



## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ & ΡΟΜΠΟΤΙΚΗΣ

Ανάλυση Τφής και Μικροδομών Εικόνων με  
Μοντέλα Διαμόρφωσης, Ενεργειακές και  
Μεταβολικές Τεχνικές: Ανίχνευση & Διαχωρισμός

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΓΕΩΡΓΙΟΥ ΕΥΑΓΓΕΛΟΠΟΥΛΟΥ

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

Αθήνα, Μάιος 2007





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

# Ανάλυση Υφής και Μικροδομών Εικόνων με Μοντέλα Διαμόρφωσης, Ενεργειακές και Μεταβολικές Τεχνικές: Ανίχνευση & Διαχωρισμός

## ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

TOU

# ΓΕΩΡΓΙΟΥ ΕΥΑΓΓΕΛΟΠΟΥΛΟΥ

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

Συμβουλευτική Επιτροπή: Καθ. Πέτρος Μαραγκός  
Καθ. Γεώργιος Καραγιάννης  
Ου. Καθ. Σπύρος Τζαφέστας

Εγκρίθηκε από την επιταμελή εξεταστική επιτροπή την 4<sup>η</sup> Μαΐου 2007.

...  
Π. Μαραγκός  
Καθηγητής Ε.Μ.Π.

...  
Γ. Καραγιάννης  
Καινοτομής Ε.Μ.Π.

...  
Σ. Τζαφέστας  
Ομ. Καθηγητής Ε.Μ.Π.

Κ. Τζαφέστας  
Λέκτωρας Ε.Μ.Π.

Σ. Κόλλιας  
Καινοτομής Ε.Μ.Π.

...  
Δ. Αργιαλάς  
Καθηγητής Ε.Μ.Π.

...  
Μ. Ζερβάκης  
Καθηγητής, Πολυτεχνείο Κορίνθου

4.06 M/ 2005



ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ  
ΕΡΕΥΝΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ  
ΥΠΟΥΡΓΕΙΟ ΑΝΑΠΤΥΞΗΣ

ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ  
Ευρωπαϊκό Κοινωνικό Ταμείο



Η παρούσα διδακτορική διατριβή πραγματοποιήθηκε στα πλαίσια του προγράμματος ΠΕΝΕΔ-2001, της Γενικής Γραμματείας Έρευνας και Τεχνολογίας. Το πρόγραμμα συγχρηματοδοτείται κατά 75% από την Ευρωπαϊκή Ένωση - Ευρωπαϊκό Κοινωνικό Ταμείο και κατά 25% από το Ελληνικό Δημόσιο.

This Ph.D. thesis was supported by grant PENED-2001 of the Greek Ministry of Development-GSRT. It is co-financed by E.U.-European Social Fund (75%) and National Resources (25%).

.....  
**ΓΕΩΡΓΙΟΣ ΕΥΑΓΓΕΛΟΠΟΥΛΟΣ**  
Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Γεώργιος Ευαγγελόπουλος, 2007.  
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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



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

# Ανάλυση Υφής και Μικροδομών Εικόνων με Μοντέλα Διαμόρφωσης, Ενεργειακές και Μεταβολικές Τεχνικές: Ανίχνευση & Διαχωρισμός

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

TOU

## ΓΕΩΡΓΙΟΥ ΕΥΑΓΓΕΛΟΠΟΥΛΟΥ

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

**Συμβουλευτική Επιτροπή:** Καθ. Πέτρος Μαραγκός  
Καθ. Γεώργιος Καραγιάννης  
Ομ. Καθ. Σπύρος Τζαφέστας

Εγκρίθηκε από την επιταμελή εξεταστική επιτροπή την ..... 2007.

TM Nagy

Π. Μαραγκός  
Καθηγητής Ε.Μ.Π.

  
Γ. Καραγιάννης  
Καθηγητής Ε.Μ.Π.

Γ. Καραγιάννης  
Καθηγητής Ε.Μ.Π.

  
Σ. Τζαφέστας  
Ομ. Καθηγητής Ε.Μ.Π.

  
Σ. Κόλλιας  
Καθηγητής Ε.Μ.Π.

Κ. Τζαφέστας  
Λέχτορας Ε.Μ.Π.

  
M. Ζεφεάκης  
Καθηγητής, Πολυτεχνείο Κρήτης

Αθήνα, Μάιος 2007



## ΠΕΡΙΛΗΨΗ

Στόχος της διατριβής είναι η ανάδειξη και ανάλυση μικροδομών της οπτικής υφής, για την εφαρμογή αποδοτικών μοντέλων, την εξαγωγή περιγραφικών χαρακτηριστικών και την αναπαράσταση εικόνων. Κύριοι δείκτες της είναι τα προβλήματα της μοντελοποίησης και ανάλυσης οπτικής υφής εικόνων σε συστήματα Όρασης Υπολογιστών, με έμφαση στα επιμέρους προβλήματα της ανίχνευσης, κατάτμησης και διαχωρισμού της σε σύνθετες εικόνες. Ερευνήθηκαν και προτείνονται προηγμένες τεχνικές μοντελοποίησης και ανάλυσης σε παράλληλες κατευθύνσεις: α) πολυκαναλικά μοντέλα στενοζωνικών συνιστωσών και χωρικών διαμορφώσεων, β) ενεργειακές μέθοδοι για εξαγωγή χαρακτηριστικών, γ) μεταβολικές τεχνικές αποδόμησης εικόνων και διαχωρισμού των συνιστωσών υφής. Η εφαρμογή των προτεινόμενων μεθόδων επιδεικνύεται σε δεδομένα από μια βάση εικόνων εδαφοτούμων με απώτερο σκοπό την κατηγοριοποίηση και την εκτίμηση της βιολογικής ποιότητας εδαφών καθώς και σε διαφορετικές κατηγορίες φυσικών εικόνων. Το αναπτυγμένο μοντέλο συνιστά την κοινή βάση προσέγγισης του φαινομένου της υφής στις διάφορες εκφάνσεις και εφαρμογές του. Συνολικά προτείνεται ένα ολοκληρωμένο σύστημα ανάλυσης και επεξεργασίας υφής που περιλαμβάνει συνεισφορές και καινοτομίες στα ύφεματα: 1) ανίχνευσης, 2) εξαγωγής χαρακτηριστικών με εφαρμογές σε κατάτμηση και ταξινόμηση υφής και πολυχριτιριακή κατάτμηση εικόνων και 3) διαχωρισμού υφής με αποσύνθεση εικόνων για καλύτερη μοντελοποίηση και ενίσχυση της πληροφορίας υφής. Τα γενικά συμπεράσματα που εξάγονται έχουν να κάνουν με τη φύση του φαινομένου, τρόπους ποσοτικοποίησης και ανάλυσης του αλλά και τη σχέση της υφής με τις υπόλοιπες δομές εικόνων. Παράπλευρα, εξετάζεται η γενίκευση των ιδεών σε μονοδιάστατα σήματα για την ανάλυση των μικροδομών ακουστικών σημάτων και την ανίχνευση φωνής.

## Λέξεις Κλειδιά

Όραση Υπολογιστών, Ανάλυση Εικόνων, Υφή Εικόνων, Υπολογιστικά Μοντέλα Υφής, Ανάλυση Υφής, Κατάτμηση Εικόνων, Εξαγωγή Χαρακτηριστικών, Κατηγοριοποίηση, AM-FM Μοντέλα, Μη-γραμμικά Μοντέλα, Μεταβολικές Τεχνικές, Πολυκαναλική Επεξεργασία, Χωρικά Φίλτρα, Ενεργειακοί Τελεστές, Αποδιαμόρφωση, Αποσύνθεση Εικόνων,  $u + v$  Μοντέλα Εικόνων, Εικόνες Εδαφοτούμων, Ανίχνευση Φωνής.



## **ABSTRACT**

The subject of the thesis is the emergence and analysis of visual texture microstructure for efficient modeling, descriptive feature extraction and image representation. Main objectives are the problems of image texture modeling and analysis in Computer Vision systems, with emphasis on the subproblems of texture detection, segmentation and separation in images. Advanced modeling and analysis method are developed in parallel directions: a) Multiband models of narrowband components and spatial modulations, b) Energy methods for texture feature extraction, c) Variational techniques of image decomposition and texture separation. The proposed methods are applied on a database of digitized soilsection images to quantify and evaluate the biological quality of soils and in different types and collections of natural images. The developed model is the common ground to approach texture in its different forms and applications. In total a complete system for texture processing and analysis is proposed that includes contributions and advances regarding: a) detection, b) feature extraction applied on segmentation and classification and multicue image segmentation, c) texture-from-structure separation through image decomposition for texture enhancement and improved modeling. The general conclusions drawn regard the nature of the phenomenon, ways to quantify and analyze it and the relationship of texture to other image structure. In parallel, the specialization of these ideas to one-dimensional signals is explored for the analysis of acoustic signal microstructures and speech detection.

## **Key Words**

Computer Vision, Image Analysis, Image Texture, Computational Texture Models, Texture Analysis, Image Segmentation, Feature Extraction, Classification, AM-FM Models, Non-linear Models, Variational Methods, Multichannel Processing, Spatial Filters, Energy Operators, Demodulation, Image Decomposition, Soilsection Images, Speech Detection.



# Πρόλογος

*It's not what you look at that matters,  
it's what you see.*

---

Henry David Thoreau

Η ανθρώπινη όραση και ακοή, πολύ περισσότερο από αισθητήριες λειτουργίες αντίληψης και αλληλεπίδρασης με τον κόσμο, είναι εκφράσεις και φορείς της ανθρώπινης νοημοσύνης. Μέσω της όρασης η ανάλυση των ιχνών του φωτός στον αμφιβληστροειδή και η επεξεργασία στον οπτικό φλοιό οδηγεί στην ερμηνεία αντικειμένων, προσώπων και φυσικών σκηνών. Μέσω της ακουστικού συστήματος, μεταβολές της πίεσης του αέρα ερμηνεύονται ως προς την προέλευση, το πλήθος, την πηγή ή το γεγονός που τις δημιουργεί. Οι σημαντικότερες δε από αυτές, οφείλονται σε ταλαντώσεις φωνητικών χορδών και μεταφέρουν ομιλία ή λόγο. Αποτελέσματα σύνθετων, παράλληλων και σε μεγάλο βαθμό ανεξερεύνητων, βιολογικών, νευρικών και γνωσιακών διεργασιών, βασίζονται σε πρώιμη νευρική επεξεργασία των ερεθισμάτων αλλά και προϋποθέτουν ή ενεργοποιούν ανώτερες νοητικές λειτουργίες όπως ο σχεδιασμός, η εκτίμηση, η απόφαση, η μνήμη, η μάθηση. Έτσι, βασιζόμενοι σε σύνθετα συστήματα επεξεργασίας και στον ανθρώπινο νου αποκτάμε αντίληψη του εξωτερικού κόσμου και συνείδηση της ατομικότητας μας. Εξάλλου η νοημοσύνη και η ανέλιξη του ανθρώπου στην εξελικτική αλυσίδα σχετίζεται με το ότι δεν κοιτάμε μόνο, αλλά βλέπουμε κι ερμηνεύουμε, δεν αντιδράμε μόνο σε ήχους αλλά κατανοούμε τον ανθρώπινο λόγο και τον (ανα-) παράγουμε.

Η αντίληψη έχει να κάνει με τη συνειδητοποίηση και την κατανόηση κάποιου είδους αισθητήριας πληροφορίας, για παράδειγμα οπτική, ακουστική, απτική, ομιλίας, ρυθμού κ.ά. Η μελέτη της οπτικής αντίληψης συγκεκριμένα επικεντρώνεται στην ικανότητα ερμηνείας της πληροφορίας οπτικού φωτός που φτάνει στα μάτια. Έτσι, οι πρωταρχικές μελέτες πραγματοποιήθηκαν από την πλευρά της οπτικής και της φυσιολογίας ζεκινώντας από τον αρχαίο κόσμο (Ελλάδα και Περσία), περνώντας στην περίοδο της Αναγέννησης (Leonardo Da Vinci) και καταλήγοντας στη σύγχρονη εποχή με σημαντικές συνεισφορές (Newton, Helmholtz και άλλοι). Ένα συμπέρασμα που προέκυψε πολύ φυσικά παρ' όλα αυτά είναι ότι αυτό που βλέπουμε, το αποτέλεσμα της όρασης δηλαδή, είναι πολύ διαφορετικό από μια «μετάφραση» της εικόνας που σχηματίζεται στον αμφιβληστροειδή. Δε βλέπουμε, κοινώς, με τα μάτια αλλά με το νου.

Οι αισθήσεις μέσω των αισθητήριων σημάτων τροφοδοτούν αρχικά αντιληπτικές ή κατώτερες διεργασίες, όπως η προσοχή ή η αντίληψη χωρίς επισκόπηση/προσπάθεια και τελικά τις ανώτερες ή γνωσιακές διεργασίες του νου. Η ψυχολογία Gestalt (γερμανική λέξη που σημαίνει φόρμα ή πρότυπο) τις δεκαετίες '20 και '30 ασχολήθηκε και διατύπωσε νόμους για την αντίληπτική οργάνωση και της προσοχής (Wertheimer, Kohler, Koffka). Εκεί συναντώνται και οι πρώτες αναφορές για στοιχειώδη «αντιληπτικά» στοιχεία καθώς και για τις ιδιότητες και τις σχέσεις «συνόλων».

Οι προσπάθειες του κλάδου της τεχνητής νοημοσύνης (artificial intelligence) τη δεκαετία του '60 για τη δημιουργία ευφυών μηχανών, ώθησαν θεωρίες για προσομοίωση της αντίληψης σε συστήματα αναγνώρισης ομιλίας και αναγνώρισης αντικειμένων, προσώπων και σκηνών. Η δεύτερη κατεύθυνση μορφοποίησε το επιστημονικό πεδίο της όρασης υπολογιστών (computer vision). Το πεδίο συνοψίζει τις προσπάθειες κατανόησης των βιολογικών μηχανισμών όρα-

σης, μοντελοποίησης νοητικών διεργασιών και απώτερο στόχο τη δημιουργία υπολογιστών με ικανότητες τεχνητής όρασης. Μηχανών που θα μπορούσαν να διούνε και να ερμηνεύσουν μια οπτική σκηνή, μηχανών που θα μπορούσαν να αλληλεπιδράσουν αυτόματα με τον ορατό κόσμο.

Με τη θεμελίωση της υπολογιστικής συλλογιστικής εμφανίζονται τα πρώτα υπολογιστικά μοντέλα όρασης τη δεκαετία του '80 (Marr, Poggio). Η λειτουργία της όρασης αντιμετωπίζεται ως ένα σύστημα επεξεργασίας πληροφορίας σε πολλαπλά στάδια. Ο David Marr στηρίζει τη θεωρία του για την όραση σε τρία επίπεδα ανάλυσης: υπολογιστικό, αλγορίθμικό και επίπεδο ανάπτυξης και υλικού. Η διαδικασία της όρασης προχωράει από την 2-Δ εικόνα στον αμφιβληστροειδή στην 3-Δ περιγραφή του χόσμου σε τρία στάδια: α) 2-Δ πρωταρχικό σκίτσο (primal sketch) (εξαγωγή χαρακτηριστικών περιοχών) β)  $2\frac{1}{2}$ -Δ σκίτσο (υφή, προσανατολισμός, βάθος) α), γ) 3-Δ μοντέλο (οπτικοποίηση της σκηνής ως 3-Δ χάρτη).

Τα πρώτα αυτά υπολογιστικά μοντέλα έθεσαν και τους στόχους της όρασης υπολογιστών. Να προσφέρει μηχανιστικές απαντήσεις σχετικά με τι βρίσκεται που και πως σε μια οπτική σκηνή, δηλαδή το είδος, τη θέση και την κατάσταση των αντικειμένων. Αυτό ανάγεται στη δημιουργία μιας συμβολικής περιγραφής μιας τρισδιάστατης σκηνής από μία ή περισσότερες δυαδικές εικόνες. Οι εικόνες είναι δείγματα φωτεινότητας από τις αντανακλάσεις των φυσικών αντικειμένων όπως αποτυπώνονται σε ένα σύστημα αισθητήρων. Για μια μηχανή το σύστημα αισθητήρων είναι μια ψηφιακή κάμερα και τα αντίστοιχα αισθητήρια σήματα, οι 2-Δ αναπαραστάσεις είναι κάποια κατηγορία ψηφιακών εικόνων.

Η όραση υπολογιστών «γεννήθηκε» από τη συμβολή κυρίως τριών επιστημονικών πεδίων: τεχνητή νοημοσύνη (υπολογιστική νοημοσύνη), επεξεργασία σημάτων (signal processing) (σήματα δύο ή περισσοτέρων μεταβλητών) και αναγνώριση προτύπων (pattern recognition) (μηχανική όραση, μάθηση). Στενά σχετιζόμενα είναι και τα πεδία της ρομποτικής (όραση ρομπότ), νευροεπιστημών (βιολογική όραση, εγκεφαλικές λειτουργίες), γνωσιακής ψυχολογίας (αντίληψη, μοντέλα του νου), της φυσικής (οπτική) και των μαθηματικών (γεωμετρία, στατιστική, βελτιστοποίηση). Καθώς η όραση βασίζεται στις εικόνες, επικαλυπτόμενα ή παράλληλα πεδία είναι και οι διάφορες επιστήμες των εικόνων, δηλαδή η επεξεργασία εικόνας (image processing), η ανάλυση εικόνας (image analysis), τα γραφικά με υπολογιστές (computer graphics) και η απεικόνιση (imaging).

Η ανάλυση και μοντελοποίηση της υφής εικόνων και επιφανειών είναι ένα από τα υποπροβλήματα της όρασης υπολογιστών. Η λέξη υφή προέρχεται από το ρήμα υφαίνω, όπως ακριβώς η αγγλική *texture* που οφείλεται στο λατινικό ρήμα *texere*, σχετίζεται δηλαδή με την ύφανση, τον τρόπο ραφής ή το πλέξιμο υφασμάτων. Ανάλογα η οπτική υφή σχετίζεται με ιδιότητες όπως η αναλυτικότητα, η περιοδικότητα, ο προσανατολισμός, η σχετική οργάνωση ή η τυχαιότητα. Πέρα από τις ιδιότητες, η έννοια της υφής είναι δύσκολο να οριστεί και συνδέεται συνήθως με έννοιες όπως η «λεπτομέρεια», η «μικροδομή» ή τα «μικροπρότυπα» των εικόνων. Η διαμόρφωση των εννοιών γύρω από το φαινόμενο της οπτικής υφής ξεχινάνε πριν ακόμη από τη θεμελίωση της υπολογιστικής συλλογιστικής ενώ οι σχηματισμοί της αποτελούν κεντρικά ζητήματα στις θεωρίες Gestalt, αρχικού σκίτσου και πρώιμης όρασης, δηλαδή μελέτης των πρώιμων σταδίων βιολογικής όρασης.

Το αντικείμενο της παρούσας διατριβής είναι η ανάλυση των μικροδομών και/ή της υφής εικόνων στα πλαίσια των θεμάτων της όρασης υπολογιστών με έμφαση στα επιμέρους προβλήματα της ανίχνευσης, κατάτμησης, ταξινόμησης και εξαγωγής/διαχωρισμού της από τις λοιπές δομές εικόνων. Για το σκοπό αυτό ερευνήθηκαν και αναπτύχθηκαν μοντέλα και τεχνικές σε τρείς παράλληλες κατευθύνσεις: α) μοντελοποίηση με χρήση ενός γενικού μαθηματικού μοντέλου, β) εξαγωγή χαρακτηριστικών και περιγραφή μέσω ενεργειακών μεθόδων, γ) αποσύνθεση εικόνων μέσω μεταβολικών μεθόδων.

Η υφή και η λεπτομέρεια βρίσκονται παντού γύρω μας στον χόσμο που μας περιβάλλει, σ' αυτά που δημιουργούμε αλλά και σ' αυτά που θέλουμε να κατανοήσουμε. Ανήκει στη γκάμα

αφηρημένων αλλά θεμελιακών ιδιοτήτων και ιδεών που συνυπάρχουν και χαρακτηρίζουν τη φύση: «χρώμα», «φόρμα», «ομορφιά», «κίνηση», «λεπτομέρεια». Η λεπτομέρεια απαντάται στο φυσικό κόσμο ως τριβή, αλληλεπίδραση και τραχύτητα, στις αισθήσεις μας (στην όραση, την ακοή, την αφή, τη γεύση), στην αισθητική και στην έκφραση μας (στη μουσική, τη ζωγραφική, στο ύφος του λόγου), στις ιδέες και στον τρόπο που αντιλαμβανόμαστε τα πράγματα ως μεταβλητήτα και διαφορετικότητα.

*Η υφή, η λεπτομέρεια είναι ρεαλισμός, ιδιαιτερότητα και τελικά ομορφιά!*

## Ευχαριστίες

Η παρούσα δουλειά πραγματοποιήθηκε στο Εθνικό Μετσόβιο Πολυτεχνείο, στο εργαστήριο Επικοινωνίας Λόγου, Επεξεργασίας Σημάτων και Όρασης Υπολογιστών (CVSP) από το Φεβρουάριο του 2001 μέχρι το Μάιο του 2007 όποτε και ολοκληρώθηκε με τη μορφή της διδακτορικής διατριβής. Με την ευκαιρία ενός κύκλου που κλείνει έτσι θα ήθελα να ευχαρίστω τους ανθρώπους που με υποστήριξαν ερευνητικά, ηθικά και ψυχολογικά αυτά τα χρόνια και για τους οποίους θεωρώ ότι είναι τόσο σημαντική η ολοκλήρωση αυτής της δουλειάς όσο είναι και για μένα.

Θα ήθελα να ευχαριστήσω καταρχήν τον επιβλέποντα καθηγητή της διατριβής μου κ. Πέτρο Μαραγκό για τις ευκαιρίες, τις δυνατότητες και τις κατευθύνσεις που μου προσέφερε. Τα μαθήματα του, προπτυχιακά αλλά και μεταπτυχιακά ήταν για μένα ξεχωριστά «διανοητικά» ερευθίσματα και με έφεραν σε επαφή με γοητευτικά πεδία όπως η όραση, η μη-γραμμικότητα και το χάος. Η ακαδημαϊκή του αφοσίωση και η «εμμονή»-επιμονή του στην ποιότητα και στα δύσκολα συνεισέφεραν άμεσα στην έρευνα, τα αποτελέσματα και την παρουσίαση αυτής δουλειάς. Η διατριβή αυτή είναι το διαδραστικό αποτέλεσμα πολλών συναντήσεων, συζητήσεων και ιδεών. Τέλος, εκτιμώ ιδιαίτερα τις ευκαιρίες που μου έδωσε να παρουσιάσω ένα μέρος αυτής της δουλειάς σε μεγάλα διευθύνη συνέδρια.

Η επαφή μου με τα μέλη της τριμελούς συμβουλευτικής επιτροπής, τον καθηγητή κ. Γεώργιο Καραγιάννη και τον ομότιμο καθηγητή κ. Σπύρο Τζαφέστα, ξεκινάει από τα πρώτα εξάμηνα των προπτυχιακών μου σπουδών και συνεχίζεται μέσω των μαθημάτων και δραστηριοτήτων του Τομέα Σημάτων, Ελέγχου και Ρομποτικής του τμήματος. Τους ευχαριστώ για την ακαδημαϊκή και ερευνητική εμπειρία που μου μετέδωσαν μέσω συμβουλών και συζητήσεων. Επίσης να ευχαριστήσω τα υπόλοιπα μέλη της επταμελούς εξεταστικής επιτροπής της διατριβής, τους καθηγητές κκ. Στέφανο Κόλλια και Δημήτρη Αργιαλά και τον λέκτορα κ. Κώστα Τζαφέστα από το Ε.Μ.Π., καθώς και τον καθηγητή κ. Μιχάλη Ζερβάκη από το Πολυτεχνείο Κρήτης για τα σχόλια και τις παρατηρήσεις ως προς το περιεχόμενο, τις ιδέες αλλά και τις μελλοντικές κατευθύνσεις αυτής της δουλειάς.

Κατά τη διάρκεια αυτών των χρόνων είχα την τύχη να συνυπάρξω και να συνεργαστώ με πολλούς χαρισματικούς και «ανοιχτούς» ανθρώπους. Ευχαριστώ πολύ τα παιδιά της ομάδας του CVSP, τους φίλους: το Βασίλη Πιτσικάλη, για τη συντροφικότητα, τη συντροφιά τις απελείωτες ώρες στο εργαστήριο, τα παράλληλα πάθη και τις κουβέντες μέσα και έξω από τα πράγματα, τη Νατάσα Σοφού για την άφογη συνεργασία στα θέματα των εδαφοτομών, την «κοριτσίστικη» αντίληψη των πραγμάτων που έφερνε στο εργαστήριο και τη διάθεση να ακούσει, τον Ιάσονα Κόκκινο για τη συνεργασία μας στο θέμα της κατάτμησης, την αφοσίωση και την ευρηματικότητα στη δουλειά του, τους Νάσσο Κατσαμάνη και Γιώργο Παπανδρέου για το «νέο αίμα» σε ιδέες, συζητήσεις, τεχνική υποστήριξη (καθώς και για δύο υπέροχα ταξίδια!), τους Τάσσο Ρούσσο και Σταμάτη Λευκιμμιάτη για το ενδιαφέρον και τον ενθουσιασμό που δείχνει να γνωρίσουν τη δουλειά μου και να μοιραστούν τη δικιά τους και τον «παλιό» Δημήτρη Δημητριάδη για τις συζητήσεις πάνω σε φωνή και μονοδιάστατα μοντέλα διαμόρφωσης. Επίσης, ευχαριστώ την Ελίζα για τη διάθεση και τη φιλολογική επισκόπηση τμημάτων του κειμένου.

Η παρούσα διδακτορική διατριβή πραγματοποιήθηκε στα πλαίσια του προγράμματος ΠΕΝΕΔ-2001, της Γενικής Γραμματείας Έρευνας και Τεχνολογίας. Θέλω να ευχαριστήσω την ομάδα του Τομέα Οικολογίας, Τμήματος Βιολογίας του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης για τη συνεργασία μας στα πλαίσια του προγράμματος και την παροχή των δεδομένων των εδαφοτομών. Ιδιαίτερα δε την κ. Αγγέλα Καραμανλή για τον εξαιρετικό χειρισμό των διαδικαστικών, γραμματειακών και οικονομικών ζητημάτων του προγράμματος.

Να ευχαριστήσω επίσης πολλούς καλούς φίλους από το Πολυτεχνείο, την Αθήνα και τα Γρεβενά. Με κάποιους είχαμε δρόμους αντίθετους, με άλλους πορείες παράλληλες προπτυχιακά, με κάποιους και μεταπτυχιακά. Τους ευχαριστώ για την τάξη και την κανονικοποίηση στο χάος της καθημερινότητάς μου αυτά τα χρόνια, και το νόημα που δώσανε σε πράγματα, ειδικά αυτοί που είναι ακόμη «εδώ»... Ιδιαίτερα ευχαριστώ τον Κανέλλο, το Θέμη, το Χάρη, το Δημήτρη και τη Θάλεια για τις πολλές και σπουδαίες κουβέντες και την αμέριστη συμπαράσταση!

Τέλος, η διατριβή είναι αφιερωμένη στους ανθρώπους που μου δίνουν ήχους, εικόνες και υφή. Τους ευχαριστώ: Τη Σάσα που άντεξε έμμεσα ένα μεγάλο μέρος από τη διαδικασία του διδακτορικού (από την καλή και απ' την ανάποδη!) και με προκαλούσε να ξεπερνάω τον εαυτό μου. Τον αδερφό μου Γιάννη (bro') για την κατανόηση, την ψυχολογική υποστήριξη, τη συμπαράσταση και την πίστη που δείχνει σε μένα. Και ξεχωριστά απ' όλους, τους γονείς μου Χρήστο και Δήμητρα που με μεγάλωσαν να σκέφτομαι «έτσι» και να μη συμβιβάζομαι, που μου έδωσαν δρόμους, ελευθερία επιλογών και κίνητρα και είναι πίσω (ή μπροστά) μου, στηρίζοντας κάθε μου προσπάθεια.

*Eυαγγελόπουλος Γιώργος  
Αθήνα, Μάιος 2007*

# Περιεχόμενα

<b>1 Εισαγωγικά</b>	<b>1</b>
1.1 Αντικείμενο και σκοπός .....	1
1.2 Η υφή στην όραση και οπτική αντίληψη: κίνητρα και σημασία .....	5
1.2.1 Αντιληπτική οργάνωση - Gestalt .....	6
1.2.2 Πρώιμη όραση .....	6
1.2.3 Αρχικό σκίτσο .....	7
1.2.4 Εφαρμογές .....	7
1.2.5 Προσεγγίσεις και μοντέλα .....	9
1.2.5.1 Στατιστικά μοντέλα .....	9
1.2.5.2 Νευροβιολογικές και ψυχοφυσικές μελέτες .....	10
1.2.5.3 Πολυσυχνοτική ανάλυση και χωρικά φίλτρα .....	11
1.2.5.4 Μοντέλα διαμόρφωσης .....	12
1.3 Εννοιολογικά και υπολογιστικά προβλήματα υφής .....	13
1.4 Συνεισφορές επιγραμματικά .....	15
1.5 Διάρθρωση κειμένου .....	16
<b>2 Οπτική Υφή και Μικροδομές Διαμόρφωσης</b>	<b>19</b>
2.1 Ορισμοί, ιδιότητες και μοντέλα .....	19
2.2 Πολυσυχνοτικός διαχωρισμός .....	23
2.2.1 Χωρικές συχνότητες .....	23
2.2.2 Γραμμικά φίλτρα και συναρτήσεις Gabor .....	24
2.2.3 Συστοιχίες Gabor φίλτρων .....	27
2.3 Μοντέλα διαμόρφωσης, ενεργειακοί τελεστές και αποδιαμόρφωση .....	30
2.3.1 Μονοδιάστατος ενεργειακός τελεστής .....	30
2.3.2 Ενεργειακός τελεστής για 2-Δ σήματα .....	31
2.3.3 Επέκταση για μιγαδικά 2-Δ σήματα .....	31
2.3.4 Χωρικά AM-FM και διαχωρισμός ενέργειας .....	32
2.3.5 Διακριτοί τελεστές και αποδιαμόρφωση εικόνων .....	33
2.3.6 Ανάλυση του ενεργειακού τελεστή 2-Δ σε θόρυβο .....	34
2.3.6.1 Στατιστικά .....	34
2.3.6.2 Ζωνοπερατή ενέργεια .....	37
<b>3 Ενέργεια και Ανάλυση Υφής</b>	<b>39</b>
3.1 Διαμορφώσεις υφής .....	39
3.1.1 Ενέργεια συνιστωσών υφής .....	40
3.2 Πολυκαναλική ανάλυση .....	41
3.2.1 Ζωνοπερατή αποδιαμόρφωση .....	42
3.2.2 Μιγαδική αποδιαμόρφωση .....	43
3.2.3 Ανάλυση συνιστωσών κατά κανάλια .....	44
3.2.3.1 Επεξεργασία και ομαλοποίηση .....	46
3.2.3.2 Ανακατασκευή .....	46

3.3	Ανάλυση κυρίαρχων συνιστωσών.....	47
3.4	Κανονικοποιημένη αποδιαμόρφωση - Gabor ESA .....	51
3.5	Ενέργεια διαμόρφωσης και χαρακτηριστικά υφής .....	53
<b>4</b>	<b>Κατάτμηση Υφής</b>	<b>57</b>
4.1	Κατάτμηση με εξέλιξη καμπυλών και κυρίαρχες διαμόρφωσεις.....	57
4.2	Χαρακτηριστικά διαμόρφωσης .....	58
4.2.1	Η επιλογή καναλιού ως στατιστική απόφαση .....	59
4.2.2	Στατιστική ανίχνευση υφής και βελτιωμένα χαρακτηριστικά.....	61
4.3	Μεταβολική κατάτμηση με εξέλιξη καμπυλών .....	62
4.4	AM-FM ανταγωνισμός περιοχών.....	63
4.5	Σταθμισμένη εξέλιξη καμπυλών .....	65
4.5.1	Αναγεννητικά μοντέλα για υφή και ακμές.....	65
4.5.2	Πιθανοτικός συνδυασμός χαρακτηριστικών .....	67
4.5.3	Αποτελέσματα σταθμισμένης κατάτμησης.....	68
4.5.4	Συμπεράσματα .....	70
<b>5</b>	<b>Συνέργεια διαχωρισμού και μοντελοποίησης: Αποσύνθεση εικόνων</b>	<b>71</b>
5.1	Δομή, υφή και μοντέλα .....	71
5.1.1	Κίνητρα και πλεονεκτήματα .....	72
5.1.2	Σύνοψη συνέργειας .....	73
5.2	Αποσύνθεση εικόνων .....	75
5.2.1	Ολική μεταβολή και ΜΔΕ .....	75
5.2.2	Κυματίδια και βάσεις .....	79
5.3	$u + Kv$ : πολλαπλές συνιστώσες υφής .....	81
5.3.1	Μαθηματική ανάλυση .....	83
5.3.1.1	Τύπαρξη .....	83
5.3.1.2	Μοναδικότητα .....	83
5.3.1.3	Σύγκλιση .....	84
5.3.2	Πολλαπλές υποσυνιστώσες υφής .....	85
5.3.3	Θεωρητικές συγχρίσεις .....	86
5.3.4	Αλγόριθμος και διακριτοποίηση .....	87
5.4	Περιορισμός του μοντέλου και ανακατασκευή υφής .....	88
5.4.1	Επιλογή συνιστωσών .....	89
5.4.1.1	Λεπτομερείς κλίμακες και υψησυχνες ζώνες .....	89
5.4.1.2	Στάθμιση πλατών .....	91
5.4.1.3	Κυρίαρχες συνιστώσες χαμηλότερης τάξης .....	95
5.4.2	Τεξτόνια κυρίαρχων διαμορφώσεων .....	96
5.4.3	Παραδείγματα και συγχρίσεις αποδόμησης .....	98
5.5	Μιγαδική αποδόμηση .....	101
5.5.1	Μιγαδική νόρμα ολικής μεταβολής .....	102
5.5.2	Συζευγμένος διαχωρισμός .....	103
5.6	Σύνοψη.....	104
<b>6</b>	<b>Συνέργεια διαχωρισμού και μοντελοποίησης: Ταξινόμηση υφής</b>	<b>107</b>
6.1	Ανάλυση της συνιστώσας υφής .....	107
6.1.1	Διανύσματα χαρακτηριστικών .....	108
6.2	Διάκριση και ομοιότητα τάξεων .....	110
6.2.1	Δοκιμές επιδόσεων στη βάση VisTex .....	111
6.3	Ταξινόμηση υφής.....	114
6.3.1	Πειράματα ταξινόμησης εικόνες (Brodatz) .....	114
6.3.1.1	Μέθοδοι ταξινόμησης .....	116

6.3.1.2	Αποτίμηση χαρακτηριστικών.....	116
6.3.1.3	Δοκιμές ταξινόμησης .....	116
6.4	Συμπεράσματα .....	121
<b>7</b>	<b>Εφαρμογές στην Ανάλυση Εικόνων Εδαφοτομών</b>	<b>123</b>
7.1	Περιγραφή εδαφοτομών και ανάλυση εδάφους .....	123
7.2	Τριγωνική εδαφοτομών, ανάλυση και κατάτμηση .....	124
7.2.1	Μορφολογική κατάτμηση .....	124
7.2.2	Συζευγμένη, πολυχριτιριακή κατάτμηση εδαφοτομών .....	125
7.3	Διαχωρισμός υφής με leveling .....	127
7.4	Αποτελέσματα κατάτμησης και συγκρίσεις .....	129
7.4.1	Βελτίωση με συγχώνευση περιοχών .....	130
7.4.2	Συγκρίσεις .....	132
7.5	Κατηγορίες και διάκριση υφής .....	135
7.6	Συμπεράσματα .....	137
<b>8</b>	<b>Μικροδομές Ακουστικών Σημάτων και Ανίχνευση Φωνής</b>	<b>139</b>
8.1	Φωνή σε θόρυβο .....	139
8.2	Ενεργειακοί τελεστές και πολυζωνικές διαμορφώσεις .....	140
8.3	Κίνητρα και ανίχνευση με στατιστικές υποθέσεις .....	141
8.4	Χαρακτηριστικά διαμόρφωσης .....	143
8.5	Εντοπισμός συνόρων φωνής σε θόρυβο .....	144
8.6	Ανίχνευση φωνητικής δραστηριότητας .....	147
8.6.1	Ανίχνευση ενέργειας διαμόρφωσης .....	148
8.6.2	Αποτελέσματα και συγκρίσεις .....	149
8.7	Συμπεράσματα .....	152
<b>9</b>	<b>Συμπεράσματα και Προεκτάσεις</b>	<b>153</b>
9.1	Συνεισφορές .....	153
9.1.1	Ενέργεια υφής και αποδοτική εκτίμηση παραμέτρων .....	153
9.1.2	Χαρακτηριστικά διαμόρφωσης για μεταβολική κατάτμηση .....	154
9.1.3	Συνέργεια μεταξύ αποδόμησης εικόνων και μοντελοποίησης υφής .....	154
9.1.4	Εφαρμογή στην ανάλυση εδαφοτομών .....	155
9.2	Συμπεράσματα .....	155
9.3	Μελλοντική έρευνα .....	156
9.4	Επίλογος .....	157
<b>Ορολογία</b>		<b>159</b>
<b>Βιβλιογραφία</b>		<b>163</b>
<b>Κατάλογος δημοσιεύσεων</b>		<b>183</b>
<b>Βιογραφικό Σημείωμα</b>		<b>185</b>



# Κατάλογος Σχημάτων

2.1	Παραδείγματα συνθετικής και φυσικής υφής εικόνων .....	22
2.2	Φίλτρα Gabor και ζωνοπερατές συνιστώσες .....	26
2.3	Συστοιχία φίλτρων Gabor .....	29
2.4	Κάλυψη του πεδίου συχνοτήτων από συστοιχίες Gabor .....	29
3.1	Παραδείγματα τοπικά στενοζωνικής υφής εικόνων .....	41
3.2	Αποδιαμόρφωση και ανάλυση υφής ανά κανάλι .....	44
3.3	Πολυκαναλική ανάλυση και συνιστώσες διαμόρφωσης υφής .....	45
3.4	Διάγραμμα ανάλυσης κυρίαρχων συνιστωσών (DCA) .....	48
3.5	DCA ανάλυση συνθετικού 2-Δ AM-FM .....	49
3.6	Σύγκριση DCA πλάτους και ενέργειας .....	50
3.7	Κανονικοποιημένη και διακριτή αποδιαμόρφωση συνθετικού 2-Δ AM-FM .....	52
3.8	Ανάλυση και σύνθεση μέσω DCA (πλάτους και ενέργειας) .....	54
3.9	Ενέργεια διαμόρφωσης και ανίχνευση υφής .....	55
4.1	Χαρακτηριστικά διαμόρφωσης (κυρίαρχων συνιστωσών) για κατάτμηση .....	59
4.2	Δοκιμές υποθέσεων για ανίχνευση υφής ανά κλίμακα .....	61
4.3	Χαρακτηριστικά διαμόρφωσης βελτιωμένα με στατιστική ανίχνευση υφής (I) (εικόνα ‘barb’) .....	62
4.4	Χαρακτηριστικά διαμόρφωσης βελτιωμένα με στατιστική ανίχνευση υφής (II) (εικόνα ‘mandrill’) .....	62
4.5	Πληροφορία υφής (κλίμακα, μέγεθος, προσανατολισμός) και κατάτμηση μέσω DCA χαρακτηριστικών διαμόρφωσης .....	64
4.6	Κατάτμηση υφής με χαρακτηριστικά διαμόρφωσης και ανταγωνισμό περιοχών (εικόνες ‘house’, ‘rice’, ‘rocks’) .....	65
4.7	Πιθανότητες (class posteriors) αναγεννητικών μοντέλων υφής και ακμών .....	67
4.8	Αποτελέσματα κατάτμησης με διαφορετικά χαρακτηριστικά και μεταβολικά σχήματα .....	68
4.9	Κατάτμηση εικόνων με DCA χαρακτηριστικά και σταθμισμένη εξέλιξη καμπυλών .....	69
5.1	Συνιστώσες εικόνας και ανακατασκευή υφής (“The Starry Night”) .....	73
5.2	Συνιστώσες cartoon και υφής εικόνας (“Kiss”) .....	76
5.3	Τυποσυνιστώσες υφής και $u + Ku$ αποσύνθεση (εικόνα ‘seaflower’)	84
5.4	Ανακατασκευή υφής με βάση την κλίμακα ή τον προσανατολισμό της περιγραφής	90
5.5	Φασματικές ζώνες κλίμακας και αποκρίσεις ισοδύναμων φίλτρων .....	91
5.6	Σύγκριση ζωνοπερατού πλάτους εικόνας και συνιστωσών cartoon και υφής .....	93
5.7	Επίδραση της στάθμισης πλάτους στην ανακατασκευή πληροφορίας υφής .....	94
5.8	Ανακατασκευή υφής και επίλογή συνιστωσών .....	97
5.9	Τεξτόνια (δομικά στοιχεία υφής) από τις αποκρίσεις των φίλτρων .....	98
5.10	Αποδόμηση συνθετικής εικόνας με ακμές, υφή και θόρυβο .....	99

5.11 Αποδόμηση $u + Kv$ και εναλλακτικοί όροι ανακατασκευής υφής .....	100
5.12 Μιγαδική αποδόμηση αναλυτικής εικόνας (Brodatz-D97, 'tree') .....	104
5.13 Αποδόμηση έγχρωμων εικόνων (εικόνα 'memorial') .....	105
5.14 Βελτιωμένη ανίχνευση ακμών από τη συνιστώσα cartoon .....	105
5.15 Αποδόμηση για αποκατάσταση εικόνων .....	105
 6.1 Στιγμιότυπα από τις τάξεις υφής της βάσης VisTex .....	111
6.2 Κατανομές του χριτηρίου Fisher στη βάση VisTex .....	112
6.3 Διαγράμματα μετρήσεων διαχωρισμότητας και ομοιότητας διαφορετικών τάξεων υφής .....	113
6.4 Ταξινόμηση (k-means ομαδοποίηση) σε εικόνες με 2 τάξεις υφής Brodatz .....	117
6.5 Ταξινόμηση (k-means ομαδοποίηση) σε εικόνες με 5 τάξεις υφής .....	118
6.6 Μεταβολή των σφαλμάτων ταξινόμησης με τις παραμέτρους της αποδόμησης ..	120
6.7 Σφάλματα ταξινόμησης (%) για τις εικόνες με 2 κατηγορίες υφής .....	121
 7.1 Πολυκλιμακωτά leveling cartoon για $u + v$ αποδόμηση εικόνων εδαφοτομών ..	127
7.2 Πολυκλιμακωτά leveling cartoon και $u + v$ αποσύνθεση εικόνας .....	128
7.3 Εκτίμηση και ανάλυση της υφής εικόνων εδαφοτομών .....	130
7.4 Κατάτμηση εδαφοτομών με συνδυασμό χριτηρίων .....	131
7.5 Βελτίωση των αποτελεσμάτων κατάτμησης με ομαδοποίηση και ένωση περιοχών ..	131
7.6 Σύγκριση μεθόδων κατάτμησης (I) .....	134
7.7 Κατάτμηση με συνδυασμό χριτηρίων (διαφορετικοί σημαδευτές) .....	134
7.8 Κατάτμηση και σύγκριση μεθόδων (II) .....	135
7.9 Λεπτομέρειες από τυπικές μικροϋφές χώματος .....	136
 8.1 Σχηματικό διάγραμμα εξαγωγής 1-Δ κυρίαρχων χαρακτηριστικών διαμόρφωσης ..	144
8.2 Ανάλυση και αποδιαμόρφωση φωνής ανά κανάλι .....	144
8.3 Χαρακτηριστικά ανάλυσης και ανίχνευσης φωνής .....	145
8.4 Παραδείγματα εντοπισμού συνόρων φωνής με χαρακτηριστικά διαμόρφωσης ..	146
8.5 Καμπύλες ανίχνευσης - ανοχής λάθους για εντοπισμό ορίων φωνής .....	147
8.6 Παραδείγματα ανίχνευσης φωνητικής δραστηριότητας .....	149
8.7 Καμπύλες ROC για αποτίμηση απόδοσης της ανίχνευσης φωνής .....	150
8.8 Νόρμες ολικών σφαλμάτων ανίχνευσης φωνητικών γεγονότων .....	152

# Κατάλογος Πινάκων

5.1	Ποσοτική σύγκριση σχημάτων αποδόμησης σε συνθετική εικόνα .....	100
6.1	Μέση σχετική αλλαγή (%) του κριτηρίου Fisher με εξαγωγή χαρακτηριστικών από τη συνιστώσα υφής .....	114
6.2	Μέση σχετική αλλαγή (%) της σχετικής εντροπίας (KL-distance) με εξαγωγή χαρακτηριστικών από τη συνιστώσα υφής .....	115
6.3	Αποτίμηση χαρακτηριστικών για εικόνες 2 και 5 ειδών υφής (μωσαϊκά Brodatz)	117
6.4	Σφάλματα ταξινόμησης (μέσες τιμές %) σε εικόνες με 2 κατηγορίες υφής .....	119
6.5	Σφάλματα ταξινόμησης (μέσες τιμές %) και παράμετροι αποδόμησης σε εικόνες με 2 κατηγορίες υφής .....	120
7.1	Ποσοτική σύγκριση μεθόδων κατάτμησης εικόνων εδαφοτομών (I) .....	133
7.2	Ποσοτική σύγκριση μεθόδων κατάτμησης εικόνων εδαφοτομών (II) .....	134
7.3	Διαχωρισμότητα τάξεων υφής εδαφοτομών με Gabor ενεργειακά χαρακτηριστικά .....	137
8.1	Ορθή ανίχνευση ορίων φωνής (%) στη βάση NTIMIT .....	146
8.2	Ακρίβεια ανίχνευσης φωνητικής δραστηριότητας, Aurora 2 .....	151
8.3	Ακρίβεια ανίχνευσης φωνητικής δραστηριότητας, Aurora 3 .....	151



# Κατάλογος Συντμήσεων

<b>1-Δ</b>	:	μονοδιάστατο
<b>2-Δ</b>	:	δισδιάστατο
<b>3-Δ</b>	:	τρισδιάστατο
<b>MΔE</b>	:	Μερική Διαφορική Εξίσωση
<b>AM</b>	:	Amplitude Modulation (διαμόρφωση πλάτους)
<b>ADCA</b>	:	Amplitude-based Dominant Components Analysis
<b>BV</b>	:	Bounded Variation (φραγμένη μεταβολή)
<b>CCA</b>	:	Channelized Components Analysis (ανάλυση συνιστωσών κατά κανάλια)
<b>CESA</b>	:	Continuous Energy Separation Algorithm (συνεχής αλγόριθμος διαχωρισμού ενέργειας)
<b>DCA</b>	:	Dominant Components Analysis (ανάλυση κυρίαρχων συνιστωσών)
<b>DESA</b>	:	Discrete Energy Separation Algorithm (διακριτός αλγόριθμος διαχωρισμού ενέργειας)
<b>EDCA</b>	:	(Teager) Energy-based Dominant Components Analysis
<b>ESA</b>	:	Energy Separation Algorithm (αλγόριθμος διαχωρισμού ενέργειας)
<b>EO</b>	:	Energy Operator (ενεργειακός τελεστής Teager-Kaiser)
<b>FC</b>	:	Fisher Criterion (χριτήριο Fisher)
<b>FM</b>	:	Frequency Modulation (διαμόρφωση συχνότητας)
<b>GAC</b>	:	Geodesic Active Contours (γεωδαιτικά ενεργά περιγράμματα)
<b>GAR</b>	:	Geodesic Active Regions (γεωδαιτικές ενεργές περιοχές)
<b>GLRT</b>	:	Generalized Likelihood Ratio Test (δοκιμή γενικευμένου λόγου πιθανοφάνειας)
<b>KLD</b>	:	Kullback-Leibler Distance (απόσταση Kullback-Leibler)
<b>MAT</b>	:	Maximum Average Teager Energy (μέγιστη μέση Teager ενέργεια)
<b>MDA</b>	:	Multiband Demodulation Analysis (πολυκαναλική ανάλυση αποδιαμόρφωσης)
<b>MDL</b>	:	Minimum Description Length
<b>MLE</b>	:	Maximum Likelihood Estimates (εκτιμήσεις μέγιστης πιθανοφάνειας)
<b>MRF</b>	:	Markov Random Fields (Μαρκοβιανά τυχαία πεδία)
<b>PCA</b>	:	Principal Components Analysis (ανάλυση πρωτευουσών συνιστωσών)
<b>PDEs</b>	:	Partial Differential Equations (μερικές διαφορικές εξισώσεις)
<b>PDF</b>	:	Probability Density Function (συνάρτηση πυκνότητας πιθανότητας)
<b>RC</b>	:	Region Competition
<b>ROF</b>	:	Rudin-Osher-Fatemi (μεταβολικό σχήμα αποθορυβοποίησης [258])
<b>RSST</b>	:	Recursive Shortest Spanning Tree (split and merge)
<b>TV</b>	:	Total Variation (ολική μεταβολή)
<b>VO</b>	:	Vese-Osher (μεταβολικό σχήμα αποδόμησης [299])
<b>WSS</b>	:	Wide Sense Stationary (στατική με την ευρεία έννοια)

<b>ZR</b>	: average Zero-crossings Rate (μέσος ρυθμός μεταβάσεων από το μηδέν)
<b>LTED</b>	: Long-term Multiband Teager Energy Divergence (μακροπρόθεσμη απόχλιση πολυκαναλικής Teager ενέργειας)
<b>LTSD</b>	: Long Term Spectral Divergence (μακροπρόθεσμη φασματική απόχλιση)
<b>mAA</b>	: mean Average Amplitude (μέσο απόλυτο πλάτος)
<b>mSA</b>	: mean Squared Amplitude (μέσο τετραγωνικό πλάτος)
<b>MTE</b>	: mean Multiband Teager Energy (πολυκαναλική μέση Teager ενέργεια)
<b>MTED</b>	: Multiband Teager Energy Divergence (απόχλιση Πολυκαναλικής Teager ενέργειας)
<b>MIA</b>	: mean Multiband Instant Amplitude (πολυκαναλικό μέσο στιγμιαίο πλάτος)
<b>MIF</b>	: mean Multiband Instant Frequency (πολυκαναλική μέση στιγμιαία συχνότητα)
<b>PTE</b>	: Power-spectrum based Teager Energy
<b>ROC</b>	: Receiver Operating Characteristics
<b>SNR</b>	: Signal to Noise Ratio (σηματοθορυβικός λόγος)
<b>STE</b>	: Short-time Teager Energy
<b>TKEO</b>	: Teager-Kaiser Energy Operator
<b>VAD</b>	: Voice Activity Detection (ανίχνευση δραστηριότητας φωνής)

# Κεφάλαιο 1

## Εισαγωγικά

Η οπτική υφή ανήκει στην κατηγορία των μικροδομών μιας εικόνας που για τις επιστήμες των εικόνων είναι η χωρική πληροφορία που εντοπίζεται στις μικρές κλίμακες και συνυπάρχει με τα μακροσκοπικά και γεωμετρικά χαρακτηριστικά των αντικειμένων ή των σκηνών. Αντίστοιχα, στις επιστήμες του ήχου η μικροδομή σχετίζεται με τη μεταβολή σε μικρές χρονικές κλίμακες. Η επεξεργασία της υφής ανήκει στις διεργασίες χαμηλού επιπέδου της όρασης και η ανάλυση της σχετίζεται με την κατανόηση της οπτικής αντίληψης από βιολογική, γνωσιακή και υπολογιστική σκοπιά. Οι εφαρμογές της καλύπτουν συναφή γνωστικά πεδία όπως η όραση υπολογιστών, η αναγνώριση προτύπων, η επεξεργασία εικόνων ή τα γραφικά υπολογιστών. Στο κεφάλαιο αυτό επιχειρείται μια εισαγωγή στα θέματα της υφής εικόνων με γνώμονα τη σημασία του φαινομένου, τα κίνητρα για την ανάπτυξη ενός γενικού πλαισίου περιγραφής αλλά και τα προβλήματα που σχετίζονται με την υπολογιστική αντικειμενική προσέγγιση του θέματος. Όπου χρίνεται σκοπόμο ανάγονται αντίστοιχες σε θέματα ανάλυσης ήχων και φωνής.

### 1.1 Αντικείμενο και σκοπός

Η οπτική πληροφορία που περιλαμβάνεται στις εικόνες του φυσικού κόσμου, όπως αυτές αναπαριστώνται σε βιολογικά αλλά και τεχνητά/υπολογιστικά συστήματα όρασης είναι δομημένη ιεραρχικά σε πολλαπλά επίπεδα οργάνωσης και κλίμακες. Οι εικόνες απαρτίζονται από δομές που αντιστοιχούνε σε διαφορετικά χαρακτηριστικά της αποτυπωμένης σκηνής και ορίζονται με βάση την ερμηνεία απεικόνισης. Η ιεραρχία «φύλλα-δέντρο-δάσος» είναι ένα τέτοιο παράδειγμα. Από αδρομερές σε λεπτομερές επίπεδο η οργάνωση συνίσταται από φόρμες (αντικείμενα, σχήματα, περιγράμματα), οι οποίες ανάγονται σε μακροδομές, στη συνέχεια σε επαναλαμβανόμενα πρότυπα και τελικά σε μικροδομές [105].

Η υφή (texture), αντικειμένων ή επιφανειών ανήκει στην κατηγορία των μικροδομών μιας εικόνας που συνυπάρχει με τις δομές μεγάλης κλίμακας και τα γεωμετρικά χαρακτηριστικά της σκηνής. Η φυσική (natural) υφή των αντικειμένων είναι η ιδιότητα που γίνεται άμεσα αντιληπτή με την αίσθηση της αφής και σχετίζεται με τις φυσικές ιδιότητες των επιφανειών όπως η ομαλότητα ή η τραχύτητα. Η οπτική υφή (visual texture) είναι μια αντιληπτική ιδιότητα περιοχών εικόνων. Σε φυσικές εικόνες αυτή μπορεί να θεωρηθεί το γνωσιακό αντίστοιχο της

φυσικής υφής καθώς συχνά σχετίζεται με χαρακτηριστικά ενδεικτικά του είδους, της σύσταση, της δομής ή της κατάστασης των αντικειμένων. Ο συνηθέστερος όρος υφή εικόνας (image texture) τυπικά περιλαμβάνει δύο κατηγορίες: υφή και πρότυπα (patterns) [24].

Η υφή είναι ανάλογη με αντιληπτικές ιδιότητες όπως η φωτεινότητα, το χρώμα, το βάθος και η κίνηση, χωρίς να είναι πλεονάζουσα ή μη συνεπής σε μια περιγραφή ενός αντικειμένου. Δηλαδή ο καυχορισμός των υπόλοιπων ιδιοτήτων του δεν περιορίζει τις ιδιότητες της υφής του. Ένα σύστημα όρασης, βιολογικό ή τεχνητό, χρησιμοποιεί υπολογιστικά μοντέλα των ιδιοτήτων αυτών σε ένα ευρύτερο πλαίσιο ανάλυσης της σύνθετης οπτικής πληροφορίας για την παραγωγή συμβόλων από 2-Δ εικόνες και την εξαγωγή συμπερασμάτων για τις 3-Δ σκηνές. Έτσι το ενδιαφέρον για την αντίληψη της υφής έχει ρίζες τόσο θεωρητικές όσο και εφαρμοσμένες.

Σε επίπεδο επιστήμης των εικόνων, η υφή αντιστοιχεί στις μεταβολές της έντασης φωτεινότητας στις μικρότερες κλίμακες, δηλαδή με ταχύτερο ρυθμό χωρικής μεταβολής ή υψηλότερη συχνότητα. Οι μεταβολές αυτές μπορεί να είναι τοπικές ή μη, τυχαίες, περιοδικές ή σχεδόν περιοδικές. Παράλληλα αντιστοιχεί σε πρότυπα που καυχορίζουν και προσδίδουν τη λεπτομέρεια των απεικονιζόμενων σκηνών και αντικειμένων. Η υφή μπορεί προσεγγιστεί είτε ως ιδιότητα των περιοχών μιας εικόνας, όπου οι περιοχές καυχορίζονται από την ομοιογένεια τους σε σχέση με τη γειτονιά με σαφή ή νοητά σύνορα και περιγράμματα, είτε ως συλλογή από στοιχειώδη πρότυπα χωρικά οργανωμένα σε κατηγορίες. Όπως προαναφέρθηκε, σε πληροφοριακό επίπεδο κωδικοποιεί πληροφορία σχετικά με τη φυσική υφή, τη δομική πολυπλοκότητα, τις παραμορφώσεις και τις μεταβολές των επιφανειών, τις ιδιότητες και τη σύσταση των υλικών καθώς και τις φυσικές ιδιότητες της σκηνής και του μέσου απεικόνισης.

Το αντικείμενο της παρούσας δουλειάς είναι ένα από τα υποπροβλήματα της όρασης υπολογιστών, η μοντελοποίηση και η ανάλυση υπολογιστικά της υφής εικόνων. Η ανάπτυξη και χρήση δηλαδή υπολογιστικών μοντέλων για την αναπαράσταση, την περιγραφή και το χειρισμό της πληροφορίας της υφής. Στόχος είναι η ανάδειξη και ανάλυση της μικροδομής ψηφιακών εικόνων για την εφαρμογή αποδοτικών μοντέλων υφής, την εξαγωγή περιγραφικών χαρακτηριστικών και την ανάπτυξη εφαρμογών μηχανικής όρασης και αυτόματης ανάλυσης εικόνων.

Η μοντελοποίηση αναφέρεται στο σχεδιασμό και την ανάπτυξη μαθηματικών, στατιστικών ή υπολογιστικών μοντέλων που ύστατα περικλείουν γνώση για την υφή σε μια χρησιμοποιήσιμη μορφή. Οι παράμετροι του μοντέλου καυχορίζουν τις διαφορετικές καταστάσεις του φαινομένου. Ένα μοντέλο μπορεί να χρησιμοποιηθεί είτε για μετρήσεις από ένα σύνολο δεδομένων (ανάλυση υφής) είτε αντίστροφα για προσομοίωση πιθανών καταστάσεων (σύνθεση υφής). Η ανάλυση και η σύνθεση περιγραφών υφής καυχορίζεται εκτιμώντας και ρυθμίζοντας αντίστοιχα τις τιμές των παραμέτρων/μεταβλητών του μοντέλου. Η μοντελοποίηση προϋποθέτει κατανόηση των σημαντικών χαρακτηριστικών της υφής αντιληπτικά αλλά και μαθηματικά, γεγονός που οδηγεί στην ανάγκη ορισμού και καυχορισμού του φαινομένου. Τι είναι και τι δεν είναι υφή σε μια εικόνα και τι ρόλο παίζουν οι διάφορες παράμετροι της αναπαράστασης, όπως για παράδειγμα η κλίμακα.

Πρωταρχικός στόχος είναι ο καυχορισμός μια συνολικής θεώρησης για την υφή σε ένα

θεωρητικό και υπολογιστικό πλαίσιο. Αυτό συμπεριλαμβάνει το λειτουργικό ορισμό του φαινομένου καθώς και την επιλογή ή την ανάπτυξη του κατάλληλου, περιγραφικού μοντέλου που θα αναδεικνύει την ύπαρξη ή απουσία υφής και τις διαφορετικές εκφάνσεις της και θα περιγράψει τις ιδιότητες της. 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 ή γενικεύσεις μοντέλων και ιδεών από διαφορετικές προσεγγίσεις.

Ιδιαίτερη έμφαση δίνεται σε επιμέρους ζητήματα που αποτελούν υποπροβλήματα αλλά και σημαντικές συνιστώσες μιας ολοκληρωμένης θεώρησης για την υφή:

**Ανίχνευση υφής (texture detection):** καθορισμός της ύπαρξης και της «έντασης» της ανά περιοχές σε μια εικόνα.

**Ανάλυση, εξαγωγή χαρακτηριστικών υφής (texture feature extraction):** κατασκευή διανυσμάτων χαρακτηριστικών για ποσοτικοποίηση και περιγραφή της υφής σε διεργασίες υψηλού επιπέδου.

**Κατάτμηση υφής (texture segmentation/segregation):** ομαδοποίηση περιοχών με ομογενή υφή, χωρισμός μιας εικόνας σε ομογενείς περιοχές με βάση την υφή τους, περιγράμματα με βάση την υφή.

**Κατάτμηση εικόνων (image segmentation):** χωρισμός εικόνας σε ομογενείς περιοχές με βάση πολλαπλά κριτήρια/ιδιότητες.

**Διαχωρισμός υφής - από - δομή (texture-from-structure separation):** αποδόμηση εικόνας στις συνθετικές τις συνιστώσες (γεωμετρικά χαρακτηριστικά, μακροδομές, υφή, όροι βοσκής).

**Ταξινόμηση, διάκριση υφής (texture classification, discrimination):** ομαδοποίηση εικόνων υφής σε κατηγορίες ανάλογα με τα χαρακτηριστικά τους.

Η ανίχνευση σχετίζεται με την ανάδειξη εκείνων των χαρακτηριστικών και την ανάπτυξη μετρήσεων ή τελεστών που να παρέχουν ενδείξεις για την ύπαρξη υφής σε μια περιοχή της εικόνας. Αυτό ισοδύναμεί με την κατασκευή ενός «ενδείκτη», με συνεχείς ή δυαδικές τιμές, για τις διαφορετικές περιοχές της εικόνας που θα χαρακτηρίζει τοπικά την ένταση και την εστίαση της υφής. Η ανάλυση αναφέρεται στη χρήση του μοντέλου για αναπαράσταση και στην ανάπτυξη αλγορίθμων για την εκτίμηση των παραμέτρων του. Τα εξαγώμενα χαρακτηριστικά ή περιγραφές της υφής μπορούν να χρησιμοποιηθούν σε πιο σύνθετα προβλήματα και εφαρμογές όπως είναι η κατάτμηση και η ταξινόμηση εικόνων, η αναγνώριση αντικειμένων και η εξαγωγή 3-Δ γεωμετρίας. Η ταξινόμηση εικόνων υφής αφορά την ομαδοποίηση εικόνων σε κατηγορίες ανάλογα με το είδος της υφής τους, όπως αυτό εκφράζεται από διανύσματα χαρακτηριστικών. Αν και η ετυμολογία του όρου διαχωρισμός υφής ταιριάζει περισσότερο με τον αγγλικό segregation, ο όρος θα χρησιμοποιείται εδώ για να δηλώσει την εξαγωγή της υφής

από τις λοιπές μακροδομές της εικόνας μέσω αποδόμησης στις συνθετικές της συνιστώσες. Η εκτιμώμενη με αυτό τον τρόπο συνιστώσα υφής μπορεί να θεωρηθεί μια ενισχυμένη εκδοχή της υφής της αρχικής εικόνας.

Το μοντέλο που χρησιμοποιείται για τη συνολική θεώρηση της υφής απαρτίζεται από τοπικά στενοζωνικά, χωρικά διαμορφωμένα κατά πλάτος και συχνότητα ημιτονοειδή σε υπέρθεση. Η ανάλυση στηρίζεται σε επεξεργασία της εικόνας σε πολλάπλα κανάλια συχνοτήτων όπως αυτά καθορίζονται από γραμμικά φίλτρα, επιλεκτικά ως προς τη συχνότητα και τον χωρικό προσανατολισμό τους. Τα χωρικά αυτά σήματα διαμόρφωσης, οι συνιστώσες του μοντέλου, χαρακτηρίζουν τη μικροδομή της εικόνας ενώ ταυτόχρονα η επιλογή χωροσυγχονοτικά των περισσότερο σημαντικών οδηγούν σε θεμελιώδη δομικά στοιχεία ή πρότυπα της υφής.

Για την προσέγγιση των επιμέρους ζητημάτων σε αυτό το κοινό πλαίσιο μοντελοποίησης, αξιοποιούνται τεχνικές και εργαλεία με κίνητρα διαισθητικά και μαθηματικά:

- α) Πολυκαναλικά μοντέλα στενοζωνικών συνιστωσών και χωρικών διαμορφώσεων.
- β) Ενεργειακές μέθοδοι για εξαγωγή χαρακτηριστικών.
- γ) Μεταβολικές μέθοδοι αποδόμησης εικόνων και διαχωρισμού της συνιστώσας υφής.

Το ερευνητικό πλαίσιο περιλαμβάνει γενικευμένους μη-γραμμικούς, 2-Δ ενεργειακούς τελεστές, μοντελοποίηση σύνθετων σημάτων/εικόνων ως υπέρθεση διαμορφωμένων ημιτονοειδών, πολυκαναλική ανάλυση μέσω ζωνοπερατών φίλτρων σε πολλαπλές συχνοτικές ζώνες, εκτίμηση με ενεργειακά κριτήρια των «ισχυρών» συνιστωσών που είναι περισσότερο σημαντικές για την υφή, λογισμό μεταβολών και μεθόδους συνεχούς βελτιστοποίησης για την αντίστροφη εκτίμηση των συνθετικών συστατικών μιας εικόνας.

Τα προτεινόμενα μοντέλα και μέθοδοι, καθώς και οι εφαρμογές στις οποίες εμπλέκονται, επιδεικνύονται σε διαφορετικές κατηγορίες εικόνων: φυσικές, δηλαδή εικόνες του φυσικού κόσμου, εικόνες υφής, δηλαδή λεπτομέρειες επιφανειών και υλικών, εικαστικές, όπως φωτογραφίες ή πίνακες ζωγραφικής, συνθετικές, δηλαδή μαθηματικές συναρτήσεις που χρησιμεύουν ως διδακτικά παραδείγματα και εικόνες μικροσκοπίου, όπως είναι οι εικόνες εδαφοτομών σε κλίμακες της τάξεως  $10^{-1} \mu\text{m}$ . Ποσοτικά και συστηματικά αποτελέσματα εξάγονται από βάσεις φυσικών εικόνων (Berkeley image database), εικόνες υφής (Brodatz textures) καθώς και μια συλλογή εικόνων εδαφοτομών με απώτερο σκοπό την ανάλυση, κατηγοριοποίηση και εκτίμηση της βιολογικής ποιότητας εδαφών.

Το κύριο τμήμα της παρούσας έρευνας έχει να κάνει με τις μικροδομές σε συστήματα εικόνων και όρασης. Σε μια επιμέρους κατεύθυνση οι ιδέες και τα μοντέλα εξειδικεύονται σε ακουστικά σήματα φωνής. Συγκεκριμένα το αντίστοιχο μοντέλο 1-Δ χρονικών διαμορφώσεων και οι ενεργειακές μέθοδοι, εφαρμόζονται για την ανάλυση και την ανίχνευση της μικροδομής ακουστικών σημάτων. Η έννοια της μικροδομής για τέτοια, χρονικά μεταβαλλόμενα σήματα ταυτίζεται με ιδιότητες όπως η σχεδόν στιγμιαία μεταβολή της συχνότητας ή του πλάτους και μπορεί να συνδέεται με ιδιότητες όπως το ηχόχρωμα ή η «ακουστική» υφή. Η υποκείμενη

εφαρμογή εδώ είναι η ανίχνευση φωνητικών γεγονότων (speech event detection). Συγκεκριμένα δίνεται έμφαση στην εξαγωγή χαρακτηριστικών που διακρίνουν την παρουσία φωνητικών γεγονότων, δηλαδή την ύπαρξη ομιλίας σε ακουστικά σήματα.

## 1.2 Η υφή στην όραση και οπτική αντίληψη: κίνητρα και σημασία

Η εικόνα που σχηματίζεται στον αμφιβληστροειδή (retinal image) αναλύεται σε διαφορετικά αντικείμενα, οι ιδιότητες των οποίων χρησιμοποιούνται στο διαχωρισμό πρώτου-πλάνου φόντου αλλά και των αντικειμένων μεταξύ τους. Η οπτική (visual) ή αμφιβληστροειδική (retinal) υφή προκύπτει από τις μεταβολές των χαρακτηριστικών ή των ιδιοτήτων μίας μεμονωμένης επιφάνεια, όπως η ανακλαστικότητα, το χρώμα ή ο 3-Δ προσανατολισμός. Αυτές οι μεταβολές μπορεί να οφείλονται στην πολυπλοκότητα της δομής των επιφανειών, την τραχύτητα και τις παραμορφώσεις (απτική (haptic) υφή [215]), τις ιδιότητες υλικών ή ουσιών και τη σύσταση (χρωστική (pigment) υφή [106]) και τις φυσικές ιδιότητες της σκηνής και του μέσου απεικόνισης. Καταλήγουν δε σε μια εικόνα με υφή στον αμφιβληστροειδή ή αντίστοιχα σε κάποιο άλλο τεχνητό σύστημα απεικόνισης. Για φυσικές εικόνες η υφή συνυπάρχει με ιδιότητες όπως το χρώμα, τη γεωμετρία της σκηνής ή την κίνηση, γεωμετρικές μακροδομές όπως σχήματα ή περιγράμματα αντικειμένων, ή ενσωματώνεται σε αδρομερείς σχηματισμούς που οφείλονται στις συνθήκες φωτισμού-σκίασης και τις ομαλές μεταβολές εμβαδού ή όγκου των αντικειμένων.

Παρ' όλα αυτά το φαινόμενο διαθέτει μια εγγενή δυσκολία ορισμού, λόγω της τεράστιας ποικιλομορφίας των προτύπων της υφής στον κόσμο και στις εικόνες του, η οποία ενισχύεται από την εξάρτηση της υφής από ένα πλήθος «υποχειμενικών» παραμέτρων όπως είναι η χλίμακα, ή η εγγύτητα, δηλαδή ο ορισμός της χωρικής γειτονιάς). Για την ανάπτυξη των πρώτων συστημάτων και εφαρμογών μηχανικής όρασης, χρησιμοποιήθηκαν ευριστικοί ορισμοί, βασισμένοι στις εκάστοτε εφαρμογές [118], ενώ οι σύγχρονες προσπάθειες επικεντρώνονται σε μαθηματικούς ορισμούς και στον καθορισμό ενιαίων πλαισίων ανάλυσης [314, 318, 325].

Η σημασία της υφής για την οπτική αντίληψη (visual perception) και την όραση υποστηρίζεται αφενός από ψυχοφυσικές (psychophysical) ενδείξεις, δηλαδή από μελέτες της ανθρώπινης συμπεριφοράς, σχετικά με την αντίληψη και το ρόλο της υφής σε αυτή και αφετέρου από τις άμεσες και έμμεσες εφαρμογές της σε θέματα μηχανικής ή βιολογικής όρασης [197, 242, 287]. Οι θεωρίες για την υφή είναι αντιροσωπευτικές της εξέλιξης των θεωριών βιολογικής και μηχανικής όρασης. Αυτό συμβαίνει αφενός λόγω τη συμμετοχής της υφής στα βασικά προβλήματα κι αφετέρου λόγω του ότι οι μέθοδοι αναπαράστασης και επεξεργασίας είναι αντιροσωπευτικές ενός υπερσυνόλου της βιβλιογραφίας σε θέματα πρώιμης όρασης. Είναι ενδεικτικό ότι η μελέτη του φαινομένου της υφής αποτελεί ένα ερευνητικό πεδίο στο οποίο απαντώνται διαφορετικά γνωστικά αντικείμενα όπως η όραση υπολογιστών, η ανάλυση και επεξεργασία εικόνων, η αναγνώριση προτύπων, η ψυχολογία και οι γνωσιακές επιστήμες, οι νευροεπιστήμες,

η στατιστική φυσική κ.ά.

Αφενός η επεξεργασία της υφής ανήκει στις διεργασίες που συνιθέτουν την οπτική αντίληψη. Καθώς η αντίληψη της υφής διαφέρει κατά κάποιο τρόπο από τη γενικότερη οπτική αντίληψη [24], παρουσιάζει ένα ιδιαίτερο ενδιαφέρον για την κατανόηση των οπτικών διεργασιών. Έτσι, σε ένα φιλοσοφικό-θεωρητικό επίπεδο έχει σημασία η κατανόηση των επιμέρους διαδικασιών που συμβάλλουν στην αντίληψη της υφής, για την κατανόηση των διεργασιών όρασης, οπτικής αντίληψης και τελικά συνείδησης. Οι συζητήσεις και η διαμόρφωση εννοιών σχετικά με το φαινόμενο της υφής ξεκινάνε πριν ακόμη τη θεμελίωση της υπολογιστικής συλλογιστικής [105], ενώ σχηματισμοί ή πρότυπα υφής (textural patterns) είναι κεντρικά ζητήματα στην ψυχολογία Gestalt [309], το πρωταρχικό σκίτσο (primal sketch) της όρασης χαμηλού επιπέδου του D. Marr [205], την αυθόρυμη όραση (preattentive vision) του B. Julesz [146] και τις θεωρίες πρώιμης όρασης (early vision) [25, 26, 145, 188].

### 1.2.1 Αντιληπτική οργάνωση - Gestalt

Μελετώντας την αντιληπτική οργάνωση (perceptual grouping) και την προσοχή (attention), οι εκπρόσωποι της ψυχολογίας Gestalt [150, 161, 309] διατύπωσαν νόμους για τις βασικές αρχές σύμφωνα με τις οποίες μεμονωμένα σύμβολα ή οντότητες (tokens) ομαδοποιούνται οπτικά. Ανάμεσα σε αυτές τις αρχές είναι η εγγύτητα, η ομοιότητα, η συμμετρία, η συνέχεια, η κλειστότητα, η συγγραμμικότητα που απαντώνται και σε περιγραφές ή ορισμούς υφής από τα δομικά της στοιχεία. Η τάση να αντιμετωπίζονται στοιχεία με παρόμοια χαρακτηριστικά σαν να ανήκουν σε ένα σχεδόν ομογενές σύνολο, υπονοεί ότι παρόμοιοι μηχανισμοί ομαδοποίησης μπορεί να εμπλέκονται και στην κατάτμηση ή το διαχωρισμό υφής [24, 252]. Επιπλέον, παρόμοια ομαδοποίηση με αυτή που προβλέπεται από τις θεωρίες Gestalt βρέθηκε ότι επιτυγχάνουν αναπαραστάσεις χώρου/χωρικών συχνοτήτων για την υφή [135, 252].

### 1.2.2 Πρώιμη όραση

Η υφή εμφανίζεται ως ερμηνεία ή αντικείμενο, σε ένα σύνολο από ψυχοφυσικές μελέτες στην προσπάθεια να αναζητηθούν οι διεργασίες οπτικής επεξεργασίας που ευθύνονται για ομαδοποίηση και την αναγνώριση αντικειμένων. Ο Beck υπέθεσε και υποστήριξε ότι το ανθρώπινο σύστημα όρασης δεν αναζητά πρότυπα ή σύνθετες φόρμες υφής αλλά στατιστικά όπως το μήκος, ο προσανατολισμός κ.ά. [20, 21]. Βασίζεται σε τοπικά χαρακτηριστικά (local features) και αναλύει τα ελάχιστα στοιχεία υφής (textural elements) που είναι “υποπρότυπα που συμβαίνουν επαναλαμβανόμενα μέσα στη συνολική υφή.” Ο Julesz διέκρινε τον τρόπο λειτουργίας της οπτικής αντίληψης σε δύο διαφορετικά μοτίβα, αυθόρυμη/χωρίς προσοχή (preattentive) και με προσοχή/με σχολαστική επισκόπηση (attentive), βασιζόμενος σε πειράματα και παρατηρήσεις σχετικά με αναγνώριση υφής από παρατηρητές [142, 143, 146]. Προχώρησε στη μελέτη συνθετικών υφών μικροπροτύπων [48, 147] και το σχηματισμό υποθέσεων σχετικά με τα απαραίτητα αντιληπτικά χαρακτηριστικά υφής, στη διατύπωση ερωτήσεων σχετικά με τη σημασία του φαινομένου σε εικόνες με κοινές άλλες ιδιότητες (χρώμα, αντίθεση, φωτεινότητα)

και στη διαμόρφωση της θεωρίας των τεξτονίων (textons) [143] ως ερμηνεία της αυθόρυμητης διάκρισης υφής (preattentive segregation). Τα τεξτόνια (blobs, terminators, crossings, color) θεωρήθηκαν οι βασικές μονάδες της αυθόρυμητης αντίληψης υφής [146]. Ανάλογα πειράματα γύρω από τη διάκριση υφής οδήγησαν επίσης σε περιγραφές με τοπικά χαρακτηριστικά που εξετάστηκαν ως προς την πυκνότητα, τη συγκέντρωση ή το μέγεθος τους [25, 303, 304]. Παράλληλα περιγραφές με γραμμικά φίλτρα βασίστηκαν σε διαπιστώσεις όπως ότι η αντίληψη συγκεκριμένων προτύπων εξαρτάται πρωτίστως από ανάλυση χωρικών συχνοτήτων και ότι ο αυθόρυμητος διαχωρισμός περιοχών συσχετίζεται με το συχνοτικό τους περιεχόμενο [22, 254].

Παράλληλες, πρώιμες νευροφυσιολογικές μελέτες έδωσαν ενδείξεις σχετικά με τους μηχανισμούς αντίληψης υφής [146]. Ερευνήθηκε η ύπαρξη αναλυτών στο οπτικό σύστημα για την ανίχνευση τεξτονίων [46], ενώ διαπιστώθηκε ότι νευρικά κύτταρα στον οπτικό φλοιό συμπεριφέρονται δεκτικά σε οπτικά χαρακτηριστικά που κατά σύμβαση σημαίνουν υφή [130].

### 1.2.3 Αρχικό σκίτσο

Στο πρώτο ολοκληρωμένο υπολογιστικό μοντέλο αρχικής όρασης του Marr [205], οι θεωρίες της υφής από την ψυχολογία Gestalt και τη θεωρία των τεξτονίων επεκτείνονται στο αρχικό ή πρωταρχικό σκίτσο (primal sketch). Το αρχικό σκίτσο προτάθηκε ως ένα πρώτο επίπεδο εσωτερικής αναπαράστασης της οπτικής πληροφορίας με όρους «αρχέγονων», στοιχειωδών χαρακτηριστικών μιας εικόνας (image primitives), που σχετίζονται με απλές γεωμετρικές δομές, όπως ακμές, γωνίες, ελλείψεις κ.ά. (elongated blobs, bars, edges, terminators). Ανιχευτές τέτοιων χαρακτηριστικών για εξαγωγή χαρακτηριστικών θεωρήθηκαν ότι καταλαμβάνουν τα πρώτα στάδια της όρασης χαμηλού επιπέδου [206]. Το 2-Δ αρχικό σκίτσο μετατρέπεται σε ένα ενδιάμεσο επίπεδο στο  $2\frac{1}{2}$ -Δ σκίτσο όπου η υφή συμμετέχει στην αναπαράσταση επιφανειών μέσω της αναπαράστασης των ασυνεχειών και τοπικού προσανατολισμού.

### 1.2.4 Εφαρμογές

Η επεξεργασία της υφής ανήκει στις διεργασίες χαμηλού επιπέδου (low-level processing) της όρασης. Η ανάλυση της θεωρείται ένα από τα υποπροβλήματα της υπολογιστικής όρασης και εντάσσεται, είτε άμεσα, σε εφαρμογές σχετικές και προσανατολισμένες στην υφή είτε έμμεσα ως χαρακτηριστικό ενός πολυχριτιριακού συστήματος αναγνώρισης και ερμηνείας. Η υφή συμμετέχει και στα τρία κύρια προβλήματα της όρασης υπολογιστών, δηλαδή κατάτμηση, αναγνώριση και ανακατασκευή 3-Δ δομής.

Ως μεμονωμένο χαρακτηριστικό, η υφή εικόνων μπορεί να χρησιμοποιηθεί σε ένα εύρος εφαρμογών:

- Τα χαρακτηριστικά μιας μοναδικής υφής μπορεί να χρησιμοποιηθούν για να αναγνωριστεί το υλικό, η ουσία ή η δομή μιας επιφάνειας, όπως π.χ. ξύλο, άμμος, μέταλλο κ.ά. (texture appearance) [67, 178, 294].
- Οι ιδιαιτερότητες στα χαρακτηριστικά ενός συνόλου εικόνων υφής ή επιφανειών, επιτρέπουν την ταξινόμηση τους και την κατηγοριοποίηση με βάση ένα σύνολο κανόνων

(texture classification) [47, 296].

3. Ανίχνευση και εντοπισμός ακμών, βασικό στάδιο του προβλήματος πρώτου πλάνου-φόντου (figure-ground), που χαρακτηρίζονται από αλλαγές σε χαρακτηριστικά υφής [152, 304] μόνο (ή σε συνδυασμό με μεταβολές σε άλλες ιδιότητες όπως η ένταση φωτεινότητας, το χρώμα ή η κίνηση [208]).
4. Η διάκριση/διαχωρισμός υφής (texture segregation) αποτελεί σημαντική συνιστώσα για την αναγνώριση αντικειμένων [291].
5. Ιδιότητες της υφής σε μια εικόνα μπορούν να χρησιμοποιηθούν για την εξαγωγή συμπερασμάτων της 3-Δ διάταξης αντικειμένων, του βάθους, του σχήματος [4, 16, 276] ή του προσανατολισμού [184, 312] (shape, orientation from texture).
6. Κατάτμηση με βάση την υφή (texture segmentation) [177, 187, 232, 253] αλλά και στα πλαίσια σύνθετης κατάτμησης εικόνων με πολλαπλά κριτήρια (multi-cue image segmentation) [185, 186].

Παράλληλα η περιγραφή και η εξαγωγή χαρακτηριστικών υφής, ως στοιχείο φυσικών σκηνών, περιλαμβάνεται σε ένα εύρος εφαρμογών επεξεργασίας εικόνων και όρασης υπολογιστών όπως η αναπαράσταση, κωδικοποίηση και μετάδοση ψηφιακών εικόνων, η αναζήτηση και ανάκτηση εικόνων (image retrieval) με ταίριασμα χαρακτηριστικών [23, 50, 80, 81, 195, 267], η αναγνώριση αντικειμένων, η μελέτη της κίνησης, η ανακατασκευή 3-Δ γεωμετρίας και η δημιουργία 3-Δ μοντέλων από 2-Δ εικόνες ή πίνακες ζωγραφικής [119]. Σε πρακτικά ή εμπορικά συστήματα η υφή εικόνων χρησιμοποιείται για την αυτόματη επίβλεψη υλικών και διεργασιών σε βιομηχανίες (υφάσματα, χαρτιά, κεραμικά, χρώματα κ.ά.), την ανάλυση βιοϊατρικών εικόνων (διάγνωση με βάση υφή και χρώμα), στις γεωεπιστήμες (ποιότητα εδαφών), την τηλεπισκόπηση και τις περιβαλλοντολογικές μελέτες, τη νανοηλεκτρονική (αναγνώριση και παρασκευή χυκλωμάτων), την επιστήμη των υλικών.

Για την δημιουργία ρεαλιστικών ή εικονικών επιφανειών, η ανάλυση και σύνθεση υφής απαντάται σε πεδία όπως γραφικά και σχεδιασμός με υπολογιστές [17, 89, 90, 128, 311], ψηφιακές τέχνες (animation, multimedia, video art), εφαρμογές εικονικής πραγματικότητας (virtual worlds). Συγχρονες τεχνικές του υπολογιστικής φωτογραφίας (computational photography) επεξεργάζονται την υφή για τη δημιουργία, την περιγραφή, την επεξεργασία ή την τροποποίηση ψηφιακών εικόνων (image abstraction, tone mapping, dynamic-range imaging) [14, 75, 88]. Η «συμπλήρωση» ή αντικατάσταση χαμένης πληροφορίας και η επιδιόρθωση φύλαμένων εικόνων, (image/texture inpainting) εξαρτώνται από τις στατιστικές ιδιότητες της υφής [28, 90] ενώ πρόσφατες τεχνικές υιοθετούν  $u+n$  μοντέλα διαχωρισμού της υφής από τη γεωμετρία της εικόνας [2, 29, 55, 91]. Ένα παράδειγμα αποτελεί η αποκατάσταση ή ανάκτηση τοιχογραφιών με προβλήματα λόγω παλαιότητας και φυλοράς, π.χ. από αρχαιολογικες ανασκαφές.

### 1.2.5 Προσεγγίσεις και μοντέλα

Η αναπαράσταση υφής εικόνων μέσω επιλογής των κατάλληλων μαθηματικών μοντέλων περιγραφής ανάγεται στην περιγραφή των μεταβολών των τιμών της εικόνας που γίνονται αντιληπτές σε μικρές κλίμακες. Το πλαίσιο περιγραφής ορίζεται από κάποιες γενικώς παραδεκτές αρχές σχετικά με την υφή [174, 287]: α) Η υφή είναι ιδιότητα περιοχών σε μια εικόνα και ορίζεται τοπικά σε χωρικές «γειτονιές», β) ορίζεται στατιστικά και περιλαμβάνει κατανομές γκρίζων ή έγχρωμων επιπέδων φωτεινότητας, γ) σε διαφορετικές κλίμακες ή επίπεδα ευχρίνειας (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. Οι αρχές αυτές καθόρισαν και κατεύθυναν τα κυριότερα μοντέλα αναπαράστασης και περιγραφής της υφής εικόνων.

#### 1.2.5.1 Στατιστικά μοντέλα

Θεωρώντας ότι η οπτική υφή είναι μια στατιστική δομή, η επιλογή στατιστικών μεθόδων προέκυψε λογικά και διαισθητικά στις πρώτες προσεγγίσεις και μοντέλα.

#### Χωρικά στατιστικά και κατανομές

Το πρώτο μοντέλο για διάκριση προτύπων υφής προτάθηκε από τον B. Julesz [142] και στηρίχθηκε σε χωρικά στατιστικά  $k$ -τάξης όπου η τάξη  $k$  είναι μεγαλύτερη του 2 (higher-order statistics). Η τάξη καθορίζει τον αριθμό των σημείων της εικόνας που σχηματίζουν ομάδες δηλαδή βασικές μονάδες για σύγκριση. Εξάλλου, η εικασία ότι το ανθρώπινο σύστημα εξάγει πληροφορία από στατιστικά μεγαλύτερα της δεύτερης τάξης ( $k \geq 2$ ), αποτυπώθηκε σε μια σειρά χαρακτηριστικών «συνεμφάνισης» (co-occurrence matrices) όπως η εντροπία, η ενέργεια, η συσχέτιση, η ομοιογένεια [118]. Μια άλλη κατηγορία περιγραφών περιλαμβάνει τις εκτιμήσεις κατανομών και στατιστικών όπως ροπές (moments), περιθώρια ιστογράμματα (marginal histograms) και από-κοινού ιστογράμματα (joint histograms) στις εξόδους γραμμικών φύλτρων [117, 128, 186, 244, 318]. Μεταξύ άλλων, στο στατιστικό πλαίσιο εντάχθηκαν διάφοροι ορισμοί, τύποι, μεγέθη και προσανατολισμοί τεξτονίων των οποίων οι κατανομές και η συχνότητα εμφάνισης προσδιορίζονται μέσα από διαδικασίες εκμάθησης [178, 185, 294, 322].

#### Τυχαία πεδία

Η θεωρία των Μαρκοβιανών τυχαίων πεδίων (Markov random fields) θεμελιώθηκε από τον Besag πάνω σε ιδέες από τη στατιστική μηχανική [30, 31]. Τέτοια πεδία χαρακτηρίζουν πρότυπα υφής μέσω της χωρικής στατιστικής των στοιχείων του πεπερασμένου πλέγματος της εικόνας, ορίζοντας σχέσεις εξάρτησης σε τοπικές γειτονιές ή κλίκες [109]. Μέσω της ισοδυναμίας με κατανομές Gibbs χρησιμοποιήθηκαν συστηματικά για την κατασκευή διακριτών μοντέλων υφής [66, 76, 104, 196, 310]. Η στατιστική θεώρηση επιτρέπει την ανάλυση, μέσω της εκτίμησης του σχετικά μικρού αριθμού παραμέτρων από παρατηρήσεις υφής αλλά και τη σύνθεση με δειγματοληψία από τις εν λόγω κατανομές. Παραλλαγές των αρχικών

μοντέλων περιλαμβάνουν περιορισμούς ως προς τη γεωμετρική μικροδομή της υφής [266]. Στην κατεύθυνση μιας ενοποιημένης και περιεκτικής θεώρησης της υφής, ερευνήθηκαν περιγραφικά (descriptive) μοντέλα που επεκτείνουν τα τυχαία πεδία αντικαθιστώντας τις κλίκες με γραμμικά φίλτρα [324, 325] ή συνδυάζοντας τα με αναγεννητικά (generative) μοντέλα [112–114, 320].

### Στοχαστικές διαδικασίες και φράκταλ

Η γεωμετρική πολυπλοκότητα της υφής των επιφανειών μπορεί να ποσοτικοποιηθεί αναλύοντας την αυτοομοιότητα στη δομή της σε πολλαπλές κλίμακες, η οποία μπορεί να είναι ντετερμινιστική ή στοχαστική. Τα μαθηματικά σύνολα φράκταλ (fractal) [193] μπορούν να αποτελέσουν μοντέλα τέτοιων διαδικασιών για την ανάλυση της υφής αλλά και του σχήματος αντικειμένων και επιφανειών. Τέτοια μοντέλα περιγράφουν με μικρό αριθμό παραμέτρων την τραχύτητα μιας επιφάνειας, την πολυπλοκότητα μιας καμπύλης ή τη στατιστική κατανομή ενός συνόλου. Ντετερμινιστικά, η φράκταλ διάσταση [18, 235] χρησιμοποιείται για μετρήσεις υφής, όπου οι επιφάνειες θεωρούνται φράκταλ σε ένα περιορισμένο εύρος κλιμάκων [242]. Στοχαστικά φράκταλ, όπως οι κλασματικές κινήσεις Brown (fractional Brownian motion), μια οικογένεια  $1/f$  θυρύβων [194] χρησιμοποιούνται για την περιγραφή σε πολλαπλές κλίμακες φυσικών σκηνών και υφής επιφανειών [151, 236, 277]. Οι  $1/f$  θύρυβοι είναι στοχαστικές διαδικασίες που μοντελοποιούν ένα πλήθος φυσικών φαινομένων. Φασματικά μεταβάλλονται αντιστρόφως ανάλογα μιας δύναμης της συχνότητας και οι γράφοι τους για δισδιάστατες συναρτήσεις είναι φράκταλ επιφάνειες. Τέτοια μοντέλα φράκταλ επιφανειών συμφωνούν με τα πολυκαναλικά μοντέλα χωρικών φίλτρων και ψυχοφυσικές μελέτες για τη σχέση τραχύτητας επιφάνειας, διάστασης και αντίληψης υφής [171, 236].

#### 1.2.5.2 Νευροβιολογικές και ψυχοφυσικές μελέτες

Η γενική αρχή πολυκαναλικής (multichannel) ή πολυζωνικής (multiband) ανάλυσης μέσω φίλτρων με επιλεκτικότητα στον προσανατολισμό και στη συχνότητα βασίστηκε σε θεωρίες για στην παράλληλη επεξεργασία της οπτικής πληροφορίας στα πρώτα στάδια όρασης [305]. Τα μοντέλα φίλτρων, ή ενεργειακά μοντέλα, δηλαδή η περιγραφή της υφής με διανύσματα χαρακτηριστικών στις εξόδους ενός συνόλου γραμμικών φίλτρων, υποστηρίζονται από μελέτες φυσιολογίας και νευροφυσιολογίας βιολογικών συστημάτων όρασης και ψυχοφυσικές ενδείξεις για συχνοτική ανάλυση από το οπτικό σύστημα [49, 254]. Η εικόνα στον αμφιβληστροειδή διαχωρίζεται σε πολλαπλές φιλτραρισμένες εκδοχές της, η ίδια μια από τις οποίες περιέχει μεταβολές έντασης σε ένα περιορισμένο, στενό εύρος συχνοτήτων [49, 77, 188].

Το σύνολο των αποτελεσμάτων πειραματικής φυσιολογίας των Hubel και Wiesel αφορούσε μεταξύ άλλων τα προφίλ και τις ιδιότητες των κυττάρων-νευρώνων (cortical cells) στον οπτικό φλοιό (primary visual cortex), διακρίνοντας τα σε «απλά» και «σύνθετα», μελετώντας τις αποχρίσεις νευρώνων σε στοιχειώδη ερεύσιμα [130–133]. Τα υποδεκτικά πεδία (receptive fields), η χωρική απόχριση δηλαδή, των «απλών» νευρώνων (simple cells) μοντελοποιήθηκαν

## Κεφάλαιο 1. Εισαγωγικά

από συναρτήσεις βέλτιστης από κοινού χωρικής και συχνοτικής τοποθέτησης, τις συναρτήσεις Gabor [71, 141, 170, 191, 204, 290, 306]. Ο Daugman ανέλυσε τις ιδιότητες των 2-Δ Gabor συναρτήσεων ως ζωνοπερατά φίλτρα που μιμούνται την βέλτιστη αναπαράσταση της πληροφορίας από τους οπτικούς νευρώνες [72]. Το γιατί τα νευρικά κύτταρα αποκρίνονται κατά τέτοιο τρόπο αποδόθηκε από τον Field στη βέλτιστη κωδικοποίηση των στατιστικών των φυσικών εικόνων [96, 224, 225].

Σχετικά με την υφή, ψυχοφυσιολογικά πειράματα μετρήσεων της απόκρισης σε ερεύνησματα, ανέδειξαν για τους θιασώτες της πολυκαναλικής θεώρησης μηχανισμούς κατανομής της ενέργειας σε πολλαπλά ισοτροπικά κανάλια [22, 25]. Αποτελέσματα διάκρισης με πολλαπλά φίλτρα Gabor προέβλεψαν με καλή ακρίβεια την ανθρώπινη απόδοση [97]. Από την άλλη, οι υποστηρικτές της ανάλυσης μέσω τοπικών χαρακτηριστικών σε αντίστοιχα πειράματα αναγνώρισαν μηχανισμούς τοπικών ανιχνευτών διαφορών σε στοιχειώδη χαρακτηριστικά [21, 130, 144, 205, 206, 303]. Τόσο η πυκνότητα, όσο και ο προσανατολισμός των διαφορετικών δομικών προτύπων παίζουν θεμελιώδη ρόλο στην αντίληψη του οπτικού φαινομένου και το γεγονός υπονοεί μια αποκλειστική σχέση ανάμεσα στην υφή και σε τοπικές χωρικές συχνότητες. Οι Julesz και Bergen παρατήρησαν την ισοδυναμία συνδυασμού φίλτρων και ανιχνευτών τεξτονίων και επεσήμαναν την αναγκαιότητα πολλαπλών φίλτρων για καλύτερη διάκριση υφής [145, 146].

### 1.2.5.3 Πολυσυχνοτική ανάλυση και χωρικά φίλτρα

Το πολυκαναλικό μοντέλο, πέρα από βιολογικά εύλογο, είναι ταυτόχρονα ελκυστικό και ευέλικτο καθώς επιτρέπει την αναζήτηση διαφορετικών μεγεθών και προσανατολισμών που κυριαρχούν δομικά σε μια υφή, ενώ προσφέρεται για κλιμακωτή, πολλαπλής ευχρίνειας ανάλυση. Τα χαρακτηριστικά υφής που εξάγονται από τις αποκρίσεις ενός συνόλου φίλτρων είναι συνήθως ενεργειακές μετρήσεις από μη-γραμμικούς τελεστές, στατιστικά γκρίζων τιμών ή κατανομές πιθανότητας και έχουν μεγαλύτερη περιγραφική ικανότητα από στατιστικά μοντέλα και μετρήσεις.

Το μοντέλο χρησιμοποιήθηκε σε διάφορες εκδοχές από το '80 και μετά, με την ίδια πάντα αρχή και φιλοσοφία των χωρικά και συχνοτικά επιλεκτικών φίλτρων για την αναπαράσταση και ανάλυση της υφής εικόνων: «πυραμίδες» από χωρικά φίλτρα Laplacian of Gaussians [45], παράγωγοι ή διαφορές συναρτήσεων Gauss [67, 139, 160, 178, 294, 295], φίλτρα Gabor [22, 35, 36, 62, 86, 87, 97, 122, 138, 243, 255, 290], φίλτρα ορθογωνισμού φάσης (quadrature filters [158, 188, 220]), ορθοκανονικές βάσεις [116, 225], κυματίδια (wavelets [56, 176, 191, 192, 244, 292]), «οδηγούμενα» (steerable) φίλτρα και πυραμίδες [26, 100, 128, 237] κ.ά. Μια συγκριτική και βιβλιογραφική παρουσίαση τέτοιων μεθόδων για περιγραφή υφής παρουσιάζεται στο [250].

Η αναπαράσταση της υφής από τις εξόδους μιας συστοιχίας φίλτρων (filterbank), δηλαδή μιας διάταξης φίλτρων με διαφορετική συχνοτική τοποθέτηση που καλύπτουν ένα εύρος χωρικών συχνοτήτων, αποσκοπεί στην κατασκευή περιγραφικών διανυσμάτων χαρακτηριστι-

κών. Η περιγραφικότητα και η ευρωστία της πολυκαναλικής ανάλυσης, αντισταθμίζεται από τον αριθμό παραμέτρων της διάταξης που πρέπει να ρυθμιστούν και την αυξημένη διάσταση των διανυσμάτων στην έξοδο. Οι διαφορετικές προσεγγίσεις για την εξαγωγή πολυκαναλικών χαρακτηριστικών καλύπτουν δύο κατηγορίες εκτιμήσεων: α) γενικευμένων ενεργειακών ποσοτήτων και β) στατιστικά κατανομών γκρίζων τιμών ή δομικών μονάδων.

Η πρώτη κατηγορία στηρίζεται σε υπολογιστικά μοντέλα βιολογικών συστημάτων όρασης. Η γραμμικότητα του μετασχηματισμού των φίλτρων στις αποκρίσεις είναι ανεπαρκής για διάκριση και αντίθετη από ψυχοφυσικές ενδείξεις [188]. Έτσι, αναζητήθηκαν μη-γραμμικοί τελεστές στις εξόδους των φίλτρων που συνδέθηκαν με τη μέτρηση διαφορετικών ενεργειακών ποσοτήτων: τελεστές ανόρθωσης μισού και πλήρους κύματος (ενέργεια [26, 188] και πλάτος [25, 34]), ενέργεια ζευγαριών ορθογωνισμού (quadrature pair energy [3, 220]) όπως είναι για παράδειγμα τα μιγαδικά Gabor φίλτρα [97, 290], περιβάλλουσα και φάση της απόκρισης μιγαδικών φίλτρων [36], τελεστές εμπνευσμένοι από βιολογικά μοντέλα για ανίχνευση στοιχειωδών δομών (blobs [138], bars [241]) ή περιοδικότητας σε κατευθυντικές δομές (grating cells [168]) και πολυφασματικά χρωματικά χαρακτηριστικά [136].

Η δεύτερη κατηγορία περιλαμβάνει ανάλυση (αλλά και σύνθεση) από τα στατιστικά των χαρακτηριστικών στις εξόδους των φίλτρων. Τέτοια είναι τα τοπικά ιστογράμματα των αποκρίσεων των γραμμικών καναλιών [113, 128, 324, 325] ή τις πυκνότητες τεξτονίων που ορίζονται με χβαντοποίηση των διανυσμάτων μέσω ομαδοποίησης στο χώρο των αποκρίσεων [178, 186]. Η ανάλυση μπορεί να περιλαμβάνει και μοντελοποίηση των κατανομών με ταίριασμα των κατάλληλων συναρτήσεων πυκνότητας-πιθανότητας, όπως μείγματα από Gaussian κατανομές [232] ενώ η σύνθεση υφής, περιορισμούς στα από-κοινού στατιστικά, όπως στα περιτώρια ιστογράμματα ή τη συσχέτιση, των συντελεστών μιας πολυκαναλικής κυματιδιακής αναπαράστασης [244].

#### 1.2.5.4 Μοντέλα διαμόρφωσης

Η συσχέτιση της υφής με τις τοπικές χωρικές συχνότητες καθώς και το γεγονός ότι οι ηλεκτροχημικές αποκρίσεις των κυττάρων του οπτικού φλοιού μεταδίδουν πληροφορία αντίθεσης φωτεινότητας και χωρικής φάσης [122] παρακινεί μια επιπλέον ερμηνεία της υφής: ως ομαλά μεταβαλλόμενες χωρικές διαμορφώσεις (spatial modulations), σε επιλεκτικά ως προς τη συχνότητα και τον προσανατολισμό κανάλια. Αναπαριστώντας την υφή με ένα τέτοιο μοντέλο πολλαπλών, στενοζωνικών συνιστώσων, οι μεταβολές των τιμών μιας εικόνας περιγράφονται στα πλαίσια τοπικών διαμορφώσεων και τα χαρακτηριστικά της υφής ερμηνεύονται από τοπικές συχνότητες.

Οι Bovik, Clark, Geisler [34, 36] μοντελοποίησαν μεμονωμένες συνιστώσες υφής στην έξοδο Gabor φίλτρων ως 2-Δ AM-FM διαμορφωμένες χωρικές συναρτήσεις, οι οποίες μπορούσαν να αποδιαμορφωθούν σε πλάτος και φάση από την απόκριση του φίλτρου. Αντί για τη φάση, η υφή που τοπικά παρουσιάζει στενοζωνική (narrowband) μεταβολή συνδέθηκε από τον Bovik κ.ά. με τοπικά αναδυόμενες συχνότητες (locally emergent frequencies [37]), και αποδιαμορφώθηκε σε πλάτος και διανυσματικά μεταβαλλόμενων συχνοτήτων.

Τέτοια 1-Δ AM-FM μοντέλα χρησιμοποιήθηκαν με επιτυχία στη μοντελοποίηση της μη-γραμμικότητας σε σήματα φωνής από τους Maragos, Kaiser, Quatieri [38, 200–202]. Το αναλυτικό πλαίσιο επεκτάθηκε και στη μοντελοποίηση υφής, αλλά και εικόνων γενικότερα με πολλαπλές AM-FM συνιστώσες από τους Maragos, Havlicek, Harding, Bovik κ.ά. [2, 74, 101, 120, 123–127, 198, 234, 307]. Σύμφωνα με το μοντέλο, οι συνιστώσες διαμόρφωσης είναι μη-στατικά ημιτονοειδή, των οποίων το χωρικά μεταβαλλόμενο πλάτος μοντελοποιεί τις τοπικές μεταβολές της αντίθεσης, ενώ το διάνυσμα της στιγμιούματα μεταβαλλόμενης φάσης τις τοπικά αναδυόμενες ταλαντώσεις υφής (κλίμακα, προσανατολισμός).

Το πρόβλημα της εκτίμησης των παραμέτρων του μοντέλου όταν εφαρμόζεται σε εικόνες υφής οδήγησε στην αναζήτηση εύρωστων υπολογιστικών αλγορίθμων για την αποδιαμόρφωση των συνιστώσων και την εξαγωγή των στιγμιούματων διαμόρφωσης. Ένας τέτοιος απλός και αποτελεσματικός αλγόριθμος, ο αλγόριθμος διαχωρισμού ενέργειας (ESA) προτάθηκε από τους Maragos, Bovik και βασίστηκε σε ένα μη-γραμμικό 2-Δ ενεργειακό τελεστή [198], γενίκευση του 1-Δ ενεργειακού τελεστή Teager-Kaiser [148, 317] που είχε προηγουμένως εφαρμοστεί στην ανάλυση και αποδιαμόρφωση σημάτων φωνής [201]. Ο τελεστής υπολογίζει την «ενέργεια» ενός ιδεατού συστήματος παραγωγής 2-Δ ταλαντώσεων στις οποίες αποδίδονται η μορφή και οι ιδιότητες της υπό ανάλυσης υφής.

Για την εφαρμογή του μοντέλου, η απομόνωση συχνοτικά των επιμέρους στενοζωνικών συνιστώσων και η αποδιαμόρφωση εφαρμόστηκε στην έξοδο συστοιχίας Gabor φίλτρων [38, 124, 125], που εκτός του διαχωρισμού, ομαλοποιούν τα σήματα ενέργειας και διαμόρφωσης. Έτσι, το πολυκαναλικό μοντέλο επεκτείνεται μέσω της περαιτέρω μοντελοποίησης των αποκρίσεων των φίλτρων και ενοποιείται με τα μοντέλα διαμόρφωσης. Η εξαγωγή διανυσμάτων χαρακτηριστικών της υφής ανάγεται στον εντοπισμό και την ανάλυση των τοπικά κυρίαρχων διαμορφώσεων της εικόνας από πολλαπλά συχνοτικά κανάλια.

### 1.3 Εννοιολογικά και υπολογιστικά προβλήματα υφής

Η μελέτη της υφής ως οπτικό φαινόμενο και αντιληπτική ιδιότητα εικόνων μέσα από την διερεύνηση των βιολογικών μηχανισμών όρασης και την δημιουργία υπολογιστικών αλγορίθμων μπορεί να θεωρηθεί ότι επικεντρώνεται σε τρία υπερκατηγορίες προβλημάτων: *αναπαράσταση, ερμηνεία και αντίληψη* της υφής.

Η αναπαράσταση σχετίζεται με τη δημιουργία περιγραφών για την υφή σε ένα σύστημα όρασης, η ερμηνεία με την επεξεργασία και την ανάλυση αυτών των περιγραφών η αντίληψη με την εξαγωγή συμβόλων για την υποκείμενη σκηνή. Οι προσπάθειες για την αναζήτηση ενός κοινού και βάσιμου εννοιολογικού, μαθηματικού και υπολογιστικού πλαισίου παρακινούνται από την πρωτογενή δυσκολία ορισμού του φαινομένου. Εξάλλου, η υφή εικόνων αναφέρεται συνήθως ως μια ιδιότητα που εύκολα αναγνωρίζεται αλλά δύσκολα ορίζεται. Ένα τέτοιο υποθετικό πλαίσιο ωμα διευκόλυνε τον ορισμό, την κατανόηση αλλά και την ερμηνεία της υφής. Θα βοηθούσε δηλαδή σε ερωτήματα σχετικά με το *τι*, το *πως* και το *γιατί*, αντίστοιχα, γύρω από την υφή και τις διαφορετικές εκφάνσεις της υφής στο φυσικό κόσμο και κατ' επέκταση, στην αναπαράσταση του μέσα από εικόνες.

Τα επιμέρους ζητήματα που προκύπτουν σχετικά με την κατανόηση και την αποδοτική εκμετάλλευση της υφής και των μικροδομών εικόνων επαφίενται σε διαφορετικά γνωστικά και επιστημονικά αντικείμενα. Στα πεδία της φυσικής και της οπτικής εντάσσονται θέματα όπως τι είναι, που οφείλεται, πως σχηματίζεται η φυσική υφή επιφανειών, ποιες είναι οι ιδιότητες της και πως μεταβάλλονται αυτές σε σχέση με τη γεωμετρία της σκηνής, τις συνθήκες φωτισμού και την απόσταση θέασης. Για τις γνωσιακές επιστήμες τα προβλήματα σχετίζονται με τις διεργασίες της εσωτερικής νοητικής αναπαράστασης και την οπτική αντίληψη της υφής, τη διάκριση της καθώς και τη σημασία της για το σύνολο της οπτικής αντίληψης. Για παράδειγμα ερωτήματα όπως σε πιο βαθύτο η πληροφορία υφής σε φυσικές σκηνές είναι χρήσιμη για κατάτμηση ή αναγνώριση αντικειμένων, με τι τρόπο η υφή βοηθάει στην ομαδοποίηση και προκαλεί την προσοχή και πως οι άνθρωποι χρησιμοποιούν την υφή για να συμπεράνουν πληροφορία σχήματος και βάθους. Υπάρχει συσχέτιση ανάμεσα στα αισθητήρια ή στις νοητικές αναπαραστάσεις οπτικής υφής και αφής<sup>1</sup>;

Στην ανάλυση και επεξεργασία εικόνων τα εννοιολογικά προβλήματα σχετίζονται με το πως εμφανίζεται η υφή στις εικόνες και πως ανταποκρίνεται στη φυσική υφή των επιφανειών, ποια τα φωτομετρικά και γεωμετρικά χαρακτηριστικά της και υπό τι προϋποθέσεις οι μεταβολές της έντασης του γκρίζου ή του έγχρωμου φάσματος ερμηνεύονται ως υφή. Τέλος για την όραση υπολογιστών και την τεχνητή νοημοσύνη τα προβλήματα έχουν να κάνουν με το ρόλο της υφής στη δημιουργία συμβολικών αναπαραστάσεων του 3-Δ κόσμου, τη συστηματική υπολογιστική μοντελοποίηση της ιδιότητας αυτής των εικόνων αλλά και τον τρόπο και το επίπεδο επεξεργασίας της στην όραση. Αρκεί μια γρήγορη, «άνωθεν» (bottom-up) προσέγγιση χαμηλού επιπέδου ή απαιτείται η ενσωμάτωση «κάτωθεν» (top-down) γνώσης;

Διαισθητικά η υφή περιλαμβάνει «έννοιες» όπως περιοχές, κατανομή τιμών, θεμελιώδη δομικά στοιχεία και κλίμακα και οι οποίες καθορίζουν πολλά από τα υπολογιστικά προβλήματα της υφής. Η υφή ως φορέας πληροφορίας περιοχής σε μια εικόνα σχετίζεται με προβλήματα και εφαρμογές όπως ανίχνευση, διάκριση, κατάτμηση, ταξινόμηση και αναγνώριση. Οι στατιστικές ιδιότητες της υφής στο χωρικό πλέγμα της εικόνας και υπεισέρχονται σε προβλήματα όπως ο καθορισμός των διαφορετικών «ειδών» υφής, η διάκριση, η σύνθεση υφής από παραδείγματα, η συσχέτιση των μικροδομών με τις μακροδομές ή την αδρή γεωμετρία των εικόνων. Ο καθορισμός των θεμελιώδων δομικών στοιχείων, αφορά στην περιγραφή των διαφορετικών κατηγοριών υφής με στοιχειώδεις χωρικές φωτομετρικές ή γεωμετρικές δομές (texture primitives), όπως είναι τα τεξτόνια ή πρωτογενή γεωμετρικά στοιχεία. Τα προβλήματα περιλαμβάνουν τον ορισμό και την ανίχνευση τέτοιων μικροδομών, την ανάπτυξη διαδικασιών εξαγωγής και μέτρησης τους αλλά και τη διερεύνηση της σχέσης συγκέντρωσης και πλήθους με την την κλίμακα και τη διάκριση της υφής.

Η σχέση κλίμακας και δομής βρίσκονται στο επίκεντρο οποιασδήποτε προσέγγισης και μελέτης της υφής. Ο ορισμός της κλίμακας είναι θεμελιώδης καθώς σε διαφορετικά επίπεδα ανάλυσης αντικείμενα, σύνολα ή λεπτομέρειες αντικειμένων γίνονται αντιληπτά ως υφή.

<sup>1</sup>Η αφή είναι η πρώτη αίσθηση που αναπτύσσεται στα έμβρυα, ενώ ερευνάται η ανάπτυξη της στα βρέφη και η σχέση με την ανάπτυξη των υπολοίπων αισθήσεων, όπως της όρασης.

## Κεφάλαιο 1. Εισαγωγικά

Σύμφωνα με το Gibson [105] η υφή είναι ένα από τα επίπεδα δομής της οπτικής πληροφορίας. Πρόκειται για τις μικροδομές (fine structure) της ύλης και συνυπάρχει μαζί με τις μακροδομές (coarse structure) 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 αντιληπτική δομή στην άλλη ακολουθεί τις παραμέτρους, τους περιορισμούς και την ευχρίνεια του συστήματος όρασης, βιολογικού ή τεχνητού.

Η πολυκλιμακωτή περιγραφή εικόνων είναι κεντρικής σημασίας για τα προβλήματα της όρασης, όπου η αναπαράσταση της πληροφορίας γίνεται σε πολλαπλά επίπεδα, με εξειδικευμένη επεξεργασία και ανάλυση της πληροφορίας για το καθένα [205]. Το αντίστροφο πρόβλημα του διαχωρισμού της υφής από τις μακροδομές μιας εικόνας εντάσσεται στο γενικότερο πρόβλημα αποσύνθεσης ή αποδόμησης (image decomposition), που στοχεύει στην εκτίμηση των συνιστώσων διαφορετικής φύσης που απαρτίζουν τις φυσικές εικόνες. Τέτοιες συνιστώσες όπως περιγράφονται από μοντέλα υπέρθεσης είναι οι μακροδομές, η υφή και ο θόρυβος και η εκτίμηση τους σχετίζεται τόσο με την κατανόηση της δομής των εικόνων όσο και με την επεξεργασία της πληροφορίας με τρόπους προσαρμοσμένους στη φύση της κάθε συνιστώσας.

Τέλος, όπως η υφή εικόνων γίνεται αντιληπτή ως η λεπτομέρεια που αποδίδεται σε επιφάνειες μιας σκηνής και επεξεργάζεται ως μικροδομή των 2-Δ σημάτων-εικόνων, ανάλογη λεπτομέρεια μπορεί να χαρακτηρίζει και άλλα αισθητήρια σήματα. Η μικροδομή σε ηχητικά ή ακουστικά 1-Δ σήματα, όπως είναι η φωνή, έχει διαφορετική φυσική σημασία και εκφράζει διαφορετικές ιδιότητες του ήχου. Στο παρελθόν αποδείχθηκε δόκιμη και επιτυχημένη η θεώρηση κοινών πλαισίων περιγραφής και ανάλυσης για εικόνες και ήχους, με βάση κοινές αρχές για επεξεργασία πολυδιάστατων σημάτων. Τέτοιες αλληλεπιδράσεις και προσεγγίσεις για την υφή σε άλλες κατηγορίες σημάτων, παρουσιάζουν ενδιαφέρον τόσο από την άποψη ανταλλαγής ιδεών, όσο και από την άποψη μιας ενιαίας θεώρησης για την αντίληψη.

## 1.4 Συνεισφορές επιγραμματικά

Κύριοι άξονες της διατριβής είναι τα προβλήματα της μοντελοποίησης και ανάλυσης οπτικής υφής εικόνων σε συστήματα όρασης υπολογιστών, με έμφαση στα επιμέρους προβλήματα της ανίχνευσης, κατάτμησης και διαχωρισμού της σε σύνθετες εικόνες. Ερευνήθηκαν και προτείνονται προηγμένες τεχνικές μοντελοποίησης και ανάλυσης σε παράλληλες κατευθύνσεις: α) πολυκαναλικά μοντέλα στενοζωνικών συνιστώσων και χωρικών διαμορφώσεων, β) ενεργειακές μέθοδοι για εξαγωγή χαρακτηριστικών, γ) μεταβολικές τεχνικές για αποδόμηση εικόνων και διαχωρισμό των συνιστώσων υφής. Οι συνεισφορές που προέκυψαν και προτείνονται συστηματικά προς την κατεύθυνση ενός εξελιγμένου συστήματος ανάλυσης υφής είναι επιγραμματικά:

- Πολυκαναλικά μοντέλα υφής AM-FM χωρικών διαμορφώσεων.
- Ορισμός, εντοπισμός και ανίχνευση της ενέργειας υφής με βάση τις κυρίαρχες διαμορ-

φώσεις.

- Επέκταση εύρωστων τεχνικών αποδιαμόρφωσης και εκτίμησης παραμέτρων του μοντέλου.
- Εξαγωγή περιγραφικών διανυσμάτων χαρακτηριστικών για ανάλυση υφής.
- Αξιοποίηση της υφής σε δύο νέα, μεταβολικά σχήματα κατάτμησης εικόνων με πολλαπλά κριτήρια.
- Ανάπτυξη σχήματος αποσύνθεσης εικόνων και διαχωρισμού της υφής από τις μακροδομές.
- Βελτιωμένη ταξινόμηση εικόνων υφής βασισμένη στην εξαγωγή χαρακτηριστικών από τη διαχωρισμένη συνιστώσα.
- Εφαρμογή των μεθόδων ανάλυσης υφής σε εικόνες εδαφοτομών.

Οι συνεισφορές από την επιμέρους εφαρμογή των μεθόδων και ιδεών σε θέματα ανάλυσης φωνής είναι:

- Εφαρμογή των μοντέλων ανάλυσης σε πολλαπλές συνιστώσες και ενέργειας των μικροδομών σε ηχητικά σήματα.
- Εξαγωγή χαρακτηριστικών για ηχητικά σήματα και ανίχνευση φωνητικών γεγονότων σε περιβάλλον θορύβου.
- Αλγόριθμος ανίχνευσης φωνής που παρουσιάζει συστηματική βελτίωση απέναντι σε καθιερωμένους αλγόριθμους αναφοράς.

## 1.5 Διάρθρωση κειμένου

Στο Κεφάλαιο 2 γίνεται μια εισαγωγή στην έννοια της υφής εικόνων, βιβλιογραφικά και διαισθητικά και ορίζεται το γενικό θεωρητικό υπόβαθρο και τα εργαλεία που θα χρησιμοποιηθούν στη συνέχεια. Περιγράφεται το μοντέλο συνιστωσών διαμορφώσεων για εικόνες και αναλύονται στατιστικά οι ενεργειακοί τελεστές ως βασικά εργαλεία εντοπισμού της ενέργειας μικροδομών και αποδιαμόρφωσης. Στο Κεφάλαιο 3 αναπτύσσεται η αναπαράσταση της υφής εικόνων με δομικά στοιχεία χωρικά σήματα διαμορφώσεων και ορίζεται η ενέργεια της. Με βάση την ενέργεια διαμορφώσεων περιγράφονται νέες μέθοδοι εξαγωγής χαρακτηριστικών που βασίζονται στην πολυκαναλική ανάλυση και τον εντοπισμό των κυρίαρχων συνιστωσών διαμόρφωσης.

Στο Κεφάλαιο 4 τα χαρακτηριστικά αυτά εντάσσονται σε μια μεταβολική μέθοδο κατάτμησης εικόνων προσανατολισμένη στην υφή (texture-oriented) και παρουσιάζεται ένα σύνολο αποτελεσμάτων σε φυσικές εικόνες. Στο Κεφάλαιο 5 προτείνεται ένα νέο μεταβολικό σχήμα αποδόμησης εικόνων και εξετάζεται η συνέργεια των διαδικασιών διαχωρισμού της υφής από

## *Κεφάλαιο 1. Εισαγωγικά*

τις μακροδομές της εικόνας και μοντελοποίησης της υφής με το προτεινόμενο μοντέλο διαμόρφωσης. Αφενός η εισαγωγή του μοντέλου στο σχήμα βοηθάει την αποσύνθεση, ενώ αφετέρου η διαχωρισμένη συνιστώσα υφής οδηγεί σε βελτιωμένα χαρακτηριστικά, όπως εξετάζεται στο Κεφάλαιο 6, σε δοκιμές ταξινόμησης εικόνων υφής με και χωρίς επίβλεψη.

Στο Κεφάλαιο 7 η ανάλυση της υφής εντάσσεται στο πλαίσιο της γενικότερης εφαρμογής της ανάλυσης ψηφιοποιημένων εικόνων εδαφοτομών, που χαρακτηρίζονται από έντονη παρουσία υφής και μικροδομών. Το προτεινόμενο πλαίσιο μοντελοποίησης εφαρμόζεται σε θέματα ανίχνευσης, ανάλυσης, κατάτμησης αλλά και κατηγοριοποίησης της υφής των εικόνων. Παρουσιάζεται μια εναλλακτική μεθοδολογία αποσύνθεσης εικόνων σε δύο συνιστώσες με βάση τη μαθηματική μορφολογία. Στις προκύπτουσες συνιστώσες υφής και γεωμετρίας στηρίζεται ένα σχήμα συζευγμένης κατάτμησης (coupled segmentation) εικόνων. Ανεξάρτητα αλλά και συνοδευτικά ως προς την παρουσίαση της υπόλοιπης διατριβής, το Κεφάλαιο 8 ασχολείται με την εφαρμογή αντίστοιχων μοντέλων και μεθόδων σε 1-Δ ακουστικά σήματα για την ανίχνευση φωνητικών γεγονότων και τον καθορισμό των ορίων της φωνής. Τέλος στο Κεφάλαιο 9, συνοψίζονται τα γενικά συμπεράσματα της διατριβής, ανοιχτά ερευνητικά ζητήματα και μελλοντικές κατευθύνσεις.

□



## Κεφάλαιο 2

# Οπτική Υφή και Μικροδομές Διαμόρφωσης

Η οπτική υφή είναι η αναπαράσταση της φυσικής υφής επιφανειών που αποτυπώνεται σε εικόνες μαζί με λοιπές ιδιότητες της σκηνής, όπως το χρώμα, το σχήμα, η θέση, η διάταξη, ο προσανατολισμός, το βάθος, η κίνηση κ.ά. Η υφή είναι δύσκολο να οριστεί αλλά υπάρχει ένα ενιαίο πλαίσιο ιδιοτήτων και περιγραφικών γνωρισμάτων που διέπει τις περισσότερες περιγραφές και μοντέλα υφής. Το ενδιαφέρον επικεντρώνεται εδώ στην κατηγορία ενεργειακών μεθόδων περιγραφής με την ερμηνεία της υφής να δίνεται μέσω τοπικών χωρικών διαμορφώσεων σε εικόνες και κατάλληλων μεθόδων επεξεργασίας. Περιγράφεται η πολυκαναλική ανάλυση με φίλτρα Gabor και ο σχεδιασμός της συστοιχίας φίλτρων για το διαχωρισμό μιας εικόνας σε στενοζωνικές συνιστώσες. Παρουσιάζεται το μοντέλο διαμορφώσεων σε πολλαπλές ζώνες συχνοτήτων που θα εφαρμοστεί για την υφή και οι ενεργειακοί τελεστές αποδιαμόρφωσης μαζί με μια θεωρητική, στατιστική ανάλυση της αναμενόμενης συμπεριφοράς τους.

### 2.1 Ορισμοί, ιδιότητες και μοντέλα

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 ή την υφή εικόνων. Τέτοιοι σχηματισμοί μπορεί να είναι αποτέλεσμα των φυσικών ιδιοτήτων της επιφάνειας όπως είναι η τραχύτητα, το βάθος, το χρώμα, η διαφορά ανακλαστικότητας κ.ά. [287]. Λόγω της απουσίας μιας ενιαίας, φορμαλιστικής περιγραφής για την υφή παρατίθενται στη συνέχεια κάποιοι 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]  
[146]

“Η ερμηνεία μέσω πεπλεγμένων σωματιδίων είναι άμεσα συνδεδεμένη με την ιδέα της ευχρίνειας της υφής την οποία μπορεί να σκεφτεί κάποιος ως το μέσο αριθμό εικονοστοιχείων (pixel) για κάθε ευδιάκριτο στοιχείο υφής (texture element).” [Ballard & Brown, 1982] [15]

“Η υφή σχετίζεται με τη χωρική κατανομή των τιμών έντασης φωτεινότητας της εικόνας και με τα διακριτά χαρακτηριστικά τονικότητας (...) τρία χαρακτηριστικά είναι κρίσιμα σε 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] [117]

“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] [99]

“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 ως μια εικόνα υφής. Τέτοιες μεταβολές υφής μπορεί να είναι κανονικές (...), τυχαίες (...) ή ενδιάμεσες (...).” [Landy, 2004] [173]

Οι ιδιότητες της υφής που αποτυπώνονται σε μια εικόνα μπορούν να ερμηνευθούν ως:

- φυσικές μεταβολές των ιδιοτήτων της επιφάνειας (τραχύτητα, στασιμότητα, γεωμετρία, ανακλαστικότητα, φυσική υφή, θέση), β)
- μεταβολές των συνθηκών φωτισμού, απεικόνισης και περιβαλλοντικών παραγόντων, γ) χημικές-βιολογικές και εξελικτικές διαδικασίες (σχηματισμοί και πρότυπα σε οργανισμούς [289,311], χημική υφή).

Η οπτική υφή στον αμφιβληστροειδή ή σε κάποιο σύστημα ψηφιακής αναπαράστασης δεν είναι μία ένα προς ένα απεικόνιση της φυσικής υφής μιας επιφάνειας λόγω της συνύπαρξης των ανωτέρω ερμηνειών και των περιορισμών της απεικόνισης.

Τηπάρχει ένα σύνολο διαισθητικών ιδιοτήτων της υφής, όπως αυτή γίνεται αντιληπτή, που συμπεριλαμβάνονται στα περισσότερα μοντέλα και ορισμούς και θεωρούνται πραγματικές [287]. Σχετίζονται με το είδος και το πλήθος των «αρχέγονων», στοιχειωδών σχηματισμών (primitives) που απαρτίζουν την υφή:

- Η υφή είναι ιδιότητα περιοχών και 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 γκρίζων ή έγχρωμων τιμών, οπότε μπορεί να οριστεί και να αναλυθεί στατιστικά.

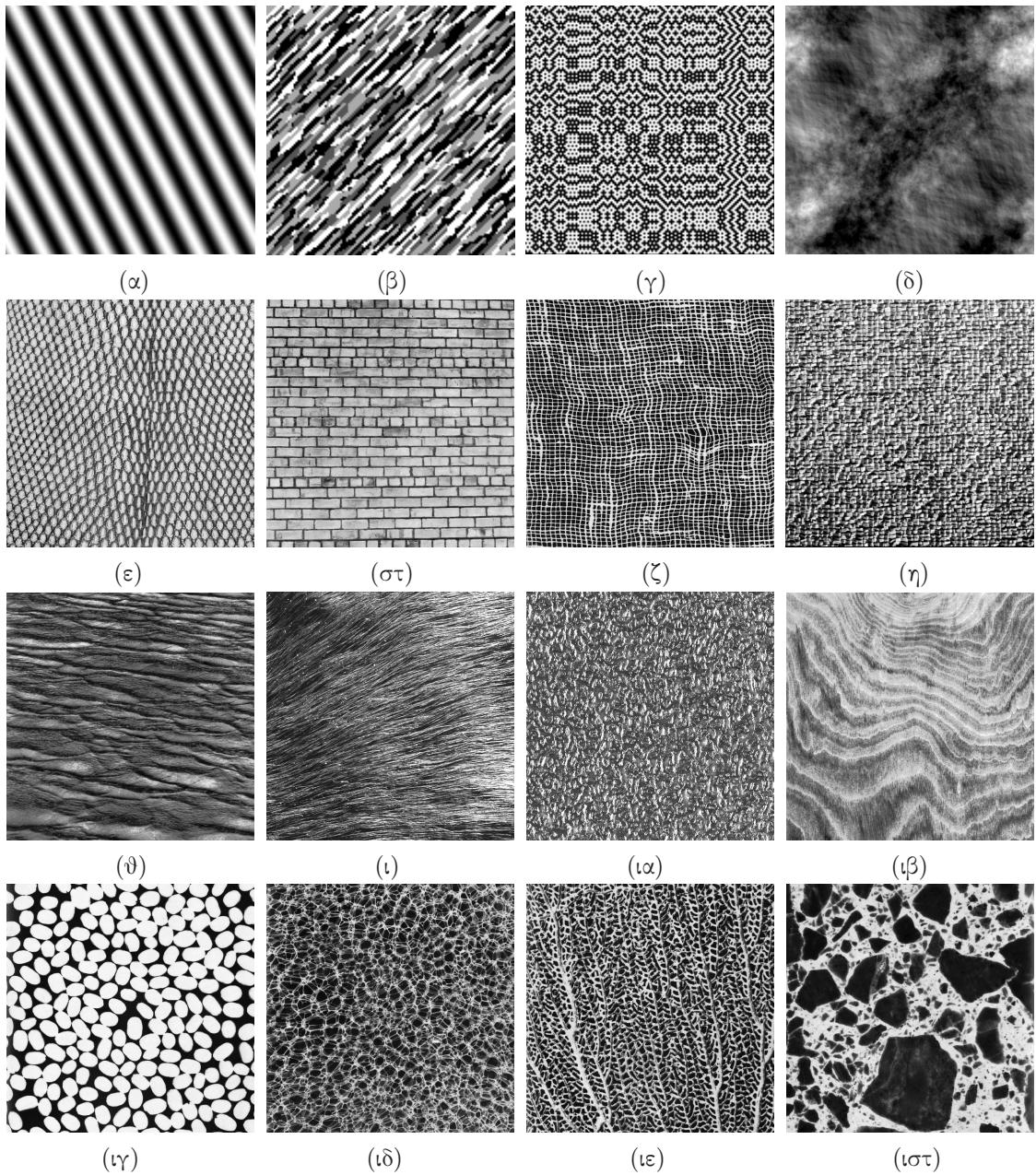
## Κεφάλαιο 2. Οπτική Υφή και Μικροδομές Διαμόρφωσης

- Η υφή γίνεται αντιληπτή σε διαφορετικές κλίμακες ή ευχρίνεια, 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.

Σημαντικά για την περιγραφή της οπτικής υφής είναι τα περιγραφικά γνωρίσματα που μπορεί να διαθέτει. Τέτοια γνωρίσματα, όχι απαραίτητα ανεξάρτητα μεταξύ τους, είναι η συγκέντρωση, η ομοιομορφία, η πυκνότητα, η τραχύτητα, η σκληρότητα, η κανονικότητα, η γραμμικότητα, η ένταση, η κατεύθυντικότητα, η κατεύθυνση, η συχνότητα, η φάση, η περιοδικότητα, η τυχαιότητα, η δομή, η φυσικότητα. Στο Σχ. 2.1 φαίνονται συνθετικές και φωτογραφικές γκρίζες εικόνες υφής, από φυσικές επιφάνειες με περισσότερο ή λιγότερο αντιληπτές ιδιότητες. Στο σύνολό τους πάντως είναι χωρικά ομογενείς, με επαναλαμβανόμενους σχηματισμούς και συχνά κάποια τυχαία μεταβολή, στη θέση, στο μέγεθος, στον προσανατολισμό ή στις στάθμες του γκρίζου.

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

- **Στατιστικές:** Χαρακτηριστικά και μοντέλα από την κατανομή των γκρίζων τιμών σε ένα διδιάστατο πλέγμα.
  - Μοντέλα γραμμικής πρόβλεψης (2-Δ μοντέλο αυτοπαλινδρόμησης, συνάρτηση αυτοσυγχέτισης) [58, 283].
  - Χαρακτηριστικά πινάκων «συνεμφάνισης» (co-occurrence matrices) (εντροπία, ενέργεια, συσχέτιση, αντίθεση, ομοιογένεια), 2ης - τάξης στατιστικά [118].
  - Τυχαία πεδία Gibbs-Markov [66, 76, 196, 310, 325].
  - Στοχαστικά φράκταλ, στατιστικά αυτοόμοιες διαδικασίες, κλασματικές κινήσεις Brown [151, 169, 182, 236, 240].
  - Τοπικές κατανομές μεγέθους και σχήματος, granulometries [272, 301].
- **Γεωμετρικές και δομικές:** Η υφή ορίζεται από τα θεμελιώδη δομικά συστατικά της.
  - Τεξτονία [67, 146, 178, 185, 296, 321].
  - Ντετερμινιστικά φράκταλ, γενικευμένες φράκταλ διαστάσεις [57, 59, 155, 242].
  - Διαγράμματα Voronoi [288].
  - Γραμματικά μοντέλα (shape, tree, array grammars) [15].
- **Ενεργειακές - Πολυκαναλικές:** Συχνοτική ανάλυση και αναπαράσταση της υφής με πολλαπλές φασματικές συνιστώσες.
  - Χαρακτηριστικά στην έξοδο γραμμικών φίλτρων (κατανομές, ενέργεια, μη-γραμμικότητες) [36, 110, 188, 250, 318].



Σχήμα 2.1: Παραδείγματα υφής εικόνων: (α),(β) συνθετικές περιοδικές, (γ),(δ) συνθετικές τυχαίες, (ε)-(η) φωτογραφικές φευδοπεριοδικές (reptile, bricks, burlap, raffia), (θ)-(ιβ) φωτογραφικές τυχαίες (water, fabric, metal, wood), (ιγ)-(ιτ) φωτογραφικές δομημένες. Οι φωτογραφικές είναι από τη βάση VisTex, MIT Media Lab [216] και τη συλλογή Brodatz [40].

- Αναπαραστάσεις χώρου-συχνότητας (space-frequency) ή κλίμακας-χώρου (scale-space) (κυματίδια, πυραμίδες, συστοιχίες φύλτρων) [116, 128, 190–192, 237, 243, 244, 292, 308].
- Χωρικά AM-FM μοντέλα, τοπικές διαμορφώσεις [37, 122, 198].

- *Μεταβολικές*: Ορισμός της υφής σε συναρτησιακούς χώρους, ελαχιστοποίηση συναρτησιακών, εκτίμηση με μερικές διαφορικές εξισώσεις (ΜΔΕ) [7–12, 29, 69, 108, 115, 214, 227, 274, 279, 298–300].

Οι σύγχρονες τάσεις επιδιώκουν 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 μεθόδικά η ανάλυση της υφής. Το ενδιαφέρον μας στρέφεται στις ενεργειακές, πολυκαναλικές μεθόδους για το μοντέλο που θα χρησιμοποιηθεί για την υφή, αλλά και στην ενοποίηση τους με μεταβολικές μεθόδους όπως θα παρουσιαστεί στη συνέχεια (κεφάλαιο 5).

## 2.2 Πολυσυχνοτικός διαχωρισμός

Η ανάλυση εικόνων σε πολλαπλές ζώνες χωρικών συχνοτήτων ή πολυσυχνοτική (multifrequency) ανάλυση [36, 191], χρησιμοποιήθηκε ευρέως στα μοντέλα αρχικής όρασης και επεξεργασίας χαμηλού επιπέδου λόγω της ομοιότητας με βιολογικές διεργασίες, της αποδοτικότητας ως προς την κωδικοποίηση πληροφορίας και των περιγραφικών πλεονεκτημάτων της αναπαράστασης σε πολλαπλά κανάλια. Φυσιολογικές και ψυχοφυσικές δοκιμές υποστηρίζουν την ύπαρξη στον οπτικό φλοιό μηχανισμών επιλεκτικών ως προς τη συχνότητα [96, 191]. Τέτοιοι μηχανισμοί αποτελούν τις βάσεις για μια «αραιή» (sparse) τεχνική οπτικής κωδικοποίησης που ελαχιστοποιεί τον πλεονασμό της πληροφορίας [224, 225]. Υπολογιστικά ο διαχωρισμός σε «κανάλια» συχνοτήτων απαντάται στις αναπαραστάσεις εικόνων πυραμίδας ή πολλαπλής ευχρίνειας (multiresolution) [128, 192] και στην ανάλυση υφής για εξαγωγή κατανομών ή χαρακτηριστικών (κεφάλαιο 1) [138, 244, 250, 325].

### 2.2.1 Χωρικές συχνότητες

Η αναπαράσταση μιας εικόνας ως πίνακα τιμών έντασης φωτεινότητας ορίζει την περιγραφή της στο χωρικό πεδίο (spatial domain) και τις μεταβολές της έντασης ανάμεσα σε διαφορετικά σημεία, όπως ορίζονται σε ένα διακριτό πλέγμα από pixel. Η μετάβαση μέσω του μετασχηματισμού Fourier στο πεδίο συχνοτήτων (frequency domain) δίνει το συχνοτικό περιεχόμενο της εικόνας ως υπέρθεση 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. Εκφράζεται συνήθως σε κύκλους ανά εικονοστοιχείο (cycles/pixel), κύκλους ανά μοίρα οπτικής γωνίας (cycles/degree) ή κανονικοποιημένα, στο πεπερασμένο διακριτό πλέγμα της εικόνας, σε κύκλους ανά 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 Σχ. 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 υψηλά επίπεδα φωτισμού είναι 50 – 60 cpd, ενώ σε χαμηλά επίπεδα 20 – 30 cpd.

Διαφορετικές χωρικές συχνότητες μεταφέρουν διαφορετική πληροφορία για το περιεχόμενο μιας εικόνας (Σχ. 2.2). Υψηλές χωρικές συνχρόνες παριστάνουν απότομες χωρικές αλλαγές στην εικόνα, όπως ακμές, ή ανταποκρίνονται σε πληροφορίες μικροδομής και λεπτομέρειες. Αντίθετα, χαμηλές χωρικές συχνότητες παριστάνουν καθολική πληροφορία για τις μακροδομές, το σχήμα, τις αναλογίες των αντικειμένων και το γενικό προσανατολισμό.

### 2.2.2 Γραμμικά φίλτρα και συναρτήσεις Gabor

Οι τυπικές φυσικές εικόνες υφής 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 υφής σε διαφορετικές περιοχές χωρικών συχνοτήτων και οποιαδήποτε χωρο-συχνοτική ανάλυση γίνεται μέσω γραμμικών μετασχηματισμών, φίλτρων ή κυματιδίων [176, 244, 250]. Αυτό περιλαμβάνει προβολές σε, ή ισοδύναμα συνελίξεις με, ένα σύνολο γραμμικών ζωνοπερατών φίλτρων.

Μια οικογένεια φίλτρων με ευνοϊκές για την ανάλυση ιδιότητες είναι οι συναρτήσεις Gabor που αποτελούν χωρικά διαμορφωμένα ημίτονα από συναρτήσεις Gauss [35, 70, 72, 102]. Τα προφίλ τους ακολουθούν τις ιδιότητες της απόχρισης των απλών κυττάρων, μιας κατηγορίας νευρώνων, στον οπτικό φλοιό των θηλαστικών. Εμπειρικές μελέτες στα προφίλ των χωρικών υποδεκτικών πεδίων γενικευμένων απλών κυττάρων συνηγορούν υπέρ της μοντελοποίησης από Gabor συναρτήσεις [65, 71, 72, 96, 170, 204, 306]. Επιπλέον, τα φίλτρα Gabor υπόκεινται σε ρύθμιση της χωρικής συχνότητας ή κλίμακας (scale), του προσανατολισμού (orientation) και του εύρους (bandwidth) ενώ στη μιγαδική τους έκφραση ικανοποιούν το κατώτατο όριο της αβεβαιότητας από κοινού για χωρική και συχνοτική τοποθέτηση [72].

Το πραγματικό και το φανταστικό μέρος του μιγαδικού Gabor φίλτρου αποτελούν ένα ζεύγος ορθογωνισμού φάσης που προσεγγίζει μια αναλυτική συνάρτηση. Η απόχριση, κανονικοποιημένη σε μοναδιαία ενέργεια ( $\|g\|_1 = 1$ ), είναι

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

και η απόχριση συχνότητας που καθορίζει και την επιλεκτικότητά του στο πεδίο των συχνοτήτων

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

όπου  $c, C$  είναι σταθερές κανονικοποίησης, το σημείο  $(x_c, y_c)$  προσδιορίζει την τοποθέτηση του φίλτρου χωρικά, οι τυπικές αποκλίσεις της 2-Δ Gaussian  $(a, b)$  καθορίζουν το ενεργό εύρος και την έκταση του φίλτρου προς τις δύο κατευθύνσεις και το διάνυσμα  $(u_c, v_c)$  τη διαμόρφωση συχνότητας ή τη 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 συμμετρία του, με  $\sigma = a = b$  για ισοτροπικά φίλτρα.

Οι σταθμερές  $c, C$  σχετίζονται με την «ενέργεια» του φίλτρου, όπως αυτή εκφράζεται μέσα από γενικευμένες  $p$ -νόρμες, και αποδεικνύεται ότι

$$c = \frac{\|g\|_p}{\sqrt[p]{2\pi ab/p}}, \quad C = 2\pi ab \frac{\|g\|_p}{\sqrt[p]{2\pi ab/p}}. \quad (2.3)$$

Έτσι, για φίλτρα μοναδιαίας  $L_1$  ενέργειας  $\|g\|_1 = 1$ , δηλαδή μοναδιαία dc απόχριση προκύπτουν  $c = 1/2\pi ab$  και  $C = 1$ . Στην περίπτωση της απόχρισης μοναδιαίου μεγίστου δηλ.  $\|g\|_\infty = 1$  οι τιμές είναι αντίθετες. Σημειώνεται ότι οι κανονικοποιήσεις των φίλτρων στο χώρο και στη συχνότητα συνδέονται από τις δυϊκές ενεργειακές νόρμες  $\|g\|_p = \|G\|_q$ , με  $1/p + 1/q = 1$ . Επομένως, μοναδιαία  $L_1$  ενέργεια στο χώρο ισοδυναμεί με μοναδιαίο πλάτος  $\|G\|_\infty = 1$  της απόχρισης συχνότητας. Για  $p = q = 2$ , η  $L_2$  ενέργεια μετράται και κανονικοποιείται ταυτόχρονα και στα δύο πεδία (θεώρημα Parseval).

Η ομοιότητα των Gabor αποχρίσεων στα πεδία χώρου και χωρικών συχνοτήτων μεταφράζεται σε ταυτόχρονη επεξεργασία της πληροφορίας και στα δύο πεδία. Για ένα φίλτρο τοποθετημένο στην αρχή, ο Fourier μετασχηματισμός του είναι μια συνάρτηση Gauß τοποθετημένη στις συχνότητες  $(u_c, v_c)$ , δηλαδή ένα ζωνοπερατό φίλτρο που επιτρέπει τη διέλευση περιορισμένου εύρους συχνοτήτων στις δύο διευθύνσεις. Τα Gabor αποτελούνε ομαλά και εύχρηστα ζωνοπερατά φίλτρα, εύκολα ελέγχιμα και ικανά να παράγουν μια ολόκληρη οικογένεια με παρόμοια χαρακτηριστικά μεταβάλλοντας ένα μικρό σύνολο παραμέτρων.

Τα μονοδιάστατα Gabor αποδείχτηκαν να ελαχιστοποιούν τη σχέση αβεβαιότητας σε μία διάσταση  $\Delta x \cdot \Delta u \geq 1/4$ , όπου το  $\Delta$  εκφράζει ενεργό rms εύρος [102]. Επεκτείνοντας στις δύο διαστάσεις [72], οι 2-Δ συναρτήσεις Gabor στη μιγαδική τους μορφή (2.1, 2.2) είναι οι μόνες που ελαχιστοποιούν ταυτόχρονα την αβεβαιότητα

$$\begin{aligned} (\Delta x)(\Delta y)(\Delta u)(\Delta v) &\geq 1/16, \\ \Delta x = \frac{a\sqrt{2}}{2}, \quad \Delta u = \frac{1}{2\sqrt{2}a} &\Rightarrow \Delta x \cdot \Delta u = 1/4 \\ \Delta y = \frac{b\sqrt{2}}{2}, \quad \Delta v = \frac{1}{2\sqrt{2}b} &\Rightarrow \Delta y \cdot \Delta v = 1/4. \end{aligned} \quad (2.4)$$

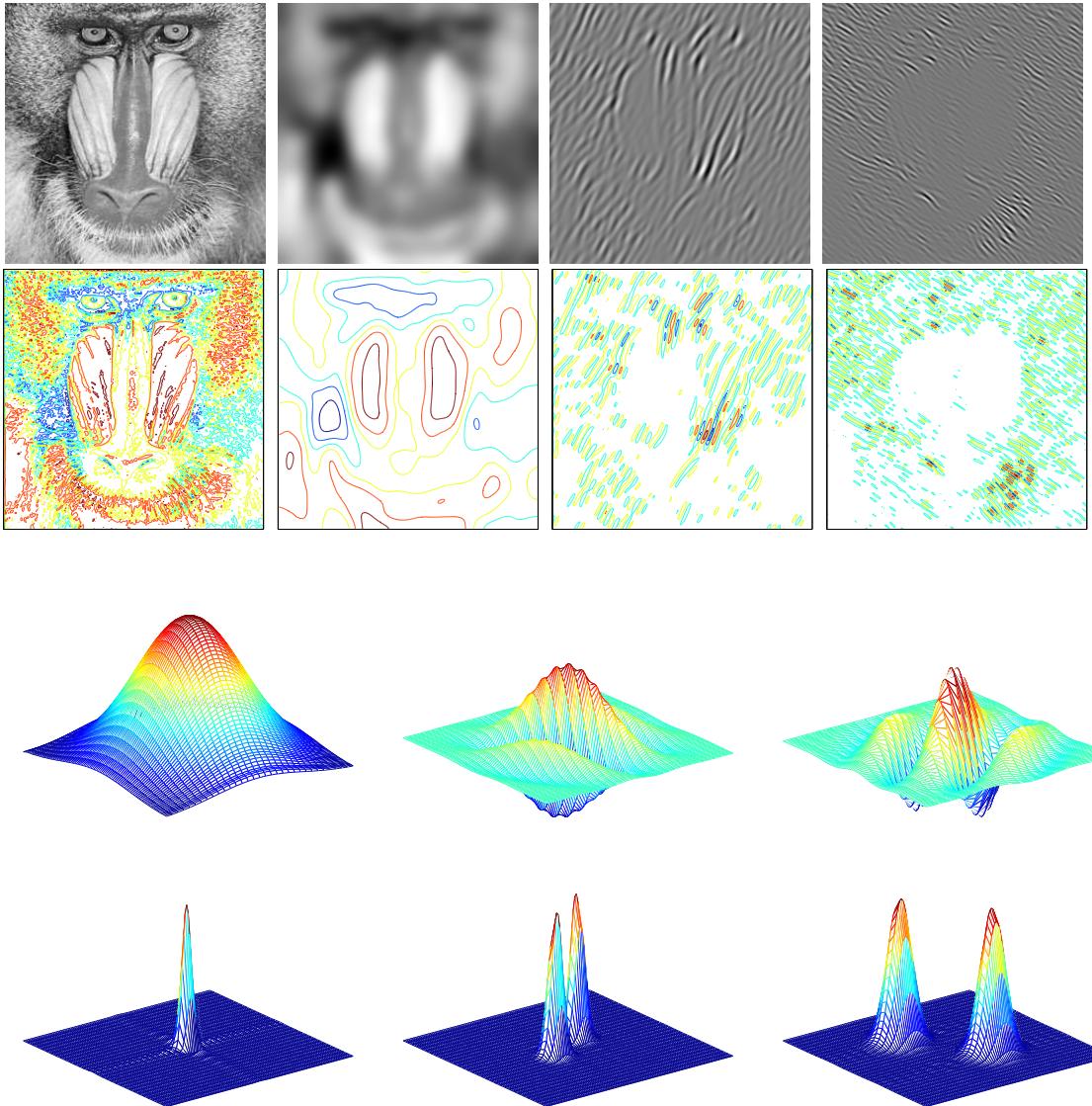
Η σημασία των παραπάνω σχέσεων είναι ότι τα φίλτρα Gabor καταφέρνουν ταυτόχρονα να δώσουν μέγιστη πληροφορία στο χωρικό και στο φασματικό πεδίο με την ελάχιστη αβεβαιότητα. Έτσι, εξάγουν με βέλτιστο τρόπο πληροφορία τόσο ως προς τη δομή («τι») όσο και προς τη θέση («που»).

Με κατάλληλους μετασχηματισμούς βάσης, οποιαδήποτε εικόνα μπορεί να αναπαρασταθεί από ένα πεπερασμένο σύνολο στοιχειωδών 2-Δ Gabor συναρτήσεων κυματιδίων [70, 73, 176, 191, 243]. Η επέκταση Gabor (expansion) ανακατασκευάζει την πληροφορία της εικόνας  $f$  ως ένα σταθμισμένο γραμμικό συνδυασμό χωρικά μετατοπισμένων κυματιδίων σταθερού εύρους και γραμμικής απόστασης στη συχνότητα

$$I(x, y) = \sum_{x_i, y_i} \sum_{w_{1j}, w_{2j}} \alpha_{ij} g_k(x - x_i, y - y_i; w_{1j}, w_{2j}), \quad (2.5)$$

με  $x_i, y_i = hi$ ,  $w_{1j}, w_{2j} = Hj$ ,  $h, H = \text{ct}$ . Τα βάρη  $\alpha_{ij}$  αποτελούν το σύνολο των συντελεστών της επέκτασης. Παρ' όλο που πρόκειται για μη-օρθογωνική προβολή, η ανακατασκευή

μπορεί να είναι ακριβής, αν τα φίλτρα διαμορφώνουν ένα «στενό πλαίσιο» (tight wavelet frame) [176] ή αν οι συντελεστές εκτιμηθούν έτσι, ώστε να αντιστοιχούν σε μια πλήρη αναπαράσταση [19, 70]. Για μικρή επικάλυψη μεταξύ τους οι συντελεστές προσεγγίζονται στενά από τις αποκρίσεις των κυματιδίων-φίλτρων  $a_{ik} \approx (I * g_k)(x_i, y_i)$  [36, 204].



Σχήμα 2.2: Φίλτρα Gabor και ζωνοπερατές συνιστώσες. 1η σειρά: αρχική εικόνα ( $256 \times 256$ ) και έξοδοι τριών πραγματικών Gabor φίλτρων σε διαφορετικές κλίμακες και προσανατολισμούς. 2η σειρά: επιπεδοσύνολα (level curves) σε έξι στάθμες. 3η και 4η σειρά: κρουστικές αποκρίσεις και αποκρίσεις συχνότητας των φίλτρων σε 3-Δ προοπτικές απεικονίσεις. Χαρακτηριστικά των τριών φίλτρων (κάθετη συχνότητα  $u_c$ , οριζόντια συχνότητα  $v_c$ , εύρος ζώνης):  $([0, 0, 27], [13, 7, 20], [-15, 21, 11])$ .

Μια τέτοια αναπαράσταση μπορεί να θεωρηθεί 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 μετατρέπεται σε μια 2-Δ συνάρτηση Δέλτα, ο μετασχηματισμός είναι στην ουσία δειγματοληψία

στο χωρικό πεδίο

$$\lim_{a,b \rightarrow \infty} [g(x,y|a,b) * I(x,y)] = \mathcal{F}\{I\}(u,v), \quad (2.6)$$

$$\lim_{a,b \rightarrow 0} [g(x,y|a,b) * I(x,y)] = \delta(x,y) * f(x,y), \quad (2.7)$$

όπου  $\mathcal{F}\{\cdot\}$  ο Fourier μετασχηματισμός της εικόνας  $I(x,y)$  και  $g(x,y|a,b)$  η απόκριση (2.1) με παραμέτρους  $(a,b)$ . Με χρήση τέτοιων πυραμίδων Gabor μπορεί να επιτευχθεί πλήρης αναπαράσταση, περιγραφή, αναγνώριση, συμπίεση και ανακατασκευή της οπτικής πληροφορίας [73].

Η εφαρμογή του μιγαδικού φίλτρου Gabor (2.1) ισοδυναμεί με φιλτράρισμα με το πραγματικό και το φανταστικό μέρος του  $h(x,y) = h_c(x,y) + jh_s(x,y)$  [34]. Τα δύο φίλτρα αποτελούν ένα ζεύγος ορθογωνισμού φάσης όπου το πραγματικό έχει άρτια και το φανταστικό περιττή συμμετρία ως προς τον προσανατολισμό τους  $\theta$ . Οι δύο εκδοχές της φιλτραρισμένης εικόνας έχουν διαφορά φάσης  $\pi/2$  κατά μήκος της  $\theta$ .

Στο Σχ. 2.2 φαίνεται η εφαρμογή σε μια εικόνα υφής τριών πραγματικών Gabor φίλτρων σε διαφορετικές χωρικές συχνότητες (ή κλίμακες και προσανατολισμούς). Οι συμμετρικοί λοβοί στο Fourier πεδίο των αποκρίσεων συχνότητας οφείλονται στη χρηση του πραγματικού μέρους, της άρτιας δηλαδή συνιστώσας των φίλτρων. Η συνάρτηση Gauss αποτελεί ένα γενικευμένο Gabor με  $(u_c = v_c = 0)$  που λειτουργεί ως βαθυπερατό φίλτρο ομαλοποίησης. Η απόκριση του είναι αναγκαία για την ανακατασκευή της αρχικής εικόνας από υπέρθεση των επιμέρους ζωνοπερατών συνιστωσών της.

### 2.2.3 Συστοιχίες Gabor φίλτρων

Τα φίλτρα Gabor χρησιμοποιήθηκαν συστηματικά για κατάτμηση, διάκριση και κατηγοριοποίηση υφής, οι παράμετροί τους (κλίμακα, προσανατολισμός, εύρος, χωρική τοποθέτηση) ερευνήθηκαν και συστηματοποιήθηκαν [35, 86] και παρουσιάστηκε η δυνατότητα αναζήτησης πεπερασμένων συνιστωσών υφής σε απλές εικόνες με χρήση ανάλογου αριθμού φίλτρων [34, 36]. 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 αλλά και ο προσανατολισμός δίνονται από αναλυτικές σχέσεις [37, 120, 122, 124, 250].

Για την ανάλυση πληροφορίας ευρείας ζώνης της υφής η εικόνα χωρίζεται σε απομονωμένες στενοζωνικές συνιστώσες (narrowband components) των οποίων το φασματικό περιεχόμενο είναι περιορισμένο σε μια μικρή περιοχή συχνοτήτων μέσω πολλαπλών καναλιών χωρικών συχνοτήτων τύπου κυματιδίων. Ο διαχωρισμός γίνεται από ένα σύνολο φίλτρων κατάλληλης δομής και διάταξης έτσι, ώστε να καλύπτεται επαρκώς όλο το εύρος των χωρικών συχνοτήτων. Ένα επιπλέον πλεονέκτημα είναι ο περιορισμός του ύφορύβου ανά ζώνη και η ομαλοποίηση των όποιων ενεργειακών μετρήσεων στην έξοδο των φίλτρων. Το πλαίσιο πολλαπλών φίλτρων, παράλληλης επεξεργασίας και εξαγωγής χαρακτηριστικών ανά ζώνη συχνοτήτων είναι γνωστό ως πολυζωνική ανάλυση.

Τα συνήθη χριτήρια 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 στενού εύρους ζώνης και στο σύνολο τους να καλύπτουν πυκνά τις σημαντικές συχνότητες της εικόνας. Για το σκοπό αυτό επιλέγονται συνήθως 2-Δ Gabor φίλτρα λόγω της βέλτιστης αξιεβαίοτητας ως προς τη χωροσυχνότική τοποθέτηση τους (2.5). Χρησιμοποιούνται ισοτροπικά φίλτρα, σε διαφορετικές κλίμακες και προσανατολισμούς, με κάποιο κανόνα σύνδεσης του εύρους ζώνης με την κεντρική συχνότητα του καθενός. Κατά κανόνα σε μεγαλύτερες κεντρικές συχνότητες αντιστοιχίζεται μεγαλύτερο εύρος ζώνης.

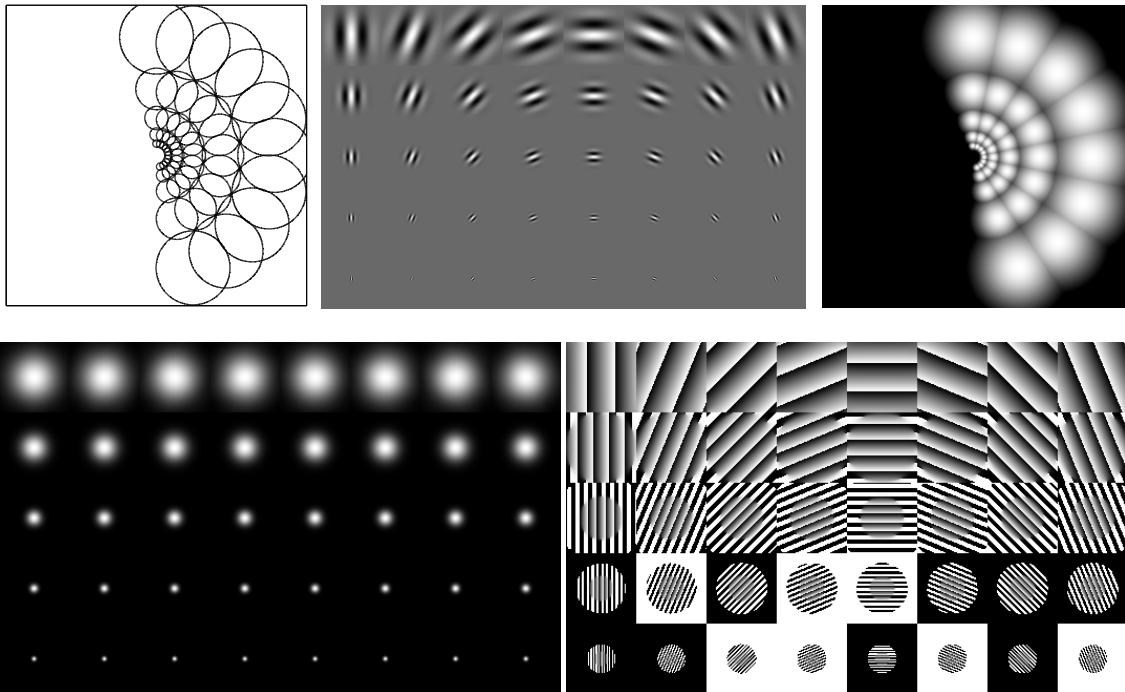
Η διάταξη των φίλτρων είναι ακτινική, με κεντρικές συχνότητες σε γεωμετρική πρόοδο και πυκνή κάλυψη του φασματικού πεδίου. Με κίνητρο και τη διάταξη των κυττάρων του φλοιού [96], η χρήση οκτάβων για το εύρος των καναλιών προσφέρει βελτιωμένο εντοπισμό στο χωρικό πεδίο (μεταβάλλοντας προοδευτικά τη χωρική υποστήριξη και το ενεργό εύρος του φίλτρου). Για κάθε κανάλι  $k$  οι ελεύθερες παράμετροι  $[\sigma_k, \vec{w}_k]$  καθορίζονται από το σχεδιασμό της συστοιχίας. Για φίλτρα που τέμνονται σε λόγο  $n$  από τις κορυφές τους ( $n$ -peak) η απόκλιση της Gaussian περιβάλλουσας είναι

$$\sigma_k = \frac{\sqrt{-2 \ln n}}{|\vec{w}_k|} \frac{2^B + 1}{2^B - 1}, \quad |\vec{w}_k| = \sqrt{u_{ck}^2 + v_{ck}^2}, \quad (2.8)$$

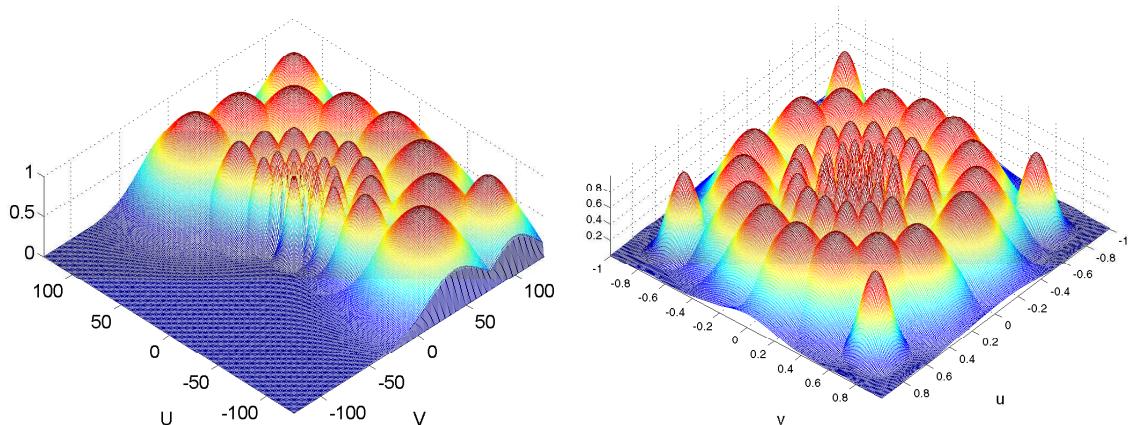
όπου  $B$  το ακτινικό εύρος (radial  $n$ -peak) σε οκτάβες και  $|\vec{w}_k|$  η κεντρική συχνότητα. Ένας τέτοιος σχεδιασμός (constant-Q) διατηρεί σταθερό το λόγο  $Q = |\vec{w}_k|/\sigma_k$  της κεντρικής συχνότητας προς το εύρος ζώνης σε κάθε φίλτρο  $k$  και εξασφαλίζει σταθερή, λογαριθμική απόσταση ανάμεσα σε διαδοχικά φίλτρα της ίδιας κατεύθυνσης  $\theta_k = \arctan(v_{ck}/u_{ck})$ . Το πλάτος της κεντρικής συχνότητας  $|\vec{w}_k|$  ταυτίζεται με την κλίμακα του φίλτρου, ενώ η κατεύθυνση της  $\theta_k$  με τον προσανατολισμό του.

Μια υλοποίηση για 40 φίλτρα σε οκτώ προσανατολισμούς και πέντε κλίμακες εύρους  $B = 1$  οκτάβα και τομή στο  $n = 1/2$  της κορυφής απεικονίζεται με τη μορφή των αποκρίσεων συχνότητας των φίλτρων στο Σχ. 2.3. Ένας τέτοιος σχεδιασμός οδηγεί σε ένα προκαλούμενο διαχωρισμό (partitioning) του πεδίου των χωρικών συχνοτήτων και σε μια προσέγγιση της εικόνας με λιγότερες συνιστώσες απ' ό,τι η επέκταση (2.5). Περισσότερες λεπτομέρειες για τον ακτινικό σχεδιασμό μπορούν να βρεθούν στο [122]. Η συμμετρική κάλυψη του πεδίου των συχνοτήτων από συστοιχία μιγαδικών και πραγματικών φίλτρων, φαίνεται από τις αποκρίσεις συχνότητας των φίλτρων στο Σχ. 2.4 ως προοπτικές 3-Δ απεικονίσεις. Τα φίλτρα είναι κανονικοποιημένα ως προς την  $L_1$  ενέργεια τους και συνοδεύονται από δύο υψηπερατά και ένα βαθυπερατό φίλτρο για την ολοκλήρωση της περιγραφής και την πληρέστερη κάλυψη και ανακατασκευή της πληροφορίας. Ανάλογα ορίζονται και συστοιχίες ισοτροπικών φίλτρων σε γραμμική διάταξη με ίδιο εύρος ζώνης και απόσταση ανά κατεύθυνση. Οι «γραμμικές» συστοιχίες (linear filterbanks) αντιμετωπίζουν με τον ίδιο τρόπο τις υψηλές και χαμηλές ζώνες συχνοτήτων, αλλά δεν εξασφαλίζουν απαραίτητα την ανακατασκευή της αρχικής εικόνας από τις επιμέρους ζωνοπερατές εκδοχές της.

Οι ζωνοπερατές συνιστώσες στην έξοδο μιας συστοιχίας έχουν έμφυτες τις έννοιες της κλίμακας και της κατεύθυντικότητας οι οποίες καθορίζονται από την Gaussian απόκλιση  $\sigma_k$



Σχήμα 2.3: Συστοιχία φίλτρων Gabor και πολυσυχνοτικός διαχωρισμός. Πάνω σειρά: χωρικά προφίλ (πραγματικό μέρος κανονικοποιημένο στο  $[0, 1]$ ) φίλτρων (μεσαία εικόνα) και ακτινική διάταξη των αποκρίσεων συχνότητας. Τα 40 φίλτρα ακτινικού εύρους  $B = 1$  οκτάβας καλύπτουν 5 κλίμακες (δακτύλιοι) και 8 προσανατολισμούς (ακτίνες). Σε κάθε δακτύλιο τα γειτονικά φίλτρα τέμνονται στο μισό της τιμής κορυφής τους (αριστερά εικόνα). Κάτω σειρά: Μέτρο και φάση των μιγαδικών φίλτρων που αντιστοιχούν στη χωρική έκταση, την κλίμακα και τον προσανατολισμό τους.



Σχήμα 2.4: Αποκρίσεις συχνότητας (3-Δ προοπτική απεικόνιση) συστοιχίας 2-Δ μιγαδικών (δεξιά) και πραγματικών (αριστερά) Gabor φίλτρων (40 φίλτρα, σε 8 κλίμακες και 5 προσανατολισμούς, ένα βαθυπεραστό και δύο υψηπεραστά). Η συμμετρική κάλυψη στα πραγματικά οφείλεται στη συμμετρία του Fourier μετασχηματισμού για άρτιες συναρτήσεις.

και την κεντρική συχνότητα των καναλιών  $\vec{w}_k$ . Με αυτό τον τρόπο μια πολυκλιμακωτή (multiscale) οργάνωση υποδηλώνεται από τα φίλτρα σε κάθε κατεύθυνση [36, 138, 244], με τις πιο λεπτομερείς κλίμακες της εικόνας να αντιστοιχούν στις ζώνες μεγαλύτερων συχνοτήτων και τις πιο αδρές στις ζώνες χαμηλών συχνοτήτων και μικρού εύρους.

## 2.3 Μοντέλα διαμόρφωσης, ενεργειακοί τελεστές και αποδιαμόρφωση

Οι ζωνοπερατές συνιστώσες της υφής στην έξοδο μιας συστοιχίας Gabor φίλτρων, είναι εικόνες υφής τοπικά στενοζωνικές κατά την έννοια που περιγράφεται στο [37]. Τέτοιες συνιστώσες αντιστοιχούν στις χωρικές διαμορφώσεις των τιμών φωτεινότητας της εικόνας σε ένα στενό εύρος συχνοτήτων και μπορούν να μοντελοποιηθούν με χωρικά AM-FM σήματα

$$f(x, y) = a(x, y) \cos[\phi(x, y)], \quad \vec{w}(x, y) = \nabla\phi(x, y), \quad (2.9)$$

δηλαδή μη-στάσιμα (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  $\vec{w}(x, y) = (\omega_1, \omega_2)(x, y)$  περιέχει πληροφορία για τις τοπικά αναδυόμενες συχνότητες. Το πλάτος και οι συχνότητες είναι ομαλά, τοπικά στενής ζώνης σήματα. Τέτοια μοντέλα διαμόρφωσης έχουν εφαρμοστεί σε ένα πλήθος προβλημάτων όρασης υπολογιστών και ανάλυσης εικόνων [2, 37, 74, 120, 124, 125, 198, 234]. Το πρόβλημα σε τέτοια μοντέλα διαμόρφωσης είναι η εκτίμηση των 2-Δ σημάτων πλάτους και συχνότητας με υπολογιστικούς αλγόριθμους χαμηλής πολυπλοκότητα και μικρού σφάλματος εκτίμησης. Μια προσέγγιση βασίζεται σε διαχωρισμό της ενέργειας τέτοιων σημάτων όπως την εκτιμάει ένας διαφορικός μη-γραμμικός τελεστής.

### 2.3.1 Μονοδιάστατος ενεργειακός τελεστής

Ο μη-γραμμικός διαφορικός τελεστής  $\Psi$  που προτάθηκε από τον Teager [281] και αναπτύχθηκε συστηματικά από τον Kaiser [148, 149] για 1-Δ σήματα συνεχούς χρόνου  $f(t)$  ορίζεται ως:

$$\Psi(f)(t) \triangleq [f'(t)]^2 - f(t)f''(t) \quad (2.10)$$

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

$$\Psi_d(f)(n) \triangleq f^2(n) - f(n-1)f(n+1) \quad (2.11)$$

για σήματα διακριτού χρόνου  $f(n)$ . Οι τελεστές  $\Psi$  και  $\Psi_d$  χαρακτηρίζονται ως ενεργειακοί τελεστές υπό την έννοια ότι εντοπίζουν την «ενέργεια» απλών αρμονικών ταλαντώσεων ημιτονικών σημάτων. Αυτή η ενεργειακή ποσότητα είναι ανάλογη τόσο του τετραγώνου του πλάτους όσο και του τετραγώνου της συχνότητας της ταλάντωσης!

Οι ενεργειακοί τελεστές μπορούν να εκτιμήσουν στιγμιαία την περιβάλλοντα διαμορφώμενων σημάτων κατά πλάτος ή συχνότητα [202], ενώ για γενικά AM-FM σήματα της μορφής  $f(t) = a(t) \cos[\phi(t)]$  η έξοδος τους είναι το ενεργειακό γινόμενο

$$\Psi[a(t) \cos(\phi(t))] \approx a^2(t)\omega^2(t), \quad (2.12)$$

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 δε μεταβάλλονται πολύ γρήγορα ή σε μεγάλο

εύρος σε σχέση με το φέρον. Εφαρμόζοντας, επιπλέον, τον  $\Psi$  στην παράγωγο  $f'(t)$  προκύπτει ένας 1-Δ αλγόριθμος αποδιαμόρφωσης AM-FM σημάτων που διαχωρίζει την ενέργεια (2.12) σε συντελεστές πλάτους και συχνότητας και εκτιμάει την περιβάλλουσα  $|a(t)|$  και τη στιγμιαία συχνότητα  $\omega(t)$  [38, 200–202]. Περισσότερα για την αποδιαμόρφωση χρονικών AM-FM σημάτων μέσω του μονοδιάστατου τελεστή παρουσιάζονται στο κεφάλαιο 8.

### 2.3.2 Ενεργειακός τελεστής για 2-Δ σήματα

Μια επέκταση του 1-Δ ενεργειακού τελεστή προτάθηκε στο [198] σε ένα γενικότερο πλαίσιο πολυδιάστατων ενεργειακών τελεστών και 2-Δ τελεστών για εικόνες. Αν  $f(\vec{x})$  ένα δισδιάστατο πραγματικό σήμα με  $\vec{x} = (x, y) \in \mathbb{R}^2$  τότε ο 2-Δ ενεργειακός τελεστής ορίζεται ως:

$$\Psi(f)(\vec{x}) \triangleq \|\nabla f(\vec{x})\|^2 - f(\vec{x})\nabla^2 f(\vec{x}), \quad \vec{x} = (x, y) \in \mathbb{R}^2, \quad (2.13)$$

όπου  $\nabla f$  η κλίση (gradient) της  $f$ ,  $\|\cdot\|$  η Ευκλείδια  $L_2$  νόρμα, και  $\nabla^2 f = \Delta f = \partial^2 f / \partial x^2 + \partial^2 f / \partial y^2$  η Laplacian του  $f$ . Ο τελεστής  $\Psi(f)$  στο δισδιάστατο σήμα  $f$  μπορεί εξ ορισμού να εκφραστεί ως το άθροισμα δύο μονοδιάστατων τελεστών στις δύο κατευθύνσεις

$$\Psi(f) = \left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2 - f \left( \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right) = \sum_{k \in \{x, y\}} \Psi_k(f) \quad (2.14)$$

με  $\Psi_k(f) = f_k^2 - f f_{kk}$ . Έτσι, η έξοδος του  $\Psi$  είναι το άθροισμα των ενεργειακών συνιστώσων του 1-Δ τελεστή στην οριζόντια και κάθετη διεύθυνση. Ένα σύνολο ιδιοτήτων του 2-Δ τελεστή  $\Psi$  μπορεί να βρεθεί στο [198].

### 2.3.3 Επέκταση για μιγαδικά 2-Δ σήματα

Η επέκταση του 2-Δ τελεστή για μιγαδικά σήματα γίνεται με βάση ένα γενικευμένο τελεστή για διανυσματικά σήματα [198]. Ο τελεστής  $C$  για ένα μιγαδικό σήμα  $f(\vec{x}) = (f_r + j f_i)(\vec{x})$  ορίστηκε ως:

$$C(f)(\vec{x}) \triangleq \|\nabla f(\vec{x})\|^2 - \text{Re}[f^*(\vec{x})\nabla^2 f(\vec{x})], \quad (2.15)$$

όπου το  $*$  ορίζει το μιγαδικό συμπλήρωμα και  $f_r = \text{Re}(f)$  και  $f_i = \text{Im}(f)$ . Εύκολα διαπιστώνεται ότι προκύπτει

$$C(f) = \Psi(f_r) + \Psi(f_i) = \Psi[\text{Re}(f)] + \Psi[\text{Im}(f)]. \quad (2.16)$$

Έτσι, η εφαρμογή του ενεργειακού τελεστή σε μιγαδικά σήματα καταλήγει σε ανεξάρτητη ανάλυση του πραγματικού και του φανταστικού τους μέρους.

Ενδιαφέρον παρουσιάζει η σύγκριση της (2.16) με το αποτέλεσμα της απευθείας εφαρμογής του  $\Psi$  της σχέσης (2.13) στο μιγαδικό σήμα  $f(\vec{x})$ . Με απλούς υπολογισμούς προκύπτει

$$\Psi(f) = \Psi(f_r + j f_i) = \Psi(f_r) - \Psi(f_i) + j \nabla^2[f_r f_i]. \quad (2.17)$$

Συγκρίνοντας την (2.16) με την (2.17) φαίνεται ότι  $C(f) \neq \text{Re}[\Psi(f)]$ . Επομένως, ο ενεργειακός τελεστής για μιγαδικά σήματα, που είναι μια πραγματική ποσότητα, δεν είναι ισοδύναμος με το πραγματικό μέρος της εφαρμογής του ενεργειακού τελεστή στο σήμα.

### 2.3.4 Χωρικά AM-FM και διαχωρισμός ενέργειας

Έστω ένα πραγματικό 2-Δ ημιτονοειδές σήμα, διαμορφωμένο κατά πλάτος και συχνότητα της μορφής (2.9)  $f(x, y) = a(x, y) \cos[\phi(x, y)]$ , όπου  $a(x, y)$  το χωρικά μεταβαλλόμενο πλάτος,  $\phi(x, y)$  το σήμα φάσης και

$$\vec{\omega}(x, y) \triangleq \nabla\phi(x, y) = [\omega_1(x, y), \omega_2(x, y)] \quad (2.18)$$

το διάνυσμα των στιγμιαίων χωρικά μεταβαλλόμενων γωνιακών συχνοτήτων

$$\omega_1(x, y) \triangleq \frac{\partial\phi}{\partial x}(x, y) , \quad \omega_2(x, y) \triangleq \frac{\partial\phi}{\partial y}(x, y). \quad (2.19)$$

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 και  $q_k(\vec{x}) \in [-1, 1]$  το σήμα διαμόρφωσης συχνότητας με  $\omega_{m,k}$  τη μέγιστη απόκλιση του  $\omega_k$  από την κεντρική του τιμή.

Εφαρμογή του  $\Psi$  σε ένα τέτοιο AM-FM σήμα  $f$  καταλήγει προσεγγιστικά στο γινόμενο

$$\Psi[a \cos(\phi)] \approx a^2 \|\vec{\omega}\|^2. \quad (2.20)$$

Το σφάλμα προσέγγισης είναι αμελητέο υπό ρεαλιστικές συνθήκες [198] 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 (2.20) 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  $\Psi$  στα μερικά διαφορικά  $\partial f / \partial x$  και  $\partial f / \partial y$  του  $f$  στις δύο διευθύνσεις

$$\Psi\left(\frac{\partial f}{\partial x}\right) \approx a^2 \omega_1^2 \|\vec{\omega}\|^2 , \quad \Psi\left(\frac{\partial f}{\partial y}\right) \approx a^2 \omega_2^2 \|\vec{\omega}\|^2. \quad (2.21)$$

Συνδυάζοντας τις (2.20),(2.21) προκύπτει ο μη-γραμμικός συνεχής αλγόριθμος διαχωρισμού ενέργειας (Continuous Energy Separation Algorithm ή CESA)

$$\sqrt{\frac{\Psi(\partial f / \partial x)}{\Psi(f)}} \approx |\omega_1(x, y)| , \quad \sqrt{\frac{\Psi(\partial f / \partial y)}{\Psi(f)}} \approx |\omega_2(x, y)|, \quad (2.22)$$

$$\frac{\Psi(f)}{\sqrt{\Psi(\partial f / \partial x) + \Psi(\partial f / \partial y)}} \approx |a(x, y)|. \quad (2.23)$$

Ο 2-Δ CESA αποτελεί επέκταση του 1-Δ αλγορίθμου αποδιαμόρφωσης χρονικά μεταβαλλόμενων σημάτων [38, 201, 202] 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, δηλαδή 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.

### 2.3.5 Διακριτοί τελεστές και αποδιαμόρφωση εικόνων

Για την ενεργειακή ανάλυση και την αποδιαμόρφωση δισδιάστατων σημάτων και συγκεκριμένα εικόνων, οι οποίες ορίζονται σε ένα 2-Δ πεπερασμένο χωρικό πλέγμα, χρησιμοποιούνται οι διακριτές εκδοχές των προηγούμενων σχέσεων. Η ερμηνεία (2.14) του  $\Psi$  επιτρέπει την επέκτασή του σε σήματα διακριτού χώρου (discrete-space)  $f(m, n)$ . Αντικαθιστώντας τις ενεργειακές συνιστώσες με τις εξόδους του 1-Δ διακριτού τελεστή  $\Psi_d$  προκύπτει ο 2-Δ διακριτός ενεργειακός τελεστής

$$\begin{aligned}\Psi_d(f)(m, n) &\triangleq \Psi_{d,1}(f)(m, n) + \Psi_{d,2}(f)(m, n) \\ &= 2f^2(m, n) - f(m-1, n)f(m+1, n), -f(m, n-1)f(m, n+1)\end{aligned}\quad (2.24)$$

όπου ο  $\Psi_{d,1}$  ενεργεί κάθετα σε όλες τις στήλες  $0 \leq m \leq M$ , ενώ ο  $\Psi_{d,2}$  ορίζόντια σε όλες τις γραμμές  $0 \leq n \leq N$  του  $f$ . Η έκφραση (2.24) είναι όμοια με το διακριτό τελεστή που αναπτύχθηκε στο [217] για βελτίωση εικόνων (image enhancement) και ανίχνευση ακμών (edge detection). Αντικαθιστώντας τα διαφορικά στο  $\Psi$  με εναλλακτικά 2-Δ σχήματα χωρικών διαφορών μπορούν να προκύψουν διαφορετικά σχήματα 2-Δ διακριτών ενεργειακών τελεστών.

Έστω ένα διακριτό AM-FM σήμα στο διακριτό  $(m, n)$  πλέγμα σημείων

$$f(m, n) = a(m, n) \cos[\phi(m, n)] \quad (2.25)$$

με κάθετη στιγμιαία συχνότητα  $\Omega_1(m, n) \triangleq \partial\phi/\partial m$  σε ακτίνια/δείγμα (radians/sample) και ορίζόντια  $\Omega_2(m, n) \triangleq \partial\phi/\partial n$  με εύρος τιμών  $[-\pi, \pi]$  και σταθερό πρόσημο για κάθε  $(m, n)$ . Υπό 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 [198, 202], τότε

$$\Psi_d[a(m, n) \cos(\phi(m, n))] \approx a^2(m, n) (\sin^2[\Omega_1(m, n)] + \sin^2[\Omega_2(m, n)]). \quad (2.26)$$

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) :

$$\arcsin \left( \sqrt{\frac{\Psi_d[f(m+1, n) - f(m-1, n)]}{4\Psi_d[f(m, n)]}} \right) \approx |\Omega_1(m, n)|, \quad (2.27)$$

$$\arcsin \left( \sqrt{\frac{\Psi_d[f(m, n+1) - f(m, n-1)]}{4\Psi_d[f(m, n)]}} \right) \approx |\Omega_2(m, n)|, \quad (2.28)$$

$$\frac{2\Psi_d[f(m, n)]}{\sqrt{\Psi_d[f(m+1, n) - f(m-1, n)] + \Psi_d[f(m, n+1) - f(m, n-1)]}} \approx |a(m, n)|. \quad (2.29)$$

Ο DESA μπορεί να εκτιμήσει σε κάθε σημείο την περιβάλλουσα και το μέτρο των στιγμιαίων συχνοτήτων του χωρικού AM-FM σήματος. Ο περιορισμός του είναι ότι  $0 \leq |\Omega_1|, |\Omega_2| \leq$

$\pi/2$ , δηλαδή εκτιμάει συχνότητες μέχρι το ένα τέταρτο της συχνότητας δειγματοληψίας. Τα πρόσημά 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 συνιστώσων AM-FM σημάτων είναι η απλότητα, η αποδοτικότητα, η μικρή υπολογιστική πολυπλοκότητα, η εξαιρετική, σχεδόν στιγμιαία, χωρική ανάλυση σε παράθυρα  $3 \times 3$  και η ικανότητά του να εντοπίζει χωρικές δομές διαμόρφωσης. Διαφορετικές προσεγγίσεις για την αποδιαμόρφωση χωρικά διαμορφωμένων σημάτων και εικόνων βασίζονται στο μετασχηματισμό Hilbert [64] και σε επεκτάσεις του για δύο διαστάσεις [74, 127, 175], όπως είναι ο πολυδιάστατος κατευθυντικός Hilbert [125]. Συγκεκριμένα, κατασκευάζεται η «αναλυτική» εικόνα ως μιγαδική επέκταση του πραγματικού σήματος, με χρήση του μέτρου και της κλίσης της οποίας λαμβάνονται υπολογιστικές εκτιμήσεις του πλάτους και του διανύσματος συχνοτήτων του σήματος [126].

Στη μονοδιάστατη περίπτωση [229], δεδομένου ενός στενοζωνικού σήματος  $f(x)$  που μοντελοποιείται από μια συνάρτηση  $a(x) \cos(\omega_0 x + \phi(x))$ , μια βέλτιστη λύση για το πρόβλημα της εκτίμησης των παραμέτρων διαμόρφωσης δόθηκε με βάση το Hilbert μετασχηματισμό και το αναλυτικό σήμα το οποίο ελαχιστοποιεί τη μέση τετραγωνική ενέργεια  $\int |a'(x)|^2 dx$  της παραγώγου του πλάτους. Στο [245] έγιναν συγχρίσεις ανάμεσα στους δύο 1-Δ αλγορίθμους και δείχθηκε ότι οι εκτιμήσεις του ESA για το στιγμιαίο πλάτος και τη συχνότητα του  $f(x)$  φέρουν ένα σφάλμα μοντελοποίησης συγχρίσιμο με αυτό του Hilbert μετασχηματισμού. Ωστόσο, ο ESA έχει καλύτερη χωροχρονική ανάλυση και μικρότερη πολυπλοκότητα.

### 2.3.6 Ανάλυση του ενεργειακού τελεστή 2-Δ σε Θόρυβο

Στην ενότητα αυτή παρουσιάζονται κάποια θεωρητικά αποτελέσματα σχετικά με την αναμενόμενη συμπεριφορά των ενεργειακών τελεστών σε 2-Δ σήματα και εικόνες. Μια τέτοια ανάλυση είναι σημαντική για την κατανόηση και την πρόβλεψη των ιδιοτήτων του ενεργειακού τελεστή σε προβλήματα εικόνων με θόρυβο, ενώ ανάλογη ανάλυση έχει γίνει και για τον 1-Δ Teager-Kaiser τελεστή στο [38]. Σε αυτό το πλαίσιο αναπτύσσονται οι βασικές στατιστικές ιδιότητες (ροπές κατώτερης τάξης) της εφαρμογής του (2.13) σε ένα τυχαίο 2-Δ σήμα  $N(x, y)$  και εξάγεται μια σχέση ανάμεσα στο εύρος ζωνοπερατού φιλτραρίσματος και στη μέση τιμή της εξόδου του τελεστή.

#### 2.3.6.1 Στατιστικά

Το  $N(x, y)$  θεωρείται ότι είναι στατική με την ευρεία έννοια (wide-sense stationary 'h WSS), μηδενικής μέσης τιμής (zero-mean) Gaussian τυχαία διαδικασία, με συνάρτηση αυτο-

συσχέτισης  $R(\tau_1, \tau_2)$  και φασματική πυκνότητα ισχύος (power spectral density)

$$\Phi(\vec{\omega}) = \int_{\mathbb{R}^2} R(\vec{\tau}) e^{-j\vec{\tau} \cdot \vec{\omega}} d\vec{\tau}, \quad \vec{\tau} = (\tau_1, \tau_2), \quad \vec{\omega} = (\omega_x, \omega_y) \quad (2.30)$$

$$\Phi(\omega_x, \omega_y) = \int_{\mathbb{R}^2} R(\tau_1, \tau_2) e^{(-j\omega_x \tau_1 - j\omega_y \tau_2)} d\tau_1 d\tau_2. \quad (2.31)$$

Επειδή η  $N(x, y)$  είναι WSS Gaussian διαδικασία, οι μερικές παράγωγοι  $N_x, N_{xx}, Ny, Nyy$ ,  $Nxy$ , όπου οι υποδείκτες δηλώνουν μερική παραγώγιση 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 των  $N$  και  $Nyy$  [229, σελ. 133]. Έτσι, η εφαρμογή του τελεστή (2.14) στη διαδικασία  $N$

$$\Psi(N) = (N_x)^2 - NN_{xx} + (N_y)^2 - NN_{yy} \quad (2.32)$$

είναι το άθροισμα τεσσάρων ανεξάρτητων διαδικασιών. Έστω συμβολικά

$$R^{(m,n)}(\tau_1, \tau_2) = \frac{\partial^k R(\tau_1, \tau_2)}{\partial \tau_1^m \partial \tau_2^n}, \quad k = m + n \quad (2.33)$$

και τα  $\mathbb{E}[\cdot]$  και  $\text{Var}[\cdot]$  δηλώνουν στατιστική αναμενόμενη τιμή (expectation) και μεταβλητότητα (variance). Ισχύουν τα εξής [229, σελ. 319]:

$$\text{Var}[N] = R(0, 0) = \gamma_0^2 \quad (2.34)$$

$$\text{Var}[N_x] = -R^{(2,0)}(0, 0) = \alpha_1^2 \quad (2.35)$$

$$\text{Var}[N_y] = -R^{(0,2)}(0, 0) = \beta_1^2 \quad (2.36)$$

$$\text{Var}[N_{xx}] = R^{(4,0)}(0, 0) = \alpha_2^2 \quad (2.37)$$

$$\text{Var}[N_{yy}] = R^{(0,4)}(0, 0) = \beta_2^2 \quad (2.38)$$

$$\mathbb{E}[NN_{xx}] = R^{(2,0)}(0, 0) = -\alpha_1^2 \quad (2.39)$$

$$\mathbb{E}[NN_{yy}] = R^{(0,2)}(0, 0) = -\beta_1^2 \quad (2.40)$$

$$\mathbb{E}[N_{xx}N_{yy}] = R^{(2,2)}(0, 0) = \gamma_1^2. \quad (2.41)$$

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

Με βάση τα παραπάνω για τις ροπές του  $\Psi$  έχουμε:

$$\begin{aligned} \mathbb{E}[\Psi(N)] &= \mathbb{E}[\Psi_x(N)] + \mathbb{E}[\Psi_y(N)] \\ &= \text{Var}[N_x] - \mathbb{E}[NN_{xx}] + \text{Var}[N_y] - \mathbb{E}[NN_{yy}] \\ &= 2\alpha_1^2 + 2\beta_1^2 \end{aligned} \quad (2.42)$$

με  $\Psi_x, \Psi_y$  τους κατευθυντικούς ενεργειακούς τελεστές ως προς  $x$  και  $y$ , αντιστοίχως. 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 (2.31)

$$\begin{aligned} \mathbb{E}[\Psi(N)] &= -R^{(2,0)}(0, \tau_2) - R^{(0,2)}(\tau_1, 0) \\ &= \frac{1}{\pi} \int_{\mathbb{R}^2} \omega_y^2 \Phi(\omega_x, \omega_y) d\omega_x d\omega_y + \frac{1}{\pi} \int_{\mathbb{R}^2} \omega_x^2 \Phi(\omega_x, \omega_y) d\omega_x d\omega_y \\ &= \frac{1}{\pi} \int_{\mathbb{R}^2} |\vec{\omega}|^2 \Phi(\omega_x, \omega_y) d\omega_x d\omega_y \end{aligned} \quad (2.43)$$

λόγω της σχέσης αυτοσυσχέτισης και φάσματος ισχύος και των ιδιοτήτων παραγώγισης του μετασχηματισμού Fourier. Επομένως, η μέση τιμή του  $\Psi(N)$  αυξάνεται γραμμικά με τη φασματική ενεργειακή μεταβλητή της του  $N$ . Κάτι τέτοιο μπορεί να αποτελεί πρόβλημα για πολύ όμορυθμη σήματα.

Η αυτοσυσχέτιση  $R_{\Psi(N)}(\tau_1, \tau_2)$  μπορεί να βρεθεί χρησιμοποιώντας τη φόρμουλα του Isserlis [134] για την αναμενόμενη τιμή γινομένων τεσσάρων τυχαίων μεταβλητών μηδενικής μέσης τιμής

$$\mathbb{E}[Y_1 Y_2 Y_3 Y_4] = \mathbb{E}[Y_1 Y_4] \mathbb{E}[Y_1 Y_3] + \mathbb{E}[Y_1 Y_3] \mathbb{E}[Y_2 Y_4] + \mathbb{E}[Y_1 Y_2] \mathbb{E}[Y_3 Y_4], \quad (2.44)$$

όπως και στο [38], με σκοπό τον περιορισμό των ροπών πολλαπλών συντελεστών:

$$\begin{aligned} R_{\Psi(N)}(\tau_1, \tau_2) &= R_{\Psi_x(N)}(\tau_1, \tau_2) + R_{\Psi_y(N)}(\tau_1, \tau_2) \\ &\quad + \underbrace{\mathbb{E}[\Psi_x(N(x + \tau_1, y)) \Psi_y(N(x, y))]}_{\mathsf{E}_1} + \underbrace{\mathbb{E}[\Psi_x(N(x, y + \tau_2)) \Psi_y(N(x, y))]}_{\mathsf{E}_2}. \end{aligned} \quad (2.45)$$

Με λογισμούς στις αναμενόμενες τιμές των τελεστών το άθροισμα στο δεύτερο σκέλος με το συμβολισμό της (2.33) καταλήγει:

$$\begin{aligned} \mathsf{E}_1 + \mathsf{E}_2 &= 8\alpha_1^2 \beta_1^2 + 2[R^{(1,1)} + 2R^{(0,1)}R^{(2,1)} \\ &\quad + 2R^{(1,0)}R^{(1,2)} + R^{(2,0)}R^{(0,2)} + RR^{(2,2)}](\tau_1, \tau_2), \end{aligned} \quad (2.46)$$

ενώ οι όροι αυτοσυσχέτισης

$$R_{\Psi_x(N)}(\tau_1, \tau_2) = 4\alpha_1^4 + [3(R^{(0,2)})^2 + 4R^{(0,1)}R^{(0,3)} + RR^{(0,4)}](\tau_1, \tau_2), \quad (2.47)$$

$$R_{\Psi_y(N)}(\tau_1, \tau_2) = 4\beta_1^4 + [3(R^{(2,0)})^2 + 4R^{(1,0)}R^{(3,0)} + RR^{(4,0)}](\tau_1, \tau_2). \quad (2.48)$$

Το σύνολο των (2.45)-(2.48) δίνει τη συνάρτηση αυτοσυσχέτισης  $R_{\Psi(N)}$ . Η ανισότητα Cauchy-Schwarz [229] καθορίζει ότι

$$|R(\tau_1, \tau_2)| \leq \frac{1}{\pi} \int_{\mathbb{R}^2} \Phi(\omega_x, \omega_y) d\omega_x d\omega_y = R(0, 0), \quad (2.49)$$

από την οποία προκύπτει ότι 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  $R^{(1,0)}(0, 0) = R^{(1,0)}(0, 0) = 0$  αλλά και  $R^{(1,1)}(0, 0) = 0$ .

## Κεφάλαιο 2. Οπτική Υφή και Μικροδομές Διαμόρφωσης

Με χρήση και των (2.34)-(2.41), η μεταβλητότητα του  $\Psi(N)$  θα είναι

$$\begin{aligned} \text{Var}[\Psi(N)] &= R_{\Psi(N)}(0,0) - E^2[\Psi(N)] \\ &= [2R^{(2,0)}R^{(0,2)} + 3((R^{(0,2)})^2 + (R^{(2,0)})^2) + R(R^{(4,0)} + R^{(0,4)} + 2R^{(2,2)})](0,0) \\ &= 2\alpha_1^2\beta_1^2 + 3(\alpha_1^4 + \beta_1^4) + \gamma_0^2(2\gamma_1^2 + \alpha_2^2 + \beta_2^2) \end{aligned} \quad (2.50)$$

και θα αυξάνεται, όπως και η αναμενόμενη τιμή, με την παρουσία μεγαλύτερων συχνοτήτων στο  $N$ . Για ένα  $1-\Delta$  σήμα  $N$ ,  $\beta_1 = \beta_2 = \gamma_1 = 0$  οπότε οι σχέσεις (2.42),(2.50) ανάγονται απευθείας στα  $1-\Delta$  αντίστοιχα τους που αναφέρονται στο [38].

Τέλος, για την εφαρμογή του ESA (2.22-2.23) απαιτούνται οι ενέργειες των παραγώγων του σήματος. Έτσι, έχουν ενδιαφέρον τα βασικά στατιστικά της εφαρμογής του τελεστή στα διαφορικά  $N_x$  και  $N_y$ . Οι αναμενόμενες τιμές είναι

$$\begin{aligned} E[\Psi(N_x)] &= 2R^{(4,0)}(0,0) - R^{(1,1)}(0,0) + R^{(2,2)}(0,0) = 2\alpha_2^2 + \gamma_1^2, \\ E[\Psi(N_y)] &= 2R^{(0,4)}(0,0) - R^{(1,1)}(0,0) + R^{(2,2)}(0,0) = 2\beta_2^2 + \gamma_1^2 \end{aligned} \quad (2.51)$$

και οι μεταβλητήτες

$$\begin{aligned} \text{Var}[\Psi(N_x)] &= 2\alpha_2^2\gamma_1^2 + 3(\alpha_2^4 + \gamma_1^4) - \alpha_2^2(2R^{(4,2)} + R^{(2,4)} + R^{(6,0)}), \\ \text{Var}[\Psi(N_y)] &= 2\beta_2^2\gamma_1^2 + 3(\beta_2^4 + \gamma_1^4) - \beta_2^2(2R^{(2,4)} + R^{(4,2)} + R^{(0,6)}). \end{aligned} \quad (2.52)$$

### 2.3.6.2 Ζωνοπερατή ενέργεια

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

$$g_\sigma(x, y) = 2h_\sigma(x, y) \sin(\omega_{c1}x + \omega_{c2}y) \quad (2.53)$$

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

$$G_\sigma(\omega_1, \omega_2) = \frac{1}{j} [H_\sigma(\omega_1 - \omega_{c1}, \omega_2 - \omega_{c2}) + H_\sigma(\omega_1 + \omega_{c1}, \omega_2 + \omega_{c2})]. \quad (2.54)$$

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

$$H_\sigma(\vec{\omega}) = \frac{1}{\sqrt{|\sigma|}} H(\vec{\omega}/\sigma) \quad (2.55)$$

Αν για λόγους απλότητας θεωρηθεί ότι η κρουστική  $h$  είναι άρτια συμμετρική έτσι, ώστε  $G_\sigma(0,0) = 0$ , οι θετικές και αρνητικές συνιστώσες του  $G_\sigma(\omega)$  δεν επικαλύπτονται και το φίλτρο είναι ισοτροπικό, δηλαδή  $\sigma_1 = \sigma_2$  τότε

$$|G_\sigma(\vec{\omega})|^2 = |H_\sigma(\vec{\omega} - \vec{\omega}_c)|^2 + |H_\sigma(\vec{\omega} + \vec{\omega}_c)|^2. \quad (2.56)$$

Το φιλτραρισμένο σήμα θα είναι  $N_\sigma(x, y) = (N * g_\sigma)(x, y)$ , όπου  $*$  δηλώνει συνέλιξη. Από τη σχέση (2.43) που δίνει την αναμενόμενη τιμή του τελεστή στη διαδικασία  $N$  προκύπτει η

γενίκευση για την ζωνοπερατά φιλτραρισμένη εκδοχή του  $\Psi(N_\sigma(x, y))$

$$\begin{aligned} \mathbb{E}[\Psi(N_\sigma)] &= \frac{1}{\pi} \int |\vec{\omega}|^2 |G_\sigma(\vec{\omega})|^2 \Phi(\vec{\omega}) d\vec{\omega} \\ &= \frac{1}{\pi} \int_{\mathbb{R}^2} |\vec{\omega}|^2 \left( |H_\sigma(\vec{\omega} - \vec{\omega}_c)|^2 + |H_\sigma(\vec{\omega} + \vec{\omega}_c)|^2 \right) \Phi(\vec{\omega}) d\vec{\omega} \\ &= \frac{1}{\pi\sigma} \int_{\mathbb{R}^2} |\vec{\omega}|^2 (|H((\vec{\omega} - \vec{\omega}_c)/\sigma)|^2 + |H((\vec{\omega} + \vec{\omega}_c)/\sigma)|^2) \Phi(\vec{\omega}) d\vec{\omega}. \end{aligned} \quad (2.57)$$

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

$$\mathbb{E}[\Psi(N_\sigma)] \approx \frac{1}{\pi\sigma} \int_{\mathbb{R}^2} |\vec{\omega}|^2 \Phi(\vec{\omega}) d\vec{\omega} = \frac{1}{\sigma} \mathbb{E}[\Psi(N)]. \quad (2.58)$$

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

$$\mathbb{E}[\Psi(N * g_\sigma)] \approx \frac{\lambda}{\sigma} \mathbb{E}[\Psi(N)], \quad (2.59)$$

δηλαδή η αναμενόμενη μέση τιμή μεταβάλλεται γραμμικά, αντιστρόφως ανάλογα με τη θετική σταθερά που καθορίζει το εύρος του ζωνοπερατού φιλτραρίσματος. Μείωση του  $\sigma$ , δηλαδή 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.

□

## Κεφάλαιο 3

# Ενέργεια και Ανάλυση Υφής

Η υφή περιοχών εικόνων αναλύεται μοντελοποιώντας τις τοπικά στενοζωνικές μεταβολές φωτεινότητας ως χωρικά διαμορφωμένα σήματα-συνιστώσες χρησιμοποιώντας φιλτράρισμα σε πολλαπλές ζώνες συχνοτήτων και μη-γραμμικούς τελεστές στιγμιαίου εντοπισμού ενέργειας. Η δραστηριότητα υφής εκφράζεται με εκτίμηση της ενέργειας χωρικών διαμορφώσεων στις κυρίαρχες ζώνες συχνοτήτων. Παρουσιάζονται οι μέθοδοι πολυκαναλικής ανάλυσης και ανάλυσης κυρίαρχων συνιστώσων για εξαγωγή χαρακτηριστικών διαμόρφωσης. Το πρόβλημα της ανάλυσης υφής ανάγεται σε ανίχνευση ενεργειακών όρων και αναπαράσταση με διανύσματα χαρακτηριστικών χαμηλής διάστασης αλλά υψηλής περιγραφικότητας.

### 3.1 Διαμορφώσεις υφής

Στοιχειώδεις συνιστώσες οπτικής υφής μπορούν να ερμηνευθούν ως τοπικά ομαλές διαμορφώσεις, και, επομένως, να μοντελοποιηθούν με μη-στάσιμα σήματα, καλά εντοπισμένα σε μια στενή ζώνη χωρικών συχνοτήτων της μορφής (2.9) που επαναλαμβάνεται εδώ για αναφορά

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

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  $\phi$ . Παρ' όλα αυτά μόνο συγκεκριμένες ερμηνείες τους έχουν νόημα στη μοντελοποίηση εικόνων τοπικά στενής ζώνης (locally narrowband), όπως αυτές που φαίνονται στο Σχ. 3.1. Το πλάτος  $a(x, y)$  μοντελοποιεί την αντίθεση φωτεινότητας και το χωρικά μεταβαλλόμενο διάνυσμα  $\vec{\omega}(x, y) = (\omega_1, \omega_2)(x, y)$  περιέχει πληροφορία για τις τοπικά αναδύομενες χωρικές συχνότητες. Τέτοιες συχνότητες μεταβάλλονται ομαλά σε μια στενή φασματική ζώνη συχνοτήτων [37].

Αν και πολλές φυσικές κατηγορίες υφής μπορούν να μοντελοποιηθούν στα πλαίσια μιας AM-FM συνιστώσας, εικόνες με 2-Δ δομή και πρότυπα παρουσιάζουν ταυτόχρονα περισσότερες από μία συνιστώσες στο τοπικό φάσμα τους. Σύμφωνα με το «πολυσυνιστωσικό» (multicomponent) AM-FM μοντέλο, μια εικόνα  $I$  μοντελοποιείται ως υπέρθεση τοπικά στε-

νοζωνικών χωρικών σημάτων AM-FM διαμόρφωσης

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

όπου ο κάθε ένας από τους  $K$  συντελεστές είναι ένα 2-Δ μη-στατικό ημίτονο με χωρικά μεταβαλλόμενο πλάτος  $a_k(x, y)$  και χωρικά μεταβαλλόμενο διάνυσμα στιγμιαίων συχνοτήτων  $\vec{\omega}_k(x, y)$ , 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 και την τραχύτητα (coarseness) της προκύπτουσας υφής. Έτσι, οι τοπικές μεταβολές της υφής σε μια εικόνα μοντελοποιούνται με πεπερασμένο αριθμό χωρικά διαμορφωμένων σημάτων τα οποία διαχωρίζονται μέσα από ζωνοπερατό φιλτράρισμα 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 της εικόνας. Τέτοια μοντέλα διαμόρφωσης πολλαπλών συνιστωσών προτάθηκαν για εικόνες με φραγμένα σφάλματα προσέγγισης γύρω από την ομαλότητα των διαμορφώσεων [124, 125].

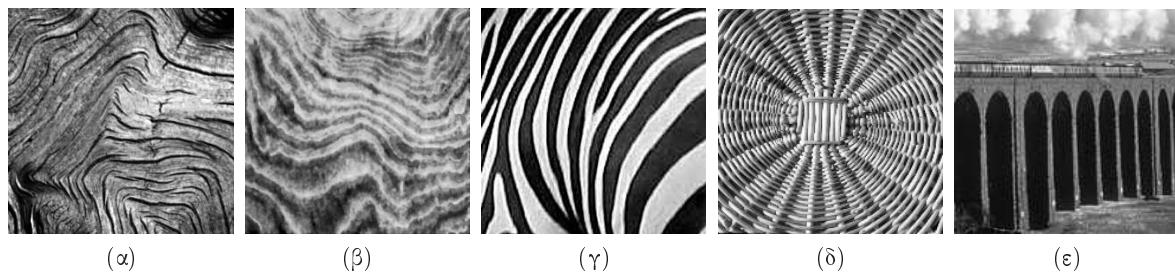
Για την ανάλυση της κάθε συνιστώσας σε συγκεκριμένο ζεύγος AM και FM συναρτήσεων [293] ξεχωρίζουν δύο προσεγγίσεις. Η πρώτη ενεργεία απευθείας στις τιμές της εικόνας  $f_k(x, y)$  και σχετίζεται με τον 2-Δ ενεργειακό τελεστή (2.13) και τον αλγόριθμο διαχωρισμού ενέργειας ESA (2.22, 2.23) που περιγράφονται στην ενότητα 2.3. Η δεύτερη περιλαμβάνει τη δημιουργία της αναλυτικής εικόνας από κάθε συνιστώσα μέσω του πολυδιάστατου κατευθυντικού Hilbert μετασχηματισμού [126]. Οι δύο προσεγγίσεις είναι σχεδόν ισοδύναμες. Περισσότερες λεπτομέρειες για αποδιαμόρφωση μέσω της 2-Δ αναλυτικής εικόνας δίνονται στα [122, 125, 127].

Οι συναρτήσεις ταλάντωσης, οι οποίες ορίστηκαν από τον Meyer [214], για μοντελοποίηση και διαχωρισμό της υφής από εικόνες μπορούν να αναζητηθούν στο πλαίσιο της AM-FM μοντελοποίησης και στην ύπαρξη διαμορφωμένων ταλαντώσεων έντασης φωτεινότητας και χωρικών συχνοτήτων. Το κίνητρο εδώ είναι οι αναλογίες ανάμεσα στα μοντέλα διαμόρφωσης και στις μαθηματικές θεωρήσεις για τις νόρμες μεταβολής της υφής [214]. Μια τέτοια μοντελοποίηση παραμένει ανοιχτό ερευνητικό ζήτημα και πεδίο μελλοντικής έρευνας.

### 3.1.1 Ενέργεια συνιστωσών υφής

Εφαρμόζοντας τον 2-Δ 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

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



Σχήμα 3.1: Παραδείγματα τοπικά στενοζωνικής υφής εικόνων ως αποτέλεσμα (α) εξελικτικών διαδικασιών, (β) παραμόρφωσης της επιφάνειας, (γ) βιολογικών προτύπων και (δ),(ε) (περιοδικών) ανθρώπινων κατασκευών.

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

Θεωρώντας ότι οι μεταβολές του πλάτους και των στιγμαίων συχνοτήτων δεν είναι πολύ γρήγορες στο χώρο ούτε πολύ μεγάλες σε τιμές σε σχέση με τις φέρουσες, το παραπάνω σφάλμα προσέγγισης γίνεται αμελητέο [198,201,202]. Ο διαχωρισμός αυτής της ενέργειας στα δύο σήματα διαμόρφωσης είναι εφικτός εφαρμόζοντας τον ESA (2.22, 2.22). Επομένως, μια μεμονωμένη, χωρικά μεταβαλλόμενη AM-FM συνιστώσα υφής χαρακτηρίζεται σε κάθε θέση από την εκτιμώμενη περιβάλλουσα πλάτους και το μέγεθος της μεταβολής των στιγμαίων συχνοτήτων.

## 3.2 Πολυκαναλική ανάλυση

Οι εικόνες υφής χαρακτηρίζονται από ενέργεια σε μία η περισσότερες διαφορετικές φασματικές ζώνες, δηλαδή πρόκειται ση γενική περίπτωση για σήματα ευρείας ζώνης. Ως συνέπεια, τα AM-FM μοντέλα δεν εφαρμόζονται απευθέας σε μια εικόνα αλλά στις ζωνοπερατά φιλτραρισμένες εκδοχές, κάθε μία από τις οποίες περιγράφεται από μια συνιστώσα της μορφής (3.1). Εκτός από τον διαχωρισμό των διαμόρφωσεων της υφής και την ανάδειξη των κυρίαρχων τοπικών συχνοτήτων [37], το ζωνοπερατό φιλτράρισμα αυξάνει την ευρωστία σε θόρυβο και ομαλοποιεί τα σήματα πλάτους και συχνότητας.

Για την απομόνωση των τοπικών συνιστωσών διαμόρφωσης χρησιμοποιούνται μηχανισμοί φιλτραρίσματος με επαρκή χωρική και φασματική τοποθέτηση. Μια συστοιχία από γραμμικά, μεταφορικά αμετάβλητα ζωνοπερατά φίλτρα καλύπτει πυκνά το πεδίο συχνοτήτων με μόνη υπόθεση ότι σε κάθε θέση της εικόνας 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 καλή από-κοινού φασματική και χωρική τοποθέτηση οδηγεί στη χρήση Gabor φίλτρων (2.1, 2.2) ως βέλτιστη επιλογή. Τα φίλτρα περιγράφονται αναλυτικά στην ενότητα 2.2, ενώ ο σχεδιασμός της συστοιχίας στην

ενότητα 2.2.3.

Οι μεθοδολογίες ανάλυσης υφής φυσικών εικόνων που σχετίζονται με ανάλυση σε πολλαπλές ζώνες συχνοτήτων ανακτούν την πληροφορία στην έξοδο κάθισε φίλτρου δημιουργώντας διανύσματα χαρακτηριστικών ή εναλλακτικά συνδυάζουν τις εξόδους ή προβάλλουν το διάνυσμα σε χαμηλότερη διάσταση. Οι δύο ξεχωριστές προσεγγίσεις αναφέρονται ως *ανάλυση συνιστωσών κατά κανάλια* (channelized component analysis ή CCA) και *ανάλυση κυρίαρχων συνιστωσών* (dominant component analysis ή DCA) αντιστοίχως [122, 125].

### 3.2.1 Ζωνοπερατή αποδιαμόρφωση

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

$$f_k(x, y) = I(x, y) * g_k(x, y) \approx I_k(x, y) * g_k(x, y). \quad (3.5)$$

Μπορεί να αποδειχθεί [127] υπό ρεαλιστικές και ασθενείς συνθήκες, όπως στο [38], ότι σε αυτή τη γειτονιά

$$\begin{aligned} \Psi[f_k(x, y)] &\approx a_k^2(x, y) |\nabla \phi_k(x, y)|^2 |G_k[\nabla \phi_k(x, y)]|^2 \\ &\approx \Psi[I_k(x, y)] |G_k[\nabla \phi_k(x, y)]|^2. \end{aligned} \quad (3.6)$$

Για την αποδιαμόρφωση μέσω ESA στα σήματα χωρικής μεταβολής ο ενεργειακός τελεστής εφαρμόζεται στις απομονωμένες συνιστώσες και στις μερικές παραγώγους τους. Για τα σήματα συχνοτήτων στις σχέσεις (2.22) ο  $\Psi$  εμφανίζεται τόσο στον αριθμητή όσο και στον παρονομαστή, οπότε ο όρος  $|G_k[\nabla \phi_k(x, y)]|^2$  ακυρώνεται λόγω διαίρεσης. Έτσι, η εφαρμογή των (2.22) για την αποδιαμόρφωση μιας συνιστώσας γίνεται απευθείας στη ζωνοπερατή συνιστώσα στην έξοδο του φίλτρου

$$\sqrt{\frac{\Psi(\partial f_k / \partial x)}{\Psi(f_k)}} \approx |\omega_{k1}(x, y)|, \quad \sqrt{\frac{\Psi(\partial f_k / \partial y)}{\Psi(f_k)}} \approx |\omega_{k2}(x, y)|. \quad (3.7)$$

Παρόμοια θεώρηση επεκτείνεται και στα διακριτά ισοδύναμά τους (2.28, 2.29). 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 (2.23) τροποποιείται με κανονικοποίηση ως εξής:

$$|G_k[(\omega_{k1}, \omega_{k2})(x, y)]|^{-1} \frac{\Psi(f_k)}{\sqrt{\Psi(\partial f_k / \partial x)} + \sqrt{\Psi(\partial f_k / \partial y)}} \approx |a_k(x, y)|, \quad (3.8)$$

όπου οι  $\omega_{k1}, \omega_{k2}$  έχουν υπολογιστεί ήδη από την (3.7). Τα πρόσημα των στιγμιαίων συχνοτήτων (3.7) 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, δηλαδή τη διακριτοποίηση των σχέσεων (3.7, 3.8) στις  $K$  ζωνοπερατές συνιστώσες στην έξοδο της συστοιχίας των 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 Μιγαδική αποδιαμόρφωση

Η απόκριση του μιγαδικού Gabor (2.1) σε μια εικόνα πραγματικών τιμών είναι μια μιγαδική εικόνα  $f_k = I * g_k = (I * \text{Re}\{g_k\}) + j(I * \text{Im}\{g_k\})$ . Σε αυτή την περίπτωση η αποδιαμόρφωση μέσω του ESA (2.22, 2.23) επεκτείνεται σε σήματα μιγαδικών τιμών εφαρμόζοντας τον ενεργειακό τελεστή για μιγαδικά σήματα (2.16).

Για ένα γενικό 2-Δ μιγαδικό σήμα  $f_k(x, y) = a_k(x, y) \exp\{j\phi_k(x, y)\}$  η απόκριση του τελεστή είναι  $C[f] = \Psi[a_k \cos(\phi_k)] + \Psi[a_k \sin(\phi_k)]$ . Από την προσέγγιση (2.20), η οποία ισχύει με φραγμένο σφάλμα [198], εύκολα φαίνεται ότι  $\Psi[a_k \cos(\phi_k)] \approx \Psi[a_k \sin(\phi_k)] \approx a_k^2 \|\vec{\omega}_k\|^2$  και επομένως,

$$C[a_k(x, y) \exp\{j\phi_k(x, y)\}] \approx 2a_k^2 \|\vec{\omega}_k\|^2 \quad (3.9)$$

που είναι δύο φορές το ενεργειακό γινόμενο το πραγματικού σήματος. Αυτό συμφωνεί και με τη διαίσθηση ότι ένα μιγαδικό σήμα διαμόρφωσης απαιτεί διπλάσια ενέργεια για να παραχθεί.

Εφαρμόζοντας τον  $C$  στα μερικά διαφορικά  $f_{kx}, f_{ky}$ , οδηγούμαστε σε ένα σχήμα αποδιαμόρφωσης, όπου οι συχνότητες δίνονται από την (2.22) με το μιγαδικό τελεστή (ή το μέσο της απόκρισης των τελεστών στο πραγματικό και φανταστικό μέρος) και το πλάτος από μια αλλαγή κλίμακας (scaling) της εκτίμησης (2.23) κατά  $\sqrt{2}$ , δηλαδή

$$|a_k(x, y)| \approx \frac{\Psi(f_k)}{\sqrt{2} \sqrt{\Psi(f_{kx}) + \Psi(f_{ky})}}. \quad (3.10)$$

Αυτό το σχήμα, το οποίο θα μπορούσε να ονομαστεί «μιγαδικός» ESA χρησιμοποιεί ουσιαστικά τις μέσες αποκρίσεις των τελεστών στις δύο συνιστώσες και καταλήγει σε ομαλότερες εκτιμήσεις των συναρτήσεων διαμόρφωσης.

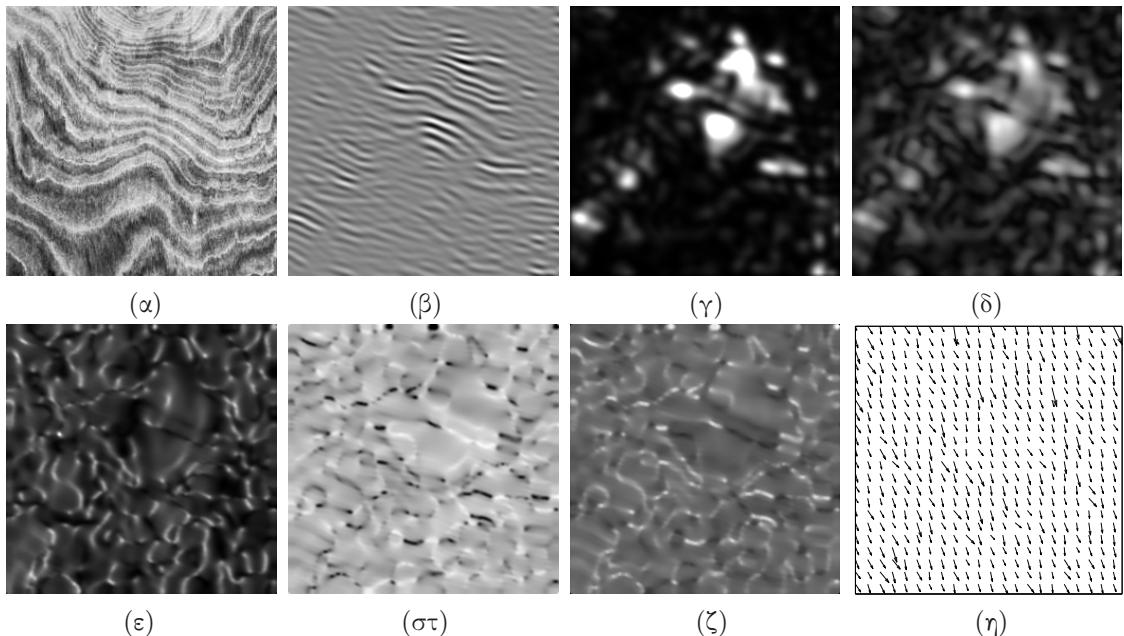
Η αποδιαμόρφωση του μιγαδικού σήματος μπορεί να γίνει επίσης χρησιμοποιώντας την αναλυτική συνάρτηση. Το μιγαδικό Gabor συνιστά ένα ζεύγος τετραγωνισμού φάσης (Εν. 2.2.2) με φάσμα ορισμένο στο «θετικό» ημιεπίπεδο του πεδίου συχνοτήτων. Αυτό καθιστά τη συνάρτηση προσεγγιστικά αναλυτική [64] της οποίας η απόκριση σε οποιαδήποτε είσοδο  $I * g_k = a_k \exp\{j\phi_k\}$  είναι επίσης αναλυτικό σήμα. Το πλάτος της περιβάλλουσας και η στιγμιαία χωρική φάση μπορούν να εκτιμηθούν από τη μιγαδική απόκριση

$$a_k(x, y) = |f_k(x, y)|, \quad \phi_k(x, y) = \arctan\left(\frac{\text{Im}\{f_k(x, y)\}}{\text{Re}\{f_k(x, y)\}}\right). \quad (3.11)$$

Τα μειονεκτήματα της αναλυτικής αποδιαμόρφωσης απαντώνται στα σφάλματα από την προσέγγιση της αναλυτικής συνάρτησης αλλά και στην έμμεση ανάγκη που υπεισέρχεται για τον υπολογισμό των στιγμιαίων συχνοτήτων.

### 3.2.3 Ανάλυση συνιστωσών κατά κανάλια

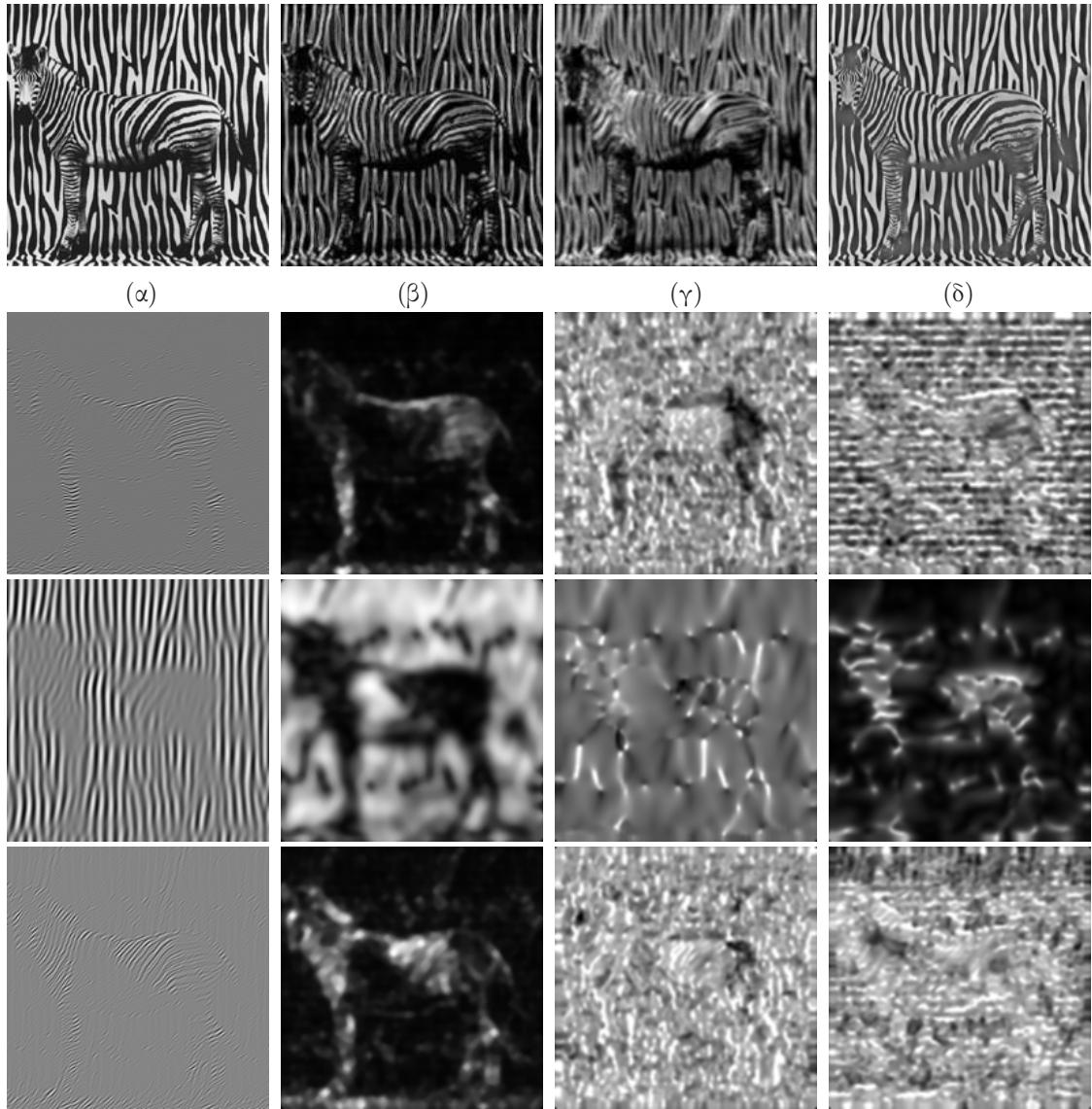
Θεωρώντας δεδομένο εκ των προτέρων ότι τα κανάλια της συστοιχίας χωρίζουν την εικόνα σε συνιστώσες με τρόπο χωρικά καθολικό (spatially global), κάθε συνιστώσα διαμόρφωσης ανήκει σε ένα κανάλι για όλο το πεδίο της εικόνας. Στο σχήμα CCA ο αριθμός των συνιστωσών στο προκύπτον μοντέλο είναι αναγκαστικά ίσος με τον αριθμό των καναλιών της συστοιχίας. Αποδιαμορφώνοντας τις προκύπτουσες συνιστώσες για την ολική εικόνα με διαχωρισμό ενέργειας (3.7, 3.8), προκύπτει μια πυκνή αναπαράσταση η οποία χαρακτηρίζει τις ισχυρές δομές της εικόνας και τα τοπικά αναδυόμενα χαρακτηριστικά της υφής. Για κάθε εικονοστοιχείο κατασκευάζεται ένα διάνυσμα χαρακτηριστικών διάστασης ίσης με τον αριθμό των συνιστωσών επί τον αριθμό των σημάτων διαμόρφωσης.



Σχήμα 3.2: Αποδιαμόρφωση και ανάλυση υφής σε ζωνοπερατό κανάλι. (α) υφή ξύλου ( $D71$ )  $256 \times 256$  [40], (β) ζωνοπερατή συνιστώσα από φίλτρο [ $u = 3, v = 14, B = 20$ ], (γ) ενέργεια διαμόρφωσης συνιστώσας υφής (έξοδος 2-Δ ενεργειακού τελεστή), (δ) περιβάλλουσα πλάτους, (ε) οριζόντια στιγμιαία συχνότητα, (στ) Κάτακόρυφη στιγμιαία συχνότητα, (ζ) Μέτρο του διανύσματος συχνοτήτων, (η) Προσανατολισμός διανύσματος συχνοτήτων.

Στο Σχ. 3.2 φαίνεται ένα παράδειγμα ανάλυσης μιας εικόνας φυσικής υφής ξύλου σε ένα κανάλι της συστοιχίας. Η εξαγόμενη ζωνοπερατή συνιστώσα στο (β) εμφανίζει διαμορφώσεις σε συγκεκριμένη κλίμακα και προσανατολισμό, οι οποίες μοντελοποιούνται από ένα μη-στατικό ημιτονοειδές, ενώ η έξοδος του 2-Δ 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 (στ) στιγμιαίας χωρικής συχνότητας. Τέλος, στο (ζ) φαίνεται το μέτρο του διανύσματος συχνοτήτων, η μέτρηση της κλίμακας της υφής και στο (η) το διάγραμμα προσανατολισμού του Η ενέργεια διαμόρφωσης ποσοτικοποιεί την ύπαρξη και την μεταβολή της υφής στη συγκεκριμένη ζώνη συχνοτήτων.

Αποτελέσματα ανάλυσης κατά κανάλια παρουσιάζονται και στο Σχ. 3.3 για μια εικόνα με έντονες διαμορφώσεις σε πολλαπλές κλίμακες και προσανατολισμούς. Οι σχηματισμοί υφής



Σχήμα 3.3: Συνιστώσες AM-FM υφής και αποδιαμόρφωση. 1η σειρά: (α) υφή κυματισμών ζέβρας  $I$  ( $200 \times 200$ ), (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. Από αριστερά προς τα δεξιά για τις υπόλοιπες: ζωνοπερατή συνιστώσα, περιβάλλουσα πλάτους  $a(x, y)$ , οριζόντια  $\omega_1(x, y)$  και κατακόρυφη  $\omega_2(x, y)$  στιγμαία συχνότητα. 2η σειρά: συνιστώσα φίλτρου [ $u = 16, v = 44, B = 5$ ], 3η σειρά: συνιστώσα φίλτρου [ $u = 14, v = 2, B = 16$ ], 4η σειρά: συνιστώσα φίλτρου [ $u = 31, v = 35, B = 5$ ].

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 2-Δ ενεργειακού τελεστή στη συνολική εικόνα ευρείας ζώνης (β), που λειτουργεί σαν φίλτρο ενίσχυσης της αντίθεσης της εικόνας [198], 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 συχνοτήτων οι τρεις συνιστώσες είναι σχεδόν συμπληρωματικές, οριοθετώντας πλήρως τις διαφορετικές περιοχές υφής στην εικόνα.

### 3.2.3.1 Επεξεργασία και ομαλοποίηση

Για τον περιορισμό των σφαλμάτων εκτίμησης της ενέργειας και της διαδικασίας αποδιαμόρφωσης καθώς και για την επιβολή κάποιας ομαλότητας στα σήματα πλάτους και συχνοτήτων συνιστάται η εφαρμογή τεχνικών μετα-επεξεργασίας (post-processing). Γενικώς, οι εικόνες υφής 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 [35,36], όπου η απόκριση πλάτους και φάσης χρησιμοποιήθηκε για κατάτμηση εικόνων, βαθυπερατά φίλτρα Gauss τυπικής απόκλισης περίπου ίσης με αυτή των Gabor καναλιών χρησιμοποιήθηκαν για ομαλοποίηση των αποκρίσεων.

Στην περίπτωση αποδιαμόρφωσης των καναλιών είναι προτιμότερο η ομαλοποίηση και η διαχείριση των σφαλμάτων να εφαρμόζεται απευθείας στις εκτιμούμενες AM και FM συναρτήσεις. Τα σφάλματα αποδιαμόρφωσης οφείλονται στα σφάλματα προσεγγίσεων του ενεργειακού τελεστή, στους λόγους με ενεργειακές ποσότητες που ενίστεις είναι πολύ μικρές και στις προσεγγίσεις παραγώγων μέσω πεπερασμένων διαφορών. Για την πρώτη περίπτωση τα τυχόν σφάλματα προσέγγισης του 2-Δ ενεργειακού τελεστή έχουν μια υψηπερατή και μια βαθυπερατή συνιστώσα [202,245]. Η υψηπερατή αφαιρείται με εφαρμογή στην έξοδο του τελεστή ενός βαθυπερατού φίλτρου π.χ. διωνυμικά ή Gauss φίλτρα. Επιπλέον, τυχόν αρνητικές τιμές στην έξοδο του τελεστή, οι οποίες 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 (3.7, 3.8) χρησιμοποιείται ένας συνδυασμός από βαθυπερατά φίλτρα ομαλοποίησης των εκτιμήσεων και στατιστικά (median) φίλτρα σε τοπικά παράθυρα. Το βαθυπερατό φίλτρο αντισταθμίζει τα σφάλματα προσέγγισης του αλγορίθμου, ενώ το median ενδείχνυται για την απομάκρυνση μεμονωμένων ακραίων μεταβολών (spikes, singularities) λόγω των διαιρέσεων του DESA. Στα παραδείγματα των Σχτ. 3.2,3.3 χρησιμοποιήθηκε ενδεικτικά διωνυμικό φίλτρο αντισταθμίσης στον ενεργειακό τελεστή, Gauss φίλτρα τυπικής απόκλισης 2 και μεγέθους 3 και median σε παράθυρα  $3 \times 3$ .

### 3.2.3.2 Ανακατασκευή

Η ανάκτηση της αρχικής εικόνας από τις παραμέτρους του μοντέλου σε κάθε θέση περιλαμβάνει ανακατασκευή των συνιστωσών (3.1) από τα χωρικά μεταβαλλόμενα σήματα διαμόρ-

φωσης 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 διάνυσμα στιγμαίων συχνοτήτων ( $\omega_{k1}, \omega_{k2}$ ) για κάθε συνιστώσα. Η ανακατασκευή φάσης είναι γενικότερα ένα δύσκολο πρόβλημα λογισμού διανυσματικών πεδίων και για την μη-αναλυτική επίλυσή του έχουν προταθεί αλγεβρικές, μεταβολικές και αριθμητικές μέθοδοι. Για τα AM-FM μοντέλα ένας πρακτικός αλλά μη-βέλτιστος αλγόριθμος προτάθηκε στο [124].

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

$$\phi_k(m, n) = \frac{1}{2}[\phi_k(m - 1, n) + \omega_{k1}(m - 1, n) + \phi_k(m, n - 1) + \omega_{k2}(m, n - 1)]. \quad (3.12)$$

Η αρχικοποίηση της (3.12) 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 [125]. Τα δείγματα της φάσης αποθηκεύονται κατά τη διαδικασία ανάλυσης και για την αποφυγή διάδοσης σφαλμάτων εκτίμησης, που υπεισέρχονται στα στιγμαία σήματα συχνοτήτων, η διαδικασία ανακατασκευής της φάσης μπορεί να γίνεται ανεξάρτητα σε περιοδικά τετράγωνα από εικονοστοιχεία. Η φάση αρχικοποιείται σε κάθε τετράγωνο, εκτιμάται ανεξάρτητα στην πρώτη στήλη και πρώτη γραμμή του και από την (3.12) ανακατασκευάζεται στο εσωτερικό.

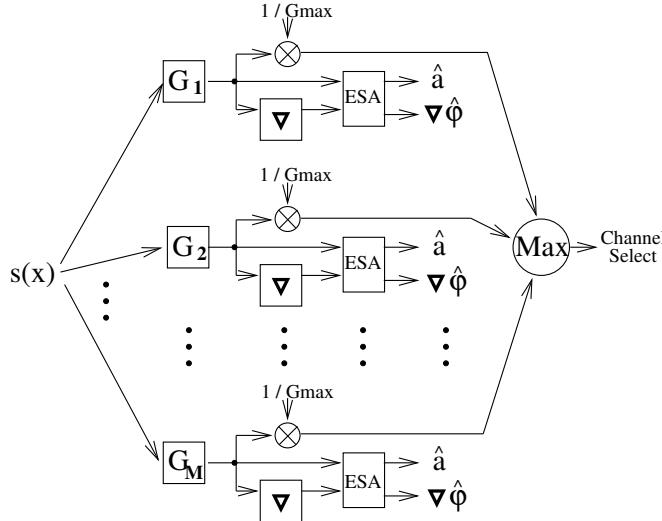
### 3.3 Ανάλυση κυρίαρχων συνιστωσών

Η πληροφορία των πολυδιάστατων χαρακτηριστικών της ανάλυσης κατά κανάλια, παρά το ότι ακολουθεί πιστά το μοντέλο διαμορφώσεων, μπορεί να καθίσταται περιττή και να αυξάνει την πολυπλοκότητα της περιγραφής. Η ανάλυση κυρίαρχων συνιστωσών (dominant components analysis ή DCA) [121, 122, 125] αποτελεί μια εναλλακτική περιγραφή με διανύσματα χαρακτηριστικών χαμηλής διάστασης. Η μέθοδος για μια εικόνα υφής, που μοντελοποιείται από το άθροισμα συνιστωσών (3.2), περιλαμβάνει την περιγραφή και την εκτίμηση των συναρτήσεων διαμόρφωσης της ισχυρότερης τοπικά συνιστώσας. 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, στιγμαίες συχνότητες που κυριαρχούν σε χωρική βάση στο φάσμα των περισσότερων φυσικών εικόνων υφής ονομάζονται αναδυόμενες και δίνονται από το διάνυσμα φάσης της συνιστώσας που είναι κυρίαρχη σε κάθε σημείο [37]. Τέτοιες συχνότητες χαρακτηρίζουν τη λεπτομέρεια, τη ροή και τον προσανατολισμό των περισσότερων φυσικών υφών (τοπικά συνεκτικές, κανονικής δομής, αποτελέσματα φυσικών διαδικασιών εξέλιξης, κοκκώδεις, προσανατολισμένες). Ως συνέπεια, εκτιμήσεις αυτών των συχνοτήτων φέρουν πληροφορία σχετική με την τοπική δομή της υφής.

Το σχήμα DCA εξάγει τις ισχυρότερες τοπικά συνιστώσες διαμόρφωσης από το AM-FM μοντέλο επιλέγοντας μία σε κάθε θέση ανάμεσα από τις διαφορετικές ζώνες συχνοτήτων.

Τα στάδια φαίνονται στο διάγραμμα του Σχ. 3.4 (από το [125]). Το σύνθετο δισδιάστατο σήμα αναλύεται από μια συστοιχία ζωνοπερατών φίλτρων για την φασματική απομόνωση των επιμέρους συνιστώσων με βάση τις αρχές της ενότητας 3.2. Οι διαφορετικές εκδοχές στενής ζώνης, που υεωρούνται μεμονωμένες συνιστώσες του μοντέλου υφής, αποδιαμορφώνονται από τις εξισώσεις διαχωρισμού ενέργειας του ESA (3.7, 3.8). Παρόμοια με την ανάλυση κατά κανάλια, η κάθε συνιστώσα δίνει ξεχωριστές εκτιμήσεις πλάτους και στιγμιαίων συχνοτήτων.



Σχήμα 3.4: Διάγραμμα ανάλυσης κυρίαρχων συνιστώσων πλάτους από το [121]. Σειριακά επιτελείται φιλτράρισμα, αποδιαμόρφωση και επιλογή του καναλιού που μεγιστοποιεί το κριτήριο κανονικοποιημένου πλάτους.

Η ανάλυση DCA υεωρεί ότι μια μοναδική συνιστώσα κυριαρχεί τοπικά στο φάσμα της εικόνας. Η κυρίαρχη συνιστώσα υφής είναι μια τοπικά στενοζωνική, ομαλά μεταβαλλόμενη συνάρτηση  $d(x, y) = a_d(x, y) \exp\{j\phi_d(x, y)\}$  η οποία εξάγεται ανακατασκευάζοντας την εικόνα σε κάθε  $\mathbf{x} = (x, y)$  από μια μοναδική συνιστώσα του αθροίσματος (3.2)

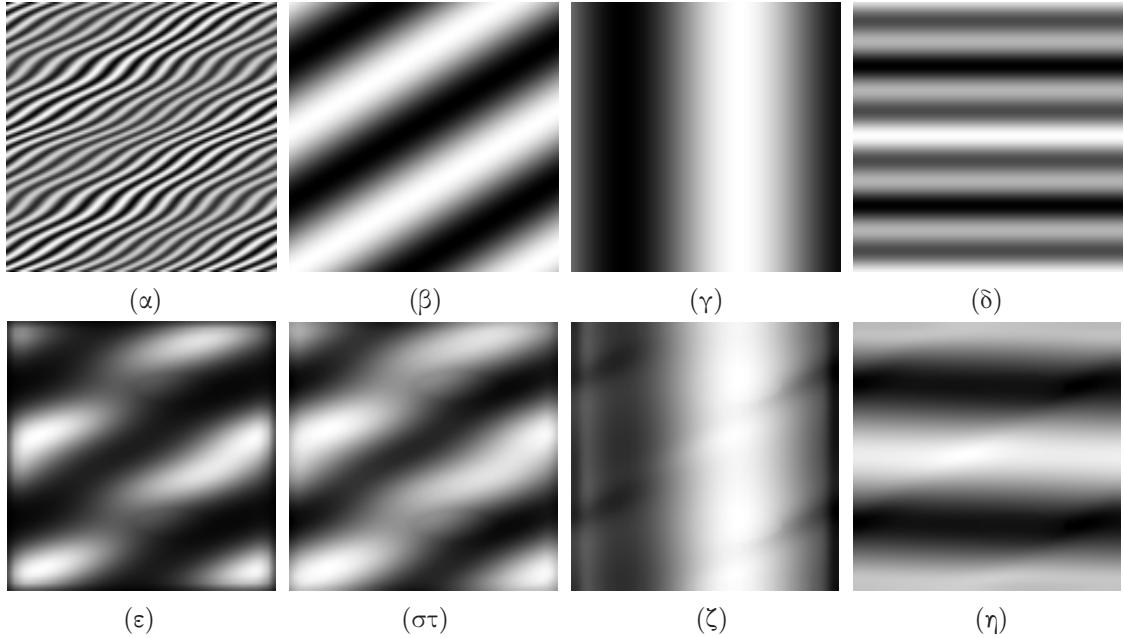
$$d(x, y) = \{t_i(x, y) : i(x, y) = \arg \max_k \{\Gamma_k(x, y)\}, k \in [1, K]\}. \quad (3.13)$$

Η επιλογή του καναλιού το οποίο δίνει την εκτίμηση για την κυρίαρχη τοπικά συνιστώσα βασίζεται στον υπολογισμό ενός μετρικού, του κριτηρίου επιλογής καναλιού  $\Gamma_k$  σε κάθε σημείο της εικόνας. Οι συναρτήσεις διαμόρφωσης της  $d(x, y)$  προκύπτουν από την  $K$ -σε-1 ανάστροφη απεικόνιση των εκτιμουμένων παραμέτρων των καναλιών  $(a_k, \nabla \phi_k)(x, y)$  στις κυρίαρχες

$$a_d(x, y) = a_{i(x, y)}(x, y), \quad \vec{\omega}_d(x, y) = \vec{\omega}_{i(x, y)}(x, y) \quad (3.14)$$

η οποία είναι υεωρητικά ισοδύναμη με την άμεση αποδιαμόρφωση της τοπικά στενοζωνικής συνιστώσας. Οι εκτιμήσεις για το κυρίαρχο πλάτος και κυρίαρχο μέτρο συχνότητας δίνονται έτσι για κάθε σημείο από το κανάλι  $i(x, y) = \arg \max_k \{\Gamma_k(x, y)\}$ , όπου  $k = 1 \dots K$  ο δείκτης των  $K$  φίλτρων της συστοιχίας.

Ανάλογα με την επιλογή του κριτηρίου  $\Gamma_k(x, y)$  το οποίο μεγιστοποιεί η συνάρτηση  $d$ , αναδεικνύονται δύο εναλλακτικά DCA σχήματα. Το DCA πλάτους ή ADCA αποτελεί την



Σχήμα 3.5: Ανάλυση κυρίαρχων συνιστώσων ενέργειας (EDCA). (α) συνθετικό 2-Δ AM-FM σήμα, (β) - (δ) πραγματική περιβάλλουσα πλάτους, οριζόντια και κατακόρυφη στιγμιαία συχνότητα, αντιστοίχως, (ε) κυρίαρχη ενέργεια Teager (στ) - (η) κυρίαρχες συνιστώσες διαμόρφωσης πλάτους και συχνοτήτων μέσω DCA.

αρχική προσέγγιση [122] και χρησιμοποιεί το κανονικοποιημένο πλάτος της περιβάλλουσας των καναλιών

$$\Gamma_k(x, y) = \frac{|a_k(x, y)|}{\max_{\vec{w}} |G_k(\vec{w})|}, \quad (3.15)$$

όπου  $G_k(\vec{w})$ ,  $w = (u, v)$  η απόκριση συχνότητας του φίλτρου  $k$ . Μπορεί να αποκτηθεί ισοδύναμα χρησιμοποιώντας το πραγματικό  $g_{ke}$  και το φανταστικό  $g_{ko}$  μέρος της απόκρισης του μιγαδικού φίλτρου (άρτιο και περιπτό φίλτρο, αντιστοίχως)

$$\Gamma_k(x, y) = \frac{\sqrt{(I * g_{ke})(x, y) + (I * g_{ko})(x, y)}}{\max_{\vec{w}} |G_k(\vec{w})|}. \quad (3.16)$$

Τα δύο κριτήρια αποδεικνύονται ότι είναι θεωρητικά ισοδύναμα και η περιβάλλουσα της μιγαδικής απόκρισης  $|(I * g_k)(x, y)|$  προσεγγιστικά είναι μια ομαλοποιημένη έκδοση του πλάτους  $|a_k(x, y)|$  [35, 36]. Η ομαλοποίηση επιβάλλεται από μια συνάρτηση Gauss τυπικής απόκλισης  $\sigma_k/\sqrt{2}$ , όπου  $\sigma_k$  η απόκλιση για το συγκεκριμένο κανάλι. Συγκεκριμένα, μπορεί να δειχθεί ότι

$$|(I * g_k)(x, y)| = cH_k(x/\sqrt{2}, y/\sqrt{2})|a_k(x, y)| \approx |a_k(x, y)|, \quad (3.17)$$

όπου  $c$  μια σταθερά και  $H_k$  η περιβάλλουσα διαμόρφωσης Gauss του φίλτρου  $k$ .

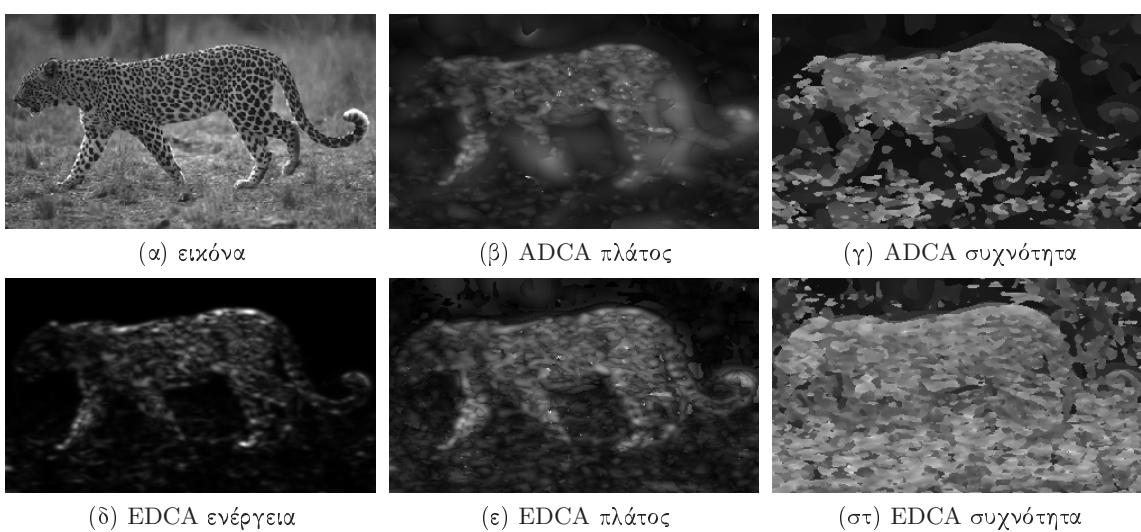
Στα πλαίσια της παρούσας διατριψής [92, 163–165, 268, 269] αναπτύχθηκε εναλλακτικά ένα νέο σχήμα DCA βασισμένο στην Teager ενέργεια των κυρίαρχων συνιστώσων το οποίο και καλείται DCA ενέργειας ή EDCA. Συγκεκριμένα, το κυρίαρχο τοπικά κανάλι επιλέγεται με βάση την έξοδο του μιγαδικού τελεστή (2.16) στην απόκριση του μιγαδικού φίλτρου

$$\Gamma_k(x, y) = C[(I * g_k)(x, y)] = \Psi[(I * g_{ke})(x, y)] + \Psi[(I * g_{ko})(x, y)]. \quad (3.18)$$

Αν χρησιμοποιηθεί το άρτιο ή το περιττό φίλτρο, το κριτήριο ανάγεται στην εφαρμογή του τελεστή σε πραγματικά σήματα. Το ενεργειακό κριτήριο καταλήγει σε μια συνιστώσα μεγαλύτερης συνοχής με βελτιωμένο εντοπισμό στα όρια υφής και αντικειμένων. Καθώς ο ενεργειακός τελεστής ευνοεί τις υψησυχνες συνιστώσες, ταιριάζει περισσότερο για επιλεκτικό ως προς την κλίμακα διαχωρισμό της υφής από τις λοιπές μακροδομές μιας εικόνας (κεφάλαιο 5).

Η ενεργειακή ανάλυση κυρίαρχων συνιστώσων 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 συνιστώσων, κάτιο που οποίο διαισθητικά φαίνεται πιο φυσικό. Εξάλλου, η ενέργεια διαμόρφωσης υφής, όπως ορίστηκε προηγούμενα, περιλαμβάνει ταυτόχρονα πληροφορία πλάτους και συσχνότητας. Επομένως, μια ερμηνεία του EDCA είναι ότι εξάγει τις κυρίαρχες συνιστώσες της υφής με βάση την ενέργεια των διαμορφώσεών τους. Μια συνιστώσα χαρακτηρίζεται κυρίαρχη όχι μόνο λόγω των μεγάλων τιμών διαμόρφωσης περιβάλλουσας, που ερμηνεύεται μέσω υψηλής τοπικής αντίθεσης αλλά και λόγω των αναδύομενων συχνοτήτων που κυριαρχούν στο φάσμα της εικόνας. Επιπλέον, οι εκτιμήσεις του EDCA, λόγω του συχνοτικού όρου, συλλαμβάνουν οξείες μεταβολές της υφής ανάμεσα σε διαφορετικές περιοχές της εικόνας και ακμές υφής (Σχ. 3.6).

Και στα δύο σχήματα οι εκτιμήσεις στις παραμέτρους διαμόρφωσης των κυρίαρχων συνιστώσων 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. Το DCA μειώνει τη διάσταση της περιγραφής από τον  $K$ -διάστατο χώρο των Gabor αποκρίσεων (ή τον  $3 \times K$  χώρο διαμορφώσεων) στο  $3 \times 1$  διάνυσμα ( $\alpha_d, \nabla\phi_d$ ), ενώ διατηρεί τα ουσιαστικά γνωρίσματα της υφής όπως είναι η έκταση, η κλίμακα και ο προσανατολισμός. Η διαδικασία επιλογής του κυρίαρχου καναλιού ερμηνεύτηκε μέσω στατιστικής ανίχνευσης σημάτων στο [163], ενώ στο [164] τα χαρακτηριστικά κυρίαρχων συνιστώσων διαμόρφωσης χρησιμοποιήθηκαν σε ένα σχήμα μεταβολικής κατάτμησης εικόνων (κεφάλαιο 4).



Σχήμα 3.6: Κυρίαρχα χαρακτηριστικά DCA με το κριτήριο πλάτους (πάνω σειρά) και ενέργειας (κάτω σειρά): κυρίαρχες συνιστώσες πλάτους, μέτρου στιγμαίων συχνοτήτων και ενέργειας.

### 3.4 Κανονικοποιημένη αποδιαμόρφωση - Gabor ESA

Ένα πρόβλημα που ανακύπτει με την αποδιαμόρφωση μέσω του ESA είναι ότι για διαχριτά δεδομένα μόνο προσεγγίσεις στα διαφορικά των σημάτων, ουσιώδη για τις μερικές διαφορίσεις, μπορούν να είναι διαθέσιμες π.χ. μέσω πεπερασμένων διαφορών. Ως αποτέλεσμα οι προσεγγιστικές έξοδοι των ενεργειακών τελεστών ο οποίες περιλαμβάνουν δύο διαφορικούς τελεστές, είναι πιθανό να εξάγουν ανακριβείς εκτιμήσεις πλάτους και συχνοτήτων. Μια θεωρητικά βάσιμη προσέγγιση που αντιμετωπίζει το πρόβλημα προκύπτει με χρήση του κανονικοποιημένου 2-Δ ενεργειακού τελεστή (regularized 2-D energy operator) και του σχετικού κανονικοποιημένου 2-Δ ESA.

Όπως αναλύθηκε στο [278] για ανίχνευση ακμών, οι δύο κανονικοποιημένες λύσεις στο πρόβλημα εκτίμησης της παραγώγου οι οποίες ελαχιστοποιούν το άθροισμα του σφάλματος προσέγγισης των δεδομένων και την ενέργεια της δεύτερης παραγώγου της προσεγγίζουσας συνάρτησης, είναι (α) παρεμβολή με «σπλίνες» (spline interpolation) και (β) συνέλιξη των δεδομένων της εικόνας με μια συνάρτηση που μπορεί αρκετά καλά να μοντελοποιηθεί από μια Gaussian. Για τις στενοζωνικές αλλά όχι απαραίτητα βαθυπερατές συνιστώσες εδώ, η Gaussian απόχριση διαμορφώνεται από ένα ημιτονοειδές με φέρουσα ίση με τη φασματική μέση θέση της συνιστώσας και προκύπτει το Gabor φίλτρο. Στο [79], η κανονικοποίηση με (splines) και Gabor του ενεργειακού τελεστή και του ESA συγχρίθηκαν για 1-Δ σήματα.

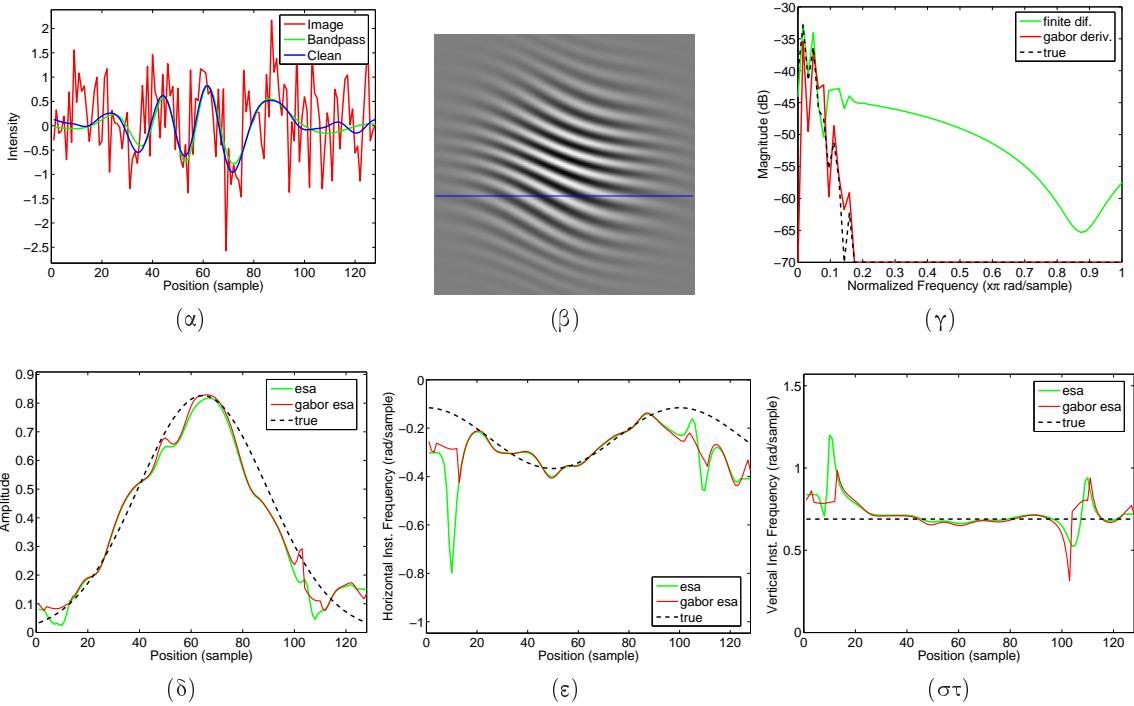
Με κίνητρο τα παραπάνω προτάθηκε ο 2-Δ Gabor ESA αλγόριθμος για ταυτόχρονο φιλτράρισμα/απομόνωση τοπικά ζωνοπερατών συνιστώσων υφής και αποδιαμόρφωση τους σε στιγμιαίο πλάτος και συχνότητα [92, 165]. Έστω  $I(x, y)$  η συνεχής εικόνα και  $f(x, y) = I(x, y) * g(x, y)$  η έξοδος της απόχρισης του Gabor φίλτρου. Εφόσον η συνέλιξη αντιμετατίθεται με τη διαφόριση ο συνεχής 2-Δ τελεστής συνδυασμένος με το ζωνοπερατό φιλτράρισμα γίνεται

$$\Psi(f) = \Psi(I * g) = \|I * \nabla g\|^2 - (I * g)(I * \nabla^2 g). \quad (3.19)$$

Έτσι, οι διαφορικοί τελεστές αντικαταστάθηκαν από παραγώγους του φίλτρου. Ο τελικός αλγόριθμος για τον Gabor ενεργειακό τελεστή (EO) συνοψίζεται στα ακόλουθα βήματα:

- (1) Βρίσκονται αναλυτικά και αποθηκεύονται όλες οι απαραίτητες διαφορικές φόρμουλες της Gabor συνάρτησης  $g(x, y)$  υπολογισμένες στις θέσεις (pixel) του πλέγματος δειγματοληψίας  $(i, j) = (i\Delta x, j\Delta y)$ . Απαιτούνται τρεις διαφορικές φόρμουλες:  $g_x, g_y, \nabla^2 g$ .
- (2) Για την εκτίμηση της στιγμιαίας ενέργειας στη θέση  $(i, j)$  χρησιμοποιείται η φόρμουλα (3.19) του συνδυασμένου συνεχούς τελεστή και του Gabor φιλτραρίσματος, όπου για κάθε όρο χρησιμοποιείται η διαχριτή συνέλιξη των δεδομένων  $I[i, j]$  και του αναγκαίου Gabor διαφορικού με δείγματα στα  $(i, j)$ .

Ομοίως, για την εκτίμηση του στιγμιαίου πλάτους και της στιγμιαίας συχνότητας ο 2-Δ Gabor ESA για την αποδιαμόρφωση του  $f = I * g$  ακολουθεί τα εξής βήματα:



**Σχήμα 3.7:** Κανονικοποιημένη και διακριτή αποδιαμόρφωση του θορυβώδους AM-FM  $f(m, n) = \exp\left(-\frac{m^2+n^2}{4\sigma^2}\right) \cos[u_c m + v_c n + 2 \sin(\frac{\pi}{50}m) + \sin(\frac{\pi}{30}n)]$ ,  $m, n = 1, \dots, 128$ . Οι παράμετροι  $(u_c, v_c) = (\frac{-2}{25}\pi, \frac{6}{25}\pi)$  rad/sample και  $\sigma = 0.08$  αντιστοιχούν σε ένα κανάλι της συστοιχίας (ακτινικής συχνότητας  $R = \frac{\pi}{4}$  και προσανατολισμού  $\theta = \frac{2\pi}{3}$ ). Η εικόνα αποδιαμορφώνεται μέσω (i) *ESA* αφού το σήμα έχει περάσει από το αντίστοιχο φίλτρο και (ii) *Gabor ESA* από την κρουστική απόκριση του καναλιού. (α) συνθετικό AM-FM σήμα, (β) 1-Δ προφίλ αρχικού, θορυβώδους ( $-10\text{dB}$ ) και ζωνοπερατού σήματος, (γ) φασματικό μέγεθος του  $I * g_x$  από πεπερασμένες διαφορές και Fourier μετασχηματισμό (η αναμενόμενη τιμή με διακεκομμένη γραμμή), (δ),(ε),(στ) αναμενόμενη διακεκομμένη και εκτιμώμενο πλάτος και συχνότητες.

- (1) Χρήση του Gabor EO για τον υπολογισμό των στιγμιαίων ενεργειών των τριών συναρτήσεων εικόνων:  $\Psi(f)$ ,  $\Psi(f_x = I * g_x) = \|I * \nabla g_x\|^2 - (I * g_x)(I * \nabla^2 g_x)$  και  $\Psi(f_y = I * g_y)$ .

- (2) Εισαγωγή των ενεργειών στη φόρμουλα του 2-Δ συνεχούς ESA.

Για τις τρεις ενέργειες που εμπλέκονται χρειάζονται επτά Gabor διαφορικές φόρμουλες:

$$g_x, g_y, g_{xx}, g_{yy}, g_{xy}, \nabla^2 g_x, \nabla^2 g_y.$$

Ο Gabor ESA είναι, επομένως, υπολογιστικά περισσότερο απαιτητικός απ' ό,τι ο αλγόριθμος με το διακριτό EO, καθώς χρειάζεται επτά συνελίξεις για κάθε Gabor της συστοιχίας, εν τούτοις προσφέρει ευρωστία και βελτιωμένη επίδοση.

Για αποδοτικότητα χρησιμοποιείται μια υλοποίηση του Gabor ESA στο πεδίο των συχνοτήτων (frequency-domain) με βάση την εξίσωση

$$\mathcal{F} \left\{ \frac{\partial^{k+l} g}{\partial x^k \partial y^l} \right\} = \mathcal{F} \{g\} (j\omega_x)^k (j\omega_y)^l \quad (3.20)$$

που σχετίζει το Fourier μετασχηματισμό  $\mathcal{F}\{\cdot\}$  ενός σήματος με τα διαφορικά του. Έτσι, το υπολογιστικό φορτίο για τις απαιτούμενες συνελίξεις μειώνεται σημαντικά, δεδομένου ότι ο

### Κεφάλαιο 3. Ενέργεια και Ανάλυση Υφής

σχεδιασμός της συστοιχίας γίνεται στο πεδίο των συχνοτήτων, ώστε να επιτρέπει φιλτράρισμα με πολλαπλασιασμό σημείου προς σημείο. Οποιοσδήποτε συνελικτικός όρος υλοποιείται από τη σχέση

$$I * \frac{\partial^{k+l} g}{\partial x^k \partial y^l} = \mathcal{F}^{-1} \left\{ \mathcal{F}\{I\} \mathcal{F}\{g\} (j\omega_x)^k (j\omega_y)^l \right\} \quad (3.21)$$

χρησιμοποιώντας το προϋπολογισμένο, κοινό για όλους τους όρους, γινόμενο  $\mathcal{F}\{I\} \mathcal{F}\{g\}$ .

Στο Σχ. 3.7 η ακριβής αναλυτική φόρμουλα είναι γνωστή για το συνθετικό AM-FM στο (α), μια 1-Δ γραμμή του οποίου φαίνεται στο (β). Το ημιτονοειδές έχει φέρουσες που συμπίπτουν με τη κεντρική ενός καναλιού της συστοιχίας και λευκός θόρυβος σε σηματοθορυβικό λόγο (SNR) ίσο με  $-10\text{dB}$  έχει προστεθεί τεχνητά στο καθαρό σήμα. Στο (γ) παρουσιάζεται μια σύγκριση ανάμεσα στο φασματικό μέγεθος 1) των  $\partial(I * g)/\partial x$ , που χρησιμοποιείται από τον ESA, υλοποιημένο με πεπερασμένες διαφορές και 2) το Fourier του (3.21) το οποίο χρησιμοποιείται από τον Gabor ESA. Το δεύτερο προσεγγίζει πολύ καλύτερα την αναμενόμενη απόκριση. Το αντίστοιχο φίλτρο  $g_x$ , το οποίο φιλτράρει στην ουσία την εικόνα  $I$ , παρουσιάζει μια μεγαλύτερη ζώνη επιτρεπτών συχνοτήτων με μικρότερη απόσβεση για συχνότητες εκτός του αρχικού εύρους. Οι εκτιμώμενες συναρτήσεις διαμόρφωσης στα (δ)-(στ) είναι περισσότερο ομαλές για τον Gabor ESA, με λιγότερες ακραίες τιμές (spikes) και ακολουθούν ακριβέστερα τις παραμέτρους διαμόρφωσης του καθαρού σήματος.

### 3.5 Ενέργεια διαμόρφωσης και χαρακτηριστικά υφής

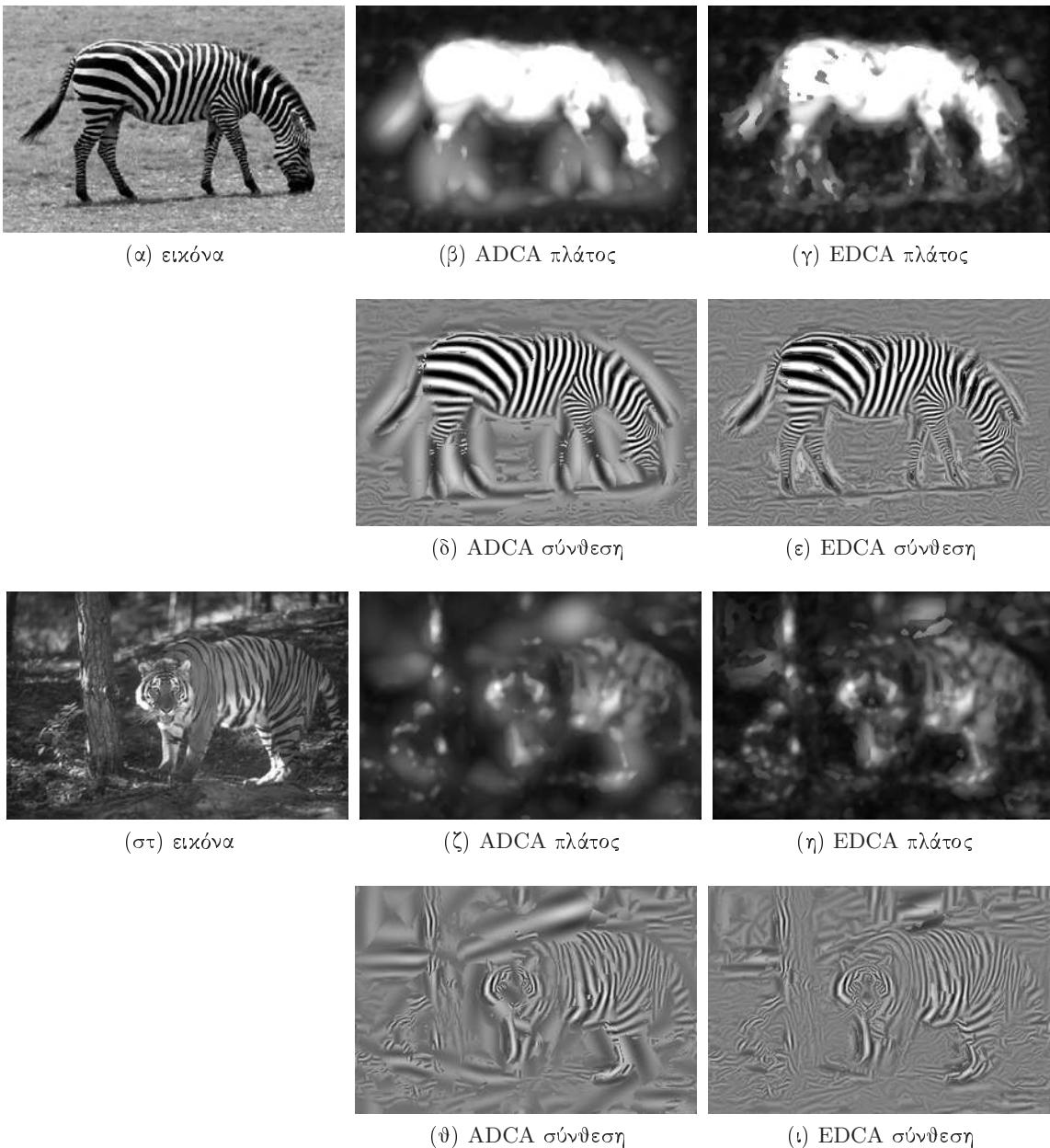
Ο μηχανισμός εντοπισμού της ενέργειας των συνιστώσων της υφής, που πραγματοποιεί η ανάλυση κυρίαρχων συνιστώσων, λειτουργεί στον πολυδιάστατο χώρο χαρακτηριστικών ο οποίος αποτελείται από τις αποκρίσεις των φίλτρων  $g_k$ . Οι φιλτραρισμένες, στενής ζώνης συνιστώσες υφής υποβάλλονται στη μη-γραμμικότητα του 2-Δ ενεργειακού τελεστή  $\Psi$  και αποδιαμορφώνονται μέσω του ESA (2.22, 2.23). Οι μετρήσεις ενέργειας ομαλοποιούνται με ένα τοπικό παράθυρο-φίλτρο μέσου όρου  $h_a$  και συγκρίνονται σημειακά. Το φίλτρο  $i$  με τη μέγιστη μέση Teager ενέργεια (maximum average Teager energy)

$$\Psi_{\text{mat}}[I(x, y)] = \max_k \{(\Psi(I * g_k) * h_a)(x, y)\}, \quad (3.22)$$

καθορίζει την κυρίαρχη συνιστώσα υφής.

**Ορισμός 3.2** *H* ενέργεια διαμόρφωσης υφής (*texture modulation energy*) ορίζεται από τη μέγιστη μέση ενέργεια διαμόρφωσης των τοπικά στενοζωνικών συνιστώσων της.

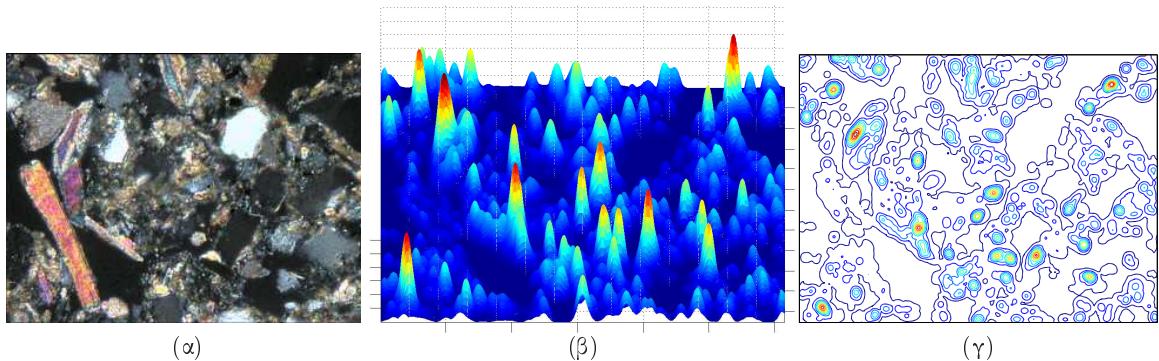
Η ενέργεια διαμόρφωσης είναι ένα μέτρο της έντασης της υφής και της χωρικής κατανομής της στα σημεία της εικόνας όπως αυτή περιγράφεται από το μοντέλο AM-FM διαμορφώσεων. Μπορεί, επομένως, να χρησιμοποιηθεί για ανίχνευση και διάκριση υφής χωρίς προσοχή (preattentively) με εφαρμογή του σύνθετου μη-γραμμικού τελεστή  $\Psi_{\text{mat}}$  στην απόκριση της συστοιχίας στην εικόνα. Καθώς οι διαφορετικές ενέργειαιακές στάθμες αντιστοιχούν σε διαφορετικά ποσά διαμόρφωσης πλάτους και συχνότητας, η ενέργεια μπορεί να χρησιμοποιηθεί



Σχήμα 3.8: DCA πλάτους και ενέργειας: συγκρίνοντας τα εκτιμώμενα πλάτη (ανάλυση) και τις συνιστώσες που ανακατασκευάζονται (σύνθεση) από το μοντέλο χρησιμοποιώντας τα εναλλακτικά κριτήρια επιλογής καναλιού (3.15) και (3.18) παρατηρούμε ότι το EDCA επικεντρώνεται σε πιο βασικές μεταβολές της υφής συνδυάζοντας πληροφορία έντασης και συχνότητας, ενώ το ADCA ευνοεί μεταβολές της εικόνας μεγάλης κλίμακας, οι οποίες δε γίνονται εύκολα αντιληπτές ως υφή.

ως χαρακτηριστικό ταξινόμησης για τη διάκριση δύο τάξεων υφής (π.χ. κατωφλιοποίηση με επίβλεψη, κεφάλαιο 6).

Παράλληλα, ένα σύνολο χαρακτηριστικών, το πλάτος κυρίαρχης διαμόρφωσης (*dominant modulation amplitude*)  $a_D(x, y)$  και το διάνυσμα συχνοτήτων κυρίαρχης διαμόρφωσης (*dominant modulation frequency vector*)  $\vec{\omega}_D(x, y)$  (3.14), εξάγονται από την εκτίμηση των παραμέτρων διαμόρφωσης του κυρίαρχου καναλιού. Σε κάθε θέση του χωρικού πεδίου της εικόνας η αναπαράσταση της υφής γίνεται από το χαμηλής διάστασης διάνυσμα κυρίαρχων διαμορφώσεων. Τα χαρακτηριστικά διαμόρφωσης (*modulation features*) τα οποία εξάγονται αναζητώντας



Σχήμα 3.9: Ενέργεια διαμόρφωσης και ανίχνευση υφής. (α) εικόνα εδαφοτομής ( $768 \times 567$ ), μέγιστη μέση Teager ενέργεια  $\Psi_{\text{mat}}$ , ως (β) 3-Δ προοπτική προβολή και (γ) υπέρθεση δέκα επιπεδοκαμπυλών της.

τις κυρίαρχες συνιστώσες υφής σε πολλαπλές ζώνες του φάσματος παρέχουν τοπική (περιοχές, ακμές, σύνορα) αλλά και ολική πληροφορία (ομοιογένεια, φασματικές ζώνες) για τη δομή της υφής και μπορούν να περιγράψουν ένα εύρος γενικών κατηγοριών υφής.

Η ενεργειακή αυτή θεώρηση εξετάστηκε σε θέματα ανίχνευσης και ανάλυσης της υφής και πολυχριτικής κατάτμηση εικόνων [268, 269], όπου η ενέργεια διαμόρφωσης μέσω του τελεστή  $\Psi_{\text{mat}}$  αποτέλεσε έναν καλά εντοπισμένο ενδείκτη των περιοχών και της έντασης της υφής (κεφάλαιο 7). Σημειώνεται ότι, όπως και στο σχήμα EDCA, όπου ο τελεστής  $\Psi_{\text{mat}}$  είναι το κριτήριο επιλογής συνιστωσών, η ενέργεια διαμόρφωσης προσφέρει βελτιωμένο εντοπισμό και διάκριση των ορίων διαφορετικών περιοχών σε σχέση με συμβατικούς ανιχνευτές υφής ή χαρακτηριστικών (πλάτος ή ενέργεια). Ευνοεί δε και αναδεικνύει τις συνιστώσες μικρότερης κλίμακας (ταχύτερης ταλάντωσης), καθώς συνυπολογίζει τη συχνότητα μαζί με το πλάτος στην ενεργειακή μέτρηση.

Στο Σχ. 3.9 παρουσιάζεται η ενέργεια διαμόρφωσης υφής (3.22) για μια πλούσια σε διαφορετικές υφές εικόνα λεπτής εδαφοτομής. Η μέτρηση εξάχθηκε από μια συστοιχία 40 φίλτρων, με μέσους όρους σε γειτονιά  $11 \times 11$  σημείων και γραμμική και μη-γραμμική ομαλοποίηση μεγέθους  $11 \times 11$  των κυρίαρχων χαρακτηριστικών. Η ενέργεια παρουσιάζεται ως προοπτική 3-Δ απεικόνιση, αλλά και με τη μορφή δέκα επιπεδοκαμπυλών της (level curves). Η ύπαρξη αλλά και τα επίπεδα έντασης της υφής ανά περιοχές στην εικόνα, χαρακτηρίζονται από την ενέργεια διαμόρφωσης.

□



## 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.ά. αποτελεί μια από τις σημαντικότερες διεργασίες όρασης χαμηλού επιπέδου. Το αντίστοιχο φιλοσοφικό και γνωσιακό πρόβλημα είναι γνωστό 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 χυρίαρχων συνιστωσών διαμόρφωσης μέσω πολυκαναλικής ανάλυσης και εντοπισμού της ενέργειας καταλήγει σε διανύσματα χαρακτηριστικών χαμηλής διαστασης που περιγράφουν την ένταση (πλάτος), την χλίμακα και τον προσανατολισμό (μέτρο και γωνία του διανύσματος) της υφής. Τέτοια χαρακτηριστικά εμπλέκονται και λειτουργούν ως κινητήριες δυνάμεις σε ένα μεταβολικό (variational) σχήμα κατάτμησης υφής χωρίς επισκόπηση με σταθμισμένη εξέλιξη καμπυλών.<sup>1</sup>

### 4.1 Κατάτμηση με εξέλιξη καμπυλών και χυρίαρχες διαμόρφωσεις

Η κατάτμηση εικόνων είναι ένα εν γένει δύσκολο πρόβλημα της υπολογιστικής όρασης καθώς απαιτεί σε κάποιο ποσοστό την σημασιολογική κατανόηση μιας εικόνας. Γενικά αναφέρεται στη διαδικασία του χωρισμού μιας εικόνας σε μη-συνεκτικές περιοχές έτσι, ώστε κάθε μια να είναι ομογενής και συνεκτική ως προς κάποια ιδιότητα, όπως η αντίθεση, το χρώμα, η υφή, η κίνηση, το βάθος. Μπορούν να διαχριθούν τρία διαφορετικά στάδια σε μια διαδικασία κατάτμησης 1) προεπεξεργασία, 2) εξαγωγή χαρακτηριστικών των περιοχών, 3) εφαρμογή κάποιου αλγορίθμου κατάτμησης, κάθε ένα από τα οποία επηρεάζει το τελικό αποτέλεσμα.

Η κατάτμηση εικόνων υφής είναι ένα από τα σημαντικότερα και πιο διάσημα προβλήματα αρχικής όρασης, που έχει αντιμετωπιστεί από διαφορετικές προοπτικές όπως στατιστικά μοντέλα [103, 196], γεωμετρικές και μορφολογικές μεθόδους [32, 266, 272] και μεταβολικές τεχνικές βασιζόμενες στα εσωτερικά των περιοχών (region-based) [177, 231, 315, 319] ή βασιζόμενες

<sup>1</sup>Η έρευνα και τα αποτελέσματα που παρουσιάζονται σε αυτή την ενότητα εντάσσονται στο πλαίσιο των εργασιών [92, 163–165] σε συνεργασία με τον Ιάσονα Κόκκινο [162].

στα όρια (boundary/edge-based) (π.χ. [188,239], snakes, deformable models [63,153], γεωδαιτικά ενεργά περιγράμματα [51,189,230]).

Στο region-based σενάριο κατάτμησης τα σημεία 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 εξελιγμένα 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 στα κεφάλαια 2 και 3. Παρά τον πληροφοριακό πλουραλισμό 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 segmentation [231]), η βέλτιστη κατάτμηση επιτυγχάνεται με την επιλογή των καναλιών που διαχωρίζουν καλύτερα τις διαφορετικές υφές. Στο σενάριο αυτόματης κατάτμησης χωρίς επισκόπηση (unsupervised segmentation), η επιλογή των καναλιών των χαρακτηριστικών και η μείωση της διάστασης του διανύσματος πραγματοποιείται συνήθως με ευρετικά χριτήρια [259,260]. Στο [256] μια διανυσματική διαδικασία διάχυσης (vector-valued diffusion) χρησιμοποιήθηκε για την απλοποίηση του διανύσματος χαρακτηριστικών.

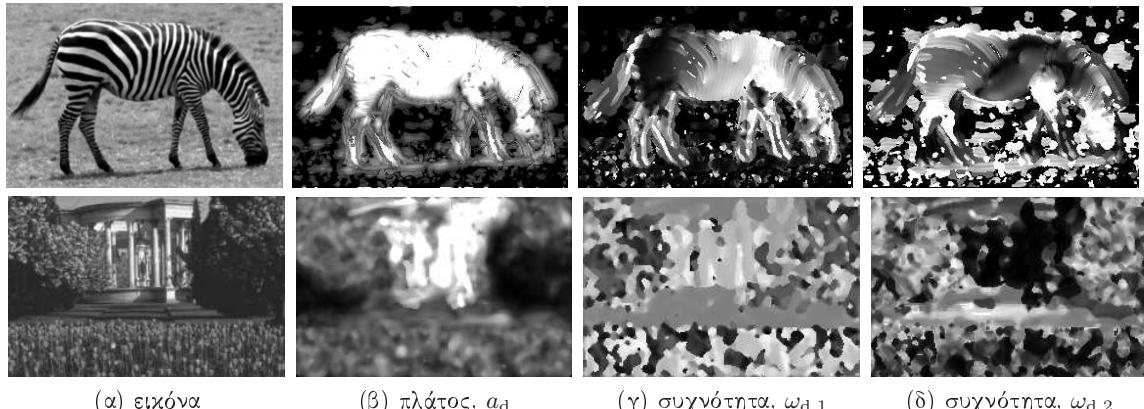
Τα χαρακτηριστικά διαμόρφωσης, μέσω του AM-FM μοντέλου, ερευνήθηκαν σε ένα σχήμα κατάτμησης χωρίς επισκόπηση [163,164]. Η ανάλυση κυρίαρχων συνιστώσων (3.3) χρησιμοποιείται για την εξαγωγή διανυσμάτων χαρακτηριστικών χαμηλής διάστασης που περικλείουν πληροφορία μεγέθους, κλίμακας και προσανατολισμού, 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 χριτήριο, με βάση τα εσωτερικά των περιοχών, όπως στα [232,319].

Μια πρόσφατη προσπάθεια σύζευξης μεταβολικών μεθόδων με χαρακτηριστικά διαμόρφωσης για κατάτμηση υφής [251,280] περιλαμβάνει ένα σχήμα γεωδαιτικών ενεργών περιγραμμάτων (geodesic active contours). Ωστόσο, το μεταβολικό κομμάτι χρησιμεύει ως βελτίωση στα αποτελέσματα κατάτμησης ενός καθαρά στατιστικού αλγορίθμου.

## 4.2 Χαρακτηριστικά διαμόρφωσης

Η υφή περιγράφεται μέσω δομών διαμόρφωσης, ψεωρώντας το σύνολο των χαρακτηριστικών που εξάγεται μέσω της ανάλυσης κυρίαρχων συνιστώσων. Εξετάστηκε τόσο η αρχική μέθοδος με βάση το πλάτος ή ADCA (εξ. 3.15), όσο και η εναλλακτική πρόταση βασισμένη στην ενέργεια ή EDCA (εξ. 3.18) που εντοπίζει ο ενεργειακός τελεστής. Στο Σχ. 4.1, φαίνεται το σύνολο των DCA χαρακτηριστικών υφής για δύο φυσικές εικόνες με περισσότερες από μία περιοχές διαφορετικής υφής (δύο για την επάνω και πέντε για την κάτω). Το πλάτος

αντιστοιχεί σε πληροφορία μεγέθους και έντασης ενώ το διάνυσμα κυρίαρχων συχνοτήτων αναδεικνύει κλίμακα και προσανατολισμό, όπως εύκολα παρατηρείται στο Σχ. 4.5.



Σχήμα 4.1: Χαρακτηριστικά διαμόρφωσης κυρίαρχων συνιστωσών ενέργειας (EDCA). Από δεξιά προς τα αριστερά και για τις δύο εικόνες: γκρίζα εικόνα, κυρίαρχο πλάτος, οριζόντια και κατακόρυφη στιγματικά συχνότητα.

#### 4.2.1 Η επιλογή καναλιού ως στατιστική απόφαση

Η επιλογή του καναλιού για την ανάλυση κυρίαρχων συνιστωσών διατυπώνεται ως πρόβλημα δοκιμών πολλαπλών υποθέσεων (multiple hypothesis testing problem) [154]. Για απλότητα εξετάζεται το μονοδιάστατο σενάριο, αν και η γενίκευση σε δύο διαστάσεις είναι απλή. Η επιλογή του καναλιού ανάγεται σε ανίχνευση σε ύδρυβο ενός ημιτόνου 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 της γειτονιάς ικανοποιούν την υπόθεση. Μακρινά σημεία μοντελοποιούνται από ένα λιγότερο περιγραφικό μοντέλο, συγκεκριμένα μια διαδικασία λευκού ύδρυβου.

Έστω η υπόθεση  $\mathcal{H}_\Omega$  ότι γύρω από ένα σημείο  $x = 0$ , το σήμα είναι ημιτονοειδές συχνότητας  $\Omega$ , άγνωστης φάσης  $\phi$ , πλάτους  $a$  και DC συνιστώσας (μέσης τιμής)  $b$ . 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,  $\sigma^2$  η μεταβλητή λευκού Gaussian ύδρυβου που παρεμβαίνει στο σήμα και το  $\sigma_G$  μοντελοποιεί την χωρική εξασθένηση της ικανότητας της υπόθεσης να εξηγεί τα δεδομένα. Ο δεύτερος όρος είναι το μοντέλο ύδρυβου παρασκηνίου (background noise), μέσης τιμής  $C$  και μεταβλητή της  $\sigma_n^2$ .

Θέτοντας το  $\sigma_G$  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  $\mathcal{H}_\Omega$  είναι:

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

Λογαριθμίζοντας και λαμβάνοντας υπόψιν 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:

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

Παρότι η τελευταία εξίσωση δίνει ένα κάτω όριο στην πιθανοφάνεια, μπορεί να χρησιμοποιηθεί για την επιλογή ανάμεσα στις δύο ανταγωνιστικές υποθέσεις. Μια δοκιμή γενικευμένου λόγου πιθανοφάνειας (Generalized Likelihood Ratio Test ή GLRT) χρησιμοποιείται για τον υπολογισμό εκτιμήσεων μέγιστης πιθανοφάνειας (Maximum Likelihood Estimates ή MLE) των παραμέτρων  $a, b, \phi$ . Μεγιστοποιώντας το (4.1) ως προς τα  $a, b, \phi$  προκύπτει

$$\hat{a} = \sqrt{\frac{(g_e * I)^2 + (g_o * I)^2}{(\sum_x G(x))^2}}, \quad \hat{\phi} = \arctan\left(-\frac{g_o * I}{g_e * I}\right), \quad \hat{b} = G * I, \quad (4.2)$$

όπου  $g_e, g_o$  ένα ζευγάρι αρτίου-περιπτού φίλτρου  $g_e(x) = G(x) \cos(\Omega x)$ ,  $g_o(x) = G(x) \sin(\Omega x)$ . Θεωρώντας τη συνάρτηση διασποράς  $G$  ίδια με αυτή του αντίστοιχου Gabor καναλιού, το  $\hat{a}$  είναι το εκτιμώμενο πλάτος απόχρισης ενώ το  $\hat{b}$  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 (3.16) για την επιλογή 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  $\mathcal{H}_0$  ότι γύρω ότι η εικόνα είναι τμηματικά σταθερή γύρω από το σημείο  $x = 0$ , απορρίπτοντας την ημιτονική συνιστώσα. Για να συγχριθεί 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) [154], 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 ως προς την ύπαρξη υφής παρουσιάζεται στο Σχ. 4.2, όπου οι δείκτες της υφής παρουσιάζονται ως γκρίζες εικόνες. Οι λευκές περιοχές

αντιστοιχούν σε περιοχές όπου η υπόθεση μη ύπαρξης υφής επικρατεί, ενώ οι γκρίζες 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 στη μύτη), ενώ ταυτόχρονα περιοχές μικρότερης κλίμακας ερμηνεύονται ως σταθερά σήματα με θόρυβο.

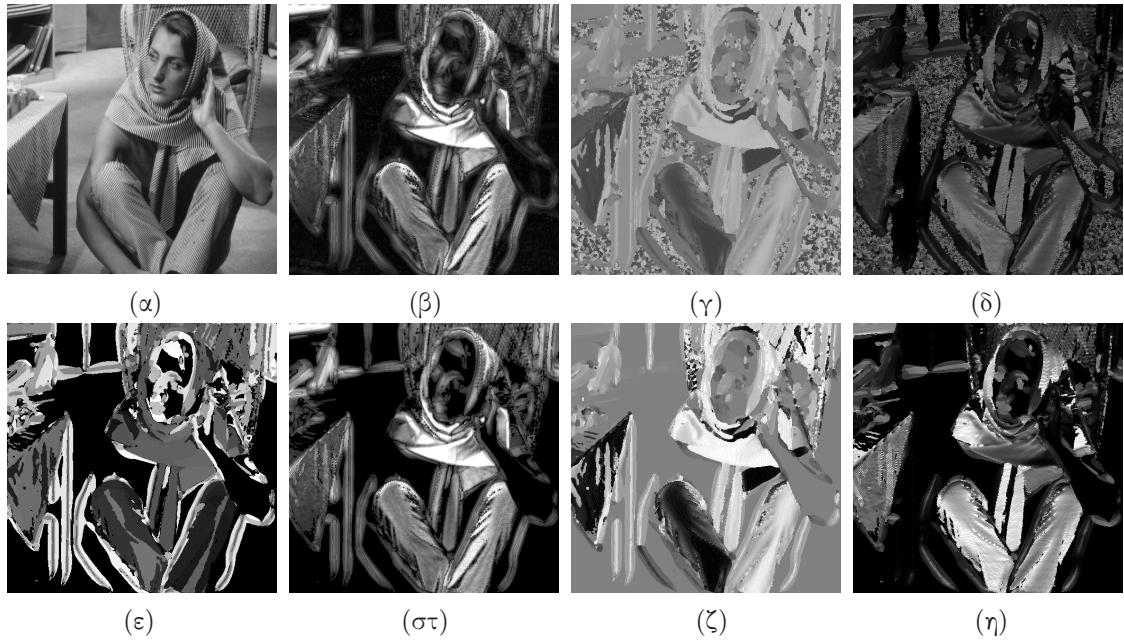


Σχήμα 4.2: Αποτελέσματα δοκιμών υποθέσεων σε τέσσερις κλίμακες για την εικόνα 'mandrill' του σχήματος 4.4(α). Οι κλίμακες, σε άυξουσα σειρά προς τα δεξιά, συμπίπτουν με αυτές της συστοιχίας των Gabor φίλτρων ενώ οι γκρίζες στάθμες είναι κβαντισμένες στον αριθμό των προσανατολισμών και λειτουργούν ως ενδείκτες υφής. Οι λευκές τιμές αντιστοιχούν σε ομαλές περιοχές, όπου τα δεδομένα δεν ερμηνεύονται ως υφή.

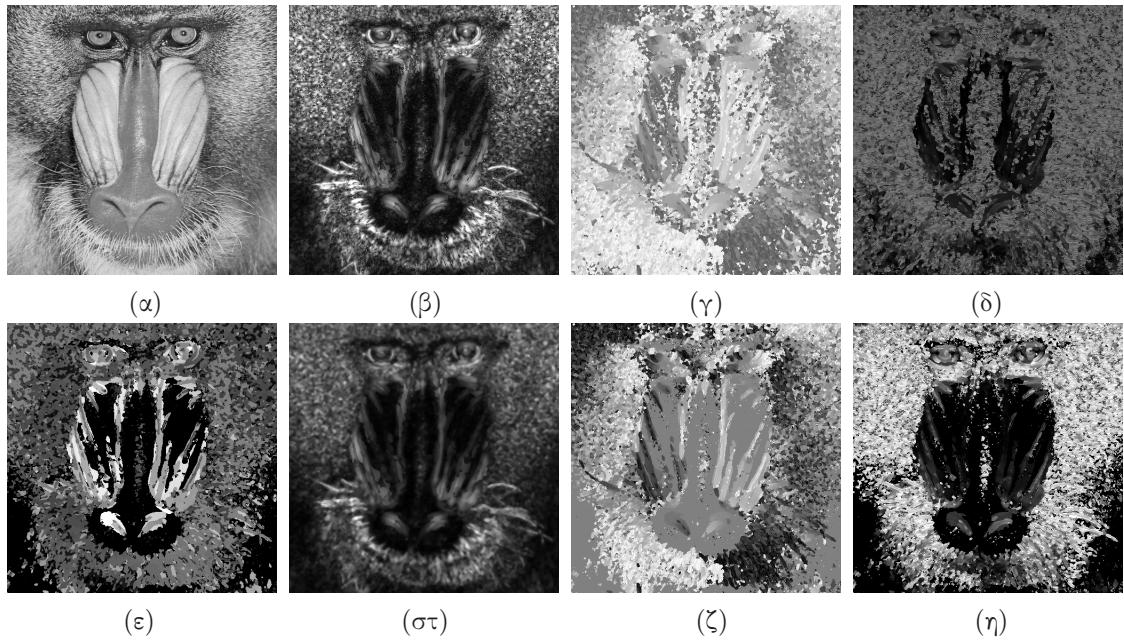
#### 4.2.2 Στατιστική ανίχνευση υφής και βελτιωμένα χαρακτηριστικά

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 σημείο της εικόνας μετά την επιλογή του καναλιού, υπολογίζεται ο λόγος πιθανοφάνειας της επιχρατούσας υπόθεσης και της υπόθεσης του μοντέλου μη-υφής να εξηγεί στην ίδια κλίμακα τα κοντινά δεδομένα. Σημεία όπου η δεύτερη υπόθεση επιχρατεί μηδενίζουν τα χαρακτηριστικά πλάτους και συγχνοτήτων τους. Τα βελτιωμένα χαρακτηριστικά μέσω DCA με βάση την ενέργεια, φαίνονται στα Σχ. 4.3,4.4. Όπως φαίνεται στη δεύτερη σειρά του Σχ. 4.3, η υπόθεση μη-ύπαρξης υφής κυριαρχεί στην περίπτωση του σταθερού δαπέδου, περιοχές οι οποίες αφαιρέθηκαν από τις αρχικές εκτιμήσεις των κυρίαρχων AM-FM συνιστώσων. Αυτό δηλώνουν και τα μαύρα pixel στο Σχ. 4.3(ψ), 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 ως ένα πρώτο στάδιο στατιστικής κατάτμησης και κατηγοριοποίησης σε περιοχές υφής και 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 [92] πραγματοποιήθηκε μια τέτοια ενσωμάτωση αλλά και μια σύμμιξη κριτηρίων υφής και ακμών, ανάλογα με την τοπική στατιστική ισχύ διαφορετικών μοντέλων.



Σχήμα 4.3: Χαρακτηριστικά διαμόρφωσης για την εικόνα ‘barb’ μέσω EDCA. (α)-(δ) γκρίζα εικόνα, κυρίαρχο πλάτος, οριζόντια και κάθετη συχνότητα, (ε) δείκτης κυρίαρχου καναλιού ανά σημείο (ως γκρίζα εικόνα), (στ)-(η) βελτιωμένα χαρακτηριστικά από την απόρριψη σημείων που χαρακτηρίζονται χωρίς υφή.



Σχήμα 4.4: Χαρακτηριστικά διαμόρφωσης για την εικόνα ‘mandrill’ μέσω EDCA. (α)-(δ) Γκρίζα εικόνα, κυρίαρχο πλάτος, οριζόντια και κάθετη συχνότητα, (ε) δείκτης κυρίαρχου καναλιού ανά σημείο (ως γκρίζα εικόνα), (στ)-(η) βελτιωμένα χαρακτηριστικά από την απόρριψη σημείων που χαρακτηρίζονται χωρίς υφή. Το πλάτος έχει υποστεί επιπλέον median φιλτράρισμα.

### 4.3 Μεταβολική κατάτμηση με εξέλιξη καμπυλών

Οι πρώτες region-based τεχνικές κατάτμησης υφής [177,319] βασίστηκαν σε εκδοχές του συναρτησιακού Mumford-Shah [221] κατάλληλα τροποποιημένες για την ενσωμάτωση πολυδιάστατων χαρακτηριστικών περιγραφής της υφής. Τρέχουσες τεχνικές κατάτμησης χωρίς

#### Κεφάλαιο 4. Κατάτμηση Υφής

επισκόπηση [41, 129, 232, 256, 297, 315] βασίζονται στη μεθοδολογία επιπεδοσυνόλων (level sets) [189, 226, 233, 263], που αποτελεί ένα κομψό και αποτελεσματικό μέσο επίλυσης προβλημάτων εξελισσόμενων καμπυλών.

Η μέθοδος ανταγωνισμού περιοχών (Region Competition ή RC [319]), 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  $R_i$ , εντός των οποίων οι παρατηρήσεις ακολουθούν υποθετικά μια παραμετρική κατανομή  $P(\cdot; \theta_i)$ . Ένας επιπλέον όρος για τα μήκη των συνόρων των περιοχών  $\Gamma = \{\Gamma_1, \dots, \Gamma_M\}$ , όπου  $\Gamma_i = \partial R_i$ , χρησιμοποιείται για να προκύψει το συναρτησιακό

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

Μέσω τεχνικών λογισμού μεταβολών προκύπτει η εξέλιξη των συνόρων των περιοχών, με κίνηση στην κατεύθυνση που κατατάσσει παρατηρήσεις-σημεία στην περιοχή που τα μοντελοποιεί καλύτερα, διατηρώντας ομαλά σύνορα

$$\frac{\partial \Gamma_i}{\partial t} = -\mu \kappa_i \vec{N}_i + \log \frac{P(F; \theta_i)}{P(F; \theta_j)} \vec{N}_i, \quad (4.5)$$

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

Στο [232] η εξέλιξη των καμπυλών συνδυάστηκε με όρους ακμών (edge-based) και μεθόδους επιπεδοσυνόλων στον αλγόριθμο γεωδαιτικών ενεργών περιοχών (Geodesic Active Regions ή GAR)

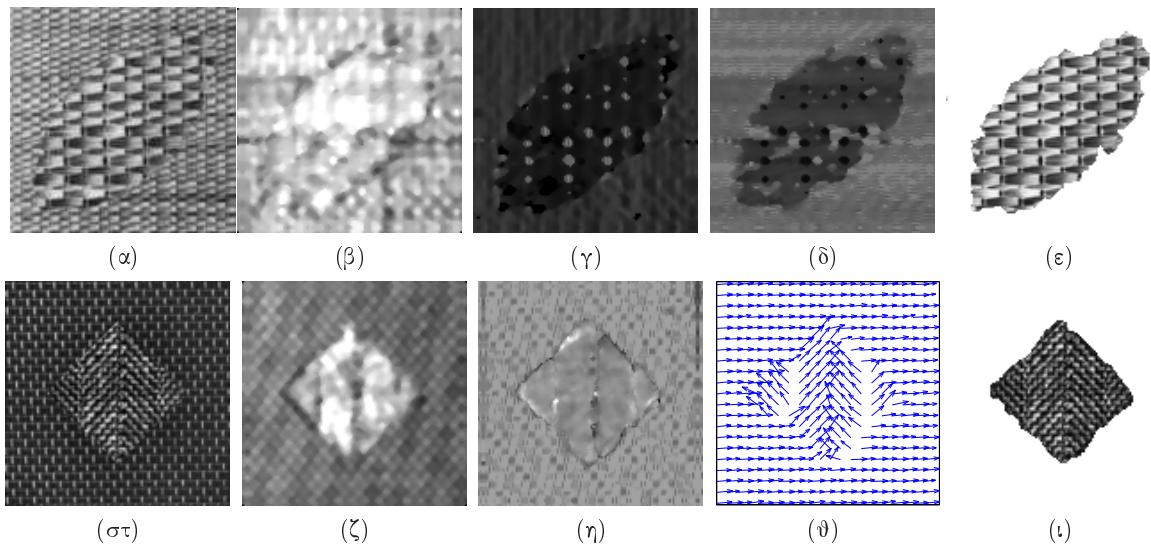
$$\frac{\partial \Gamma_i}{\partial t} = \lambda \log \frac{P(F; \theta_i)}{P(F; \theta_j)} \vec{N}_i - (1 - \lambda) \left[ g(I_E) \kappa_i \vec{N}_i + (\nabla g(I_E) \cdot \vec{N}_i) \vec{N}_i \right], \quad (4.6)$$

όπου  $g(\cdot)$  είναι μια μονοτονικά αύξουσα συνάρτηση,  $P_c$  η πιθανότητα ενός pixel να ανήκει σε ένα σύνορο και το  $\lambda$  καυθορίζει τα σχετικά βάρη των όρων περιοχών και των όρων ακμών. Ο τελευταίος όρος είναι εμπνευσμένος από το μοντέλο γεωδαιτικών ενεργών περιγραμμάτων (Geodesic Active Contours ή GAC [51, 156]) και ωθεί τα σύνορα κοντά σε περιοχές έντονων ακμών. Άλλοι αλγόριθμοι όπως οι [297, 315] μπορούν να υπερηφανούν της ίδιας φύσης με τον αρχικό RC.

#### 4.4 AM-FM ανταγωνισμός περιοχών

Το σχήμα κατάτμησης ανταγωνισμού περιοχών προσαρμόζεται και η αναζήτηση της βέλτιστης κατανομής σε περιοχές των σημείων της εικόνας, γίνεται με βάση τα κυρίαρχα χαρακτηριστικά διαμόρφωσης (AM-FM Region Competition). Η εξαγωγή γίνεται με την EDDA προσέγγιση (3.18) και τον εντοπισμό της κυρίαρχης ενέργειας διαμόρφωσης. Το διάνυσμα χαρακτηριστικών αποτελείται από την ένταση φωτεινότητας της εικόνας, το κυρίαρχο πλάτος διαμόρφωσης, μέτρο και τον προσανατολισμό του διανύσματος στιγμιαίων συχνοτήτων (3.14)

$$F = [a_d, |\vec{\omega}_d|, \angle \vec{\omega}_d, I].$$



Σχήμα 4.5: Πληροφορία υφής (μέγεθος, κλίμακα, προσανατολισμός) και κατάτμηση μέσω DCA. Πάνω σειρά: (α) υφές με διαφορετική κλίμακα, (β) - (δ) κυρίαρχο πλάτος και συχνότητες, (ε) κατάτμηση σε δύο περιοχές. Κάτω σειρά: (στ) υφές με διαφορετικό προσανατολισμό, (ζ)-(θ) κυρίαρχο πλάτος, μέτρο και προσανατολισμός του διανύσματος κυρίαρχης συχνότητας, (ι) αποτέλεσμα κατάτμησης.

Η ένταση  $I$  της εικόνας συμπεριλαμβάνεται καθώς εξακολουθεί να παραμένει σημαντικό κριτήριο για τις περιοχές χωρίς υφή [256].

Η εξέλιξη καμπυλών όπως καθορίζεται από την (4.4) υλοποιείται με μεθόδους επιπεδοσυνόλων παρόμοιας αρχιτεκτονικής με το [232]. Η κατανομή του διανύσματος χαρακτηριστικών σε κάθε περιοχή μοντελοποιείται, για τα τρία πρώτα χαρακτηριστικά, ως γινόμενο πολυμεταβλητών Gauss κατανομών της μορφής

$$P(F; \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} e^{-\frac{1}{2}(F-\mu_i)^T \Sigma_i^{-1} (F-\mu_i)} \quad (4.7)$$

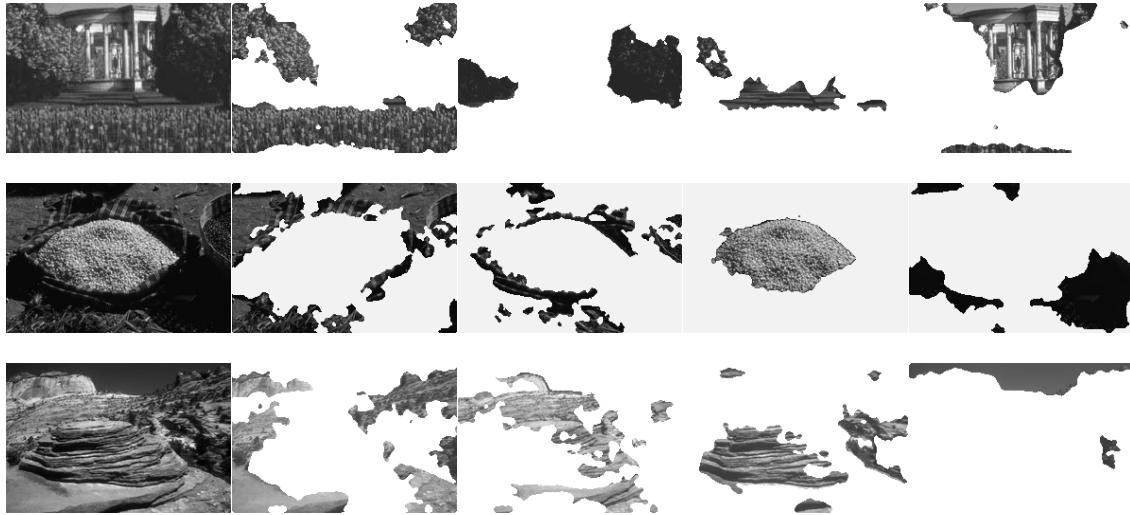
και  $d$  πλήνος χαρακτηριστικών και για τον προσανατολισμό ως κατανομή von-Mises. Η τελευταία είναι η αντίστοιχη της Gaussian για δεδομένα που περιέχουν κατεύθυνση.

Η εκτίμηση των παραμέτρων  $a_i = \{\mu_i, \Sigma_i\}$  των κατανομών και η κατανομή των δεδομένων σε κάθε περιοχή «μαθαίνεται» παράλληλα με την εξέλιξη των καμπυλών (fronts), με αποτέλεσμα ένα προσαρμοστικό, χωρίς επισκόπηση σχήμα κατάτμησης. Παραδείγματα μη-επιβλεπόμενης κατάτμησης για το αρχικό αυτό σχήμα με AM-FM χαρακτηριστικά διαμόρφωσης φαίνονται στα Σχ. 4.5,4.6. Οι καμπύλες αρχικοποιούνται με τυχαία τοποθέτηση, σε προεπιλεγμένο πλήνος.

Στις εικόνες 4.5(ε),(χ) φαίνεται το αποτέλεσμα της διαδικασίας στις δύο συνθετικές εικόνες υφής, όπου τα κεντρικά τμήματα διαφέρουν ως προς την κλίμακα στη μία και ως προς τον προσανατολισμό στην άλλη περίπτωση. Στην πρώτη σειρά της 4.6 ('house') οι στήλες λόγω του σχεδόν σταθερού πλάτους, Σχ. 4.1(β), ανιχνεύονται ως ξεχωριστή περιοχή όπως και τα σκαλιά λόγω της έντονης κατακόρυφης κυρίαρχης συχνότητας, Σχ. 4.1(γ). Παρόμοια συμπεράσματα και παρατηρήσεις εξάγονται και από τις άλλες δύο εικόνες ('rocks', 'rice'). Σημειώνεται ότι σημαντικό ρόλο παίζει η επιλογή του αρχικού αριθμού καμπυλών. Η αυτόματη

## Κεφάλαιο 4. Κατάτμηση Υφής

αρχικοποίηση αποτελεί δύσκολο πρόβλημα και σε κάθε περίπτωση διαφορετική επιλογή οδηγεί σε διαφορετική ομαδοποίηση των σημείων της εικόνας.



Σχήμα 4.6: Κατάτμηση υφής φυσικών εικόνων σε τέσσερις περιοχές ομοιόμορφων AM-FM κυρίαρχων χαρακτηριστικών διαμόρφωσης.

## 4.5 Σταθμισμένη εξέλιξη καμπυλών

Η εξαγωγή χαρακτηριστικών για κατάτμηση μπορεί να καταστεί προβληματική, όταν η υπόθεση για την αναλυόμενη κατηγορία σημάτων δεν ανταποκρίνεται στο πραγματικό περιεχόμενο της εικόνας (π.χ. τα χαρακτηριστικά υφής στα όρια των αντικειμένων, χαρακτηριστικά προσανατολισμού σε ομαλές περιοχές ή ένταση και ανίχνευση ακμών σε περιοχές υφής) [112, 186]. Χρησιμοποιώντας ημιτονοειδείς οικογένειες αναγεννητικών μοντέλων (generative models), οι διαφορετικές περιοχές χαρακτηρίζονται από ένα μετρικό βεβαιότητας (confidence measure) ως προς τις διαφορετικές κατηγορίες περιεχομένου που χρησιμοποιούνται για εξαγωγή χαρακτηριστικών. Το πλαίσιο για την επιλογή του καναλιού (4.2.1), γενικεύεται έτσι ώστε να συμπεριλαμβάνει και μοντέλα ακμών [166] για τη εξαγωγή εκτιμήσεων πιθανοτήτων που λειτουργούν ως βάρη σε ένα τροποποιημένο μεταβολικό σχήμα RC κατάτμησης [92, 162, 165].

### 4.5.1 Αναγεννητικά μοντέλα για υφή και ακμές

Τιοθετούμε την υπόθεση ότι κάθε γειτονιά μια εικόνας μπορεί να περιγραφεί σαν να ανήκει σε μια από τις τάξεις υφής, ακμών, και ομαλών περιοχών, χρησιμοποιώντας προκαθορισμένα αναγεννητικά μοντέλα [112, 225, 320, 321]. Θεωρώντας τις 3 πιθανές τάξεις  $C_i, i \in 1, 2, 3$  και την παρατήρηση (δεδομένα εικόνας)  $O$ , κάθε τάξη χρησιμοποιεί ένα διάνυσμα παραμέτρων  $\mathbf{E}_i$  και προσεγγίζει τις παρατηρήσεις μέσω μιας εξίσωσης σύνθεσης της μορφής  $O(x) \simeq I_i(x|\mathbf{E}_i)$ .

Η σύνθεση (πρόβλεψη) για το μοντέλο υφής γύρω από το σημείο  $x$  εκφράζεται ως ημιτονοειδές συχνότητας (κλίμακας)  $\omega$ , σταθεράς φάσης  $\phi$ , πλάτους  $A$  και DC τιμής  $B$ :

$$O(x) \simeq I_T(x|\{A, \phi, B\}) = A \cos(\omega x + \phi) + B, \quad \mathbf{E} = \{A, \phi, B\}. \quad (4.8)$$

Η ακρίβεια της πρόβλεψης μειώνεται με την απομάκρυνση από το σημείο  $x$  θεώρηση που εισάγεται ως

$$P(O(x)|x, \mathbf{E}) = G(x)P(O(x)|I_T(x|\mathbf{E})) + (1 - G(x))c, \quad (4.9)$$

όπου  $G(x)$  ένας χωρικά φθίνων όρος στάθμισης που αντιπροσωπεύει την τοπικότητα της διαδικασίας μοντελοποίησης,  $P(O(x)|I_T(x|\mathbf{E}))$  μια Gaussian κατανομή μέσης τιμής  $I_T(x)$  και άγνωστης μεταβλητότητας και  $c$  ένας σταθερός όρος ο οποίος συνεισφέρεται από το ομοιόμορφο μοντέλο του παρασκηνίου. Χρησιμοποιώντας ένα ζεύγος Gabor φίλτρων ορθογωνισμού φάσης  $g_e = \cos(\omega x)G(x)$  και  $g_o = \sin(\omega x)G(x)$  στην ίδια συχνότητα  $\omega$ , αποδείχθηκε [163, 166] ότι η εκτίμηση πλάτους  $\sqrt{(O * g_e)^2 + (O * g_o)^2}$  σχετίζεται με την πιθανοφάνεια των παρατηρήσεων της εικόνας σε αυτό το πιθανοτικό μοντέλο (Εν. 4.2.1). Η επιλογή του καναλιού με το μεγαλύτερο πλάτος (DCA) ισοδυναμεί με την επιλογή του καναλιού που «εξηγεί» καλύτερα τα δεδομένα. Η γενίκευση για την επιλογή του καναλιού με βάση τον ενεργειακό τελεστή (Ε-DCA) γίνεται θεωρώντας Brownian θόρυβο στα δεδομένα της εικόνας, όποτε και η εκτίμηση πλάτους γίνεται  $A\omega^2$  [166].

Η εξίσωση σύνθεσης ακμών βασίζεται στην παρατήρηση ότι οι ακμές γίνονται αντιληπτές ως σήματα με συμφωνία φάσης (*phase congruency*) [220], δηλαδή προσεγγίζονται από μια σειρά Fourier της μορφής

$$O(x) \simeq I_E(x) = A \sum_k a_k \cos(\omega_0 kx + \phi) + B, \quad (4.10)$$

όπου  $I_E$  το μοντέλο για την τάξη ακμών και  $\omega_0$  η θεμελιώδης συχνότητα της σειράς. Για  $\phi = \pi/2$  και  $\phi = 0$ , βηματικές ακμές (step edges) και ακμές γραμμές (bar, line edges) γίνονται αντιληπτές αντιστοίχως.

Χρησιμοποιώντας ένα ζεύγος ενός άρτιου και ενός περιπτού φίλτρου  $(h_e, h_o) = (G(x) \sum_k a_k \cos(\omega_0 kx), G(x) \sum_k a_k \sin(\omega_0 kx))$ , καταλήγουμε σε μια εκτίμηση του πλάτους που σχετίζεται με την πιθανοφάνεια των παρατηρήσεων κάτω από το μοντέλο (4.10) [166]. Θεωρώντας το μέγεθος του φάσματος του  $h_e + jh_o$  αμελητέο για αρνητικές συχνότητες, μπορούμε να θεωρήσουμε τα φίλτρα αποτελούν ζεύγος Hilbert, κι επομένως ζεύγος ορθογωνισμού φάσης. Τέτοια φίλτρα χρησιμοποιούνται για ανίχνευση ακμών [219, 239] μέσω της εκτίμησης της τοπικής ενέργειας  $(h_e * O)^2 + (h_o * O)^2$ , μιας ποσότητας ανεξάρτητης του  $\phi$ . Έτσι, μπορεί να δειχθεί ότι η τετραγωνισμένη τοπική ενέργεια  $(h_e * I)^2 + (h_o * I)^2$  μετράει τοπικά την προσαρμογή του μοντέλου ακμών (4.10) στα δεδομένα.

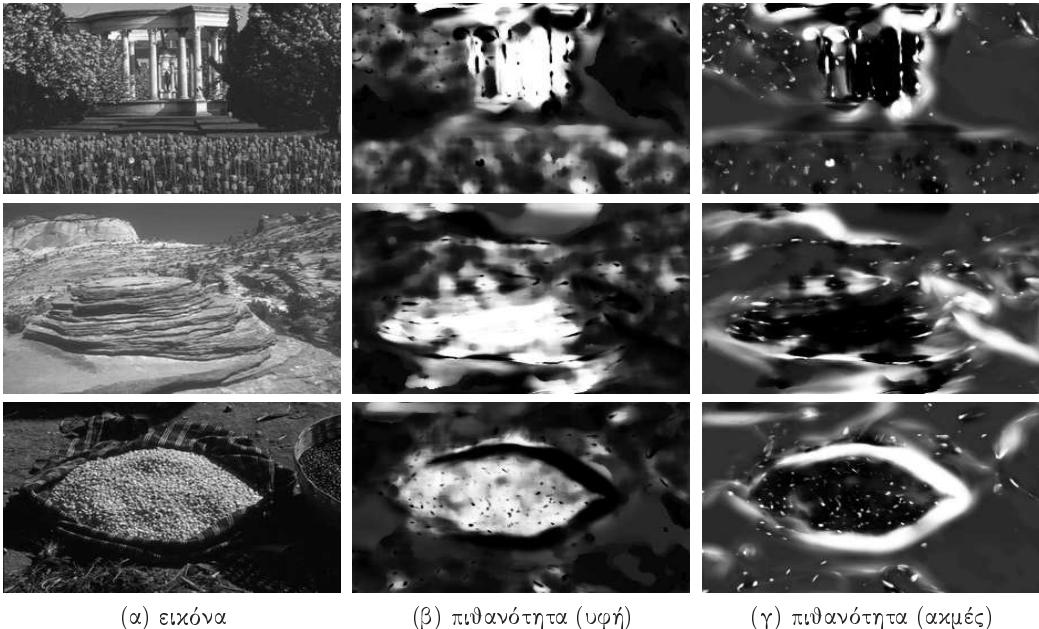
Οι δύο τάξεις μοντέλων συμπληρώνονται από την τάξη για «ομαλά» σήματα που μοντελοποιούνται ως σταθερά σήματα

$$O(x) \simeq I_S(x) = B. \quad (4.11)$$

Από τις τρεις τάξεις μοντέλων μπορούν να υπολογιστούν οι εκ των υστέρων πιθανότητες (class posterior probabilities) ενός εικονοστοιχείου να έχει «αναγεννηθεί» από μια από τις τρεις πιθανές υποθέσεις, μέσω του κανόνα του Bayes [165]. Εφαρμογή σε φυσικές εικόνες δίνει αντιληπτικά λογικά αποτελέσματα (Σχ. 4.7). Σε αυτές τις εικόνες η πιθανότητα ακμών

#### Κεφάλαιο 4. Κατάτμηση Υφής

είναι μεγαλύτερη στα σύνορα των αντικειμένων, ενώ η πιθανότητα υφής είναι μειωμένη καθώς το μοντέλο ακμών εξηγεί καλύτερα τη μεταβολή του σήματος.



Σχήμα 4.7: Εικόνες και εκ των υστέρων πιθανότητες (class posteriors) του μοντέλου (β) υφής και (γ) ακμών (από το [165]).

#### 4.5.2 Πιθανοτικός συνδυασμός χαρακτηριστικών

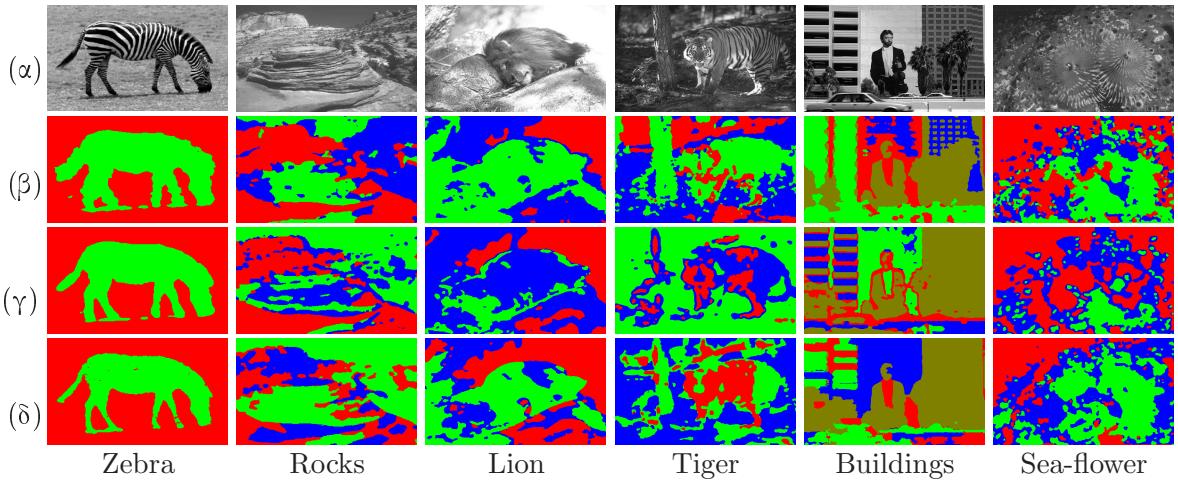
Οι εκ των υστέρων πιθανότητες των τάξεων χρησιμοποιούνται στον αλγόριθμο ανταγωνισμού περιοχών. Καθώς οι λεπτομέρειες της μεθόδου είναι έξω από το αντικείμενο της παρούσας διατριβής, ωστε παρουσιαστεί εδώ η βασική ιδέα παραπέμποντας για πλήρη διατύπωση και αιτιολόγηση στο [165].

Η ιδέα βασίζεται σε Bayesian μεθόδους σύμμικης (fusion), που συνδυάζουν τις εξόδους διαφορετικών ταξινομητών ως τυχαίες μεταβλητές θεωρώντας την κοινή κατανομή τους. Εν συντομίᾳ, θεωρούνται τρεις ταξινομητές που καθορίζουν την εξέλιξη της καμπύλης: ένας βασισμένος στη φωτεινότητα, ένας στα χαρακτηριστικά υφής και ένας που ομαδοποιεί τα σημεία σε διαφορετικές τάξεις κατά μήκος των ακμών (όπως ο όρος  $g\kappa\vec{N} + (\nabla g \cdot \vec{N})\vec{N}$  στο (4.6)).

Για δύο τάξεις, δηλαδή περιοχές ανάμεσα στην καμπύλη  $i$  και  $j$  και ένα σύνολο χαρακτηριστικών  $F$  κατασκευάζονται οι λόγοι πιθανοφάνειας  $\log \frac{P(F|i)}{P(F|j)}$  που θεωρούνται τυχαίες μεταβλητές Gaussian κατανομής. Τέτοιοι λόγοι αποφάσεων εμφανίζονται στο RC σχήμα (4.6). Οι κατανομές τους ποσοτικοποιούν την βεβαιότητα που συνοδεύει την απόφαση του ταξινομητή για την απόδοση των δεδομένων στη μια τάξη-περιοχή ή στην άλλη. Οι λόγοι για τους δύο πρώτους ταξινομητές είναι

$$\mathcal{L}_T = \log \frac{P(F_T; \theta_{T,i})}{P(F_T; \theta_{T,j})}, \quad \mathcal{L}_S = \log \frac{P(F_S; \theta_{S,i})}{P(F_S; \theta_{S,j})}, \quad (4.12)$$

όπου  $P(F_M; a_{M,i})$  η πιθανοφάνεια των χαρακτηριστικών  $F_M$  που σχετίζεται με την τάξη  $M$  (υφή-T ή ομαλή-S) κατανομής  $P$  για την περιοχή  $i$ , της οποίας οι παράμετροι είναι  $\theta_{M,i}$ .



Σχήμα 4.8: Αποτελέσματα κατάτμησης χρησιμοποιώντας διαφορετικά χαρακτηριστικά και μεταβολικά σχήματα: (α) αρχική εικόνα και κατάτμηση με (β) DCA χαρακτηριστικά και φωτεινότητα με το αρχικό AMFM-RC σχήμα (Εν. 4.4), (γ) τα χαρακτηριστικά διάχυσης του [256], (δ) DCA χαρακτηριστικά και το σταθμισμένο RC σχήμα (4.13).

Η απόφαση του κάθε ταξινομητή σταθμίζεται από την εκ-των υστέρων πιθανότητα της κάθε υπόθεσης· έτσι για περιοχές υφής τα χαρακτηριστικά υφής έχουν μεγαλύτερο αντίκτυπο στην εξέλιξη της καμπύλης από ότι η φωτεινότητα και το αντίθετο για ομαλές περιοχές. Από την εξίσωση (4.6) προκύπτει έτσι το σταθμισμένο σχήμα εξέλιξης καμπυλών :

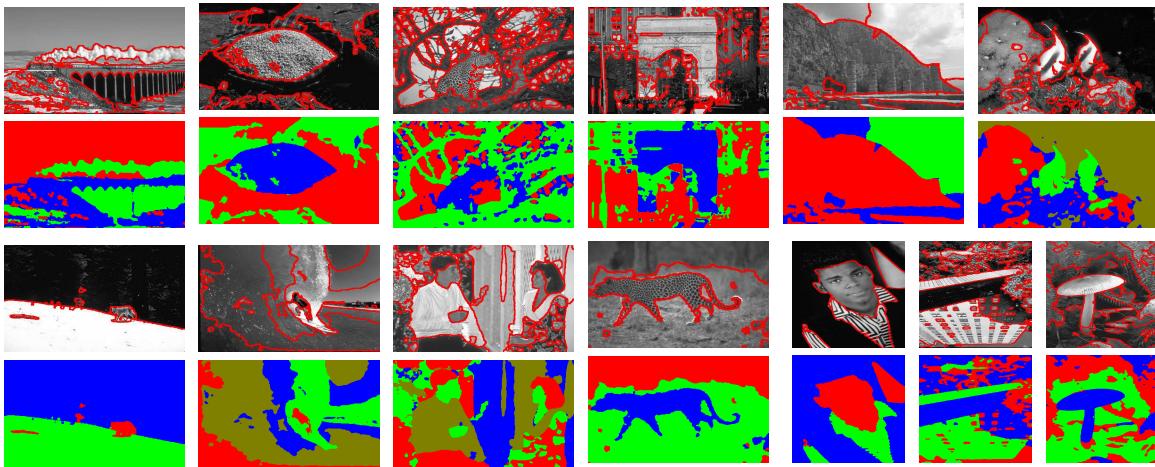
$$\frac{\partial \Gamma_i}{\partial t} = \left[ \sum_{c \in T, S} w_c \log \frac{P(F_c; \alpha_{c,i})}{P(F_c; \alpha_{c,j})} - w_E [g \kappa \vec{N} + (\nabla g \cdot \vec{N}) \vec{N}] \right] \vec{N}, \quad (4.13)$$

όπου ως βάρη  $w_E, w_T, w_S$  χρησιμοποιούνται οι εκ των υστέρων πιθανότητες των υποθέσεων ακμών, υφής και ομαλών περιοχών. Η στάθμιση με το  $w_E$  δεν επηρεάζει τη διαδικασία ανίχνευσης ακμών που χρησιμοποιείται για την εκτίμηση του  $\nabla g$ , καθώς απλά καθορίζει τη σημαντικότητα της πληροφορίας ακμών.

#### 4.5.3 Αποτελέσματα σταθμισμένης κατάτμησης

Το σχήμα σταθμισμένης κατάτμησης δοκιμάστηκε σε ένα πλήθος φυσικών εικόνων από τη βάση Berkeley Segmentation Dataset [207, 208]. Παρουσιάζονται ενδεικτικά τα αποτελέσματα κατάτμησης με τα DCA χαρακτηριστικά που αναφέρθηκαν στο [165] όπου συνοδεύονται από αναλυτική ποσοτική αποτίμηση σε δοκιμές επιδόσεων (benchmark).

Στα Σχτ. 4.8, 4.9 παρουσιάζονται αποτελέσματα κατάτμησης χρησιμοποιώντας τρία εναλλακτικά σχήματα: στη δεύτερη σειρά, τα αποτελέσματα του μεταβολικού σχήματος AM-FM RC [164], όπως περιγράφηκε στην ενότητα 4.4, χρησιμοποιώντας τα EDCA χαρακτηριστικά υφής, ενώ στην τρίτη χρησιμοποιείται το σύνολο χαρακτηριστικών μη-γραμμικής διάχυσης [256] με το ίδιο μεταβολικό σχήμα. Στην τελευταία σειρά φαίνονται τα αποτελέσματα του σταθμισμένου σχήματος με την πιθανοτική σύμμικη των χαρακτηριστικών. Για όλα τα σχήματα ο αριθμός των καμπυλών επιλέχθηκε με επισκόπηση αλλά τα αποτελέσματα είναι παρόμοια και για διαφορετικό αριθμό καμπυλών.



Σχήμα 4.9: Κατάτμηση εικόνων με DCA χαρακτηριστικά και σταθμισμένη εξέλιξη καμπυλών.

Σχετικά με την πρώτη προσέγγιση [164] παρατηρούμε ότι ενώ τα αποτελέσματα είναι ικανοποιητικά σε εικόνες με έντονη υφή (π.χ. tiger, sea-flower), σε ομαλές περιοχές η εσφαλμένη συμπεριφορά του διανύσματος προσανατολισμού οδηγεί σε μη-ομαλά σύνορα των περιοχών. Επιπλέον, στην εικόνα zebra το μεγάλο πλάτος της υφής στα σύνορα της φιγούρας «μεγενθύνει» την περιοχή πρώτου πλάνου, καθώς η ανταγωνιστική υπόθεση του φόντου χαρακτηρίζεται από χαρακτηριστικά μικρού πλάτους και δε μπορεί να «εξηγήσει» τις παρατηρήσεις αυτές.

Τα χαρακτηριστικά διάχυσης [256] οδηγούν σε συγκρίσιμα αποτελέσματα με αυτά των EDCA χαρακτηριστικών, αλλά πάσχουν από παρόμοια προβλήματα στις ομαλές περιοχές. Σε πολλές εικόνες (rocks, lion, tiger Sq. 4.8), τα σύνορα των αντικειμένων ομαδοποιούνται σε μια ζεχωριστή περιοχή, καθώς οι παράγωγοι στα σύνορα των αντικειμένων αντιμετωπίζονται ως υφή.

Τα αποτελέσματα του σταθμισμένου σχήματος φαίνονται ανεπηρέαστα από την σύγχυση των χαρακτηριστικών υφής και ακμών· καλύτερες κατατμήσεις λαμβάνονται, με τα όρια των περιοχών να εντοπίζονται με ακρίβεια τα σύνορα των αντικειμένων. Η επίδραση των χαρακτηριστικών προσανατολισμού σε ομαλές περιοχές περιορίζεται, λόγω του μικρότερου βάρους που αποδίδεται στη υπόθεση υφής εκεί. Σημειώνονται και κάποιες εξαιρέσεις, όπως για παράδειγμα στην εικόνα ‘bulidings’ στο Σχ. 4.8, όπου η υφή δεν είναι αρκετά έντονη και οδηγεί σε υπερκατάτμηση στο σταθμισμένο σχήμα.

Τέλος αναφορικά με τις ποσοτικές δοκιμές στη βάση Berkeley, χρησιμοποιήθηκαν και συγκρίθηκαν διαφορετικά χαρακτηριστικά αλλά και σχήματα κατάτμησης ως προς τις τιμές μετρικών που εκφράζουν α) τη σύμπτωσης ανθρώπινης και υπολογιστικής κατάτμησης, β) την ακρίβεια των συνόρων των περιοχών κατάτμησης. Και για τα δύο, τα αποτελέσματα πιστοποίησαν τόσο την αποδοτικότητα των DCA χαρακτηριστικών σε σχέση με τα χαρακτηριστικά διάχυσης [256], όσο και τη χρησιμότητα του σταθμισμένου σχήματος συνδυασμού των χαρακτηριστικών. Τόσο ζεχωριστά όσο και σε συνδυασμό επιφέρουν συστηματικές βελτιώσεις της απόδοσης. Ακόμη, η επίδοση της βασική μεθόδου AM-FM RC ξεπέρασε αυτή της μεθόδου αναφοράς Normalized Cuts [208].

#### **4.5.4 Συμπεράσματα**

Τα χαρακτηριστικά κυρίαρχων διαμορφώσεων δοκιμάστηκαν ως περιγραφείς της υφής μιας εικόνας σε ένα σχήμα μεταβολικής κατάτμησης με ανταγωνισμό περιοχών. Τα πλεονεκτήματα της μεθόδου πηγάζουν από τρεις κυρίως κατευθύνσεις: α) τα χαρακτηριστικά διαμόρφωσης εμπεριέχουν πληροφορία κλίμακας, αντίθεσης και προσανατολισμού της υφής σε ένα διάνυσμα χαμηλής διάστασης, β) η ερμηνεία μέσω αναγεννητικών μοντέλων ενοποιεί την αντιμετώπιση των διαφορετικών κατηγοριών δεδομένων (υφή, ακμές, ομαλές περιοχές) και την εξαγωγή πιθανοτήτων για τις τάξεις υφής και ακμών στα διαφορετικά σημεία της εικόνας, γ) το μοντέλο εξαγωγής μπορεί να χρησιμοποιηθεί για να συνθέσει την πληροφορία υφής της εικόνας. Η οπική αλλά και ποσοτική αποτίμηση, τόσο των χαρακτηριστικών όσο και του πιθανοτικού σταθμισμένου σχήματος αναδεικνύει την ικανότητα των μεθόδων για κατάτμηση φυσικών εικόνων με υφή, αλλά και το δυναμικό του AM-FM μοντέλου για ανάλυση της υφής σε εφαρμογές όρασης ανωτέρου επιπέδου.

□

## Κεφάλαιο 5

# Συνέργεια διαχωρισμού και μοντελοποίησης: Αποσύνθεση εικόνων

Κάθε εικόνα αποτελεί ένα σύνολο σημασιολογικών δομών και οντοτήτων, από τις οποίες η κάθε μια έχει τη δική της περιγραφή και γίνεται αντιληπτή διαφορετικά. Η υφή είναι η μικροδομή που συνυπάρχει με μακροδομές όπως είναι τα αντικείμενα, τα περιγράμματα και τα σχήματα τους. Τα δύο είδη πληροφορίας μοντελοποιούνται από μαθηματικές συναρτήσεις με διαφορετικά χαρακτηριστικά και σύμφωνα με το  $u + v$  μοντέλο, υπερτίθενται για το σχηματισμό μιας εικόνας  $f$ . Η μία είναι τμηματικά ομαλή, με απότομες μεταβολές ανάμεσα σε επίπεδα σχεδόν σταθερής φωτεινότητας και η άλλη, η υφή, από ταχύτατες μεταβολές, πρότυπα σε μικρές κλίμακες και ταλαντώσεις. Η διαδικασία διαχωρισμού υφής από την τμηματικά ομαλή συνιστώσα και τις μακροδομές ονομάζεται αποσύνθεση ή αποδόμηση εικόνων (image decomposition). Το πολυκαναλικό μοντέλο υφής και το μεταβολικό πλαίσιο αποδόμησης εικόνων, ενώνονται σε ένα σχήμα συνέργειας όπου η μία διαδικασία χρησιμοποιεί την άλλη για βελτιωμένο διαχωρισμό και ανάλυση υφής.

### 5.1 Δομή, υφή και μοντέλα

Σε φυσικές εικόνες η πληροφορία της υφής συνυπάρχει με γεωμετρικές μακροδομές όπως τα σύνορα των αντικειμένων ή οι ακμές ή εμφανίζεται εμβυθισμένη σε περισσότερο αδρές δομές και κλίμακες που σχηματίζονται από τις συνθήκες φωτισμού και σκίασης ή από μεταβολές ομαλών περιοχών/όγκου. Η αποδόμηση εικόνων [221] αναφέρεται στη διαδικασία διαχωρισμού μιας δοσμένης εικόνας σε εννοιολογικά και θεωρητικά διαφορετικές συνιστώσες, κυρίως ως προς δύο προοπτικές 1) την αντιμετώπιση προβλημάτων ανάλυσης εικόνων και όρασης (π.χ. ανάλυση δομής και λεπτομερειών) και 2) τη μελέτη των διαδικασιών σχηματισμού προτύπων εικόνων (π.χ. πολυεπίπεδες αναπαραστάσεις). Πολλά ανάστροφα προβλήματα όρασης μπορούν να θεωρηθούν ειδικές περιπτώσεις αποσύνθεσης: σχήματα διάχυσης (diffusion schemes) και

κλίμακας-χώρου (scale-spaces) [107, 179, 212, 239, 286, 308], τεχνικές απλοποίησης [221, 258], χυματίδια (wavelets) και μετασχηματισμοί ridgelets/curvelets [176, 191, 214], και λεξικοπαγείς αναπαραστάσεις σε βάσεις [69, 112, 178, 213, 273, 274, 296, 321] (μετασχηματισμοί, τεξτόνια, primitives). Σε αυτές τις περιπτώσεις οι ξεχωριστές συνιστώσεις είναι αντίστοιχα πολυεπίπεδες πυραμίδες, κανονικοποιημένες συναρτήσεις και residuals, μετασχηματισμοί και αρχέγονα δομικά στοιχεία μέσω εχμάλησης. Ομοίως, οι περισσότερες αντίστροφες διαδικασίες αποκατάστασης όπως η απομάκρυνση θορύβου [6, 53, 108, 223, 258] ή clutter [323] εντάσσονται στο πλαίσιο διαχωρισμού υφής από (μακρό)δομή [214, 221, 298].

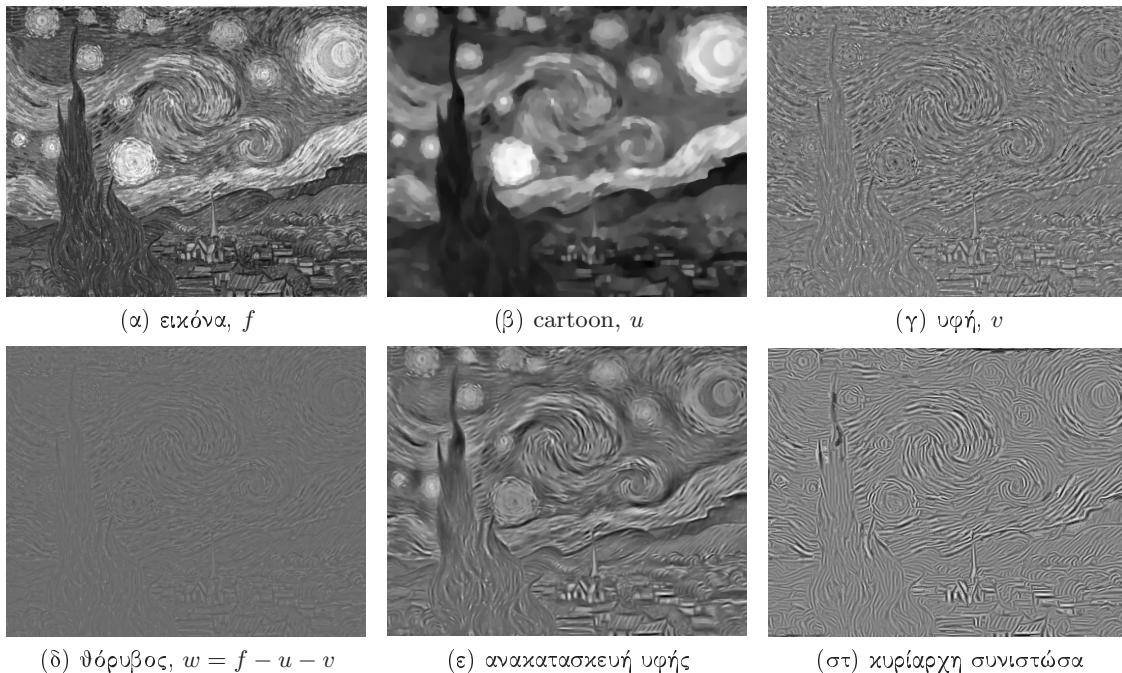
Γενικά τα σχήματα αποσύνθεσης θεωρούν μια εικόνα ως το άθροισμα μιας γεωμετρικής συνιστώσας, που αποκαλείται εύστοχα *cartoon* και μιας συνιστώσας ταλαντώσεων υψηλής συχνότητας που περιγράφει την υφή και το θόρυβο. Οι δύο συνιστώσες λόγω της διαφορετικής φύσης τους, απαιτούν ξεχωριστή επεξεργασία και παρακινούν μια ανάλυση σε δύο παράλληλα κανάλια. Η συνιστώσα της υφής μπορεί να χρησιμοποιηθεί για κατάτμηση και ταξινόμηση υφής, ανάλυση επιφανειών, εκτίμηση σχήματος και προσανατολισμού (shape/orientation from texture), ενώ η συνιστώσα δομής για ανίχνευση ακμών, κατάτμηση, ανάλυση σχήματος και αναγνώριση αντικειμένων. Επιπλέον, η μελέτη του διαχωρισμού έχει βρεί εφαρμογή σε ανερχόμενες εφαρμογές όπως ο (ψηφιακός) επιχρωματισμός [29, 55, 91, 111] και η υπολογιστική φωτογραφία [14, 75, 88] και έχει εμπνεύσει λύσεις σε κλασσικά προβλήματα με αλγορίθμους προσαρμοσμένους διαφορετικά στα μέρη υφής και δομής, όπως αποκατάσταση εικόνων [12, 108], παρεμβολή, κατάτμηση [268], ανάκτηση, ταξινόμηση [9, 298], συμπίεση [213] και αναγνώριση προσώπων [61].

Παράλληλα, οποιοδήποτε μοντέλο υφής έρχεται αντιμέτωπο με θέματα όπως α) η ερμηνεία μιας γενικής τάξης εικόνων υφής, β) ανίχνευση των περιοχών υφής σε μια εικόνα και γ) η απόκριση σε άλλα χαρακτηριστικά και περιοχές. Μοντέλα που περιγράφουν την υφή με στοιχειώδη σωματίδια (π.χ. διαμορφώσεις, τεξτόνια, αρχέγονα) μπορούν με την ίδια ευελιξία που ερμηνεύουν ένα μεγάλο εύρος προτύπων υφής να εξηγήσουν και διαφορετικές δομές εικόνων. Σε αυτές τις περιπτώσεις οι μηχανισμοί για μοντελοποίηση επιλεκτική ως προς την υφή αναζητούνται σε χαρακτηριστικά όπως η κλίμακα, η κατεύθυνση, η τυχαιότητα ή σε τοπικά μετρικά πιθανότητας ύπαρξης υφής [92, 165, 186].

### 5.1.1 Κίνητρα και πλεονεκτήματα

Οι δύο διαδικασίες, δηλαδή η μοντελοποίηση και η αποδόμηση εικόνων ή ο διαχωρισμός της υφής, θεωρούνται συσχετισμένες και μπορούν να τοποθετηθούν υπό ένα κοινό πρίσμα. Προς αυτή την κατεύθυνση, αναπτύσσεται ένα σχήμα συνέργειας ανάμεσα στις δύο, με κίνητρο τις εννοιολογικές ομοιότητες ανάμεσα στον αυστηρά μαθηματικό ορισμό της υφής σε ένα χώρο ταλαντευόμενων συναρτήσεων [214, 298] και το μοντέλο χωρικών διαμορφώσεων της υφής σε πολλαπλές ζώνες συχνοτήτων [37, 125, 198].

Στη μια κατεύθυνση συνέργειας, προτείνεται ένα νέο σχήμα αποδόμησης, όπου οι ταλαντώσεις της υφής περιγράφονται με το AM-FM μοντέλο (2.9). Σε αντίθεση με τα τρέχοντα



Σχήμα 5.1: Αποσύνθεση εικόνων: Η εικόνα στο (α) είναι ένα μείγμα από γεωμετρικές δομές και πρότυπα υφής. Το  $u + v$  μοντέλο ορίζει την υπέρθεση μας απλοποιημένης προσέγγισης της εικόνας ("cartoon") στο (β) που αποτελείται από σχήματα περιγράμματα, αντικείμενα και φόρμες και μια ταλαντευόμενη συνιστώσα (υφή) στο (γ). Παρατηρείστε τις αυστηρές γεωμετρικές φόρμες (κύκλοι, σπείρες) και τα επίπεδα πλατώ στο (β) καθώς και τις λεπτομέρειες μικρής κλίμακας και τα προσανατολισμένα πρότυπα (εξπρεσιονιστικές γραμμές) στο (γ). Τα σχήματα γίνονται αντιληπτά ως οι αφηρημένες φόρμες στην εικόνα και οι προσανατολισμένες γραμμές ως η υφή τους. "The Starry Night", Vincent van Gogh, 1889, (The Museum of Modern Art, New York).

σχήματα, όπου η υφή λαμβάνεται από τη διαφορά ανάμεσα σε μια κανονικοποιημένη προσέγγιση της αρχικής εικόνας ή αναζητείται σε δυαδικούς συναρτησιακούς χώρους, πρότερη πληροφορία του μοντέλου εισάγεται αποκλειστικά σε ένα συναρτησιακό, η ελαχιστοποίηση του οποίου δίνει ταυτόχρονα τις συνιστώσες cartoon και υφής. Παράπλευρα, το τμήμα της υφής αποδομείται περαιτέρω σε συνιστώσες επιλεκτικές ως προς τις χωρικές τους συχνότητες (προσανατολισμό και κλίμακα) και κατασκευάζεται μια πυραμίδα πολυκλιμακωτών (multiscale) και πολυκατευμνητικών (multiorientation) αναπαραστάσεων της υφής.

Στην αντίθετη κατεύθυνση, μόνο η συνιστώσα της εικόνας που κατά την αποδόμηση αποδίδεται στην υφή, διατηρείται για ανάλυση (εξαγωγή χαρακτηριστικών). Η συνιστώσα αυτή, ιδιαίτερα απελευθερωμένη από δομές μεγάλης κλίμακας, μοντελοποιείται καλύτερα από το μοντέλο στενοζωνικών υποσυνιστωσών (narrowband subcomponents), οι παράμετροι του οποίου μπορούν να εκτιμηθούν κατά τη φάση αποδόμησης. Έτσι, ο διαχωρισμός πριν την ανάλυση μπορεί να θεωρηθεί ένα στάδιο προεπεξεργασίας και ενίσχυσης της πληροφορίας υφής.

### 5.1.2 Σύνοψη συνέργειας

Τιοθετείται το μοντέλο  $f = u + v$  για εικόνες και η μεταβολική προσέγγιση ελαχιστοποίησης ενέργειας για αποδόμηση, αναζητώντας μια λύση σε ένα πρόβλημα ολικής μεταβολής

(total variation, TV) [258] με περιορισμούς. Για το σκοπό αυτό κατασκευάζεται ένα συναρτησιακό με έναν όρο πιστότητας (fidelity) και έναν όρο που πριμοδοτεί την ομοιότητα ανάμεσα στη συνιστώσα της υφής και σε ένα άθροισμα στενοζωνικών συνιστωσών της εικόνας. Η ελαχιστοποίηση του με μεταβολικές μεθόδους οδηγεί σε ένα σύστημα Μερικών Διαφορικών Εξισώσεων (ΜΔΕ), η λύση του οποίου είναι το ζευγάρι συναρτήσεων  $(u, v)$ . Το πρόβλημα μπορεί να αναχθεί ως ειδική περίπτωση μιας γενικότερης αποδόμησης σε  $K + 1$  συνιστώσες  $(u, v_1, \dots, v_K)$ , όπου το σύστημα  $K + 1$  εξισώσεων απλοποιείται, κάτω από ορισμένες προϋποθέσεις, στη  $u + v$  λύση με  $v = \sum_k v_k$ . Το γραμμικό άθροισμα συνιστωσών δίνεται από την ανακατασκευή πληροφορίας από τις αποκρίσεις φίλτρων Gabor (2.1, 2.2) σε πολλαπλές ζώνες συχνοτήτων με παραλλαγές του σχήματος να προκύπτουν από τη μορφή αυτής της ανακατασκευής.

Μια προσέγγιση είναι μέσω της επιλογής εκ των προτέρων του εύρους κλίμακας της υφής στην εικόνα και η διατήρηση των μεσαίων και υψηλότερων ζωνοπερατών συνιστωσών, όπου ανήκει η υφή κατά κύριο λόγο, σε σχέση με τις χαμηλότερες όπου κυριαρχούν οι ακμές. Κάτι τέτοιο απαιτεί κάποιο μηχανισμό επιλογής κλίμακας [42, 108, 163] ή κάποια ευρετική μέθοδο, προσαρμοσμένη για κάθε εικόνα. Η ανάγκη για αποκλειστική γνώση κλίμακας παρακάμπτεται με την στάθμιση των όλων των ζωνοπερατών συνιστωσών με κάποιο μέτρο της συνεισφοράς τους στο συνολικό άθροισμα υφής. Το πλάτος της περιβάλλουσας αντιστοιχεί στις μεταβολές τοπικά της αντίθεσης και επομένως σε μέσες μεταβολές υφής ή προτύπων στις διαφορετικές ζώνες συχνοτήτων. Τέλος, σε μια άλλη παραλλαγή, επιλέγονται σε κάθε θέση οι συνιστώσες που κυριαρχούν στο άθροισμα με βάση κάποιο ενεργειακό κριτήριο στο χώρο απόκρισης των φίλτρων. Επιλέγοντας και διατάσσοντας ιεραρχικά τις συνιστώσες χρησιμοποιώντας ανάλυση κυρίαρχων (ιεραρχικών) συνιστωσών (DCA, HCA) [124, 163], το άθροισμα καθολικά στενοζωνικών συνιστωσών μετατρέπεται σε άθροισμα τοπικά στενοζωνικών φθίνουσας με το δείκτη άθροισης σημαντικότητας στην αναπαράσταση.

Οι στενοζωνικές συνιστώσες είναι μιγαδικές συναρτήσεις, που προέρχονται από τις αποκρίσεις μιγαδικών καναλιών με πραγματικό και φανταστικό μέρος σε ορθογώνια φάση [36, 72]. Η μιγαδική πληροφορία λαμβάνεται υπόψιν γενικεύοντας την αποδόμηση σε μιγαδικές συναρτήσεις - εικόνες. Οι μιγαδικές εικόνες αντιμετωπίζονται ως διανυσματικές συναρτήσεις στον πραγματικό-φανταστικό χώρο και η αποδόμηση στα δύο κανάλια είναι συζευγμένη με μηγραμμικό τρόπο όπως στις εξισώσεις διάχυσης διανυσματικών συναρτήσεων [285, 286]. Για το λόγο αυτό μια κατάλληλη επέκταση της νόρμας ολικής μεταβολής, η μιγαδική TV ορίζεται σύμφωνα με το πολυδιάστατο [33] και το διανυσματικό [54] πλαίσιο ολικής μεταβολής.

Το κομμάτι της υφής που έχει διαχωριστεί από την εικόνα, υπονοεί σε ένα δεύτερο επίπεδο μια αποδόμηση στις υποσυνιστώσες που ορίζονται από το πλέγμα των φίλτρων. Κάθε τέτοια υποσυνιστώσα υφής μοντελοποιείται περαιτέρω από διαμορφώσεις ομαλού τοπικά πλάτους και συχνότητας. Η εφαρμογή το μοντέλου σε μια δοσμένη εικόνα γίνεται με την εκτίμηση των παραμέτρων από το κομμάτι της υφής μόνο. Έτσι, παραπλανητικά στοιχεία δομής όπως οι ακμές απορρίπτονται, οδηγώντας σε πιο συστηματική ανάλυση και εύρωστα χαρακτηριστικά υφής.

## 5.2 Αποσύνθεση εικόνων

Η αποσύνθεση ή αποδόμηση εικόνων (image decomposition) αποσκοπεί στο διαχωρισμό μιας εικόνας σε συνιστώσες διαφορετικής (μαθηματικής, εννοιολογικής, σημασιολογικής) φύσης. Θεμελιώδως, το μοντέλο γραμμικής υπέρθεσης προβλέπει ότι κάθε εικόνα  $f$  δίνεται από το άθροισμα δύο ανεξάρτητων συνιστωσών

$$f = u + v, \quad f, u, v : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (5.1)$$

όπου το  $u$  είναι μια τυμηματικά ομαλή συνάρτηση, με επίπεδα φωτεινότητας και απότομες ασυνέχειες και το  $v$  μια ταλαντευόμενη συνάρτηση γρήγορης μεταβολής. Στο παραπάνω μοντέλο, γνωστό ως μοντέλο  $u + v$ , οι δύο συνιστώσες είναι υπολογιστικά και αντιληπτά συμπληρωματικές, όπως για παράδειγμα στα Σχτ. 5.1, 5.2.

Η συνιστώσα  $u$  αποκαλείται εύστοχα συνιστώσα καρτούν (cartoon component) και είναι μια προσέγγιση της αρχικής εικόνας [221] που περιέχει γεωμετρική πληροφορία όπως ακμές, περιγράμματα αντικειμένων, χαρακτηριστικά μεγάλης κλίμακας, σχήματα και φαινόμενα φωτισμού. Στην τέχνη [210]

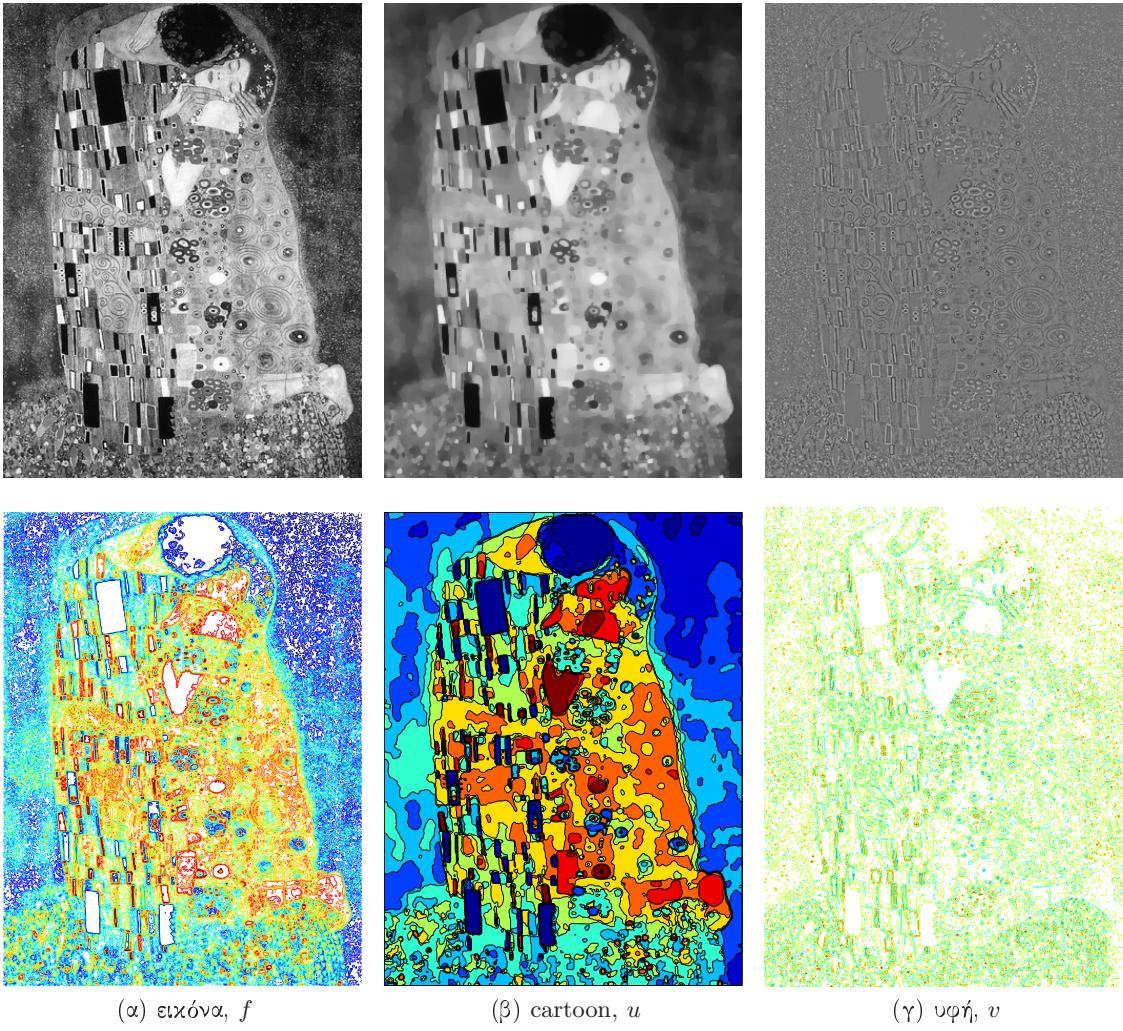
«το cartoon είναι ένας τρόπος ενίσχυσης μέσω απλοποίησης (amplification through simplification). Μετατρέποντας σε cartoon μια εικόνα, δεν αφαιρούμε λεπτομέρεια αλλά επικεντρωνόμαστε σε συγκεκριμένες λεπτομέρειες. Με το να ανάγει μια εικόνα στο ουσιαστικό της νόημα, ο καλλιτέχνης μπορεί να ενισχύσει αυτό το νόημα με τρόπο που η ρεαλιστική τέχνη δεν μπορεί.»

Η συνιστώσα  $v$  ή «ταλαντώσεις υφής», μοντελοποιεί τα χαρακτηριστικά μικρής κλίμακας, προσανατολισμένα πρότυπα, περιοδικότητες και τυχαίο θόρυβο. Η έννοια της κλίμακας καθορίζει την ποσότητα λεπτομέρειας και τη συσχέτιση ανάμεσα στις δύο συνιστώσες και μια πυραμίδα από ζευγάρια αποδόμησης μπορεί να κατασκευαστεί με κατάλληλο ορισμό της, π.χ. από το επίπεδο ομοιότητας του cartoon με την αρχική εικόνα [108, 213, 268, 279]. Σε μια παραλλαγή του μοντέλου, η υφή  $v$  διαχωρίζεται από τη συνιστώσα θορύβου  $w$  [8, 274] σε ένα μοντέλο τριών συνιστωσών  $f = u + v + w$  (στα  $u + v$  μοντέλα ο θόρυβος δίνεται από το παραμένον  $f - u - v$ ).

Οι κύριες προσεγγίσεις για αποδόμηση εικόνων, δηλαδή οι μέθοδοι με ΜΔΕ και οι προβολές σε βάσεις, ενοποιούνται και αντιμετωπίζονται από κοινού κάτω από το πλαίσιο μεθόδων ελαχιστοποίησης ενέργειας (energy minimization).

### 5.2.1 Ολική μεταβολή και ΜΔΕ

Το μεταβολικό πλαίσιο (variational paradigm) [7, 8, 11, 12, 227, 298, 299], γενικεύει τα μοντέλα Mumford-Shah (MS) [221] και Rudin-Osher-Fatemi (ROF) [257, 258] για απλοποίηση και αποκατάσταση εικόνων. Πολλές από τις μη-γραμμικές τεχνικές ομαλοποίησης με διατήρηση ακμών, όπως η ανισοτροπική διάχυση [239] δίνουν προσεγγιστικά μια συνιστώσα  $u$ .



Σχήμα 5.2: Ο πίνακας (α) είναι μια μείζη (β) γραμμετρικών δομών cartoon και (γ) προτύπων υφής.  
1η σειρά: γκρίζες εικόνες, 2η σειρά: επιπεδοσύνολα ("The Kiss", Gustav Klimt, 1908).

Μια σύνοψη τέτοιων τεχνικών μπορεί να βρεθεί στο [218] και αποδεικνύεται ότι είναι ειδικές περιπτώσεις ή στενά συνδεδεμένες με το συναρτησιακό MS [221]

$$E_{\text{MS}}(u, C) = \int_{\Omega \setminus C} \left( \|\nabla u\|^2 + \lambda(u - f)^2 \right) dx dy + \mu \text{Len}(C), \quad (5.2)$$

η ελαχιστοποίηση του οποίου δίνει μια τμηματικά ομαλή  $u \in \Omega$  και ένα ελάχιστο μήκος ακμών  $C$ .

Η μέθοδος ελαχιστοποίησης της ολικής μεταβολής (total variation (TV) minimization) [257, 258] προτάθηκε για αποκατάσταση και αποθορυβοποίηση εικόνων. Το  $u$  είναι μια συνάρτηση σε ένα χώρο  $BV(\Omega)$  φραγμένης μεταβολής (Bounded Variation)  $u \in BV(\Omega)$  που επιτρέπει ασυνέχειες κατά μήκος καμπυλών και επομένως ακμές και περιγράμματα [5]. Το παραμένον (residual)  $v = f - u$  που σε αυτή την περίπτωση αντιπροσωπεύει το θόρυβο περιορίζεται να είναι μικρό ως προς την Ευκλειδία νόρμα  $L^2(\Omega)$ . Τελικά, το cartoon υπολογίζεται με ελαχιστοποίηση του συναρτησιακού

$$E_{\text{ROF}}(u) = \int_{\Omega} \|\nabla u\| dx dy + \lambda \int_{\Omega} (f - u)^2 dx dy, \quad (5.3)$$

*Κεφάλαιο 5. Συνέργεια διαχωρισμού και μοντελοποίησης: Αποσύνθεση εικόνων*

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

$$\|u\|_{TV} = \int_{\Omega} \|\nabla u\| dx dy \quad (5.4)$$

είναι η  $BV$  ημι-νόρμα (semi-norm) [5] ή η ολική μεταβολή του  $u$ . Η νόρμα στον  $BV$  χώρο δίνεται ισοδύναμα από τη σχέση

$$\|u\|_{BV} = \|u\|_{L^1} + \|u\|_{TV}, \quad (5.5)$$

ενώ η συνάρτηση  $u$  ανήκει στον  $BV$  αν  $\|u\|_{BV} < \infty$ . Η ελαχιστοποίηση του (5.3) αντιστοιχεί στην επίλυση ενός προβλήματος βελτιστοποίησης με περιορισμούς (constrained optimization) που ανήκει στην κατηγορία κανονικοποίησης αντίστροφων προβλημάτων (Tikhonov regularization [1, 284]).

Στο γενικό τους πλαίσιο, τα περισσότερα σχήματα αποσυνθέτουν μια εικόνα χρησιμοποιώντας περιγραφές των διαφορετικών συνιστωσών σε ξεχωριστούς συναρτησιακούς χώρους, εφόδιασμένους με εξειδικευμένες νόρμες - μετρικά [167]. Το πρόβλημα τίθεται σαν ένα πρόβλημα ελαχιστοποίησης ενός χυρτού (convex) συναρτησιακού [39] των άγνωστων συνιστωσών της γενικής μορφής

$$\inf_{(u,v) \in (U \times V)(\Omega)} \{E(u,v) = J(u) + \lambda F(u,v) + \mu L(v)\} \quad (5.6)$$

δεδομένου του  $f = u+v$ , όπου  $U, V$  οι συναρτησιακοί χώροι για τις δύο συνιστώσες και  $\lambda, \mu \leq 0$  ρυθμιστικές σταθερές. Ο πρώτος όρος  $J(u)$  είναι συνήθως ένας όρος κανονικοποίησης μέσω TV ελαχιστοποίησης. Ένας δεύτερος όρος πιστότητας (fidelity)  $F(u,v) = \|f - u - v\|_X^2$  καθορίζει την ακρίβεια της προσέγγισης της  $f$  μέσω  $u+v$  και περιορίζει το παραμένον, και  $L(v)$  είναι μια νόρμα που μοντελοποιεί τις μεταβολές της υφής  $v$  σε κάποιο χώρο (Banach [214], Sobolev [8, 298] ή γενικευμένο Hilbert [10, 11, 227]). Αυτοί οι διαφορετικοί χώροι για την υφή πριμοδοτούν με μικρά μετρικά το  $v$ .

Στο ROF (5.3), οι όροι πιστότητας και ο περιορισμός στην υφή συμπίπτουν καθώς  $\|f - u\|_{L^2} = \|v\|_{L^2}$ , και το  $v$  είναι η διαφορά της εικόνας  $f$  από το  $u$ , δηλαδή  $E_{ROF}(u) = \|u\|_{TV} + \|f - u\|_{L^2}$ . Ο αλγόριθμος ROF χρησιμοποιήθηκε για να οριστούν επίσημα οι εικόνες που μπορούν να αποτελούν cartoon [115].

Η συνιστώσα της υφής ορίστηκε από τον Y. Meyer [214] σε ένα Banach χώρο  $G$  ταλαντευόμενων συναρτήσεων  $v$  έτσι ώστε

$$v = \text{div}(\vec{g}) = \partial_x g_1 + \partial_y g_2, \quad (5.7)$$

δηλαδή ως η απόκλιση ενός διανύσματος  $\vec{g} = (g_1, g_2)$  με  $g_1, g_2 \in L^\infty(\mathbb{R}^2)$ . Ο χώρος  $G$  είναι εφοδιασμένος με τη νόρμα  $\|\cdot\|_G$ , που ορίζεται ως η ελάχιστη (infimum)  $L^\infty$  νόρμα του  $|\vec{g}|$  όλων των δυνατών αποδομήσεων (5.7):

$$\|v\|_G = \inf \{ \|\vec{g}\|_{L^\infty} : v = \text{div}(\vec{g}) \}, \quad |\vec{g}(x,y)| = \sqrt{g_1^2(x,y) + g_2^2(x,y)}. \quad (5.8)$$

Κατά μια έννοια ο χώρος  $G$  είναι ο δυϊκός του  $BV$  χώρου με  $\dot{BV} \subset L^2 \subset G$  (όπου  $\dot{BV}$  ο ομογενής  $BV^1$ ) [214] και η νόρμα του η δυϊκή της  $\|\cdot\|_{TV}$  (5.4). Οι συναρτήσεις  $v \in G$  είναι γενικές, περιοδικές συναρτήσεις, μηδενικής μέσης τιμής που μπορούν να παρουσιάζουν ταλαντώσεις μεγάλου εύρους (πλάτους), αλλά μικρές τιμές του  $\|\cdot\|_G$ , γεγονός που διαπιστώνεται αριθμητικά [8] και αποδεικνύεται ψεωρητικά [115]. Αν η συνιστώσα  $v$  αντιπροσωπεύει υφή ή/και ύδρυβο σε μια εικόνα, τότε μια αποδόμηση/αποκατάσταση μπορεί να επιτευχθεί από ένα σχήμα

$$\frac{E(u, v)}{(u, v) \in BV \times G(\Omega)} = \int_{\Omega} \|\nabla u\| dx dy + \lambda \|v\|_G. \quad (5.9)$$

Για την υπολογιστική αξιοποίηση του, το μετρικό  $\|v\|_G$  προσεγγίστηκε πρακτικά από τους Vese και Osher [227, 298, 299] με χρήση Sobolev νορμών  $L^p$

$$\|\vec{g}\|_{L^p} = \left( \int_{\Omega} |\vec{g}|^p dx dy \right)^{1/p}, \quad 1 \leq p \leq \infty. \quad (5.10)$$

Για  $p \rightarrow \infty$  η  $L^p$  νόρμα προσεγγίζει την  $L^\infty$  και επομένως το μετρικό του χώρου  $G$  5.8. Η προσέγγιση χρησιμοποιήθηκε σε ένα σχήμα αποδόμησης σε τρεις συνιστώσες με ελαχιστοποίηση του ενεργειακού

$$\frac{E(u, \vec{g})}{(u, \vec{g}) \in BV \times L^p} = \int_{\Omega} \|\nabla u\| dx dy + \lambda \int_{\Omega} |f - (u + \text{div}(\vec{g}))|^2 dx dy + \mu \|\vec{g}\|_{L^p} \quad (5.11)$$

με το παραμένον  $f - u - v$  να μοντελοποιεί το ύδρυβο  $w$ . Στο όριο που  $\lambda, p \rightarrow \infty$  το παραπάνω σχήμα προσεγγίζει το διαχωρισμό των ταλαντώσεων της υφής (5.9). Επιπλέον, το διάνυσμα  $\vec{g}$  που περιγράφει τη μεταβολή της υφής  $v$  στην οριζόντια και κάθετη διεύθυνση της εικόνας αποτελεί κάποιο τρόπο τη ροή υφής (texture flow) που μπορεί να χρησιμεύσει για περαιτέρω ανάλυση [55, 299]. Μια ειδική περίπτωση του (5.11) προκύπτει για  $p = 2$  και  $\lambda = 0$ , ισοδύναμη με την ελαχιστοποίηση της νόρμας  $\|v\|_{H^{-1}}^2 = \int_{\Omega} |\nabla(\Delta^{-1})v|^2$  [227, 228]. Σ' αυτή την περίπτωση αποδεικνύεται ότι το παραμένον  $v = f - u$ , που εκφράζει υφή ή ύδρυβο ικανοποιεί την απαίτηση για μηδενική μέση τιμή.

Γύρω από την ίδια λογική, μια άλλη προσέγγιση στην επίλυση του (5.9) είναι με τη χρήση του ακριβούς ορισμού της νόρμας  $\|\cdot\|_G$  και ένα επιπλέον όρο πιστότητας [7, 8]:

$$\frac{E(u, v)}{(u, v) \in BV \times G} = \left\{ J(u) + \lambda \|f - u - v\|_{L^2}^2 : \|v\|_G \leq \mu \right\}. \quad (5.12)$$

Η δημιουργία του τελικού κυρτού συναρτησιακού περιλαμβάνει τη συνάρτηση δείκτη (indicator function) του συνόλου  $\|v\|_G \leq \mu$  ως περιορισμό, όπου η σταθερά  $\mu$  καθορίζει το βαθμό ταλαντώσεων του  $v$ , δηλαδή μικρές τιμές επιτρέπουν ταλαντώσεις μεγάλου εύρους στην υφή. Η αριθμητική επίλυση γίνεται με τη μέθοδο επαναλήψεων σταθερού σημείου και έναν αλγόριθμο ελαχιστοποίησης της ολικής μεταβολής με μη-γραμμική προβολή [52].

Ο διαχωρισμός σε τρεις συνιστώσες  $u + v + w$  με το ίδιο μοντέλο εξετάστηκε στο [8] κάτω από ένα κοινό πρίσμα ελαχιστοποίησης κυρτών συναρτησιακών της γενικής μορφής (5.6),

---

<sup>1</sup> $\dot{BV} = BV / \{u \in BV / \nabla u = 0\}$ .

## Κεφάλαιο 5. Συνέργεια διαχωρισμού και μοντέλοποίησης: Αποσύνθεση εικόνων

επιλύοντας ισοδύναμα δυϊκά προβλήματα που περιλαμβάνουν δυϊκές νόρμες (dual norms). Η συνιστώσα του θορύβου ω λαμβάνεται στο δυικό ενός χώρου Besov με συρρίκνωση κυματιδίων (wavelet shrinkage) [82,83], δηλαδή κατωφλιοποίηση στους συντελεστές wavelet της εικόνας. Με αυτό τον τρόπο ένα  $u + w$  μοντέλο εφαρμόστηκε για αποθορυβοποίηση εικόνων υφής (texture denoising).

Τέλος, μια ομαδοποίηση των μοντέλων αποδόμησης επιχειρήθηκε στο [11] σε σχέση με διαφορετικές κατηγορίες υφής. Για γενικά ταλαντεύομενα πρότυπα το  $(TV - L^2, G)$  (5.3, 5.9), για δομημένες το  $(TV - L^1)$ , μια παραλλαγή του ROF με  $L^1$  αντί για  $L^2$  περιορισμό/πιστότητα [223], και για ομαλές - περιοδικές ή κατευθυντικές υφές το  $(TV - Gabor)$ . Η τελευταία περίπτωση παρουσιάζει ιδιαίτερο ενδιαφέρον και είναι μια ειδική περίπτωση TV-Hilbert κανονικοποίησης [10], όπου ο όρος πιστότητας  $\|f - u\|_X$  ορίζεται σε κάποιο Hilbert χώρο  $\mathcal{H}$ . Η ελαχιστοποίηση τότε του ενεργειακού

$$\underset{(u,v) \in BV \times \mathcal{H}}{E(u,v)} = \int_{\Omega} \|\nabla u\| dx dy + \lambda \|v\|_{\mathcal{H}}^2 \quad (5.13)$$

ανάγεται στο ROF σχήμα για  $\mathcal{H} = L^2$  και στο [227] για  $\mathcal{H} = H^{-1}$ . Η γενική νόρμα  $\|f - u\|_{\mathcal{H}}$ , υπολογισμένη ως η προβολή της συνιστώσας  $v = f - u$  σε συγκεκριμένα Gabor κυματίδια  $g_k$ , δηλαδή  $\|f - u\|_{\mathcal{H}} = \langle f - u, g_k * (f - u) \rangle$  επιτρέπει ένα σχήμα αποδόμησης επιλεκτικό ως προς τη συχνότητα και την κατεύθυνση της υφής στην εικόνα.

Σε όλα τα προηγούμενα σχήματα προκύπτει το θέμα της βέλτιστης επιλογής των παραμέτρων αποδόμησης  $\lambda, \mu$  αλλά και η συγκριτική και οριακή συμπεριφορά των διαφόρων μεθόδων σε σχέση με το εύρος αυτών των παραμέτρων. Το  $\lambda$  σχετίζεται με το βαθμό πιστότητας της προσέγγισης  $f = u + v$ , με την κλίμακα των λεπτομερειών που παραμένουν στο cartoon και ενίστε με το θόρυβο στην εικόνα. Το  $\mu$  σχετίζεται κατά κάποιο τρόπο με το εύρος των ταλαντώσεων υφής και την ποσότητα πληροφορίας που συμπεριλαμβάνεται στο  $v$  [8,214]. Η βέλτιστη ρύθμιση των παραμέτρων έχει ερευνηθεί μέσω διαφόρων μεθόδων ανάλογα με την εφαρμογή, όπως ο σηματοθορυβικός λόγος (SNR) για θέματα αποκατάστασης [9,10,108,258], το γράφημα συσχέτισης των  $u, v$  για διαχωρισμό συνιστωσών [11], ποσοτικά κριτήρια πιστότητας [299] και πρότερες απαίτησεις για την  $L_2$  ενέργεια του παραμένοντος [7].

Γενικά και με αναφορά στο γενικό σχήμα (5.6), μεγάλες τιμές του  $\lambda$  καταλήγουν σε μικρές νόρμες παραμενόντων και λεπτομέρειες μικρής κλίμακας στο  $u$ , π.χ. κλίμακες μεγαλύτερες από  $1/\lambda$ . Το  $\mu$  λειτουργεί συμπληρωματικά ως προς το  $\lambda$  αλλά μικρότερες τιμές του εξασφαλίζουν μεγαλύτερες ταλαντώσεις στο  $v$ . Μια λογική απαίτηση θα ήταν να επιλεχθούν  $\lambda \gg \mu$ . Σημειωτέον ότι τα περισσότερα από τα σχήματα που έχουν αναπτυχθεί ανάγονται στο μοντέλο ROF στο όριο  $\lambda \rightarrow 0$ .

### 5.2.2 Κυματίδια και βάσεις

Η αποδόμηση εικόνων με βάση «κυματίδια» και «λεξικά» [69,91,183,213,214,274] (όπου ο όρος λεξικά αναφέρεται σε βάσεις από στοιχειώδεις δομές εικόνας που μαθαίνονται ή είναι γνωστές εκ των προτέρων) αφορά στην αναπαράσταση των διαφορετικών συνιστωσών της ει-

κόνας με εξειδικευμένες δομές - βάσεις εικόνων (wavelets, packets, textons). Ο λεξικοπαγής (wavelet-dictionary based) διαχωρισμός των διαφορετικών πηγών πληροφορίας συστηματοποιείται ως ένα πρόβλημα βελτιστοποίησης και εύρεσης της πιο αποδοτικής αναπαράστασης. Η αποδοτικότητα μπορεί να σχετίζεται με την αραιότητα (sparseness) [213, 274], την ανεξαρτησία (independence) ή την «ικανότητα σκιαγράφησης» (“sketchability”) [112]. Η ελκυστικότητα της περιγραφικής ικανότητας τέτοιων μεθόδων που ψάχνουν για τις κατάλληλες βάσεις σταθμίζεται από την ανάγκη να οριστούν αποκλειστικά και εκ των προτέρων τα κατάλληλα υποψήφια λεξικά καθώς και τις αυξημένες απαιτήσεις σε υπολογιστικό φόρτο.

Στο απλούστερο σενάριο, η μέθοδος αποθρυβοποίησης με ήπια κατωφλιοποίηση (soft-thresholding), αφαιρεί λεπτομέρειες από την εικόνα κάτω από μια συγκεκριμένη κλίμακα και δημιουργεί cartoon προσεγγίσεις [82]. Η αναλογία με τις μεταβολικές μεθόδους προκύπτει από το γεγονός ότι η συρρίκνωση κυματιδίων [83] είναι ισοδύναμη με την TV κανονικοποίηση και το ROF σχήμα [274, 275]. Η αποδόμηση με κυματίδια [190–192] και η ελαχιστοποίηση της ολικής μεταβολής αντιμετωπίζονται υπό ένα κοινό μεταβολικό πλαίσιο στο [183] για αποκατάσταση εικόνων, όπου το παραμένον  $f - u$  περιορίζεται σε ένα χώρο (ήπια-) κατωφλιοποιημένων κυματιδιακών συντελεστών. Οι συντελεστές είναι το αποτέλεσμα της αναπαράστασης σε ένα λεξικό που αποτελείται από wavelets, wavelet bases και wavelet packets. Στο [213] μια πολυεπίπεδη (multilayered) αποδόμηση προτάθηκε για συμπίεση, με περιγραφές των περιοδικών προτύπων υφής χρησιμοποιώντας τοπικές βάσεις συνημιτόνων και ορθοκανονικά κυματίδια για τις γεωμετρικές ακμές των cartoon.

Κυματιδιακή συρρίκνωση για  $BV$  ελαχιστοποίηση φραγμένης μεταβολής είναι και η προσέγγιση [69] όπου ο διαχωρισμός του cartoon από τις ταλαντώσεις λαμβάνεται επιλύοντας το πρόβλημα βελτιστοποίησης

$$\inf_{u,v} \left\{ \lambda \|f - u - v\|_{L^2}^2 + \gamma \|v\|_{H^{-1}}^2 + 2\alpha \|u\|_{B_1^1(L_1)} \right\}. \quad (5.14)$$

Η κανονικοποίηση γίνεται σε ένα εναλλακτικό, μικρότερο Besov χώρο  $B_1^1 \subset BV$  και το  $v$  περιορίζεται να έχει μικρή νόρμα στον  $H^{-1}$ , δηλαδή το σχήμα (5.11) για  $p = 2$  [227].

Η αραιότητα της αναπαράστασης μέσω προβολής σε διαφορετικές βάσεις έχει επίσης χρησιμοποιηθεί για διαχωρισμό περιεχομένου [274]. Η αναπαράσταση, η αποδόμηση και η συμπίεση εικόνας αντιμετωπίστηκαν από κοινού μέσω του αλγόριθμου Basis Pursuit [60, 113] με επιπλέον TV κανονικοποίηση

$$\inf_{u,v} \left\{ \|\mathbf{T}_u u\|_{L^1} + \|\mathbf{T}_v v\|_{L^1} + \lambda \|f - u - v\|_{L^2}^2 + \gamma J(u) \right\}, \quad (5.15)$$

όπου τα  $\mathbf{T}_u, \mathbf{T}_v$  είναι οι μετασχηματισμοί βάσης για το cartoon και την υφή αντίστοιχα, επιλεγμένοι από κατάλληλα λεξικά. Για το γεωμετρικό, ομαλό περιεχόμενο οι μετασχηματισμοί είναι τέτοιοι ώστε να αποκαλύπτουν δομή (wavelets, ridgelets, curvelets [273]) και για την υφή τοπικές συναρτήσεις Gabor και συνημιτόνων. Ο ύσρυβος λαμβάνεται ως το παραμένον  $f - u - v$ . Η μέθοδος, που ονομάστηκε ανάλυση μορφολογικών συνιστώσων (morphological components analysis), χρησιμοποιήθηκε για ανάκτηση χαμένης πληροφορίας και επιχρωματισμό [91]. Ο όρος πιστότητας στο (5.15) τροποποιείται ώστε να λαμβάνεται υπόψιν μόνο στα

σημεία της εικόνας στα οποία όπου υπάρχει πληροφορία, δηλαδή  $\|\mathbf{M}(f - u - v)\|_{L^2}^2$ , όπου ο πίνακας - μάσκα  $\mathbf{M}$  κωδικοποιεί την απουσία ή την παρουσία πληροφορίας. Με αυτό τον τρόπο τα  $u$  και  $v$  στα «χαμένα» pixel ανακατασκευάζεται από τα στοιχεία των βάσεων για την κάθε συνιστώσα. Ενοποιείται έτσι η αποδόμηση, η αποθορυβοποίηση και ο επιχρωματισμός της εικόνας  $f = u + v$ .

Τέλος σε ένα μαθηματικό μοντέλο για το πρωταρχικό σκίτσο [112,113] το περιεχόμενο μιας εικόνας αναλύεται γενικά χρησιμοποιώντας δύο κατηγορίες περιεχομένου, τη συνιστώσα που μπορεί να σκιαγραφηθεί (παρόμοια με το cartoon) και τη συμπληρωματική της. Η ικανότητα σκιαγράφησης ορίζεται με βάση το λόγο της πιθανοφάνειας ανάμεσα σε ένα γραμμικό μοντέλο αραιής αναπαράστασης, όπως προηγούμενα, και ένα στοχαστικό MRF μοντέλο. Ακολουθιακά, το περιεχόμενο σκιαγράφησης εξάγεται αναζητώντας τη βέλτιστη αναπαράσταση (matching pursuit) σε ένα λεξικό από τοπικά primitives, που έχει προκύψει μετά από εκπαίδευση. Η υφή εξάγεται μεγιστοποιώντας την πιθανοφάνεια ή την στατιστική εντροπία της περιγραφής με όρους τοπικών ιστογραμμάτων των αποκρίσεων μιας οικογένειας φίλτρων (filter pursuit, FRAME) [324,325].

### 5.3 $u + Kv$ : πολλαπλές συνιστώσες υφής

Θεωρείστε μια τυχαία συνάρτηση-εικόνα  $f : \Omega \rightarrow \mathbb{R}$ , με  $\Omega$  ένα ανοιχτό και φραγμένο υποσύνολο του  $\mathbb{R}^2$  που μπορεί να γραφτεί ως γραμμικός συνδυασμός δύο συνιστώσων της μορφής  $f = u + v$ . Αόριστα ορισμένα το  $u \in \Omega$  είναι μια τμηματικά ομαλή, αφηρημένη προσέγγιση της εικόνας και το  $v \in \Omega$  μια ταλαντεύομενη συνάρτηση μικρού πλάτους. Η cartoon συνιστώσα  $u$  περιέχει πληροφορία για τη γεωμετρία των περιοχών της εικόνας, τις ασυνέχειες και την οργάνωση των μεγάλων δομών της εικόνας, ενώ η συνιστώσα  $v$  τα πρότυπα της υφής.

Για την ανάκτηση των δομικών συνιστώσων της  $f$  προτείνεται η επίλυση του ακόλουθου προβλήματος ελαχιστοποίησης χυρτών συναρτησιακών [39]

$$\begin{aligned} & \inf_{(u,v) \in BV \times \mathcal{H}} \{E(u,v)\}, \\ E(u,v) &= \|u\|_{TV} + \lambda \|f - u - v\|_{L^2}^2 + \mu \|v - \sum_k t_k\|_{L^2}^2 \\ &= \int_{\Omega} \|\nabla u\| dx dy + \lambda \int_{\Omega} |f - u - v|^2 dx dy + \mu \int_{\Omega} |v - \sum_k t_k|^2 dx dy, \end{aligned} \quad (5.16)$$

όπου το  $u$  ανήκει στο χώρο φραγμένης μεταβολής  $BV(\Omega)$  και το  $v$  σε ένα γενικευμένο Hilbert χώρο  $\mathcal{H}(\Omega)$ . Το τελικό ενεργειακό υπό ελαχιστοποίηση είναι της μορφής (5.6), δηλαδή κανονικοποίησης με περιορισμούς. Ο πρώτος όρος είναι η νόρμα ολικής μεταβολής (5.4) του  $u$  και ο δεύτερος η πιστότητα της ανακατασκευής  $u + v$  στην αρχική εικόνα  $f \in L^2(\Omega)$ . Ο τρίτος όρος είναι ένας επιπλέον περιορισμός για τη συνιστώσα της υφής. Εκφράζει την συγγένεια (proximity) του  $v$  με μια ανακατασκευή της υφής της εικόνας από ένα σύνολο στενοζωνικών συνιστώσων  $t_k$  μιας συνάρτησης  $t \in L^2(\Omega)$ , δεδομένου ενός συνόλου φίλτρων διαφορετικής

συχνότητας και προσανατολισμού  $\{g_k, k \in \{1 \cdots K\}\} \in L^2(\Omega)$ . Η συνάρτηση  $t$ , για την ώρα, ας θεωρηθεί το αποτέλεσμα μιας γραμμικής  $t = \mathbf{T}f$  ή μη-γραμμικής  $f \mapsto t$  απεικόνισης  $\Omega \mapsto \Omega$ . Έτσι, ο έξιτρα όρος αναγκάζει το  $v$  να είναι κοντά, κατά την  $L^2$  έννοια, σε μια πολυσυχνοτική αποδόμηση του  $t$ .

Η αποδόμηση του  $t$  στις  $t_k$  συνιστώσες μπορεί να αντιμετωπισθεί είτε ως προβολή του σε ένα λεξικό από  $K$  τοπικές βάσεις ή ως πολυκαναλικό φίλτροισμα από  $K$  αποχρίσεις φίλτρων. Τα χαρακτηριστικά των συνιστωσών (αριθμός, χωρική τοποθέτηση, ρύθμιση σε συχνότητα και προσανατολισμό, εύρος) καθορίζονται από τις παραμέτρους της οικογένειας των συναρτήσεων  $g_k$ . Αν αυτές είναι Gabor συναρτήσεις (2.1) διατεταγμένες ως συστοιχία φίλτρων και  $t = f$ , δηλαδή  $\mathbf{T} = \mathbf{I}$  τότε τα  $t_k$  είναι οι στενοζωνικές συνιστώσες του  $f$ . Κάθε συνιστώσα, και επομένως κάθε προβολή σε μια βάση, είναι το εσωτερικό γινόμενο

$$t_k = \langle t(\xi, \eta), g_k(\xi - x, \eta - y) \rangle = \int_{\Omega} f(\xi, \eta) g_k(x - \xi, y - \eta) d\xi d\eta = (t * g_k)(x, y) \quad (5.17)$$

ή ισοδύναμα η συνέλιξη του  $t$  με την κρουστική απόκριση του  $g_k$ .

Το συναρτησιακό  $E(u, v) = \int_{\Omega} \Phi(u, v, u_x, u_y) dx dy$  ελαχιστοποιείται από το σχετικό ζεύγος των Euler-Lagrange εξισώσεων [6, 98]. Εφαρμόζοντας φυσικές (Neumann) συνοριακές συνθήκες, η gradient-descent ΜΔΕ-λύση είναι

$$\frac{\partial u}{\partial t} + u = (f - v) + \frac{1}{2\lambda} \operatorname{div} \left( \frac{\nabla u}{\|\nabla u\|} \right), \quad (5.18)$$

$$\frac{\partial v}{\partial t} + v = \frac{\lambda}{\mu + \lambda} (f - u) + \frac{\mu}{\mu + \lambda} \sum_{k=1}^K t_k, \quad (5.19)$$

$$(u, v)(x, y, 0) = (f, 0), \quad \frac{\partial u}{\partial \vec{N}} = 0, \quad u \in \partial \Omega,$$

όπου  $\vec{N}$  το εξωτερικό μοναδιαίο κάθετο διάνυσμα στο σύνορο  $\partial \Omega$  και  $\kappa(u) = \operatorname{div} (\nabla u / \|\nabla u\|)$  η επίπεδη καμπυλότητα (level curvature) ή η απόκριση μοναδιαίας κλίσης της  $u = u(x, y)$  που δίνεται από την αναλυτική διαφορική σχέση

$$\kappa(u) = \frac{u_{xx}u_y^2 + u_{yy}u_x^2 - 2u_{xy}u_xu_y}{(u_x^2 + u_y^2)^{3/2}}. \quad (5.20)$$

Η εξίσωση (5.18) είναι η λύση που ελαχιστοποιεί το ROF μοντέλο (5.3). Απαλείφοντας το  $v$  από την (5.18), οι λύσεις μόνιμης κατάστασης (steady-state) δίνονται σε συμπαγή μορφή από τις σχέσεις

$$u = f - \sum_{k=1}^K t_k + \frac{\mu + \lambda}{2\mu\lambda} \kappa(u), \quad (5.21)$$

$$v = \frac{1}{\mu + \lambda} \left( \lambda(f - u) + \mu \sum_{k=1}^K t_k \right). \quad (5.22)$$

Από αυτό το ζεύγος εξισώσεων μπορούν να γίνουν οι εξής παρατηρήσεις: Η λύση δίνεται από μια μόνο ΜΔΕ, την εξίσωση (5.21) με τη δεύτερη εξίσωση να δίνει τη συνιστώσα της υφής ως ένα ενισχυμένο και σταθμισμένο παραμένον. Το παραμένον  $f - \sum_k t_k$  είναι μια

προσέγγιση της βαθυπερατής συνιστώσας της εικόνας. Η cartoon συνιστώσα (5.21) μπορεί να γραφεί προσεγγιστικά ως  $g_0 * f + \lambda' \kappa(u)$ , όπου  $g_0$  μια Gaussian συνάρτηση/περιβάλλουσα, που ερμηνεύεται ως το άθροισμα των μεγάλων μεταβολών φωτεινότητας της εικόνας και η κανονικοποίηση των ακμών με βάση την καμπυλότητα. Η εξίσωση της υφής (5.22) μπορεί να θεωρηθεί ως το παραμένον του ROF, διορθωμένο από το άθροισμα των στενοζωνικών συνιστωσών.

Οι σταθερές παράμετροι  $\lambda, \mu > 0$  καθορίζουν τα παραμένοντα της αποδόμησης και της μοντελοποίησης. Μεγάλα  $\lambda$  οδηγούν σε καλές προσεγγίσεις καθώς  $u + v \rightarrow f$  για  $\lambda \rightarrow \infty$  και λεπτομερέστερες κλίμακες χαρακτηριστικών διατηρούνται στο  $u$ . Το παραμένον μπορεί σε κάθε περίπτωση να θεωρηθεί ως ο υψηπερατός θόρυβος που δεν συμπεριλαμβάνεται στο  $v$ . Ανάλογα, το  $\mu$  «ποινικοποιεί» το σφάλμα μοντελοποίησης της υφής και για μεγάλες τιμές  $\mu \rightarrow \infty$  αυτό το σφάλμα είναι σχεδόν ταυτόσημα μηδενικό καθώς  $v \rightarrow \sum_k t_k$ .

### 5.3.1 Μαθηματική ανάλυση

**Πρόταση 5.1** Για το πρόβλημα ελαχιστοποίησης (5.16) υπάρχει μια λύση  $(\hat{u}, \hat{v})$ , η λύση αυτή είναι μοναδική και η ακολουθία  $(u_n, v_n)$  που την προσεγγίζει αριθμητικά συγκλίνει στην  $(\hat{u}, \hat{v})$  καθώς  $n \rightarrow \infty$ . Η απόδειξη της πρότασης ακολουθεί ξεχωριστά για την ύπαρξη, τη μοναδικότητα και τη σύγκλιση.

#### 5.3.1.1 Ύπαρξη

Το διατυπωμένο συναρτησιακό  $E(u, v)$  είναι κυρτό (convex) και συνεκτικό (coercive). Η κυρτότητα προκύπτει από το γεγονός ότι το πρόβλημα (TV κανονικοποίηση) και οι περιορισμοί του είναι κυρτές συναρτήσεις ή διαφορετικά το υπό εξέταση συναρτησιακό είναι ένα άθροισμα κυρτών όρων [39]. Είναι απλό να επιβεβαιώσει κανείς από την ανισότητα Minkowski για  $L^1, L^2$  ότι

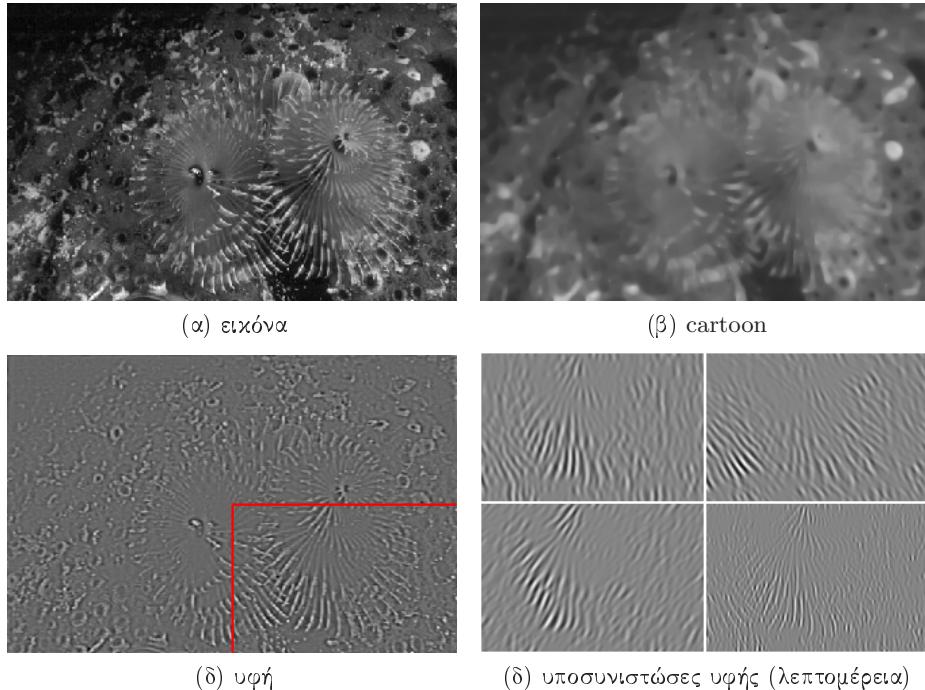
$$E((1-l)u_1 + lu_2, (1-l)v_1 + lv_2) \leq (1-l)E(u_1, v_1) + lE(u_2, v_2), \quad (5.23)$$

$$\forall (u_1, v_1), (u_2, v_2) \in L^2 \times L^2, l \in [0, 1].$$

Επιπρόσθετα, εφόσον το  $E(u, v)$  είναι συνεχές είναι και κάτω-ημισυνεχές (lower-semi continuous) στο  $L^2 \times L^2$ . Η συνεκτικότητα (coerciveness) είναι επίσης εύκολο να διαπιστωθεί καθώς  $E(u, v) \rightarrow +\infty$  για  $|u, v|_{L^2} \rightarrow +\infty$ . Συγκεκριμένα, το  $|(u, v)|_{L^2} = (|u|_{L^2} + |v|_{L^2})^{1/2} \rightarrow +\infty$  υπονοεί είτε ότι  $|u|_{L^2} \rightarrow +\infty$  ή  $|v|_{L^2} \rightarrow +\infty$  ή και τα δύο. Εφόσον  $E(u, v) \geq \lambda \|f - u - v\|_{L^2}^2 + \mu \|v - \sum_k t_k\|_{L^2}^2$  και τα  $f, f_k$  δεν αλλάζουν, εξάγονται ότι  $E(u, v) \rightarrow +\infty$  και επομένως το  $E$  είναι συνεκτικό στο  $L^2 \times L^2$ .

#### 5.3.1.2 Μοναδικότητα

Το συναρτησιακό είναι αυστηρά κυρτό στο  $L^2 \times L^2$ , έτσι μια λύση  $(\hat{u}, \hat{v})$  του (5.16) είναι μοναδική, εκτός από την κατεύθυνση  $(u, -u)$ . Θεωρείστε την τιμή του  $E(\hat{u} + t\hat{u}, \hat{v} - t\hat{u})$  για  $t \neq$



Σχήμα 5.3: Αποσύνθεση της εικόνας (α) σε cartoon (β) και υφή (γ) με βάση το σχήμα  $K + 1$  εξισώσεων. Στο (δ) παρουσιάζονται 4 από τις  $K$  υποσυνιστώσες της υφής για την κάτω δεξιά λεπτομέρεια της εικόνας.

0. Χρησιμοποιώντας ανισότητες στο συναρτησιακό υπό ελαχιστοποίηση (5.16) καταλήγουμε

$$E(\hat{u} + t\hat{u}, \hat{v} - t\hat{u}) \leq E(\hat{u}, \hat{v}) + (|1+t|-1)J(\hat{u}) + |t|\|\hat{u}\|_{L^2}^2 + 2|t|\|\hat{u}\|_{L^2}|\hat{v} - \sum_k t_k|_{L^2}. \quad (5.24)$$

Εφόσον το  $E(\hat{u}, \hat{v})$  είναι ελάχιστο και οι άλλοι όροι είναι πάντα θετικοί για  $t \neq -1$ , τότε αν το  $E(\hat{u} + t\hat{u}, \hat{v} - t\hat{u})$  ήταν επίσης ελάχιστο ύμα ίσχυε η ισότητα για  $\hat{u} = 0$ . Για  $t = -1$  και  $E(0, \hat{u} + \hat{v}) = E(\hat{u}, \hat{v})$  έχουμε επίσης ότι  $\hat{u} = 0$ . Έτσι,  $\forall t \neq 0, (\hat{u} + t\hat{u}, \hat{v} - t\hat{u}) = (\hat{u}, \hat{v})$  και οι δύο λύσεις που ελαχιστοποιούν το  $E$  ταυτίζονται.

### 5.3.1.3 Σύγκλιση

Μια ακολουθία ελαχιστοποίησης  $u_n$  που λύνει την ΜΔΕ (5.18) μέσω gradient descent ή την (5.21) μέσω επαναλήψεων σταθερού σημείου συγκλίνει στην μοναδική λύση  $(\hat{u}, \hat{v})$  του (5.16). Στη δεύτερη περίπτωση θέτουμε  $u_0 = f$  και κατασκευάζουμε την ακολουθία επαναληπτικών λύσεων  $u_n$ ,  $n \rightarrow +\infty$ . Η δε ακολουθία  $v_n$  μπορεί να εκτιμηθεί έμμεσα σε κάθε βήμα από την (5.19). Ορίζουμε με  $(u_{n+1}, v_{n+1})$  τη λύση των (5.21, 5.19) με  $v_{n+1} = v(u_n)$ , όπου η  $u_n$  δίνεται από την (5.21). Εφόσον το  $E(u_n, v_n)$  είναι μια μη-φθίνουσα ακολουθία, που συγκλίνει στο  $\mathbb{R}$ , καταλήγουμε από τη μοναδικότητα της λύσης ότι  $\eta (u_n, v_n)$  συγκλίνει στην  $(\hat{u}, \hat{v})$  για  $n \rightarrow +\infty$ .

### 5.3.2 Πολλαπλές υποσυνιστώσες υφής

Θεωρούμε το πιο σύνθετο, τροποποιημένο πρόβλημα της αποδόμησης μιας εικόνας  $f$  σε  $K+1$  συνιστώσες, δηλαδή ένα  $u + Kv$  μοντέλο

$$f = u + v, \quad v = \sum_k^K v_k, \quad (5.25)$$

όπου  $u$  είναι η συνήθης cartoon συνιστώσα και κάθε μια από τις  $v_k$  είναι μια στενοζωνική συνιστώσα υφής με άγνωστη τοποθέτηση στο φασματικό πεδίο. Στην ουσία πρόκειται για μια διπλή αποδόμηση της εικόνας, όπου το  $u+v$  μοντέλο εξηγεί τη χωρική υπέρθεση και το πολυχαναλικό μοντέλο υφής την αποδόμηση του φάσματος της υφής. Αναζητούμε ένα διαχωρισμό της εικόνας σε συνιστώσες, τέτοιο ώστε το  $f = u + v$  μοντέλο να είναι η ειδική περίπτωση για  $v = \sum_K v_k$ . Για το λόγο αυτό αναδιατυπώνουμε το συναρτησιακό του προβλήματος (5.16) για τις  $K+1$  συνιστώσες

$$\inf_{(u, \{v_k\})} \{E(u, \{v_k\}), k \in \{1, K\}, (u, v_k) \in (BV \times \mathcal{H})\},$$

$$E(u, \{v_k\}) = \|\nabla u\|_{L^1} + \lambda \|f - u - \sum_k^K v_k\|_{L^2}^2 + \mu \sum_k^K \|v_k - t_k\|_{L^2}^2, \quad (5.26)$$

όπου  $(u, \{v_k\}) = (u, v_1, \dots, v_k)$  είναι οι  $K+1$  άγνωστες μεταβλητές. Σημειώνεται ότι για τον όρο συγγένειας στη μοντελοποίηση χρησιμοποιήθηκε μια νόρμα διανύσματος, δηλαδή μια σύνθεση τετραγωνισμένων νορμών  $l^2(k) \circ L^2(\Omega)(v_k - f_k)$  [33]. Μια εναλλακτική λογική επιλογή θα ήταν μια νόρμα  $L^2(\Omega) \circ l^2(k)$ .

Εφαρμόζοντας Euler-Lagrange ελαχιστοποίηση στο παραπάνω συναρτησιακό καταλήγουμε στο σύνολο  $K+1$  εξισώσεων σταθερής κατάστασης

$$u = (f - \sum_{k=1}^K v_k) + \frac{1}{2\lambda} \kappa(u), \quad (5.27)$$

$$v_k = \frac{\lambda}{\mu} (f - u - \sum_{k=1}^K v_k) + t_k, \quad k \in \{1, K\}, \quad (5.28)$$

όπου κάθε  $v_k$  αναφέρεται σε μια μοναδική υποσυνιστώσα υφής. Προσθέτοντας τις  $K$  μεμονωμένες εξισώσεις (5.28) των συνιστωσών της υφής λαμβάνουμε την ανακατασκευή του  $u$  ως την υπέρθεση

$$\sum_{k=1}^K v_k = \frac{1}{\mu + K\lambda} \left( K\lambda(f - u) + \mu \sum_{k=1}^K t_k \right). \quad (5.29)$$

Σχετικά με την τελευταία αυτή εξίσωση ανακατασκευής μπορούν να γίνουν κάποιες ενδιαφέρουσες παρατηρήσεις:

- (α) Η ΜΔΕ για το  $u$  δίνεται σε κλειστή, αναλυτική μορφή αντικαθιστώντας την (5.29) στην (5.27) και εξαρτάται μόνο από τα  $(f, u, t_k)$ , όπως στην  $u+v$  περίπτωση (5.21). Υποστηρίζουμε ότι η συνιστώσα της υφής  $u$  είναι προσεγγιστικά το άθροισμα της (5.29), δηλαδή  $u \approx \sum_{k=1}^K v_k$ .
- (β) Κάθε μεμονωμένη υποσυνιστώσα  $v_k$ , έμεσα εξαρτώμενη από τον ορισμό και τις παραμέτρους του  $t$ , μπορεί να εκτιμηθεί απευθείας από τη λύση  $u$  χρησιμοποιώντας τις (5.29, 5.28).

(γ) Οι παράμετροι της αποδόμησης  $\lambda, \mu$  μπορούν να συσχετισθούν απαιτώντας  $v \geq 0$ . Αυτό υπονοεί ότι  $\mu/\lambda > K$ , δηλαδή ότι το  $\mu$  πρέπει να είναι κατά προσέγγιση μια τάξη μεγέθους μεγαλύτερο από το  $\lambda$ , με τυπικές τιμές για το  $K$ , όπως θα δούμε  $K \in \{5, 40\}$ .

(δ) Αν αφήσουμε το  $\mu \gg \lambda K$ , τότε η υπέρθεση (5.29) γίνεται

$$\sum_{k=1}^K v_k \approx \sum_{k=1}^K t_k, \quad \mu \gg \lambda K \quad (5.30)$$

δηλαδή το  $v$  προσεγγίζεται από την ανακατασκευή του  $t$  αλλά η προσέγγιση μπορεί να ερμηνευθεί και αντίστροφα. Έστω ότι στο  $u + v$  σχήμα των (5.21, 5.22) οι υποσυνιστώσες μοντελοποιούνται με κάποιες από τις αποχρίσεις της συστοιχίας, δηλαδή  $v_k = g_k * u$ . Τότε, σε κάθε επανάληψη  $n$  μιας αριθμητικής λύσης ο όρος  $\sum_k t_k$  μπορεί να προσεγγιστεί από την προβολή της τρέχουσας εκτίμησης για την υφή  $u_n$  στο προκαθορισμένο σύνολο φίλτρων (Εν. 2.2.3). Παρατηρούμε δε ότι το πρόβλημα (5.16) εξάγεται ως ειδική περίπτωση της γενικής αποδόμησης της εικόνας σε  $K+1$  συνιστώσες  $(u, v_1, \dots, v_K)$ .

### 5.3.3 Θεωρητικές συγκρίσεις

Με βάση το επιχείρημα ότι η υφή είναι η ταλαντευόμενη συνιστώσα της εικόνας [214], τα τρέχοντα σχήματα αποδόμησης [7, 8, 228, 298] όπως διατυπώνονται από τις εξισώσεις (5.9, 5.11) και (5.12) αναζητούν συναρτήσεις με συγκεκριμένες ιδιότητες όπως μικρή ολική μεταβολή, μεγάλες ταλαντώσεις, περιοδικότητα, τυχαιότητα κ.ά. Αντίθετα, το  $u + K_v$  σχήμα ορίζει αποκλειστικά μια τάξη γενικών ταλαντώσεων και συγκεκριμένα την οικογένεια των διαμορφωμένων κατά πλάτος και συχνότητα συναρτήσεων  $a(x, y) \exp i\phi(x, y)$ .

Η υφή μοντελοποιείται από την υπέρθεση ταλαντώσεων σε διαφορετικές συχνότητες, με ρυθμό που μεταβάλλεται σύμφωνα με το σήμα φάσης  $\phi(x, y)$  και πλάτος ομαλά μεταβαλλόμενο σύμφωνα με το  $a(x, y)$ . Παράπλευρα επιτυγχάνεται μια αποδόμηση σε δεύτερο επίπεδο σε υποσυνιστώσες φασματικά διαχωρισμένες καθώς και μια αναπαράσταση της υφής ως προς τις συναρτήσεις διαμόρφωσης πλάτους και συχνότητας.

Στο TV-Gabor μοντέλο [10, 11] που περιγράφεται από το πρόβλημα (5.13) ο όρος πιστότητας περιλαμβάνει προβολή της διαφοράς  $v = f - u$  σε προκαθορισμένες Gabor συναρτήσεις, έστω  $g'_k$ . Η νόρμα στον Hilbert χώρο  $\mathcal{H}$  ορίζεται με εσωτερικό γινόμενο, αλλά για τις ανάγκες της σύγκρισης η διατύπωση της μπορεί να γραφτεί με συνελικτικούς όρους

$$\begin{aligned} \|f - u\|_{\mathcal{H}} &= \langle f - u, g'_k * (f - u) \rangle \\ &= \|(f - u) * (\delta(t) - g_k)^{1/2}\|_{L^2}^2 \\ &= \|f - u\|_{L^2}^2 - \|(f - u) * g_k\|_{L^2}^2 \\ &= \|v\|_{L^2}^2 - \|v * g_k\|_{L^2}^2. \end{aligned} \quad (5.31)$$

Η ποσότητα είναι πάντοτε θετική καθώς μπορεί να δειχθεί ότι για κανονικοποιημένα  $L_2$  φίλτρα,  $\|v\|_{L^2}^2 \geq \|v * g_k\|_{L^2}^2$ . Ο Gabor πυρήνας  $g_k$  είναι *a-priori* επιλεγμένος με βάση πληροφορία για τη συχνότητα ή τον προσανατολισμό της ταλάντωσης της υφής στην εικόνα. Πολλαπλοί

## Κεφάλαιο 5. Συνέργεια διαχωρισμού και μοντελοποίησης: Αποσύνθεση εικόνων

προσανατολισμοί μπορούν να ληφθούν υπόψιν αυθοίζοντας ένα σύνολο από προβολές σε Gabor

$$\|f - u\|_{\mathcal{H}} = K\|v\|_{L^2}^2 - \sum_k^K \|v * g_k\|_{L^2}^2. \quad (5.32)$$

Στο  $u + Kv$  σχήμα μέσω στενοζωνικής Gabor αποδόμησης που αναπτύχθηκε, αν υποθέσουμε ότι  $t = f$  ή ( $\mathbf{T} = \mathbf{I}$ ) που σημαίνει ότι  $t_k = g_k * f$  και προσεγγίσουμε τις υποσυνιστώσες της υφής από τις αποκρίσεις  $v_k \approx g_k * v$ , τότε ο περιορισμός στο (5.26) γίνεται

$$\sum_k \|v_k - t_k\|_{L^2}^2 \approx \sum_k \|(f - v) * g_k\|_{L^2}^2 = \sum_k \|u * g_k\|_{L^2}^2. \quad (5.33)$$

Επομένως, ενώ το TV-Gabor ελαχιστοποιεί την διαφορά στην  $L_2$ -ενέργεια ανάμεσα στο  $v$  και στις επιλεγμένες *a-priori* συνιστώσες, το προτεινόμενο  $u + Kv$  μπορεί να θεωρηθεί ότι ελαχιστοποιεί την ενέργεια του *cartoon* στα «κανάλια υφής». Ένα σημαντικό πλεονέκτημα που αναδεικνύεται έτσι είναι ότι το νέο σχήμα λειτουργεί όπως ένα μοντέλο  $u + v + w$ , όπου ο παραμένον ύδρυθος  $w$  θα ελαχιστοποιηθεί μαζί με το  $u$  στα κανάλια της υφής, ενώ στο TV-Gabor θα συμπεριληφθεί πιο εύκολα στο  $v$ , ειδικά στην περίπτωση λευκού ύδρυθου.

### 5.3.4 Αλγόριθμος και διακριτοποίηση

Για την προσέγγιση της λύσης των (5.21, 5.22) ή (5.27, 5.28) χρησιμοποιείται ένα αριθμητικό σχήμα επαναλήψεων σταθερού σημείου (fixed point Gauss Seidel iteration), με διακριτοποίηση παρόμοια με αυτή που προτείνεται στο [298]: πεπερασμένες συμμετρικές διαφορές για τις παραγώγους με την τρέχουσα, πιο πρόσφατη τιμή  $u^{(n)}$  να χρησιμοποιείται σε κάθε σημείο και μια διακριτή εκδοχή της καμπυλότητας  $\kappa^{(n)}(u^{(n)}, u^{(n+1)})$ . Μοναδικότητες (singularities) που προκύπτουν για  $\|\nabla u\| = 0$ , καθώς και από τον μη-ορισμό της παραγώγου της TV νόρμας (5.4) για  $u = 0$  λαμβάνονται υπόψιν στη διακριτή νόρμα κανονικοποιώντας το μέτρο του gradient [1, 108, 299] ( $J_e(u) = \int_{\Omega} \sqrt{|\nabla u|^2 + e}$ , με  $e$  πολύ μικρή σταθερά).

Λεπτομέρειες για τη διακριτοποίηση και δειγματοληψία των Gabor φίλτρων μπορούν να βρεθούν στα [36, 250] και για την υλοποίηση των διακριτών ενεργειακών τελεστών και των σχημάτων αποδιαμόρφωσης στα [92, 165, 198]. Για αποτελεσματικότητα και ταχύτητα το φιλτράρισμα γίνεται στο πεδίο των συχνοτήτων με χρήση FFT, όπως και η υλοποίηση του Gabor ESA με χρήση της σχέσης

$$\mathcal{F} \left\{ \frac{\partial^{m+\ell} g}{\partial x^m \partial y^\ell} \right\} = \mathcal{F}\{g\} (j\omega_x)^m (j\omega_y)^\ell \quad (5.34)$$

που σχετίζει το Fourier μετασχηματισμό  $\mathcal{F}\{\cdot\}(\omega_x, \omega_y)$  της συνάρτησης  $g$  και τις μερικές παραγώγους τάξης  $m$  και  $\ell$ .

Ο αλγόριθμος επαναληπτικά υπολογίζει για κάθε ύδεση  $(i, j)$  την επόμενη τιμή  $u^{(n+1)}$  από τη διακριτοποιημένη (5.21) και την επόμενη  $v^{(n+1)}$  από την αντίστοιχη (5.22), αν χρησιμοποιείται για την ανανέωση της εκτίμησης του αυθοίσματος ανακατασκευής  $\sum_k t_k^{(n)}$  (5.30). Εκτός από την ανανέωση ή μη της τιμής του, υπάρχουν εναλλακτικές επιλογές και ως προς τον τρόπο

υπολογισμού του αυθροίσματος. Συγκεκριμένα, έστω  $\mathcal{T}$  ο διακριτός τελεστής της απεικόνισης (γραμμικής ή μη-γραμμικής) της  $f$  στο  $t$ , τότε στο βήμα  $(n+1)$  η κάθε υποσυνιστώσα υπολογίζεται από το μετασχηματισμό

$$t_k^{(n+1)} = \mathcal{T}\{g_k * f\} \quad \text{ή} \quad t_k^{(n+1)} = \mathcal{T}\{g_k * v^{(n)}\} \quad (5.35)$$

ανάλογα με το αν επιλέγεται τα  $t_k^{(n)}$  να είναι προϋπολογισμένα και αμετάβλητα ή να επανεκτιμούνται σε κάθε επανάληψη (5.30). Ο υπολογισμός τους, και ως αποτέλεσμα ο υπολογισμός του αυθροίσματος χρησιμοποιεί διακριτές συνελίξεις και/ή αποδιαμόρφωση (για την επιλογή των  $\mathcal{T}, K$  βλέπε Εν. 5.4).

Τελικά, οι δύο διακριτές συνιστώσες υπολογίζονται σε κάθε βήμα από τις γραμμικοποιημένες εξισώσεις

$$u^{(n+1)} = f - \sum_{k=1}^K t_k^{(n+1)} + 0.5 \left( \frac{1}{\mu} + \frac{1}{\lambda} \right) \kappa^{(n)}(u^{(n)}, u^{(n+1)}), \quad (5.36)$$

$$v^{(n+1)} = \frac{1}{\mu + \lambda} \left( \lambda(f - u^{(n+1)}) + \mu \sum_{k=1}^K t_k^{(n+1)} \right). \quad (5.37)$$

Οι επαναλήψεις σταματάνε μόλις η μέση μεταβολή στις τιμές γίνει μικρότερη από κάποια τιμή  $\epsilon$ , δηλαδή:

$$\max \left\{ \sum_{i,j} (u^{(n+1)} - u^{(n)}), \sum_{i,j} (v^{(n+1)} - v^{(n)}) \right\} \leq \epsilon. \quad (5.38)$$

Το  $u + Kv$  σχήμα υλοποιείται ανάλογα, με τις  $K$  υποσυνιστώσες υφής να δίνονται με διακριτοποίηση των σχέσεων (5.28).

## 5.4 Περιορισμός του μοντέλου και ανακατασκευή υφής

Η επιλογή του αυθροίσματος συνιστωσών στις (5.22, 5.29), δηλαδή ο περιορισμός που εισάγει το μοντέλο για το  $v$ , ισοδυναμεί με την επιλογή των καναλιών ή των συνιστωσών εκείνων από τον πολυσυχνοτικό διαχωρισμό της εικόνας που αναπαριστούν πιο πιστά την υφή. Πληροφορία υφής απαντάται σε διαφορετικές κλίμακες, που εκδηλώνεται από την εμφάνιση ενός εύρους (μικρο)δομών, από primitives ή τεξτόνια δύο σημείων μέχρι αργά μεταβαλλόμενα, οργανωμένα πρότυπα. Το ίδιο ισχύει για τη γεωμετρική πληροφορία όπως οι ακμές έντασης φωτεινότητας και τα περιγράμματα, που ανιχνεύονται σε μεγάλο εύρος από φυσικές και οπτικές κλίμακες χαρακτηριστικών [179]. Επομένως τα δύο είδη πληροφορίας εικόνας συνυπάρχουν και η ερμηνεία τους εξαρτάται από την κλίμακα της περιγραφής. Θεωρητικά αλλά και αντιληπτικά, η ανάθεση ζωνών συχνοτήτων σε μία από τις δύο κατηγορίες χαρακτηριστικών δεν είναι προφανής, διαφορετικά ο διαχωρισμός δομής-υφής θα λυνόταν απλά στο φασματικό πεδίο. Υπολογιστικά μια τέτοια ανάθεση θα απαιτούσε πρότερη γνώση ή μέτρα της φασματικής κατανομής χαρακτηριστικών όπως οι ακμές και τα δομικά στοιχεία της υφής.

### 5.4.1 Επιλογή συνιστωσών

Στο συναρτησιακό (5.16) το άθροισμα  $\sum t_k$  στην πιο ακατέργαστη μορφή του είναι μια ανακατασκευή του  $t$  από το σύνολο των αποκρίσεων των φίλτρων  $g_k$ , όπου ο αριθμός των συνιστωσών που καθορίζεται από τη συστοιχία είναι  $K$ . Με επιλογή ενός υποσυνόλου του  $K$  και μιας απεικόνισης  $f : \Omega \mapsto t$ , μπορεί να ανακατασκευαστεί μόνο η σχετική με την υφή πληροφορία της εικόνας, να μειωθεί η πλεονάζουσα πληροφορία και να καθοδηγηθεί η διαδικασία αποδόμησης. Προσέγγιση μιας εικόνας με λιγότερες ζωνοπερατές συνιστώσες μπορεί να γίνει με τεχνικές ελαχιστοποίησης ενεργειακών [138] καθώς και τεχνικές επιλογής όπως η ανάλυση πρωτευουσών συνιστωσών (principal components analysis ή PCA) [84].

#### 5.4.1.1 Λεπτομερείς κλίμακες και υψησυχνες ζώνες

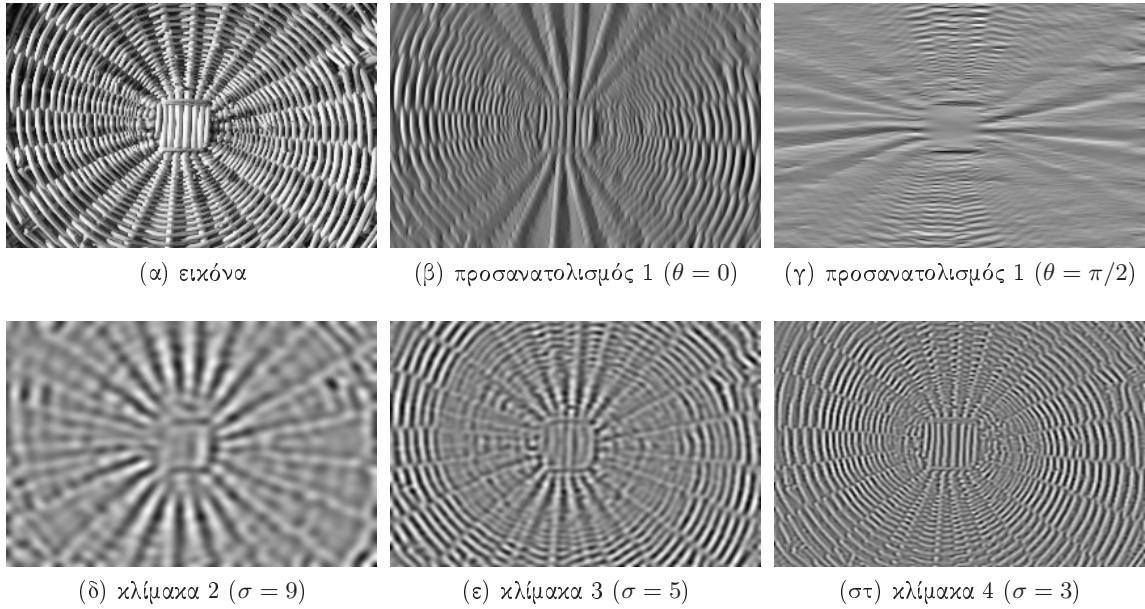
Η πληροφορία της εικόνας που ανακατασκευάζεται από το άθροισμα συνιστωσών μπορεί να εμφανίζεται ανομοιόμορφα σε όλο το εύρος συχνοτήτων. Μια εύλογη ερώτηση είναι αν μπορεί κανείς να αναγνωρίσει ή να ανιχνεύσει την πληροφορία υφής και ακμών στις διαφορετικές ζώνες συχνοτήτων. Διασυνθητικά αλλά και πειραματικά, η υφή σχετίζεται με τις ζώνες μεσαίων και μεγάλων συχνοτήτων [56] (και επομένως στις μικρότερες κλίμακες) και οι σημαντικές για την οργάνωση της σκηνής ακμές, που αντιστοιχούν σε εξέχοντα χαρακτηριστικά (π.χ. περιγράμματα αντικειμένων) στις μεσαίες και μικρότερες συχνότητες (μεγάλες, αδρομερείς κλίμακες). Υπολογιστικά, οι μικροακμές ή ακμές φωτεινότητας στο εσωτερικό περιοχών υφής διαχωρίζονται από τα όρια των διαφορετικών περιοχών (texture boundaries) ή μακροακμές [118,152]. Αντιληπτικά, σύνολα από ακμές μικρής κλίμακας (μικροακμές) ομαδοποιούνται δίνοντας σχηματισμούς υφής ενώ μια μικρή κλίμακα οπτικής οξύτητας (visual acuity) καθορίζει τη διάκριση υφής/περιγραμμάτων [110]. Έτσι, προτρέπεται κανείς σε έναν χονδρικό διαχωρισμό του χώρου των κλιμάκων σε μικρές και κυρίαρχες για την υφή και μεγάλες και σημαντικές για τις δομές της εικόνας.

Έστω ότι το άθροισμα ξαναγράφεται ως προς τις  $R$  κλίμακες και τους  $P$  προσανατολισμούς της συστοιχίας των  $K = RP$  φίλτρων

$$\sum_{k=1}^K t_k(x, y) = \sum_{\sigma=\sigma_1}^{\sigma_R} \sum_{\theta=\theta_1}^{\theta_P} t_{\sigma, \theta}(x, y), \quad k = R(\theta - 1) + \sigma, \quad (5.39)$$

όπου  $\sigma, \theta$  οι δείκτες κλίμακας και προσανατολισμού. Το  $\sigma$  προχωράει από τις ζώνες χαμηλών στις ζώνες υψηλών συχνοτήτων και το  $\theta$  από 0 σε  $\pi$  (rad). Αυτό καθιστά το άθροισμα μια ίσης-κλίμακας ανακατασκευή από ισο-κατευθυντικές συνιστώσες. Ως προς το χώρο των φίλτρων, αυτό ισοδυναμεί με αυθοίσματα αποκρίσεων ρυθμισμένα σε μία κλίμακα ή σε ένα προσανατολισμό.

Τέτοια αυθοίσματα φαίνονται στο Σχ. 5.4 για έναν προσανατολισμό, και τρεις κλίμακες (μεγάλη, μεσαία, μικρή). Η εικόνα ενός προσανατολισμού ( $\beta$ ) περιλαμβάνει όλες τις κλίμακες των φίλτρων και κάθε κλίμακα ( $\delta$ ), ( $\epsilon$ ), ( $\sigma$ ) όλους τους προσανατολισμούς της συστοιχίας. Μια τέτοια αποδόμηση-ανακατασκευή είναι εξαρτώμενη από τα σχεδιαστικά χαρακτηριστικά της συστοιχίας και επομένως μη-μοναδική.



Σχήμα 5.4: Αθροίσματα στενοζωνικών συνιστωσών της εικόνας (α), ως προς όλες τις κλίμακες για κάθετο προσανατολισμό φίλτρων στο (β) και οριζόντιο στο (γ), και ως προς όλες τις κατευθύνσεις για μεγάλη (δ), μεσαία (ε) και μικρή (στ) κλίμακα. Η κλίμακα καθορίζεται από την τυπική απόκλιση  $\sigma$  της κοινής Gauss περιβάλλουσας των καναλιών, και ο προσανατολισμός  $\theta$  από την κοινή γωνία του διανύσματος κεντρικής συνχότητας σε μια ακτινική κατεύθυνση της συστοιχίας. Μικρές κλίμακες αντιστοιχούν σε μεγάλες (κεντρικές) συχνότητες ταλάντωσης και σε δείγματα από μια Gabor πυραμίδα κλίμακας-χώρου (scale-space).

Στην περίπτωση Gabor καναλιών (2.1), η περιβάλλουσα καθορίζει την κλίμακα της συνιστώσας και το εκθετικό φέρον των προσανατολισμό της. Γράφοντας  $g_{\sigma,\theta} = G_\sigma C_{\sigma,\theta}$ , όπου  $G_\sigma = (2\pi\sigma^2)^{-1} \exp\{-(x^2 + y^2)/2\sigma_k^2\}$  η Gaussian κλίμακας  $\sigma$  και  $C_{\sigma,\theta} = \exp\{|\omega|(j \cos(\theta)x + j \sin(\theta)y)\}$  με  $|\omega| = |\omega|(\sigma)$  το μιγαδικό εκθετικό γωνίας  $\theta$  τότε

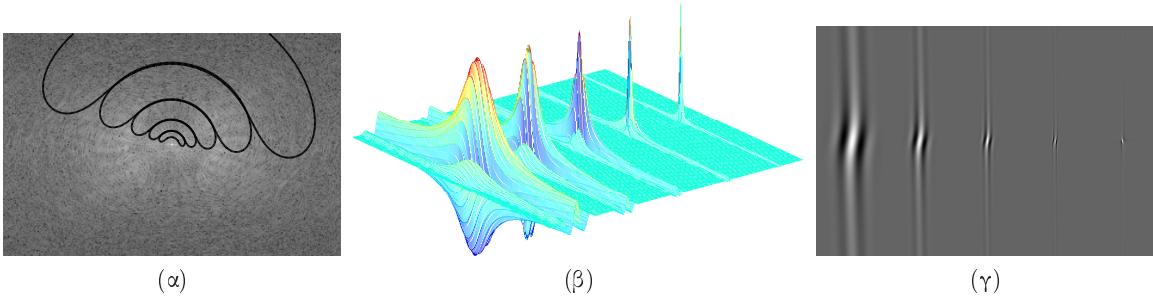
$$\sum_{k=1}^K t_k = \sum_{\sigma=\sigma_1}^{\sigma_R} G_\sigma \sum_{\theta=\theta_1}^{\theta_P} (C_{\sigma,\theta} * f) = \sum_{\sigma=\sigma_1}^{\sigma_R} (G_s * f) \sum_{\theta=\theta_1}^{\theta_P} C_{\sigma,\theta}. \quad (5.40)$$

Σε αυτό το «άθροισμα κατά κλίμακες», το  $(G_\sigma * f)$  είναι η βαθυπερατή προσέγγιση της αρχικής εικόνας  $f$  σε κλίμακα  $s$ . Έτσι, η ανακατασκευή ερμηνεύεται ως ένα διακριτό Gaussian scale-space [13, 159, 179], που διαμορφώνεται από το άθροισμα των φερόντων σε κάθε προσανατολισμού. Εννοιολογικά ισοδυναμεί με δειγματοληψία μιας 1-Δ πυραμίδας Gabor scale-space καθώς η άθροιση ως προς τον προσανατολισμό ολοκληρώνει (integrates-out) τη μία από τις δύο μεταβλητές χωρικής συχνότητας. Στο Σχ. 5.5 παρουσιάζονται στο πεδίο των συχνοτήτων της 5.4 (α) οι «ζώνες» κλίμακας που σχηματίζονται από τη διαμορφωμένη βαθυπερατή περιβάλλουσα μαζί με τις ισοδύναμες αποχρίσεις των σύνθετων φίλτρων

$$G_\sigma \sum_{\theta} C_{\sigma,\theta} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \sum_{\theta} \exp\{|\omega|(j \cos(\theta)x + j \sin(\theta)y)\}. \quad (5.41)$$

Οι συνιστώσες των ζωνών χαμηλότερης συχνότητας, κάτω από ένα κατώφλι κλίμακας  $s_T$ , καταπιέζονται ορίζοντας

$$\sum_{k=1}^K t_k \triangleq \sum_{\sigma=\sigma_T}^{\sigma_R} (G_\sigma * f) \sum_{\theta=\theta_1}^{\theta_P} C_{\sigma,\theta} \quad (5.42)$$



Σχήμα 5.5: (α) Μετασχηματισμός Fourier της εικόνας 5.4 (α), με τις ζωνοπερατές ζώνες ίσης κλίμακας, που καθορίζονται από τις Gaussian περιβάλλουσες των καναλιών και (β),(γ) αποκρίσεις των ισοδύναμων φίλτρων.

με σκοπό την ανακατασκευή πρώτιστα της πληροφορίας υφής από τις ζώνες υψηλότερων συχνοτήτων και την ελαχιστοποίηση του rms σφάλματος  $\|v - \sum_k^K t_k\|_{L^2}$ . Ο αριθμός των συνιστωσών επομένως μειώνεται σε ένα πολλαπλάσιο του αριθμού των συνολικών προσανατολισμών της συστοιχίας. Η διατήρηση ενός εύρους κλίμακων που επιτελείται μπορεί να παραλληλιστεί με «συρρίκνωση» συνιστωσών βασισμένη στην κλίμακα. Η επιλογή της αδρομερέστερης κλίμακας υφής, δηλαδή το κατώφλι, μπορεί να γίνει με χωρικά προσαρμοστικά μετρικά κλίμακας [23, 42, 95, 108, 163, 180]. Περιορίζουμε την ανάλυση εδώ στην περίπτωση μιας καθολικής επιλογής ενός εύρους πάνω από ένα ευρετικό κατώφλι-κλίμακα. Μια πιθανή επέκταση θα ήταν η εφαρμογή μιας ήπιας κατώφλιοποίησης της κλίμακας στο άθροισμα των συνιστωσών. Αυτόματη επιλογή κλίμακας θα παρουσιαστεί στη συνέχεια μέσω των κυρίαρχων συνιστωσών.

#### 5.4.1.2 Στάθμιση πλατών

Οι στενοζωνικές συνιστώσες  $f_k = f * g_k$  είναι αποτέλεσμα της συμβολής τόσο του περιεχομένου υφής όσο και του λοιπού περιεχομένου της εικόνας, δηλαδή  $f_k = u * g_k + v * g_k$ . Σύμφωνα με το σενάριο της σταθμισμένης ανακατασκευής κάθε συνιστώσα κανονικοποιείται από ένα χωρικά προσαρμοστικό μετρικό που ενισχύει τοπικά τη συνεισφορά της υφής στο άθροισμα.

Ένα εύλογο μετρικό είναι η περιβάλλουσα του πλάτους της συνιστώσας, η οποία είναι άμεσα διαθέσιμη από την αποδιαμόρφωση του καναλιού (2.23). Πολλαπλασιάζοντας με το κανονικοποιημένο πλάτος του αντίστοιχου καναλιού το άθροισμα ανακατασκευής μετατρέπεται σε μια σταθμισμένη, μέση συνιστώσα

$$\sum_{k=1}^K t_k \triangleq \sum_k \frac{\alpha_k(f * g_k)}{\sum_k \alpha_k}. \quad (5.43)$$

Η ανακατασκευή είναι γραμμική ως προς τις επιμέρους συνιστώσες  $f_k$ , δίνει έμφαση στις μεγάλες ενεργειακά συνιστώσες σε κάθε σημείο και αυξάνει τη διαχωρισμότητα των συνιστωσών με κοντινό φασματικό περιεχόμενο. Το πλάτος  $\alpha_k(x, y)$  λαμβάνει μεγάλες τιμές στις περιοχές

υψηλής αντίθεσης που υποδηλώνουν μεγάλες ταλαντώσεις, συντονισμένες στη συχνότητα και τον προσανατολισμό της συνιστώσας και μικρές στις σχετικά ομαλές περιοχές. Κατά συνέπεια, η στάθμιση εύρους προωθεί τη συμβολή από το ταλαντευόμενο κομμάτι της υφής ακυ και περιορίζει το τμηματικά ομαλό ακυ. Ωστόσο, το πλάτος της συνιστώσας μπορεί να ανταποκρίνεται και σε απότομες μεταβάσεις ακμές καθώς το μοντέλο τις ερμηνεύει ως ταλαντώσεις μιας περιόδου, οι οποίες όμως απομακρύνονται από το ν κατά την TV κανονικοποίηση.

Ας εξετάσουμε την εσωτερική στάθμιση για μια στενοζωνική συνιστώσα  $f_k = u * g_k + v * g_k = v_k + u_k$ , η οποία αποτελείται από τις δύο υποσυνιστώσες cartoon  $u_k$  και υφής  $v_k$  στην ίδια ζώνη  $k$ . Η κάθε μια μπορεί να γραφεί ως αναλυτικό σήμα

$$u_k = \alpha_{uk} e^{\phi_{uk}}, \quad \alpha_{uk} = |u_k| = |u * g_k|, \quad (5.44)$$

$$v_k = \alpha_{vk} e^{\phi_{vk}}, \quad \alpha_{vk} = |v_k| = |v * g_k|. \quad (5.45)$$

Η αρχική συνιστώσα  $f_k$ , επομένως, είναι το άθροισμα των δύο αναλυτικών σημάτων

$$f_k = \alpha_{uk} e^{j\phi_{uk}} + \alpha_{vk} e^{j\phi_{vk}}, \quad (5.46)$$

ενώ για τη σταθμισμένη με πλάτη ανακατασκευή, η προκύπτουσα συνιστώσα είναι

$$\tilde{f}_k = \frac{\alpha_{uk} u_k + \alpha_{vk} v_k}{\alpha_{uk} + \alpha_{vk}} = \frac{\alpha_{uk}^2 e^{j\phi_{uk}} + \alpha_{vk}^2 e^{j\phi_{vk}}}{\alpha_{uk} + \alpha_{vk}}. \quad (5.47)$$

Πολλαπλασιάζοντας την (5.46) δύο φορές με τα πλάτη των  $u_k$  και  $v_k$  και αντικαθιστώντας στον αριθμητή της (5.47) προκύπτει μια τροποποιημένη έκφραση για την ανακατασκευασμένη συνιστώσα

$$\tilde{f}_k = \frac{(\alpha_{uk} + \alpha_{vk}) f_k - \alpha_{vk} \alpha_{uk} (e^{j\phi_{uk}} + e^{j\phi_{vk}})}{\alpha_{uk} + \alpha_{vk}} \Rightarrow \quad (5.48)$$

$$\tilde{f}_k = f_k - \frac{\alpha_{vk} u_k + \alpha_{uk} v_k}{\alpha_{uk} + \alpha_{vk}}. \quad (5.49)$$

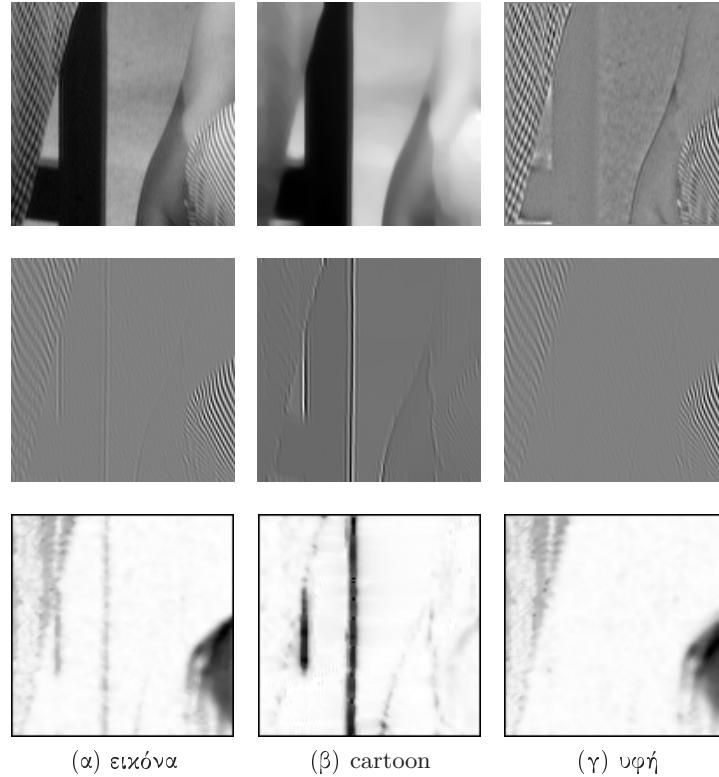
Ερμηνεύοντας την παραπάνω σχέση μπορούνε να σημειωθούν τα εξής: Η σταθμισμένη συνιστώσα είναι η πραγματική μείον το σταθμισμένο με αντίστροφα βάρη άθροισμα των  $u_k, v_k$ . Σε κάθε σημείο της  $\tilde{f}_k$  τονίζει ή καταπιέζει γεωμετρικά χαρακτηριστικά ή υφή. Συγκεκριμένα, εάν ένα σημείο  $(x, y)$  ανήκει στο σύνολο των ακμών (edge set)  $C_k$  της συγκεκριμένης ζώνης συχνοτήτων  $k$ , τότε  $\alpha_{uk} \gg \alpha_{vk}$  και η εξίσωση (5.49) γίνεται

$$\tilde{f}_k \approx f_k - v_k, \quad \forall (x, y) \in C_k. \quad (\text{edge set}) \quad (5.50)$$

Αντίθετα σε περιοχές υφής  $T_k$  στη ζώνη  $k$  όπου  $\alpha_{vk} \gg \alpha_{uk}$  η εξίσωση προσεγγίζει

$$\tilde{f}_k \approx f_k - u_k, \quad \forall (x, y) \in T_k. \quad (\text{texture}) \quad (5.51)$$

Επομένως, η στάθμιση με τα επιμέρους πλάτη απομακρύνει εσωτερικά σε κάθε συνιστώσα τη συμβολή του cartoon στα σημεία υφής και τις ταλαντώσεις της υφής από σημεία μεταβάσεων και ακμών. Το ζωνοπερατό cartoon ταυτίζεται με τις ακμές αφού οι ομαλές μεταβολές, κυρίως στις μεγαλύτερες συχνότητες, απομακρύνονται από τα φίλτρα μοναδιαίας ενέργειας και



Σχήμα 5.6: Εικόνα (λεπτομέρεια  $128 \times 128$  pixel), cartoon και υφή από το σχήμα Vese-Osher με  $(\lambda, \mu) = (5, 0.01)$  (πάνω σειρά), οι φιλτραρισμένες εκδοχές τους (μεσαία) και τα πλάτη (κάτω, σε «ανεστραμμένη» κλίμακα γκριζου) στην τέταρτη κλίμακα ( $\sigma = 3$ ) και στον όγδοο προσανατολισμό ( $\theta = -\pi/8$ ) της συστοιχίας Gabor φίλτρων

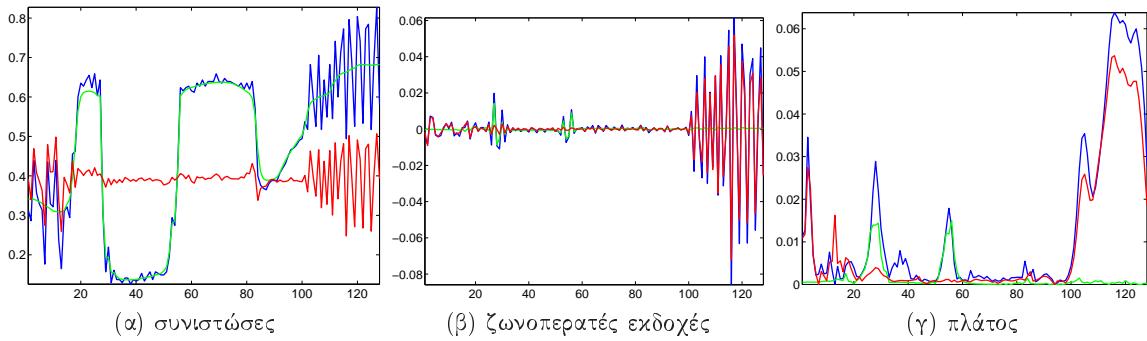
μηδενικής DC. Η επίδραση της στάθμισης πλάτους μπορεί να παρατηρηθεί στα Σχτ. 5.6,5.7, για μια εικόνα και τις 1-Δ γραφικές παραστάσεις για μια γραμμή της.

Κατ' επέκταση, θεωρώντας  $K$  στενοζωνικές συνιστώσες η ανακατασκευή που προκύπτει με απλή υπέρθεση τους είναι

$$\begin{aligned} \sum_k \tilde{f}_k &= \sum_k f_k - \sum_k \frac{\alpha_{vk} u_k + \alpha_{uk} v_k}{\alpha_{uk} + \alpha_{vk}} \\ &\approx f - \sum_k \frac{\alpha_{vk}}{\alpha_{uk} + \alpha_{vk}} u_k + \frac{\alpha_{uk}}{\alpha_{uk} + \alpha_{vk}} v_k \\ &= f - \sum_k \frac{1}{1 + \lambda_k} u_k + \frac{\lambda_k}{1 + \lambda_k} v_k, \quad \lambda_k = \frac{\alpha_{uk}}{\alpha_{vk}}, \end{aligned} \quad (5.52)$$

δηλαδή η υπέρθεση ακολουθεί παρόμοιες προσεγγίσεις για την ανάκτηση της εικόνας πλήρους εύρους. Η χωρικά προσαρμοστική παράμετρος  $\lambda_k = \lambda_k(x, y)$  ρυθμίζει για κάθε  $k$  σε κάθε σημείο τη συμβολή των  $u_k, v_k$  στην ανακατασκευή. Αποτελεί έτσι ένα ποσοτικό μέτρο της συσχέτισης cartoon και υφής σε κάθε ζώνη.

Λίγο διαφορετικά είναι τα πράγματα εξετάζοντας την ανακατασκευή μέσω της εξίσωσης (5.43), δηλαδή με στάθμιση των πλατών των στενοζωνικών συνιστώσων  $f_k$  σε αντίθεση με τη στάθμιση ατομικά από τα πλάτη των  $u_k, v_k$ . Πράγματι, παρατηρώντας από την (5.46) ότι



Σχήμα 5.7: Προφίλ συναρτήσεων για μια γραμμή (84) της εικόνας στο Σχ. 5.6. (α) αποδόμηση της  $f$  σε cartoon  $u$  και υφή  $v$  χρησιμοποιώντας το σχήμα Vese-Osher (5.11), (β) απόκριση Gabor φίλτρου ( $\sigma = 3, \theta = -\pi/8$ ) στις συναρτήσεις  $f, u, v$ , (γ) αποδιαμορφωμένο πλάτος (μπλε: αρχική, πράσινο: cartoon, κόκκινο: συνιστώσα υφής).

$\alpha_k^2 = \alpha_{uk}^2 + \alpha_{vk}^2$  και εφαρμόζοντας την (5.43) με τετραγωνικά βάρη

$$\sum_{k=1}^K t_k = \sum_k \frac{\alpha_k^2 f_k}{\sum_k \alpha_k^2} = \sum_k \frac{(\alpha_{uk}^2 + \alpha_{vk}^2) f_k}{\sum_k \alpha_k^2}. \quad (5.53)$$

Εισάγοντας τις σταθμισμένες συνιστώσες (5.49) η προσέγγιση γίνεται

$$\sum_{k=1}^K t_k = \sum_k \frac{(\alpha_{uk}^2 + \alpha_{vk}^2) \tilde{f}_k + \alpha_{vk}^2 u_k + \alpha_{uk}^2 v_k}{\sum_k \alpha_k^2} \quad (5.54)$$

$$= \sum_k (\ell_k + \mu_k) \tilde{f}_k + \mu_k u_k + \ell_k v_k, \quad \ell_k = \frac{\alpha_{uk}^2}{\sum_k \alpha_k^2}, \quad \mu_k = \frac{\alpha_{vk}^2}{\sum_k \alpha_k^2}. \quad (5.55)$$

Επομένως, για κάθε  $k$  η συνιστώσα που ανακατασκευάζεται είναι της μορφής  $t_k(x, y) = (\ell_k + \mu_k)\tilde{f}_k(x, y) + \mu_k u_k(x, y) + \ell_k v_k(x, y)$ , με τα  $\ell_k, \mu_k$  χωρικά προσαρμοστικές σταθερές που εκφράζουν την «ένταση» σε υφή ή ακμές στη ζώνη  $k$ ,  $\ell_k + \mu_k = \alpha_k^2 / \sum_k \alpha_k^2$  το βάρος της κάθε στενοζωνικής συνιστώσας και  $\ell_k/\mu_k = \lambda_k^2$ .

Τέλος, ας σημειωθεί ότι η στάθμιση με τα πλάτη θυμίζει τον υπολογισμό της στιγμιαίας συχνότητας  $\omega_i = \phi'$  για σήματα  $f = ae^{j\phi} = \sum_k f_k$  πολλαπλών συνιστωσών

$$\omega_i = \sum_k \frac{\alpha_k^2 \omega_{ik}}{\sum_k \alpha_k^2}, \quad (5.56)$$

από τη μέση τιμή των σταθμισμένων με τα πλάτη ακ συχνοτήτων  $\omega_{ik}$  της κάθε συνιστώσας  $f_k$  [140, 307]. Σε αυτή την περίπτωση η εκτίμηση της στιγμιαίας συχνότητας είναι πιο ομαλή, χωρίς τις μη-γραμμικότητες που εισάγονται από τον υπολογισμό της από τη ροπή πρώτης τάξης ως προς τη συχνότητα της αναπαράστασης χώρου-συχνότητας [64, 135]. Η παράγωγος της φάσης, δηλαδή η στιγμιαία συχνότητα, είναι ακριβώς ίση με αυτή την ποσότητα όταν οι συνιστώσες είναι συμμετρικές.

Σε αντιδιαστολή με την διατήρηση των λεπτότερων κλιμάκων (5.40) (φασματική επιλογή συνιστωσών), η στάθμιση εύρους ενισχύει τις συνιστώσες υφής σε κάθε σημείο εικόνας, χωρίς την ανάγκη για κάποιο κατώφλι. Επιτελεί μια αυτόματη ενεργειακή επιλογή, «καταπιέζοντας» τις αδύναμες τοπικά ταλαντευόμενες συνιστώσες. Περαιτέρω, τα ομαλοποιημένα βάρη  $\alpha_k^2 / \sum_k \alpha_k^2$  προσεγγίζουν την ανάλυση κυρίαρχων συνιστωσών (3.13), εφόσον εάν  $\forall(x,y) \exists k : \alpha_k(x,y) \gg \alpha_{k'}, \forall k' \neq k$  τότε η  $t_k$  είναι η κυρίαρχη συνιστώσα υφής.

#### 5.4.1.3 Κυρίαρχες συνιστώσες χαμηλότερης τάξης

Έστω η σημειακή αναδιάταξη των συνιστωσών μιας εικόνας έτσι ώστε να προκύπτει μια ιεραρχική διάταξη συνιστωσών, στην οποία αύξηση της τάξης υποδηλώνει λιγότερο κυρίαρχη ή σημαντική συνιστώσα (Εν. ;;). Θεωρούμε τότε την ανακατασκευή

$$\sum_{k=1}^K t_k(x, y) \triangleq \sum_{k=1}^D d_k(x, y), \quad (5.57)$$

όπου  $D < K$  είναι οι περισσότερο σημαντικές πρώτες συνιστώσες στο χώρο κυρίαρχων συνιστωσών. Η ανάλυση DCA εφαρμόζεται σε αυτό το χώρο όπου οι προκύπτουσες συνιστώσες θεωρούνται καθολικά απομονωμένες σε κάθε σημείο μέσω της συστοιχίας φίλτρων. Έστω η 2-Δ διανυσματική συνάρτηση  $\mathbf{F} = [f_1, \dots, f_K]$  στον  $K$ -διάστατο χώρο των  $N \times 1$  αποκρίσεων  $f_k$ , δηλαδή ο πίνακας κάθε στήλη του οποίου είναι το διάνυσμα-απόκριση του φίλτρου  $k$  στην εικόνα  $f$ . Αλγεβρικά η απεικόνιση  $\mathbf{F} \mapsto \mathbf{F}_D$  από τον χώρο στενοζωνικών στο χώρο κυρίαρχων, τοπικά στενοζωνικών συνιστωσών δίνεται από μια σχέση  $\mathbf{F}_D = \mathbf{F} \cdot \mathbf{W}$ , όπου  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_K]$  ο  $K \times K$  πίνακας μετασχηματισμού (transformation matrix).

Ο μετασχηματισμός ορίζεται από τις διατεταγμένες συνιστώσες  $\mathbf{w}_k$  του πίνακα που αποτελούν συναρτήσεις διανυσματικών τιμών. Κάθε  $\mathbf{w}_k : \{w_{k1}, \dots, w_{kN}\}$  μπορεί να πάρει  $N$  τιμές που αντιστοιχούν σε μοναδιαία διανύσματα διάστασης  $K$ . Οι τιμές του πρώτου, κυρίαρχου διανύσματος  $\mathbf{w}_1$  ορίζονται από τη μεγιστοποίηση

$$w_{1i} = \arg \max_{\|\mathbf{w}_1\|=1} \{\Gamma(\tilde{f}_i)\mathbf{w}_1\}, \quad \forall i \in [1, N], \quad (5.58)$$

όπου  $\tilde{f}_i$  τα διανύσματα-γραμμές του πίνακα  $\mathbf{F}$ , δηλαδή  $\mathbf{F} = [f_1, \dots, f_K] = [\tilde{f}_1, \dots, \tilde{f}_N]^T$  και  $\Gamma(\cdot)$  είναι μια συνάρτηση κόστους που ταυτίζεται με το κριτήριο που μεγιστοποιείται στο σύνολο των αποκρίσεων των φίλτρων (πλάτος ή απόκριση του ενεργειακού τελεστή, Εν. 3.3).

Η κυρίαρχη συνιστώσα είναι τότε

$$d_1 = [\tilde{f}_1 w_{11}, \dots, \tilde{f}_N w_{1N}], \quad (5.59)$$

ενώ οι συνιστώσες ανώτερης τάξης δίνονται από

$$d_k = [\tilde{f}_1 w_{k1}, \dots, \tilde{f}_N w_{kN}], \quad (5.60)$$

όπου τα  $w_{ki}$  ορίζονται από τη μεγιστοποίηση στο σημείο  $i$  των δεδομένων που προκύπτουν αφαιρώντας τα προηγούμενα  $k - 1$  μέγιστα  $\hat{f}_i = \tilde{f}_i - \sum_{c=1}^{k-1} w_{ci}^T \tilde{f}_c w_{ci}$ , δηλαδή

$$w_{ki} = \arg \max_{\|\mathbf{w}_k\|=1} \{\Gamma(\hat{f}_i)\mathbf{w}_k\}, \quad \hat{f}_i = \tilde{f}_i - \sum_{c=1}^{k-1} w_{ci}^T \tilde{f}_c w_{ci}, \quad \forall i \in [1, N]. \quad (5.61)$$

Ανάμεσα στις συνιστώσες προκύπτει αυτόματα μια σχέση διάταξης καθώς  $\Gamma(d_{k+1}(\mathbf{x})) < \Gamma(d_k(\mathbf{x}))$ ,  $\forall \mathbf{x}$ . Η διάταξη αυτή είναι παρόμοια με την PCA προβολή [84], με τις κυρίαρχες συνιστώσες να είναι οι ιδιοτιμές του μετασχηματισμού  $\mathbf{W}$  με ιδιοδιανύσματα  $\mathbf{w}_k$ . Ενώ το κριτήριο διάταξης είναι μη-γραμμικό, η προβολή  $\mathbf{F}_D = \mathbf{F} \cdot \mathbf{W}$  είναι γραμμική σε μια βάση

που αναπαριστά βέλτιστα τα δεδομένα  $\mathbf{F}$  ως προς την ενέργεια της απόκρισης τους σε κάποιο τελεστή  $\Gamma$ .

Η διάσταση του χώρου χαρακτηριστικών των φίλτρων μειώνεται με τη διατήρηση των πρώτων διατεταγμένων συνιστωσών και την απόρριψη των συνιστωσών ανώτερης τάξης. Οι κυρίαρχες συνιστώσες χαμηλότερης τάξης είναι οι περισσότερο σημαντικές για την αναπαράσταση της πληροφορίας σύμφωνα με την επιλογή της συνάρτησης κόστους  $\Gamma$  (το πλάτος περιβάλλουσας ή την ενέργεια διαμόρφωσης των συνιστωσών). Ένας τρόπος επιλογής των συνιστωσών αυτών είναι η απόρριψη των συνιστωσών που συνεισφέρουν λιγότερο στη συνολική μεταβλητότητα ή τυπική απόκλιση της εικόνας  $f$ , δηλαδή

$$\sum_{k=1}^K t_k(x, y) \triangleq \sum_{k=1}^D d_k(x, y), \quad D : \left\{ \frac{\text{Var}\{d_k\}}{\text{Var}\{f\}} > \epsilon, \forall k < D \right\}, \quad (5.62)$$

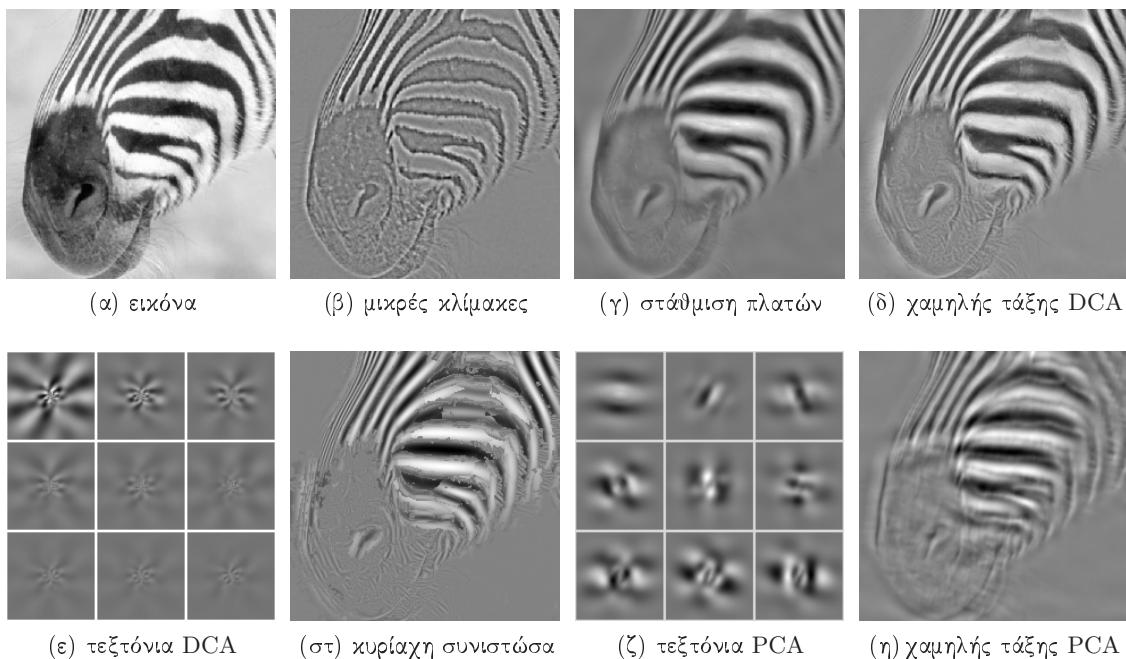
όπου  $\epsilon$  μια σταθερά κοντά στη μονάδα, (π.χ.  $\epsilon > 0.8$ ). Μια ενδιαφέρουσα παρατήρηση είναι ότι χρησιμοποιώντας την ενέργεια διαμόρφωσης (μέσω του ενεργειακού τελεστή) για το  $\Gamma$ , ο αριθμός των συνιστωσών που συνεισφέρουν περισσότερο στη μεταβλητότητα είναι πολύ μικρότερος από ότι με το κριτήριο πλάτους. Αυτό σημαίνει ότι χρειάζονται λιγότερες συνιστώσες για την αναπαράσταση της εικόνας και ακόμη και στην πρώτη, κυρίαρχη συνιστώσα του DCA η πληροφορία που διατηρείται είναι περισσότερη.

Μια τέτοια ανάλυση ιεραρχικών συνιστωσών (HCA) καταλήγει σε μια αναπαράσταση μειωμένης διάστασης, σε σχέση με την αναπαράσταση από το σύνολο των στενοζωνικών συνιστωσών, αλλά ταυτόχρονα καλύτερης προσέγγισης και ανακατασκευής του  $f$  σε σύγκριση με το DCA [92, 125], που διατηρεί μόνο την κυρίαρχη, πρώτη συνιστώσα  $d_1$ . Μία σύγκριση των δύο σχημάτων παρουσιάζεται για ανακατασκευή υφής στο Σχ. 5.8 (δ), (στ). Στο (η) παρουσιάζεται η ανακατασκευή από τις πρώτες  $D$  συνιστώσες της προβολής στην PCA βάση του χώρου των φιλτραρισμένων διανυσμάτων.

#### 5.4.2 Τεξτόνια κυρίαρχων διαμορφώσεων

Τα δομικά στοιχεία της υφής, οι στοιχειώδεις σχηματισμοί που συντελούν στο σχηματισμό των προτύπων υφής και ονομάστηκαν τεξτόνια, ορίστηκαν εξαρχής με γεωμετρικούς όρους (π.χ. blobs, crosses, junctions, bars) [143, 146, 303, 321]. Η αναπαράσταση μέσω των αποκρίσεων της συστοιχίας και στενοζωνικών συνιστωσών χρησιμοποιήθηκε για ένα λειτουργικό ορισμό των τεξτονίων από ένα σύνολο «πρωτότυπων» αποκρίσεων [178, 186, 296]. Αυτά τα πρωτότυπα, αποτέλεσμα ομαδοποίησης σε τάξεις (clustering) στο χώρο των διανυσμάτων απόκρισης, αντιστοιχούν στις κυρίαρχες τοπικές δομές στην εικόνα (patches).

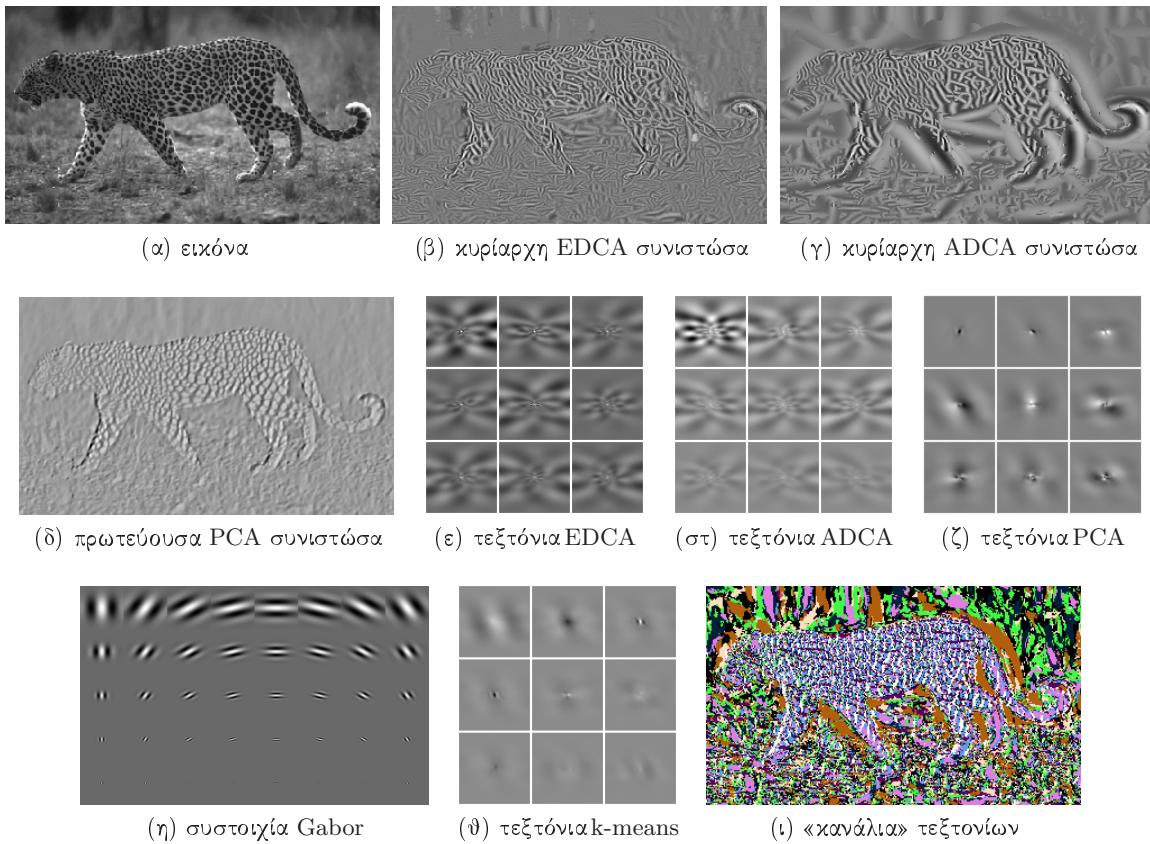
Με την ίδια λογική, ο ρόλος της ομαδοποίησης για τη δημιουργία πρωτότυπων μπορεί να αντικατασταθεί από τις κυρίαρχες συνιστώσες (5.59, 5.60). Συγκεκριμένα, τα κυρίαρχα διανύσματα του μετασχηματισμού  $\mathbf{W}$  της απεικόνισης  $\mathbf{F}_D = \mathbf{F} \cdot \mathbf{W}$ , μπορούν να θεωρηθούν τεξτόνια κυρίαρχων, τοπικά στενοζωνικών AM-FM συνιστωσών. Τα διανύσματα  $\mathbf{W} = [\{\mathbf{w}_k\}_{k=1 \dots D}]$  μπορούν να ανακτηθούν από τις στήλες του πίνακα χρησιμοποιώντας τον ψευδο-αντίστροφο  $\mathbf{F}^{-1}$  του πίνακα αποκρίσεων των φίλτρων [139] και την απεικόνιση  $\mathbf{W} = \mathbf{F}^{-1} \cdot \mathbf{F}_D$ .



Σχήμα 5.8: Επιλογή συνιστώσων και ανακατασκευή υφής από συστοιχία 40 φίλτρων (πέντε κλίμακες, οκτώ προσανατολισμοί). πάνω σειρά: (α) εικόνα και η ανακατασκευή (υφής) από (β) λεπτότερες (τρεις) κλίμακες ( $K = 24$ ), (γ) στάθμιση πλατών ( $K = 40$ ), (δ) κυρίαρχες συνιστώσες χαμηλότερης τάξης ( $D=9$ ). κάτω σειρά: (ε) κυρίαρχες συνιστώσες ανακατασκευασμένες ως τοπικά φίλτρα, (σ) κυρίαρχη (ενεργειακά) συνιστώσα, (ζ) πρωτεύουσες συνιστώσες μέσω PCA ως τοπικά φίλτρα, (η) ανακατασκευή από τις  $D$  πρώτες πρωτεύουσες συνιστώσες PCA .

Τα  $w_k$  είναι πρωτότυπα φιλτραρισμένων συνιστωσών της εικόνας και η πληροφορία που χωδικοποιούν, δηλαδή οι αντίστοιχες πρωτότυπες αποχρίσεις των φίλτρων μπορούν να οπτικοποιηθούν με αντίστροφη προβολή (back-projection) στο χώρο των φίλτρων της συστοιχίας. Αν  $\mathbf{G} = [\{g'_k\}_{k=1 \dots K}]$  είναι ο συνδυασμένος (concatenated) πίνακας των αποχρίσεων, όπου  $g'_k$  το διάνυσμα-απόχριση του  $g_k$ , τότε οι δομές που αντιστοιχούν στις  $D$  κυρίαρχες συνιστώσες ανακατασκευάζονται με όρους ελαχίστων τετραγώνων από τα  $p_k = w_k \cdot \mathbf{G}^{-1}$ , όπου  $\mathbf{G}^{-1}$  ο ψευδο-αντίστροφος πίνακας του  $\mathbf{G}$ . Στο Σχ. 5.8 (ε) τα ανακτημένα φίλτρα/δομές αντιστοιχούν, κατά ιεραρχική σειρά σημαντικότητας, στις πρώτες  $D=9$  συνιστώσες που χρησιμοποιήθηκαν για την ανακατασκευή της εικόνας στο (δ).

Τα τεξτόνια της αναπαράστασης με κυρίαρχες συνιστώσες, είναι σχεδόν αφηρημένες κυματοειδείς δομές, οι οποίες δεν αντιστοιχούν ή δεν απεικονίζονται οπτικά σε ευδιάκριτα τοπικά χαρακτηριστικά της εικόνας. Τα τεξτόνια των PCA συνιστωσών, με αντίστοιχο back-projection φαίνονται στο Σχ. 5.8 (ζ) για τις συνιστώσες που συγχροτούν την εικόνα (η). Σε αντίθεση με τις κυρίαρχες, τοπικά στενής ζώνης, συνιστώσες του DCA οι δομές εδώ αντιστοιχούν στις περισσότερο ανεξάρτητες, ολικά στενοζωνικές συνιστώσες. Είναι σαφώς περισσότερο δομημένες και ταυτίζονται ευκολότερα με χαρακτηριστικά της εικόνας (π.χ. κυματώσεις, ακμές υφής, bars). Ο ορισμός αυτός για τα δομικά στοιχεία της υφής με βάση κρίαρχες συνιστώσες μπορεί να χρησιμοποιηθεί στη συγχρότηση λεξικών τεξτονίων και την κατασκευή μοντέλων υφής με βάση τη συχνότητα εμφάνισης σε δεδομένα εκπαίδευσης [296].

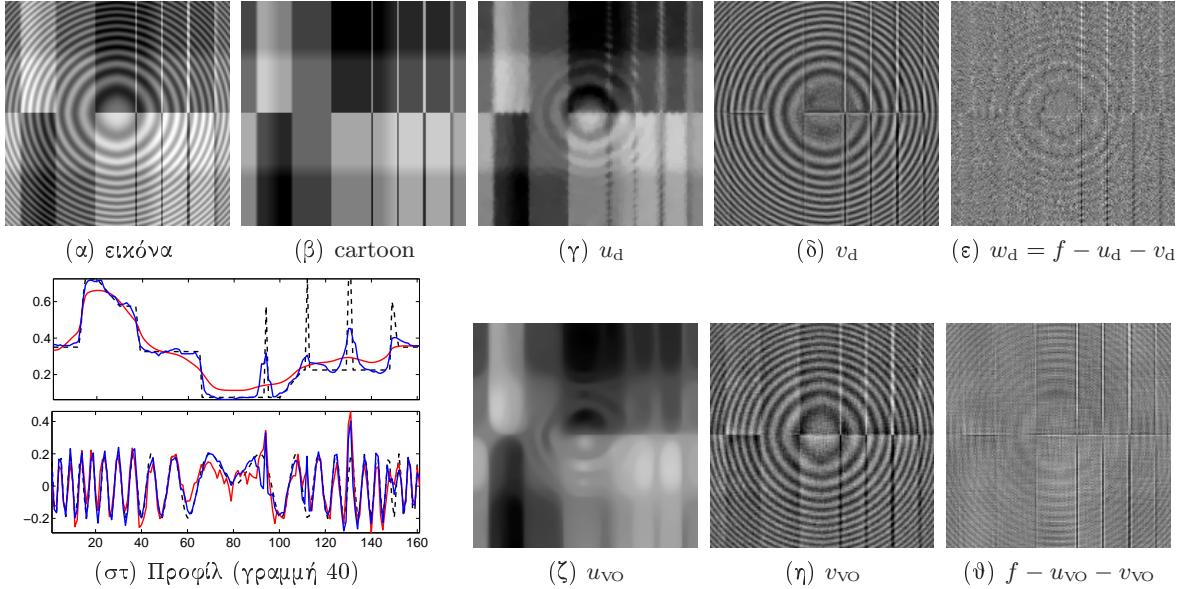


Σχήμα 5.9: Τεξτόνια από τις αποκρίσεις των φίλτρων της συστοιχίας. (α),(β),(γ) εικόνα και κυρίαρχες συνιστώσες, κατά ενέργεια (EDCA) και πλάτος (ADCA), (δ) πρωτεύουσα συνιστώσα μέσω PCA , (ε),(στ),(ζ) τεξτόνια (textons) από τα κυρίαρχα διανύσματα DCA και PCA που αντιστοιχούν σε τοπικές δομές της εικόνας, (η) αποκρίσεις των 40 φίλτρων Gabor της συστοιχίας (δεν συμπεριλαμβάνονται το βαθυπερατό ή τα υψηπερατά κανάλια) κατά κλίμακα και προσανατολισμό, (θ) τεξτόνια μέσω k-means ομαδοποίησης στο χώρο των αποκρίσεων της συστοιχίας στην εικόνα (διανύσματα απόκρισης), (ι) «κανάλια» ή απεικόνιση τεξτονίων-ανά-σημείο από την ομαδοποίηση, όπου κάθε σημείο ανατίθεται σε μια από τις δομές του (θ).

#### 5.4.3 Παραδείγματα και συγκρίσεις αποδόμησης

Η ακρίβεια των σχημάτων αποσύνθεσης εικόνων είναι δύσκολο να εκτιμηθεί πειραματικά, καθώς οποιαδήποτε μέθοδος αντιμετωπίζει τα ίδια προβλήματα σχετικά με την εκτίμηση των αποτελεσμάτων, όπως π.χ. ποσοτικά μετρικά, κοινό πλαίσιο για συγκρίσεις, ρύθμιση παραμέτρων, αντικειμενικά χριτήρια αρτιότητας, δοκιμές επιδόσεων (benchmark evaluations) [207] κ.ά. Τα αποτελέσματα που αναφέρονται εδώ αντιστοιχούν στην υλοποίηση του σχήματος (5.21, 5.22) με τις προτεινόμενες μεθόδους ανακατασκευής της υφής, δηλαδή καθορισμό του  $\sum_k t_k$  που περιγράφονται στην Εν. 5.4.1. Τα διαφορετικά σχήματα, ανάλογα με τον όρο ανακατασκευής, θα συμβολίζονται με  $(u_p, v_p)$  για το μερικό (partial) άθροισμα στις υψηλότερες συχνότητες (5.40),  $(u_w, v_w)$  για τη στάθμιση (weighting) πλατών (5.43) και  $(u_d, v_d)$  για τις κυρίαρχες (dominant) συνιστώσες (5.57). Οι συγκρίσεις παρουσιάζονται ως προς το σχήμα αναφοράς Vese-Osher (VO) [298] που ορίζεται από την (5.11).

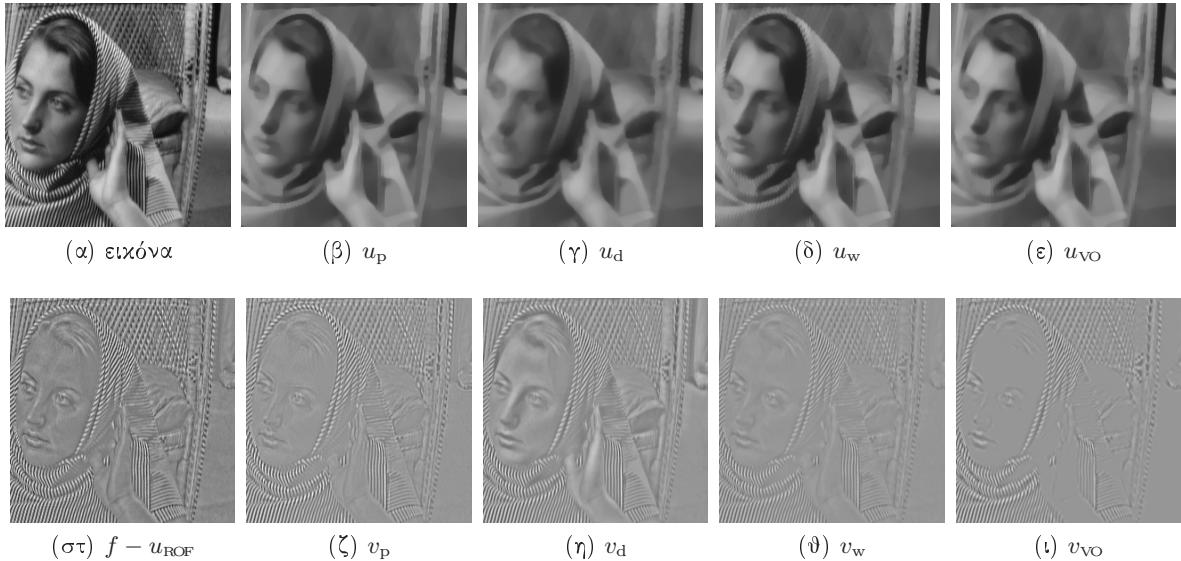
Μια συνθετική εικόνα, δύσκολη να αποδομηθεί αντιληπτικά, παρουσιάζεται στο Σχ. 5.10. Η συνιστώσα γεωμετρικής δομής (β) περιέχει διάφορα είδη ακμών (steps, lines, ramps) [205,



Σχήμα 5.10: Αποδόμηση συνθετικής εικόνας, που περιλαμβάνει δομές (ακμές steps, lines, ramps), υφή (χωρική ταλάντωση chirp) και λευκό θόρυβο (20dB). (α) «καθαρή» εικόνα με υφή ( $u + v$ ), (β) cartoon αναφοράς. Οι συνιστώσες και ο παραμένων θόρυβος με χρήση (γ)-(ε) του νέου σχήματος και ανακατασκευή από τις κυρίαρχες συνιστώσες με  $(\lambda, \mu) = (10, 100)$  και (η)-(ι) του αλγορίθμου Vese-Osher της (5.11) με  $(\lambda_{vo}, \mu_{vo}) = (0.75, 0.001)$ . (στ) Προφίλ της εικόνας για μια σειρά (40) φαίνονται για τις συνιστώσες αναφοράς (διακεκομμένες) και τις ανακτημένες συνιστώσες (μπλε για το προτεινόμενο, κόκκινο για το VO). Οι παράμετροι των μοντέλων επιλέχθηκαν έτσι ώστε τα δύο σχήματα να δίνουν παρόμοιες  $L_2$  νόρμες υφής, δηλ.  $\|v_{vo}\|_{L^2} = \|v_d\|_{L^2}$ . Οι τιμές των εικόνων παρουσιάζονται σε πλήρη κλίμακα του γκρίζου για καλύτερη οπτικοποίηση.

206, 238, 239] και περιοχές επίπεδης φωτεινότητας για απλότητα. Η συνιστώσα υφής είναι ένα χωρικό chirp σήμα σταθερού πλάτους με χωρική συχνότητα που μεταβάλλεται από  $\pi/80$  μέχρι  $\pi/2$  και προσθετικό λευκό θόρυβο 20dB ώστε η εικόνα να προσομοιώνει ένα  $u+v+w$  μοντέλο. Αν εξαιρέσει κανείς την πληροφορία ως προς το περιεχόμενο της εικόνας που καθορίζεται από την υπέρθεση των συνθετικών συνιστωσών αναφοράς, η εικόνα (α) είναι δύσκολο να διαχωριστεί οπτικά σε υφή και cartoon. Για παράδειγμα οι κεντρικές, μεγάλης κλίμακας ταλαντώσεις του chirp θα μπορούσαν να καταταχθούν στο cartoon, ενώ οι κάθετες γραμμές, οι οποίες εξ' ορισμού ανήκουν στις ακμές της εικόνας θα μπορούσαν να θεωρηθούν υφή (ταλάντωση δύο σημείων). Έτσι, μια επιπλέον πτυχή της σύγκρισης είναι να διαπιστωθεί ποιες κλίμακες και χαρακτηριστικά εντάσσει ο κάθε αλγόριθμος στις τρεις συνιστώσες (υφή, μακροδομή, θόρυβο).

Η προτεινόμενη αποδόμηση με ανακατασκευή μέσω των κυρίαρχων συνιστωσών υλοποιήθηκε με παραμέτρους  $(\lambda, \mu) = (10, 100)$ , έτσι ώστε να διατηρηθούν οι μικρής κλίμακας ακμές γραμμών στο (γ). Όλες οι κατηγορίες ακμών διατηρούνται και εντοπίζονται στο cartoon μαζί με τις χαμηλόσυχνες μεταβολές του chirp και τις ταλαντώσεις που υπερτίθενται στις κάθετες ακμές. Αντιληπτικά, ωστόσο, το κεντρικό blob και οι πρώτοι ομόκεντροι διακτύλιοι της υφής chirp ερμηνεύονται πιο εύκολα ως τμήμα της συνιστώσας μακροδομής, ενώ τα υπολείμματα των κάθετων ταλαντώσεων «φαίνονται» στο μοντέλο ως ακμές μικρής κλίμακας. Οι παράμετροι του αλγορίθμου αναφοράς VO ρυθμίστηκαν έτσι ώστε να επιτευχθούν οι ίδιες  $\|v\|_{L^2}$  νόρμες,



Σχήμα 5.11: Εναλλακτικοί όροι ανακατασκευής της υφής για αποδόμηση με το  $u + Kv$  σχήμα. πάνω σειρά: cartoon συνιστώσες της εικόνας (α) με το προτεινόμενο σχήμα και εναλλακτικά αθροισματα ανακατασκευής  $\sum_k t_k$  στα (β)-(δ) και το σχήμα Vese-Osher στο (ε). κάτω σειρά: παραμένον του σχήματος Rudin-Osher-Fatemi (στ) και συνιστώσες υφής των αντίστοιχων cartoon. Το  $(u_p, v_p)$  αντιστοιχεί στο μερικό στις ανώτερες συχνότητες (5.40), το  $(u_w, v_w)$  στο σταθμισμένο με τα πλάτη (5.43) και το  $(u_d, v_d)$  στο άθροισμα κυρίαρχων, χαμηλότερης τάξης (5.57). Οι παράμετροι για το  $u + Kv$  σχήμα είναι  $(\lambda, \mu) = (10, 5)$  και για το VO  $(\lambda_{vo}, \mu_{vo}) = (5, 0.1)$ .

δηλαδή το ίδιο πόσο πληροφορίας στο  $v$  [7]. Με το συγκεκριμένο σύνολο παραμέτρων το cartoon (ζ) είναι θολό με ακμές «ράμπες» και γραμμές διαταραγμένες και περισσότερο ορατές στο (θ). Επίσης το προτεινόμενο μοντέλο αποδόμησης διαχωρίζει το θόρυβο στο παραμένον (ε), ενώ το VO σχήμα των αναμειγνύει με τη συνιστώσα της υφής.

Μερικά απλά μετρικά που χρησιμοποιούνται για ποσοτικές συγκρίσεις είναι ο cartoon-to-noise σηματοθορυβικός λόγος  $SNR = \|u - \bar{u}\|_{L^2}^2 / \|w\|_{L^2}^2$ , και οι Ευκλείδεις νόρμες των συνιστώσων (χανονικοποιημένες % ως προς την αναφορά) όπως φαίνονται ενδεικτικά στον Πίν. 5.1. Τα rms σφάλματα ανάμεσα στις συνιστώσες αναφοράς και τις εκτιμώμενες από τα δύο σχήματα είναι περίπου μηδενικά και για τις τρεις συνιστώσες της εικόνας.

	SNR	$\ u\ _{L^2}$	$\ v\ _{L^2}$	$\ f - u - v\ _{L^2}$
αναφορά	14.96	100	100	100
$u + v$ (VO)	12.96	96.6	100	125.7
$u + Kv$	17.16	98.1	100	77.5

Πίνακας 5.1: Ποσοτική σύγκριση αποδόμησης για τη συνθετική εικόνα 5.10 (α).

Η λεπτομέρεια από την εικόνα ‘barb’ που παρουσιάζεται στο Σχ. 5.11 (α) χαρακτηρίζεται από προσανατολισμένες, υψησυχνες ταλαντώσεις υφής, οι οποίες εντοπίζονται και εντάσσονται στη συνιστώσα υφής από όλα τα εναλλακτικά σχήματα ανακατασκευής. Ωστόσο, η έκταση των υπολειμμάτων δομής στο  $v$  και υφής στο  $u$  διαφέρουν. Η καλύτερη οπτικά αποδόμηση επιτυγχάνεται από το σχήμα της μερικής ανακατασκευής, το οποίο οδηγεί σε ένα ομαλό cartoon (β) με οξείες μεταβάσεις και καλά εντοπισμένες ακμές και μια ταλαντευόμενη, μικρής

κλίμακας υφή (ζ). Το σχήμα με ανακατασκευή μέσω κυρίαρχων συνιστωσών διατηρεί κάποιες δομές μεγάλης φωτεινότητας και ακμές στο  $u$  ενώ το σταθμισμένο σχήμα (δ) κάποια ορατά στοιχεία υφής. Ωστόσο, και τα δύο λαμβάνονται χωρίς την ανάγκη για ευρετικές επιλογές (όπως οι κλίμακες υφής) και συμπεριφέρονται ανώτερα σε εικόνες με περισσότερο σύνθετη υφή (μείζεις συχνοτήτων, έντασης και προσανατολισμού).

## 5.5 Μιγαδική αποδόμηση

Μια ενδιαφέρουσα επέκταση είναι η μιγαδική αποδόμηση ή αποδόμηση μιγαδικών εικόνων, δηλαδή ο  $f = u + v$  διαχωρισμός με την αρχική εικόνα να είναι μια συνάρτηση μιγαδικών τιμών  $f : \mathbb{R}^2 \rightarrow \mathbb{C}$ . Παρόμοιες ιδέες έχουν εμφανιστεί και για εξισώσεις διάχυσης με μιγαδικούς συντελεστές [107] ή αποκατάσταση και αποδόμηση διανυσματικών εικόνων [12, 157, 286]. Σε πρακτικές περιπτώσεις ανάγκη για μιγαδική αποδόμηση μπορεί να προκύψει αν τα διαθέσιμα δεδομένα είναι αναλυτικά ή εμβυθισμένα σε μιγαδικό ύδρυβο. Η ειδική περίπτωση που προκύπτει εδώ είναι λόγω των μιγαδικών αποκρίσεων των Gabor φίλτρων που αναδεικνύουν τις υποσυνιστώσες του  $v$ . Συγκεκριμένα, αν  $t_k \in \mathbb{C}$  στο άνθροισμα ανακατασκευής της (5.16), κάθε μιγαδική συνιστώσα μπορεί να γραφεί ως

$$t_k \approx t * g_k = \operatorname{Re}\{t_k\} + j\operatorname{Im}\{t_k\} = t_{kr} + jt_{ki}, \quad (5.63)$$

όπου  $\operatorname{Re}\{\cdot\}$  και  $\operatorname{Im}\{\cdot\}$  συμβολίζουν το πραγματικό και φανταστικό μέρος αντίστοιχα. Αν  $t_k \in \mathbb{R}$  και  $g_k = g_{ke} + jg_{ko}$  η χρονιστική απόκριση του μιγαδικού Gabor (2.1), τότε  $t_{kr} = t_k * g_{ke}$  είναι άρτια και  $t_{ki} = t_k * g_{ko}$  η περιττή φιλτραρισμένη συνιστώσα.

Η μοντελοποίηση, ανακατασκευή και αποδόμηση μπορεί να πραγματοποιηθεί χρησιμοποιώντας μεμονωμένα και ανεξάρτητα την άρτια ή την περιττή συνιστώσα. *Καναλικά* (channel-wise) σχήματα διάχυσης με ροές Beltrami χρησιμοποιώντας μιγαδικά Gabor wavelets χρησιμοποιήθηκαν για βελτίωση (enhancement) υφής [157]. Ελλείψει πρακτικών επιχειρημάτων υπέρ των συνιστωσών που εξάγονται από το άρτιο ή το περιττό τμήμα των συμμετρικών φίλτρων, επιλέγεται για την πληρότητα της θεώρησης να συμπεριληφθεί πληροφορία και από τα δύο κανάλια. Επιπρόσθετα, ένα συζευγμένο σχήμα επωφελείται από τα πλεονεκτήματα των ανιχνευτών χαρακτηριστικών φίλτρων τετραγωνισμού φάσης (π.χ. ανίχνευση και εντοπισμός με βάση τη φάση ή την ενέργεια [219, 220]).

Πραγματικά δεδομένα  $f \in \mathbb{R}$  καθορίζουν εκ των προτέρων ότι  $u, v \in \mathbb{R}$ . Ο περιορισμός αυτός καταλήγει στο απλοϊκό αποτέλεσμα ότι η φανταστική συνιστώσα, δηλαδή η περιττή συνιστώσα του φίλτρου, δε συνεισφέρει στις Euler-Lagrange (5.18, 5.19) της (5.16) για μιγαδικά  $t_k$ . Το ίδιο ισχύει και για απόλυτα φανταστικά δεδομένα. Έστω τώρα μια μιγαδική εικόνα  $f \in \mathbb{C}$ , η οποία θεωρείται μια διανυσματική συνάρτηση  $\mathbf{f} \in \mathbb{R}^2$  τέτοια ώστε  $\{f = f_r + jf_i : \mathbf{f} = [f_r, f_i]\}$ . Κάθε φίλτρο τότε είναι το ζευγάρι ορθογωνισμού  $\mathbf{g}_k = [g_{ke}, g_{ko}]$ , ενώ ένα μοντέλο μιγαδικού διαχωρισμού θα απαιτούσε  $\mathbf{f} = \mathbf{u} + \mathbf{v} = [u_r, u_i] + [v_r, v_i]$ . Αναπτύσσεται εδώ ένα σχήμα συζευγμένης αποδόμησης (coupled decomposition) όπου ο διαχωρισμός στο ένα κανάλι, πραγματικό ή φανταστικό, επηρεάζει τη ροή (flux) στο άλλο.

### 5.5.1 Μιγαδική νόρμα ολικής μεταβολής

Το πρόβλημα ελαχιστοποίησης (5.16) τροποποιείται για διανυσματικές συναρτήσεις

$$E(\mathbf{u}, \mathbf{v}) = J_j(\mathbf{u}) + \lambda \|\mathbf{f} - \mathbf{u} - \mathbf{v}\|_{L^2}^2 + \mu \|\mathbf{v} - \sum_k \mathbf{t}_k\|_{L^2}^2, \quad (5.64)$$

όπου  $J_j(\mathbf{u})$  είναι ένας ορισμός της TV νόρμας για μιγαδικές συναρτήσεις σε διανυσματική αναπαράσταση και οι  $\|\cdot\|_{L^2}$  είναι νόρμες σε διανύσματα (vector norms).

**Ορισμός 5.1** *Μιγαδική νόρμα ολικής μεταβολής (complex Total-Variation norm, CTV): Για κάθε  $u = u_r + ju_i : \Omega \rightarrow \mathbb{C}$  ορίζεται η μιγαδική TV νόρμα για τη διανυσματική συνάρτηση  $\mathbf{u} = [u_r, u_i] : \Omega \rightarrow \mathbb{R}^2$ , ως η  $L^1$  του διανυσματικού μέτρου του gradient*

$$J_j(\mathbf{u}) = J_j(u) \triangleq \int_{\Omega} \left( \|\nabla u_r\|^2 + \|\nabla u_i\|^2 \right)^{1/2}. \quad (5.65)$$

Η μιγαδική νόρμα είναι περιστροφικά αμετάβλητη (rotational invariant), επιτρέπει ασυνέχειες και ακμές ενώ για αμιγώς πραγματικά ή φανταστικά δεδομένα ανάγεται στην απλή TV (5.4). Μπορεί να ερμηνευθεί από τη σύνθεση γενικευμένων νορμών  $J_i(\mathbf{u}) \triangleq [L^1(\Omega) \circ l^2(i) \circ l^2(j)](\mathbf{u})$  [33], όπου το  $i$  αναφέρεται στο κανάλι και το  $j$  στη χωρική διάσταση.

Μια αντίστοιχη νόρμα, η διανυσματική διακριτή TV, ορίστηκε για διανυσματικές διακριτές εικόνες [54] με πεπερασμένες διαφορές για τα gradient και χρησιμοποιήθηκε για αποκατάσταση και αποδόμηση έγχρωμων εικόνων [12, 261]. Εναλλακτικά για τη μιγαδική νόρμα όταν μπορούσε να χρησιμοποιηθεί ο ορισμός της χρωματικής TV (color TV) για διανυσματικές εικόνες  $J_c(\mathbf{u}) = (J(u_r)^2 + J(u_i)^2)^{1/2}$ , δηλαδή η Ευκλειδια νόρμα των TV για τις μεμονωμένες συνιστώσες [33].

Σε ένα πλαίσιο Riemannian γεωμετρίας μπορεί να δειχθεί ότι εξάγεται από το ίχνος (trace) του πίνακα μεταβολής  $G$  της πρώτης θεμελιώδους μορφής<sup>2</sup> (first fundamental form)

$$d\mathbf{u}^2 = \begin{bmatrix} dx & dy \end{bmatrix} G \begin{bmatrix} dx \\ dy \end{bmatrix}, \quad G = \nabla u_r \nabla u_r^T + \nabla u_i \nabla u_i^T. \quad (5.66)$$

Το ίχνος του τετραγωνικού πίνακα  $G$  ισούται με το άθροισμα των ιδιοτιμών του, δηλαδή

$$\text{trace}[G] = \|\nabla u_r\|^2 + \|\nabla u_i\|^2 = \lambda^+ + \lambda^-, \quad (5.67)$$

όπου  $(\lambda^+, \lambda^-)$  είναι η μεγαλύτερη και η μικρότερη ιδιοτιμή αντίστοιχα [78, 261, 285]. Επομένως,

$$J_j(\mathbf{u}) = \int_{\Omega} \sqrt{\lambda^+ + \lambda^-} \quad (5.68)$$

με τις κατευθύνσεις των ιδιοδιανυσμάτων των  $\lambda^+$  και  $\lambda^-$ , να είναι οι κατευθύνσεις μεγαλύτερης και μικρότερης αλλαγής του  $\mathbf{u}$ . Με ορισμένη μια τέτοια διανυσματική γεωμετρία της εικόνας η  $J_j(\mathbf{u})$  κανονικοποιεί το cartoon διατηρώντας ακμές και γωνίες που εντοπίζονται σε οποιοδήποτε από τα δύο κανάλια.

<sup>2</sup>Η πρώτη θεμελιώδης μορφή, στη διαφορική γεωμετρία, είναι το εσωτερικό γινόμενο των εφαπτόμενων διανυσμάτων στην επιφάνεια  $\mathbf{u}(u_r, u_i) : \mathbb{R} \rightarrow \mathbb{R}^2$  και καθορίζει το arc length των καμπυλών στην επιφάνεια.

### 5.5.2 Συζευγμένος διαχωρισμός

Παρατηρώντας ότι  $\|\mathbf{f}\|_{L^2}^2 = \|f_r\|_{L^2}^2 + \|f_i\|_{L^2}^2$ , η ελαχιστοποίηση του (5.64) γίνεται ισοδύναμη με κανονικοποίηση της εικόνας με ξεχωριστούς περιορισμούς στο πραγματικό και στο μιγαδικό μέρος της συνάρτησης

$$E(\mathbf{u}, \mathbf{v}) = J_j(u_r, u_i) + I(f_r, u_r, v_r, t_r) + I(f_i, u_i, v_i, t_i), \quad (5.69)$$

όπου  $I(f, u, v, t) = \lambda \|f - u - v\|_{L^2}^2 + \mu \|v - \sum_k^K t_k\|_{L^2}^2$ . Η αντίστοιχη λύση προκύπτει από τις εξισώσεις Euler-Lagrange για τέσσερις ανεξάρτητες μεταβλητές  $(u_r, u_i, v_r, v_i)$ . Είναι ανάλογη με τις εξισώσεις αποδόμησης πραγματικών εικόνων (5.18, 5.19), με ένα τροποποιημένο όρο καμπυλότητας, την απόκλιση μοναδιαίου gradient της διανυσματικής συνάρτησης  $\mathbf{u}$

$$\begin{aligned} u_r &= f_r - \sum_{k=1}^K t_{kr} + \frac{\mu+\lambda}{2\mu\lambda} \operatorname{div} \left( \frac{\nabla u_r}{\|\nabla \mathbf{u}\|} \right), \\ \|\nabla \mathbf{u}\| &= \sqrt{|u_x|^2 + |u_y|^2} = \sqrt{|\nabla u_r|^2 + |\nabla u_i|^2}. \end{aligned} \quad (5.70)$$

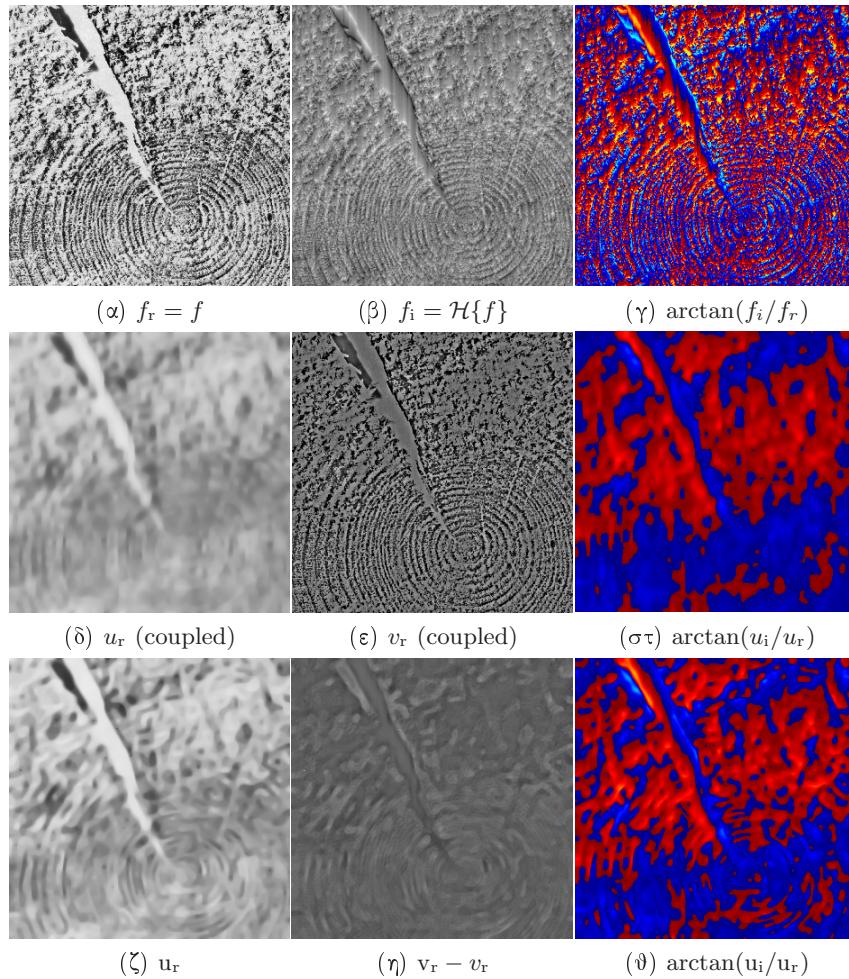
Η πραγματική συνιστώσα υφής  $v_r$  προκύπτει ως  $(u_r, f_r, t_{kr})$  στην αντίστοιχη πραγματική (5.22), ενώ οι εξισώσεις για τις μιγαδικές συνιστώσες  $(u_i, v_i)$  είναι ανάλογες (αντικαθιστώντας τους υποδείκτες  $r$  με  $i$ ). Έτσι, η λύση είναι ένα σύστημα τεσσάρων εξισώσεων, δύο εκ των οποίων είναι μη-γραμμικές ΜΔΕ, συζευγμένων στους όρους καμπυλότητας και στα αυροίσματα ανακατασκευής. Η εξίσωση (5.70) είναι μια κανονικοποίηση ανισοτροπικού τύπου, με το ρυθμό διάχυσης να καθορίζεται ισόποσα από τα gradient της πραγματικής και της φανταστικής συνιστώσας.

Η ανακατασκευή των πραγματικών και φανταστικών συνιστωσών της υφής, για την απλούστερη περίπτωση που  $t = f$ , γράφεται αναλυτικά

$$\begin{aligned} \sum_k t_{kr} &= \operatorname{Re} \left\{ \sum_k f * g_k \right\} = f_r * \sum_k g_{ke} - f_i * \sum_k g_{ko}, \\ \sum_k t_{ki} &= \operatorname{Im} \left\{ \sum_k f * g_k \right\} = f_i * \sum_k g_{ke} + f_r * \sum_k g_{ko}. \end{aligned} \quad (5.71)$$

Για πραγματικά δεδομένα  $t_{kr} = f_r * g_{ke}$ , μόνο η άρτια-συμμετρική απόχριση των φίλτρων επηρεάζει την αποδόμηση, ενώ η μιγαδική συνιστώσα  $t_{ki} = f_r * g_{ke}$  είναι η αδιάφορη, περιττά-συμμετρική απόχριση.

Στο Σχ. 5.12 μιγαδικά δεδομένα δημιουργήθηκαν τεχνητά με χρήση του 2-Δ πολυδιάστατου Hilbert μετασχηματισμού [126, 127] της αρχικής εικόνας (α). Το αποτέλεσμα είναι μια αναλυτική εικόνα  $f + j\mathcal{H}\{f\}$ , για την οποία το πραγματικό και το φανταστικό μέρος αποτελούν το διάνυσμα  $\mathbf{f}_0$  προσανατολισμός του οποίου φαίνεται στο (γ). Οι μιγαδικές συνιστώσες  $\mathbf{u} = [u_r, u_i]$  και  $\mathbf{v} = [v_r, v_i]$  ανακτώνται με δύο τρόπους: 1) τη συζευγμένη, μιγαδική αποδόμηση (5.70) (δεύτερη σειρά) και 2) τον διαχωρισμό σε κάθε κανάλι ξεχωριστά χρησιμοποιώντας τις αρχικές εξισώσεις (5.21, 5.22), δηλαδή το άρτιο Gabor φίλτρο με πραγματικά δεδομένα εισόδου  $f_r$  (τρίτη σειρά) και το περιττό με φανταστικά  $f_i$ . Και στα δύο σχήματα οι παράμετροι  $(\lambda, \mu)$  ήταν οι ίδιοι και χρησιμοποιήθηκε η ανακατασκευή με επιλογή των κυρίαρχων



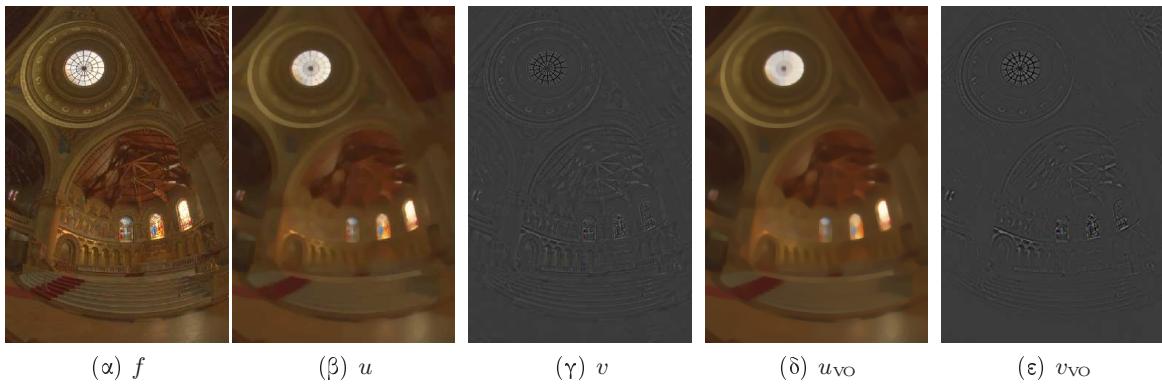
Σχήμα 5.12: Μιγαδική αποδόμηση (Brodatz-D97). (α),(β) Hilbert ζεύγος ( $f, \mathcal{H}\{f\}$ ) και (γ) προσανατολισμός του αρχικού διανύσματος δεδομένων  $\mathbf{f} = [f, \mathcal{H}\{f\}]$ . (δ),(ε) πραγματικό cartoon  $u_r$  και υφή  $v_r$  (συν μια προσθετική σταθερά) από το συζευγμένη μιγαδική αποδόμηση, και (στ) προσανατολισμός του διανύσματος  $\mathbf{u} = [u_r, u_i]$ , (ζ)-(θ) «καναλικό» πραγματικό cartoon  $u_r$  (με πραγματική είσοδο και άρτιο Gabor, διαφορά των πραγματικών συνιστωσών υφής  $v_r$  και προσανατολισμός του διανύσματος. Ο διαχωρισμός πραγματοποιήθηκε με επιλογή κυρίαρχων συνιστωσών και  $(\lambda, \mu) = (1, 0.1)$ .

συνιστωσών (5.57). Οι προσανατολισμοί των χανονικοποιημένων cartoon διανυσμάτων φαίνονται στις εικόνες (στ) και (θ) αντίστοιχα. Το συζευγμένο σχήμα δίνει περισσότερο ομαλά αποτελέσματα και καλύτερο διαχωρισμό της υφής (π.χ. οι δακτύλιοι στην υφή του δέντρου). Ωστόσο, η συνιστώσα cartoon είναι λιγότερο οξεία με «θολωμένες» περιοχές. Σημειώνεται επίσης ότι οι φωτεινές περιοχές στην εικόνα της διαφοράς των πραγματικών συνιστωσών υφής σηματοδοτούν τον περιορισμό των επιπέδων φωτεινότητας στο συζευγμένο σενάριο.

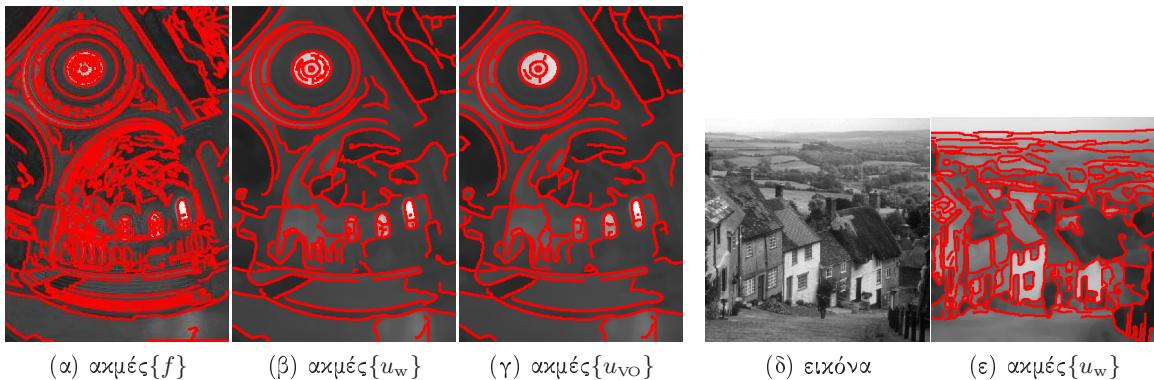
## 5.6 Σύνοψη

Η ευελιξία και η περιγραφικότητα του πολυκαναλικού μοντέλου υφής μέσω πολυσυχνοτικού διαχωρισμού χρησιμοποιήθηκε σε ένα νέο σχήμα αποδόμησης εικόνων. Το νέο  $u + K_v$  σχήμα είναι μια προσπάθεια στην κατεύθυνση εισαγωγής πρότερης πληροφορίας μοντέλων υ-

Κεφάλαιο 5. Συνέργεια διαχωρισμού και μοντελοποίησης: Αποσύνθεση εικόνων



Σχήμα 5.13: Αποσύνθεση σε cartoon και υφή με το  $u + Kv$  σχήμα με ανακατασκευή στάθμισης πλατών στα (β),(γ), και το Vese-Osher σχήμα στα (δ),(ε), με παραμέτρους  $(\lambda, \mu) = (10, 5)$  ανδ  $(\lambda_{vo}, \mu_{vo}) = (5, 0.1)$ . Η αποδόμηση έγινε σε κάθε χρωματικό κανάλι ξεχωριστά και τα αποτελέσματα παρουσιάζονται ως έγχρωμες εικόνες. Οι εικόνες υφής παρουσιάζονται με την προσθήκη σταθεράς ίσης με τη μέση τιμή του cartoon (*Stanford Memorial Church* (high-dynamic-range image) [88]).



Σχήμα 5.14: Ανίχνευση ακμών από την αρχική εικόνα Σχ. 5.13(α) και τις συνιστώσες cartoon Σχ. 5.13(β)(δ).



Σχήμα 5.15: Αποσύνθεση σε cartoon και υφή για αποκατάσταση εικόνων (η εικόνα αριστερά είναι τμήμα της τοιχογραφίας 'Πότνια' από την ανασκαφή στο Ακρωτήρι Θήρας).

φής σε μοντέλα της οικογένειας  $u + v$ . Το  $K$  του νέου σχήματος, που το διαχωρίζει από τα  $u + v$  μοντέλα, συμβολίζει την επιπλέον αποδόμηση της συνιστώσας της υφής σε  $K$  υποσυνιστώσες αλλά και την ισοδυναμία του προβλήματος με ένα πρόβλημα εκτίμησης  $K + 1$  συνιστώσων της εικόνας. Το νέο μεταβολικό σχήμα βασίζεται έμμεσα στην ανακατασκευή της πλησιέστερης στην υφή πληροφορίας της εικόνας με όρους χυρίως της σχέσης υφής και κλίμακας περιγραφής. Στον πυρήνα της βρίσκεται η περιγραφή των στενοζωνικών συνιστω-

σών μέσω του AM-FM μοντέλου, η αναπαράσταση μέσω κυρίαρχων διαμορφώσεων αλλά και η αποδιαμόρφωση τους για την εξαγωγή χωρικών μετρικών έντασης της υφής όπως είναι το πλάτος διαμόρφωσης. Το ίδιο πλαίσιο οδήγησε και σε μια περιγραφή της υφής με στοιχειώδη χωρικά δομικά στοιχεία ανάλογα με τις περιγραφές τεξτονίων. Τέλος, αναπτύχθηκε μια επέκταση του σχήματος για μιγαδικές εικόνες όπου η μοντελοποίηση γίνεται με μιγαδικές συνιστώσες και μιγαδικά φίλτρα. Το σχήμα μπορεί να εφαρμοστεί για εξειδικευμένη επεξεργασία των συνιστώσων όπως για παράδειγμα βελτιωμένη ανίχνευση ακμών από τη συνιστώσα ι που φαίνεται στο Σχ. 5.14, αποκατάσταση εικόνων που φαίνεται στο Σχ. 5.15, εξαγωγή χαρακτηριστικών και ταξινόμηση υφής από το ι που θα παρουσιαστούν στο επόμενο κεφάλαιο. Επεκτάσεις του μοντέλου, όπως η μιγαδική αποσύνθεση συμπεριλαμβάνουν το χειρισμό συναρτήσεων διανυσματικών τιμών, όπως είναι οι έγχρωμες εικόνες (Σχ. 5.13).

□

## Κεφάλαιο 6

# Συνέργεια διαχωρισμού και μοντελοποίησης: Ταξινόμηση υφής

Τα διαφορετικά επίπεδα οργάνωσης των φυσικών εικόνων επιβάλλουν τη διαφορετική επεξεργασία και ανάλυση των κατηγοριών οπτικών χαρακτηριστικών που αναζητούνται. Η υφή θα μπορούσε να υφεωρηθεί το χαρακτηριστικό που παραμένει «αφαιρώντας» από την εικόνα τις αργές μεταβολές φωτεινότητας, το χρώμα, τα σχήματα και τα περιγράμματα των αντικειμένων, τη γεωμετρία της σκηνής ή γενικότερα τις δομές μεγάλης κλίμακας. Το σχήμα συνέργειας βασίστηκε σε με τέτοια παραδοχή εκ των προτέρων για την αποδόμηση, με ανάθεση στην υφή χαρακτηριστικών όπως μικρή ολική μεταβολή, μηδενική μέση τιμή, ταλαντώσεις μεγάλης συχνότητας και χωρικά διαμορφωμένο πλάτος και φάση. Ο διαχωρισμός από μακροδομές της εικόνας πριν από την εφαρμογή του μοντέλου, δηλαδή την εκτίμηση παραμέτρων, την εξαγωγή χαρακτηριστικών και την ανάλυση, αποτελεί ένα στάδιο προεπεξεργασίας και βελτίωσης της πληροφορίας υφής. Η αποδοτικότητα των χαρακτηριστικών που εξάγονται από τη διαχωρισμένη συνιστώσα αναδεικνύεται σε εφαρμογές ταξινόμησης υφής, όπου συγχρίνεται το σφάλμα ταξινόμησης και η διαχωρισμότητα των τάξεων που επιτυγχάνουν διαφορετικά διανύσματα χαρακτηριστικών.

### 6.1 Ανάλυση της συνιστώσας υφής

Η διαδικασία μοντελοποίησης της υφής θα μπορούσαμε να υφεωρήσουμε ότι απαρτίζεται από δύο διακριτά στάδια: την επιλογή του κατάλληλου μοντέλου και την επιλογή των δεδομένων στα οποία θα εφαρμοστεί. Στο πρώτο στάδιο οι απαιτήσεις και τα χριτήρια επιλογής καθορίζονται από τη γενικότητα του μοντέλου, δηλαδή τη δυνατότητα του να ερμηνεύει τις διαφορετικές αντιληπτικές τάξεις υφής (δομημένες, στοχαστικές, περιοδικές, συνθετικές, φυσικές) και τη διακριτική ικανότητα του, δηλαδή την ευαισθησία του στα λοιπά χαρακτηριστικά της εικόνας. Το δεύτερο περιλαμβάνει πιθανούς μετασχηματισμούς και προεπεξεργασία της

αρχικής εικόνας έτσι ώστε να προσαρμοστεί στους μηχανισμούς ανάλυσης και στο ταίριασμα του μοντέλου.

Το πολυυσχυνοτικό, στενοζωνικό μοντέλο ανταποκρίνεται σε πολλές διαφορετικές οικογένειες εικόνων υφής (περιοδικές, προσανατολισμένες, δομημένες, στοχαστικές υφές αποκρίνονται σε ένα ή περισσότερα κανάλια της συστοιχίας). Η προσαρμογή του μοντέλου είναι περισσότερο δικαιολογημένη και εύρωστη όταν γίνεται στη συνιστώσα υφής  $u$  που προέρχεται από το σχήμα αποδόμησης. Η ταλαντεύομενη συνιστώσα διατηρεί τα ουσιώδη χαρακτηριστικά της υφής όπως η κλίμακα, ο προσανατολισμός, ο ρυθμός ταλάντωσης και οι τοπικά αναδυόμενες συχνότητες και είναι απαλλαγμένη από προσθετικό υψηπερατό ύδρυμα και μεγάλες δομές της εικόνας, διαφορές στα επίπεδα φωτεινότητας, απότομες μεταβάσεις και ταλαντώσεις μεγάλης περιόδου. Αυτή η προεπεξεργασία λειτουργεί ως μηχανισμός επιλογής της υφής και καταπίεσης των αδιάφορων ή ασυσχέτιστων με την υφή χαρακτηριστικών, όπως είναι οι ακμές και τροφοδοτεί το μοντέλο με ταλαντεύομενες συναρτήσεις, μηδενικής μέσης τιμής και χαμηλού αλλά μεταβαλόμενου πλάτους. Κάτι τέτοιο είναι εμφανές και στο Σχ. 5.6 όπου φαίνεται η απόκριση ενός γραμμικού φίλτρου στην αρχική εικόνα και στις συνιστώσες της.

Ένα παράδειγμα πρακτικής εφαρμογής της θεώρησης για μοντελοποίηση μετά την αποδόμηση μπορεί να βρεθεί στην ανίχνευση υφής με τη βοήθεια μη-γραμμικών τελεστών. Η ύπαρξη και τα χαρακτηριστικά της υφής ανιχνεύονται με την εφαρμογή κατάλληλων τελεστών στις εξόδους των φίλτρων [110, 136, 168, 268]. Ένας τέτοιος τελεστής πρέπει να δίνει απόκριση σε περιοχές υφής και να μην αποκρίνεται σε χαρακτηριστικά που σχετίζονται με τα αντικείμενα και τη γεωμετρία της σκηνής όπως περιγράμματα ή ακμές [110, 186].

### 6.1.1 Διανύσματα χαρακτηριστικών

Για να αναδειχθεί η αξία και τα αποτελέσματα της αποδόμησης στη μοντελοποίηση θεωρούμε πέντε διαφορετικά διανύσματα χαρακτηριστικών για ανάλυση υφής. Η ακρίβεια του μοντέλου καθορίζεται από την αποτελεσματικότητα των χαρακτηριστικών με όρους ταξινόμησης και διάκρισης υφής. Τα διανύσματα βασίζονται στο πολυκαναλικό μοντέλο στενοζωνικών συνιστώσων και εξάγονται στην έξοδο μιας συστοιχίας Gabor. Η περιγραφή της υφής ανάγεται στο χώρο των  $K$  αποχρίσεων των φίλτρων, όπου σε κάθε σημείο της εικόνας αποδίδεται ένα σύνολο τιμών, αποτέλεσμα γραμμικού φιλτραρίσματος και μη-γραμμικών τελεστών ή μετασχηματισμών επιλογής χαρακτηριστικών.

Η εξαγωγή χαρακτηριστικών γίνεται τόσο από την αρχική εικόνα  $f$  όσο και από τη συνιστώσα υφής  $u$  μετά την αποδόμηση με το  $f = u + Ku$  μοντέλο (τα διανύσματα συμβολίζονται με  $\vec{I}_{\{\cdot\}}$  και  $\vec{V}_{\{\cdot\}}$  αντίστοιχα):

- α) **DCA** ( $4 \times 1$ ) [92, 164]: διάνυσμα των κυρίαρχων χαρακτηριστικών διαμόρφωσης (3.14), ενισχυμένο από τις τιμές φωτεινότητας (από την αρχική  $f$  ή την κανονικοποιημένη cartoon εικόνα  $u$ )

$$\vec{V}_D(x, y) = [a_d(v), \vec{w}_d(v), u](x, y), \quad (6.1.1\alpha)$$

$$\vec{I}_D(x, y) = [a_d(f), \vec{w}_d(f), f](x, y). \quad (6.1.1\beta)$$

Κεφάλαιο 6. Συνέργεια διαχωρισμού και μοντελοποίησης: Ταξινόμηση υφής

- β) **Gabor** ( $K \times 1$ ) [3, 36, 110, 219, 220, 250, 265, 290]: ενέργεια ορθογωνισμού των  $K$  αποχρίσεων των μιγαδικών φίλτρων (2.1)

$$\vec{V}_G(x, y) = [\{(g_{ke} * v)^2 + (g_{ko} * v)^2\}_{k=1:K}](x, y), \quad (6.1.1\alpha)$$

$$\vec{I}_G(x, y) = [\{(g_{ke} * f)^2 + (g_{ko} * f)^2\}_{k=1:K}](x, y). \quad (6.1.1\beta)$$

- γ) **Teager** ( $K \times 1$ ) [269]: μιγαδική Teager ενέργεια (2.10), δηλαδή η εφαρμογή του μιγαδικού, μη-γραμμικού τελεστή στις εξόδους των φίλτρων

$$\vec{V}_T(x, y) = [\{C(g_k * v)\}_{k=1:K}](x, y), \quad (6.1.1\alpha)$$

$$\vec{I}_T(x, y) = [\{C(g_k * f)\}_{k=1:K}](x, y). \quad (6.1.1\beta)$$

- δ) **pooled** ( $(K+4) \times 1$ ): κοινό διάνυσμα από το συνδυασμό των DCA και Gabor χαρακτηριστικών

$$\vec{V}_P(x, y) = [\vec{V}_G, \vec{G}_D](x, y), \quad (6.1.1\alpha)$$

$$\vec{I}_P(x, y) = [\vec{I}_G, \vec{I}_D](x, y). \quad (6.1.1\beta)$$

- ε) **pooled/PCA** ( $M \times 1$ ,  $M < K+4$ ): γραμμική εξαγωγή χαρακτηριστικών από συνδυασμένο διάνυσμα μέσω PCA ή επέκταση Karhunen-Loéve (προβολή σε ένα γραμμικό υποχώρο μέσω κριτηρίων ελαχίστων τετραγώνων) [84, 137] και επιλογή των συνιστωσών με τις μεγαλύτερες ιδιοτιμές που συνεισφέρουν περισσότερο στη συνολική μεταβλητότητα.

Τα διανύσματα «πλήρων συνιστωσών» **Gabor** και **Teager** είναι  $K$ -διάστατες πολυκαναλικές ενεργειακές μετρήσεις από μη-γραμμικούς τελεστές στο σύνολο των στενοζωνικών συνιστωσών. Οι υποσυνιστώσες υφής από το στάδιο αποδόμησης μπορούν να χρησιμοποιηθούν για επιλογή χαρακτηριστικών και μείωση της διάστασης, δηλαδή εξαγωγή από ένα υποσύνολο ή γραμμικό συνδυασμό των συνιστωσών. Το διάνυσμα κυρίαρχων χαρακτηριστικών συνιστά μια «πυκνή», χαμηλής διάστασης ( $4 \times 1$ ), συνεχή περιγραφή της υφής [92, 125, 165]. Αποτελείται από τις παραμέτρους διαμόρφωσης της κυρίαρχης συνιστώσας, οι οποίες λαμβάνονται μέσω DCA στο  $v$  ή ως υποπροϊόν της διαδικασίας αποδόμησης μέσω (5.28)  $v_d = \frac{\lambda}{\mu}(f - u - v) + d$ , με  $d$  την κυρίαρχη συνιστώσα της εικόνας (3.13) ή διαφορετικά την πρώτη συνιστώσα στο ιεραρχικά διατεταγμένο άθροισμα (5.57).

Η ανάλυση μέσω κυρίαρχων συνιστωσών και η ανακατασκευή υφής μπορεί να τεθεί σε ένα αυστηρό μαθηματικό πλαίσιο, θεωρώντας μια σταθμισμένη υπέρθεση των διατεταγμένων κυρίαρχων συνιστωσών  $d_k$ . Τα βάρη είναι διακριτά δείγματα από μια οικογένεια σιγμοειδών συναρτήσεων (sigmoid functions)

$$w(k) = 1 / (1 + \exp\{c(k-D)\}), \quad k = 1 \dots K, \quad (6.1)$$

όπου  $k$  ο δείκτης της συνιστώσας, το κέντρο  $D$  της σιγμοειδούς δηλώνει τον αριθμό των σημαντικών συνιστωσών χαμηλότερης τάξης και  $c \geq 0$  είναι μια σταθερά που καθορίζει την κλίση στη συνιστώσα  $k = D$ . Η ανακατασκευή υφής μπορεί να γραφτεί στη γενική μορφή

$$\sum_k^K v_k \triangleq \sum_k^K w(k; c, D) d_k. \quad (6.2)$$

Αφήνοντας το  $c \geq 0$  να μεταβληθεί δημιουργείται μια συνεχής γέφυρα ανάμεσα στη μη-ορθογωνική, ατελή ανακατασκευή Gabor της εικόνας  $\sum_k^K f_k$  για  $c = 0$  και στο άθροισμα DCA χαμηλότερης τάξης  $\sum_k^D d_k$  για  $c = +\infty$ . Επιπλέον, η επιλογή της πρώτης μόνο συνιστώσας θέτοντας  $D = 2$  ισοδυναμεί με την ανάλυση μέσω DCA και ομοίως με την ανακατασκευή με στάθμιση πλατών (5.43).

## 6.2 Διάκριση και ομοιότητα τάξεων

Ο σχηματισμός των τάξεων υφής με τα εν λόγω διανύσματα χαρακτηριστικών αξιολογείται για θέματα όπως η ανάκτηση εικόνων με βάση την υφή [23, 80, 81, 195, 267] και η κατηγοριοποίηση χρησιμοποιώντας δύο κατάλληλα ποσοτικά κριτήρια, τη σχετική εντροπία και το κριτήριο Fisher. Και για τα δύο θεωρούμε συγκρίσεις τάξεων υφής (texture classes) ανά ζεύγη, με ένα ή περισσότερα στιγμιότυπα της τάξης (class instance) διαθέσιμα.

Ένα πιθανοτικό μέτρο ομοιότητας είναι η σχετική εντροπία (relative entropy) [80, 81, 84] ανάμεσα σε δύο συναρτήσεις πυκνότητας - πιθανότητας, που εκφράζεται από την απόσταση Kullback-Leibler

$$\text{KLD}_{i,j}(\theta_i, \theta_j) = \int_{-\infty}^{+\infty} p(\mathbf{x}; \theta_i) \log \frac{p(\mathbf{x}; \theta_i)}{p(\mathbf{x}; \theta_j)} d\mathbf{x}, \quad (6.3)$$

όπου οι δύο πυκνότητες είναι ίδιες, με διαφορετικά διανύσματα παραμέτρων  $\theta$ . Οι παράμετροι εκφράζονται με εκτιμήσεις μέγιστης πιθανοφάνειας (Maximum Likelihood estimates) και η ελαχιστοποίηση του (6.3) μεγιστοποιεί την λογαριθμική πιθανοφάνεια (log-likelihood)  $p(\mathbf{x}; \theta_j)$ . Έτσι, όσο μικρότερη είναι η KL απόσταση, τόσο ποιο όμοια είναι τα στατιστικά των δύο τάξεων ( $\text{KLD}_{i,j}(\theta_i, \theta_j) = 0 \iff \theta_i = \theta_j$ ). Σημειώνεται εδώ ότι  $\text{KLD}_{i,j} \neq \text{KLD}_{j,i}$ , δηλαδή το μετρικό KLD ποσοτικοποιεί την ομοιότητα, δεδομένης της μίας εκ των δύο τάξεων. Οι κατανομές θεωρούνται πολυμεταβλητές Gaussian  $\mathcal{N}_d \sim (\mu, \Sigma)$  στον  $d$ -διάστατο χώρο χαρακτηριστικών, με μέσο διάνυσμα (mean vector)  $\mu$  και πίνακα συμμεταβλητότητας (covariance matrix)  $\Sigma$ .

Το κριτήριο Fisher είναι ένα μετρικό που βασίζεται στη διακριτική ανάλυση (discriminant analysis) και ποσοτικοποιεί για δύο τάξεις  $C_i$  και  $C_j$  το πόσο συμπαγείς είναι οι τάξεις (compactness) και την απόσταση ανάμεσα τους (inter-cluster distance), από τη γραμμική προβολή των  $d$ -διάστατων δεδομένων  $\mathbf{x}$  σε μια γραμμή  $\mathbf{w}$ :

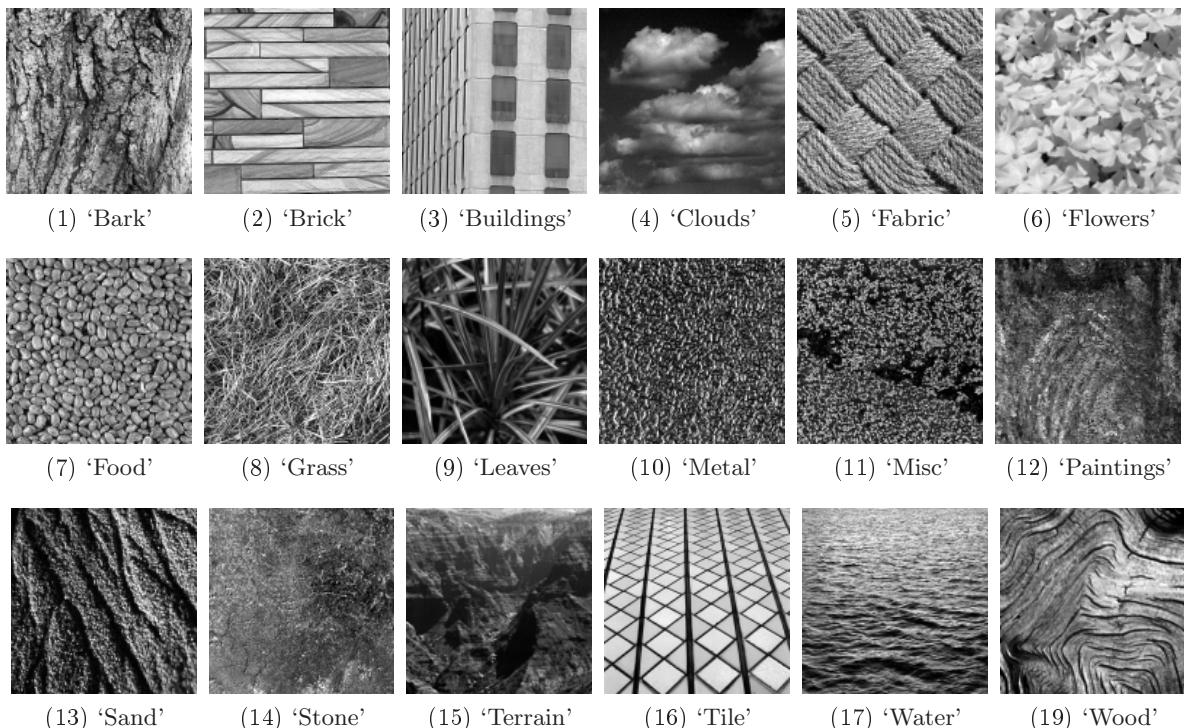
$$\text{FC}(\mu_i, \mu_j) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}, \quad \mathbf{w} = \mathbf{S}_w^{-1}(\mu_i - \mu_j). \quad (6.4)$$

Οι πίνακες  $\mathbf{S}_W, \mathbf{S}_B$  είναι οι εσω- και δια- ταξικοί πίνακες διασποράς (intra, inter-class scatter matrices) αντίστοιχα [84], δηλαδή ο πίνακας διασποράς εσωτερικά της τάξης

$$\mathbf{S}_W = \sum_{\mathbf{x} \in C_i} (\mathbf{x} - \mu_i)(\mathbf{x} - \mu_i)^T + \sum_{\mathbf{x} \in C_j} (\mathbf{x} - \mu_j)(\mathbf{x} - \mu_j)^T, \quad (6.5)$$

ανάλογος στον πίνακα συμμεταβλητότητας των συνδυασμένων  $d$ -διάστατων δεδομένων, και ο πίνακας διασποράς ανάμεσα στις τάξεις

$$\mathbf{S}_B = (\mu_i - \mu_j)(\mu_i - \mu_j)^T. \quad (6.6)$$



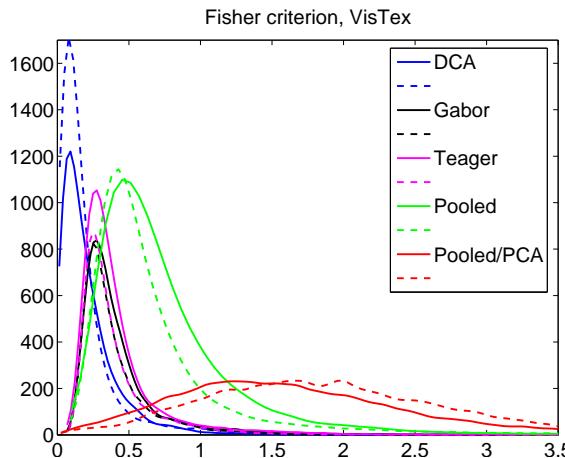
Σχήμα 6.1: Στιγμότυπα από τις τάξεις της υφής της βάσης VisTex [216]. Οι εικόνες είναι αντιπροσωπευτικά δείγματα από τα στιγμότυπα  $128 \times 128$  των τάξεων που έχουν σχηματιστεί εμπειρικά, ανάλογα με τη φυσική επιφάνεια ή σκηνή που αποτυπώνουν.

Το χριτήριο βασίζεται στην απόσταση των μέσων των δεδομένων μετά την προβολή τους σε μια γραμμή  $\mathbf{w}$  σε σχέση με το άθροισμα από τις μεταβλητές τους και η μεγιστοποίηση του δίνει το βέλτιστο γραμμικό διαχωριστή για τις δύο τάξεις. Έτσι, μεγαλύτερες τιμές δηλώνουν καλύτερη διαχωρισμότητα των δεδομένων  $\mathbf{x}$  που ανήκουν στη μία τάξη από αυτά που ανήκουν στην άλλη. Επιπλέον, το χριτήριο ή μια παραλλαγή του, η inter-to-intra απόσταση των τάξεων  $\mathbf{S}_B/\mathbf{S}_W$ , μπορεί να εφαρμοστεί για αποτίμηση των επιμέρους χαρακτηριστικών (feature evaluation) ενός διανύσματος [42, 110].

### 6.2.1 Δοκιμές επιδόσεων στη βάση VisTex

Τα διανύσματα χαρακτηριστικών εξετάσθηκαν ως προς τη διαχωρισμότητα και την ομοιότητα που αποδίδουν σε διαφορετικές τάξεις υφής, με τη βοήθεια δοκιμών επιδόσεων (benchmark evaluation) στις εικόνες αναφοράς της βάσης VisTex [216]. Πρόκειται για ένα σύνολο  $N = 167$  εικόνων διαστάσεων  $128 \times 128$ , που έχουν ταξινομηθεί χονδρικά, με ανθρώπινη επί-βλεψη σε 19 τάξης υλικών/σκηνών (όπως π.χ. ξύλο, άμμος, μέταλλο) στιγμότυπα των οποίων φαίνονται στο Σχ. 6.1. Η αποτίμηση της επίδοσης έγινε με μέτρηση των δύο αποστάσεων στα διαφορετικά διανύσματα χαρακτηριστικών που εξάγονται τόσο από την αρχική εικόνα  $f$  όσο και από την εκτιμώμενη συνιστώσα υφής  $v$  μετά από αποδόμηση με το  $u + Kv$  σχήμα.

Η αποδόμηση πραγματοποιήθηκε στο σύνολο των εικόνων της βάσης χρησιμοποιώντας ανακατασκευή από το άθροισμα κυρίαρχων συνιστωσών (5.57), με αυτόματη επιλογή του αριθμού των συνιστωσών μέσω της μεταβλητής τους (Εν. 5.4.1.3). Η συστοιχία των φίλτρων

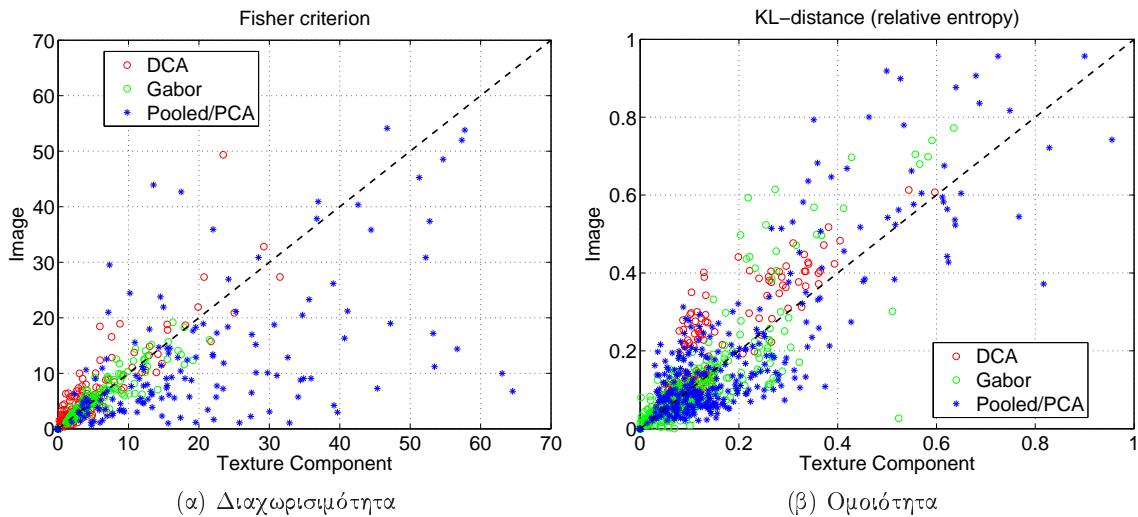


Σχήμα 6.2: Κατανομές του κριτήριου Fisher στη βάση VisTex(167 εικόνες, 19 διαμορφωμένες κατηγορίες υφής). Κάθε εικόνα θεωρήθηκε στιγμιότυπο από ξεχωριστή τάξη και όλα τα ζευγάρια εκτιμήθηκαν ως προς τη διαχωρισμό της τάξης της εικόνας. Οι χαμηλές τιμές οφείλονται στην πόλωση των εκτιμήσεων ως προς τον τρόπο που ορίζονται οι τάξεις. Οι συμπαγείς γραμμές αντιστοιχούν σε χαρακτηριστικά από την αρχική εικόνα  $f = u + v$  και οι διακεκομμένες σε εξαγωγή από τη συνιστώσα  $v$ .

(Εν. 2.2.3) συγκροτήθηκε από 40 κανάλια, σε πέντε κλίμακες και οκτώ προσανατολισμούς και οι παράμετροι του μεταβολικού σχήματος διατηρήθηκαν σταθερές  $(\lambda, \mu) = (0.1, 0.01)$  (έτσι ώστε να επιβληθεί κάποια έξτρα ομαλότητα στα cartoon). Η επιλογή των παραμέτρων παιζει σημαντικό ρόλο στα αποτελέσματα και την ομοιότητα ή τη διαχωρισμό των χαρακτηριστικών αφού ελέγχει το ποσό της πληροφορίας που «παραβλέπεται» από την υφή.

Αρχικά, κάθε εικόνα θεωρήθηκε στιγμιότυπο από μια διαφορετική τάξη και όλα τα πιθανά  $N!$  ζευγάρια δοκιμάστηκαν. Στο Σχ. 6.2 τα διαγράμματα αντιστοιχούν στις κατανομές των τιμών του χριτηρίου Fisher για τα διανύσματα χαρακτηριστικών  $\vec{I}_{\{j\}}$  και  $\vec{V}_{\{j\}}$  που εξάγονται από τα  $f$  και  $v$  αντίστοιχα. Οι κατανομές συγκριτικά με τους δύο διαφορετικούς τρόπους εξαγωγής είναι παρόμοιες, με μικρές διαφορές στις μέσες τιμές και τυπικές αποκλείσεις μικρότερες για το  $v$  εκτός από το DCA διάνυσμα. Από τη σκοπιά των διανυσμάτων, τα DCA χαρακτηριστικά λαμβάνουν τις χαμηλότερες τιμές, ωστόσο οι κατανομές είναι περισσότερο συμπαγείς, πράγμα που δηλώνει συνέπεια και ευρωστία σε σχέση με τον τυχαίο σχηματισμό των τάξεων. Εξετάζοντας τους τρόπους εξαγωγής, οι μετρήσεις με το pooled/PCA διάνυσμα μετατοπίζονται σε μεγαλύτερες τιμές όταν εφαρμόζεται στο  $v$ . Σημειώνεται ωστόσο ότι οι εκτιμήσεις εδώ είναι πολωμένες (biased), λόγω του ότι πολλά από τα στιγμιότυπα-τάξεις είναι πολύ παρόμοια (π.χ. στιγμιότυπα της ίδιας αντιληπτικής ή φυσικής τάξης ή τμήματα από την ίδια εικόνα που έχουν υποστεί κλιμάκωση ή περιστροφή).

Το σύνολο των δεδομένων δοκιμών βελτιώθηκε και ισχορίζοντας «χειροκίνητα» τις τάξεις υφής από τα πιο αντιπροσωπευτικά και παρόμοια αντιληπτικά στιγμιότυπα των τάξεων αναφοράς, π.χ. εξαιρώντας με οπτική επίβλεψη στιγμιότυπα-μοναδικότητες (singularities) των τάξεων. Οι μετρήσεις των αποστάσεων επαναλήφθηκαν για τις 19 τάξεις της βάσης, όπου δείγματα εξάχθηκαν από όλα τα επιλεγμένα στιγμιότυπα της κάθε τάξης. Στο Σχ. 6.3 παρουσιάζονται διαγράμματα διασποράς (scatter plots) των αποστάσεων στο  $(v, f)$  χώρο τιμών, στα οποία



Σχήμα 6.3: Διαχωρισμότητα και ομοιότητα διαφορετικών τάξεων υφής από τρία διανύσματα χαρακτηριστικών (Gabor, DCA, Pooled/PCA) στη βάση VisTex (19 κατηγορίες, υποσύνολα των στιγμότυπων τους επιλέχθηκαν με επίβλεψη για την κάθε τάξη). Οι τιμές των αποστάσεων Fisher και KL απεικονίζονται ως scatter διαγράμματα στον  $(v, f)$  χώρο. Κάθε σημείο αποτυπώνει τις τιμές για ένα ζεύγος στιγμότυπων από διαφορετικές τάξεις, με εξαγωγή χαρακτηριστικών από το τη συνιστώσα  $v$  και την αρχική  $f$ . (α) κριτήριο Fisher (σημεία κάτω από τη διαγώνιο παρουσιάζουν βελτιωμένη διαχωρισμότητα στις μετρήσεις από την  $v$ ), (β) σχετική εντροπία (σημεία πάνω από τη διαγώνιο αντιστοιχούν σε περισσότερα πλησιέστερες τάξεις για τις μετρήσεις με την  $f'$ ). Σχετικές μεταβολές των τιμών των αποστάσεων για τις 19 τάξεις, στιγμότυπα των οποίων φαίνονται στο Σχ. 6.1, δίνονται στους Πίνακες 6.1, 6.2.

κάθε σημείο αποτυπώνει για ένα ζευγάρι στιγμότυπων από διαφορετικές τάξεις την τιμή για το  $v$  σε σχέση με την τιμή για το  $f$ . Η διαγώνια ευθεία είναι η γραμμή-σύνορο που δηλώνει ίδια απόδοση και για τα δύο.

Οι μετρήσεις ερμηνεύονται «αντίστροφα» για τα δύο διαγράμματα. Η διαχωρισμότητα και η διαφορετικότητα είναι σημαινόμενες από μεγάλες τιμές του χριτηρίου Fisher και μικρές τιμές της εντροπίας. Στο διάγραμμα (α) ο μικρότερος αριθμός των Gabor και PCA σημείων κάτω από τη διαγώνιο μεταφράζεται σε βελτιωμένη διαχωρισμότητα, δηλαδή μικρότερη τιμή του χριτηρίου για περισσότερα ζευγάρια τάξεων χρησιμοποιώντας τη συνιστώσα υφής μόνο. Αντίθετα στο (β), το διάνυσμα DCA έχει περισσότερα σημεία πάνω από τη διαγώνιο, πράγμα που σημαίνει μικρότερη σχετική εντροπία και επομένως λιγότερο «χοντινές» ως προς τις κατανομές τους τάξεις στο  $v$ . Στον Πίν. 6.1 παρουσιάζεται η σχετική % αλλαγή ανά κατηγορία ή τάξη υφής των τιμών του χριτηρίου Fisher χρησιμοποιώντας χαρακτηριστικά από την  $v$  συνιστώσα σε σχέση με την εξαγωγή τους από την εικόνα  $f$ . Τα ποσοστά είναι μέσες τιμές για όλες τις δοκιμές που περιλαμβάνουν μια συγκεκριμένη τάξη. Οι μεγάλες σχετικές αλλαγές που επιφέρει το pooled/PCA και στα δύο χριτήρια οφείλονται αφενός στο «άπλωμα» της κατανομής σε σχέση με τα άλλα διανύσματα και αφετέρου στην εμφανή μετατόπιση της προς μεγαλύτερες τιμές όταν εκτιμάται από το  $v$  (Σχ. 6.2).

Τάξη (αρ. δειγμάτων)	Fisher χριτήριο (σχετικό %)				
	DCA	Gabor	Teager	pooled*	PCA†
‘Bark’ (5)	-30.0	-3.4	-0.97	-19.0	83
‘Brick’ (5)	-15.0	4.9	1.6	-5.8	150
‘Buildings’(2)	9.0	46.0	29.0	24.0	810
‘Clouds’ (2)	-4.6	-2.7	-4.6	-7.4	350
‘Fabric’ (2)	-19.0	3.8	5.4	-9.5	280
‘Flowers’ (3)	-27.0	6.7	-6.8	-12.0	100
‘Food’ (3)	7.9	-3.9	1.6	-8.8	210
‘Grass’ (2)	-41.0	-5.2	-13.0	-28.0	9
‘Leaves’ (2)	3.2	- 21.0	- 18.0	-18.0	90
‘Metal’ (4)	-32.0	1.7	- 4.9	-9.0	260
‘Misc’ (2)	-2.5	1.1	3.7	-3.8	37
‘Paintings’(3)	-35.0	15.0	15.0	17.0	240
‘Sand’ (5)	-7.2	13.0	19.0	13.0	51
‘Stone’ (3)	-10.0	-6.7	-2.3	-6.9	100
‘Terrain’ (4)	-1.9	1.5	2.8	-2.9	160
‘Tile’ (3)	-16.0	23.0	9.3	2.8	230
‘Water’ (5)	5.2	8.0	3.9	9.0	190
‘Waldo’ (2)	7.0	-7.2	-11.0	-7.1	140
‘Wood’ (1)	140.0	-12.0	-11.0	18.0	-11
<b>Μέση τιμή</b>	<b>-3.9</b>	<b>3.3</b>	<b>1.0</b>	<b>-2.9</b>	<b>180</b>

Πίνακας 6.1: Μέση σχετική αλλαγή (%) του κριτηρίου Fisher με εξαγωγή χαρακτηριστικών από τη συνιστώσα υφής  $v$  σε σχέση με τα χαρακτηριστικά από την αρχική εικόνα  $f$ .

### 6.3 Ταξινόμηση υφής

Η αποδόμηση μιας εικόνας με σκοπό το διαχωρισμό συνιστωσών καταλήγει σε πληροφορία δύο καναλιών για την εικόνα, κάνοντας έτοιμη δυνατή την εξειδικευμένη και προσανατολισμένη στο περιεχόμενο διαχείριση και ανάλυση [9, 29, 92, 186, 268]. Η δομή  $u$ , για παράδειγμα, μπορεί να χρησιμοποιηθεί για την ανίχνευση των ιδιοτήτων αντικειμένων μεγάλης κλίμακας και την γενική ανάλυση σκηνών ενώ  $v$  για την ομαδοποίηση των χαρακτηριστικών και των στατιστικών της υφής. Η συνιστώσα  $u$  της εικόνας είναι βελτιωμένη και απαλλαγμένη από τις ασυνέχειες αντικειμένων, τις αργές μεταβάσεις της έντασης, τις μεταβολές της σκίασης και της φωτεινότητας και καταλληλότερη για περαιτέρω επεξεργασία της υφής.

Η συνέργεια που αναπτύχθηκε περιλαμβάνει σαφή, βασισμένη στο μοντέλο ανακατασκευή της υφής κατά τη διάρκεια του σταδίου αποδόμησης και διαχωρισμό της υφής από τη δομή πριν από τη μοντελοποίηση και την ανάλυση της. Η δεύτερη, «αντίστροφη» κατεύθυνση επιδεικνύεται σε μια εφαρμογή ταξινόμησης υφής στην οποία το συνολικό λάθος μειώνεται εξάγοντας τα διανύσματα των χαρακτηριστικών από το  $v$ .

#### 6.3.1 Πειράματα ταξινόμησης εικόνες (Brodatz)

Τα χαρακτηριστικά που αποτιμήθηκαν ως προς την ικανότητα τους για διάχριση και ταίριασμα εικόνων υφής έχουν κοινό παρονομαστή το στάδιο του γραμμικού φίλτραρισματος σε

Τάξη (αρ. δειγμάτων)	KL-distance (σχετικό %)				
	DCA	Gabor	Teager	pooled*	PCA†
‘Bark’ (5)	- 10.0	5.7	5.0	22.0	50.0
‘Brick’ (5)	- 4.8	17.0	30.0	6.9	31.0
‘Buildings’(2)	- 16.0	640.0	- 130.0	34.0	- 2.6
‘Clouds’ (2)	- 49.0	- 310.0	- 200.0	- 29.0	- 6.0
‘Fabric’ (2)	4.7	63.0	- 20.0	47.0	5.3
‘Flowers’ (3)	- 12.0	37.0	20.0	11.0	28.0
‘Food’ (3)	- 10.0	- 19.0	- 17.0	43.0	21.0
‘Grass’ (2)	- 4.0	- 150.0	47.0	26.0	57.0
‘Leaves’ (2)	- 19.0	3.3	- 18.0	3.9	24.0
‘Metal’ (4)	- 8.0	- 90.0	10.0	- 0.46	17.0
‘Misc’ (2)	- 6.3	35.0	16.0	- 62.0	55.0
‘Paintings’(3)	- 29.0	86.0	51.0	44.0	53.0
‘Sand’ (5)	- 8.3	69.0	4.5	- 100.0	8.6
‘Stone’ (3)	- 10.0	560.0	11.0	7.4	51.0
‘Terrain’ (4)	- 2.1	- 24.0	29.0	7.5	67.0
‘Tile’ (3)	- 1.0	11.0	17.0	- 26.0	- 23.0
‘Water’ (5)	- 19.0	68.0	- 2.4	42.0	84.0
‘Waldo’ (2)	- 3.4	13.0	15.0	8.9	0.87
‘Wood’ (1)	- 4.1	- 17.0	- 1.4	- 2.0	16.0
<b>Μέση τιμή</b>	- 11.0	52.0	- 7.2	4.3	28.0

Πίνακας 6.2: Μέση σχετική αλλαγή (%) της σχετικής εντροπίας (KL-distance) με εξαγωγή χαρακτηριστικών από τη συνιστώσα υφής  $v$  σε σχέση με τα χαρακτηριστικά από την αρχική εικόνα  $f$ .

πολλαπλές ζώνες συχνοτήτων. Μια αναλυτική επισκόπηση πολυκαναλικών χαρακτηριστικών για θέματα ταξινόμησης υφής μπορεί να βρεθεί στο [250].

Θεωρούμε την ταξινόμηση φωτογραφικών εικόνων υφής επιφανειών και υλικών (π.χ. το σύνολο Brodatz [40]). Πρόκειται για εικόνες χωρίς αντικείμενα που θα μπορούσαν, τουλάχιστον οπτικά, να θεωρηθούν αμιγώς εικόνες υφής. Ωστόσο, η εφαρμογή ενός αλγορίθμου απλοποίησης ή αποδόμησης οδηγεί σε δύο συνιστώσες: α) μια τμηματικά ομαλή που περιγράφει τα πρότυπα υφής μεγάλης κλίμακας, ροές υφής και αντίθεση στην εικόνα. Αυτή είναι η πληροφορία δομής της υφής (structure inside texture), σε μια ειδική περίπτωση η βαθυπερατή συνιστώσα της εικόνας. β) ένα αμιγώς μικρής-κλίμακας σήμα μεταβολής που περιγράφει τις ταλαντώσεις μηδενικού μέσου της υφής. Το παραμένον της  $u + v$  αποδόμησης, όπως στη γενική περίπτωση, μοντελοποιεί το ύδρυβο.

Οι τέσσερις εικόνες ζευγαριών υφής Brodatz στο Σχ. 6.4 μεγέθους ( $128 \times 256$  pixel) χρησιμοποιήθηκαν για δοκιμές ταξινόμησης στο [250] μαζί με «μωσαϊκά» από 5, 10 και 16 είδη υφής. Ένα παράδειγμα αποσύνθεσης για ταξινόμηση σε δύο κατηγορίες φαίνεται στο Σχ. 6.4 (α) (δ). Αυτές οι εικόνες αποτέλεσαν τα δεδομένα εισόδου στα πειράματα ταξινόμησης, στα οποία χρησιμοποιήθηκαν ταξινομητές με επίβλεψη και εκπαίδευση (supervised trained classifiers) και ομαδοποίηση χωρίς επίβλεψη (unsupervised clustering).

### 6.3.1.1 Μέθοδοι ταξινόμησης

Στις δοκιμές χωρίς επίβλεψη τα δείγματα εκπαίδευσης (train samples) λαμβάνονται από διαφορετικές περιοχές της εικόνας από τα δείγματα δοκιμών (test samples), τα οποία παρουσιάζονται στο Σχ. 6.4. Έξι παραμετρικοί και μη-παραμετρικοί ταξινομητές χρησιμοποιήθηκαν [84, 137]: Bayes/Maximum Likelihood γραμμικός (Bayes NL) και ορθογωνικός (Bayes NQ), κοντινότερου μέσου (nearest mean, NM), 1-κοντινότερου γείτονα (1-nearest neighbor, 1-NN), γραμμικού διαχωριστή Fisher (linear discriminant, FLC), μείγμα από (2) Gaussian (mixture of Gaussians, MoG) και ταξινομητής με βάση τα διανύσματα υποστήριξης (support vector classifier, SVC). Για μη-επιβλεπόμενη ταξινόμηση, επιλέχθηκε ομαδοποίηση μέσω του αλγορίθμου k-means με γνωστό εκ των προτέρων τον αριθμό των ομάδων-τάξεων. Η υλοποίηση των αλγορίθμων και οι δοκιμές έγιναν με χρήση του πακέτου λογισμικού **prtools** [85].

### 6.3.1.2 Αποτίμηση χαρακτηριστικών

Αρχικά, τα πέντε διανύσματα χαρακτηριστικών (Εν. 6.1.1) αποτιμήθηκαν ως προς την αποδοτικότητα τους για ταξινόμηση χρησιμοποιώντας την inter-to-intra απόσταση  $S_B/S_W$  (6.5, 6.6), με τις τιμές που αναφέρονται στον Πίνακα 6.3. Μεγάλες αποστάσεις υποδηλώνουν περισσότερο αποδοτικά χαρακτηριστικά, ανεξάρτητα από την επιλογή του ταξινομητή, ωστόσο η μέτρηση δεν σχετίζεται απαραίτητα και με την πραγματική απόδοση ταξινόμησης. Συγκρινούντας τις τιμές κάθετα σε κάθε στήλη παίρνουμε μια ένδειξη της δυσκολίας των εικόνων για ταξινόμηση με τα συγκεκριμένα χαρακτηριστικά (δηλαδή η D5D92 είναι η πιο σύνθετη από τα ζευγάρια και οι Nat-5v, Nat-5v2 από τις πεντάδες). Παρατηρούμε ότι τέτοιες ενδείξεις συμφωνούνε και με την δυσκολία αντιληπτικής, οπτικής κατάτμησης τέτοιων εικόνων.

Σχετικά με τα μεμονωμένα διανύσματα χαρακτηριστικών, η επίδοση αυξάνεται με την αύξηση της διάστασης (το 43-διάστατο κοινό (pooled) διάνυσμα φαίνεται καλύτερο των Gabor, PCA, και DCA διανυσμάτων διαστάσεων αντιστοίχως 40, 9 και 4). Το ενδιαφέρον είναι η αύξηση στις τιμές των αποστάσεων χρησιμοποιώντας τη εκτιμώμενη συνιστώσα υφής  $v$ . Σε όλα τα ζευγάρια πλην ενός, η σχετική μεταβολή είναι σημαντική φτάνοντας σε πολλές περιπτώσεις περισσότερο από 100% αύξηση (σημειώστε τα διανύσματα Gabor και Teager στη «δύσκολη» περίπτωση D5D92 ή τις σύνθετες περιπτώσεις Nat-5v\* στις πεντάδες). Για το διάνυσμα DCA το περισσότερο σημαντικό μεμονωμένο χαρακτηριστικό είναι με συνέπεια το μέγεθος της στιγμιαίας συχνότητας και το λιγότερο σημαντικό η τοπική τιμή της φωτεινότητας που εξάγεται από την αρχική εικόνα ή το cartoon (για τα χαρακτηριστικά αποδόμησης  $\vec{V}$ ).

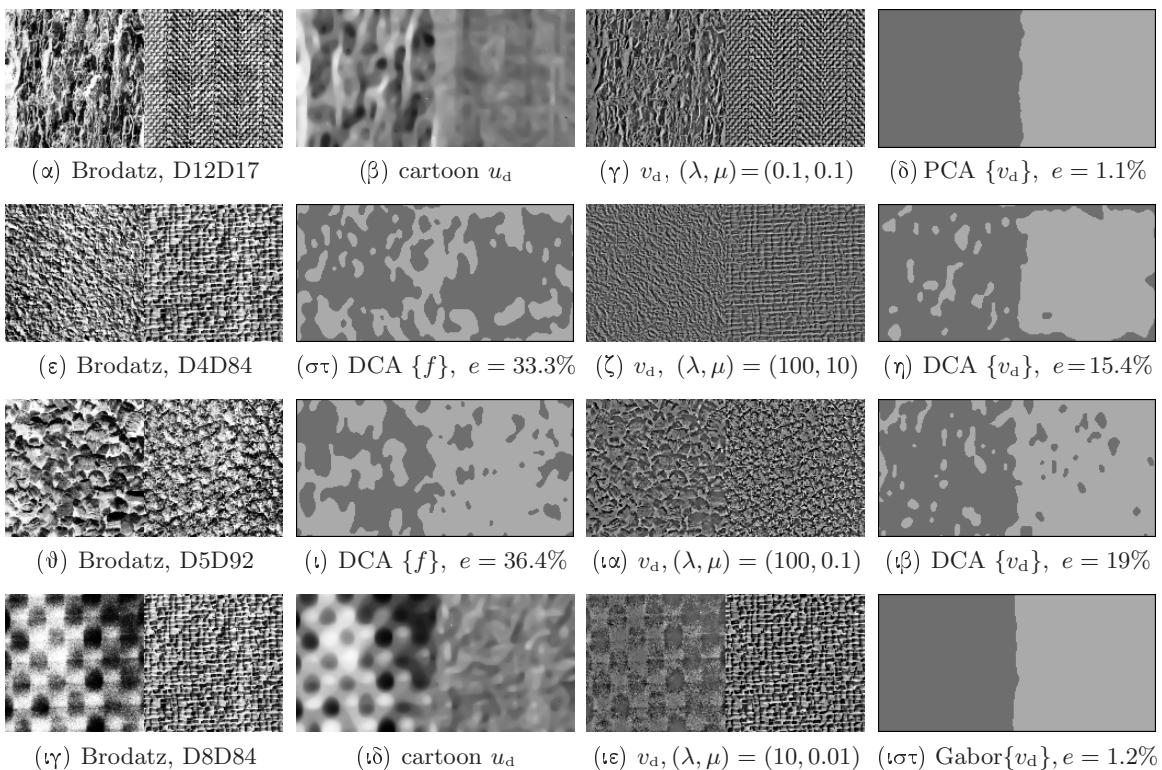
### 6.3.1.3 Δοκιμές ταξινόμησης

Οι δοκιμές ταξινόμησης πραγματοποιήθηκαν στις εικόνες αναφοράς με χρήση των χαρακτηριστικών που εξάγονται από την αρχική εικόνα και τη συνιστώσα υφής, τα οποία από εδώ και στο εξής θα καλούνται «χαρακτηριστικά στο  $f$ » και «χαρακτηριστικά στο  $v$ » αντίστοιχα. Η επίδοση ποσοτικοποιήθηκε για όλες τις μεθόδους με το μέσο σφάλμα ταξινόμησης (classification error) στο σενάριο με επισκόπηση [137]. Για το σφάλμα ομαδοποίησης (clustering

Εικόνες *	Inter-Intra Cluster Αποστάσεις									
	DCA		Gabor		Teager		pooled		PCA	
	<i>f</i>	<i>v</i>	<i>f</i>	<i>v</i>	<i>f</i>	<i>v</i>	<i>f</i>	<i>v</i>	<i>f</i>	<i>v</i>
D12D17	8.68	9.35	15.3	46.0	15.6	45.2	25.3	49.3	14.9	26.1
D4D84	1.60	6.05	10.5	46.4	10.8	44.8	11.5	52.2	9.34	25.4
D5D92	1.15	1.28	3.02	37.8	2.85	39.1	3.21	38.0	2.39	20.0
D8D84	27.8	18.7	20.2	50.3	19.5	46.6	41.3	55.9	21.8	28.6
Nat-5c	7.45	9.06	18.4	28.2	19.1	28.4	20.9	29.6	15.5	19.7
Nat-5m	5.50	4.49	9.06	24.3	8.22	24.7	11.8	24.6	9.93	14.8
Nat-5v	1.80	2.63	7.64	25.3	7.34	25.5	8.36	25.7	6.18	16.1
Nat-5v2	2.90	2.33	6.42	21.3	6.67	22.0	7.76	21.4	6.42	11.6
Nat-5v3	3.28	2.83	8.43	22.6	8.55	22.1	9.71	23.8	6.83	13.9

*f*: αρχική εικόνα, *v*: διαχωρισμένη συνιστώσα υφής, \* εικόνες [250]

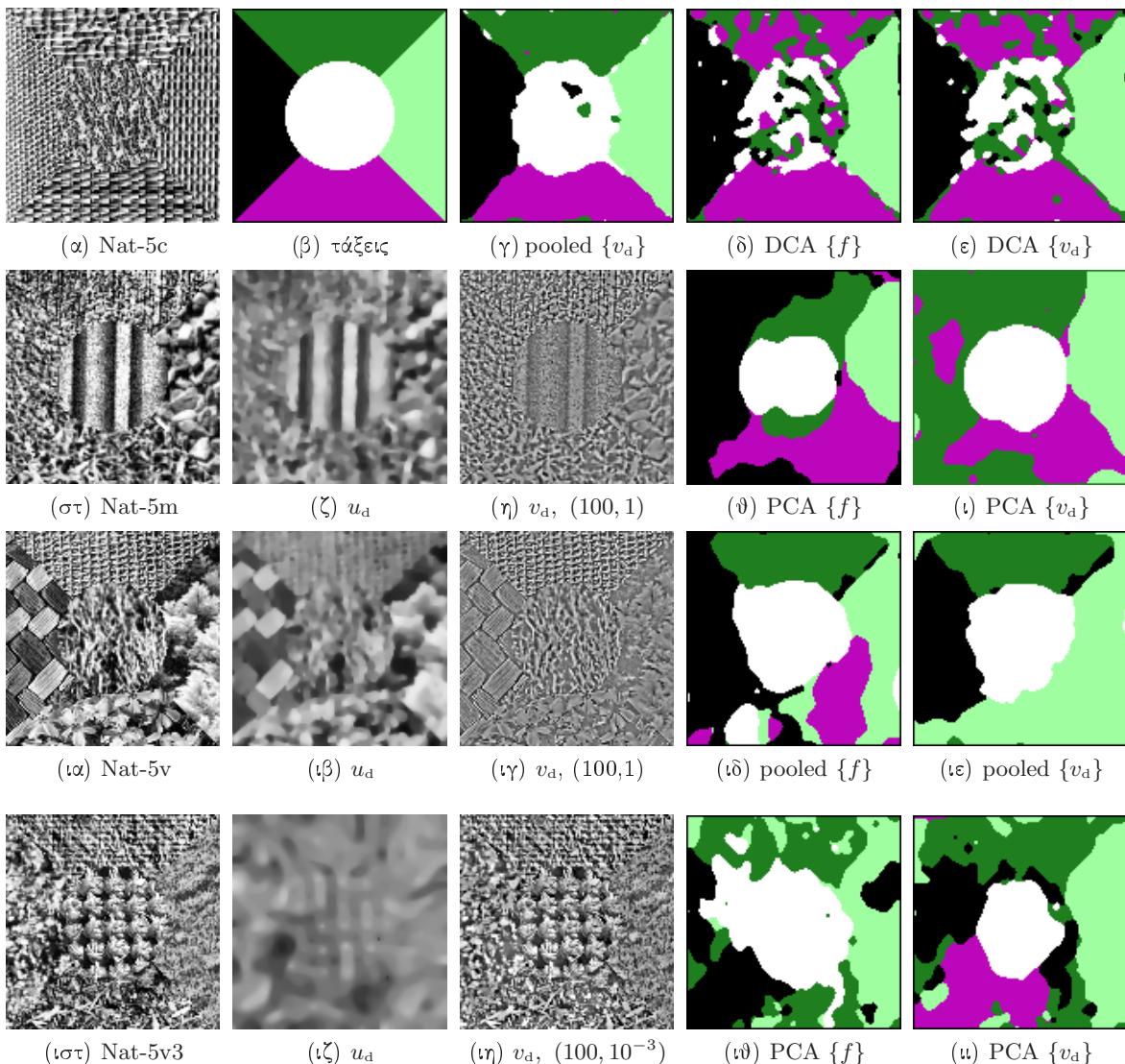
Πίνακας 6.3: Αποτίμηση χαρακτηριστικών για εικόνες 2 και 5 ειδών υφής (μωσαϊκά Brodatz).



Σχήμα 6.4: Ταξινόμηση (*k-means* ομαδοποίηση) στις εικόνες ζευγαριών υφής Brodatz ( $128 \times 256$ ). Η πάνω και η τελευταία σειρά αντιστοιχούν στα αποτελέσματα ελάχιστου σφάλματος από όλες τις τιμές παραμέτρων και διανύσματα χαρακτηριστικών. Οι μεσαίες αντιστοιχούν στη μέγιστη σχετική μείωση του σφάλματος για εξαγωγή χαρακτηριστικών από το *v*. Οι αποδομήσεις έγιναν με το *u* +  $Kv$  σχήμα και την ανακατασκευή κυρίαρχων συνιστωσών.

error) χρησιμοποιήθηκε η συνολική πιθανότητα λάθους μετά από ευθυγράμμιση των ομάδων (cluster alignment) με γνώμονα το ελάχιστο σφάλμα.

Στον Πίνακα 6.4 παρουσιάζονται τα μέσα σφάλματα για τις τέσσερις εικόνες του 6.4 με χαρακτηριστικά στο *f*, μαζί με τα σφάλματα και το ποσοστό σχετικής μεταβολής (αύξηση ή



Σχήμα 6.5: Ταξινόμηση (k-means ομαδοποίηση) σε εικόνες με 5 τάξεις υφής ( $128 \times 128$ ). Οι δύο τελευταίες στήλες αντιστοιχούν στη μέγιστη σχετική μείωση του σφάλματος για εξαγωγή χαρακτηριστικών από το  $v$ . Οι αποδομήσεις έγιναν με το  $u + Kv$  σχήμα και ανακατασκευή κυρίαρχων συνιστωσών με παραμέτρους  $(\lambda, \mu)$  που φαίνονται στις εικόνες υφής.

μείωση) χρησιμοποιώντας τα χαρακτηριστικά στο  $v$ , που λαμβάνονται με  $u + Kv$  αποδόμηση. Οι παράμετροι του διαχωρισμού  $(\lambda, \mu)$  ελέγχουν το «πέρασμα» των λεπτομερειών από τη μια συνιστώσα στην άλλη και επομένως επηρεάζουν την απόδοση ταξινόμησης. Τα σφάλματα που αναφέρονται στη συνιστώσα  $v$  στον Πίνακα 6.4 είναι τα ελάχιστα σφάλματα που επιτεύχθηκαν για κάθε ταξινομητή και διάνυσμα χαρακτηριστικών, σε ένα διακριτό εύρος από τιμές των παραμέτρων.

Οι παράμετροι  $\lambda$  και  $\mu$  αφέθηκαν εναλλακτικά να λαμβάνουν τιμές από το λογαριθμικό σύνολο  $(l, m) \in \{-3, 2\}$  με  $(l, m) = 10 \log_{10}(\lambda, \mu)$  και  $-\infty$  να αντιστοιχεί στο 0. Για  $\mu = 0$  η αποδόμηση συμπίπτει με το σχήμα ROF, ενώ για  $\lambda = 0$  με μια TV κανονικοποίηση με περιορισμούς από το μοντέλο της υφής. Τα διαγράμματα στο Σχ. 6.6 παρουσιάζουν τα μέσα σφάλματα ταξινόμησης που απαρτίζονται από τη μέση τιμή των σφαλμάτων ομαδοποίησης και

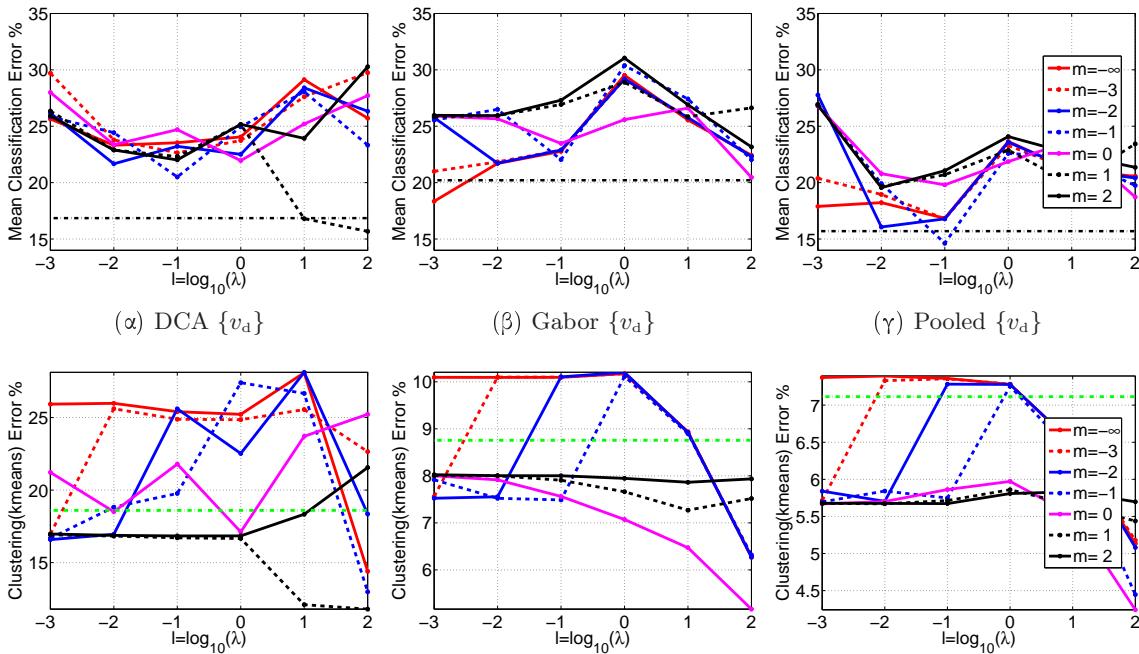
Ταξινόμηση (επιβλεπόμενη/ ομαδοποίηση)	Χαρακτηριστικά, συνιστώσες & σχετικά σφάλματα								
	DCA			Gabor			Teager		
	{f}	{v}	D	{f}	{v}	D	{f}	{v}	D
Bayes NL	<b>13.4</b>	<b>16.6</b>	24.0	<b>29.0</b>	<b>25.4</b>	-12.7	<b>28.6</b>	<b>24.9</b>	-12.9
Bayes NQ	18.2	18.1	-0.37	36.5	27.4	-24.9	37.4	28.3	-24.5
NM	17.4	16.9	-3.09	31.9	26.8	-16.1	32.2	27.1	-15.8
1-NN	13.4	16.6	24.0	29.0	25.4	-12.7	28.6	24.9	-12.9
FLD	16.4	19.5	18.3	37.1	30.7	-17.4	40.4	31.9	-21.2
MoG	13.4	16.6	24.0	29.0	25.4	-12.7	28.6	24.9	-12.9
SVC	13.4	16.6	24.0	29.0	25.4	-12.7	28.6	24.9	-12.9
μέσοι	15.1	17.3	15.8	31.7	26.6	-15.6	32.1	26.7	-16.1
<b>k-means</b>	18.6	11.8	-36.6	8.76	5.17	-41.0	9.06	6.18	-31.8

f: αρχική εικόνα, v: συνιστώσα υφής, D: σχετική μεταβολή του σφάλματος

(επιβλεπόμενη/ ομαδοποίηση)	pooled			PCA		
	{f}	{v}	D	{f}	{v}	D
Bayes NL	23.5	20.3	-13.5	49.5	39.4	-20.3
Bayes NQ	26.9	21.5	-20.0	48.6	41.1	-15.5
NM	<b>18.8</b>	<b>19.6</b>	4.42	<b>48.1</b>	<b>29.9</b>	-37.8
1-NN	23.5	20.3	-13.5	49.5	39.4	-20.3
FLD	30.2	25.7	-15.0	45.5	34.4	-24.3
MoG	23.5	20.3	-13.5	49.5	39.4	-20.3
SVC	23.5	20.3	-13.5	49.5	39.4	-20.3
μέσοι	24.3	21.2	-12.1	48.6	37.6	-22.7
<b>k-means</b>	7.11	4.24	-40.4	7.16	4.34	-39.4

Πίνακας 6.4: Σφάλματα ταξινόμησης (μέσες τιμές %) στις εικόνες με 2 κατηγορίες υφής (παράμετροι ( $\lambda, \mu$ ) ελάχιστου σφάλματος για κάθε περίπτωση ταξινομητή και διανύσματος).

του μέσου σφάλματος των εκπαιδευμένων ταξινομητών (οι δύο τελευταίες γραμμές του Πίνακα 6.4). Παρότι δεν διακρίνεται φαίνομενικά κάποια συνεπής συμπεριφορά που να καθοδηγεί μια επιλογή των παραμέτρων με βάση την απόδοση μπορούν να γίνουν μερικές παρατηρήσεις: Και για τα τρία διανύσματα, υπάρχουν πάντα σημεία στις καμπύλες που προέρχονται από το διαχωρισμένη υ συνιστώσα, στις οποίες το μέσο συνολικό σφάλμα είναι μικρότερο απ' ότι με τα χαρακτηριστικά στο f. Αυτό είναι περισσότερο εμφανές αν κοιτάξουμε τις καμπύλες για τα σφάλματα της ομαδοποίησης μόνο (κάτω σειρά), όπου πολλοί διαφορετικοί συνδυασμοί ( $\lambda, \mu$ ) μειώνουν το συνολικό σφάλμα. Το χαμηλής διάστασης DCA διάνυσμα στο v παρουσιάζει ελάχιστο για μεγάλες τιμές (100, 10), δηλαδή μεγάλη πιστότητα στο παραμένον και το μοντέλο), ενώ το Gabor στην περίπτωση ( $10^{-3}, 0$ ) που ανάγεται στο σχήμα ROF. Το κοινό διάνυσμα στο ( $\gamma$ ) ελαχιστοποιεί το σφάλμα στις μεσαίες τιμές (0.1, 0.1), η τιμή του οποίου αποτελεί επίσης την καλύτερη μέση απόδοση ταξινόμησης/ομαδοποίησης (14.6% σφάλμα στο v, 7% σχετική μείωση). Μεμονωμένη επίδοση για διαφορετικές μεθόδους ταξινόμησης με χαρακτηριστικά στο v, για τα τρία αυτά ζεύγη χαρακτηριστικών παραμέτρων φαίνονται στον Πίνακα 6.5.

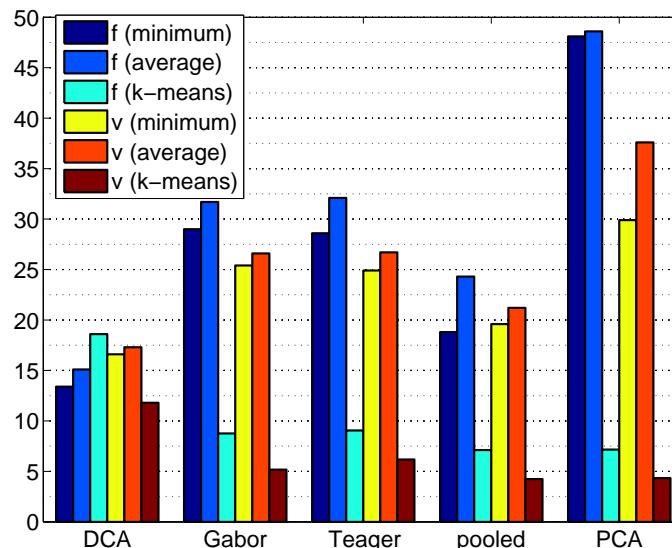


Σχήμα 6.6: Μέσα σφάλματα ταξινόμησης (ομαδοποίηση και ταξινόμηση με επιβλεψη) στα ζευγάρια του Σχ. 6.4, χρησιμοποιώντας τρία διανύσματα χαρακτηριστικών που έχουν εξαχθεί από το  $v$ , για διαφορετικές τιμές των παραμέτρων αποδόμησης  $(l, m) = 10 \log_{10}(\lambda, \mu)$ . Οι τιμές αυξάνονται λογαριθμικά κατά 10, από  $10^{-3}$  μέχρι  $10^2$  και τα διαγράμματα απεικονίζονται ως προς το  $\lambda$  με κάθε υποπαράσταση να αντιστοιχεί σε άλλη τιμή του  $\mu$ . Η οριζόντια (διακεκομένη) είναι το μέσο σφάλμα των χαρακτηριστικών στο  $f$ . κάτω σειρά: οι αντίστοιχες καμπύλες με τα σφάλματα ομαδοποίησης μόνο.

Ταξινόμηση (επιβλεπόμενη/ ομαδοποίηση)	Χαρακτηριστικά & παράμετροι $(l, m) = \log_{10}(\lambda, \mu)$								
	DCA			Gabor			pooled		
	-3,-∞	-1,-1	2,1	-3,-∞	-1,-1	2,1	-3,-∞	-1,-1	2,1
Bayes NL	24.1	21.4	<b>16.6</b>	<b>25.4</b>	33.9	44.7	27.7	20.3	39.9
Bayes NQ	27.0	22.5	26.7	27.4	43.2	51.4	31.4	26.5	51.3
NM	23.1	16.9	18.5	26.8	34.0	39.8	<b>20.0</b>	20.5	28.8
1-NN	24.1	21.4	16.6	25.4	33.9	44.7	27.7	20.3	39.9
FLD	31.1	23.7	25.2	30.7	43.5	50.0	36.6	36.1	50.4
MoG	24.1	21.4	16.6	25.4	33.9	44.7	27.7	20.3	39.9
SVC	24.1	21.4	16.6	25.4	33.9	44.7	27.7	20.3	39.9
μέσοι	25.4	21.2	19.6	26.6	36.6	45.7	28.4	23.5	41.4
<b>k-means</b>	25.9	19.8	<b>11.8</b>	10.1	<b>7.49</b>	7.52	7.37	5.75	<b>5.44</b>

Πίνακας 6.5: Σφάλματα ταξινόμησης (μέσες τιμές %) στις εικόνες με 2 κατηγορίες υφής (παράμετροι  $(\lambda, \mu)$  ελάχιστου σφάλματος για κάθε διάνυσμα).

Επιστρέφοντας στον Πίνακα 6.4 των βέλτιστων αποδόσεων για αυτό το εύρος των παραμέτρων αποδόμησης, τονίζονται ιδιαίτερα τα ελάχιστα σφάλματα για κάθε συνιστώσα (bold), το μέσο ως προς τις μεθόδους συνολικό σφάλμα ταξινόμησης και το σφάλμα ομαδοποίησης. Εκτός από το διάνυσμα DCA, η μέση επίδοση βελτιώνεται εξάγοντας πολυκαναλικά χαρακτηριστικά από το  $v$ , φτάνοντας μια σχετική μείωση του σφάλματος κατά 22.7% επί όλων των μεθόδων. Για ομαδοποίηση μέσω *k-means*, τα αποτελέσματα είναι συνεπή για όλα τα διανύ-



Σχήμα 6.7: Σφάλματα (%) ταξινόμησης, μέση και ελάχιστη τιμή, και ομαδοποίησης στις εικόνες Brodatz με δύο κατηγορίες υφής, για διαφορετικά διανύσματα χαρακτηριστικών από την εικόνα  $f$  και τη συνιστώσα  $v$ .

σματα με καλύτερη επίδοση το σφάλμα 4.24% για το κοινό (pooled) διάνυσμα και μέγιστη μείωση 41% για το Gabor διάνυσμα. Η επιβλεπόμενη ταξινόμηση μέσω DCA, παρ' ότι δεν βελτιώθηκε σε κάθε περίπτωση από την αποδόμηση, έδωσε με συνέπεια τα καλύτερα αποτελέσματα (μικρότερο σφάλμα) και για τις δύο συνιστώσες, ως προς την ταξινόμηση με εκπαίδευση για διαφορετικούς ταξινομητές. Τα συμπεράσματα αυτά συνοψίζονται συγκεντρωτικά και οπτικοποιούνται στο διάγραμμα του Σχ. 6.7.

Τα παραδείγματα στο Σχ. 6.4 δείχνουν το χαρακτηρισμό (labeling) μέσω k-means για κάθε μια από τις εικόνες δοκιμών. Αντιστοιχούν στις περιπτώσεις ελάχιστου σφάλματος (πρώτη και τελευταία σειρά) και μέγιστης σχετικής μείωσης (μεσαίες σειρές). Η τελευταία στήλη απεικονίζει τις τάξεις που προκύπτουν από το  $v$ . Παρατηρείστε τη βελτιωμένη ταξινόμηση στη δεύτερη και ειδικά στη δύσκολη τρίτη περίπτωση (D5D92) που επιτυγχάνεται μέσω χαρακτηριστικών DCA. Αντίστοιχες δοκιμές έγιναν και για τις εικόνες με τις πέντε κατηγορίες υφής (Brodatz και VisTex δείγματα). Τα συνολικά αποτελέσματα παρουσίασαν χειρότερη επίδοση για τις συγκεκριμένες μεθόδους ταξινόμησης, οι οποίες πάντως βελτιώθηκαν από την ανάλυση στη  $v$  συνιστώσα δίνοντας για την k-means ομαδοποίηση μέγιστη σχετική μείωση του σφάλματος κατά 15.9% στο κοινό διάνυσμα. Παραδείγματα της βελτίωσης φαίνονται στο Σχ. 6.5.

## 6.4 Συμπεράσματα

Η αποδόμηση εικόνων με το μοντέλο  $u + K$  νεφαρμόστηκε για την καλύτερη μοντελοποίηση και ανάλυση της συνιστώσας  $v$ . Αυτή η συνιστώσα, που προκύπτει από την απομάκρυνση των μακροδομών και των χαρακτηριστικών μεγάλης κλίμακας μιας εικόνα είναι περισσότερο κατάλληλη για τη δημιουργία περιγραφών και την αναπαράσταση χαρακτηριστικών της υφής. Το σχήμα συνέργειας εξετάστηκε προς την «αντίθετη» κατεύθυνση κατά την οποία η διαδικασία της αποδόμησης βοήθησε το στάδιο της μοντελοποίησης. Χρησιμοποιώντας το μοντέλο πολυ-

## *Κεφάλαιο 6. Συνέργεια διαχωρισμού και μοντελοποίησης: Ταξινόμηση υφής*

καναλικής ανάλυσης που εμπλέκεται και στην αποδόμηση της εικόνας εξετάστηκαν παραλλαγές διανυσμάτων χαρακτηριστικών ως προς την διάκριση, το ταίριασμα και την ταξινόμηση υφής. Η εξαγωγή χαρακτηριστικών με το πολυκαναλικό μοντέλο αποδείχτηκε αποτελεσματικότερη στη συνιστώσα  $u$ , μετά από βελτιστοποίηση των παραμέτρων αποδόμησης ( $\lambda, \mu$ ), σε όλες τις περιπτώσεις πλην της επιβλεπόμενης ταξινόμησης για τα χαρακτηριστικά κυρίαρχης διαμόρφωσης. Για μη-επιβλεπόμενη ομαδοποίηση τα αποτελέσματα ήταν συνεπή για όλα τα διανύσματα χαρακτηριστικών, με τη μέγιστη μείωση του σφάλματος να φτάνει το 41%. Παράλληλα, αναδείχθηκε και η περιγραφική ικανότητα του μειωμένης διάστασης διανύσματος κυρίαρχων διαμορφώσεων, το οποίο τόσο στην αρχική εικόνα όσο και στην εκτιμώμενη συνιστώσα υφής έδωσε κατά πολύ τα μικρότερα σφάλματα ταξινόμησης.

□

## Κεφάλαιο 7

# Εφαρμογές στην Ανάλυση Εικόνων Εδαφοτομών

Μια κατηγορία εικόνων με έντονη υφή και παρουσία διαφορετικών δομών είναι οι ψηφιακές εικόνες λεπτών τομών εδάφους ή εδαφοτομών (soil sections). Η αυτόματη επεξεργασία τέτοιων εικόνων μπορεί να αποκαλύψει στοιχεία της δομής του εδάφους και να εξάγει εκτιμήσεις βιοοικολογικής 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<sup>1</sup>.

### 7.1 Περιγραφή εδαφοτομών και ανάλυση εδάφους

Ο σκοπός της μικρομορφολογικής ανάλυσης (soil micromorphology), ως χλάδος 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 [44]. Η μελέτη της δομής του εδάφους αφορά στο μέγεθος, το σχήμα, την οξύτητα, την αντίθεση, τη συχνότητα εμφάνισης και τη χωρική τοποθεση 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

<sup>1</sup>Η έρευνα και τα αποτελέσματα που παρουσιάζονται σε αυτή την ενότητας εντάσσονται στο πλαίσιο των εργασιών [268, 269] σε συνεργασία με την Νατάσα Σοφού.

που χρησιμοποιείται.

Οι εικόνες εδαφοτομών, οι οποίες 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 1-Δ όπως είναι οι ακμές ή οι καμπύλες, είτε 2-Δ όπως φωτεινά ή σκοτεινά blobs (μικρές ομογενείς περιοχές τυχαίου σχήματος), την χωρική τοποθέτηση και απόσταση συστατικών και την υφή τους.

Η μέχρι τώρα αυτόματη ανάλυση εικόνων εδαφοτομών περιορίζεται σε κατωφλιοποίησεις, ευρετικά, μετρήσεις ιστογραμμάτων, απλές τεχνικές ψηφιακής επεξεργασίας και εκτιμήσεις φράκταλ διάστασης [43,68,246,282]. 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 ικανού να αναλύει και να εκτιμά την γονιμότητα και την βιοοικολογική ποιότητα του εδάφους μέσω των εικόνων τομών [203]. Έτσι, η ακριβή και χρονοβόρα διαδικασία της παραδοσιακής βιοχημικής ανάλυσης να αυτοματοποιηθεί, να ενισχυθεί και πιθανόν να συνδυαστεί με τεχνικές όρασης.

Η κατάτμηση εικόνων βρίσκεται στον πυρήνα της ανάλυσης των εικόνων 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 εμβαδόν, ο προσανατολισμός, η συγκέντρωση, η επιμήκυνση, το μέγεθος, η καμπυλότητα κ.ά.

## 7.2 Υφή εδαφοτομών, ανάλυση και κατάτμηση

### 7.2.1 Μορφολογική κατάτμηση

Από το πλήθος μεθόδων κατάτμησης εικόνων ο μορφολογικός μετασχηματισμός πλημμυρισμού (watershed transform) έχει αποδειχθεί πολύ ισχυρός και αποτελεσματικός ιδιαίτερα μέσω σύζευξης με μη-γραμμικούς μορφολογικούς τελεστές [32,302]. 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 υεωρείται μια τοπογραφική επιφάνεια βυθισμένη στο νερό. Οι σημαδευτές παιζουν το ρόλο των πηγών πλημμυρισμού από 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) [302] και ιεραρχικές ουρές προτεραιότητας (hierarchical queues) [32].

Εκτός από τη μορφολογική διαδικασία πλημμυρισμού, το watershed μοντελοποιήθηκε επίσης με συνεχή τρόπο μέσω της eikonal Μερικής Διαφορικής Εξίσωσης (ΜΔΕ) [211, 222], χρησιμοποιώντας ιδέες από το πεδίο της εξέλιξης καμπυλών [199]. 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 αναδυόμενου κύματος αντικειτωπίζεται ως καμπύλη η οποία εξελίσσεται με προκαθορισμένη ταχύτητα. Συγκεκριμένα το σύνορο του σημαδευτή θεωρείται ως μια ομαλή, κλειστή καμπύλη  $\vec{C}(p, t)$  όπου το  $p \in [0, 1]$  παραμετροποιεί την καμπύλη και  $t$  μια τεχνητή παράμετρος εξέλιξης. Η ΜΔΕ που υλοποιεί τον watershed πλημμυρισμό είναι

$$\frac{\partial \vec{C}}{\partial t} = \frac{c}{A(t)\|\nabla I\|} \cdot \vec{N}, \quad (7.1)$$

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) [270].

Σύμφωνα με την προσέγγιση μέσω επιπεδοσυνόλων [226], όπου η εξελισσόμενη καμπύλη εμβυθίζεται ως το μηδενικό επιπεδοσύνολο  $\Gamma(t) = \{(x, y) : \Phi(x, y, t) = 0\}$  μιας χωροχρονικής συνάρτησης μεγαλύτερης διάστασης  $\Phi(x, y, t)$ , η ΜΔΕ που διέπει την εξέλιξη είναι

$$\frac{\partial \Phi}{\partial t} = \frac{c}{A(t)\|\nabla I(x, y)\|} \|\nabla \Phi\|. \quad (7.2)$$

Αποτελεσματικοί αλγόριθμοι για την επίλυση χρονικά εξαρτώμενων eikonal ΜΔΕ είναι η μέθοδος επιπεδοσυνόλων στενής ζώνης (narrow-band level sets) και η μέθοδος ταχείας πρόλασης (fast marching) [264].

### 7.2.2 Συζευγμένη, πολυκριτιριακή κατάτμηση εδαφοτομών

Με βάση την αποδόμηση μιας εικόνας στις δύο συνιστώσες της  $u, v$  και την ανάγκη για εξειδικευμένη επεξεργασία της κάθε συνιστώσας ξεχωριστά αναπτύχθηκε ένα σχήμα πολυκριτιριακής κατάτμησης, με σύζευξη πληροφορίας γεωμετρίας και υφής. Οι μεταβολές της αντίθεσης (ή του μεγέθους) λαμβάνονται υπόψιν μέσω του  $u$  κομματιού και οι ταλαντώσεις υφής περιγράφονται από τη  $v$  συνιστώσα. Η συμβολή και η ποσοτικοποίηση της υφής γίνεται μέσω της διαδικασίας ανίχνευσης και μέτρησης της μέγιστης μέσης Teager ενέργειας ή ενέργειας ταλαντώσεων υφής  $\Psi_{\text{mat}}$  που ορίστηκε στην ενότητα 3.5 και δίνεται από την (3.22). Με τον τρόπο αυτό συμπεριλαμβάνεται στη διαδικασία της μεταβολικής κατάτμησης ένα μετρικό που σχετίζεται με το είδος και την ένταση της υφής στις περιοχές της εικόνας.

Η προτεινόμενη διαδικασία κατάτμησης βασίζεται στην εξέλιξη καμπύλης που καθορίζει τη ακόλουθη ΜΔΕ :

$$\frac{\partial \vec{C}}{\partial t} = \left( \frac{\lambda_1}{A(t) \|\nabla f_1\|} + \lambda_2 \Psi_{\text{mat}}(f_2) \right) \vec{N}, \quad (7.3)$$

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 δύο eikonal όρους, οι οποίοι 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. Ο δεύτερος όρος αντιστοιχεί σε ροή που μεγιστοποιεί την ενέργεια διαμόρφωσης υφής

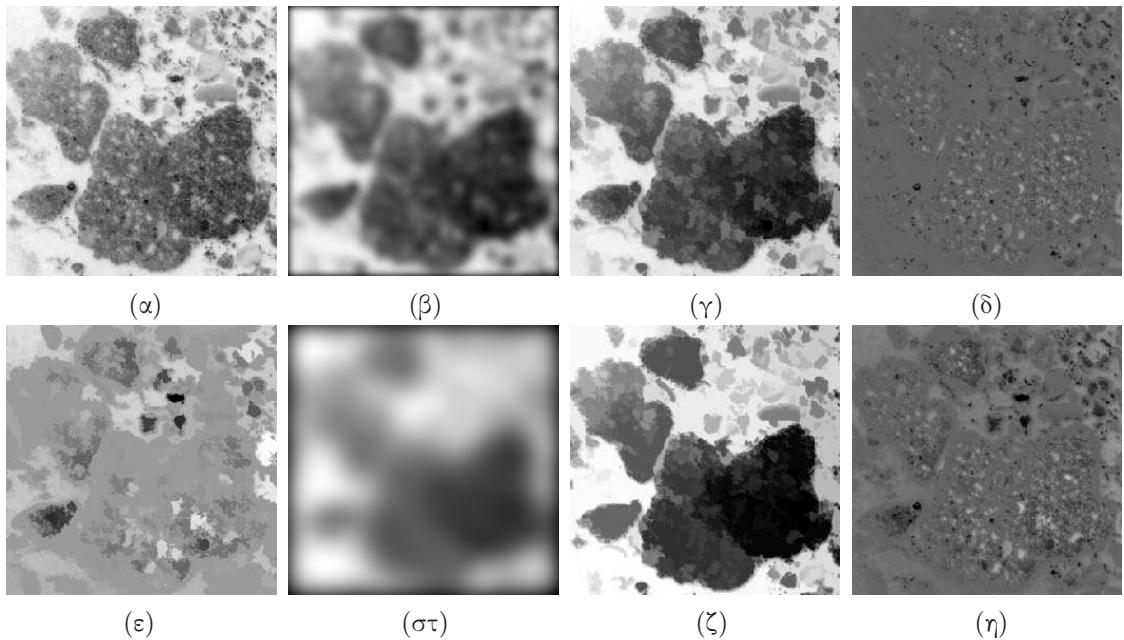
$$\max \iint_{R(C)} \Psi_{\text{mat}}(f) \implies \partial \vec{C} / \partial t = \Psi_{\text{mat}}(f) \vec{N}, \quad (7.4)$$

με αποτέλεσμα να ωθεί την καμπύλη σε περιοχές με έντονη παρουσία υφής. Ακολουθώντας την θεώρηση επιπεδοσυνόλων [226], η συνάρτηση εμβυθίζεται ως το μηδενικό επιπεδοσύνολο μιας χώρο-χρονικής συνάρτησης  $\Phi(x, y, t)$ . Εφόσον η ΜΔΕ είναι καθαρού eikonal τύπου, η υλοποίηση βασίζεται στη μέθοδο fast marching [264] που εξασφαλίζει υπολογιστική ταχύτητα.

Με βάση τη ΜΔΕ (7.3) μπορούν να προκύψουν διαφορετικά σενάρια επιλέγοντας τα σήματα  $f_1$  και  $f_2$ . Η πιο προφανής επιλογή είναι με  $f_1 = I$ ,  $f_2 = I$  που αντιστοιχεί σε εξέλιξη με ταχύτητα αντιστρόφως ανάλογη της αντίθεσης φωτεινότητας και ανάλογη της  $\Psi_{\text{mat}}$  ενέργειας της εικόνας. Στο πρωτότυπο σενάριο  $f_1 = u$ ,  $f_2 = v$ , η εξέλιξη της καμπύλης είναι αντιστρόφως ανάλογη της αντίθεσης φωτεινότητας (ή μεγέθους) της cartoon συνιστώσας και ανάλογη της  $\Psi_{\text{mat}}$  ενέργειας της συνιστώσας υφής της εικόνας. Αυτό το δεύτερο σενάριο παρουσιάζει ιδιαίτερο ενδιαφέρον και υιοθετείται καθώς ενοποιεί πληροφορία ακμών και υφής, συνδυάζοντας τα διαφορετικά σήματα που προκύπτουν από τον  $u + v$  διαχωρισμό της εικόνας. Σκόπιμα αποφεύγεται η αναφορά ως τώρα σε κάποιο συγκεκριμένο σχήμα διαχωρισμού, καθώς για την αποδόμηση μπορεί να χρησιμοποιηθεί είτε το  $u + K u$  σχήμα που έχει αναπτυχθεί (Εν. 5.3), είτε οποιοδήποτε από τα υπόλοιπα δοκιμασμένα σχήματα (π.χ. Rudin-Osher-Fatemi, Vese-Osher).

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

Στο προτεινόμενο σχήμα υπάρχουν πολλαπλές καμπύλες που πρόκειται να εξελιχθούν ταυτόχρονα και αρχικοποιούνται ως περιγράμματα ενός συνόλου σημαδευτών (markers), ενδεικτικών των σημαντικών ή(και) ομογενών περιοχών της εικόνας που μπορεί, εναλλακτικά, 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  $\Psi_{\text{mat}}$ , εντοπίζοντας περιοχές υψηλής ενέργειας υφής. (iii) συνδυασμός κριτηρίων αντίθεσης και υφής και (iv) «χειροκίνητα» (manually) τοποθετημένοι με επίβλεψη σε περιοχές



Σχήμα 7.1: Πολυκλιμακωτά leveling cartoon για  $u + v$  αποδόμηση. πάνω σειρά: (α) λεπτομέρεια εδαφοτομής (αναφορά), (β) Gauss σημαδευτής 1 ( $\sigma = 4$ ), (γ) leveling 1 ( $u_1$ ), (δ) διαφορά 1 ( $v_1 = f - u_1 + 100$ ). κάτω σειρά: (ε) διαφορά  $u_1 - u_2$ , (στ) Gauss σημαδευτής 2 ( $\sigma = 12$ ), (ζ) leveling 2 ( $u_2$ ), (η) διαφορά 2 ( $v_2 = f - u_2 + 100$ ).

ενδιαφέροντος.

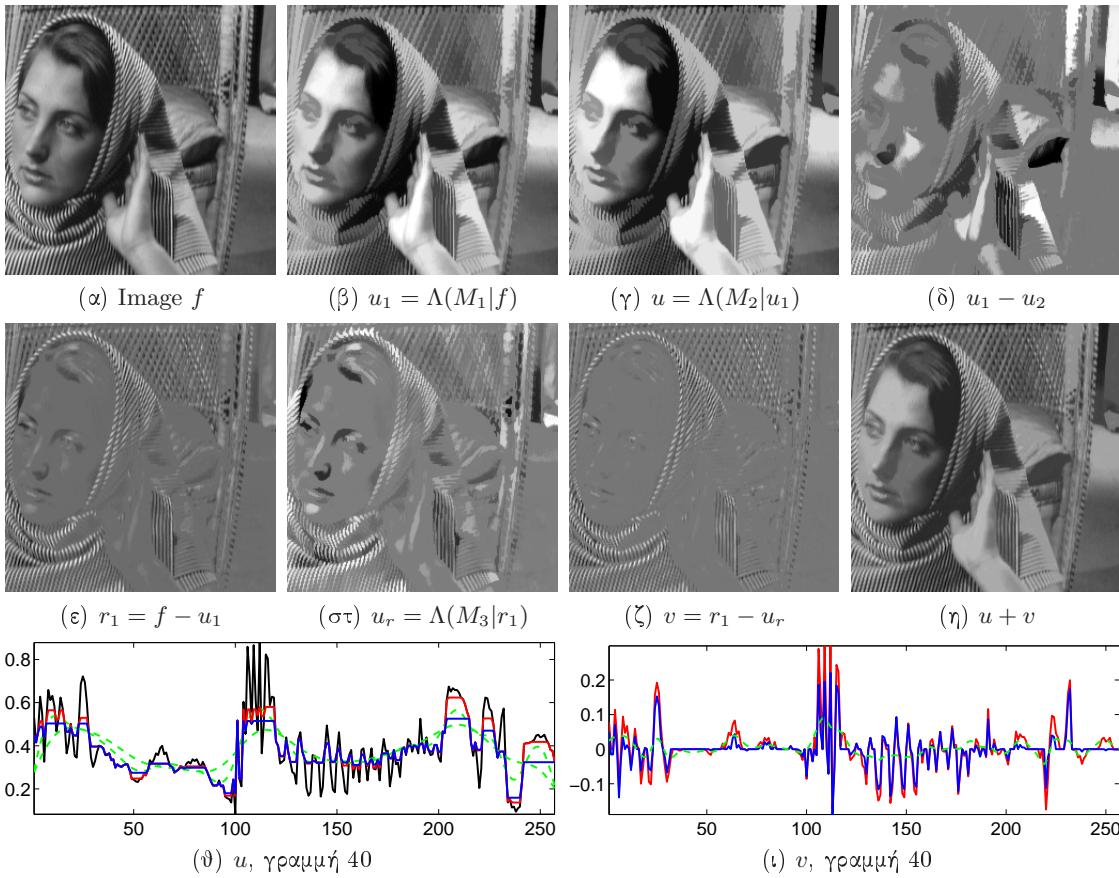
### 7.3 Διαχωρισμός υφής με levelings

Στο κεφάλαιο 5 έγινε εκτενής αναφορά στη  $u + v$  οικογένεια μοντέλων. Ένας εναλλακτικός διαχωρισμός της συνιστώσας υφής  $v$  από τις μακροδομές δίνεται μέσω απλοποίησης της εικόνας από μια κατηγορία μη-γραμμικών, αντικειμενοστραφών φίλτρων που διατηρούν τα περιγράμματα· τα levelings [268, 269], τα οποία βασίζονται στη χρήση μορφολογικών τελεστών [212]. Η απλοποίηση της  $f$  γίνεται επεκτείνοντας και συστέλλοντας τοπικά μια αρχική εικόνα, τον σημαδευτή (marker)  $M$ , και περιορίζοντας την εξέλιξη του σε καθολικό επίπεδο με την εικόνα αναφοράς  $f$ . Πιο συγκεκριμένα, επαναλήψεις του τελεστή εικόνων  $\lambda(F|f) = (\delta(F) \wedge f) \vee \varepsilon(F)$ , όπου  $\delta(F)$  και  $\varepsilon(F)$  είναι τελεστές dilation και erosion αντίστοιχα μιας εικόνας  $F$  με ένα μικρό δίσκο [262], δίνουν στο όριο το leveling της  $I$  σε σχέση με τον  $M$ , που συμβολίζεται ως

$$\Lambda(M|f) = \lim_{k \rightarrow \infty} F_k, \quad F_k = \lambda(F_{k-1}|f), \quad F_0 = M. \quad (7.5)$$

Τα levelings έχουν πολλές ενδιαφέρουσες ιδιότητες όπως ότι διατηρούν τις ακμές, δεν εισάγουν νέα ελάχιστα ή μέγιστα περιοχής και αποτελούν αυξανόμενα 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 σημαδευτή, με την εικόνα αναφοράς να παίζει το ρόλο ενός καθολικού περιορισμού.

Με γνώμονα τις παραπάνω ιδιότητες, μια εναλλακτική μέθοδος για  $u + v$  διαχωρισμό μιας εικόνας προκύπτει υιοθετώντας ως cartoon συνιστώσα το leveling της εικόνας και ως



Σχήμα 7.2: Πολυκλιμακωτά leveling cartoon και  $u + v$  αποδόμηση. πάνω σειρά: (α) εικόνα  $f$  (αναφορά), (β) leveling 1 ( $u_1$ ) με σημαδευτή  $f * g$ , όπου  $g$  μια Gauss απόκριση ( $\sigma_1 = 5$ ), (γ) cartoon/leveling 2 ( $u_2$ ), με ( $\sigma_2 = 10$ ), (δ) διαφορά ( $u_2 - u_1$ ). κάτω σειρά: (ε) παραμένον  $r_1 = f_1 - u_1$ , (στ) Leveling του  $r_1$  με Gauss σημαδευτή ( $\sigma_3 = \sigma_1/2$ ), (ζ) υφή / παραμένον ( $v = r_1 - u_r$ ), (η) ανακατασκευή ( $v + u$ ), (θ) προφίλ (σειρά 40) των cartoon (μάυρο:  $f$ , κόκκινο:  $L_1$ , μπλε:  $u = L_2$ , πράσινο: markers  $M_1, M_2$ ), (ι) προφίλ (σειρά 40) λεπτομερειών (κόκκινο: παραμένον  $r_1$ , μπλε: υφή  $v$ , πράσινο: marker  $M_3$ ).

συνιστώσα υφής τη διαφορά του από αυτή, δηλαδή

$$u \triangleq \Lambda(M|f), \quad v \triangleq f - u. \quad (7.6)$$

Για το  $u$  η επιλογή του σημαδευτή  $M$  προσφέρει ευελιξία για το τελικό leveling και μπορεί να οριστεί μέσω ανάλυσης σε πολλαπλές χλίμακες. Συγκεκριμένα κατασκευάζονται πολυκλιμακωτά leveling cartoon  $u_i = \Lambda(M_i|u_{i-1})$ ,  $i = 1, 2, 3, \dots$ , με  $u_0 = f$  από μια ακολουθία σημαδευτών  $M_i$ , που λαμβάνονται δειγματοληπτώντας μια Gaussian πυραμίδα (scale-space). Τα πολυκλιμακωτά cartoon διαμέτουν την ιδιότητα της αιτιατότητας, καθώς το  $u_j$  είναι leveling του  $u_i$  για  $j > i$ . Τα αντίστοιχα παραμένοντα  $v_i = f - u_i$  συνιστούν μια ιεραρχία από συνιστώσες υφής σε πολλαπλές χλίμακες.

Εναλλακτικά για την επιλογή των σημαδευτών εξετάστηκε και η χρήση ανισοτροπικής διάχυσης [239], όπου σε κάθε βήμα της ακολουθίας ο leveling σημαδευτής λαμβάνεται από μια εκδοχή της εικόνας με θολωμένες περιοχές αλλά αρκετά καλά διατηρημένα σύνορα, λόγω της διαδικασίας περιορισμένης διάχυσης. Στο Σχ. 7.3 φαίνονται δύο από τα στάδια της leveling πυραμίδας, στα οποία οι ακμές και τα σύνορα των κυρίαρχων γεωμετρικών δομών της εικόνας

διατηρούνται στις cartoon συνιστώσες σε διαφορετικές κλίμακες.

Οι συνιστώσες της υφής μπορούν να βελτιωθούν χρησιμοποιώντας ανάλογο φιλτράρισμα και στη διαφορά  $r_1 = f - u_1$ . Συγκεκριμένα, παίρνοντας το leveling του  $r_1$  με marker μια μικρή κλίμακα της πυραμίδας  $\{M_r(\sigma_r) : \sigma_r < \sigma_i\}$ , π.χ.  $\sigma_r = 0.5\sigma_1$ , κρατάμε το νέο παραμένον ως υφή καταλήγοντας σε μια περισσότερο εκλεπτυσμένη εκτίμηση του  $v$

$$u \triangleq \Lambda(M|f), \quad v \triangleq f - u - \Lambda(M_r|f - u). \quad (7.7)$$

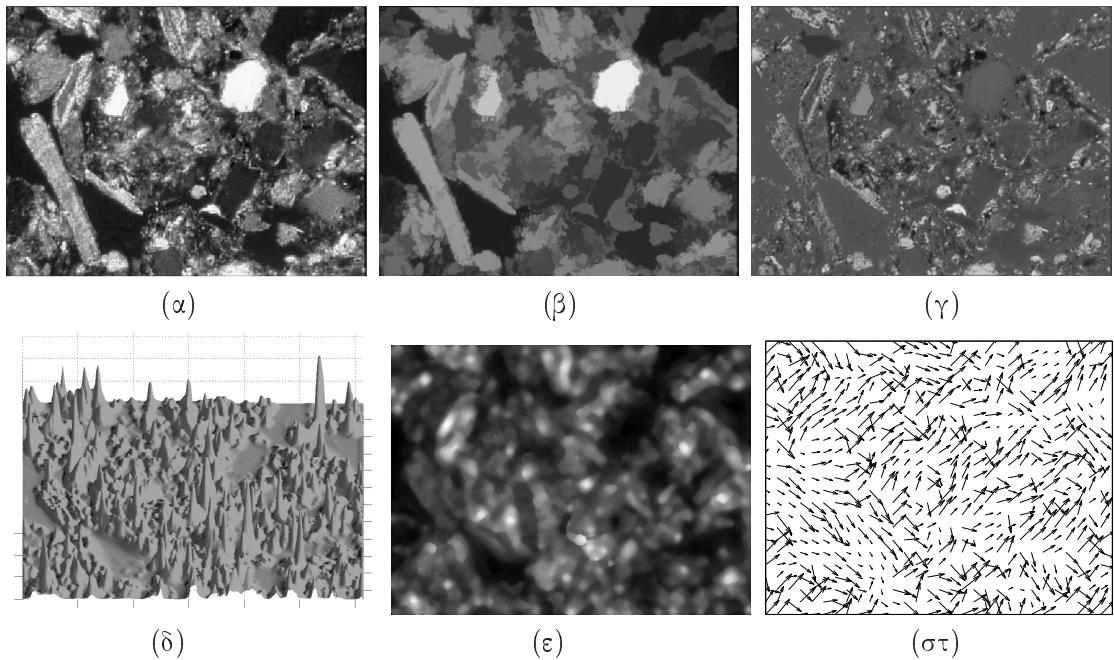
Η διπλή αυτή leveling διαδικασία αφαιρεί από το  $v$  υπόλοιπα δομής μικρότερης κλίμακας από την κλίμακα που κρατάει το leveling στο  $u$  κομμάτι. Η βελτίωση του κομματιού της υφής μπορεί να παρατηρηθεί στο Σχ. 7.2. Έκτος από Gauss σημαδευτές μπορεί να κανείς να χρησιμοποιήσει περισσότερο εξειδικευμένους markers υφής. Μια επιλογή είναι η ένδειξη για τις περιοχές υφής που προσφέρει η κυρίαρχη ενέργεια ταλαντώσεων υφής.

Συγκρίνοντας τα leveling cartoon με τα αυτά που λαμβάνονται με την επίλυση του προβλήματος TV ελαχιστοποίησης 5.3 μπορούν να αναφερθούν τα εξής: (α) Τα levelings διατηρούν τα ακρότατα των περιοχών και δεν δημιουργούν νέα ενώ τα TV cartoon 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 cartoon ευελιξία και πολυκλιμακωτές δυνατότητες σε σχέση με τις απλές σταθερές κανονικοποίησης που ελέγχουν την κλίμακα του TV cartoon.

## 7.4 Αποτελέσματα κατάτμησης και συγχρίσεις

Οι μέθοδοι για διαχωρισμό, ανίχνευση και ανάλυση υφής μαζί το κοινό σχήμα κατάτμησης εφαρμόστηκαν σε δύο κατηγορίες εικόνων εδαφοτομών: μεγάλης κλίμακας από σάρωση, ανάλυσης 47 pixel/μμ, όπως η λεπτομέρεια του Σχ. 7.1 και μικρής κλίμακας, από μικροσκόπιο πολωμένου φωτός σε ανάλυση 1 pixel/μμ. Στο Σχ. 7.3 παρουσιάζεται η διαδικασία διαχωρισμού, ανάλυσης της υφής και εξαγωγής κυρίαρχων χαρακτηριστικών της μέσω πολυκαναλικού εντοπισμού της ενέργειας διαμόρφωσης. Στο Σχ. 7.3(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 cartoon συνιστώσα  $u$  και συνιστώσα υφής  $v$ , στα (β) και (γ) αντίστοιχα, γίνεται με χρήση του δευτέρου επιπέδου της πυραμίδας leveling από Gauss σημαδευτές ( $\sigma_1 = 4$ ,  $\sigma_2 = 8$ ).

Η καλά εντοπισμένη μέτρηση ενέργειας στο (γ) είναι ενδεικτική των ισχυρών συγκεντρώσεων και μεταβολών της υφής 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  $3 \times 3$ , ενώ η φασματική εξαρτάται από το εύρος των φίλτρων. Τα χαρακτηριστικά κυρίαρχων διαμορφώσεων της υφής εξάγονται μέσω EDCA. Το κυρίαρχο πλάτος που παρουσιάζεται στο (δ) μετά από median φιλτράρισμα είναι ομοιόμορφο σε περιοχές ανάλογης έντασης και αντίθεσης. Στο (ε) φαίνεται



Σχήμα 7.3: Ανάλυση και χαρακτηριστικά υφής εδαφοτομών: (α) εικόνα εδαφοτομής ( $768 \times 567$  pixels) ψηφιοποιημένη σε  $1\text{pixel}/\mu\text{m}$  (γκρίζα κλίμακα), (β) cartoon συνιστώσα  $U$ , (γ) συνιστώσα υφής  $v$ , (δ) προοπτική 3-Δ απεικόνιση της ενέργειας διαμόρφωσης υφής  $\Psi_{\text{mat}}$ , (ε) κυρίαρχο πλάτος, (στ) διανύσματα προσανατολισμού κυρίαρχων συχνοτήτων. Τα (δ)-(στ) έχουν εξαχθεί από την συνιστώσα.

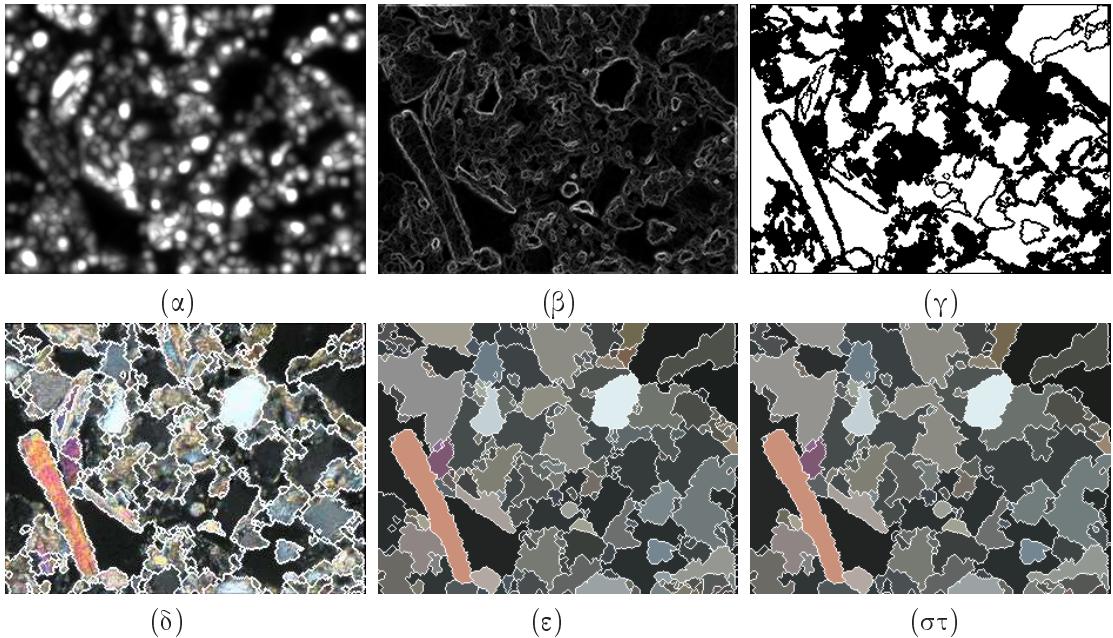
Ο προσανατολισμός του διάνυσματος κυρίαρχων ή τοπικά αναδυόμενων συχνοτήτων, ο οποίος αναδεικνύει την κατεύθυνση μεταβολής της φάσης των ταλαντώσεων υφής στην εικόνα. Όλες οι μετρήσεις έχουν γίνει στην  $u$  συνιστώσα της εικόνας (γ) μετά τον leveling διαχωρισμό.

Μετά τον μονοχρωματικό  $u + v$  διαχωρισμό και την εξαγωγή των χαρακτηριστικών της υφής, στο Σχ. 7.3(α)-(δ), ένα κατάλληλο σύνολο από αυτόματα αποκτημένους σημαδευτές, που φαίνεται στο Σχ. 7.4(γ), εξελίσσεται με βάση τη ΜΔΕ (7.3), όπου  $f_1 = u$  και  $f_2 = v$ . Λόγω της πολυ-δομικής φύσης τέτοιων εικόνων, τα αρχικά αποτελέσματα κατάτμησης στο Σχ. 7.4 (δ)-(ε) 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 Σχ. 7.4(στ) με καλύτερα ομαδοποιημένες δομές (ειδικά μικρομάζας) και μικρότερο αριθμό τελικών περιοχών.

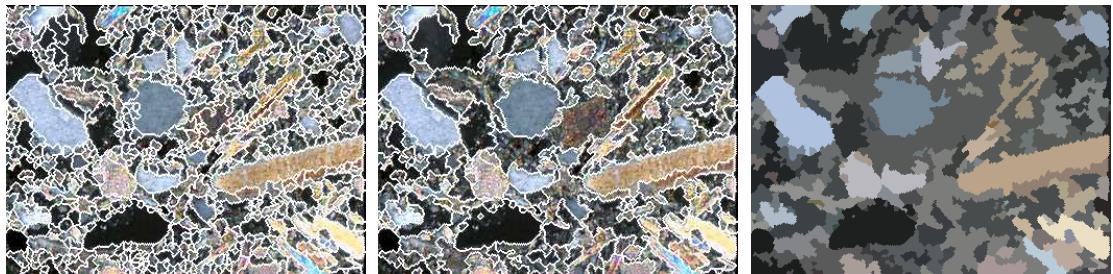
#### 7.4.1 Βελτίωση με συγχώνευση περιοχών

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

Η διαδικασία συγχώνευσης περιοχών βασίζεται στον υπολογισμό της απόστασης των μέ-



Σχήμα 7.4: Κατάτμηση της εδαφοτομής του 7.3(α) με συνδυασμό κριτηρίων: (α) ενέργεια διαμορφώσεων υφής  $\Psi_{\text{mat}}(v)$ , (β) μέτρο gradient του  $u$  (Σχ. 7.3(β)), (γ) σημαδευτές, (δ) συζευγμένη κατάτμηση ( $A(t) = 1$ ), (ε) περιοχές κατάτμησης, (στ) βελτιωμένες περιοχές μετά από ένωση εκ των υστέρων (region post-merging). Από το [269].



Σχήμα 7.5: Βελτίωση των αποτελεσμάτων κατάτμησης με ένωση περιοχών. Το αρχικό αποτέλεσμα, (αριστερά εικόνα) του συζευγμένου σχήματος, βελτιώνεται με ομαδοποίηση με βάση το χρώμα και τις συνιστώσες γεωμετρίας και υφής (μεσαία εικόνα) Οι τελικές ομαδοποιημένες περιοχές κατάτμησης φαίνονται δεξιά.

σων των χαρακτηριστικών ανάμεσα σε γειτονικές περιοχές. Τέτοια χαρακτηριστικά μπορεί να είναι το χρώμα, η ένταση, η αντίθεση των συνιστώσων cartoon και υφής 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 Ευκλείδια απόσταση

$$D_{R_1 R_2} = \left( \sum_j w_j |\mu_{1j} - \mu_{2j}|^2 \right)^{1/2}, \quad (7.8)$$

όπου  $\mu_{ij}$  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  $\mu_{ij}$  και τη μεταβλητή  $\sigma_{ij}^2$  κάθε περιοχής  $i$  και για κάθε χαρακτηριστικό  $j$ . Αναζητείται το ζεύγος ελάχιστης απόστασης με περιορισμό να ικανοποιείται

το τροποποιημένο Fisher χριτήριο

$$F = \text{rank}_k[F_j] = \text{rank}_k \left[ \frac{(N_1 + N_2)\sigma_{0j}^2}{N_1\sigma_{1j}^2 + N_2\sigma_{2j}^2} - 1 \right] > c, \quad (7.9)$$

όπου το  $\text{rank}_k[\cdot]$  συμβολίζει την  $k$ -οστή διάταξη ενός συνόλου,  $N_i$  είναι ο αριθμός των σημείων της περιοχής  $i = \{1, 2\}$ ,  $F_j$  είναι ο περιορισμός για τη μεταβλητή της  $\sigma_{0j}^2$  του χαρακτηριστικού  $j$  στην ενωμένη περιοχή και  $c$  ένα κατώφλι-σταθερά (συνήθως τίθεται 0). Αν το χριτήριο (7.9) ικανοποιείται για το ζεύγος ελάχιστης απόστασης, οι περιοχές ενώνονται δίνοντας μια υπερπεριοχή, μικρής μέσης απόστασης χαρακτηριστικών και μικρής μεταβλητήτας συγκρινόμενης με τις αρχικές. Σε αντίθετη περίπτωση, το ζεύγος της επόμενης μικρότερης απόσταση ελέγχεται και ούτω καθεξής.

Μια σταθερή ελάχιστη επιτρεπόμενη απόσταση ένωσης επιλέγεται κατά την αρχικοποίηση, χρησιμοποιώντας ένα από τα χαμηλότερα κέντρα του ιστογράμματος, για να εγγυηθεί η σύγκλιση της διαδικασίας και να αποτραπεί η ένωση «ανομοιόμορφων» περιοχών. Στο (7.5) φαίνεται ένα παράδειγμα της διαδικασίας ένωσης για ένα αποτέλεσμα κατάτμησης μιας εικόνας. Τα χαρακτηριστικά που χρησιμοποιήθηκαν είναι το χρώμα και η ένταση των  $u$  και  $v$  συνιστώσων ξεχωριστά, με κατώφλι του χριτηρίου (7.9) τιμής  $c = 0.5$  και αυτόματη σύγκλιση.

#### 7.4.2 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. Οποιαδήποτε τέτοια σύγκριση παρουσιάζει δύσκολίες καθώς η αποδοτικότητα της διαδικασίας κατάτμησης, απουσία δεδομένων ελέγχου, εξαρτάται κυρίως από την εφαρμογή και δεν υπάρχει ένας οικουμενικά βέλτιστος αλγόριθμος που να μπορεί να χρησιμοποιηθεί ως αναφορά σε κάθε περίπτωση. Η έλλειψη ενός ολικού χριτηρίου ποιότητας για την αποτίμηση των διαφορετικών μεθόδων, παράλληλα με την απουσία δεδομένων από ανθρώπινη κατάτμηση εικόνων εδαφοτομών κάνει δύσκολη την ποσοτική αποτίμηση των αποτελεσμάτων.

Εκτός από απλή επισκόπηση του τελικού διαχωρισμού των περιοχών, ως μέσα ποσοτικοποίησης των αποτελεσμάτων και συγκρίσεων χρησιμοποιήθηκαν σε αυτή την περίπτωση κάποια εμπειρικά χριτήρια ποιότητας. Επιλέχθηκαν τα ακόλουθα μετρικά ελαχιστοποίησης για τα οποία μικρότερες τιμές υποδηλώνουν καλύτερο αποτέλεσμα κατάτμησης:

- A. Συνάρτηση ολικού κόστους (*Liu Yang Global Cost*) (LYGC) [181], ένας συγκερασμός (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 Ευκλείδια απόσταση ανάμεσα στην αρχική και στην κατατμημένη περιοχή για κάθε σημείο και  $A_{R_i}$  το εμβαδόν (area) της  $i$  (σε pixels).

B. Συναρτησιακό ενέργειας (*Mumford Shah Energy functional*) (MSF) [221], που πραγματοποιεί κατάτμηση ελαχιστοποιώντας το ολοκληρωτικό μέσο χρωματικό σφάλμα, το gradient και το μήκος περιγραμμάτων καμπυλών

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

όπου  $u$  η κατατμημένη εικόνα-μωσαϊκό και  $\Gamma$  τα σύνορα κατάτμησης. Το  $\Gamma$  πρέπει να είναι όσο το δυνατόν μικρότερο ενώ τα  $\mu$ ,  $\nu$  είναι σταθερές.

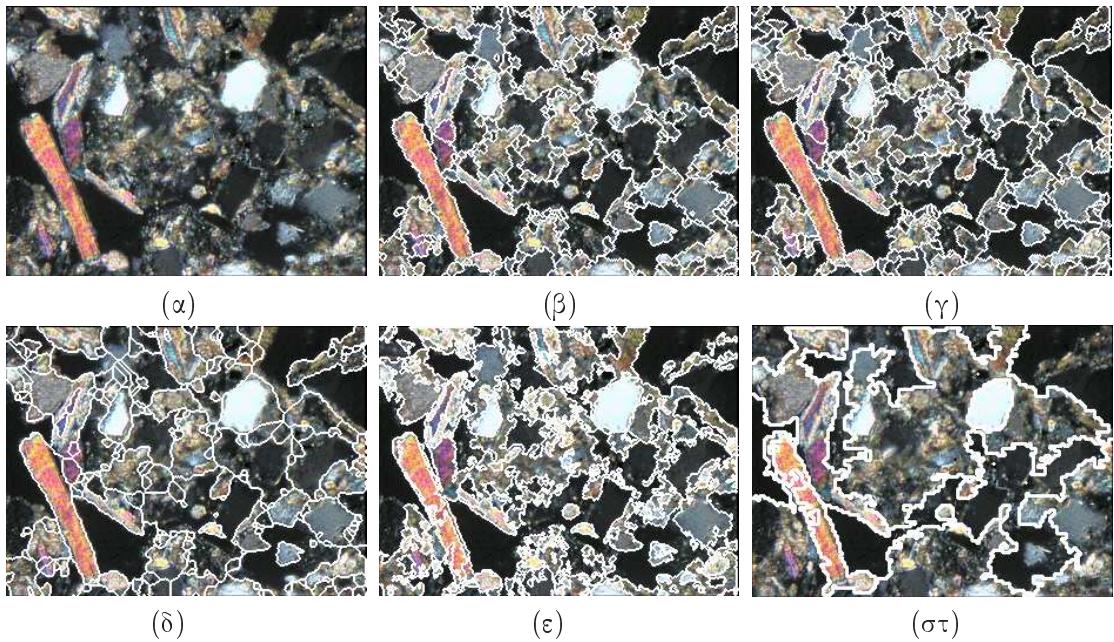
Το συζευγμένο σχήμα, το οποίο θα αναφέρεται ως CTS (contrast-texture) ή VTS (volume-texture) συγκρίθηκε με τις εξής μεθόδους: (α) μετασχηματισμός watershed (WT), καθώς η μέθοδος επεκτείνει την κλασσική προσέγγιση περιλαμβάνοντας πληροφορία υφής, (β) μέθοδος διαχωρισμού και ένωσης Recursive Shortest Spanning Tree (RSST) [172], μια καθιερωμένη region-based τεχνική κατάτμησης, (γ) ομαδοποίηση *k-means clustering* (KM) [84].

Αποτελέσματα συγκριτικά φαίνονται στο Σχ. 7.6, ενώ τα μετρικά ποιότητας φαίνονται στον Πίνακα 7.1. Είναι φανερό ότι το συζευγμένο σχήμα ξεπερνάει σε απόδοση το κλασσικό watershed, Σχ. 7.6(δ), γεγονός αναμενόμενο λόγω της διπλής πληροφορίας αντίθεσης φωτεινότητας και υφής. Τα αποτελέσματα του watershed βελτιώνονται σχετικά με την καλύτερη τοποθέτηση ακμών και τον περιορισμό των συνόρων σε λάθος ακμές (λόγω της παρουσίας υφής). Αν αντί του  $A(t) = 1$  στη ΜΔΕ (7.3) χρησιμοποιηθεί  $A(t) \neq 1$  και πλημμυρισμός όγκου [270], η τοποθέτηση των ακμών είναι περαιτέρω βελτιωμένη, Σχ. 7.6(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 (στ) αντίστοιχα, φαίνονται υποκατατμημένα (undersegmented) με μέτρια τοποθέτηση ακμών, οι οποίες στην περίπτωση του RSST είναι και ισχυρά κβαντισμένες.

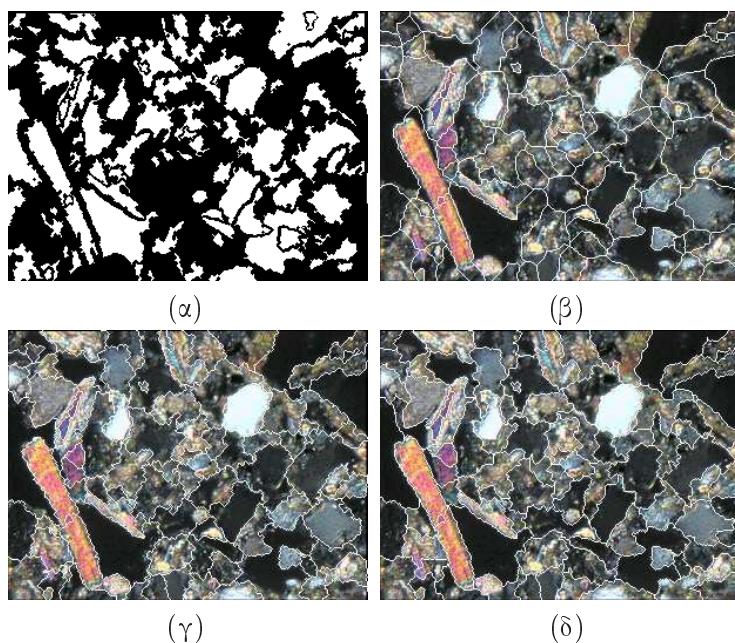
Από τον Πίνακα 7.1 φαίνεται ότι τα καλύτερα αποτελέσματα με ποσοτικούς όρους επιτεύχθηκαν από το volume-based συζευγμένο σχήμα πλημμυρισμού με δεύτερο καλύτερο το εναλλακτικό contrast-based. Ένα δεύτερο σύνολο συγκρίσεων για την ίδια εικόνα, ανάμεσα στα δύο σχήματα και το απλό watershed παρουσιάζεται στο Σχ. 7.7 με ένα διαφορετικό αρχικό σύνολο σημαδευτών, η επιλογή του οποίου φαίνεται στο (α). Τέλος η όλη διαδικασία παρουσιάζεται στο Σχ. 7.8 και για μια διαφορετική εικόνα εδαφοτομής με όλα τα ενδιάμεσα στάδια και με το σύνολο των σχετικών συγκρίσεων που ποσοτικοποιούνται στον Πίνακα 7.2.

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 7.1: Ποσοτική αποτίμηση μεθόδων κατάτμησης για την εδαφοτομή 7.6(α).



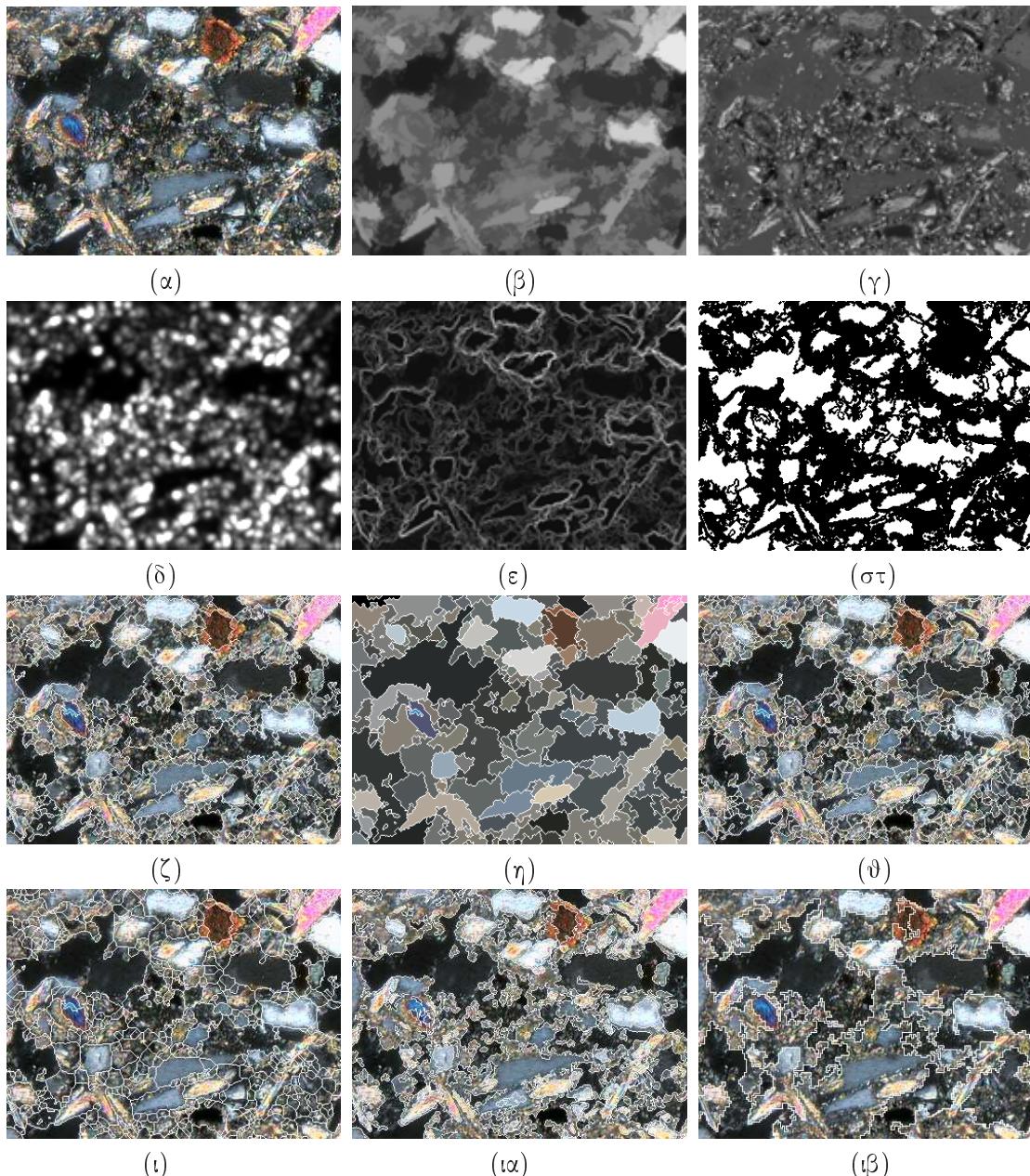
Σχήμα 7.6: Διαφορετικές μέθοδοι κατάτμησης: (α) εδαφοτομή (έγχρωμη) του 7.3(α), (β) συζευγμένη κατάτμηση (contrast-based), (γ) συζευγμένη κατάτμηση (volume-based), (δ) κατάτμηση watershed, (ε) κατάτμηση διαχωρισμού και ένωσης (split and merge) βασισμένη στο RSST, (στ) ομαδοποίηση  $k$ -μέσων ( $k$ -means clustering). Από το [269].



Σχήμα 7.7: Κατάτμηση όπως στο σχήμα 7.4 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).

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

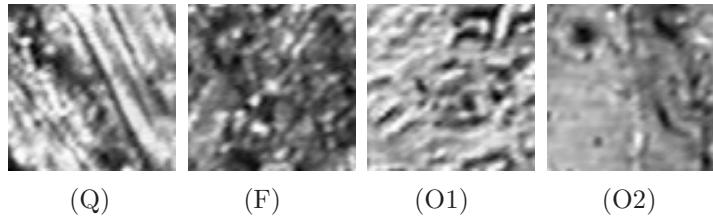
Πίνακας 7.2: Ποσοτική αποτίμηση μεθόδων κατάτμησης για την εδαφοτομή 7.8(α).



Σχήμα 7.8: Κατάτμηση και συγκρίσεις: (α) εδαφοτομή (έγχρωμη), (β) συνιστώσα cartoon  $u$ , (γ) συνιστώσα υφής  $v$ , (δ) Ενέργεια διαμορφώσεων υφής  $\Psi_{\text{mat}}(v)$ , (ε) μέτρο gradient του  $u$ , (στ) σημαδευτές, (ζ) συζευγμένη κατάτμηση (contrast-based), (η) περιοχές κατάτμησης βελτιωμένες με συγχωνεύσεις, (θ) συζευγμένη κατάτμηση (volume-based), (ι) κατάτμηση διαχωρισμού και ένωσης (split and merge) βασισμένη στο RSST, (ιβ) ομαδοποίηση κ-μέσων (k-means clustering). Από το [268].

## 7.5 Κατηγορίες και διάκριση υφής

Η ταξινόμηση της υφής που απαντάται στις εικόνες των εδαφοτομών συνίσταται στην κατασκευή διανυσμάτων χαρακτηριστικών για κάθε χωρικό σημείο της εικόνας και στην επιλογή ενός κατάλληλου ταξινομητή (classifier) (π.χ. k-means, Bayes, SVC) [84]. Για να εκτιμηθεί η αποτελεσματικότητα των εξαγόμενων χαρακτηριστικών για ανάλυση υφής ανεξάρτητα από την επιλογή του ταξινομητή εξετάζεται η διαχωρισμότητα των τάξεων σε ένα σύνολο από τυπικές εικόνες μικροϋφής εδάφους, σύμφωνα με το κριτήριο Fisher (6.4). Όπως αναφέρθηκε ξανά το



Σχήμα 7.9: Τυπικές μικροϋφέρες χώματος: κρύσταλλοι (quartz grains - crystals) (Q), μικρομάζα (fine matter) (F) και οργανικό υλικό (organic matter) (O1, O2).

χριτήριο αυτό ποσοτικοποιεί την απόσταση ανάμεσα στις διαφορετικές τάξεις που κατασκευάζει ένας ταξινομητής και το βαθμό στον οποίο αυτές είναι συμπαγείς [84] (βλέπε αναλυτικά Εν. 6.1.1).

Ελέγχονται τρία διανύσματα χαρακτηριστικών και η αποτελεσματικότητα τους ποσοτικοποιείται από την τιμή του χριτηρίου. Τα χαρακτηριστικά, όπως και στην Εν. 6.1.1, βασίζονται στην πολυκαναλική Gabor ανάλυση και στην περιγραφή της υφής στον χώρο των αποχρίσεων των φίλτρων. Σε κάθε σημείο αποδίδεται ένα σύνολο τιμών, αποτέλεσμα ενός τελεστή ή μιας διαδικασίας στην έξοδο της συστοιχίας.

**α) Gabor** ( $K \times 1$ ): ενέργεια (quadrature) των  $K$  αποχρίσεων των μιγαδικών φίλτρων

$$\vec{I}_G(x, y) = [\{(g_{ke} * f)^2 + (g_{ko} * f)^2\}_{k=1:K}](x, y). \quad (7.10)$$

**β) Teager** ( $K \times 1$ ): Teager ενέργεια των αποχρίσεων, δηλ. η εφαρμογή του μη-γραμμικού τελεστή στις εξόδους των φίλτρων

$$\vec{I}_T(x, y) = [\{C(g_k * f)\}_{k=1:K}](x, y). \quad (7.11)$$

**γ) EDC** ( $3 \times 1$ ): κυρίαρχη ενέργεια διαμόρφωσης, προσανατολισμός κυρίαρχης συχνότητας και ένταση φωτεινότητας

$$\vec{I}_D(x, y) = [\Psi_{mat}(I), \angle \vec{\omega}_D, f](x, y). \quad (7.12)$$

Το τελευταίο διάνυσμα, είναι ένα εναλλακτικό διάνυσμα χαμηλής διάστασης, που έχει εξαχθεί μέσω εντοπισμού ενέργειας διαμόρφωσης και DCA. Η μέτρηση της ενέργειας εμπεριέχει τις μετρήσεις κυρίαρχου πλάτους και συχνότητας.

Οι τέσσερις τυπικές τάξεις υφής που φαίνονται στο Σχ. 7.9 έχουν ληφθεί από τις λεπτομέρειες των κατατημένων εδαφοτομών και αντιπροσωπεύουν κρύσταλλους/ανόργανο υλικό (Q), μικρομάζα ή συσσωματώματα αδιάκριτου υλικού (F) και οργανικό υλικό (O). Οι κατηγορίες 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 7.3 παρουσιάζονται οι τιμές των αποστάσεων  $F$ .

Οι μέγιστες και ελάχιστες αποστάσεις με τα αντίστοιχα ζεύγη τάξεων ανά μέθοδο είναι:  $F_G=[1.41 \text{ (Q-O1)}, 1.03 \text{ (Q-F)}]$ ,  $F_T=[1.56 \text{ (Q-O1)}, 1.10 \text{ (Q-F)}]$ ,  $F_D=[1.65 \text{ (F-O1)}$ ,

	Gabor, $\vec{I}_G$				Teager, $\vec{I}_T$				EDC, $\vec{I}_D$			
	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 7.3: Αποστάσεις Fisher για χαρακτηριστικά υφής μετά από Gabor πολυκαναλική ανάλυση, για τα δείγματα τυπικής υφής στο σχήμα 7.9.

1.01 (F-O2)]. Το διάνυσμα Teager ενεργειών των Gabor αποκρίσεων πετυχαίνει μεγαλύτερη διαχωρισμότητα για κάθε ζεύγος με μέση απόσταση 1.33, συγχρινόμενο με την τυπική Gabor ενέργεια μέσης απόστασης 1.23. Το χαμηλής διάστασης διάνυσμα κυρίαρχων χαρακτηριστικών, το οποίο φέρει πληροφορία πλάτους και συχνότητας διαμόρφωσης δίνει τη μέγιστη απόσταση, (F-O1), και μια μέση τιμή 1.27 για όλα τα ζεύγη.

Στο μικρό δείγμα εικόνων που εξετάστηκε, ο τελεστής Teager είναι περισσότερο αποδοτικός από την κλασσική τοπική ενέργεια Gabor σε επίπεδο διαχωρισμότητας των τάξεων και μπορεί να χρησιμοποιηθεί εναλλακτικά σαν ενεργειακός τελεστής ή μη-γραμμικότητα μετά από πολυκαναλικό φιλτράρισμα. Η ενέργεια διαμορφώσεων υφής ισορροπεί το συγκερασμό ανάμεσα στη διάσταση και την περιγραφικότητα και αναδεικνύει το γεγονός ότι τα κυρίαρχα χαρακτηριστικά διαμόρφωσης μπορούν να χρησιμοποιηθούν για ανάλυση των κατατμημένων εικόνων εκ των υστέρων, με ταξινόμηση των περιοχών που έχουν προκύψει ανάλογα με την υφή τους ή εκτίμηση στατιστικών μεγεθών σχετικών με την κατηγορία των διαφορετικών περιοχών.

## 7.6 Συμπεράσματα

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

*Κεφάλαιο 7. Εφαρμογές στην Ανάλυση Εικόνων Εδαφοτομών  
τηλεπισκόπισης.*

□

## Κεφάλαιο 8

# Μικροδομές Ακουστικών Σημάτων και Ανίχνευση Φωνής

Η ομιλία αποτελεί ίσως την περισσότερο αναπτυγμένη μορφή ανθρώπινης επικοινωνίας και ακολούθησε τη μουσική και τις κάθε είδους εικόνες. Από την πρώτη εισαγωγή αναλογικών χυματομορφών σε μηχανές και τους πρώτους κωδικοποιητές φωνής, φτάσαμε στην 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 [281] 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 φωνής, όπως η παρουσία διαμορφώσεων και η πολυκαναλική ανάλυση, προτείνονται νέες περιγραφές ενέργειας και φασματικού περιεχομένου, αντίστοιχες με αυτές που αναπτύχθηκαν για την ανάλυση της υφής εικόνων. Έτσι, 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 αποτελεσματικότητα 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) [93, 94].

### 8.1 Φωνή σε Θόρυβο

Η ανίχνευση ύπαρξης φωνής σε ακολουθίες ηχητικών σημάτων μπορεί να αντιμετωπίσει είτε άμεσα ως αναγνώριση και ονοματισμός 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, το διαχωρισμό και την προσπέλαση μεγάλων βάσεων και σε τεχνικές ενίσχυσης όπως η εκτίμηση φάσματος θορύβου (noise spectrum estimation) [209], η απόρριψη πλαισίων επεξεργασίας (frame dropping) [248], η μείωση θορύβου, η ακύρωση ηχούς (echo cancellation), η κανονικοποίηση ενέργειας και η συμπίεση διαστημάτων σιωπής. Στις τηλεπικοινωνίες εφαρμόζεται για επιλεκτική, πραγματικού χρόνου, μετάδοση φωνής πάνω από δίκτυα.

Οι σύγχρονες προσεγγίσεις επικεντρώνονται είτε στην ανάπτυξη εξελιγμένων χαρακτηριστικών είτε σε εξυπότερες και πολυπλοκότερες λογικές απόφασης. Πρωτότυπα χαρακτηριστικά για ταυτοποίηση 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 [316], χρονική περιβάλλουσα ισχύος [209], φασματική απόκλιση [248, 249] και χρονοσυχνοτικές παραμέτρους μέσω πολυκαναλικής ανάλυσης [313]. Πρόσφατα το στατιστικό πλαίσιο ανίχνευσης αποκτά ενδιαφέρον με χρήση ιδιοτήτων των στατιστικών της φωνής και ως καθώς και βέλτιστους κανόνες λόγων πιθανοφάνειας [271].

## 8.2 Ενεργειακοί τελεστές και πολυζωνικές διαμορφώσεις

Πειραματικές και θεωρητικές ενδείξεις για την ύπαρξη διαμορφώσεων σε πολλαπλές κλίμακες κατά τη διάρκεια παραγωγής φωνής οδήγησαν στην πρόταση του AM-FM μοντέλου διαμορφώσεων [201]. Η αποδιαμόρφωση ενός πραγματικού AM-FM σήματος

$$x(t) = a(t) \cos \left( \int_0^t \omega(\tau) d\tau \right) \quad (8.1)$$

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 [148, 281]. Για σήματα συνεχούς χρόνου αυτός ο τελεστής είναι

$$\Psi[x(t)] \equiv [\dot{x}(t)]^2 - x(t)\ddot{x}(t), \quad (8.2)$$

όπου  $\dot{x}(t) = dx(t)/dt$ . Ο Teager-Kaiser ενεργειακός τελεστής  $\Psi$  εντοπίζει τη στιγμιαία ενέργεια πηγής που παράγει μια ταλάντωση. Εφαρμοζόμενος σε ένα AM-FM σήμα της μορφής (8.1), ο  $\Psi$  δίνει τη στιγμιαία ενέργεια της πηγής, δηλαδή την ποσότητα  $\Psi[x(t)] \approx a^2(t)\omega^2(t)$ ,

## Κεφάλαιο 8. Μικροδομές Ακουστικών Σημάτων και Ανίχνευση Φωνής

όπου το σφάλμα προσέγγισης είναι αμελητέο [201] αν το στιγμιαίο πλάτος  $a(t)$  και η στιγμιαία συχνότητα  $\omega(t)$  δε μεταβάλλονται πολύ ή γρήγορα σε σχέση με τη μέση τιμή του  $\omega(t)$ .

Ο 1-Δ αλγόριθμος AM-FM αποδιαμόρφωσης [201] (*energy separation algorithm*), βασίζεται στον  $\Psi$  και στον διαχωρισμό της στιγμιαίας ενέργειας στα σήματα πλάτους και συχνότητας

$$\sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}} \approx \omega(t) \quad , \quad \frac{\Psi[x(t)]}{\sqrt{\Psi[\dot{x}(t)]}} \approx |a(t)|. \quad (8.3)$$

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

$$x[n] = x(nT) = A[n] \cos \left( \int_0^n \Omega[k] dk \right), \quad (8.4)$$

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 (8.4) 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*) [201]:

$$\arccos \left( 1 - \frac{\Psi_d[x_n - x_{n-1}] + \Psi_d[x_{n+1} - x_n]}{4\Psi_d[x_n]} \right) \approx \Omega[n], \quad (8.5)$$

$$\sqrt{\frac{\Psi_d[x_n]}{\sin^2(\Omega[n])}} \approx |A[n]|. \quad (8.6)$$

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

Όπως και στην 2-Δ περίπτωση, για να εφαρμοστεί η αποδιαμόρφωση μέσω του ESA σε ένα AM-FM σήμα, απαιτούνται διαδικασίες φιλτραρίσματος για την απομόνωση στενών ζωνών συχνοτήτων του σήματος όπου κυριαρχεί μία μόνο τέτοια συνιστώσα. Το σχήμα πολυκαναλικής ανάλυσης (MDA) προτάθηκε στο [38] ως τρόπος εξαγωγής διαμορφώσεων παρουσία θιρύβου. Συμφωνα με το MDA, η αποδιαμόρφωση επιτελείται στην έξοδο μιας συστοιχίας 1-Δ Gabor φίλτρων, των οποίων η κρουστική απόκριση  $h(t)$  και η απόκριση συχνότητας  $H(\omega)$  δίνονται από τις σχέσεις

$$h(t) = \exp(-\alpha^2 t^2) \cos(\omega_c t) \quad (8.7)$$

$$H(\omega) = \frac{\sqrt{\pi}}{\alpha} \left[ \exp\left(-\frac{(\omega - \omega_c)^2}{4\alpha^2}\right) + \exp\left(-\frac{(\omega + \omega_c)^2}{4\alpha^2}\right) \right], \quad (8.8)$$

όπου  $\omega_c$  η κεντρική συχνότητα του φίλτρου και  $\alpha$  το rms εύρος του.

### 8.3 Κίνητρα και ανίχνευση με στατιστικές υποθέσεις

Για την αποτελεσματική ανίχνευση φωνής προϋποτίθεται η εξέταση τόσο του ενεργειακού επιπέδου της διέγερσης όσο και του συχνοτικού της περιεχομένου. Ο ορισμός της ενέργειας ενός σήματος κατά τον Teager, που ταυτίζεται με την ενέργεια που παράγεται από την πηγή του, ταιριάζει σε ένα τέτοιο πλαίσιο. Η Teager ενέργεια μετράει ταυτόχρονα και στιγμιαία

το μέγεθος και των δύο ποσοτήτων που εκφράζουν τόσο τη φασματική όσο και τη χρονική πληροφορία μεταβολής.

Θεωρώντας το άνθροισμα διαμορφωμένων ημιτόνων ως το μοντέλο φωνής [201], η περιγραφή διακριτού χρόνου είναι

$$s[n] = \sum_{k=1}^K A_k[n] \cos(\Omega_{ck} \cdot n + \Phi_k[n]), \quad (8.9)$$

όπου  $k$  ο δείκτης συντονισμού (resonance) και  $K$  ο αριθμός των θεμελιωδών συντονισμών (formants). Τα στιγμαία μεταβαλλόμενα σήματα πλάτους  $A_k$  και φάσης  $\Phi_k$  (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, δηλαδή  $K = 1$  με την απομόνωση μιας ζώνης διαμόρφωσης μέσω ενός αρκετά στενού Gabor φίλτρου. Η φέρουσα  $\Omega_{ck}$  μπορεί να θεωρηθεί γνωστή και να προσεγγιστεί μέσω της κεντρικής συχνότητας του φίλτρου. Διατύπωνοντας 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 [154]:

$$\begin{aligned} H_0 : \quad X[n] &= W[n], \\ H_1 : \quad X[n] &= W[n] + A \cos(\Omega_c n + \Phi) + B, \end{aligned} \quad (8.10)$$

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 [94] ότι η συνάρτηση λογαριθμικής πιθανοφάνειας για τη δεσμευμένη στο  $H_1$  κατανομή Gauss  $p(X|H_1)$  είναι:

$$\ln p(X|H_1) \approx N \frac{\hat{A}^2}{4\hat{\sigma}_1^2} + \frac{N}{2\hat{\sigma}_1^2} (\hat{B}^2 - \hat{\sigma}_0^2) - \frac{N}{2} \ln 2\pi\hat{\sigma}_1^2. \quad (8.11)$$

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  $MDL(i) = \ln p(X|H_i) - \frac{n_i}{2} \ln N$ , όπου  $n_i = [1, 4]$  η διάσταση του διανύσματος παραμέτρων για  $i = 1, 2$ .

Η εκτίμηση του  $\Omega_c$  χρησιμοποιώντας ένα Gaussian παράθυρο εύρους συχνότητας  $\sigma_g$  χαρακτηρίζεται από κάποιες σχέσεις αβεβαιότητας των μέσων διάρκειας και συχνότητας [64], που καταλήγουν στο  $\Omega_c^2 + \sigma_g^2 \geq 1/4N^2$  [94]. Χρησιμοποιώντας το κατώτερο όριο αβεβαιότητας με  $n_1 = 4$  για το  $MDL(1)$  μπορεί να κατασκευαστεί ένας κανόνας για την ανίχνευση ημιτονοειδών συνιστωσών φωνής

$$N \frac{\hat{A}^2}{4\hat{\sigma}_1^2} + \ln(\Omega_c^2 + \sigma_g^2) \stackrel{H_1}{\underset{H_0}{\gtrless}} \mathcal{O}(\hat{B}, \hat{\sigma}_1^2, \hat{\sigma}_0^2, N), \quad (8.12)$$

όπου  $\mathcal{O}$  μια συνάρτηση στατιστικών στο παράθυρο ανάλυσης, το  $\sigma_g$  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

## Κεφάλαιο 8. Μικροδομές Ακουστικών Σημάτων και Ανίχνευση Φωνής

μεγιστοποιώντας το MDL χριτήριο ενώ ένα παράθυρο ανάλυσης χαρακτηρίζεται ως θόρυβος αν  $MDL(0) > MDL(i), \forall i \neq 0$ .

Από το [38] 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  $\Omega_c$  και  $H$  είναι αντίστοιχα η κεντρική και η απόχριση συχνότητας του φίλτρου και  $\Gamma_c$  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

$$\ln \Psi(X[n]) \approx \ln A^2 + \ln(\Omega_c^2 + \Gamma_c) + \text{const.} \quad (8.13)$$

Συγκρίνοντας τις εξισώσεις (8.12) και (8.13) προκύπτει η αναλογία ανάμεσα στους όρους πλάτους και συχνότητας και τις σταθερές που βασίζονται στο μέσο ζωνοπερατό θόρυβο μέσα στους λογαρίθμους. Αυτές οι ομοιότητες, παρά τις προσεγγίσεις που εφαρμόστηκαν 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 1-Δ AM-FM διαμόρφωσεις είναι: (α) η στατιστική ερμηνεία ενός ανιχνευτή κατωφλίων (8.12), ο οποίος 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.

## 8.4 Χαρακτηριστικά διαμόρφωσης

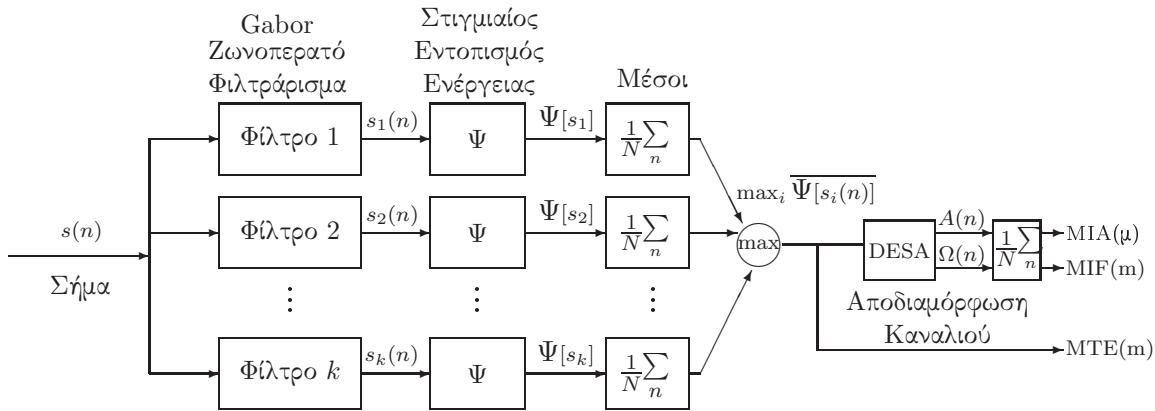
Οι ζώνες διαμόρφωσης εξάγονται μέσω μιας συστοιχίας  $K$  Gabor φίλτρων και ο διακριτός ενεργειακός τελεστής  $\Psi_d$  εφαρμόζεται ως μη-γραμμική ενεργειακή μέτρηση. Η διαδικασία αποτυπώνεται σχηματικά στο Σχ. 8.1. Για κάθε παράθυρο ανάλυσης  $m$ , λαμβάνονται αναπαραστάσεις βραχέως χρόνου των κυρίαρχων συνιστωσών διαμόρφωσης εντοπίζοντας στον πολυδιάστατο χώρο των αποκρίσεων των φίλτρων στο σήμα  $s$ , τη μέγιστη μέση Teager ενέργεια (*mean Multiband Teager Energy*) ή MTE :

$$\text{MTE}(m) = \max_{1 \leq k \leq K} \frac{1}{N} \sum_{n=1}^N \Psi_d[(s * h_k)(n)], \quad (8.14)$$

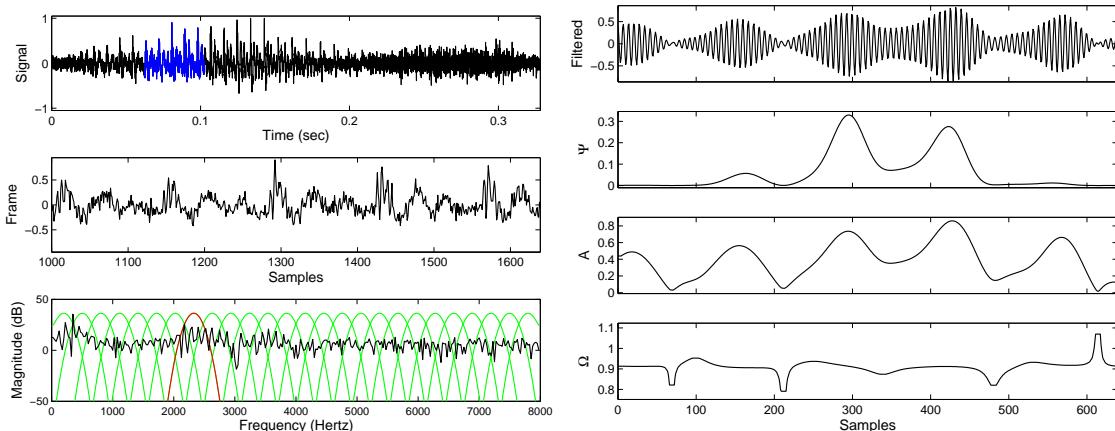
όπου το  $*$  συμβολίζει συνέλιξη,  $n$  είναι ο δείκτης δείγματος με  $(m-1)N+1 \leq n \leq mN$  και  $h_k$  η απόχριση του  $k$ -οστού φίλτρου.

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)

$$\text{MIA}(m) = \frac{1}{N} \sum_{n=1}^N |A_i(n)|, \quad \text{MIF}(m) = \frac{1}{N} \sum_{n=1}^N \Omega_i(n). \quad (8.15)$$



Σχήμα 8.1: Εξαγωγή κυρίαρχων χαρακτηριστικών διαμόρφωσης: Πολυκαναλικό φιλτράρισμα, εντοπισμός ενέργειας διαμόρφωσης και αποδιαμόρφωση του φίλτρου με την μέγιστη μέση Teager ενέργεια του τελεστή  $\Psi$ . 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.



Σχήμα 8.2: 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.

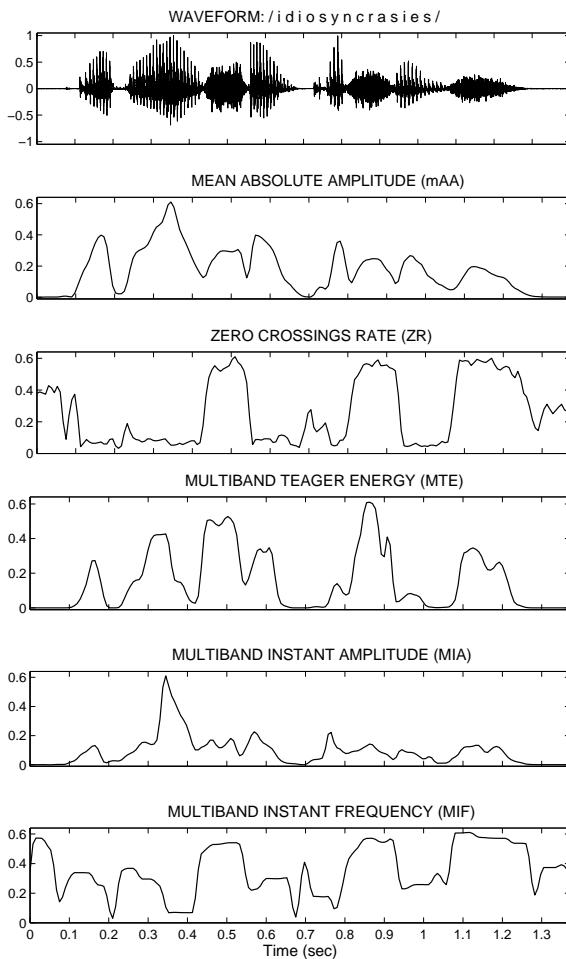
Το MTE ερμηνεύεται ως η κυρίαρχη ενέργεια διαμόρφωσης (dominant modulation energy), που εντοπίζει από κοινού πληροφορία πλάτους και συχνότητας των σημάτων φωνής.

Στο Σχ. 8.2 φαίνεται για ένα παράθυρο ανάλυσης ενός 1-Δ σήματος, φωνής στην προκειμένη περίπτωση, η διαδικασία κάλυψης του φάσματος από τη συστοιχία (25 Gabor φίλτρα με rms ενεργό εύρος 160 Hz και grammik' h topoj' ethsh) και η επιλογή και αποδιαμόρφωση του κυρίαρχου καναλιού. Τα νέα χαρακτηριστικά βραχέως χρόνου για ολόκληρο το σήμα, συγχρινόμενα με κλασσικές αναπαραστάσεις ενέργειας (μέσο απόλυτο πλάτος ή mAA) και συχνοτικής μεταβολής (μέσος ρυθμός μεταβάσεων από το μηδέν ή ZR) φαίνονται στο Σχ. 8.3.

## 8.5 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 κλασσική μέθοδο ανίχνευσης ορίων φωνής (endpoint detection) με κατώφλια [247]. Από τα πρώτα 100

*Κεφάλαιο 8. Μικροδομές Ακουστικών Σημάτων και Ανίχνευση Φωνής*



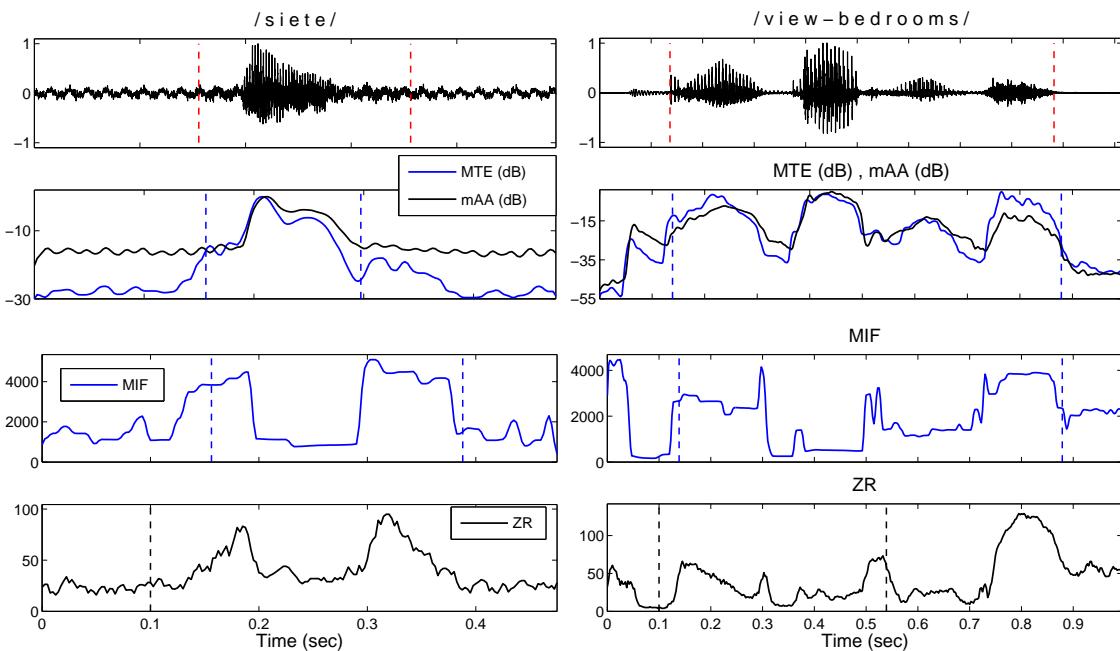
Σχήμα 8.3: Χαρακτηριστικά ανάλυσης και ανίχνευσης φωνής σε 15 ms παρ'ajura Hamming, με 1/2 της διάρκειας επικάλυψη, στα 16 kHz. Κυμματομορφή, κλασσικά χαρακτηριστικά (mAA, ZR) και τα νέα χαρακτηριστικά διαμόρφωσης MTE, MIA & MIF. Όλες οι παραστάσεις είναι κανονικοποιημένες και ομαλοποιημένες με συνδυασμό γραμμικών και μη-γραμμικών φίλτρων (median 3 σημείων, Hanning 3 σημείων).

ms, που θεωρούνται εκ των προτέρων σιωπή ή θόρυβος περιβάλλοντος, η μέση τιμή  $\mu_{\text{sif}}$  και η τυπική απόκλιση  $\sigma_{\text{sif}}$  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  $\kappa, \lambda$ , σύμφωνα με τις σχέσεις

$$\gamma_f = \mu_{\text{sif}} + \kappa \sigma_{\text{sif}}, \quad \gamma_d = \min(T_1, T_2), \quad \gamma_u = 5 \cdot \gamma_d, \quad (8.16)$$

$$T_1 = \lambda P_{\max} + (1 - \lambda) S_{\max}, \quad T_2 = 3 \cdot S_{\max}. \quad (8.17)$$

Ένας διπλός έλεγχος της «ενεργειακής» μέτρησης του σήματος, αναζητώντας τα ακραία σημεία όπου ξεπερνιέται ένα υψηλό κατώφλι  $\gamma_u$  αμέσως μετά από ένα χαμηλότερο  $\gamma_d$ , ανιχνεύει το κύριο, συνήθως έμφωνο τιμήμα της διαταραχής. Τα αρχικά εκτιμώμενα σύνορα μετατοπίζονται έτσι ώστε να συμπεριλάβουν τυχόν ισχυρή φασματική δραστηριότητα, χρησιμοποιώντας το  $\gamma_f$  κατώφλι μετρήσεων «συχνοτικού» περιεχομένου. Για τις εδώ δοκιμές δόθηκαν στα βάρη οι τιμές  $\lambda = 0.02$ ,  $\kappa = 1$ . Στο Σχ. 8.4 δίνονται παραδείγματα του αλγορίθμου με τα κλασσικά και τα νέα χαρακτηριστικά διαμόρφωσης.



Σχήμα 8.4: 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 αποτελέσματα τους. Οι «ενεργειακές» μετρήσεις (mAA, MTE) παρουσιάζονται σε dB σε κοινή παράσταση για λόγους σύγκρισης.

Κλασσικά	Πολυκαναλικά	Διαμόρφωσης	Teager
mAA, ZR 56.1	mSA, ZR 66.6	MAA, MZR 51.5	MTE, MIF 73.5

Πίνακας 8.1: Ορθή ανίχνευση ορίων φωνής (%) στη βάση NTIMIT

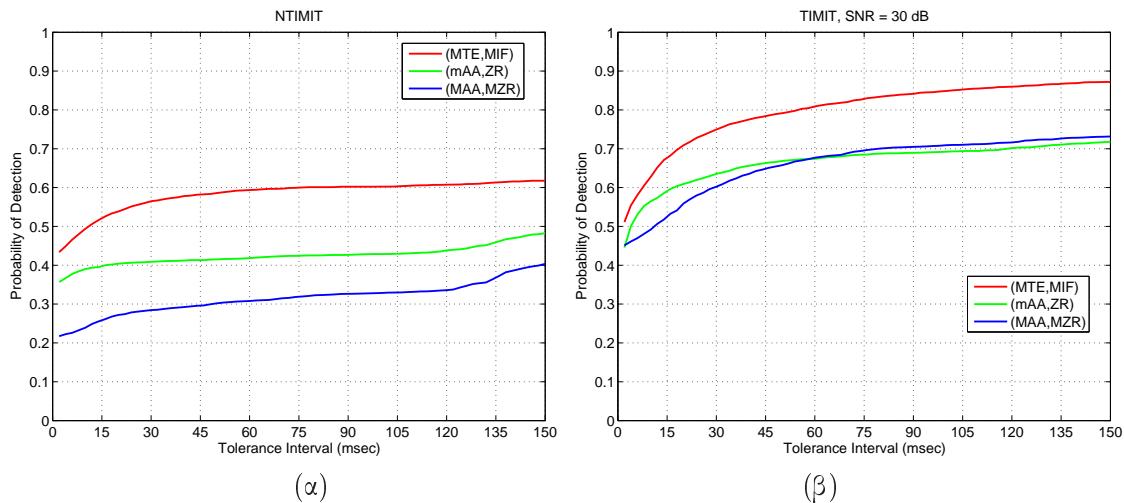
Τα χαρακτηριστικά για ανίχνευση ορίων φωνής δοκιμάστηκαν κάτω από πραγματικές συνθήκες τηλεφωνικού θορύβου στη βάση NTIMIT. Το εγχείρημα ήταν η ανίχνευση ορίων φράσεων, ανεξάρτητα από το είδος της ενδιάμεσης δραστηριότητας, συγχρινόμενα με τα όρια που έχουν προκύψει από ανθρώπινη επισκόπηση της βάσης. 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 8.1. 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 πολυκαναλικές παραλλαγές των κλασσικών και αποτελούν αντιστοιχα 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 [316], ενώ η STE είναι η μέση έξοδος του ενεργειακού τελεστή ανά παράθυρο, χωρίς πολυκαναλική ανάλυση. Συμπερασματικά, η αποτελεσματικότητα των νέων χαρακτηριστικών μπορεί να αποδοθεί στα πλεονεκτήματα τόσο της πολυκαναλικής ανάλυσης όσο και των διαφορετικών μετρήσεων ενέργειας διαμορφώσεων. Η μείωση του μέσου

## Κεφάλαιο 8. Μικροδομές Ακουστικών Σημάτων και Ανίχνευση Φωνής

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 καθυστέρησης τέλους (spoke-too-soon error) να μεταβάλλεται στο διάστημα 2 με 150 ms. Αυτό το διάστημα, χανονικοποιημένο στο μέγιστο του, δίνει ένα μέτρο που σχετίζεται με την πιθανότητα εσφαλμένου συναγερμού. Οι δύο ποσότητες δεν είναι ίσες αλλά συνδέονται με μια ένα προς ένα, μονοτονικά αύξουσα, άγνωστη συνάρτηση καθώς αύξηση στην ανοχή, αυξάνει την πιθανότητα κατά κάποιο ποσοστό [94]. Στο Σχ. 8.5 φαίνονται τέτοιες καμπύλες ανίχνευσης - ανοχής για τα κλασσικά χαρακτηριστικά, τις πολυκαναλικές εκδοχές τους και τα χαρακτηριστικά κυρίαρχων διαμορφώσεων.



Σχήμα 8.5: 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.

## 8.6 Ανίχνευση φωνητικής δραστηριότητας

Τα συστήματα ανίχνευσης φωνητικής δραστηριότητας (*voice activity detection*, VAD), κατηγοριοποιούν εισερχόμενα παράθυρα ενός σήματος με βάση εξαγώμενα χαρακτηριστικά σε δύο τάξεις χονδρικά: φωνητικά και μη-φωνητικά γεγονότα (παύσεις, σιωπή, θόρυβος, αλλά και ψίθυροι, ασθενή άφωνα κ.ά.). Ένα τέτοιο σύστημα, σχεδιασμένο για ακρίβεια και ανοχή σε θόρυβο, προτάθηκε στα [248, 249] και περιλαμβάνει τη χρήση της μακροπρόθεσμης φασματικής απόκλισης (*long-term spectral divergence*, LTSD). 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. Ο αλγόριθμος βασίζεται σε προσαρμοζόμενα

κατώφλια και ενημέρωση των παραμέτρων του θορύβου. Το VAD με LTSD δοκιμάστηκε εξαντλητικά και επιτυχημένα κάτω από διάφορες, πραγματικές συνθήκες θορύβου και συγκρίθηκε με πρότυπα VAD [27], χρησιμοποιώντας κοινές μεθόδους αποτίμησης αλλά και ακρίβεια αναγνώρισης.

### 8.6.1 Ανίχνευση ενέργειας διαμόρφωσης

Η αποτελεσματικότητα του MTE ως ανιχνευτή γεγονότων φωνής σε ένα VAD σύστημα εκμεταλλεύεται προσαρμόζοντας τον βασισμένο στο LTSD αλγόριθμο και αλλάζοντας τη βασική του μέτρηση με την κυρίαρχη ενέργεια διαμόρφωσης MTE σε δύο εναλλακτικές εκφράσεις απόκλισης: μακράς και βραχέας. Κατά τη διάρκεια μιας περιόδου αρχικοποίησης μαθαίνονται τα χαρακτηριστικά του θορύβου. Μετά την εξαγωγή των χαρακτηριστικών (Εν. 8.4) η διαφορά επιπέδου σε dB σε σχέση με την αντίστοιχη μέτρηση θορύβου συγχρίνεται με ένα προσαρμοζόμενο κατώφλι  $\gamma \in [\gamma_0, \gamma_1]$ :

$$\gamma = \gamma_0 + (\gamma_1 - \gamma_0)(E - E_0)/(E_1 - E_0), \quad (8.18)$$

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 υπολογίζοντας τοπικούς μέσους σε μια μικρή γειτονιά παραθύρων. Με τον τρόπο αυτό λαμβάνονται υπόψιν συνηθισμένες αλλά ρεαλιστικές περιπτώσεις με μη-στατικές συνθήκες θορύβου.

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

1. Απόκλιση πολυκαναλικής Teager ενέργειας (*multiband Teager energy divergence*, MTED), η μέγιστη ανά κανάλι μέση ενέργεια MTE, όπως ορίστηκε προηγούμενα συγχρινόμενη με το αντίστοιχο μέγεθος MTEW για το θόρυβο

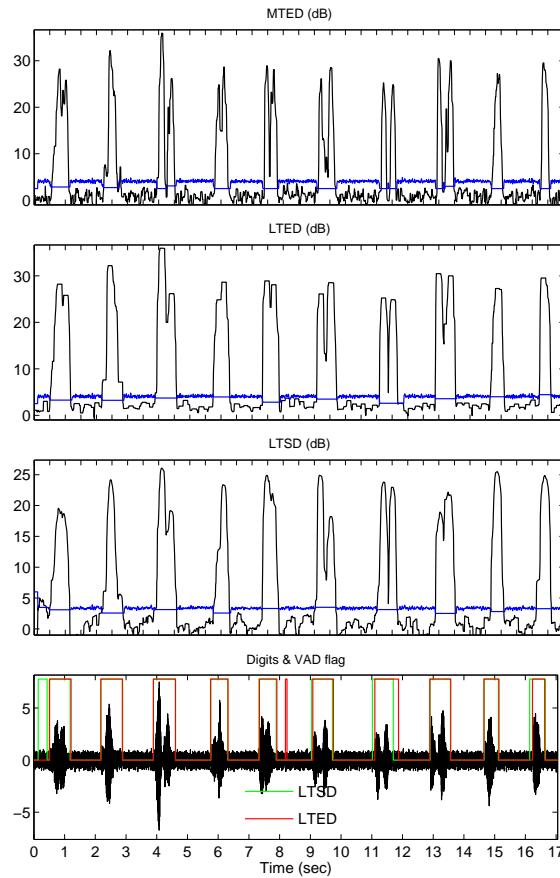
$$\text{MTED}(m) = 10 \log_{10} (\text{MTE}(m)/\text{MTEW}). \quad (8.19)$$

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 Εν. 8.5 που συγχρίνει τη διαφορά μεγέθους του MTE.

2. Μακροπρόθεσμη απόκλιση πολυκαναλικής Teager ενέργειας (*long-term multiband Teager energy divergence*, LTED), όπου το MTE μεγιστοποιείται σε μια γειτονιά  $2L$  παραθύρων, καταλήγοντας σε μια διεσταλμένη και κανονικοποιημένη ως προς τον θόρυβο εκδοχή του:

$$\text{LTED}(m) = 10 \log_{10} \left( \max_l \{\text{MTE}(m+l)\}/\text{MTEW} \right) \quad (8.20)$$

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



Σχήμα 8.6: Παραδείγματα ανίχνευσης κυρίαρχης ενέργεια διαμόρφωσης MTE για VAD και το χαρακτηριστικό αναφοράς LTSD [248]. Τα διαστήματα δραστηριότητας ανιχνεύθηκαν παρόμοια και με τα δύο χαρακτηριστικά, LTED και MTED.

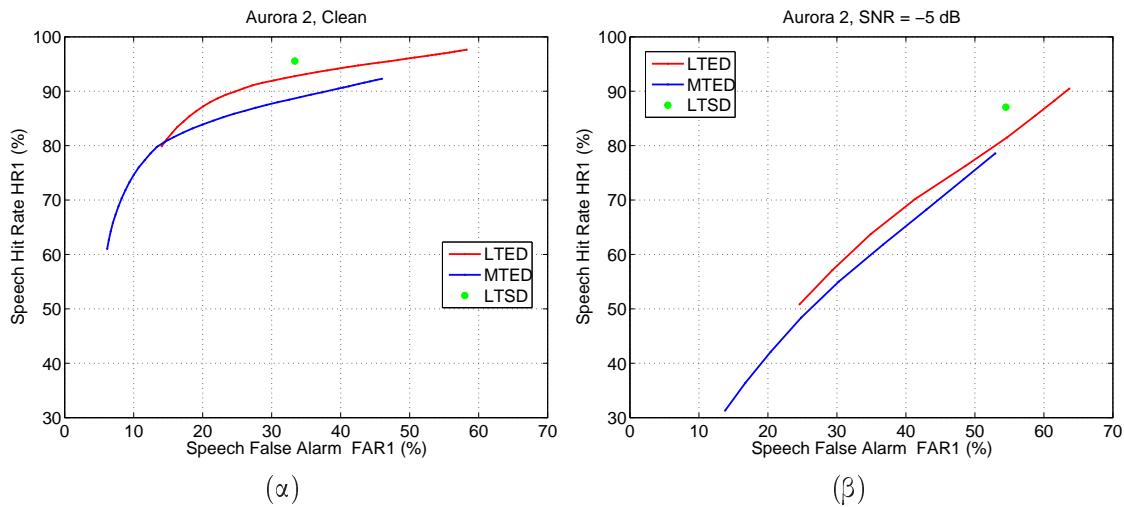
Στο Σχ. 8.6 παρουσιάζεται ένα παράδειγμα των προτεινόμενων χαρακτηριστικών για VAD σε ένα παράδειγμα ακολουθίας ψηφίων από τη βάση Aurora 3 (quiet, hands-free mic.), για παράθυρα 25 ms με 10 ms μετατοπίσεις και μια συστοιχία 25 Gabor. Μαζί φαίνεται και το προσαρμοζόμενο κατώφλι που σηματοδοτεί τη φωνητική δραστηριότητα. Διαισθητικά, το MTE χαρακτηριστικό αναμένεται να δουλεύει καλύτερα καθώς μεγιστοποιεί τη μέτρηση ανά πλαίσιο ανάλυσης τόσο συχνοτικά (μέσω της πολυκαναλικής ανάλυσης) όσο και χρονικά (μέσω της τοπικής γειτονιάς). Έτσι, εξασφαλίζεται η κυρίαρχη τοπικά χρονο-φασματική περιγραφή του σήματος.

### 8.6.2 Αποτελέσματα και συγκρίσεις

Το πειραματικό πλαίσιο περιλαμβάνει συγκρίσεις απόδοσης ανίχνευσης του LTSD σχήματος αναφοράς και των βασισμένων στο MTE VAD (8.19, 8.20), κάτω από διαφορετικές συνθήκες ψηφίου στις βάσεις Aurora 2 (70070 φράσεις) και Aurora 3 (4914 φράσεις). Η αποτίμηση γίνεται σύμφωνα με τα λάθη κατηγοριοποίησης σε διαφορετικά SNR [27, 209, 248] χρησιμοποιώντας κάποιο χαρακτηρισμό ή ονοματισμό αναφοράς του καθαρού συνόλου δεδομένων. Εδώ χρησιμοποιήθηκε αυτόματη αναγνώριση φωνής (automatic speech recognition) για κατάτμηση και ονοματισμό των γεγονότων φωνής και σιωπής στις βάσεις. Αποτελέ-

σματα υψηλού ποσοστού αναγνώρισης στις «καθαρές» ακολουθίες ψηφίων καθορίζουν το ground truth. Εν συντομίᾳ, για το σύνολο Aurora 2, η εκπαίδευση του συστήματος γίνεται χρησιμοποιώντας 32 mixtures, 18 καταστάσεις και το μήκους 39 διάνυσμα χαρακτηριστικών  $[MFCC, \log E, \Delta, \Delta\Delta]^T$  στο σενάριο “clean-train” με τις δοκιμές στα καθαρά δεδομένα να πετυχαίνουν ένα ποσοστό 99.6% ακρίβειας αναγνώρισης λέξεων. Για την Aurora 3, η εκπαίδευση γίνεται με 16 mixtures, 16 καταστάσεις και το ίδιο διάνυσμα χαρακτηριστικών. Το υποσύνολο 1522 φράσεων για το “well-matched” σενάριο δοκιμών χρησιμοποιήθηκε με 93.7% ακρίβεια αναγνώρισης.

Για το VAD αναφοράς LTSD χρησιμοποιήθηκαν οι προδιαγραφές που αναφέρονται στο [248], ενώ για τα προτεινόμενα VAD τα βέλτιστα κατώφλια (8.18) εκτιμήθηκαν μέσω καμπυλών ROC. Στο Σχ. 8.7 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 ιδιαίτερου 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.



Σχήμα 8.7: 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.

Η απόδοση των VAD εκτιμήθηκε σύμφωνα με το ποσοστό ορθών ανιχνεύσεων φωνής (speech hit rate, HR1), το οποίο 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, HR0) ορισμένο αντίστοιχα 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 8.2 και 8.3 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 και στις δύο κατηγορίες γεγονότων

*Κεφάλαιο 8. Μικροδομές Ακουστικών Σημάτων και Ανίχνευση Φωνής*

Πίνακας 8.2: 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 8.3: 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	

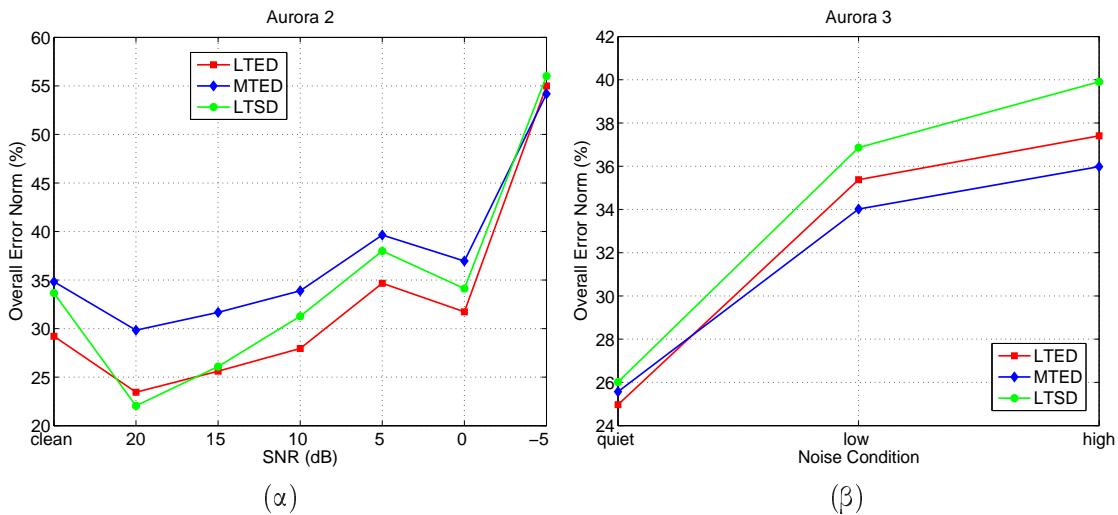
HR1: speech Hit Rate, HR0: non-speech Hit Rate

(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 συναγερμών. Για την καλύτερη λειτουργία επιδιώκεται η ελαχιστοποιήση της νόρμας λάθους (*overall false alarm error norm*):

$$\|(FAR0, FAR1)\| = \left[ (1 - HR0)^2 + (1 - HR1)^2 \right]^{1/2}. \quad (8.21)$$

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 Ευκλείδια απόσταση από το ιδανικό σημείο λειτουργίας στην ROC καμπύλη (HR=100, FAR=0) (Sq. 8.7). Στο Σχ. 8.8 οι νόρμες λάθους παρουσιάζονται για τα δύο σύνολα δεδομένων και τα τρία διαφορετικά VAD ως συνάρτηση της μείωσης του SNR.

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 (8.21) με 7.6% σχετική μείωση του ολικού λάθους έναντι του LTSD. Από το Σχ. 8.8 (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 ανιχνεύουνται τα ίδια το LTED πετυχαίνει υψηλότερη απόδοση



Σχήμα 8.8: Ολικά σφάλματα (overall false alarm error norms) ανίχνευσης φωνητικών γεγονότων κάτω από διαφορετικά SNR στις βάσεις (α) Aurora 2 και (β) Aurora 3.

σε μεμονωμένα hit rate από το LTSD και μια συνολική σχετική μείωση σφάλματος 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 μοντέλα και χαρακτηριστικά κυρίαρχων διαμορφώσεων ξεπερνάνε το LTSD στο συνολικό σφάλμα κάτω από όλες τις συνθήκες θορύβου, όπως φαίνεται και στο Σχ. 8.8(β). Σημειώνεται ότι το LTED είναι με συνέπεια καλύτερο και στα δύο σύνολα δοκιμών.

## 8.7 Συμπεράσματα

Η ανίχνευση φωνητικών γεγονότων προσεγγίστηκε μέσω της ανίχνευσης μικροδομών διαμόρφωσης, εντοπίζοντας και μετρώντας κατάλληλα και σε πολλαπλές ζώνες την κυρίαρχη ενέργεια τους και εξάγοντας αργά μεταβαλλόμενη πληροφορία κυρίαρχου πλάτους και συχνότητας. Η μεθοδολογία είναι ανάλογη με την εξαγωγή των κυρίαρχων συνιστώσων της υφής σε 2-Δ εικόνες και την ανίχνευση των μικροδομών της. Τα χαρακτηριστικά διαμόρφωσης σε 1-Δ σήματα διαπιστώθηκε ότι βελτιώνουν συστηματικά την ανίχνευση των ορίων της φωνής σε μεμονωμένες λέξεις ή φράσεις και μειώνουν το μέσο ολικό σφάλμα ενός σύνεναρού και μεγάλης ακρίβειας ανιχνευτή φωνητικής δραστηριότητας σε μεγάλες βάσεις αναφοράς. Η επέκταση αυτών των ιδεών για γενικευμένη ανίχνευση και ανάλυση ακουστικών γεγονότων μπορεί να γίνει μέσω μιας συστηματικής περιγραφής με όρους μικροδιαμορφώσεων των διαφορετικών κατηγοριών ηχητικών σημάτων.

□

## Κεφάλαιο 9

# Συμπεράσματα και Προεκτάσεις

*I may not have gone where I intended to go, but I think I have ended up where I needed to be.*

---

Douglas Adams

Η μελέτη της υφής με όρους AM-FM διαμορφώσεων ανέδειξε την ικανότητα τέτοιων μοντέλων να εξάγουν και να περιγράφουν δομές σε πολλαπλές κλίμακες με εφαρμογές σε ζητήματα αναπαράστασης εικόνων και υπολογιστικής όρασης. Η ιδιότητα-χαρακτηριστικό της υφής εικόνων αντιμετωπίστηκε σφαιρικά μέσα από ένα υποσύνολο προβλημάτων όπως είναι η ανίχνευση, η αποσύνθεση εικόνων, η εξαγωγή χαρακτηριστικών, η κατάτμηση και η ταξινόμηση.

### 9.1 Συνεισφορές

Το θέμα της οπτικής υφής αναπτύχθηκε στα διάφορα επίπεδα του με κοινή βάση το πολυκαναλικό μοντέλο στενοζωνικών διαμορφώσεων. Συνολικά προτάθηκε ένα ολοκληρωμένο σύστημα ανάλυσης και επεξεργασίας υφής που περιλαμβάνει συνεισφορές και καινοτομίες στα επιμέρους προβλήματα: 1) ανίχνευσης, 2) εξαγωγής χαρακτηριστικών με εφαρμογές σε κατάτμηση-ταξινόμηση υφής και πολυχριτική κατάτμηση εικόνων και 3) διαχωρισμού υφής με αποσύνθεση εικόνων για μοντελοποίηση και ενίσχυση της πληροφορίας.

#### 9.1.1 Ενέργεια υφής και αποδοτική εκτίμηση παραμέτρων

Για την περιγραφή της υφής αξιοποιήθηκε και εμπλουτίστηκε το μοντέλο πολυζωνικών, χωρικών AM-FM διαμορφώσεων σύμφωνα με το οποίο οι μεταβολές της φωτεινότητας μιας εικόνας περιγράφονται από τοπικά στενοζωνικά ημιτονοειδή διαμορφωμένα κατά πλάτος και συχνότητα. Οι συνιστώσες διαχωρίζονται χρησιμοποιώντας συστοιχίες γραμμικών φίλτρων, ανάλυση σε πολλαπλές ζώνες συχνοτήτων και μη-γραμμικούς τελεστές τοπικού εντοπισμού ενέργειας. Μια ενεργειακή ποσότητα, που στις κυρίαρχες ζώνες συχνοτήτων ορίζει την ενέργεια υφής και χρησιμοποιείται για ανίχνευση και εξαγωγή χαρακτηριστικών. Μια θεωρητική

ανάλυση τέτοιων τελεστών παρουσιάστηκε σε ζωνοπερατές εικόνες με υόρυβο. Εξετάστηκαν δύο μέθοδοι ανάλυσης με μοντέλα διαμόρφωσης, η ανάλυση πολλαπλών και η ανάλυση κυρίαρχων συνιστωσών (DCA). Στη δεύτερη περίπτωση προτάθηκε μια νέα, εναλλακτική μεθοδολογία εντοπισμού των κυρίαρχων συνιστωσών της υφής (EDCA) αλλά και μια κανονικοποιημένη προσέγγιση στην αποδιαμόρφωση των διακριτών AM-FM σημάτων με χρήση Gabor φίλτρων (Gabor-ESA). Η υλοποίηση των αλγορίθμων βασίστηκε σε αποδοτικές, γρήγορες υλοποιήσεις στο πεδίο του Fourier μετασχηματισμού. Έτσι, το πρόβλημα της ανάλυσης υφής ανάγεται σε ανίχνευση ενεργειακών όρων και εξαγωγή διανυσμάτων χαρακτηριστικών (κυρίαρχο πλάτος, συχνότητα & προσανατολισμός) χαμηλής διάστασης, των οποίων οι υπολογισμοί γίνονται μέσω εύρωστων αλγορίθμων.

### 9.1.2 Χαρακτηριστικά διαμόρφωσης για μεταβολική κατάτμηση

Η κατάτμηση μιας εικόνας σε περιοχές με ομοιογενή χαρακτηριστικά όπως η γεωμετρία, η φωτεινότητα, το χρώμα, η υφή, η κίνηση κ.ά. είναι μια από τις σημαντικές διεργασίες όρασης χαμηλού επιπέδου και ένα από τα μεγάλα προβλήματα της όρασης υπολογιστών. Περιλαμβάνει τα ανεξάρτητα στάδια της εξαγωγής χαρακτηριστικών, της επιλογής ή μείωσης της διάστασης τους και ενός αλγορίθμου κατάτμησης. Το μοντέλο και τα εξαγώμενα διανύσματα χαρακτηριστικών, χρησιμοποιήθηκε ως κινητήρα δύναμη για μεταβολική κατάτμηση υφής χωρίς επισκόπηση με βάση ένα τροποποιημένο σχήμα ανταγωνισμού περιοχών και γεωδαιτικών ενεργών περιοχών. Η πιθανοτική ερμηνεία της διαδικασίας επιλογής καναλιού-φίλτρου κατά το στάδιο ανάλυσης καθώς και η γενίκευση του για ανίχνευση ακμών και υφής με χρήση αναγεννητικών ημιτονοειδών μοντέλων κάνει δυνατή την εκτίμηση πιθανοτήτων για τις διαφορετικές κατηγορίες σημάτων σε μια εικόνα (υφή, ακμές, ομαλές περιοχές). Επιπλέον επιτρέπει τη σύζευξη των ετερογενών χαρακτηριστικών και οδηγεί σε βελτιωμένα αποτελέσματα κατάτμησης. Παρουσιάστηκαν παραδείγματα σε ένα πλήθος φυσικών εικόνων με υφή, συνοδευόμενα από ποσοτικά αποτελέσματα και συγκρίσεις σε μια βάση εικόνων αναφοράς.

### 9.1.3 Συνέργεια μεταξύ αποδόμησης εικόνων και μοντελοποίησης υφής

Η κατηγορία *α+ν* μοντέλων εικόνων προβλέπει δύο είδη πληροφορίας με τη μορφή ανεξάρτητων συνιστωσών που υπερτίθενται για το σχηματισμό μιας εικόνας. Η μία είναι τμηματικά ομαλή, με απότομες μεταβολές ανάμεσα σε επίπεδα τμηματικά σταθερής φωτεινότητας που περιγράφουν τις μακροδομές και τα γεωμετρικά χαρακτηριστικά της εικόνας και η άλλη, η υφή, αποτελείται από ταχύτατες μεταβολές, πρότυπα σε μικρές κλίμακες και ταλαντώσεις. Η διαδικασία διαχωρισμού υφής από την τμηματικά ομαλή συνιστώσα, η αποσύνθεση ή αποδόμηση εικόνων ανήκει στα αντίστροφη προβλήματα αποκατάστασης ή ανάκτησης. Τα προβλήματα της αποδόμησης εικόνων, του διαχωρισμού υφής και της μοντελοποίησης υφής αντιμετωπίστηκαν σε ένα πλαίσιο συνέργειας, δηλαδή μια διαδραστική σχέση όπου η μία διαδικασία χρησιμοποιεί την άλλη για βελτίωση τόσο του διαχωρισμού όσο και της ανάλυσης. Συγκεκριμένα,

αναπτύχθηκε και προτάθηκε ένα νέο σχήμα αποσύνθεσης εικόνων της κατηγορίας  $u + v$  βασισμένο σε μεταβολικές μεθοδολογίες ελαχιστοποίησης ενεργειακών με χρήση ενός όρου ανακατασκευής της υφής από τις πολυκαναλικές συνιστώσες της. Η αποδόμηση στηρίζεται στην επίλυση ενός προβλήματος ολικής μεταβολής με περιορισμούς για σχετικά με την ανακατασκευή της πληροφορίας υφής από την εικόνα. Τα πλεονεκτήματα του νέου σχήματος, το οποίο συγχρίθηκε θεωρητικά και πειραματικά με σχήματα αναφοράς της βιβλιογραφίας είναι α) η αποκλειστική αξιοποίηση ενός μοντέλου υφής, β) η υποκείμενη έννοια της κλίμακας και της πολυεπίπεδης αποδόμησης, γ) η περαιτέρω αποδόμηση της υφής σε ένα δεύτερο επίπεδο σε στενοζωνικές υποσυνιστώσες διαφορετικής συχνότητας και προσανατολισμού. Το προτεινόμενο σχήμα αποτελεί έτσι ένα  $u + Kv$  μοντέλο, με  $K$  υποσυνιστώσες διαμορφωμένων ταλαντώσεων. Στην αντίθετη κατεύθυνση συνέργειας, ο διαχωρισμός των μικροδομών από τα χαρακτηριστικά μεγάλης κλίμακας, πριν από την εφαρμογή του μοντέλου αντιμετωπίστηκε ως ένα στάδιο προεπεξεργασίας και ενίσχυσης της πληροφορίας υφής. Η αποδοτικότητα των χαρακτηριστικών που εξάγονται από τη διαχωρισμένη συνιστώσα, αναδειχθηκε μέσα από μια εφαρμογή ταξινόμησης υφής, όπου συγχρίθηκε το σφάλμα ταξινόμησης και η διαχωρισμότητα των τάξεων διαφορετικών διανυσμάτων χαρακτηριστικών, με έμφαση στο προτεινόμενο σύνολο των κυρίαρχων χαρακτηριστικών διαμόρφωσης.

#### **9.1.4 Εφαρμογή στην ανάλυση εδαφοτομών**

Μια κατηγορία εικόνων με έντονη υφή και παρουσία διαφορετικών δομών είναι οι ψηφιακές εικόνες λεπτών τομών του εδάφους ή εδαφοτομών (soil sections). Η αυτόματη επεξεργασία τέτοιων εικόνων αποκαλύπτει στοιχεία της σύστασης και της δομής του εδάφους και μπορεί, σχετιζόμενη με ανθρώπινη γνώση και παρατηρήσεις, να βοηθήσει σε εκτιμήσεις βιοοικολογικής και περιβαλλοντικής σημασίας (γονιμότητα, αλλαγές σε οικοσυστήματα κ.ά.). Οι αναπτυχθείσες μέθοδοι εφαρμόστηκαν συστηματικά για εξαγωγή χαρακτηριστικών, ανάλυση υφής και κατάτμηση τέτοιων εικόνων σε ομογενείς περιοχές, σχετικές με την μικρομορφολογία του εδάφους. Αρχικά, με παράλληλη, χωριστή αξιοποίηση της καμποδομής και της μικροδομής τέτοιων εικόνων, αναπτύχθηκε ένα γενικό σχήμα συζευγμένης πολυκριτικακής κατάτμησης. Πρόκειται για ένα μεταβολικό σχήμα που βασίζεται στην εξέλιξη καμπυλών επιπεδοσυνόλων των οποίων η ταχύτητα εξαρτάται τόσο από τα γεωμετρικά χαρακτηριστικά της εικόνας όσο και από την ανιχνευόμενη ενέργεια διαμόρφωσης της συνιστώσας υφής. Παρουσιάστηκαν ποιοτικές και ποσοτικές συγχρίσεις κατάτμησης σε επιλεγμένα, χαρακτηριστικά δείγματα της βάσης. Η ανάλυση των κατατμημένων εικόνων μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση των διαφορετικών φυσικών δομών του εδάφους, π.χ. κρύσταλλοι, ανόργανο και οργανικό υλικό, με βάση το μικροδομή της υφής τους.

## **9.2 Συμπεράσματα**

- Το μοντέλο διαμορφώσεων σε πολλαπλές ζώνες αποτελεί ένα γενικό μοντέλο για την οπτική υφή που οδηγεί σε περιγραφικά διανύσματα χαρακτηριστικών. Αυτό επιβεβαιώ-

θηκε θεωρητικά και πρακτικά σε εφαρμογές κατάτμησης και ταξινόμησης εικόνων.

- Η υφή μπορεί να οριστεί ενεργειακά, από το πλάτος και τη συχνότητα των κυρίαρχων ζωνοπερατών συνιστωσών της. Αυτή η ενέργεια αποτελεί και έναν ενδείκτη για ανίχνευση της ύπαρξης υφής σε εικόνες
- Η ανάλυση και η εξαγωγή χαρακτηριστικών γίνεται αποδοτικότερη χρησιμοποιώντας την κυρίαρχη αυτή ενεργειακή ποσότητα και τη μέθοδο κανονικοποιημένης αποδιαμόρφωσης που παρακάμπτει τα προβλήματα διακριτοποίησης
- Τα δομικά στοιχεία της υφής με βάση το μοντέλο είναι ταλαντώσεις με διαμορφωμένα πλάτη και φάσεις σε διαφορετικές ζώνες συχνοτήτων. Οι υποσυνιστώσεις αυτές διαχωρίζονται από την πληροφορία μακροδομών των εικόνων και συνθέτουν τη συνιστώσα της υφής.
- Οι εφαρμογές της υφής ευνοούνται από το διαχωρισμό και τη διαφορετική επεξεργασία της πληροφορίας μικροδομής.
- Οι ιδέες αυτές βρίσκουν εφαρμογή και στην ανάλυση της μικροδομής γενικών σημάτων, όπως διαπιστώθηκε για ακουστικά σήματα και την ενεργειακή ανίχνευση φωνητικής δραστηριότητας.

### 9.3 Μελλοντική έρευνα

Ένα ανοιχτό ζήτημα σε σχέση με την περιγραφή και ανάλυση της υφής είναι η αναζήτηση μεγαλύτερης ευελιξίας στο μοντέλο χωρικών διαμορφώσεων. Αυτό σχετίζεται με παραμετρους όπως είναι η επιλογή του αριθμού των συνιστωσών, ο σχεδιασμός της συστοιχίας των φίλτρων αλλά και η προσαρμοστικότητα για διαφορετικές κατηγορίες εικόνων. Επιπλέον, τα ημιτονοειδή μοντέλα, παρ' ότι καλύπτουν πολλές κατηγορίες υφής, δυσκολεύονται να ανταποκριθούν σε δομημένες ή μη κατευθυντικές υφές.

Η μαθηματική και φυσική σύνδεση των χωρικών διαμορφώσεων με τις συναρτήσεις ταλάντωσης που ορίστηκαν από το Meyer [214] για μοντελοποίηση και διαχωρισμό της υφής, είναι πολλά υποσχόμενη καθώς οι αναλογίες ανάμεσα στα δύο μοντέλα είναι χαρακτηριστικές. Οι ταλαντώσεις υφής, που περιγράφονται σε ένα χώρο μαθηματικών συναρτήσεων με αυστηρές μαθηματικές ιδιότητες, θα μπορούσαν να ταυτιστούν με τις μη-στατικές ταλαντώσεις της έντασης φωτεινότητας και των τοπικά αναδυόμενων συχνοτήτων. Επιπλέον, μελλοντικές κατευθύνσεις στα θέματα της αποσύνθεσης εικόνων περιλαμβάνουν διαφορετικούς περιορισμούς και μοντέλα ανακατασκευής, καλύτερο διαχωρισμό της υφής από το θόρυβο και αποδοτικότερα σχήματα διακριτοποίησης και επίλυσης των εμπλεκόμενων συστημάτων ΜΔΕ.

Προς μια ενιαία αντιμετώπιση της υφής με βάση τα δομικά της στοιχεία έγινε μια πρώτη προσέγγιση στο πλαίσιο των χωρικών διαμορφώσεων και των κυρίαρχων συνιστωσών της. Τα θεμελιώδη στοιχεία ή primitives της υφής θεωρήθηκαν οι χωρικά τοπικά εμφανιζόμενες

συνιστώσες διαμόρφωσης σε στενές ζώνες συχνοτήτων. Τέτοιες συνιστώσες εντοπίζονται και οριοθετούν περιοχές, καλύπτουν διαφορετικές κλίμακες και μπορούν να ορισθούν μέσω των παραμέτρων διαμόρφωσης σε ένα χωρικό πλέγμα. Μέσω της κυρίαρχης ενέργειας διαμορφώσεων υφής, ποσοτικοποιείται η συγκέντρωση και η παρουσία των κυρίαρχων primitives διαμόρφωσης σε κάθε θέση και καθορίζεται αν μια περιοχή είναι ή διαμέτει κατ' αυτόν τον τρόπο υφή. Τέτοια οικουμενικά primitives της υφής μπορούνε να «μαθαίνονται» από σύνολα εικόνων και να χρησιμοποιηθούν για αναγνώριση υφής και αντικειμένων.

Τα επιμέρους στάδια και διαδικασίες που εξετάστηκαν για την υφή μπορούν να ενσωματώθούν σε ένα ολοκληρωμένο σύστημα ανάλυσης και προεπεξεργασίας υφής (texture front-end) βασισμένο στα μοντέλα διαμόρφωσης, το οποίο θα τροφοδοτεί εφαρμογές υπολογιστικής όρασης ανωτέρου επιπέδου (higher-level vision), όπως είναι η αναγνώριση αντικειμένων και η ανακατασκευή 3-Δ σκηνών.

Για την εφαρμογή στις εικόνες εδαφοτομών είναι αναγκαία η εύρεση επιπλέον τρόπων συσχετισμού της υπολογιστικής, αυτόματης ανάλυσης με γεωλογικές παραμέτρους, βιολογικές παρατηρήσεις και ανθρώπινες εκτιμήσεις. Αρχετά ζητήματα ανάμεσα στα πεδία της υπολογιστικής όρασης και της γεωλογίας χρειάζεται να διερευνηθούν όπως ποιες ιδιότητες της υφής (π.χ. τραχύτητα, προσανατολισμός, κλίμακα) χαρακτηρίζουν τις ξεχωριστές δομές του εδάφους όπως απεικονίζονται στις τομές, ποια εξαγώμενα χαρακτηριστικά τις περιγράφουν καλύτερα, ποια η σημασία διαφορετικής «πυκνότητας» και παρουσίας υφής κ.ά. Προς αυτή την κατεύθυνση είναι αναγκαίος ο χειροκίνητος καθορισμός στη βάση των διαφορετικών κατηγοριών δομών και πιθανότατα και των ορίων τους σε κάθε εικόνα μεμονωμένα (manual segmentation and annotation). Αυτό θα επιτρέψει τη συστηματική αποτίμηση των μεθόδων στο σύνολο των δεδομένων και την εξαγωγή συγκεντρωτικών αποτελεσμάτων.

Στο θέμα της ανίχνευσης των μικροδομών της φωνής, το οποίο εξετάστηκε ανεξάρτητα σε ξεχωριστό κεφάλαιο, 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 υφής, με εφαρμογές σε θέματα αποθηρυβοποίησης, αποκατάστασης χαμένης πληροφορίας, κατάτμησης φωνής και αναγνώρισης ομιλίας. Για την ώρα, πάντως, δεν υπάρχει η φυσική ερμηνεία ενός ακουστικού φαινομένου αντίστοιχου της οπτικής υφής.

## 9.4 Επίλογος

Η υφή επιφανειών είναι ένα από τα οπτικά χαρακτηριστικά του φυσικού κόσμου που αναπαριστώνται μέσω συμβόλων σε βιολογικά συστήματα όρασης. Συμμετέχει τόσο σε ένα πλήθος αναπαραστάσεων χαμηλού επιπέδου όσο και στα μεγάλα προβλήματα της αναγνώρισης και ανακατασκευής σκηνών σε τεχνητά συστήματα όρασης. Κωδικοποιεί πληροφορία σχετική με το είδος, τη σύσταση, τη χωρική τοποθέτηση αλλά και την προέλευση ή την εξέλιξη των αντικειμένων και των προτύπων που χαρακτηρίζει. Και ενώ έχει μελετηθεί με όρους βιολο-

γικούς, ψυχοφυσικούς, αντιληπτικούς και γνωσιακούς και υπολογιστικούς, εξακολουθεί να χαρακτηρίζεται από αβεβαιότητα σχετικά με τον ορισμό και την περιγραφή του ως ένα οπτικό φαινόμενο. Το γεγονός αυτό από μόνο του προσδίδει βαρύτητα στην έρευνα που σχετίζεται με την επεξεργασία, την ανάκτηση και την αξιοποίηση της πληροφορίας υφής σε εικόνες. Επιπλέον, η σημασία της λεπτομέρειας σε εικόνες αιτιολογείται και από το πλήθος των εφαρμογών, στις οποίες εμπλέκεται άμεσα ή ως ενδιάμεσο στάδιο η ανάλυση της υφής. Όπως οποιαδήποτε επιστήμη του νου που φηλαρίζει αιρετικά την τεχνητή νοημοσύνη, μια ολοκληρωμένη υπολογιστική προσέγγιση της υφής μπορεί να συμβάλλει στην κατανόηση των μηχανισμών όρασης, αντίληψης, συνειδητότητας του εαυτού και συνειδητοποίησης του φυσικού κόσμου.

□

# Ορολογία

**amplitude (of) oscillation, modulation:** πλάτος ή εύρος ταλάντωσης, διαμόρφωσης

**attentive (vision, perception):** προσεχτική (όραση, αντίληψη), με σχολαστική επισκόπηση

**benchmark (test):** (δοκιμή) επιδόσεων

**bottom-up (processing):** συνθετική ή άνωθεν (επεξεργασία)

**boundaries:** όρια, σύνορα

**cognitive (sciences):** γνωσιακές επιστήμες

**classification:** ταξινόμηση

**clustering:** ομαδοποίηση

**complex (cortical) cells:** σύνθετα κύτταρα (νευρώνες)

**component:** συνιστώσα

**contrast (color, intensity):** αντίθεση (χρώματος, έντασης φωτεινότητας)

**contours:** περιγράμματα

**curve evolution:** εξέλιξη καμπυλών

**curvature:** καμπυλότητα

**decomposition (image):** αποσύνθεση, αποδόμηση (εικόνας), ανάλυση στα συνθετικές της συνιστώσες

**dominant component(s):** κυρίαρχη(ες) συνιστώσα(ες)

**diffusion (or heat) equation:** εξίσωση διάχυσης

**fronts (evolving):** καμπύλες (εξελισσόμενες)

**early vision:** πρώιμη όραση

**edges:** ακμές

**emergent frequencies:** εμφανιζόμενες ή αναδυόμενες συχνότητες

**feature:** γνώρισμα, χαρακτηριστικό

**filterbank:** συστοιχία φίλτρων

**Geodesic Active Contours (Regions):** γεωδαιτικά ενεργά περιγράμματα (περιοχές)

**level sets:** επιπεδοσύνολα

**likelihood:** πιθανοφάνεια

**localization:** εντοπισμός, τοπικοποίηση

**low-level (processing, vision):** χαμηλού επιπέδου (επεξεργασία, όραση)

**modulations:** διαμορφώσεις

**multiband:** πολυζωνικό

**multichannel:** πολυκαναλικό

**multipfrequency:** πολυσυχνοτικό

**multivariate:** πολυμεταβλητό

**narrowband (component):** στενοζωνική (συνιστώσα)

**oscillations:** ταλαντώσεις

**orientation:** προσανατολισμός

**perception:** αντίληψη, αντιληπτικότητα

**preattentive (processing):** (επεξεργασία) χωρίς σχολαστική επισκόπηση

**primal sketch:** πρωταρχικό σκίτσο

**primary visual cortex:** πρωτεύων οπτικός φλοιός

**primitives (image, texture):** στοιχειώδη (αρχέγονα) χαρακτηριστικά (εικόνας, υφής)

**principal (component):** πρωτεύουσα συνιστώσα

**quadrature (pair) filters:** φίλτρα ορθογωνισμού φάσης

**(spatial) receptive fields:** (χωρικά) υποδεκτικά πεδία

**(photo)receptors:** (Φωτο)ϋποδοχείς, (φωτο)δέκτες

**region:** Περιοχή

**Region Competition (segmentation):** κατάτμηση με Ανταγωνισμό Περιοχών

## *Oρολογία*

**residual:** παραμένον

**retina:** αμφιβληστροειδής (χιτώνας)

**scale (coarse, fine):** κλίμακα (αδρομερής, λεπτομερής)

**segmentation:** κατάτμηση

**segregation:** διαχωρισμός

**simple (cortical) cells:** απλά κύτταρα (νευρώνες)

**soilsection (image):** (εικόνα) εδαφοτομής

**striate cortex (V1, primary vis. cortex):** γραμμωτός φλοιός

**supervised (learning, segmentation, classification):** επιβλεπόμενη (μάθηση, κατάτμηση, ταξινόμηση)

**textons:** τεξτόνια, βασικές μονάδες αυθόρυμητης αντίληψης υφής

**texture (visual, natural):** υφή (οπτική, φυσική)

**texture (optical, haptic, retinal):** υφή

**top-down (processing):** κάτωθεν (επεξεργασία)

**wavelets:** κυματίδια

**event detection:** ανίχνευση γεγονότων ενδιαφέροντος

**formants:** θεμελιώδεις συντονισμοί (φωνής)

**(speech) endpoints:** σύνορα, τερματικά σημεία (φωνής, λέξεων, φράσεων)

**resonance:** συντονισμός (φωνής)

**voice activity detection:** ανίχνευση φωνητικής δραστηριότητας



# Βιβλιογραφία

- [1] ACAR, R., AND VOGEL, C. R. Analysis of Bounded Variation penalty methods for ill-posed problems. *Inverse Problems* 10 (1994), 1217–1229.
- [2] 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.
- [3] ADELSON, E., AND BERGEN, J. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A* 2, 2 (Feb. 1985), 284–299.
- [4] ALOIMONOS, Y. Shape from texture. *Biological Cybernetics* 58, 5 (Feb. 1988), 345–360.
- [5] AMBROSIO, L., FUSCO, N., AND PALLARA, D. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, 2000.
- [6] AUBERT, G., AND KORNPROBST, P. *Mathematical problems in image processing: Partial Differential Equations and the Calculus of Variations*, 2 ed., vol. 147 of *Applied Mathematical Sciences*. Springer, New York, 2006.
- [7] 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.
- [8] AUJOL, J.-F., AND CHAMBOLLE, A. Dual norms and image decomposition models. *Intl. Journal of Computer Vision* 63, 1 (June 2005), 85–104.
- [9] AUJOL, J.-F., AND CHAN, T. F. Combining geometrical and textured information to perform image classification. *Journal of Visual Communication and Image Representation* 17 (2006), 1004–1023.
- [10] AUJOL, J.-F., AND GILBOA, G. Constrained and SNR-based solutions for TV-Hilbert space image denoising. *Journal of Mathematical Imaging and Vision* 26, 1-2 (Nov. 2006), 217–237.
- [11] AUJOL, J.-F., GILBOA, G., CHAN, T., AND OSHER, S. Structure-texture image decomposition-modeling, algorithms and parameter selection. *Intl. Journal of Computer Vision* 67, 1 (Feb. 2006), 111–136.
- [12] AUJOL, J.-F., AND KANG, S.-H. Color image decomposition and restoration. *Journal of Visual Communication and Image Representation* 17 (2006), 916–928.
- [13] BABAUD, J., WITKIN, A., BAUDIN, M., AND DUDA, R. Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1, 8 (Aug. 1986), 26–33.

- [14] BAE, S., PARIS, S., AND DURAND, F. Two-scale tone management for photographic look. In *SIGGRAPH 06* (2006), ACM Press.
- [15] BALLARD, D. H., AND BROWN, C. M. *Computer Vision*. Prentice-Hall, 1982.
- [16] 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.
- [17] BAR-JOSEPH, Z., EL-YANIV, R., LISCHINSKI, D., AND WERMAN, M. Texture mixing and texture movie synthesis using statistical learning. *IEEE Transactions on Visualization and Computer Graphics* 07, 2 (2001), 120–135.
- [18] BARNESLEY, M. F. *Fractals Everywhere*. Academic Press, 1988.
- [19] BASTIAANS, M. Gabor's expansion of a signal into Gaussian elementary signals. *Proceedings of the IEEE* 68, 4 (Apr. 1980), 538–539.
- [20] BECK, J. Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology* 85 (1972), 1–19.
- [21] BECK, J. Textural segmentation, second-order statistics, and textural elements. *Biological Cybernetics* 48, 2 (1983), 125–130.
- [22] 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.
- [23] BELONGIE, S., CARSON, C., GREENSPAN, H., AND MALIK, J. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (Bombay, India, 1998).
- [24] BERGEN, J. R. Texture. In *The MIT Encyclopedia of the Cognitive Sciences*, R. A. Wilson and F. C. Keil, Eds. MIT Press, Cambridge, MA, 1999.
- [25] BERGEN, J. R., AND ADELSON, E. H. Early vision and texture perception. *Nature* 333 (May 1988), 363–364.
- [26] BERGEN, J. R., AND LANDY, M. S. Computational modeling of visual texture segregation. In *Computational Models of Visual Processing*, J. A. Movshon and M. S. Landy, Eds. MIT Press, Cambridge, MA, 1991, pp. 253–271.
- [27] 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.
- [28] BERTALMIO, M., SAPIRO, G., CASELLES, V., AND BALLESTER, C. Image inpainting. In *SIGGRAPH 00* (2000), ACM Press, pp. 417–424.
- [29] 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.
- [30] BESAG, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society B* 36 (1974), 192–236.

- [31] BESAG, J. On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society B* 48 (1986), 259–302.
- [32] 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.
- [33] BLOMGREN, P., AND CHAN, T. F. Color TV: total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing* 7, 3 (Mar. 1998), 304–309.
- [34] BOVIK, A., CLARK, M., AND GEISLER, W. Texture segmentation using a class of narrowband filters. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (Dallas, TX, Apr. 1987).
- [35] BOVIK, A. C. Analysis of multichannel narrow-band filters for image texture segmentation. *IEEE Transactions on Signal Processing* 39, 9 (Sept. 1991), 2025–2043.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] BOYD, S., AND VANDENBERGHE, L. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
- [40] BRODATZ, P. *Textures: A Photographic Album for Artists and Designers*. Dover, New York, 1966.
- [41] BROX, T., AND WEICKERT, J. A TV flow based local scale measure for texture discrimination. In *European Conf. on Computer Vision (ECCV)* (Prague, Czech Republic, 2004).
- [42] BROX, T., AND WEICKERT, J. A TV flow based local scale estimate and its application to texture discrimination. *Journal of Visual Communication and Image Representation* 17 (2006), 1053–1073.
- [43] 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.
- [44] BULLOCK, P., FEDOROFF, N., JONGERIUS, A., STOOPS, G., AND TURSINA, T. *Handbook for Soil Thin Section Description*. Waine Research Publications, Albrighton, Woverhampton, England, 1985.
- [45] 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.
- [46] CAELLI, T. *Visual Perception*. Pergamon Press, 1981.

- [47] CAELLI, T. Texture classification and segmentation algorithms in man and machines. *Spatial Vision* 7, 4 (1993), 277–292.
- [48] CAELLI, T., AND JULESZ, B. On perceptual analyzers underlying visual texture discrimination: Part I. *Biological Cybernetics* 28, 3 (Sept. 1978), 167 – 175.
- [49] CAMPBELL, F., AND ROBSON, J. Application of fourier analysis to the visibility of gratings. *The Journal of Physiology* 197, 3 (1968), 551–566.
- [50] CARSON, C., BELONGIE, S., GREENSPAN, H., AND MALIK, J. Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 8 (Aug 2002), 1026–1038.
- [51] CASELLES, V., KIMMEL, R., AND SAPIRO, G. Geodesic Active Contours. *Intl. Journal of Computer Vision* 22, 1 (Feb. 1997), 61–79.
- [52] CHAMBOLLE, A. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision* 20, 1-2 (Jan. 2004), 89–97.
- [53] CHAMBOLLE, A., AND LIONS, P.-L. Image recovery via total variation minimization and related problems. *Numerische Mathematik* 76, 2 (Apr. 1997), 167–188.
- [54] CHAN, T. F., OSHER, S., AND SHEN, J. The digital TV filter and nonlinear denoising. *IEEE Transactions on Image Processing* 10, 2 (Feb. 2001), 231–241.
- [55] CHAN, T. F., AND SHEN, J. Variational image inpainting. *Communications on Pure and Applied Mathematics* 58, 5 (2005), 579–619.
- [56] CHANG, T., AND KUO, C.-C. Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on Image Processing* 2, 4 (Oct. 1993), 429–441.
- [57] 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.
- [58] 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.
- [59] 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.
- [60] CHEN, S. S., DONOHO, D. L., AND SAUNDERS, M. A. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* 20, 1 (1998), 33–61.
- [61] CHEN, T., YIN, W., ZHOU, X. S., COMANICIU, D., AND HUANG, T. S. Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 9 (2006), 1519–1524.
- [62] COGGINS, J., AND JAIN, A. K. A spatial filtering approach to texture analysis. *Pattern Recognition Letters* 3 (1985), 195–203.
- [63] COHEN, L. On active contour models and balloons. *CVGIP: Image Understanding* 53, 2 (May 1991), 211–218.
- [64] COHEN, L. *Time-Frequency Analysis*. Prentice Hall, 1995.

- [65] CORMACK, L. K. Computational models of early human vision. In *Handbook of Image and Video Processing*, A. Bovik, Ed., 2 ed. Academic Press, 2005, pp. 325–345.
- [66] 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.
- [67] CULA, O. G., AND DANA, K. J. 3D texture recognition using bidirectional feature histograms. *Intl. Journal of Computer Vision* 59, 1 (Aug. 2004), 33 – 60.
- [68] 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.
- [69] DAUBECHIES, I., AND TESCHKE, G. Variational image restoration by means of wavelets. *Applied and Computational Harmonic Analysis* 19, 1 (2005), 1–16.
- [70] 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.
- [71] DAUGMAN, J. G. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research* (1980), 847–856.
- [72] 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.
- [73] 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.
- [74] DAUGMAN, J. G., AND DOWNING, C. J. Demodulation, predictive coding, and spatial vision. *Journal of the Optical Society of America A* 12, 4 (Apr. 1995), 641–660.
- [75] DECARLO, D., AND SANTELLA, A. Stylization and abstraction of photographs. In *SIGGRAPH 02* (2002), ACM Press, pp. 769–776.
- [76] 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.
- [77] DEVALOIS, R. L., AND DEVALOIS, K. K. *Spatial Vision*. Oxford University Press, New York, 1988.
- [78] DI-ZENZO, S. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing* 33, 1 (1986), 116–125.
- [79] DIMITRIADIS, D., AND MARAGOS, P. Robust energy demodulation based on continuous models with application to speech recognition. In *Eurospeech* (2003).
- [80] DO, M., AND VETTERLI, M. Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models. *IEEE Transactions on Multimedia* 4, 4 (Dec. 2002), 517–527.
- [81] DO, M., AND VETTERLI, M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing* 11, 2 (Feb. 2002), 146–158.

- [82] DONOHO, D. L. De-noising by soft-thresholding. *IEEE Transactions on Image Processing* 41, 3 (May 1995), 613–627.
- [83] DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G., AND PICARD, D. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society. Series B* 57, 2 (1995), 301–369.
- [84] DUDA, R. O., HART, P. E., AND STORK, D. G. *Pattern Classification*, 2 ed. Wiley-Interscience, 2001.
- [85] DUIN, R., JUSZCZAK, P., PACLIK, P., PEKALSKA, E., DE RIDDER, D., AND TAX, D. PRTools4, A Matlab Toolbox for Pattern Recognition, Delft University of Technology, 2004. [www.prtools.org](http://www.prtools.org).
- [86] DUNN, D., AND HIGGINS, W. E. Optimal Gabor filters for texture segmentation. *IEEE Transactions on Image Processing* 4, 7 (July 1995), 947–964.
- [87] 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.
- [88] DURAND, F., AND DORSEY, J. Fast bilateral filtering for the display of high-dynamic-range images. In *SIGGRAPH 02* (2002), ACM Press, pp. 257–266.
- [89] EFROS, A. A., AND FREEMAN, W. T. Image quilting for texture synthesis and transfer. In *ACM SIGGRAPH 01* (Los Angeles, CA, USA, Aug. 2001), pp. 341–346.
- [90] EFROS, A. A., AND LEUNG, T. K. Texture synthesis by non-parametric sampling. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (Corfu, Greece, Sept. 1999), pp. 1033–1038.
- [91] ELAD, M., STARCK, J.-L., QUERRE, P., AND DONOHO, D. L. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Applied and Computational Harmonic Analysis* 19, 3 (Nov. 2005), 340–358.
- [92] EVANGELOPOULOS, G., KOKKINOS, I., AND MARAGOS, P. Advances in variational image segmentation using AM-FM models: Regularized demodulation and probabilistic cue integration. In *Proc. VLSM 2005, LNCS* (2005), vol. 3275, pp. 121–136.
- [93] EVANGELOPOULOS, G., AND MARAGOS, P. Speech event detection using multiband modulation energy. In *Proc. Interspeech '05 - Eurospeech* (Lisbon, Portugal, 2005).
- [94] EVANGELOPOULOS, G., AND MARAGOS, P. Multiband modulation energy tracking for noisy speech detection. *IEEE Trans. Audio Speech Language Processing* 14, 6 (Nov. 2006), 2024–2038.
- [95] FDEZ-VALDIVIA, J., GARCIA, J., MARTINEZ-BAENA, J., AND FDEZ-VIDAL, X. R. The selection of natural scales in 2D images using adaptive Gabor filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 5 (May 1998), 458–469.
- [96] FIELD, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A* 4, 12 (Dec. 1987), 2379–2394.
- [97] FOGEL, I., AND SAGI, D. Gabor filters as texture discriminator. *Biological Cybernetics* 61 (1989), 103–113.

- [98] FOMIN, S. V., AND GELFAND, I. M. *Calculus of Variations*. Dover Publications, 2000.
- [99] FORSYTH, D. A., AND PONCE, J. *Computer Vision: A Modern Approach*. Prentice-Hall, 2003.
- [100] FREEMAN, W. T., AND ADELSON, E. H. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13, 6 (1991), 891–906.
- [101] FRIEDLANDER, B., AND FRANCOS, J. Estimation of amplitude and phase parameters of multicomponent signals. *IEEE Transactions on Signal Processing* 43, 4 (Apr. 1995), 917–926.
- [102] GABOR, D. Theory of communication. *Journal Inst. of Elec. Eng. London* 93, III (1946), 429–457.
- [103] 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.
- [104] 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.
- [105] GIBSON, J. J. *The perception of the visual world*. Houghton Mifflin, Boston, MA, 1950.
- [106] GIBSON, J. J. *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1987.
- [107] GILBOA, G., SOCHEN, N., AND ZEEVI, Y. Y. Image enhancement and denoising by complex diffusion processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 8 (Aug. 2004), 1020–1036.
- [108] GILBOA, G., SOCHEN, N., AND ZEEVI, Y. Y. Variational denoising of partly-textured images by spatially varying constraints. *IEEE Transactions on Image Processing* 15, 8 (Aug. 2006), 2281–2289.
- [109] GOUTSIAS, J. Markov random fields: Interacting particle systems for statistical image modeling and analysis. Tech. Rep. 96-01, John Hopkins University, ECE, 1996.
- [110] 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.
- [111] GROSSAUER, H. A combined PDE and texture synthesis approach to inpainting. In *European Conf. on Computer Vision (ECCV)* (Prague, Czech Republic, 2004), pp. 214–224.
- [112] GUO, C. E., ZHU, S. C., AND WU, Y. N. Towards a mathematical theory of primal sketch and sketchability. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (Nice, France, 2003), vol. 2, pp. 1228–1235.
- [113] GUO, C. E., ZHU, S. C., AND WU, Y. N. Primal sketch: Integrating structure and texture. *Computer Vision and Image Understanding* 106, 1 (Apr. 2007), 5–19.
- [114] GUO, G.-E., ZHU, S.-C., AND WU, Y. Modeling visual patterns by integrating descriptive and generative methods. *Intl. Journal of Computer Vision* 53, 1 (June 2003), 5 – 29.

- [115] HADDAD, A., AND MEYER, Y. Variational methods in image processing. Tech. Rep. 2006-20, CMLA, 2006.
- [116] 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.
- [117] HARALICK, R., AND SHAPIRO, L. *Computer and Robot Vision*. Addison Wesley, 1992.
- [118] HARALICK, R. M. Statistical and structural approaches to texture. *Proc. IEEE* 67 (1979), 786–804.
- [119] HARTLEY, R., AND ZISSEMAN, A. *Multiple view geometry in computer vision*. Cambridge University Press, New York, NY, USA, 2000.
- [120] HAVLICEK, J., AND BOVIK, A. C. Multi-component AM-FM image models and wavelet-based demodulation with component tracking. In *Intl. Conf. on Image Processing (ICIP)* (Austin, TX, Nov. 1994), pp. I41–I45.
- [121] 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.
- [122] 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.
- [123] HAVLICEK, J. P., HARDING, D. S., AND BOVIK, A. C. Reconstruction from the multi-component AM-FM image representation. In *Intl. Conf. on Image Processing (ICIP)* (Washington, DC, Oct. 1995), pp. II280–II283.
- [124] HAVLICEK, J. P., HARDING, D. S., AND BOVIK, A. C. The multi-component AM-FM image representation. *IEEE Transactions on Image Processing* 5, 6 (June 1996), 1094–1100.
- [125] 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.
- [126] HAVLICEK, J. P., HAVLICEK, J. W., AND BOVIK, A. C. The analytic image. In *Intl. Conf. on Image Processing (ICIP)* (Santa Barbara, CA, Oct. 1997), pp. 446–449.
- [127] 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. Academic Press, 2005, pp. 377–395.
- [128] HEEGER, D. J., AND BERGEN, J. R. Pyramid-based texture analysis/synthesis. In *SIGGRAPH 95* (1995), ACM Press, pp. 229–238.
- [129] HEILER, M., AND SCHNORR, C. Natural image statistics for natural image segmentation. *Intl. Journal of Computer Vision* 63, 1 (June 2005), 5–19.
- [130] HUBEL, D. *Eye, Brain and Vision*. Scientific American Library, 1988.
- [131] 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.

- [132] HUBEL, D. H., AND WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)* 195 (1968), 215–243.
- [133] HUBEL, D. H., AND WIESEL, T. N. *Brain and Visual Perception: The Story of a 25-Year Old Collaboration*. Oxford University Press, 2005.
- [134] 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.
- [135] JACOBSON, L. D., AND WECHSLER, H. Joint spatial/spatial-frequency representation. *Signal Processing* 14, 1 (1988), 37–68.
- [136] 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.
- [137] JAIN, A. K., DUIN, R. P., AND MAO, J. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 1 (2000), 4–37.
- [138] JAIN, A. K., AND FARROKHNA, F. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition* 24, 12 (1991), 1167–1186.
- [139] JONES, D. G., AND MALIK, J. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *European Conf. on Computer Vision (ECCV)* (1992), pp. 395–410.
- [140] JONES, G., AND BOASHASH, B. Instantaneous frequency, instantaneous bandwidth, and the analysis of multicomponent signals. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (1990), pp. 2467–2470.
- [141] JONES, J., AND PALMER, L. A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology* 58, 6 (Dec. 1987), 1233–1258.
- [142] JULESZ, B. Visual pattern discrimination. *IRE Trans. on Information Theory* IT-8 (Feb. 1962), 84–92.
- [143] JULESZ, B. Textons, the elements of texture perception and their interactions. *Nature* 290 (Mar. 1981), 91–97.
- [144] JULESZ, B. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics* 41 (1981), 131–181.
- [145] JULESZ, B. *Dialogues on Perception*. MIT Press, 1995.
- [146] 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.
- [147] 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.
- [148] 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.
- [149] 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).

- [150] KANIZSA, G. *Organization in Vision: Essays on Gestalt Perception*. Praeger Publishers, New York, 1979.
- [151] KAPLAN, L. M. Extended fractal analysis for texture classification and segmentation. *IEEE Transactions on Image Processing* 8, 11 (Nov. 1999), 1572–1585.
- [152] KASHYAP, R., AND EOM, K. Texture boundary detection based on the long correlation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 1 (Jan. 1989), 58–67.
- [153] KASS, M., WITKIN, A., AND TERZOPOULOS, D. Snakes: Active Contour Models. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (London, UK, 1987).
- [154] KAY, S. M. *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1993.
- [155] 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.
- [156] KICHENASSAMY, S., KUMAR, A., OLVER, P., TANNENBAUM, A., AND YEZZI, A. Gradient flows and geometric active contour models. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (Boston, MA, USA, 1995).
- [157] KIMMEL, R., MALLADI, R., AND SOCHEN, N. Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images. *Intl. Journal of Computer Vision* 39, 2 (Sept. 2000), 111–129.
- [158] KNUTSSON, H., AND GRANLUND, G. H. Texture Analysis Using Two-Dimensional Quadrature Filters. In *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management (CAPAIDM)* (Pasadena, Oct. 1983).
- [159] KOENDERINK, J. J. The structure of images. *Biological Cybernetics* 50, 5 (Aug. 1984), 363–370.
- [160] KOENDERINK, J. J., AND VAN DOORN, A. J. Representation of local geometry in the visual system. *Biological Cybernetics* 55, 6 (Mar. 1987), 367–375.
- [161] KOFFKA, K. *Principles of Gestalt Psychology*. Harcourt, Brace, New York, 1963.
- [162] KOKKINOS, I. *Synergy between Image Segmentation and Object Recognition using Geometrical and Statistical Computer Vision Techniques*. PhD thesis, National Technical University of Athens, May 2006.
- [163] KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Advances in texture analysis: Energy dominant component and multiple hypothesis testing. In *Intl. Conf. on Image Processing (ICIP)* (Singapore, 2004), vol. 3, pp. 1509–1512.
- [164] KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Modulation-feature based textured image segmentation using curve evolution. In *Intl. Conf. on Image Processing (ICIP)* (Singapore, 2004), vol. 2, pp. 1204–1207.
- [165] KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Texture analysis and segmentation using modulation features, generative models and weighted curve evolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1 (Jan. 2009), 142–157.

- [166] KOKKINOS, I., AND MARAGOS, P. A Detection-Theoretic Approach to Texture and Edge Discrimination. In *3rd Int.l Texture Workshop* (2005).
- [167] KOLMOGOROV, A. N., AND FOMIN, S. V. *Elements of the Theory of Functions and Functional Analysis*. Dover Publications, 1999.
- [168] KRUIZINGA, P., AND PETKOV, N. Nonlinear operator for oriented texture. *IEEE Transactions on Image Processing* 8, 10 (Oct. 1999), 1395–1407.
- [169] 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.
- [170] KULIKOWSKI, J. J., MARCELJA, S., AND BISHOP, P. O. Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics* 43, 3 (Apr. 1982), 187–198.
- [171] 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.
- [172] 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.
- [173] LANDY, M. S. Texture perception. In *Encyclopedia of Neuroscience*, G. Adelman and B. H. Smith, Eds., 3 ed. Elsevier, Amsterdam, 2004.
- [174] 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.
- [175] LARKIN, K. G., BONE, D. J., AND OLDFIELD, M. A. Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform. *Journal of the Optical Society of America A* 18, 8 (Aug. 2001), 1862–1870.
- [176] LEE, T. S. Image representation using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, 10 (1996), 959–971.
- [177] LEE, T. S., MUMFORD, D., AND YUILLE, A. Texture segmentation by minimizing vector valued energy functionals. In *European Conf. on Computer Vision (ECCV)* (Santa Margherita Ligure, Italy, 1992).
- [178] 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.
- [179] LINDEBERG, T. *Scale-space theory in computer vision*. Kluwer Academic, 1994.
- [180] LINDEBERG, T. Feature detection with automatic scale selection. *Intl. Journal of Computer Vision* 30, 2 (Nov. 1998), 79–116.
- [181] LIU, J., AND YANG, Y. Multiresolution color image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, 7 (1994), 689–700.
- [182] 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.

- [183] MALGOUVRES, F. Minimizing the total variation under a general convex constraint for image restoration. *IEEE Transactions on Image Processing* 11, 12 (Dec. 2002), 1450–1456.
- [184] 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.
- [185] MALIK, J., BELONGIE, S., LEUNG, T., AND SHI, J. Textons, Contours and Regions: Cue integration in image segmentation. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (Corfu, Greece, Sept. 1999).
- [186] 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.
- [187] MALIK, J., AND PERONA, P. A computational model of texture segmentation. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (San Diego, CA, June 1989), pp. 326–332.
- [188] MALIK, J., AND PERONA, P. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A* 7, 5 (May 1990), 923–932.
- [189] 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.
- [190] MALLAT, S. *A Wavelet Tour of Signal Processing*, 2 ed. Academic Press, 1999.
- [191] 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.
- [192] MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 7 (July 1989), 674–693.
- [193] MANDELBROT, B. B. *The Fractal Geometry of Nature*. Freeman, San Francisco, 1982.
- [194] MANDELBROT, B. B., AND NESS, J. W. V. Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10 (Oct. 1968), 422–437.
- [195] MANJUNATH, B., AND MA, W. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8, 18 (Aug. 1996), 837–842.
- [196] 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.
- [197] MARAGOS, P. *Image Analysis and Computer Vision*. National Technical University of Athens, June 2005.
- [198] 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.
- [199] 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.

- [200] MARAGOS, P., KAISER, J. F., AND QUATIERI, T. F. On separating amplitude from frequency modulations using energy operators. In *Intl. Conf. on Speech and Signal Processing* (San Francisco, CA, Mar. 1992), pp. II:1–4.
- [201] 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.
- [202] 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.
- [203] MARAGOS, P., SOFOU, A., STAMOU, G. B., TZOUVARAS, V., STAMOU, G. P., AND PAPATHEODOROU, E. Image analysis of soil micromorphology: Feature extraction, segmentation and quality inference. *J. Applied Signal Processing* 6 (2004), 902–912.
- [204] 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.
- [205] MARR, D. *Vision*. W.H. Freeman, San Francisco, CA, 1982.
- [206] MARR, D., AND HILDRETH, E. Theory of edge detection. *Proc. R. Soc. Lond. B*, 207 (1980), 187 – 217.
- [207] MARTIN, D., FOWLKES, C., TAL, D., AND MALIK, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (Vancouver, BC, Canada, 2001), vol. 02, pp. 416–423.
- [208] 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.
- [209] 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.
- [210] MCCLOUD, S. *Understanding Comics, the invisible art*. Harper Paperbacks, 1994.
- [211] 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.
- [212] MEYER, F., AND MARAGOS, P. Nonlinear scale-space representation with morphological leveling. *Journal of Visual Communication and Image Representation* 11, 2 (June 2000), 245–265.
- [213] MEYER, F. G., AVERBUCH, A. Z., AND COIFMAN, R. R. Multilayered image representation: Application to image compression. *IEEE Transactions on Image Processing* 11, 9 (Sept. 2002), 1072–1080.
- [214] MEYER, Y. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, vol. 22 of *University Lecture Series*. AMS, 2001.

- [215] MINSKY, M., YOUNG MING, O., STEELE, O., BROOKS, F. P., AND BEHENSKY, M. Feeling and seeing: issues in force display. In *Symposium on Interactive 3D Graphics* (Snowbird, Utah, USA, 1990), ACM Press, pp. 235 – 241.
- [216] MIT VISION AND MODELLING GROUP. VisTex Database, 1995. <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>.
- [217] MITRA, S. K., LI, H., LIN, I.-S., AND YU, T.-H. A new class of nonlinear filters for image enhancement. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (Toronto, Canada, May 1991), pp. 2525–2528.
- [218] MOREL, J., AND SOLIMINI, S. *Variational Methods in Image Segmentation*. Birkhauser, Boston, 1995.
- [219] MORRONE, M., AND OWENS, R. Feature detection from local energy. *Pattern Recognition Letters* 6 (Dec. 1987), 303–313.
- [220] MORRONE, M. C., AND BURR, D. Feature detection in human vision: a phase-dependent energy model. *Proc. of the Royal Society of London. Series B* 235, 1280 (Dec. 1988), 221–245.
- [221] MUMFORD, D., AND SHAH, J. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics* 42, 5 (1989), 577–685.
- [222] NAJMAN, L., AND SCHMITT, M. Watershed of a continuous function. *Signal Processing* 38, 7 (1994), 99–112.
- [223] NIKOLOVA, M. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision* 20, 1-2 (Jan. 2004), 99–120.
- [224] OLSHAUSEN, B. A., AND FIELD, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 6583 (June 1996), 607–609.
- [225] OLSHAUSEN, B. A., AND FIELD, D. J. Sparse coding with an overcomplete basis set: A strategy employed by V1 ? *Vision Research* 37, 23 (Nov. 1997), 3311–3325.
- [226] 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.
- [227] OSHER, S., SOLE, A., AND VESE, L. Image decomposition, image restoration, and texture modeling using total variation minimization and the  $H^{-1}$  norm. *SIAM Journal of Multiscale Modeling and Simulation* 1, 3 (2003), 349–370.
- [228] OSHER, S., SOLE, A., AND VESE, L. Image decomposition, image restoration, and texture modeling using total variation minimization and the  $h^{-1}$  norm. In *Intl. Conf. on Image Processing (ICIP)* (Barcelona, Spain, Nov. 2003).
- [229] PAPOULIS, A. *Probability, Random Variables and Stochastic Processes*, 3 ed. McGraw-Hill, 1991.
- [230] PARAGIOS, N., AND DERICHE, R. Geodesic Active Contours for supervised texture segmentation. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (1999).

- [231] 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.
- [232] 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.
- [233] PARAGIOS, N., AND OSHER, S. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer, 2002.
- [234] PATTICHIS, M. S., AND BOVIK, A. C. Analyzing image structure by multidimensional frequency modulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 5 (May 2007), 753–766.
- [235] PEITGEN, H.-O., AND SAUPE, D. *The Science of Fractal Images*. Springer-Verlag, Bremen, 1988.
- [236] PENTLAND, A. P. Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 6 (Nov. 1984), 661–674.
- [237] PERONA, P. Deformable kernels for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 5 (1995), 488–499.
- [238] PERONA, P., AND MALIK, J. Detecting and localizing edges composed of steps, peaks and roofs. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (1990), pp. 52–57.
- [239] 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.
- [240] PESQUET-POPESCU, B., AND VEHEL, J. Stochastic fractal models for image processing. *IEEE Signal Processing Magazine* 19, 5 (Sept. 2002), 48–62.
- [241] 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 (Feb. 1997), 83–96.
- [242] PETROU, M., AND SEVILLA, P. G. *Image Processing: Dealing With Texture*. Wiley, 2006.
- [243] 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.
- [244] 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.
- [245] 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.
- [246] PROTZ, R., SHIPITALO, M., MERMUT, A., AND FOX, C. Image-analysis of soils-present and future. *Geoderma* 40, 1-2 (1987), 115–125.
- [247] 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.

- [248] 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.
- [249] 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.
- [250] RANDEN, T., AND HUSOY, J. H. Filtering for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 4 (Apr. 1999), 291–310.
- [251] RAY, N., HAVLICEK, J., ACTON, S., AND PATTICHIS, M. Active contour segmentation guided by AM-FM DCA. In *Intl. Conf. on Image Processing (ICIP)* (Thessaloniki, Greece, Oct. 2001), pp. 78–81.
- [252] REED, T., AND WECHSLER, H. Segmentation of textured images and Gestalt organization using spatial/spatial-frequency representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 1 (Jan. 1990), 1–12.
- [253] 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.
- [254] RICHARDS, W., AND POLIT, A. Texture matching. *Kybernetik* 16, 3 (Sept. 1974), 155 – 162.
- [255] ROSENFIELD, 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.
- [256] ROUSSON, M., BROX, T., AND DERICHE, R. Active unsupervised texture segmentation on a diffusion-based feature space. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (Madison, Wisconsin, 2003).
- [257] RUDIN, L., AND OSHER, S. Total variation based image restoration with free local constraints. In *Intl. Conf. on Image Processing (ICIP)* (Austin, Texas, Nov. 1994), vol. 1, pp. 31–35.
- [258] 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.
- [259] 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).
- [260] 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.
- [261] SAPIRO, G., AND RINGACH, D. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Transactions on Image Processing* 5, 11 (Nov. 1996), 1582–1586.
- [262] SERRA, J. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [263] SETHIAN, J. A. A fast marching method for monotonically advancing fronts. *Proceedings of National Academy of Sciences* 93, 4 (1996), 1591–1595.

- [264] SETHIAN, J. A. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.
- [265] SHI, M., AND HEALEY, G. Hyperspectral texture recognition using a multiscale opponent representation. *IEEE Transactions on Geoscience and Remote Sensing* 41, 5 (2003), 1090–1095.
- [266] 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.
- [267] SMEULDERS, A. W., WORRING, M., SANTINI, S., GUPTA, A., AND JAIN, R. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 12 (2000), 1349–1380.
- [268] SOFOU, A., EVANGELOPOULOS, G., AND MARAGOS, P. Coupled geometric and texture PDE-based segmentation. In *Intl. Conf. on Image Processing (ICIP)* (Genova, Italy, 2005), pp. II–650–3.
- [269] 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), 394–398.
- [270] SOFOU, A., AND MARAGOS, P. PDE-based modelling of image segmentation using volumic flooding. In *Intl. Conf. on Image Processing (ICIP)* (Barcelona, Spain, 2003), vol. 2, pp. 431–434.
- [271] 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.
- [272] SOILLE, P. *Morphological Image Analysis*, 2 ed. Springer-Verlag, 2003.
- [273] STARCK, J.-L., CANDES, E., AND DONOHO, D. The curvelet transform for image denoising. *IEEE Transactions on Image Processing* 11, 6 (June 2002), 670–684.
- [274] 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.
- [275] STEIDL, G., WEICKERT, J., BROX, T., MRAZEK, P., AND WELK, M. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs. *SIAM Journal of Numerical Analysis* 42, 2 (2004), 686–713.
- [276] 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.
- [277] 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.
- [278] T. A. POGGIO, H. V., AND YUILLE, A. A Regularized Solution to Edge Detection. *J. Complexity* 4, 2 (Oct. 1988), 106–128.

- [279] TADMOR, E., NEZZAR, S., AND VESE, L. A multiscale image representation using hierarchical ( $BV, L_2$ ) decompositions. *SIAM Journal of Multiscale Modeling and Simulation* 2, 4 (2004), 554–579.
- [280] TANGSUKSON, T., AND HAVLICEK, J. P. AM-FM image segmentation. In *Intl. Conf. on Image Processing (ICIP)* (Vancouver, BC, Canada, 2000).
- [281] 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.
- [282] 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.
- [283] THERRIEN, C. W. An estimation-theoretic approach to terrain image segmentation. *Computer Vision, Graphics, and Image Processing* 22, 3 (June 1983), 313–326.
- [284] TIKHONOV, A., AND ARSENIN, V. *Solutions of Ill-Posed Problems*. Wiley, New York, 1977.
- [285] TSCHUMPERLE, D., AND DERICHE, R. Diffusion PDEs on vector-valued images. *IEEE Signal Processing Magazine* 19, 5 (Sept. 2002), 16–25.
- [286] TSCHUMPERLE, D., AND DERICHE, R. Vector-valued image regularization with PDEs: a common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 4 (2005), 506–517.
- [287] 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.
- [288] 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.
- [289] TURING, A. M. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond.* 327, B (1952), 37–72.
- [290] TURNER, M. R. Texture discrimination by Gabor functions. *Biological Cybernetics* 55, 2-3 (Nov. 1986), 71–82.
- [291] ULLMAN, S. *High-Level Vision: Object-Recognition and Visual Cognition*. MIT Press, Cambridge, MA, 2000.
- [292] UNSER, M. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing* 4, 11 (Nov. 1995), 1549–1560.
- [293] 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.
- [294] VARMA, M., AND ZISSERMAN, A. Classifying images of materials: Achieving viewpoint and illumination independence. In *European Conf. on Computer Vision (ECCV)* (Copenhagen, Denmark, May 2002), pp. 255–271.
- [295] VARMA, M., AND ZISSERMAN, A. Texture classification: Are filter banks necessary? In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (Madison, Wisconsin, 2003), vol. 02, p. 691.

- [296] VARMA, M., AND ZISSEMAN, A. A statistical approach to texture classification from single images. *Intl. Journal of Computer Vision* 62, 1 (Apr. 2005), 61–81.
- [297] VESE, L., AND CHAN, T. Active contours without edges. *IEEE Transactions on Image Processing* 10, 2 (Feb. 2001), 266–277.
- [298] VESE, L. A., AND OSHER, S. J. Modeling textures with total variation minimization and oscillating patterns in image processing. *SIAM Journal on Scientific Computing* 19, 1-3 (Dec. 2003), 553–572.
- [299] 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, 1-2 (Jan. 2004), 7–18.
- [300] VESE, L. A., AND OSHER, S. J. Image decomposition: Separation of texture from piece-wise smooth content. In *Eighth Intl. Symp. on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)* (Timisoara, Sept. 2006), pp. 103–110.
- [301] VINCENT, L. Granulometries and opening trees. *Fundamenta Informaticae* 41, 1-2 (Jan. 2000), 57–90.
- [302] 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.
- [303] VOORHEES, H., AND POGGIO, T. Detecting textons and texture boundaries in natural images. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (London, UK, 1987), pp. 250–258.
- [304] VOORHEES, H., AND POGGIO, T. Computing texture boundaries from images. *Nature* 333 (May 1988), 364–367.
- [305] WANDELL, B. A. *Foundations of Vision*. Sinauer Press, 1995.
- [306] WEBSTER, M. A., AND DE VALOIS, R. L. Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *Journal of the Optical Society of America A* 2, 7 (July 1985), 1124–1131.
- [307] WEI, D., AND BOVIK, A. On the instantaneous frequencies of multicomponent AM-FM signals. *IEEE Signal Processing Letters* 5, 4 (Apr. 1998), 84–86.
- [308] WEICKERT, J. Multiscale texture enhancement. In *LNCS, Computer Analysis of Images and Patterns*, vol. 970. Springer, 1995, pp. 230–237.
- [309] 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.
- [310] WINKLER, G., AND LUTZ, G. F. H. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction.*, vol. 27 of *Applications of Mathematics*. Springer-Verlag, 2003.
- [311] WITKIN, A., AND KASS, M. Reaction-diffusion textures. In *SIGGRAPH 91* (1991), ACM Press, pp. 299–308.
- [312] WITKIN, A. P. Recovering surface shape and orientation from texture. *Artificial Intelligence* 17, 1-3 (Aug. 1981), 17–45.

- [313] 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.
- [314] 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.
- [315] YEZZI, A., TSAI, A., AND WILLSKY, A. A statistical approach to snakes for bimodal and trimodal imagery. In *IEEE Intl. Conf. on Computer Vision (ICCV)* (Corfu, Greece, Sept. 1999).
- [316] YING, G., MITCHELL, C., AND JAMIESON, L. Endpoint detection of isolated utterances based on a modified Teager energy measurement. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (Minneapolis, MN, Apr. 1993), pp. 732–735.
- [317] YU, T.-H., MITRA, S. K., AND KAISER, J. F. Novel Algorithm for Image Enhancement. In *Proc. SPIE/SPSE Conf. on Image Processing Algorithms and Techniques II* (San Jose, CA, Feb. 1991).
- [318] 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.
- [319] 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.
- [320] ZHU, S.-C. Statistical modeling and conceptualization of visual patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 6 (2003), 691–712.
- [321] 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.
- [322] ZHU, S. C., GUO, C. E., WU, Y., AND WANG, Y. What are textons. In *European Conf. on Computer Vision (ECCV)* (Copenhagen, Denmark, May 2002).
- [323] ZHU, S. C., AND MUMFORD, D. Prior learning and Gibbs reaction-diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 11 (Nov. 1997), 1236–1250.
- [324] ZHU, S. C., WU, Y. N., AND MUMFORD, D. Minimax entropy principle and its application to texture modeling. *Neural Computation* 9, 8 (1997), 1627–1660.
- [325] 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

## ΠΕΡΙΟΔΙΚΑ:

1. Sofou, A., Evangelopoulos, G., and Maragos, P. "Soil image segmentation and texture analysis: a computer vision approach". *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 4, pp. 394-398, Oct. 2005.
2. Evangelopoulos, G., and Maragos, P. "Multiband modulation energy tracking for noisy speech detection". *IEEE Trans. on Speech and Audio Processing*, vol. 14, no. 6, pp. 2024-2038, Nov. 2006.
3. Kotti, M., Ververidis, D., Evangelopoulos, G., Panagakis, I., Kotropoulos, C., Maragos, P., Pitas, I. "Audio-Assisted Movie Dialogue Detection", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1618 - 1627, Nov. 2008.
4. Kokkinos, I., Evangelopoulos, G., and Maragos, P. "Texture Analysis and Segmentation Using Modulation, Generative and Variational Models", *IEEE Trans. Pattern Analysis & Machine Intelligence*, in press, Jan. 2009.
5. Maragos, P. and Evangelopoulos, G. "Leveling Cartoons, Texture Markers and Multiscale Image Decomposition", *Image and Vision Computing*, (submitted).
6. Evangelopoulos, G., and Maragos, P. "Synergy between Image Decomposition and Texture Modeling: Oscillations and Modulations", *IEEE Trans. on Image Processing*, (under submission).

## ΚΕΦΑΛΑΙΑ ΒΙΒΛΙΩΝ:

1. Evangelopoulos, G., Rapantzikos, K., Maragos, P. and Avrithis, Y., "Audio-visual attention modeling and salient event detection", on *Multimodal Processing and Interaction: Audio, Video, Text*, Eds. Maragos P., Potamianos A. and Gross P., Springer, 2008.

## ΣΥΝΕΔΡΙΑ:

1. Evangelopoulos, G., Kokkinos, I., and Maragos, P. "Advances in variational image segmentation using AM-FM models: Regularized demodulation and probabilistic cue integration". In Proc. *IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision* (VLSM-05), Beijing, China, 2005, LNCS, vol. 3275, pp. 121-136.
2. Sofou, A., Evangelopoulos, G., and Maragos, P. "Coupled geometric and texture PDE-based segmentation", In Proc. *IEEE Intl. Conf. on Image Processing* (ICIP-05), Genova, Italy, 2005, pp. II. 600-603.

3. Evangelopoulos, G., and Maragos, P., “Speech event detection using multiband modulation energy”. *In Proc. ISCA Interspeech* (Eurospeech-05), Lisbon, Portugal, 2005, pp. 685-688.
4. 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* (ICIP-04), Singapore, 2004, vol. 3, pp. 1509-1512.
5. Kokkinos, I., Evangelopoulos, G., and Maragos, P. “Modulation-feature based textured image segmentation using curve evolution”, *In Proc. IEEE Intl. Conf. on Image Processing* (ICIP-04), Singapore, 2004, vol. 2, pp. 1204-1207.
7. Maragos, P and Evangelopoulos, G. “Leveling Cartoons, Texture Energy Markers, and Image Decomposition”, *8th Intl. Symposium on Mathematical Morphology* (ISMM-07), Rio de Janeiro, Brazil, Oct. 10-13, 2007, pp. 125-138.
8. Rapantzikos, K., Evangelopoulos, G., Maragos, P. and Avrithis, Y. “An audiovisual saliency model for movie summarization”, *Intl. Workshop on Multimedia Signal Processing* (MMSP-07), Chania, Greece, 2007.
9. Evangelopoulos, G., and Maragos, P. “Image decomposition into structure and texture subcomponents with multifrequency modulation constraints”, *In Proc. IEEE Conf. on Computer Vision & Pattern Recognition* (CVPR-08), Alaska, 2008.
10. Evangelopoulos, G. and Maragos, P. “Texture Modulation-constrained Image Decomposition”, *In Proc. IEEE Intl. Conf. on Image Processing* (ICIP-08), San Diego, CA, U.S.A., Oct. 12-15, 2008, pp. 793-796.
11. Evangelopoulos, G., Rapantzikos, K., Potamianos, A., Maragos, P., Zlatintsi, A., Avrithis, Y. “Movie Summarization based on Audiovisual Saliency Detection”, *In Proc. IEEE Intl. Conf. on Image Processing* (ICIP-08), San Diego, CA, U.S.A., Oct. 12-15, 2008, pp. 2528-2531.

□

# Βιογραφικό Σημείωμα

Ο Γιώργος Ευαγγελόπουλος γεννήθηκε στα Γρεβενά το 1978 όπου και ολοκλήρωσε στο 2ο Γενικό Λύκειο τις εγκύλιες σπουδές. Εισήχθηκε στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Ευνικού Μετσόβιου Πολυτεχνείου το 1996. Το Σεπτέμβριο του 2001 έλαβε το Δίπλωμα Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών με κατεύθυνση Μηχανικού Επικοινωνιών και ειδίκευση στους κύκλους μαθημάτων Τηλεπικοινωνιών, Σημάτων-Ελέγχου-Ρομποτικής και Δικτύων Επικοινωνιών. Η διπλωματική του εργασία ήταν στο πεδίο της Ανάλυσης και Επεξεργασίας Σημάτων και Φωνής με έμφαση στη διάκριση φωνής, σιωπής και υφορύβου. Κατά τη διάρκεια των ετών 2002-2007 συνέχισε τις σπουδές του ως Υποψήφιος Διδάκτωρ της Σχολής ΗΜΜΥ στον τομέα Σημάτων, Ελέγχου και Ρομποτικής δουλεύοντας στην ερευνητική ομάδα CVSP στα γενικά πεδία της Ανάλυσης Εικόνων και Όρασης Υπολογιστών. Η διατριβή του με τίτλο «Ανάλυση Υφής και Μικροδομών Εικόνων με Μοντέλα Διαμόρφωσης, Ενεργειακές και Μεταβολικές Τεχνικές: Ανίχνευση & Διαχωρισμός», επικεντρώθηκε στη μοντελοποίηση υφής εικόνων για ανάλυση εικόνων και διαδικασίες όρασης χαμηλού επιπέδου. Η Διδακτορική του Διατριβή εντάχθηκε στο ερευνητικό πρόγραμμα της ΓΓΕΤ, ΠΕΝΕΔ-2001 στο πεδίο «Ανάλυση Εικόνων και Έλεγχος της Ποιότητας Εδαφών» ενώ παράλληλα συμμετείχε μερικά και σε άλλα προγράμματα βασικής έρευνας, ως νέος ερευνητής στο πρόγραμμα «Αρχιψήδης» για την ανάπτυξη τεχνικών επεξεργασίας βιοιατρικών σημάτων και ως ερευνητής στο ΕΠΙΣΕΥ στα ερευνητικά προγράμματα της Ευρωπαϊκής Ένωσης ‘HIWIRE’ και ‘MUSCLE NoE’, σε θέματα ανίχνευσης φωνητικών γεγονότων και αυτόματης δημιουργίας περιλήψεων από βίντεο δεδομένα αντιστοίχως. Τα ερευνητικά του ενδιαφέροντα βρίσκονται στις περιοχές της μη-γραμμικής επεξεργασίας σημάτων, ανάλυσης οπτικής υφής, ανάλυσης ήχου και εικόνας και γενικευμένη ανίχνευση γεγονότων σε πολυμεσικά συστήματα με εφαρμογές στην Επεξεργασία Εικόνων, την Όραση Υπολογιστών, την Επεξεργασία Πολυμέσων και τα Γραφικά με Υπολογιστές. Από το 2001 είναι μέλος του Τεχνικού Επιμελητηρίου Ελλάδας (ΤΕΕ) και του διεθνούς τεχνικού οργανισμού IEEE. Έχει δημοσιεύσει 16 άρθρα σε διεθνή περιοδικά και διεθνή συνέδρια με κριτές, ενώ είναι τακτικός χριτής για περιοδικά των οργανισμών IEEE και OSA. (<http://cvsp.cs.ntua.gr/gevag>).

□