31, 3 (Sept. 1978), 137 – 140.
- [91] KAISER, J. F. On a simple algorithm to calculate the ‘energy’ of a signal. In *Intl. Conf. on Speech and Signal Processing* (Albuquerque, NM, Apr. 1990), pp. 381–384.
- [92] KAISER, J. F. On Teager’s energy algorithm and it’s generalization to continuous signals. In *Proc. 4th IEEE Digital Signal Processing Workshop* (Sept. 1990).
- [93] KAPLAN, L. M. Extended fractal analysis for texture classification and segmentation. *IEEE Transactions on Image Processing* 8, 11 (Nov. 1999), 1572–1585.
- [94] KASS, M., WITKIN, A., AND TERZOPOULOS, D. Snakes: Active Contour Models. In *Intl. Conf. on Computer Vision* (London, UK, 1987).
- [95] KAY, S. M. *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1993.
- [96] KELLER, J. M., CHEN, S., AND CROWNOVER, R. M. Texture description and segmentation through fractal geometry. *Computer Vision, Graphics, and Image Processing* 45, 2 (Feb. 1989), 150–166.
- [97] KICHENASSAMY, S., KUMAR, A., OLVER, P., TANNENBAUM, A., AND YEZZI, A. Gradient flows and geometric active contour models. In *Intl. Conf. on Computer Vision* (Boston, MA, USA, 1995).
- [98] KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Advances in texture analysis: Energy dominant component and multiple hypothesis testing. In *Proc. Intl. Conf. on Image Processing* (Singapore, 2004), vol. 3, pp. 1509–1512.
- [99] KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Modulation-feature based textured image segmentation using curve evolution. In *Proc. Intl. Conf. on Image Processing* (Singapore, 2004), vol. 2, pp. 1204–1207.
- [100] KRUIZINGA, P., AND PETKOV, N. Nonlinear operator for oriented texture. *IEEE Transactions on Image Processing* 8, 10 (Oct. 1999), 1395–1407.
- [101] KUBE, P., AND PENTLAND, A. On the imaging of fractal surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10, 5 (Sept. 1988), 704–707.

- [102] KUMAR, T., ZHOU, P., AND GLASER, D. A. Comparison of human performance with algorithms for estimating fractal dimension of fractional Brownian statistics. *Journal of the Optical Society of America A* 10, 6 (June 1993), 1136–1146.
- [103] KWOK, S. H., AND CONSTANTINIDES, A. G. A fast Recursive Shortest Spanning Tree for image segmentation and edge detection. *IEEE Transactions on Image Processing* 6, 2 (1997), 328–332.
- [104] LAKSHMANAN, S., AND DERIN, H. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 8 (Aug. 1989), 799–813.
- [105] LANDY, M. S. Texture perception. In *Encyclopedia of Neuroscience*, G. Adelman and B. H. Smith, Eds., 3 ed. Elsevier, Amsterdam, 2004.
- [106] LANDY, M. S., AND GRAHAM, N. Visual perception of texture. In *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, Eds. MIT Press, Cambridge, MA, 2004, pp. 1106–1118.
- [107] LEE, T. S., MUMFORD, D., AND YUILLE, A. Texture segmentation by minimizing vector valued energy functionals. In *European Conf. on Computer Vision* (Santa Margherita Ligure, Italy, 1992).
- [108] LEUNG, T., AND MALIK, J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Intl. Journal of Computer Vision* 43, 1 (June 2001), 29–44.
- [109] LIU, J., AND YANG, Y. Multiresolution color image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, 7 (1994), 689–700.
- [110] LUNDAHL, T., OHLEY, W. J., KAY, S. M., AND SIFFERT, R. Fractional Brownian motion: A maximum likelihood estimator and its application to image texture. *IEEE Transactions on Medical Imaging MI-5*, 3 (Sept. 1988), 152–161.
- [111] MALIK, J., , AND ROSENHOLTZ, R. Computing local surface orientation and shape from texture for curved surfaces. *Intl. Journal of Computer Vision* 23, 2 (June 1997), 149–168.
- [112] MALIK, J., BELONGIE, S., LEUNG, T., AND SHI, J. Textons, Contours and Regions: Cue integration in image segmentation. In *Intl. Conf. on Computer Vision* (Corfu, Greece, Sept. 1999).
- [113] MALIK, J., BELONGIE, S., LEUNG, T., AND SHI, J. Contour and texture analysis for image segmentation. *Intl. Journal of Computer Vision* 43, 1 (June 2001), 7–27.
- [114] MALIK, J., AND PERONA, P. Preatentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A* 7, 5 (May 1990), 923–932.

- [115] MALLADI, R., SETHIAN, J., AND VEMURI, B. Shape modeling with front propagation: A Level-Set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 2 (Feb. 1995), 158–175.
- [116] MALLAT, S. G. Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 12, 37 (Dec. 1989), 2091–2110.
- [117] MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 7 (1989), 674–693.
- [118] MANDELBROT, B. B. *The Fractal Geometry of Nature*. Freeman, San Francisco, 1982.
- [119] MANDELBROT, B. B., AND NESS, J. W. V. Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10 (Oct. 1968), 422–437.
- [120] MANJUNATH, B. S., AND CHELLAPPA, R. Unsupervised texture segmentation using Markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13, 5 (May 1991), 478–482.
- [121] MARAGOS, P. Fractal signal analysis using mathematical morphology. In *Advances in Electronics and Electron Physics, vol.88*, P. Hawkes and B. Kazan, Eds. Academic Press, 1994, pp. 199–246.
- [122] MARAGOS, P. *Image Analysis and Computer Vision*. National Technical University of Athens, June 2005.
- [123] MARAGOS, P., AND BOVIK, A. C. Image demodulation using multidimensional energy separation. *Journal of the Optical Society of America A* 12, 9 (Sept. 1995), 1867–1876.
- [124] MARAGOS, P., AND BUTT, M. A. Curve evolution, differential morphology, and distance transforms applied to multiscale and eikonal problems. *Fundamenta Informaticae* 41 (2000), 91 – 129.
- [125] MARAGOS, P., KAISER, J. F., AND QUATIERI, T. F. Energy separation in signal modulations with application to speech analysis. *IEEE Transactions on Signal Processing* 41, 10 (Oct. 1993), 3024–3051.
- [126] MARAGOS, P., KAISER, J. F., AND QUATIERI, T. F. On amplitude and frequency demodulation using energy operators. *IEEE Transactions on Signal Processing* 41, 4 (Apr. 1993), 1532–1550.
- [127] MARAGOS, P., SOFOU, A., STAMOU, G. B., TZOUVARAS, V., STAMOU, G. P., AND PAPANICOLAOU, E. Image analysis of soil micromorphology: Feature extraction, segmentation and quality inference. *J. Applied Signal Processing* 6 (2004), 902–912.
- [128] MARCELJA, S. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America A* 70, 11 (Nov. 1980), 1297–1300.

- [129] MARR, D. *Vision*. W.H. Freeman, San Francisco, CA, 1982.
- [130] MARR, D., AND HILDRETH, E. Theory of edge detection. *Proc. R. Soc. Lond. B*, 207 (1980), 187 – 217.
- [131] MARTIN, D. R., FOWLKES, C. C., AND MALIK, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 5 (May 2004), 530 – 549.
- [132] MARZINZIK, M., AND KOLLMEIER, B. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE Transactions on Speech and Audio Processing* 10, 2 (Feb. 2002), 109–118.
- [133] MCCLOUD, S. *Understanding Comics, the invisible art*. Harper Paperbacks, 1994.
- [134] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., AND TELLER, E. Equations of state calculations by fast computing machine. *Journal of Chemical Physics* 21 (1953), 1087–1091.
- [135] MEYER, F., AND MARAGOS, P. Multiscale morphological segmentations based on watershed, flooding, and eikonal pde. In *Proc. of Scale-Space '99* (1999), Springer-Verlang, pp. 351–362.
- [136] MEYER, F., AND MARAGOS, P. Nonlinear scale-space representation with morphological levelings. *Journal of Visual Communication and Image Representation* 11, 2 (June 2000), 245–265.
- [137] MEYER, Y. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, vol. 22 of *University Lecture Series*. AMS, 2001.
- [138] MITRA, S. K., LI, H., LIN, I.-S., AND YU, T.-H. A new class of nonlinear filters for image enhancement. In *Intl. Conf. on Acoustics Speech and Signal Processing* (Toronto, Canada, May 1991), pp. 2525–2528.
- [139] MOREL, J., AND SOLIMINI, S. *Variational Methods in Image Segmentation*. Birkhauser, Boston, 1995.
- [140] MUMFORD, D., AND SHAH, J. Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure & Appl. Math.* 42, 5 (1989), 577–685.
- [141] NAGEL, T. What is it like to be a bat? *The Philosophical Review* LXXXIII, 4 (Oct. 1974), 435–450.
- [142] NAJMAN, L., AND SCHMITT, M. Watershed of a continuous function. *Signal Processing* 38, 7 (1994), 99–112.
- [143] OSHER, S., AND SETHIAN, J. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics* 79 (1988), 12–49.

- [144] OSHER, S., SOLE, A., AND VESE, L. Image decomposition, image restoration, and texture modeling using total variation minimization and the H^{-1} norm. In *Journal of Multiscale Modeling and Simulation*. SIAM, 2003, pp. 349–370.
- [145] PAPOULIS, A. *Probability, Random Variables and Stochastic Processes*, 3 ed. McGraw-Hill, 1991.
- [146] PARAGIOS, N., AND DERICHE, R. Geodesic Active Contours for supervised texture segmentation. In *Intl. Conf. on Computer Vision & Pattern. Recognition* (1999).
- [147] PARAGIOS, N., AND DERICHE, R. Geodesic Active Regions: A new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation* 13, 1-2 (Mar. 2002), 249–268.
- [148] PARAGIOS, N., AND DERICHE, R. Geodesic Active Regions and Level Set methods for supervised texture segmentation. *Intl. Journal of Computer Vision* 46, 3 (Feb. 2002), 223–247.
- [149] PARAGIOS, N., AND OSHER, S. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer, 2002.
- [150] PEITGEN, H.-O., JURGENS, H., AND SAUPE, D. *Chaos and Fractals*. SpringerVerlag, 2004.
- [151] PEITGEN, H.-O., AND SAUPE, D. *The Science of Fractal Images*. Springer-Verlag, Bremen, 1988.
- [152] PENTLAND, A. P. Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 6 (Nov. 1984), 661–674.
- [153] PERONA, P., AND MALIK, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 7 (Jan. 1990), 629–639.
- [154] PESQUET-POPESCU, B., AND VEHEL, J. Stochastic fractal models for image processing. *IEEE Signal Processing Magazine* 19, 5 (Sept. 2002), 48–62.
- [155] PETKOV, N., AND KRUIZINGA, P. Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells. *Biological Cybernetics* 76, 2 (1997), 83–96.
- [156] PORAT, M., AND ZEEVI, Y. Y. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10, 4 (July 1988), 629–639.
- [157] PORTILLA, J., AND SIMONCELLI, E. A parametric texture model based on joint statistics of complex wavelet coefficients. *Intl. Journal of Computer Vision* 40, 1 (Oct. 2000), 49–70.
- [158] POTAMIANOS, A., AND MARAGOS, P. A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation. *Signal Processing* 37, 1 (May 1994), 95–120.

- [159] PROTZ, R., SHIPITALO, M., MERMUT, A., AND FOX, C. Image-analysis of soils-present and future. *Geoderma* 40, 1-2 (1987), 115–125.
- [160] RABINER, L. R., AND SAMBUR, M. R. An algorithm for determining the endpoints of isolated utterances. *Bell Systems Tech. J.* 54, 2 (Feb. 1975), 297–315.
- [161] RAMIREZ, J., SEGURA, J. C., BENITEZ, C., DE LA TORRE, A., AND RUBIO, A. Efficient voice activity detection algorithms using long-term speech information. *Speech Communication* 42, 3-4 (Apr. 2004), 271–287.
- [162] RAMIREZ, J., SEGURA, J. C., BENITEZ, C., DE LA TORRE, A., AND RUBIO, A. J. A new Kullback-Leibler VAD for speech recognition in noise. *IEEE Signal Processing Letters* 11, 2 (Feb. 2004), 266–269.
- [163] RANDEN, T., AND HUSOY, J. H. Filtering for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 4 (Apr. 1999), 291–310.
- [164] RAY, N., HAVLICEK, J., ACTON, S., AND PATTICHIS, M. Active contour segmentation guided by AM-FM DCA. In *Proc. Intl. Conf. on Image Processing* (Thessaloniki, Greece, Oct. 2001), pp. 78–81.
- [165] REED, T. R., AND DU BUF, J. M. H. A review of recent texture segmentation and feature extraction techniques. *CVGIP: Image Understanding* 57, 3 (May 1993), 359–372.
- [166] ROSENFELD, B. S., AND SAGI, D. Spatial variability as a limiting factor in texture discrimination tasks: Implication for performance asymmetries. *Journal of the Optical Society of America A* 7, 9 (Sept. 1990), 1632–1643.
- [167] ROUSSON, M., BROX, T., AND DERICHE, R. Active unsupervised texture segmentation on a diffusion-based feature space. In *Intl. Conf. on Computer Vision & Pattern. Recognition* (2003).
- [168] RUDIN, L., OSHER, S., AND FATEMI, E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60, 1-4 (Nov. 1992), 259–268.
- [169] SAGIV, C., SOCHEN, N. A., AND ZEEVI, Y. Y. Texture segmentation via a diffusion-segmentation scheme in the Gabor feature space. In *2nd Intl. Workshop on Texture Analysis and Synthesis* (Copenhagen, Denmark, 2002).
- [170] SANDBERG, B., CHAN, T., AND VESE, L. A Level-Set and Gabor-based active contour algorithm for segmenting textured images. Tech. Rep. 02-39, UCLA CAM, 2002.
- [171] SERRA, J. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [172] SETHIAN, J. *Level Set Methods*. Cambridge University Press, 1996.
- [173] SETHIAN, J. A. Fast marching methods. *SIAM Review* 41, 7 (1999).

- [174] SETHIAN, J. A. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.
- [175] SHI, M., AND HEALEY, G. Hyperspectral texture recognition using a multi-scale opponent representation. *IEEE Transactions on Geoscience and Remote Sensing* 41, 5 (2003), 1090–1095.
- [176] SIVAKUMAR, K., AND GOUTSIAS, J. Morphologically constrained GRFs: Applications to texture synthesis and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 2 (Feb. 1999), 99–113.
- [177] SOFOU, A., EVANGELOPOULOS, G., AND MARAGOS, P. Coupled geometric and texture PDE-based segmentation. In *Proc. Intl. Conf. on Image Processing* (Genova, Italy, 2005).
- [178] SOFOU, A., EVANGELOPOULOS, G., AND MARAGOS, P. Soil image segmentation and texture analysis: a computer vision approach. *IEEE Geosci. Remote Sensing Let.* 2, 4 (Oct. 2005), to appear.
- [179] SOFOU, A., AND MARAGOS, P. PDE-based modelling of image segmentation using volumic flooding. In *Proc. Intl. Conf. on Image Processing* (Barcelona, Spain, 2003), vol. 2, pp. 431–434.
- [180] SOHN, J. S., KIM, N. S., AND SUNG, W. A statistical model-based voice activity detection. *IEEE Signal Processing Letters* 6, 1 (Jan. 1999), 1–3.
- [181] SOILLE, P. *Morphological Image Analysis*, 2 ed. Springer-Verlag, 2003.
- [182] STARCK, J.-L., ELAD, M., AND DONOHO, D. L. Image decomposition: Separation of texture from piece-wise smooth content. In *SPIE annual meeting* (San Diego, California, Aug. 2003).
- [183] STARCK, J.-L., ELAD, M., AND DONOHO, D. L. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing* 14, 10 (Oct. 2005), 1570–1582.
- [184] SUPER, B. J., AND BOVIK, A. Shape from texture using local spectral moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 4 (Apr. 1995), 333–343.
- [185] SUPER, B. J., AND BOVIK, A. C. Localized measurement of image fractal dimension using Gabor filters. *Journal of Visual Communication and Image Representation* 2, 2 (June 1991), 114–128.
- [186] TEAGER, H. M., AND TEAGER, S. M. Evidence of nonlinear sound production mechanisms in the vocal tract. In *Speech Production and Speech Modelling*. Kluwer Academic, 1990, pp. 241–261.
- [187] TERRIBILE, F., AND FITZPATRICK, E. The application of multilayer digital image-processing techniques to the description of soil thin-sections. *Geoderma* 55, 1-2 (Oct. 1992), 159–174.

- [188] THERRIEN, C. W. An estimation-theoretic approach to terrain image segmentation. *Computer Vision, Graphics, and Image Processing* 22, 3 (June 1983), 313–326.
- [189] TUCERYAN, M., AND JAIN, A. Texture analysis. In *The Handbook of Pattern Recognition and Computer Vision*, C. Chen, L.F.Pau, and P. Wang, Eds., 2 ed. World Scientific, 1998, pp. 207–248.
- [190] TUCERYAN, M., AND JAIN, A. K. Texture segmentation using Voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 2 (Feb. 1990), 211–216.
- [191] TURING, A. M. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond.* 327, B (1952), 37–72.
- [192] TURNER, M. R. Texture discrimination by Gabor functions. *Biological Cybernetics* 55, 2-3 (Nov. 1986), 71–82.
- [193] VAKMAN, D. On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency. *IEEE Transactions on Signal Processing* 44, 4 (Apr. 1996), 791–797.
- [194] VESE, L., AND CHAN, T. Active contours without edges. *IEEE Transactions on Image Processing* 10, 2 (Feb. 2001), 266–277.
- [195] VESE, L. A., AND OSHER, S. J. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing* 19, 1-3 (2003), 553–572.
- [196] VESE, L. A., AND OSHER, S. J. Image denoising and decomposition with total variation minimization and oscillatory functions. *Journal of Mathematical Imaging and Vision* 20 (2004), 7–18.
- [197] VINCENT, L. Granulometries and opening trees. *Fundamenta Informaticae* 41, 1-2 (Jan. 2000), 57–90.
- [198] VINCENT, L., AND SOILLE, P. Watershed in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13, 6 (June 1991), 583–598.
- [199] VOORHEES, H., AND POGGIO, T. Detecting textons and texture boundaries in natural images. In *Intl. Conf. on Computer Vision* (London, UK, 1987), pp. 250–258.
- [200] VOORHEES, H., AND POGGIO, T. Computing texture boundaries from images. *Nature* 333 (May 1988), 364–367.
- [201] VOSS, R. F. Fractals in nature: from characterisation to simulation. In *The science of fractal images*, H.-O. Peitgen and D. Saupe, Eds. Springer-Verlag, Bremen, 1988, pp. 21–70.
- [202] WANDELL, B. A. *Foundations of Vision*. Sinauer Press, 1995.

- [203] WERTHEIMER, M. Principles of perceptual organization. In *Readings in Perception*, D. C. Beardslee and M. Wertheimer, Eds. van Nostrand, Princeton NJ, 1958, pp. 115–135.
- [204] WINKLER, G., AND LUTZ, G. F. H. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Applications of Mathematics. Springer-Verlag, 2003.
- [205] WITKIN, A., AND KASS, M. Reaction-diffusion textures. In *SIGGRAPH 91* (1991), ACM Press, pp. 299–308.
- [206] WITKIN, A. P. Recovering surface shape and orientation from texture. *Artificial Intelligence* 17, 1-3 (Aug. 1981), 17–45.
- [207] WU, G. D., AND LIN, C. T. Word boundary detection with mel-scale frequency bank in noisy environment. *IEEE Transactions on Speech and Audio Processing* 8, 5 (Sept. 2000), 541–553.
- [208] WU, Y., ZHU, S., AND LIU, X. Equivalence of Julesz ensembles and FRAME models. *Intl. Journal of Computer Vision* 38, 3 (July 2000), 247–265.
- [209] YEZZI, A., TSAI, A., AND WILLSKY, A. A statistical approach to snakes for bimodal and trimodal imagery. In *Intl. Conf. on Computer Vision* (Corfu, Greece, Sept. 1999).
- [210] YING, G., MITCHELL, C., AND JAMIESON, L. Endpoint detection of isolated utterances based on a modified Teager energy measurement. In *Intl. Conf. on Acoustics Speech and Signal Processing* (Minneapolis, MN, Apr. 1993), pp. 732–735.
- [211] ZHU, S., LIU, X., AND WU, Y. Exploring texture ensembles by efficient Markov Chain Monte Carlo: toward a “trichromacy” theory of texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 6 (June 2000), 554 – 569.
- [212] ZHU, S., AND YUILLE, A. Region Competition: Unifying Snakes, Region Growing and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, 9 (Sept. 1996), 884–900.
- [213] ZHU, S. C. Embedding Gestalt laws in Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 11 (Nov. 1999), 1170 – 1187.
- [214] ZHU, S. C., GUO, C. E., WANG, Y., AND XU, Z. What are textons? *Intl. Journal of Computer Vision* 62, 1-2 (Apr. 2005), 121 – 143.
- [215] ZHU, S. C., GUO, C. E., WU, Y., AND WANG, Y. What are textons. In *European Conf. on Computer Vision* (Copenhagen, Denmark, May 2002).
- [216] ZHU, S. C., WU, Y. N., AND MUMFORD, D. Filters, random field and maximum entropy (FRAME): Towards a unified theory for texture modeling. *Intl. Journal of Computer Vision* 27, 2 (Feb. 1998), 1–20.

Κατάλογος δημοσιεύσεων του συγγραφέα

Ανάλυση και Κατάτμηση Υφής

- KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Advances in texture analysis: Energy dominant component and multiple hypothesis testing. In *Proc. IEEE Intl. Conf. on Image Processing* (Singapore, 2004), vol. 3, pp. 1509–1512.
- KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Modulation-feature based textured image segmentation using curve evolution. In *Proc. IEEE Intl. Conf. on Image Processing* (Singapore, 2004), vol. 2, pp. 1204–1207.
- EVANGELOPOULOS, G., KOKKINOS, I., AND MARAGOS, P. Advances in variational image segmentation using AM-FM models: Regularized demodulation and probabilistic cue integration. In *Proc. 3rd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision* (Beijing, China, 2005).

Υφή και Ανάλυση Εδαφοτομών

- SOFOU, A., EVANGELOPOULOS, G., AND MARAGOS, P. Coupled geometric and texture PDE-based segmentation. In *Proc. IEEE Intl. Conf. on Image Processing* (Genova, Italy, 2005).
- SOFOU, A., EVANGELOPOULOS, G., AND MARAGOS, P. Soil image segmentation and texture analysis: a computer vision approach. *IEEE Geosci. Remote Sensing Let.* 2, 4 (Oct. 2005), to appear.

Ανίχνευση και Ανάλυση Φωνής

- EVANGELOPOULOS, G., AND MARAGOS, P. Speech event detection using multiband modulation energy. In *Proc. Interspeech 05 - Eurospeech* (Lisbon, Portugal, 2005).
- EVANGELOPOULOS, G., AND MARAGOS, P. Multiband modulation energy tracking for noisy speech detection. *IEEE Transactions on Speech and Audio Processing* (2006), to appear.

□