



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Ανάλυση Εικόνας και Όραση Υπολογιστών:
Θεωρία και Εφαρμογές στην
Αποκατάσταση Αρχαίων Τοιχογραφιών**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΤΟΥ

ΓΕΩΡΓΙΟΥ Ν. ΠΑΠΑΝΔΡΕΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π.

Αθήνα, Ιούνιος 2009



ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Ανάλυση Εικόνας και Όραση Υπολογιστών: Θεωρία και Εφαρμογές στην Αποκατάσταση Αρχαίων Τοιχογραφιών

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

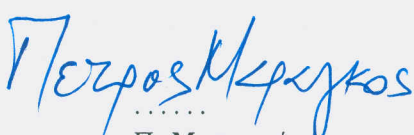
ΤΟΥ

ΓΕΩΡΓΙΟΥ Ν. ΠΑΠΑΝΔΡΕΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π.

Συμβουλευτική Επιτροπή: Καθ. Πέτρος Μαραγκός (Επιβλέπων)
Καθ. Στέφανος Κόλλιας
Επ. Καθ. Κωνσταντίνος Τζαφέστας

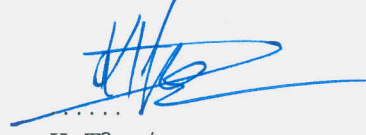
Εγκρίθηκε από την επταμελή εξεταστική επιτροπή στις 22 Ιουλίου 2009:


.....

Π. Μαραγκός
Καθηγητής Ε.Μ.Π.


.....

Σ. Κόλλιας
Καθηγητής Ε.Μ.Π.


.....

Κ. Τζαφέστας
Επ. Καθηγητής Ε.Μ.Π.


.....

Γ. Παπαβασιλόπουλος
Καθηγητής Ε.Μ.Π.


.....

Τ. Σελλής
Καθηγητής Ε.Μ.Π.


.....

Α. Βλαχόπουλος
Επ. Καθ. Παν. Ιωαννίνων


.....

Γ. Ποταμιάνος
Ερευν. Α', Δημόκριτος

Αθήνα, Ιούνιος 2009



Ελλάδα
ανταγωνιστική
ΥΠΟΥΡΓΕΙΟ ΑΝΑΠΤΥΞΗΣ

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

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

...

ΓΕΩΡΓΙΟΣ Ν. ΠΑΠΑΝΔΡΕΟΥ

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

Copyright © Γεώργιος Παπανδρέου, 2009.

Με επιφύλαξη παντός δικαιώματος. All rights reserved

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

Περιεχόμενα

1	Εισαγωγή	1
2	Προϊστορικές Τοιχογραφίες στο Ακρωτήριο Θήρας: Φωτογράφιση και Ψηφιακή Αποκατάσταση	9
2.1	Τοιχογραφίες με Ελλείποντα Τμήματα και Ενδοσυμπλήρωση Πληροφορίας .	10
2.2	Φωτογράφιση Τοιχογραφιών	14
2.2.1	Αντικείμενο Φωτογράφισης	14
2.2.2	Στόχοι και Διαδικασία Φωτογράφισης	16
2.3	Σύνθεση Τοιχογραφικών Μωσαϊκών Υπερ-Υψηλής Ανάλυσης	19
3	Πολυπλεγματικοί Αλγόριθμοι για Κατάτμηση και Ενδοσυμπλήρωση με Μοντέλα Μερικών Διαφορικών Εξισώσεων	25
3.1	Κατάτμηση με Μερικές Διαφορικές Εξισώσεις: Γεωμετρικά Μοντέλα Ενεργών Περιγραμμάτων	26
3.1.1	Ενοποιημένη Θεώρηση Γεωμετρικών Μοντέλων Ενεργών Περιγραμμάτων	28
3.2	Ενδοσυμπλήρωση με Μερικές Διαφορικές Εξισώσεις	30
3.2.1	Μοντέλα Ανισοτροπικής Διάχυσης	31
3.3	Αριθμητικά Σχήματα και Πολυπλεγματικοί Αλγόριθμοι	34
3.3.1	Διακριτοποίηση Χωρικών Παραγώγων	34
3.3.2	Σχήματα Χρονικής Ολοκλήρωσης	37
3.3.3	Πολυπλεγματική Επίλυση Ανισοτροπικών Προβλημάτων	39
3.4	Πειράματα Κατάτμησης με Γεωμετρικά Ενεργά Περιγράμματα και Πολυπλεγματικούς Αλγορίθμους	41
3.5	Πειράματα Ενδοσυμπλήρωσης με Μερικές Διαφορικές Εξισώσεις Ανισοτροπικής Διάχυσης και Πολυπλεγματικούς Αλγορίθμους	43
4	Αυτόματη Επιλογή Κλίμακας Μέσω Στατιστικής Διεπικύρωσης	49
4.1	Αποθορυβοποίηση με Διαδικασίες Μη-Γραμμικής Διάχυσης	50
4.2	Αποθορυβοποίηση με Μορφολογικά Φίλτρα	50
4.3	Αυτόματη Επιλογή Κλίμακας με Στατιστική Διεπικύρωση	51
4.3.1	Ορισμός Προβλήματος	52
4.3.2	Προτεινόμενος Αλγόριθμος Επιλογής Κλίμακας	53
4.3.3	Πειράματα και Συγκρίσεις	55
4.3.4	Συμπεράσματα	57

5 Αναπαράσταση και Στατιστικές Ιδιότητες Φυσικών Εικόνων στο Πεδίο των Κυματιδίων	61
5.1 Κλασική Μοντελοποίηση Φυσικών Εικόνων ως Στάσιμων Στοχαστικών Διαδικασιών και Αναπαράσταση στο Πεδίο Μετασχηματισμού Fourier	61
5.2 Χωρικά Εντοπισμένοι Μετασχηματισμοί Εικόνων	63
5.2.1 Κυματιδιακή Αναπαράσταση Χώρου-Κλίμακας για Σήματα και Εικόνες	64
5.2.2 Αναπαράσταση Σημάτων σε Πλαίσια και ο Δίδεντρος Μιγαδικός Κυματιδιακός Μετασχηματισμός	67
5.2.3 Τοπικά Αρμονική Αναπαράσταση Χώρου-Συχνότητας και ο Διαμορφωμένος Μιγαδικός Επικαλυπτόμενος Μετασχηματισμός	71
5.3 Στατιστικές Ιδιότητες Φυσικών Εικόνων στο Πεδίο των Κυματιδίων	80
5.3.1 Ιδιότητες Μέτρου Συντελεστών Κυματιδίων: Κυρτωτικά Περιθώρια Ιστογράμματα και Αλληλεξαρτήσεις Συντελεστών	81
5.3.2 Ιδιότητες Φάσης Συντελεστών Κυματιδίων	82
6 Ενδοσυμπλήρωση με Κυματίδια Οργανωμένα σε Δενδρικό Στατιστικό Μοντέλο	87
6.1 Η Ενδοσυμπλήρωση ως Πρόβλημα Στατιστικής Συμπερασματολογίας	87
6.2 Μοντελοποίηση Συντελεστών Μιγαδικών Κυματιδίων Εικόνων με Κρυφό Δένδρο Markov	88
6.2.1 Το HMT Μοντέλο για Περιγραφή 1-Δ και 2-Δ Πραγματικών Κυματιδίων	89
6.2.2 Προσαρμογή του HMT Μοντέλου σε 2-Δ Μιγαδικά Κυματίδια: Δενδρικές Εξαρτήσεις στις Αναπαραστάσεις DT-CWT και MCLT	90
6.2.3 Περιγραφή Εικόνων με το HMT Μοντέλο: Παραμετροποίηση και Βασικοί Αλγόριθμοι για Συμπερασματολογία και Μάθηση	91
6.3 Αποθορυβοποίηση με Κρυφό Δένδρο Markov στο Χώρο των Μιγαδικών Κυματιδίων	95
6.4 Σύνθεση Υφής με Κρυφό Δένδρο Markov στο Χώρο των Μιγαδικών Κυματιδίων	100
6.5 Ενδοσυμπλήρωση με Κρυφό Δένδρο Markov στο Χώρο των Μιγαδικών Κυματιδίων	101
6.5.1 Ντετερμινιστική Επίλυση με Εναλλασσόμενη Βελτιστοποίηση	105
6.5.2 Στοχαστική Επίλυση με Δειγματοληψία Gibbs	105
6.5.3 Αρχικοποίηση με μεθόδους Μερικών Διαφορικών Εξισώσεων	108
6.6 Επεκτάσεις του Βασικού Μοντέλου	108
6.6.1 Αποκατάσταση Συνιστωσών Καρτούν και Υφής με Χρήση Πολλαπλών Λεξικών	108
6.6.2 Σύζευξη Μοντέλων Κυματιδίων και Τεχνικών ΜΔΕ στο Χωρικό Πεδίο .	109
6.6.3 Τεχνικές για Ενσωμάτωση Πληροφορίας Μιγαδικής Φάσης Κατά την Αποκατάσταση	109
6.7 Πειράματα Ενδοσυμπλήρωσης	110
6.8 Σύνοψη και Προοπτικές Βελτίωσης του Δενδρικού Μοντέλου Κυματιδίων . .	111
6.Α' Επαναληπτική Επίλυση Συστημάτων Γραμμικών Εξισώσεων: Μέθοδοι Καθόδου Μέγιστης Κλίσης και Συζυγών Κλίσεων	112
6.Β' Αριθμητική Επίλυση Προβλημάτων Κυρτού Τετραγωνικού Προγραμματισμού με Περιορισμούς Ανισότητας	114

7 Σύνοψη Διατριβής και Μελλοντική Έρευνα	121
7.1 Συνεισφορές Διατριβής	121
7.2 Κατευθύνσεις Μελλοντικής Έρευνας	122
Α΄ Προσαρμοστική Σύμμιξη Πληροφορίας για Συνδυασμένη Οπτική-Ακουστική Αναγνώριση Ομιλίας	125
A.1 Εισαγωγή	126
A.2 Αβεβαιότητα Χαρακτηριστικών και Πολυτροπική Σύμμιξη	127
A.2.1 Αβεβαιότητα Παρατήρησης Χαρακτηριστικών και η Αντιστάθμισή της Κατά την Κατηγοριοποίηση	127
A.2.2 Αβεβαιότητα Παρατήρησης και Πολυτροπική Σύμμιξη	128
A.3 Εκπαίδευση με Θορυβώδη Δεδομένα Χρησιμοποιώντας τον Αλγόριθμο EM	130
A.3.1 Θορυβώδης EM Εκπαίδευση GMM	132
A.3.2 Θορυβώδης EM Εκπαίδευση HMM	133
A.4 Συνδυασμένη Οπτική-Ακουστική Αναγνώριση Ομιλίας	133
A.4.1 Εξαγωγή και Εκτίμηση Αβεβαιότητας Οπτικών Χαρακτηριστικών Ομιλίας	133
A.4.2 Εξαγωγή Ακουστικών Χαρακτηριστικών και Εκτίμηση της Αβεβαιότητάς τους	136
A.4.3 Σύγχρονες και Ασύγχρονες Αρχιτεκτονικές Σύμμιξης Ακολουθιακών Δεδομένων	136
A.5 Πειραματική Αξιολόγηση	137
A.5.1 Βάση Δεδομένων και Μεθοδολογία Αξιολόγησης	137
A.5.2 Μονο-Τροπικά Πειράματα Αναγνώρισης Φωνής	139
A.5.3 Πειράματα Συνδυασμένης Οπτικής-Ακουστικής Αναγνώρισης	141
A.6 Συμπεράσματα	144
Β΄ Δημοσιεύσεις Διατριβής	147
Βιβλιογραφία	150

Κατάλογος Σχημάτων

1.1	Αμφισημία στην ερμηνεία οπτικών ερεθισμάτων: Κύβος του Necker	2
1.2	Απατηλά περιγράμματα στην ανθρώπινη όραση: Το τρίγωνο του Kanizsa.	4
1.3	Κλασικά κριτήρια Gestalt ομαδοποίησης.	4
1.4	Επανάγγιγμα φωτογραφίας: Έπαρση της Σοβιετικής σημαίας στο Reichstag του Βερολίνου.	5
2.1	Το πρόβλημα της ενδοσυμπλήρωσης. Ω : κενό, Ω^c : γνωστή περιοχή, $\partial\Omega$: σύνορο περιοχών.	11
2.2	Παράδειγμα ανθρώπινης επισημείωσης των ελλειπουσών περιοχών σε τοιχογραφίες από το Ακρωτήρι Θήρας. (α) Αρχική εικόνα της τοιχογραφίας. (β) Επισημείωση όπου ο χρήστης έχει υπερθέσει στην αρχική εικόνα τη μάσκα που αντιστοιχεί στις ελλείπουσες περιοχές.	12
2.3	Παράδειγματα αλλοίωσης σε τμήματα της τοιχογραφίας ‘Ακολουθία Ανδρών’ από το Ακρωτήρι Θήρας. Λεπτομέρεια από το κεφάλι (α) και τον κορμό (β) του νεαρού άνδρα. Εκτός από τα ελλείποντα τμήματα παρατηρούμε αλλοιώσεις όπως ξεθώριασμα του χρώματος και μερική απόξεση της ζωγραφικής επιφάνειας.	13
2.4	Τμήμα τοπογραφικού διαγράμματος του αρχαιολογικού χώρου στο Ακρωτήρι Θήρας	15
2.5	Γενική άποψη των φωτογραφημένων τοιχογραφιών, από την Ξεστή 3 και από το Κτίριο Β. Οι τοιχογραφίες βρίσκονται στο εργαστήριο τοιχογραφιών στο Ακρωτήρι Θήρας και στο Εθνικό Αρχαιολογικό Μουσείο.	17
2.6	Παράδειγμα πολυκλιμακωτής φωτογράφισης της τοιχογραφίας ‘Κροκοσυλλέκτρια & Πότνια’ από το Ακρωτήρι Θήρας	20
2.7	Παράδειγματα πανοραμικής φωτογράφισης: Κυλινδρικό και γραμμικό πανόραμα.	21
2.8	Κύρια βήματα κατά τη σύνθεση φωτογραφικών μωσαϊκών υπερ-υψηλής ανάλυσης.	22
2.9	Παράδειγμα πλοήγησης εντός της τοιχογραφίας ‘Κροκοσυλλέκτρια & Πότνια’.	24
3.1	Δομικός τανυστής και τανυστής διάχυσης κοντά σε ακμή	32
3.2	Το ίχνος 5-σημείων για υπολογισμό του χωρικού όρου $\text{div}(g\nabla u)$	35
3.3	Το ίχνος 9-σημείων για υπολογισμό του χωρικού όρου $\text{div}(D\nabla u)$	36
3.4	Το 3×3 ίχνος για υπολογισμό του χωρικού όρου $\text{trace}(TH)$	36
3.5	Συνηθισμένες πολυπλεγματικές τακτικές. Το βήμα πλέγματος δίνεται στην αριστερή στήλη. Διάγραμμα για την περίπτωση χρήσης 4 επιπέδων.	41
3.6	Ανισοτροπικά ψεύδεργα του διαχωρίσιμου AOS σχήματος.	44

3.7	Πλήρως 2-Δ IMEX σε σύγκριση με τη διαχωρίσιμη AOS διακριτοποίηση. Το AOS σχήμα δεν κανονικοποιεί επαρκώς την καμπύλη και ευνοεί το σχηματισμό γραμμών παράλληλα προς τους άξονες.	45
3.8	Παράδειγμα κατάτμησης εικόνας με τη μέθοδο Chan-Vese.	46
3.9	Παράδειγμα κατάτμησης υφής με τη μέθοδο των Rousson-Brox-Deriche. . .	47
3.10	Ενδοσυμπλήρωση με ΜΔΕ Laplace και ανισοτροπικής διάχυσης σε τμήμα της τοιχογραφίας 'Κροκοσυλλέκτρια & Πότνια'.	48
4.1	Εναλλακτικά σχέδια επαναδειγματοληψίας κατά τη διεπικύρωση: Quadruple-CV και Double-CV	54
4.2	Παράδειγμα αποθορυβοποίησης με αυτόματη επιλογή κλίμακας. Θορυβώδης εικόνα, εκτίμηση με αυτόματα επιλεγμένα κλίμακα, σφάλμα μοντέλου και εκτίμηση του σφάλματος πρόβλεψης με τους αλγορίθμους διεπικύρωσης	56
4.3	Παραδείγματα αποθορυβοποίησης με μη γραμμική διάχυση και μορφολογικό φίλτρο επιπέδωσης. Κλίμακα αυτόματα επιλεγμένα μέσω διεπικύρωσης.	57
4.4	Παράδειγμα αποθορυβοποίησης σε εικόνες τοιχογραφιών από το Ακρωτήρι.	58
4.5	Σχετική αύξηση του σφάλματος μοντελοποίησης $\Delta(ME)/ME$ ως συνάρτηση του επιπέδου του θορύβου nl για τους εξεταζόμενους αλγορίθμους επιλογής κλίμακας.	59
5.1	Δείγματα κλασματικής κίνησης Brown για διάφορες τιμές της φράκταλ διάστασης $D = 3.5 - \beta/2$. Η σύνθεση έγινε με μεθόδους στο πεδίο Fourier . .	63
5.2	Ιδιότητες χρονικού/φασματικού εντοπισμού της κυματιδιακής ανάλυσης με τη βοήθεια διαγραμμάτων ορθογωνίων Heisenberg, επίσης γνωστών και ως διαγραμμάτων κουτιών χρόνου-συχνότητας.	64
5.3	Συστοιχίες φίλτρων που αντιστοιχούν σε ένα βήμα κυματιδιακής ανάλυσης/σύνθεσης με τον αναδρομικό DWT αλγόριθμο του Mallat.	66
5.4	Τα τρία κυματίδια που αντιστοιχούν στον διαχωρίσιμο 2-Δ (DWT) και οι συντελεστών κυματιδίων που αντιστοιχούν στο 2-Δ DWT της εικόνας Lena. .	68
5.5	Η συστοιχία φίλτρου για τον υπολογισμό του 1-Δ δίδεντρου μιγαδικού κυματιδιακού μετασχηματισμού (DT-CWT).	69
5.6	Τα έξι μιγαδικά κατευθυντικά κυματίδια που αντιστοιχούν στον 2-Δ DT-CWT και παράδειγμα ανάλυσης της 512×512 εικόνας Lena.	72
5.7	Αλληλεπικαλυπτόμενα παράθυρα ανάλυσης μεταξύ διαδοχικών μπλοκ. . . .	75
5.8	Τα MCLT άτομα που αντιστοιχούν σε ανάλυση με μήκος μπλοκ $l = 16$	78
5.9	Μιγαδικοί συντελεστές που αντιστοιχούν στο 2-Δ MCLT της εικόνας Barbara, για ανάλυση σε μπλοκ των 16×16 εικονοστοιχείων.	79
5.10	Παράδειγμα εικόνας από τη βάση του van Hateren [vHvdS98] και ιστογράμμο κυματιδιακών αποκρίσεων μαζί με ταιριασμένες γκαουσιανή και γενικευμένη γκαουσιανή κατανομές.	81
5.11	Ιστογράμματα συντελεστών κυματιδίων για διάφορα ζευγάρια φίλτρων εφαρμοσμένα σε μία εικόνα.	83
5.12	Η σημασία της Fourier φάσης στην αντίληψη εικόνων.	84
5.13	Η σημασία της τοπικής κυματιδιακής φάσης στην αντίληψη εικόνων.	85
6.1	Η δομή του κρυφού δένδρου Markov στη μοντελοποίηση συντελεστών κυματιδίων εικόνων.	89

6.2	Η δομή του δενδρικού γραφικού μοντέλου στο προτεινόμενο στατιστικό μοντέλο εικόνας για τις δύο μιγαδικές κυματιδιακές αναπαραστάσεις DT-CWT και MCLT.	92
6.3	Οι εικόνες που χρησιμοποιήθηκαν στα πειράματα αποθορυβοποίησης με δενδρικά κυματιδιακά μοντέλα.	96
6.4	Ποιοτική αξιολόγηση αποτελέσματος αποθορυβοποίησης με δενδρικά στατιστικά μοντέλα κυματιδίων.	99
6.5	Πειράματα ανασύνθεσης υφής από εικόνες της συλλογής Brodatz με μιγαδικά κυματίδια και τα HMT/HIT μοντέλο.	102
6.6	Πειράματα ανασύνθεσης υφής από εικόνες τοιχογραφιών του Ακρωτηρίου.	103
6.7	Αποτέλεσμα ενδοσυμπλήρωσης σε τμήμα της τοιχογραφίας 'Γυναίκες' από το Ακρωτήρι.	117
6.8	Αποτέλεσμα ενδοσυμπλήρωσης σε τμήμα της θηραϊκής τοιχογραφίας 'Ακολουθία Ανδρών'.	118
6.9	Συγκριτικά αποτελέσματα ενδοσυμπλήρωσης σε τεχνητά επικαλυμμένη εικόνα.	119
A'.1	Γραφική αναπαράσταση εναλλακτικών σεναρίων μέτρησης χαρακτηριστικών.	128
A'.2	Σύνορα απόφασης κατά την κατηγοριοποίηση μιας θορυβώδους παρατήρησης για διάφορες τιμές μεταβλητότητας θορύβου.	129
A'.3	Η πολυτροπική αντιστάθμιση αβεβαιότητας οδηγεί σε προσαρμοστική σύμμιξη.	131
A'.4	Υποσύστημα εξαγωγής οπτικών χαρακτηριστικών.	135
A'.5	Ενδεικτικά καρτέ από όλα τα 36 υποκείμενα της βάσης δεδομένων CUAVE.	137
A'.6	Αποτελέσματα κατηγοριοποίησης ψηφίων με μόνο ηχητικά χαρακτηριστικά για διάφορα επίπεδα babble θορύβου.	140
A'.7	Αποτελέσματα κατηγοριοποίησης ψηφίων με μόνο AAM και PCA οπτικά χαρακτηριστικά για μεταβαλλόμενο αριθμό συντελεστών υφής.	141
A'.8	Οπτικό-ακουστική Multistream-HMM κατηγοριοποίηση ψηφίων για διάφορα επίπεδα babble θορύβου.	143
A'.9	Κέρδος στην απόδοση λόγω αντιστάθμισης της αβεβαιότητας για την οπτικό-ακουστική κατηγοριοποίησης ψηφίων για διάφορα επίπεδα babble θορύβου και τις 5 επαναλήψεις του πειράματος για διαφορετικά σύνολα ελέγχου.	144
A'.10	Αποτελέσματα οπτικό-ακουστικής κατηγοριοποίησης ψηφίων με Product-HMM για διάφορα επίπεδα babble θορύβου.	145

Κατάλογος Πινάκων

2.1	Κατάλογος φωτογραφημένων τοιχογραφιών από το Ακρωτήρι Θήρας και το Εθνικό Αρχαιολογικό Μουσείο της Αθήνας.	18
3.1	Υπολογιστικό κόστος κύριων αλγοριθμικών συνιστωσών κατά την κατάτμηση εικόνας διαστάσεων 256×256 εικονοστοιχείων με πολυπλεγματικούς αλγόριθμους.	42
6.1	Πείραμα αποθορυβοποίησης για λευκό προσθετικό θόρυβο Gauss μεταβλητότητας σ_n^2 . Εξετάζουμε την επίδοση, μετρημένη με όρους PSNR (dB), του δενδρικού μοντέλου HMT συνδυασμένου με τον DT-CWT ή τον MCLT σε σχέση με αντίστοιχες εκδοχές του ανεξάρτητου μοντέλου HIT δίχως δενδρικές εξαρτήσεις. Επίσης περιλαμβάνουμε αποτελέσματα εναλλακτικών τεχνικών (Wiener, BLS-GSM, BM3D). Για κάθε μέθοδο δείχνουμε επίσης μέσο κέρδος αποθορυβοποίησης για κάθε εικόνα ως προς όλα τα επίπεδα θορύβου και για κάθε επίπεδο θορύβου ως προς όλες τις εικόνες.	98

ΠΡΟΛΟΓΟΣ

Η παρούσα διατριβή αποτελεί τον τελευταίο σταθμό ενός μακρινού μα συνάμα συναρπαστικού ταξιδιού. Σηματοδοτεί το τέλος μιας δύσκολης εξερευνητικής αποστολής όπου ο προορισμός διαμορφώθηκε σταδιακά κατά την πορεία. Πέρα από το τελικό αποτέλεσμα, άξιζε η διαδρομή καθαυτή, η οποία μου χάρισε πλήθος σπουδαίων εμπειριών και πολύτιμα εφόδια για να συνεχίσω το ταξίδι παραπέρα ύστερα από αυτή τη στάση.

Σε αυτήν την πορεία δεν ήμουν μόνος. Έδρασα σε ένα περιβάλλον που συνδυάζει με μοναδικό τρόπο υψηλό επίπεδο ακαδημαϊκής αριστείας και ανθρώπινων σχέσεων.

Οφείλω πολλά στον επιβλέποντα καθηγητή μου και επικεφαλής του εργαστηρίου κ. Πέτρο Μαραγκό. Ήταν διαρκώς δίπλα μου στην ερευνητική μου προσπάθεια, ενώ ταυτόχρονα μου παρείχε πλήρη ελευθερία να ακολουθήσω τα δικά μου ερευνητικά μονοπάτια. Τα προπτυχιακά και μεταπτυχιακά μαθήματα που διδάσκει με αξιοζήλευτη ενεργητικότητα αποτέλεσαν σημαντική πηγή έμπνευσης και ιδεών. Η εμπειρία του και οι συμβουλές του σε ακαδημαϊκά και επαγγελματικά ζητήματα αποδείχθηκαν ιδιαίτερα χρήσιμες. Το ήθος, η εργατικότητα, και ο ερευνητικός του ζήλος αποτελούν παράδειγμα προς μίμηση.

Η εμπειρία του διδακτορικού θα ήταν απείρως φτωχότερη δίχως την παρουσία των υπόλοιπων παιδιών του εργαστηρίου, ενός συνδυασμού ανθρώπων με σπάνιο ήθος, συντρόφων σε πορείες που συχνά τέμνονταν. Οι σχέσεις μας έχουν σφυρηλατηθεί μέσα από ατέλειωτες ώρες συνύπαρξης στο εργαστήριο, ερευνητική συνεργασία, αλληλοϋποστήριξη, συζητήσεις πάνω σε επιστημονικά ή κοινωνικά θέματα αμοιβαίου προβληματισμού, αμέτρητα κοινά ξενύχτια τις παραμονές καταληκτικών ημερομηνιών συνεδρίων (με 'επαθλο' αξέχαστα ταξίδια σε μακρινά μέρη), καλοκαιρινές διακοπές στα νησιά και μοναδικές βραδιές εξόδου στην Αθήνα. Ο Νάσος είναι συνοδοιπόρος από τα προπτυχιακά χρόνια, με παρακίνησε ως πιο αποφασιστικός να μείνω στο εργαστήριο για διδακτορικό, κοντινός συζητητής, συγκάτοικος (αυτή τη φορά κυριολεκτικά) και μετά το διδακτορικό. Ο Βασίλης ευαίσθητος και πάντα παρών, πρόθυμος να βοηθήσει μα ταυτόχρονα διακριτικός. Από τους δυο τους έμαθα πάρα πολλά για την ανθρώπινη ομιλία στη συνεργασία μας για την οπτική-ακουστική περιγραφή της. Ο Σταμάτης φιλότιμος και επίμονος, σχολαστικός· συνεργαστήκαμε άψογα σε θέματα κυματιδίων και στατιστικής, και είμαστε στενοί φίλοι έκτοτε. Ο Ιάσοντας χάραξε πρώτος το δρόμο, αφοσιωμένος στην επιστήμη και στους φίλους του, ήταν και είναι πολύτιμος σύμβουλος μου έχοντας βαθιά γνώση και πλούσια διαίσθηση στην όραση υπολογιστών. Ο Γιώργος εμψυχωτής και ιδεαλιστής, με αντισυμβατική προσέγγιση σε θέματα επιστήμης και ζωής, θέτει το 'γιατί;'. Ο Τάσος προσεγγίζει τα ζητήματα από πιο μαθηματική σκοπιά· οι συζητήσεις μας σε μοντέλα ανισοτροπικής διάχυσης μου φάνηκαν ιδιαίτερα χρήσιμες. Η Νάνσυ έχει αναδειχθεί σε σταθερή αξία και οργανώτρια της καθημερινότητας στο γραφείο. Ο Δημήτρης, ο παλιότερος του εργαστηρίου, και ο Σταύρος, ο νεώτερος, γεμάτος ενεργητικότητα. Η Νατάσα, η Χάρης, και η

Όλγα, συνομιλήτριες και φίλες σε διάφορες φάσεις. Η Δέσποινα και η Βίκυ, αναπόσπαστα μέλη του εργαστηρίου τα τελευταία χρόνια. Τους ευχαριστώ όλους για τη συντροφιά τους, απολογούμαι που τα πειράγματά μου ήταν κάποιες φορές υπερβολικά τραχιά. Η καθημερινότητα στο μικρόκοσμο του εργαστηρίου θα μου λείπει πολύ.

Ήμουν τυχερός που το θέμα του διδακτορικού μου στα πλαίσια του προγράμματος ΠΕΝΕΔ με έφερε σε επαφή με τον ιδιαίτερα ενδιαφέροντα κόσμο της αρχαιολογίας και τους ανθρώπους της. Η υποψήφια διδάκτορας Φραγκούλα Γεώρμα και ο καθηγητής Ανδρέας Βλαχόπουλος πρόσφεραν απλόχερα το χρόνο τους για να με εισαγάγουν στο χώρο και τις τοιχογραφίες του Ακρωτηρίου Θήρας, ενώ υπήρξαν εξαιρετικά φιλόξενοι κατά τις επισκέψεις μου επιτόπου στο χώρο της ανασκαφής. Οι καθηγητές Κλαίρη Παλυβού και Κώστας Τζαφέστας και οι υπόλοιποι υποψήφιοι διδάκτορες Μάνθος Αληφραγκής και Κώστας Αθανασίου συνέβαλαν ουσιαστικά στην πρόοδο της κοινής μας προσπάθειας. Η εξωστρέφεια και η διεπισημονική προσέγγιση της ανασκαφής οφείλονται στο όραμα του Διευθυντή της, καθηγητή Χρίστου Ντούμα. Είχα επίσης την ευκαιρία να συμμετέχω και σε άλλα ελληνικά και ευρωπαϊκά προγράμματα του εργαστηρίου μας. Εξαιρετικά χρήσιμη ήταν η εμπλοκή μου στα Έργα MUSCLE και ASPI – επωφελήθηκα ιδιαίτερα συνεργαζόμενος με τους καθηγητές Αλέξανδρο Ποταμιάνο και Anil Kokaram. Υποτροφίες από το Ίδρυμα Κρατικών Υποτροφιών και το Κοινωφελές Ίδρυμα Αλέξανδρος Σ. Ωνάσης προσέφεραν σε διάφορες φάσεις της διδακτορικής μου έρευνας σημαντική οικονομική και ηθική υποστήριξη.

Είμαι ευγνώμων στα μέλη της τριμελούς επιτροπής του διδακτορικού μου, καθηγητές Στέφανο Κόλλια και Κώστα Τζαφέστα για τις χρήσιμες συμβουλές τους καθόλη τη διάρκεια αυτής της πορείας· στα επιπλέον μέλη της πενταμελούς επιτροπής καθηγητές Γεώργιο Παπαβασιλόπουλο και Τίμο Σελλή για τα κριτικά τους σχόλια κατά την παρουσίαση της ενδιάμεσης προόδου· στον καθηγητή Ανδρέα Βλαχόπουλο και στον Δρ. Γεράσιμο Ποταμιάνο που συμπλήρωσαν την επταμελή επιτροπή εξέτασης. Όλοι διάβασαν προσεκτικά το κείμενο της διατριβής και συνεισέφεραν με τις εποικοδομητικές τους παρατηρήσεις.

Όλη αυτήν την περίοδο είχα τη στήριξη πολλών δικών μου ανθρώπων εκτός εργαστηρίου, για τους οποίους το ακριβές θέμα του διδακτορικού μου είχε ευτυχώς μικρή σημασία. Από τα σχολικά χρόνια ο Δημήτρης, ο Αλέξης, η Νότα, ο Θωμάς. Με κοινή αφετηρία τις προπτυχιακές σπουδές, ο Φίλιππος, ο Θεόδωρος, ο Νίκος, και ο Κωστής. Μετέπειτα (μέσω Σταμάτη) η Ελίζα και η Στέλλα. Ο Αλέξης, φίλος κολλητός από μικρά παιδιά. Η Λίντα που ήταν πάντα εκεί παρότι συνήθως βρισκόταν μακριά. Η ευρύτερη οικογένειά μου: τα αδέρφια μου Ανδρέας και Κώστας· η Έφη και η Λίζα· τα ανιψάκια μου Μαρία, Νίκος, και Δανάη που φέρναν απίστευτη ζωτικότητα στο σπίτι· οι θείοι Ελένη και Γρηγόρης. Οι γονείς μου Νίκος και Μαίρη που μου πρόσφεραν τα πάντα. Τους ευχαριστώ.

Γιώργος Παπανδρέου

ΠΕΡΙΛΗΨΗ

Η εργασία μας κινείται στην περιοχή της ανάλυσης εικόνων και όρασης υπολογιστών και εμπνέεται από το πρόβλημα της ψηφιακής αποκατάστασης των τοιχογραφιών του Ακρωτηρίου Θήρας. Λόγω της θεματολογίας της, η εργασία μας εντάσσεται στο ευρύτερο πλαίσιο πρωτοβουλιών για αξιοποίηση σύγχρονων ψηφιακών τεχνολογιών στην καταγραφή, αποκατάσταση, μελέτη, και διάδοση της πολιτισμικής μας κληρονομιάς. Τα προβλήματα που πραγματεύεται η διατριβή κινούνται τόσο στο επίπεδο των πρακτικών αναγκών της συγκεκριμένης εφαρμογής, όσο και στο επίπεδο θεωρητικών ζητημάτων και θεμάτων βασικής έρευνας στην κατεύθυνση ανάπτυξης μαθηματικών μοντέλων κατάλληλων για την περιγραφή εικόνων του φυσικού κόσμου.

Σε αυτό το πλαίσιο, περιγράφουμε καταρχήν τη λεπτομερή φωτογραφική αποτύπωση σημαντικού αριθμού από τις τοιχογραφίες του Ακρωτηρίου Θήρας. Η λύση που έχουμε δώσει στο πρόβλημα της αποτύπωσης μιας ολόκληρης τοιχογραφίας σε μία ενοποιημένη σύνθεση υπερ-υψηλής ανάλυσης συνίσταται στη λήψη μεγάλου πλήθους φωτογραφιών, με καθεμία από αυτές να απεικονίζει μέρος του τοιχογραφικού συνόλου, με μετέπειτα σύνθεση των επιμέρους στιγμιότυπων σε ένα γιγαντιαίο φωτογραφικό μωσαϊκό ανά σύνθεση στο οποίο μπορεί εύκολα να περιηγηθεί ο μελετητής των τοιχογραφιών. Η προσέγγιση αυτή, αυτοματοποιημένη σε μεγάλο βαθμό χάρη στη χρήση σύγχρονων τεχνικών όρασης υπολογιστών, έχει αποδειχθεί ιδιαίτερα επιτυχημένη και αναδεικνύεται ως μια πρακτική και ιδιαίτερα αποδοτική μέθοδος φωτογραφικής αποτύπωσης έργων τέχνης, δίχως την ανάγκη ύπαρξης ιδιαίτερα ακριβούς εξειδικευμένου εξοπλισμού.

Με εφαρμογή την ψηφιακή αποκατάσταση των ελλειπουσών τμημάτων στις τοιχογραφίες, στο επίκεντρο της ερευνητικής μας προσπάθειας βρίσκεται το πρόβλημα της ενδοσυμπλήρωσης εικόνων, που αναφέρεται στην συμπλήρωση κενών σε εικόνες. Με αναφορά στο πρόβλημα της ενδοσυμπλήρωσης, μελετούμε δύο ευρείες οικογένειες μεθόδων μοντελοποίησης εικόνων. Η πρώτη προσέγγιση συνίσταται σε γεωμετρικές μεθόδους περιγραφής εικόνων εμπνευσμένες από φυσικές διαδικασίες διάχυσης και εκφρασμένες με φορμαλισμό μεταβολικού λογισμού και μερικών διαφορικών εξισώσεων. Η κύρια πρωτότυπη ερευνητική μας συνεισφορά σε αυτήν την περιοχή συνίσταται στην ανάπτυξη ιδιαίτερα αποδοτικών πολυπλεγματικών αλγορίθμων για την αριθμητική επίλυση μοντέλων μη γραμμικής ανισοτροπικής διάχυσης, με εφαρμογή τόσο στο πρόβλημα της ενδοσυμπλήρωσης, όσο και στο πρόβλημα της κατάτμησης εικόνας με τεχνικές γεωμετρικών ενεργών περιγραμμάτων. Επίσης προτείνουμε αλγορίθμους στατιστικής χροιάς για το ζήτημα της βέλτιστης επιλογής κλίμακας στο πρόβλημα της αποθορυβοποίησης εικόνων.

Η δεύτερη προσέγγιση στο πρόβλημα της ενδοσυμπλήρωσης που μελετάμε στη διατριβή συνίσταται στην περιγραφή εικόνων στο πεδίο του μετασχηματισμού κυματιδίων, που αποτελεί προνομιακό πεδίο για την στατιστική περιγραφή εικόνων. Σε αυτό το πλαι-

σιο συζητάμε τις στατιστικές ιδιότητες φυσικών εικόνων στην κυματιδιακή αναπαράσταση και μελετάμε σε βάθος το πολυκλιμακωτό δενδρικό μαρκοβιανό HMT στατιστικό μοντέλο που τις περιγράφει ικανοποιητικά. Η κύρια πρωτότυπη ερευνητική μας συνεισφορά σε αυτήν την περιοχή συνίσταται στην ανάπτυξη πιθανοτικών μεθόδων ενδοσυμπλήρωσης βασισμένων στην κυματιδιακή HMT αναπαράσταση. Επιπλέον προτείνουμε νέους αλγόριθμους για εκπαίδευση των παραμέτρων του HMT μοντέλου από θορυβώδη δεδομένα, με εφαρμογή στην αποθορυβοποίηση εικόνας. Με βάση την πειραματική διερεύνηση των προτεινόμενων τεχνικών, εντοπίζουμε πλεονεκτήματα και αδυναμίες τους και συζητάμε πιθανές μελλοντικές βελτιώσεις τους.

Στο Παράρτημα της διατριβής συζητάμε τέλος μια άλλη συνιστώσα της διδακτορικής μας έρευνας στην περιοχή της συνδυασμένης οπτικής-ακουστικής αναγνώρισης ομιλίας. Οι ερευνητικές συνεισφορές μας εντοπίζονται στην ανάπτυξη ενός θεωρητικού πλαισίου για βέλτιστη προσαρμοστική σύμμιξη πολυαισθητηριακών δεδομένων βασισμένη στη διαχείριση της αβεβαιότητας των επιμέρους θορυβωδών μετρήσεων (συνεργατική εργασία), καθώς και στην ανάπτυξη τεχνικών όρασης υπολογιστών για εξαγωγή οπτικών χαρακτηριστικών ομιλίας από το πρόσωπο του ομιλητή.

ABSTRACT

Our work lies in the area of image analysis and computer vision and is motivated by the problem of digitally restoring the wall paintings excavated at Akrotiri, Thera. Its subject makes it part of a broader recent trend to utilize modern digital technologies in order to facilitate capturing, restoring, studying, and disseminating our cultural heritage. The thesis tackles problems related both to the specific practical needs of the Akrotiri Thera application as well as broader theoretical basic research issues in the direction of developing mathematical models for describing natural images.

In this framework, we start by describing the process we followed for the detailed photographic acquisition of a large number of Thera wall paintings. We have addressed the problem of digitally capturing a whole wall painting into a single ultra high resolution composition. Our approach consists of, first, capturing a large number of photographs, each depicting in detail parts of the composition, which we then stitch into a huge photographic mosaic. A web-based interface facilitates browsing through these high resolution panoramic mosaics. This approach, largely automated thanks to the use of modern computer vision techniques, has proven particularly successful and stands out as a practical and effective method for capturing high resolution photographs of artwork without the need for using highly specialized and expensive equipment.

With application to the digital restoration of missing parts in the Thera wall paintings, our research has focused on image inpainting, the problem of filling-in missing image parts. In relation to image inpainting, we study two broad classes of image modeling methods. The first class consists of geometric image description methods inspired by physical diffusion processes and formulated as variational or partial differential equation models. Our main novel contribution in this area is the development of fast multigrid algorithms for the numerical solution of nonlinear anisotropic diffusion models, with application not only to image inpainting but also to image segmentation with geometric active contour models. We also propose statistically motivated algorithms for optimal scale selection in the context of image denoising with nonlinear scale-space methods.

The thesis also studies a second statistical wavelet-based approach to image inpainting. In this framework, we discuss important statistical properties of natural images in the wavelet representation, captured to a good extent by the multiscale hidden Markov tree (HMT) model which we study in depth. Our main novel contribution in this area is the development of probabilistic image inpainting methods based on the HMT representation. We also propose new algorithms for learning the HMT model parameters from noisy data, with application to image denoising. Through a

series of experiments involving the proposed models, we highlight their strengths and weaknesses and discuss possible future improvements.

In the thesis Appendix we also discuss another aspect of our doctoral thesis research on audiovisual speech recognition. Our main research contribution in this area is the development of a theoretical framework for optimal adaptive fusion of multimodal data based on careful modeling of the individual feature measurement uncertainty (collaborative work), as well as the development of computer vision techniques for visual speech feature extraction from the speaker's face.

Κεφάλαιο 1

Εισαγωγή

Όραση Υπολογιστών

Η όραση υπολογιστών ασχολείται με την αυτοματοποιημένη εξαγωγή χρήσιμων πληροφοριών από εικόνες ή ακολουθίες εικόνων [Mar82, Mar02, FP03]. Το πεδίο εφαρμογών της είναι ιδιαίτερα ευρύ, αφού η εικόνα περιέχει μεγάλα ποσά πληροφορίας που μπορούν να αξιοποιηθούν ποικιλότροπα από τις μηχανές. Απώτερος στόχος είναι η κατασκευή συστημάτων τα οποία θα αντιλαμβάνονται τον ορατό κόσμο με ανθρώπινη ευφυΐα. Αν και η έρευνα στην περιοχή έχει κάνει τα τελευταία χρόνια σημαντική πρόοδο, απέχουμε ακόμα πάρα πολύ από την εκπλήρωση αυτού του οράματος.

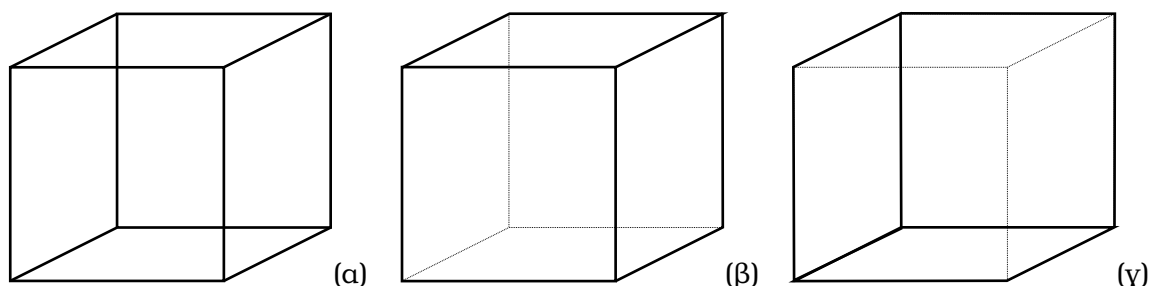
Τα πρώτα βήματα στην όραση υπολογιστών έγιναν τη δεκαετία του '60. Σχεδόν εξαρχής έγινε αντιληπτό ότι η εξαγωγή χρήσιμων συμπερασμάτων με συστήματα μηχανικής όρασης είναι μια εξαιρετικά δύσκολη υπόθεση. Πλήθος παραγόντων (γεωμετρία και ανακλαστικότητα αντικειμένου, σχετική θέση αντικειμένου-παρατηρητή, φωτισμός σκηνής, μεταξύ άλλων) αλληλεπιδρούν και συνδιαμορφώνουν την τελική εικόνα/2-Δ προβολή στον αμφιβληστροειδή χιτώνα του ανθρώπινου ματιού ή στους φωτογραφικούς αισθητήρες. Εξαιτίας αυτής της πολυπλοκότητας, οι πρώτες προσπάθειες για κατασκευή συστημάτων μηχανικής όρασης είχαν περιορισμένη μόνο επιτυχία και μάλιστα υπό ελεγχόμενες συνθήκες και σε πολύ συγκεκριμένα πεδία εφαρμογών. Ανάλογες εξάλλου δυσκολίες συνάντησαν οι πρώτοι αισιόδοξοι ερευνητές σε συναφή πεδία της τεχνητής νοημοσύνης, όπως στην επεξεργασία φυσικού λόγου [JM08]. Καταδείχθηκε έτσι ότι είναι δύσκολο να μιμηθούμε μηχανιστικά λειτουργίες που για τον άνθρωπο φαίνονται εύκολες αφού εκτελούνται μη συνειδητά.

Μία προσέγγιση για να αντιμετωπιστεί η πολυπλοκότητα αυτών των φαινομένων, μερικώς εμπνευσμένη από τη μελέτη βιολογικών συστημάτων όρασης, είναι να αποσυντεθεί το συνολικό πρόβλημα της όρασης σε πιο βατά υποπροβλήματα, οργανωμένα σε μια ιεραρχία *χαμηλού-μέσου-υψηλού* επιπέδου. Με αυτόν τον τρόπο, ξεκινώντας από τις απλές μετρήσεις φωτεινότητας στο 2-Δ επίπεδο της εικόνας και μέσω ενδιάμεσων συμβολικών αναπαραστάσεων ('σκίτσο', αποτελούμενο από ακμές, μονάδες υψής, κτλ.) καταλήγουμε σε ανακατασκευή 3-Δ αντικειμένων στα οποία μπορούν στη συνέχεια να αποδωθούν σημασιολογικές ιδιότητες [Mar82]. Η αρχική διατύπωση αυτής της θεωρίας όρασης από τον Marr προέβλεπε ότι οι υπολογισμοί σε αυτήν την αρχιτεκτονική ακολουθούν αυστηρά *άνωθεν* (bottom-up) ροή. Πολλοί ερευνητές αργότερα ανέδειξαν τη σημασία επαύξησης του μοντέλου με *κάτωθεν* (top-down) μηχανισμούς ανάδρασης πληροφορίας, που είναι

απαραίτητοι για την άρση αμφισημιών στα χαμηλότερα επίπεδα, ειδικά υπό δύσκολες οπτικές συνθήκες.

Αντίστροφα Προβλήματα Όρασης και Μηχανισμοί Κανονικοποίησης

Από μια άποψη, η διαδικασία της όρασης αποτελεί ένα *αντίστροφο* πρόβλημα: δεδομένης της 2-Δ πληροφορίας που φθάνει στους αισθητήρες και αποτελεί την εικόνα, επιθυμούμε να ανακτήσουμε τις γεωμετρικές και οπτικές ιδιότητες 3-Δ αντικειμένων. Καθότι συχνά πολλές διατάξεις της οπτικής σκηνής είναι συμβατές με τη 2-Δ εικόνα, η διαδικασία αντιστροφής χαρακτηρίζεται από έντονη αμφισημία. Ο ανθρώπινος εγκέφαλος φαίνεται ότι υπερβαίνει τέτοιου είδους αμφισημίες συνδυάζοντας τις οπτικές μετρήσεις με πρότερη γνώση του κόσμου, μια αντιληπτική διαδικασία που πρώτος ο von Helmholtz περιέγραψε καλώντας την *ασύνειδο συμπερασμό* (unconscious inference) [vH25, Τόμος 3]. Ένα γνωστό παράδειγμα που αναδεικνύει την αμφισημία στην ανακατασκευή 3-Δ αντικειμένων από 2-Δ πληροφορία είναι ο 'κύβος του Necker' που παρουσιάζεται στο Σχ. 1.1. Πέρα από την απώλεια πληροφορίας λόγω προβολής 3-Δ αντικειμένων σε 2-Δ επιφάνεια, πολλές λειτουργίες όρασης είναι δύσκολο να επιτελεσθούν με χρήση αποκλειστικά τοπικών μετρήσεων. Ένα τέτοιο παράδειγμα πολυσημίας, το *φαινόμενο διαφράγματος* (aperture effect), παρουσιάζεται κατά τον υπολογισμό της οπτικής ροής σε ακολουθίες εικόνων. Άλλες αιτίες, όπως μειωμένος φωτισμός, *αβλήθηλεπικαλύψεις* (occlusion) αντικειμένων και θόρυβος μέτρησης, δυσχεραίνουν περαιτέρω την κατανόηση του περιβάλλοντος με οπτικά μέσα.



Σχήμα 1.1: Αμφισημία στην ερμηνεία οπτικών ερεθισμάτων: Κύβος του Necker. Με το 2-Δ κύβο του σχήματος (α) είναι συμβατές και οι δύο 3-Δ γεωμετρικές ερμηνείες των σχημάτων (β) και (γ). Εμπνευσμένο από αντίστοιχο σχήμα στο [Mar82].

Για να ανταπεξέλθουν στο δύσκολο πρόβλημα οπτικής αντιστροφής, τα συστήματα όρασης πρέπει να περιλαμβάνουν μηχανισμούς *κανονικοποίησης* (regularization) [PTK85, MMP87]. Αυτοί οι μηχανισμοί επαυξάνουν την πληροφορία προερχόμενη από οπτικές μετρήσεις με εμπειρική γνώση σχετικά με κανονικότητες του περιβάλλοντος και περιορίζουν κατάλληλα την λύση προβλημάτων οπτικής αντιστροφής. Τέτοιοι μηχανισμοί συνήθως εντάσσονται σε μοντέλα όρασης είτε με τη μορφή όρων κανονικοποίησης σε φορμαλισμό μεταβολικού λογισμού/μερικών διαφορικών εξισώσεων [TA77], είτε με τη μορφή πρότερης γνώσης σε πιθανοτικούς φορμαλισμούς της όρασης [KR96]. Όπως έχει αναδειχθεί κυρίως μέσα από τη *θεωρία προτύπων* (pattern theory) του Grenander [Gre93, Mum96, GM07], ο πιθανοτικός φορμαλισμός της όρασης τονίζει την ανάγκη για ανάπτυξη πιστών *αναγεννητικών* (generative) πιθανοτικών μοντέλων για σύνθεση εικόνων. Εισάγοντας τις οπτικές μετρήσεις στο αναγεννητικό μοντέλο και συνδυάζοντάς τες με την

πρότερη πιθανοτική περιγραφή οδηγούμαστε, μέσω του θεωρήματος Bayes, σε τεχνικές ‘ανάλυσης-μέσω-σύνθεσης’ που αυτόματα ενσωματώνουν λειτουργίες κανονικοποίησης.

Ενδοσυμπλήρωση Οπτικής Πληροφορίας

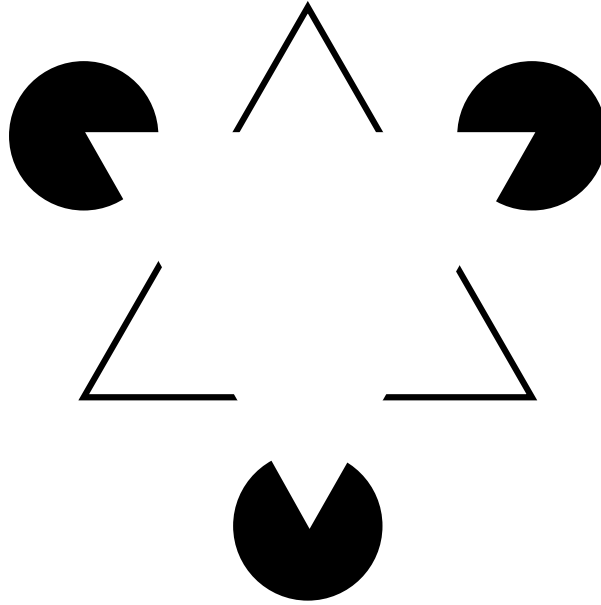
Μια σημαντική αντίστροφη λειτουργία όρασης που βρίσκεται στο επίκεντρο της διδακτορικής μας έρευνας είναι το πρόβλημα της *ενδοσυμπλήρωσης* (inpainting/infilling) εικόνων. Η ενδοσυμπλήρωση έγκειται στην ανάκτηση ελλειπουσών περιοχών εικόνας με βάση τη διαθέσιμη πληροφορία στις προσκείμενες γνωστές περιοχές της εικόνας. Εμπλουτίζοντας την πληροφορία μέτρησης με γνώση για τις κανονικότητες των εικόνων μπορούμε να μαντέψουμε το περιεχόμενο της εικόνας στις αποκρυμμένες περιοχές. Θα εξετάσουμε τεχνικές ενδοσυμπλήρωσης βασισμένες τόσο σε μοντελοποίηση μερικών διαφορικών εξισώσεων, όσο και σε στοχαστική μοντελοποίηση στο πεδίο των κυματιδίων.

Όπως έχει αναδειχθεί με κλασικά ψυχοφυσιολογικά πειράματα, μηχανισμοί σχετιζόμενοι με οπτική ενδοσυμπλήρωση ενυπάρχουν στο ανθρώπινο σύστημα όρασης [PTN98, Pal03]. Τέτοιοι μηχανισμοί αποκαλύπτονται με γλαφυρό τρόπο στο φαινόμενο των *απατηλών περιγραμμάτων* (illusory contours) όπου ο παρατηρητής προσλαμβάνει την ύπαρξη περιγραμμάτων παρά την απουσία ακμών από την εικόνα. Ένα χαρακτηριστικό τέτοιο παράδειγμα είναι το τρίγωνο του Kanizsa [Kan79], που δείχνουμε στο Σχ. 1.2. Αυτά τα φαινόμενα μπορούν να ερμηνευθούν με χρήση της έννοιας της *αντιληπτικής οργάνωσης* (perceptual organization) στο πλαίσιο ιδεών της ψυχολογικής Σχολής του Gestalt [Kof35, MGKP08]. Η αντιληπτική οργάνωση σχετίζεται με την ανάλυση της δομής των αντικειμένων της εικόνας και κατανόηση των μεταξύ τους σχέσεων (π.χ. ιεραρχική συσχέτιση μεταξύ των αντικειμένων της εικόνας, διαχωρισμός προσκηνίου/παρασκηνίου). Στόχος είναι μια εύρωστη και οικονομική περιγραφή της εικόνας, γνωστό και ως Gestalt αρχή του ‘Prägnanz’ [Kof35], σε στενή εννοιολογική συνάφεια με το *ξυράφι του Occam* (Occam’s razor) και την *ελαχιστοποίηση του μήκους περιγραφής* (minimum description length–MDL), ως αρχές για την επιλογή στατιστικού μοντέλου. Μέσο για να επιτευχθεί η αντιληπτική οργάνωση είναι η *αντιληπτική ομαδοποίηση* (perceptual grouping) με βάση κριτήρια όπως η εγγύτητα, ομοιότητα χρώματος/μεγέθους/κατεύθυνσης, η συνέχιση, και το κλείσιμο που συνοψίζονται στο Σχ. 1.3. Διαδικασίες αντιληπτικής οργάνωσης φαίνεται να επενεργούν τόσο στο χαμηλό, όσο και σε υψηλότερα επίπεδα όρασης [PBN03]. Υπό το πρίσμα της αντιληπτικής οργάνωσης, μηχανισμοί οπτικής ομαδοποίησης επενεργούν κατά την ενδοσυμπλήρωση εικόνων και βοηθούν στην εξάλειψη των ελλειπόντων τμημάτων.

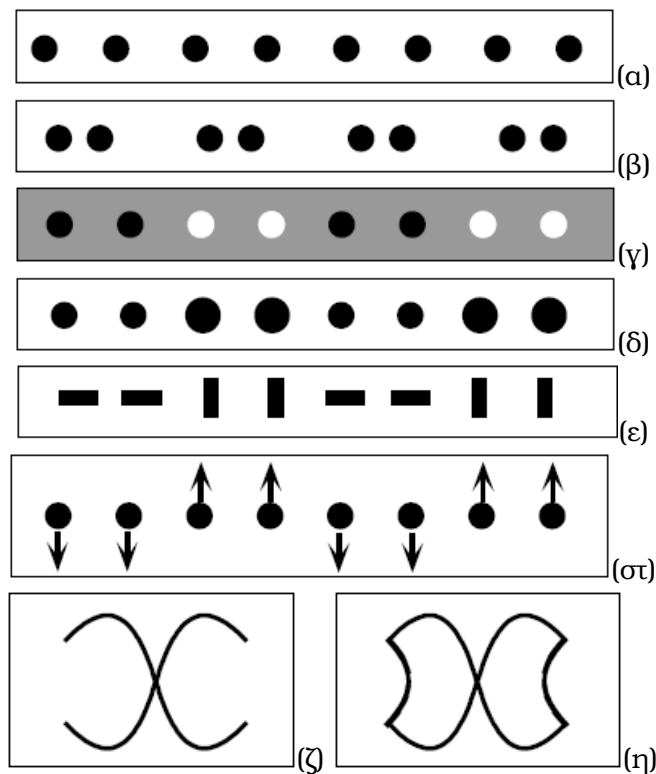
Τεχνικές ενδοσυμπλήρωσης παρουσιάζουν ενδιαφέρον σε πολλά πεδία εφαρμογών. Θα αναφέρουμε ενδεικτικά την ανάκτηση απωλεσθέντων μπλοκ εικόνας σε βιντεοροές που μεταδίδονται μέσω τηλεπικοινωνιακών καναλιών με απώλειες, την ψηφιακή αποκατάσταση αρχειοθετημένων φιλμ σε ταινιοθήκες, το ψηφιακό *επανάγγιγμα/ρετουσάρισμα* (retouching) φωτογραφιών (δες το Σχ. 1.4 για ένα διάσημο παράδειγμα), και τις ψηφιακές επεμβάσεις για διόρθωση ατελειών και ειδικά εφέ κατά την *τελική επεξεργασία* (post-production) στην παραγωγή κινηματογραφικών ταινιών.

Αντικείμενο και Συνεισφορές της Διατριβής

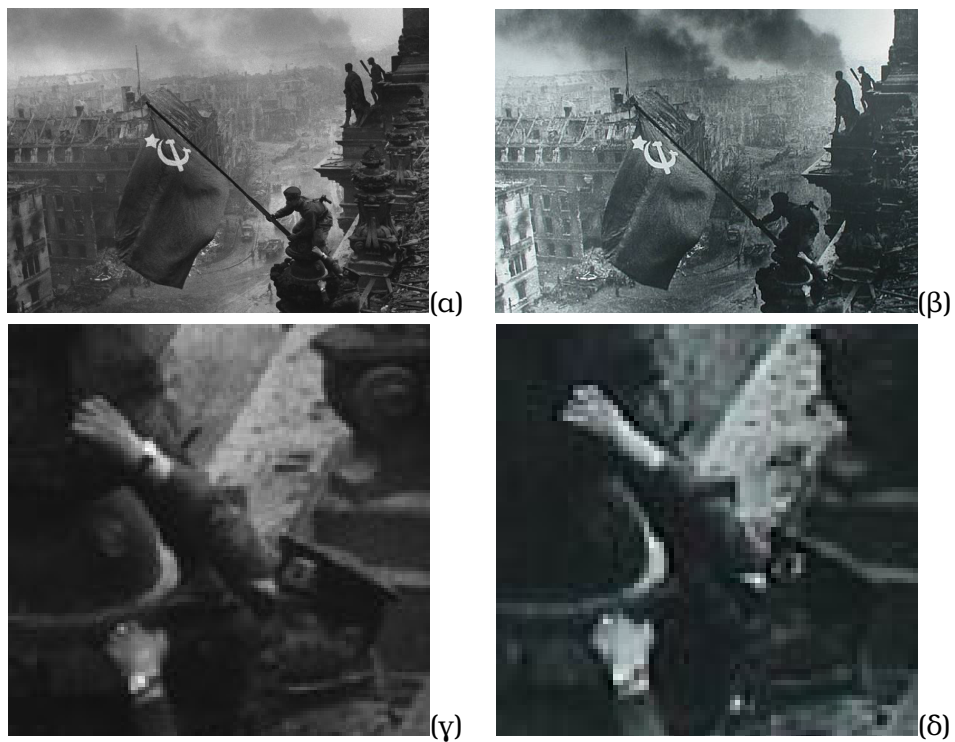
Βασικό πεδίο εφαρμογής στο οποίο επικεντρώθηκε η εργασία μας είναι η ψηφιακή αποκατάσταση προϊστορικών τοιχογραφιών από την ανασκαφή του Ακρωτηρίου Θήρας·



Σχήμα 1.2: Απατηλά περιγράμματα στην ανθρώπινη όραση: Το τρίγωνο του Kanizsa [Kan79].



Σχήμα 1.3: Κλασικά κριτήρια Gestalt ομαδοποίησης κατά τον Wertheimer [Wer23]. Σχήμα προσαρμοσμένο από το [Pal03]. Καμία ομαδοποίηση (α), εγγύτητα (β), ομοιότητα χρώματος (γ) / μεγέθους (δ) / κατεύθυνσης (ε), κοινή μοίρα (στ), συνέχιση (ζ), και κλείσιμο (η).



Σχήμα 1.4: Επανάγγιγμα φωτογραφίας: Έπαρση της Σοβιετικής σημαίας στο Reichstag του Βερολίνου. Η αρχική φωτογραφία (α, λεπτομέρεια στο γ) επαναγγίχθηκε ώστε στη δημοσιευμένη φωτογραφία (β, λεπτομέρεια στο δ) να μην είναι ορατό το δεύτερο ρολόι (πιθανότατα λάφυρο) στο δεξί καρπό του στρατιώτη. Φωτογράφος: Υ. Khaldei, 1945.

παραδείγματα αυτών των τοιχογραφιών φαίνονται στο Σχ. 2.5, σελ. 17. Η δραστηριότητα αυτή εμπίπτει στο γενικότερο πλαίσιο πρωτοβουλιών για χρήση ψηφιακών τεχνικών στη συντήρηση, διατήρηση, και ανάδειξη της πολιτισμικής κληρονομιάς. Με χρήση τεχνικών ενδοσυμπλήρωσης προσπαθούμε να αποκαταστήσουμε ψηφιακά τα μικρά ή μεγαλύτερα ελλείποντα τμήματα των τοιχογραφιών. Τα αντίστοιχα *σπαράγματα* (fragments) είτε έχουν για πάντα καταστραφεί/θρυμματιστεί, είτε δεν έχουν ακόμα αντιστοιχηθεί με τις υπό συντήρηση τοιχογραφίες. Πέρα από ολόκληρα τμήματα που λείπουν, πρόσθετες φθορές παρατηρούνται υπό τη μορφή είτε ισχυρών ρωγμών στο επίχρισμα (στα σημεία επαφής των συγκολληθέντων σπαραγμάτων), είτε ελαφρύτερων ρωγμών στο χρώμα. Λόγω της μεγάλης έκτασης και ποικιλίας αυτών των αλλοιώσεων (μεγάλα ελλείποντα τμήματα, ισχυρές/ελαφρές ρωγμές), η ψηφιακή τους αποκατάσταση παρουσιάζει αυξημένη δυσκολία.

Ως ένα πρώτο ουσιαστικό βήμα για την πρόοδο του διδακτορικού, προχωρήσαμε στη λεπτομερή φωτογραφική αποτύπωση σημαντικού αριθμού από τις τοιχογραφίες του Ακρωτηρίου Θήρας. Μια και δεν ήταν εφικτό να αποτυπώσουμε ολόκληρες τις τοιχογραφίες σε μια φωτογραφία με επαρκή χωρική ανάλυση, αποφασίσαμε να προχωρήσουμε σε λήψη μεγάλου πλήθους φωτογραφιών, με καθεμία από αυτές να απεικονίζει μέρος του τοιχογραφικού συνόλου. Προχωρήσαμε μετέπειτα στην ημιαυτόματη συνένωση των επιμέρους στιγμιότυπων σε ένα γιγαντιαίο φωτογραφικό μωσαϊκό ανά τοιχογραφία, στο οποίο είναι εύκολο να περιηγηθεί κανείς και να μελετήσει με λεπτομέρεια κάθε τοιχογραφική σύνθεση. Αν και η παραπάνω διαδικασία δεν εμπεριείχε στοιχεία βασικής έρευνας, αναδεικνύει τη συγκεκριμένη τεχνική ως μια πρακτική και ιδιαίτερα αποδοτική μέθοδο φωτογραφικής αποτύπωσης έργων τέχνης, δίχως την ανάγκη ύπαρξης ιδιαίτερα ακριβούς εξειδικευμένου εξοπλισμού.

Στο επίπεδο της βασικής έρευνας, ένα σημαντικό μέρος από τις πρωτότυπες ερευνητικές μας συνεισφορές στην ενασχόλησή μας με το πρόβλημα της αποκατάστασης εικόνων μπορεί να ιδωθεί υπό το ενοποιητικό πλαίσιο της πολυκλιμακωτής ανάλυσης εικόνας. Πιο συγκεκριμένα, η ιδέα της πολυκλιμακωτής ανάλυσης ανακύπτει υπό διάφορες μορφές στην εργασία μας:

- Με τη μορφή των *πολυπλεγματικών* (multigrid) αλγορίθμων, δείχνουμε πως η πολυπλεγματική ανάλυση μπορεί να αποδειχθεί ένα ισχυρό υπολογιστικό εργαλείο για την ταχεία αριθμητική επίλυση μοντέλων κατάτμησης και ενδοσυμπλήρωσης εκφρασμένων σε φορμαλισμό μερικών διαφορικών εξισώσεων – Κεφάλαιο 3.
- Ο πολυκλιμακωτός μετασχηματισμός *κυμαιδίων* (wavelet) εμφανίζει σημαντικά προτερήματα ως πεδίο ιεραρχικής ανάλυσης εικόνων. Ο συνδυασμός του με αναδρομικές ως προς την κλίμακα μαρκοβιανές πιθανοτικές δομές αποφέρει εκφραστικά και αποδοτικά μοντέλα για περιγραφή φυσικών εικόνων, τα οποία και δείχνουμε πως μπορούν να χρησιμοποιηθούν για ενδοσυμπλήρωση εικόνων – Κεφάλαια 5 και 6.
- Η περιγραφή της εικόνας σε διαδοχικές κλίμακες συνιστά το *χώρο κλίμακας* (scale space). Αν και η πλήρης περιγραφή της δομής της εικόνας απαιτεί ολόκληρο το χώρο κλίμακας, για ορισμένες εφαρμογές αρκεί να επιλέξει κανείς ένα μόνο αντιπροσωπευτικό του στιγμιότυπο. Για την εφαρμογή της αποθρομβοποίησης εικόνων, δείχνουμε πως η επιλογή του κατάλληλου στιγμιότυπου από το χώρο κλίμακας

μπορεί να φορμαλιστεί ως ένα πρόβλημα επιλογής στατιστικού μοντέλου και να αντιμετωπιστεί αποδοτικά με μεθόδους στατιστικής *διεπικύρωσης* (cross-validation) – Κεφάλαιο 4.

Πολλά από τα προβλήματα που έχουμε μελετήσει στο πλαίσιο της διατριβής έχουν έντονη στοχαστική διάσταση, όπως παρουσία ελλειπών ή θορυβωδών δεδομένων. Η χρήση μπεϋζιανών στατιστικών τεχνικών για την αναπαράσταση και διαχείριση της σχετιζόμενης αβεβαιότητας διατρέχει έτσι την εργασία μας. Πρωτότυπες ερευνητικές συνεισφορές μας που μπορούν να ιδωθούν υπό αυτήν την οπτική γωνία είναι:

- Για μια ευρεία κατηγορία παραμετρικών δενδρικών μαρκοβιανών μοντέλων εικόνων, δείχνουμε πώς μπορούν να εκπαιδευθούν από θορυβώδεις παρατηρήσεις και, κυρίως, πώς μπορούν να χρησιμοποιηθούν για συμπλήρωση κενών σε εικόνες. Οι προτεινόμενες τεχνικές βασίζονται σε καλά θεμελιωμένες τεχνικές μπεϋζιανής συμπερασματολογίας και οδηγούν σε σειρά πρακτικών ντετερμινιστικών και στοχαστικών αλγορίθμων για ενδοσυμπλήρωση εικόνων – Κεφάλαιο 6.
- Με αφετηρία το πρόβλημα της οπτικής-ακουστικής αναγνώρισης ομιλίας, έχουμε αναπτύξει μια θεωρία για βέλτιστη προσαρμοστική σύμμειξη πολυαισθητηριακών δεδομένων βασισμένη στη διαχείριση της αβεβαιότητας των επιμέρους θορυβωδών μετρήσεων – Παράρτημα Α'. Αν και η εφαρμογή είναι τελείως διαφορετική και περιλαμβάνει 1-Δ μοντέλα μαρκοβιανών αλυσίδων, το σύνολο των μαθηματικών εργαλείων που αναπτύξαμε στο πλαίσιο αυτό αποδείχθηκε ιδιαίτερα χρήσιμο ως αφετηρία για την πληρέστερη κατανόηση και ανάπτυξη τεχνικών εκπαίδευσης υπό θορυβώδεις μετρήσεις για τα δενδρικά μαρκοβιανά μοντέλα εικόνας του Κεφαλαίου 6.

Άλλη αξιοσημείωτη συνεισφορά της διδακτορικής μας έρευνας οι οποία όμως δεν έχει συνάφεια με το πρόβλημα της αποκατάστασης των τοιχογραφιών και απλά θα σκιαγραφηθεί στο Παράρτημα Α' είναι η ανάπτυξη τεχνικών για οπτική μοντελοποίηση και εξαγωγή χαρακτηριστικών ομιλίας από το ανθρώπινο πρόσωπο. Αυτή εντάσσεται σε μια ευρύτερη συνεργατική προσπάθεια εντός του Εργαστηρίου με στόχο τη συνδυασμένη οπτική/ακουστική μελέτη της ανθρώπινης ομιλίας με εφαρμογές σε ανάλυση και σύνθεση φωνής.

Οργάνωση της Διατριβής

Το υπόλοιπο της διατριβής είναι οργανωμένο ως εξής:

Στο **Κεφάλαιο 2** δίνεται μια συνοπτική περιγραφή της αρχαιολογικής ανασκαφής στο Ακρωτήρι Θήρας και του ρόλου που μπορούν να παίξουν οι ψηφιακές τεχνολογίες στη διατήρηση και ανάδειξη των αρχαιολογικών ευρημάτων. Δίνουμε έμφαση στις τοιχογραφίες και στην ψηφιακή τους αποκατάσταση με τεχνικές ενδοσυμπλήρωσης. Επίσης περιγράφουμε τη διαδικασία με την οποία προχωρήσαμε στη λεπτομερή φωτογραφική αποτύπωση επιλεγμένων τοιχογραφιών, καθώς και τεχνικές που χρησιμοποιήσαμε για τη δημιουργία φωτογραφικών μωσαϊκών υπερ-υψηλής ανάλυσης με συρραφή μεγάλου πλήθους επιμέρους φωτογραφικών στιγμιότυπων.

Στο **Κεφάλαιο 3** περιγράφουμε μια ευρεία οικογένεια μοντέλων μερικών διαφορικών εξισώσεων που βρίσκουν εφαρμογή σε κατάτμηση και ενδοσυμπλήρωση εικόνων.

Έμφαση δίνεται στην ταχεία επίλυση αυτών των μοντέλων με πολυπλεγματικές τεχνικές, οι οποίες οδηγούν σε αλγορίθμους με υπολογιστική πολυπλοκότητα που αυξάνεται μόνο γραμμικά με τον αριθμό των εικονοστοιχείων. Αξιολογούμε ποσοτικά και ποιοτικά τους προτεινόμενους αλγορίθμους και παρουσιάζουμε ενδεικτικά αποτελέσματα ενδοσυμπλήρωσης τοιχογραφιών από το Ακρωτήρι. Σχετικές δημοσιεύσεις μας είναι οι [PM04, PM07].

Στο **Κεφάλαιο 4** περιγράφουμε το πρόβλημα επιλογής αντιπροσωπευτικού στιγμιοτύπου στην πολυκλιμακωτή ανάλυση εικόνας σε μη-γραμμικούς χώρους κλίμακας με όρους επιλογής βέλτιστου στατιστικού μοντέλου. Επικεντρωνόμαστε στην εφαρμογή της αποθορυβοποίησης εικόνων με χρήση μερικών διαφορικών εξισώσεων ή μορφολογικών τελεστών, για την οποία παρουσιάζουμε αλγορίθμους βέλτιστης επιλογής κλίμακας βασισμένους στην τεχνική της στατιστικής διεπικύρωσης. Αξιολογούμε την επίδοση της προτεινόμενης τεχνικής σε σχέση με εναλλακτικές μεθόδους επιλογής κλίμακας. Σχετικές δημοσιεύσεις μας είναι οι [PM05a, PM05b].

Στο **Κεφάλαιο 5** συζητάμε την αναπαράσταση εικόνων με χρήση χωρικά εντοπισμένων κυματιδιακών μετασχηματισμών. Παρουσιάζουμε εναλλακτικές εκδοχές αναπαράστασης με κυματίδια χώρου-κλίμακας και χώρου-συχνότητας, δίνοντας ιδιαίτερη έμφαση σε υπερπλήρεις μιγαδικούς κυματιδιακούς μετασχηματισμούς επειδή χαρακτηρίζονται από βελτιωμένες ιδιότητες μεταφορικής και περιστροφικής συμμετρίας. Συζητούμε επίσης ποιοτικά τις στατιστικές ιδιότητες εικόνων του φυσικού κόσμου όπως αυτές αναδεικνύονται στο πεδίο των κυματιδίων.

Στο **Κεφάλαιο 6** περιγράφουμε δενδρικά αναγεννητικά στατιστικά μοντέλα περιγραφής των κυματιδιακών συντελεστών, τα οποία αναπαράγουν σε σημαντικό βαθμό βασικές στατιστικές ιδιότητες των κυματιδιακών αποκρίσεων φυσικών εικόνων. Συζητάμε πώς το συγκεκριμένο μοντέλο χρησιμοποιείται για σύνθεση υψής και αποθορυβοποίηση εικόνων, όπου και παρουσιάζουμε τεχνικές εκπαίδευσης μέγιστης πιθανοφάνειας για τις οποίες αρκεί η γνώση της θορυβώδους εκδοχής της εικόνας. Ως κύρια συνεισφορά σε αυτό το πλαίσιο, προτείνουμε μια μέθοδο ενδοσυμπλήρωσης βασισμένη σε αυτό το μοντέλο, η οποία συνίσταται σε στατιστική συμπερασματολογία για τα ελλείποντα τμήματα της εικόνας με βάση πληροφορία που εξάγεται από τις γνωστές περιοχές, και δίνουμε παραδείγματα εφαρμογής της. Σχετική δημοσίευση είναι η [PMK08].

Στο **Κεφάλαιο 7** συνοψίζουμε τις συνεισφορές της διατριβής και σκιαγραφούμε κατευθύνσεις στις οποίες μπορεί να εξελιχθεί στο μέλλον η σχετική ερευνητική μας προσπάθεια. Στο **Παράρτημα Α'** περιγράφουμε με συντομία την έρευνά μας στο πρόβλημα της οπτικής-ακουστικής αναγνώρισης ομιλίας, που έχει περιληφθεί στις δημοσιεύσεις [PKPM09, KPM09]. Αν και το πεδίο εφαρμογής είναι τελείως διαφορετικό, ορισμένα από τα μαθηματικά εργαλεία που αναπτύχθηκαν στα πλαίσια της συγκεκριμένης εργασίας αποδείχθηκαν χρήσιμα κατά την ανάπτυξη των δενδρικών κυματιδιακών μοντέλων του Κεφαλαίου 6. Κλείνοντας, στο **Παράρτημα Β'** παραθέτουμε σχολιασμένη λίστα με τις δημοσιεύσεις από τη διδακτορική μας έρευνα, ορισμένες από τις οποίες περιγράφουν ερευνητική δραστηριότητα που δεν περιλαμβάνεται στην παρούσα αναφορά.

Κεφάλαιο 2

Προϊστορικές Τοιχογραφίες στο Ακρωτήρι Θήρας: Φωτογράφιση και Ψηφιακή Αποκατάσταση

Σημαντική ώθηση της ερευνητικής μας δραστηριότητας στο πρόβλημα της ενδοσυμπλήρωσης εικόνων δίνει η συμμετοχή μας στο διεπιστημονικό ερευνητικό Έργο με τίτλο ‘Όραση Υπολογιστών και Εικονική Πραγματικότητα για Τριδιάστατη Ανακατασκευή Αρχαιολογικών Ευρυμάτων με Εφαρμογές σε Τοιχογραφίες και Κτίρια του Προϊστορικού Οικισμού Ακρωτηρίου Θήρας’, που εντάσσεται στην πρωτοβουλία ΠΕΝΕΔ-2003 και χρηματοδοτείται από τη ΓΓΕΤ. Η συμβολή της δικής μας διδακτορικής διατριβής έγκειται στην ανάπτυξη τεχνικών όρασης υπολογιστών για ψηφιακή αποκατάσταση των τοιχογραφιών. Στους κύριους ερευνητές με άμεση εμπλοκή στο Έργο περιλαμβάνονται οι καθ. Π. Μαργακός (επιστημονικός υπεύθυνος του Έργου) και επ. καθ. Κ. Τζαφέστας από το ΕΜΠ· ομ. καθ. Χ. Ντούμας και επ. καθ. Α. Βλαχόπουλος από την Ανασκαφή Ακρωτηρίου Θήρας· αν. καθ. Κ. Παλυβού από το ΑΠΘ. Άλλες διδακτορικές διατριβές που υποστηρίζονται από το Έργο επικεντρώνονται στην εικονική 3-Δ ανακατασκευή κτιρίων του αρχαιολογικού χώρου Ακρωτηρίου (Μ. Αληφραγκής, Σχ. ΗΜΜΥ, ΕΜΠ), στην αρχαιολογική μελέτη του Κτιρίου Β’ του οικισμού (Φρ. Γεώργια, Τμ. Ιστορίας-Αρχαιολογίας, Παν. Ιωαννίνων), και στη χρήση αυτοματοποιημένων τεχνικών για την αρχιτεκτονική τεκμηρίωση ανασκαφών (Κ. Αθανασίου, Τμ. Αρχιτεκτόνων, Πολυτεχνική Σχολή ΑΠΘ).

Οι τοιχογραφίες που ανακαλύφθηκαν στον προϊστορικό οικισμό του Ακρωτηρίου Θήρας είναι διάσημες για τη θεματογραφία τους, την καλλιτεχνική τους αξία, και για τον πλούτο πληροφορίας που παρέχουν για τον Αιγαιακό πολιτισμό πριν από περίπου 3500 χρόνια. Οφείλουν τη διατήρησή τους ως τις μέρες μας στην ισχυρότατη έκρηξη του ηφαιστείου της Θήρας κατά τον 17ο αιώνα π.Χ., που είχε ως συνέπεια να καταπλακωθεί ολόκληρος ο προϊστορικός οικισμός από παχύ στρώμα ηφαιστειακής τέφρας. Ο οικισμός ανακαλύφθηκε και άρχισε να ανασκάπτεται το 1967 υπό την καθοδήγηση του καθ. Σ. Μαρινάτου, τον οποίο διαδέχθηκε το 1974 ο καθ. Χ. Ντούμας, ως και σήμερα διευθυντής της ανασκαφής. Η αποκατάσταση αυτών των μοναδικών τοιχογραφιών είναι ιδιαίτερα δύσκολη και χρονοβόρα. Η διαδικασία συντήρησης περιλαμβάνει, μεταξύ άλλων, συλλογή/αποκόλληση των τμημάτων της τοιχογραφίας από τον αρχαιολογικό χώρο, καθαρισμό και στερέωση των σπαραγμάτων στο εργαστήριο, καθώς και ανασύνθεση κάθε τοιχογραφίας από τα ευρεθέντα τμήματά της.

Πλούσιο πληροφοριακό υλικό τόσο για την ιστορία, όσο και για τη σημασία της ανασκαφής, με ιδιαίτερη έμφαση τις τοιχογραφίες, περιέχεται στο [Dou92]. Σημαντικό πρωτογενές αρχαιολογικό υλικό από την ανασκαφή μπορεί να βρεθεί στα [Mar99a, Mar99b, Mar99c]. Οι τοιχογραφίες της Θήρας έχουν βρεθεί στο επίκεντρο πολλών διεθνών συμποσίων και συνεδρίων· το τελευταίο σχετικό διεθνές συνέδριο έλαβε χώρα το 1997 και στα πρακτικά του [She00] δημοσιεύεται πλήθος σχετικών ανακοινώσεων. Περαιτέρω αρχαιολογικές αναφορές που βρήκαμε ιδιαίτερα χρήσιμες είναι οι [Nto92, Παλ99, Βλα03, Vla08, NMT09].

2.1 Τοιχογραφίες με Ελλείποντα Τμήματα και Ενδοσυμπλήρωση Πληροφορίας

Ψπάρχει έντονο ενδιαφέρον τα τελευταία χρόνια για ανάπτυξη αυτοματοποιημένων τεχνικών που θα επιβοηθούν το έργο των αρχαιολόγων και των συντηρητών στην ανασκαφή Ακρωτηρίου Θήρας, που εντάσσεται στην ευρύτερη τάση για αξιοποίηση τεχνολογικών καινοτομιών με στόχο τη διατήρηση της πολιτισμικής κληρονομιάς. Σημαντικές ερευνητικές προσπάθειες έχουν γίνει με σκοπό την ψηφιακή καταγραφή των τοιχογραφικών σπαραγμάτων και την υποβοήθηση της διαδικασίας συναρμολόγησης τοιχογραφικών συνόλων από τα επιμέρους διατηρηθέντα τμήματά τους· αναφέρουμε ενδεικτικά τις δημοσιεύσεις [PPE⁺02, BTFN⁺08]. Στην κατεύθυνση της πληροφοριακής υποστήριξης και τεκμηρίωσης της ανασκαφής με συστήματα βάσεων δεδομένων, αξίζει να αναφέρουμε το Έργο ARCHEOTOOL (ΓΓΕΤ, 1995-1996) με συμμετοχή του Εργαστηρίου Συστημάτων Βάσεων Γνώσεων και Δεδομένων του ΕΜΠ.

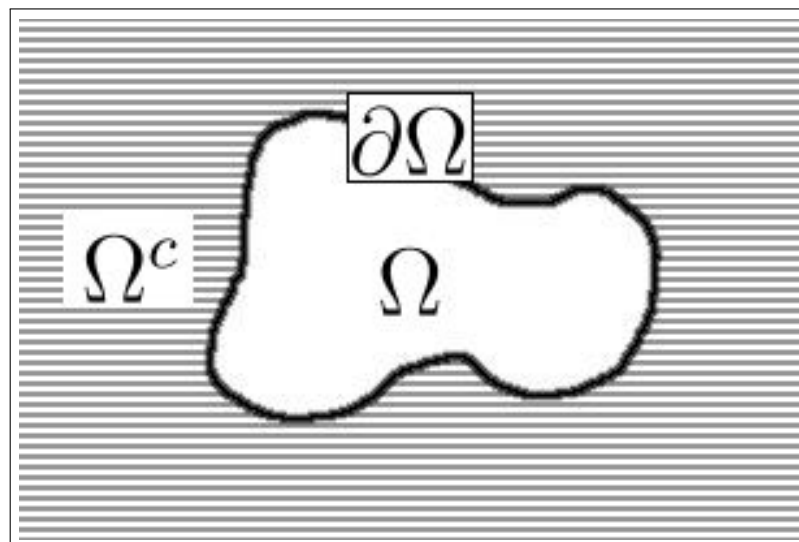
Στο επίκεντρο της δικής μας ερευνητικής προσπάθειας έχει βρεθεί το πρόβλημα της ψηφιακής ενδοσυμπλήρωσης των τοιχογραφιών. Μετά το πέρας της διαδικασίας συντήρησης, κάποια τμήματα κάθε τοιχογραφίας παραμένουν κενά, είτε επειδή τα αντίστοιχα θραύσματα δεν έχουν ακόμη ταυτοποιηθεί, είτε γιατί έχουν πλέον τελείως καταστραφεί. Σκοπός μας είναι η αυτοματοποιημένη ψηφιακή συμπλήρωση αυτών των κενών, είτε για αισθητική αποκατάσταση της τοιχογραφίας με σκοπό την παρουσίασή της στο ευρύ κοινό, είτε για υποβοήθηση του ερμηνευτικού έργου των ειδικών αρχαιολόγων. Επίσης, μέθοδοι ψηφιακής ενδοσυμπλήρωσης μπορεί να αποδειχθούν ιδιαίτερα χρήσιμες και για το πρόβλημα της αυτοματοποιημένης επανασυναρμολόγησης των τοιχογραφιών. Συγκεκριμένα, επεκτείνοντας τις τιμές της εικόνας σε μια στενή ζώνη γύρω από κάθε σπάραγμα μας επιτρέπει να χρησιμοποιήσουμε το απεικονιστικό περιεχόμενο της εικόνας ως χαρακτηριστικό επιβοηθητικό του σχήματος του περιγράμματος κατά την αξιολόγηση της ποιότητας ταιριάσματος δύο πιθανά εφαπτόμενων σπαραγμάτων.

Παίρνοντας την τοιχογραφία του Σχ. 2.6 ως παράδειγμα, θα μπορούσαμε να διακρίνουμε τις ελλείψεις που παρατηρούμε στις συντηρημένες τοιχογραφίες του Ακρωτηρίου Θήρας σε διάφορες κατηγορίες, ανάλογα με την έκταση της βλάβης:

- *Τύπος I*: Αλλοιώσεις μικρής έκτασης, π.χ. ρωγμές. Συναντώνται στις επαφές συγκολλημένων σπαραγμάτων, ή οφείλονται σε απόξεση/σκάσιμο του χρώματος από το επίχρισμα. Δεν θα αποτελούσαν σε καμία περίπτωση εμπόδιο στην αποκατάσταση της σύνθεσης από ζωγράφο-συντηρητή, με την έννοια ότι αυτή θα μπορούσε να γίνει με σχεδόν ντετερμινιστικό τρόπο. Η παρουσία τους δίνει στην τοιχογραφία την ψευδαίσθηση του ψηφιδωτού, και για αυτό είναι θεμιτή η ψηφιακή τους εξάλειψη.

- *Τύπος II*: Κενά μεσαίας έκτασης. Οφείλονται στην απώλεια ή αδυναμία αντιστοίχισης μικρού αριθμού διάσπαρτων σπαραγμάτων από τμήματα της τοιχογραφίας. Για το ανθρώπινο μάτι, ακόμη και για μη ειδικούς, η συμπλήρωση των κενών αυτών και η σκιαγράφιση των ελλειπόντων μοτίβων δεν είναι ιδιαίτερα δύσκολη, αν και σε ορισμένες περιπτώσεις περισσότερες από μία παρεμφερείς εκδοχές είναι αποδεκτές.
- *Τύπος III*: Μεγάλα ελλείποντα τμήματα. Οφείλονται στην απώλεια ή αδυναμία αντιστοίχισης μεγάλου αριθμού σπαραγμάτων από μία περιοχή της τοιχογραφίας. Καθιστούν την αποκατάσταση του συγκεκριμένου τμήματος της τοιχογραφίας ιδιαίτερα προβληματική, με την έννοια ότι απαιτείται διεξοδική μελέτη και ερμηνεία της όλης σύνθεσης από αρχαιολόγους και ζωγράφους για να αποτολμηθεί και να αιτιολογηθεί οποιαδήποτε απόπειρα αποκατάστασης. Σε ορισμένες περιπτώσεις η έκταση της αλλοίωσης μπορεί να είναι τέτοια που, ακόμη και ύστερα από τέτοια μελέτη, η προτεινόμενη αποκατάσταση να είναι μόνο ποιοτική και ενδεικτική (έντονη πολυσημία).

Είναι κατανοητό ότι η ψηφιακή αποκατάσταση των τοιχογραφιών πιθανόν να απαιτεί διαφορετικού είδους τεχνικές, ανάλογα με τον βαθμό της αλλοίωσης σύμφωνα με την παραπάνω κατηγοριοποίηση. Για παράδειγμα, βλάβες Τύπου-I μπορούν να ανιχνευθούν και να αντιμετωπιστούν με σχετικά απλές μη-γραμμικές τεχνικές επεξεργασίας εικόνας, π.χ. [GNP06]. Βλάβες Τύπου-II απαιτούν την εφαρμογή αρκετά πιο σύνθετων σύγχρονων μεθόδων της Όρασης Υπολογιστών, γνωστών ως μεθόδων αυτόματης ψηφιακής ενδοσυμπλήρωσης, και βρίσκονται στο επίκεντρο της παρούσας μελέτης. Τέλος, βλάβες Τύπου-III είναι εξαιρετικά πιο δύσκολο να αντιμετωπιστούν με πλήρως αυτόματες ψηφιακές τεχνικές, καθότι απαιτούν την ενσωμάτωση λεπτομερούς ειδικευμένης γνώσης (expert domain knowledge) στο ψηφιακό σύστημα αποκατάστασης. Αν και δεν έχουμε κινηθεί προς αυτήν την κατεύθυνση στα πλαίσια της διατριβής, είναι ενδιαφέρουσα και θα μας απασχολήσει στο μέλλον η ανάπτυξη μεθόδων για επιβλεπόμενη ψηφιακή αποκατάσταση Τύπου-III βλαβών, όπου ψηφιακό σύστημα και ειδικός αλληλεπιδρούν για την αποκατάσταση της βλάβης.



Σχήμα 2.1: Το πρόβλημα της ενδοσυμπλήρωσης. Ω : κενό, Ω^c : γνωστή περιοχή, $\partial\Omega$: σύνορο περιοχών.

Ανεξάρτητα πάντως από τις έντονες διαφορές μεταξύ των παραπάνω κατηγοριών βλαβών, μπορεί κανείς να θεωρήσει ένα πιο αφηρημένο μαθηματικό πλαίσιο για την ενιαία θεωρητική αντιμετώπιση και φορμαλισμό του προβλήματος της αποκατάστασης των ελλειπόντων τμημάτων των τοιχογραφιών. Με αναφορά στο Σχ. 2.1, το πρόβλημα της *ευδοσυμπλήρωσης* (inpainting/infilling) συνίσταται στην εκτίμηση της εικόνας στο κενό Ω , με βάση τη γνώση των τιμών της στο συμπληρωματικό χωρίο Ω^c και το σύνορο $\partial\Omega$. Στα πλαίσια της ως τώρα εργασίας μας το χωρίο Ω παρέχεται από το χρήστη ως δυαδική μάσκα, δηλαδή θεωρούμε ότι έχουμε διαθέσιμη μια ανθρώπινη *επισημείωση αναφοράς* (reference annotation) που οριοθετεί τις ελλείπουσες περιοχές στις τοιχογραφίες μας, όπως φαίνεται στα παράδειγμα του Σχ. 2.2.



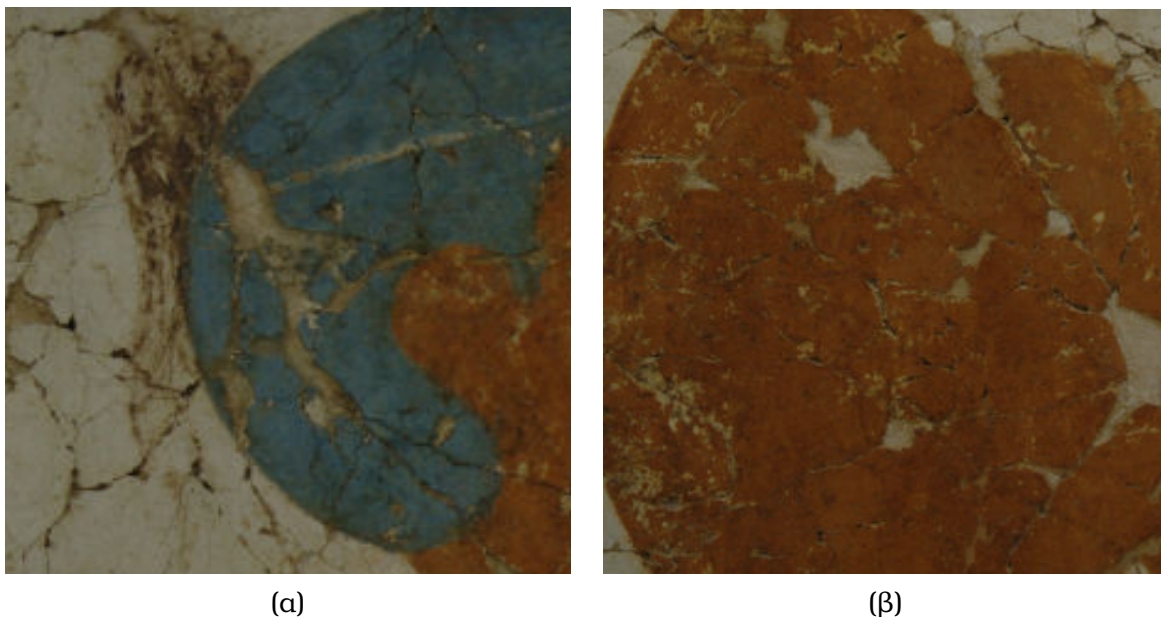
(α)

(β)

Σχήμα 2.2: Παράδειγμα ανθρώπινης επισημείωσης των ελλειπουσών περιοχών σε τοιχογραφίες από το Ακρωτήρι Θήρας. (α) Αρχική εικόνα της τοιχογραφίας. (β) Επισημείωση όπου ο χρήστης έχει υπερθέσει στην αρχική εικόνα τη μάσκα που αντιστοιχεί στις ελλείπουσες περιοχές.

Αντί για την ανθρώπινη επισημείωση, θα ήταν χρήσιμη η αυτόματη ανίχνευση των ελλειπουσών περιοχών στις ψηφιακές φωτογραφίες των τοιχογραφιών ώστε να επιτυγχάνεται η πλήρως αυτοματοποιημένη αποκατάστασή τους. Αυτό είναι ιδιαίτερα επιθυμητό δεδομένου του μεγάλου όγκου του φωτογραφικού υλικού που έχουμε στη διάθεσή μας, το οποίο καθιστά ιδιαίτερα χρονοβόρα την ανθρώπινη επισημείωσή τους. Δυστυχώς, η αυτόματη

ανίχνευση των κενών στις φωτογραφίες των τοιχογραφιών είναι δυσχερής. Η χρωματική ομοιότητα μεταξύ του κονιάματος που έχει τοποθετηθεί από τους συντηρητές στις ελλείπουσες περιοχές και του ασβεστοκονιάματος στις αρχικές τοιχογραφίες δυσκολεύει τη διάκριση των κενών με χρήση αποκλειστικά χρωματικής πληροφορίας. Η ποικιλομορφία στη γεωμετρία των ελλειπουσών περιοχών δυσχεραίνει περαιτέρω την αυτόματη ανίχνυσή τους. Από την άλλη πλευρά, υπάρχει μία ιδιαιτερότητα της διαδικασίας συντήρησης των τοιχογραφιών που θα μπορούσε πιθανώς να διευκολύνει την ανίχνευση των κενών. Συγκεκριμένα, κατά το τελικό στάδιο της συντήρησης της τοιχογραφίας όπου τα συνανήκοντα κομμάτια προσκολλώνται σε φορητό πίνακα, οι συντηρητές φροντίζουν ώστε τα ευρεθέντα σπαράγματα να προεξέχουν ελαφρά (μερικά χιλιοστά) του κονιάματος που τοποθετείται στις ελλείπουσες περιοχές [ΓΜΣ92]. Έτσι, χαρακτηριστικά ασυνέχειας βάθους εξαγμένα με τεχνικές 3-Δ υπολογιστικής όρασης θα μπορούσαν να συνδυαστούν με χαρακτηριστικά χρώματος και να διευκολύνουν την ανίχνευση των κενών. Τεχνικές στερεοσκοπικής όρασης θα μπορούσαν να χρησιμοποιηθούν για αυτόν το σκοπό [HZ03], αν και τεχνικές ανίχνευσης ασυνεχειών βάθους μέσω ελεγχόμενου φωτισμού όπως η [RTF⁺04] πιθανά να ταιριάζουν καλύτερα στο πρόβλημα¹ [Rus09]. Πάντως η εφαρμογή τέτοιων τεχνικών 3-Δ όρασης προϋποθέτει τη λήψη νέων φωτογραφικών δεδομένων υπό καλά ελεγχόμενες συνθήκες φωτισμού, οπότε η αυτοματοποιημένη ανίχνευση των κενών αφήνεται για μελλοντική εργασία.



Σχήμα 2.3: Παραδείγματα αλλοίωσης σε τμήματα της τοιχογραφίας *‘Ακολουθία Ανδρών’* από το Ακρωτήρι Θήρας. Λεπτομέρεια από το κεφάλι (α) και τον κορμό (β) του νεαρού άνδρα. Εκτός από τα ελλείποντα τμήματα παρατηρούμε αλλοιώσεις όπως ξεθώριασμα του χρώματος και μερική απόξεση της ζωγραφικής επιφάνειας.

Μπορεί κανείς να θεωρήσει διάφορες επεκτάσεις του προβλήματος, για παράδειγμα να υποθέσει ότι η γνώση μας στην περιοχή Ω^c περιορίζεται σε μια θορυβώδη εκδοχή της αρχικής εικόνας. Μια προσέγγιση για αυτήν την περίπτωση είναι να αποθορυβοποιήσει

¹Οι καθηγητές Szymon Rusinkiewicz και Tim Weyrich προσέφεραν χρήσιμες συμβουλές σχετικά με τη δυνατότητα ανάκτησης του τοιχογραφικού αναγλύφου με φωτομετρικές μεθόδους.

κάνεις την αρχική εικόνα σε ένα στάδιο προεπεξεργασίας με αυτοματοποιημένες μεθόδους αποθορυβοποίησης σαν κι αυτές που συζητάμε στο Κεφάλαιο 4. Μια άλλη προσέγγιση για το χειρισμό θορυβωδών παρατηρήσεων είναι να θεωρήσει κανείς σταθμισμένη μάσκα ενδοσυμπλήρωσης, όπου οι χωρικές παρατηρήσεις σταθμίζονται με βάρη αντιστρόφως ανάλογα της έντασης του θορύβου, όπως συζητάμε στο Κεφάλαιο 6. Βέβαια, όπως φαίνεται στα παραδείγματα του Σχ. 2.3, συχνά ο μηχανισμός αλλοίωσης της τοιχογραφίας περιλαμβάνει πολυπλοκότερες διαδικασίες όπως ξεθώριασμα του χρώματος ή μερική απόξεση της ζωγραφικής επιφάνειας. Θεωρώντας αυτές τις περιοχές ως ελλείπουσες και περιλαμβάνοντάς τες στη μάσκα ενδοσυμπλήρωσης Ω οδηγεί σε απώλεια πιθανά χρήσιμης πληροφορίας. Μοντελοποιώντας τέτοιου είδους αλλοιώσεις ως προσθήκη προσθετικού θορύβου αποτελεί επίσης φτωχή προσέγγιση του μηχανισμού αλλοίωσης. Ιδανικά θα επιθυμούσε κανείς να αναπτύξει εξειδικευμένα μοντέλα παρατήρησης που να ανταποκρίνονται στη φυσική διαδικασία φθοράς του τοιχογραφικού υποστρώματος και χημικής αλλοίωσης των χρωστικών ουσιών της τοιχογραφίας, αλλά οι σχετικές χημικές μελέτες [PKSC00] είναι δύσκολο να οδηγήσουν σε συγκεκριμένα υπολογιστικά μοντέλα χημικής αλλοίωσης των χρωμάτων της εικόνας.

2.2 Φωτογράφιση Τοιχογραφιών

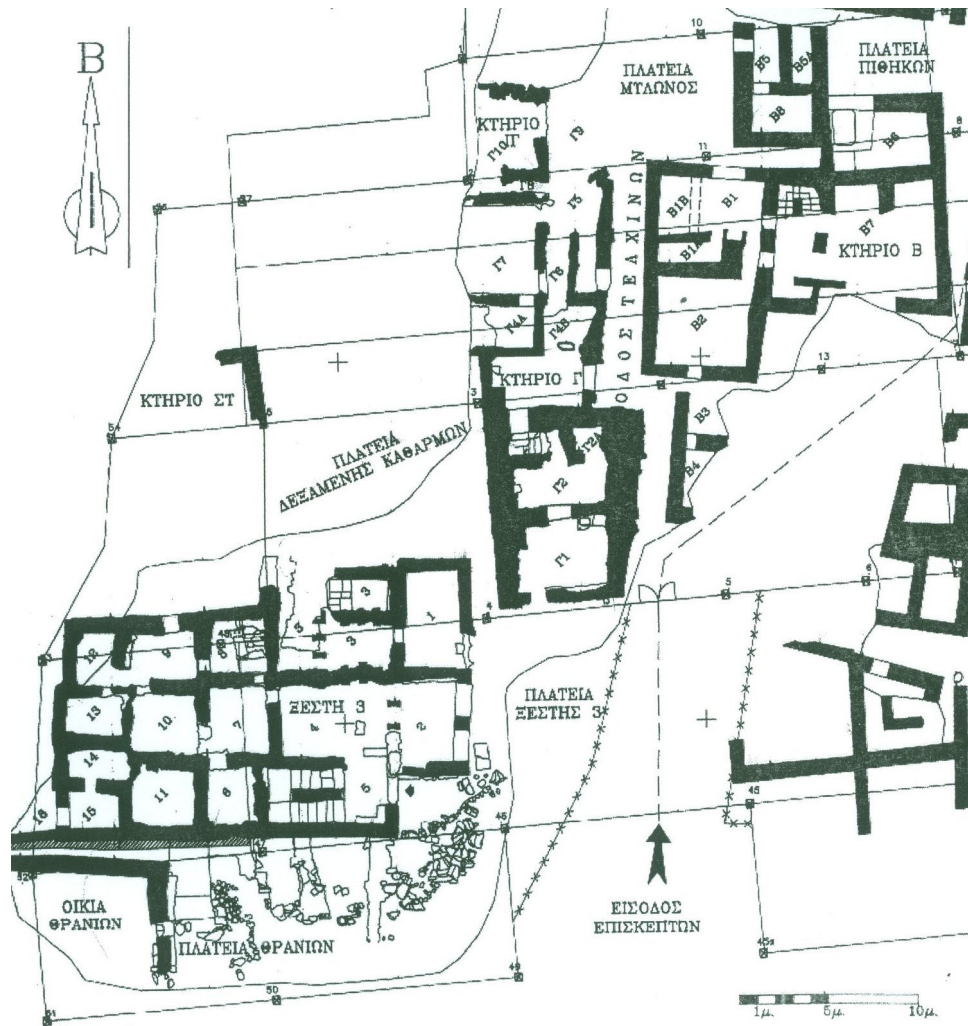
Απαραίτητη προϋπόθεση για την πρόοδο της διατριβής ήταν η ύπαρξη λεπτομερούς φωτογραφικού υλικού από επιλεγμένες τοιχογραφίες προερχόμενες από την ανασκαφή του Ακρωτηρίου Θήρας. Αν και προϋπήρχε εκτεταμένο φωτογραφικό υλικό στη διάθεση της αρχαιολογικής ομάδας του Ακρωτηρίου Θήρας, αυτό είχε ληφθεί για διαφορετικούς σκοπούς, με αποτέλεσμα να μην είναι άμεσα κατάλληλο για τους σκοπούς της ερευνητικής μας εργασίας. Πιο συγκεκριμένα, το υπάρχον φωτογραφικό υλικό

1. είτε είχε ληφθεί για συγκεκριμένες ανάγκες *αρχαιολογικής τεκμηρίωσης*, οπότε ήταν ανομοιογενές και κάλυπτε μέρος μόνο των τοιχογραφιών, σε διάφορα στάδια αποκατάστασής τους, σε διάφορες χρονικές περιόδους στα περίπου 40 χρόνια από την έναρξη της ανασκαφής ως σήμερα.
2. είτε είχε ληφθεί με σκοπό την *ευρύτερη προβολή/δημοσιοποίηση* των τοιχογραφιών, και χαρακτηριζόταν από περιορισμένη λεπτομέρεια ή/και ύστερη ψηφιακή παρέμβαση/αποκατάσταση των χρωμάτων των τοιχογραφιών.

Για αυτούς τους λόγους κρίθηκε απαραίτητη η λήψη νέου φωτογραφικού υλικού για τις ανάγκες της διδακτορικής διατριβής.

2.2.1 Αντικείμενο Φωτογράφισης

Για το σκοπό της φωτογράφισης των τοιχογραφιών πραγματοποιήσαμε τρεις επισκέψεις στο εργαστήριο τοιχογραφιών της ανασκαφής Ακρωτηρίου Θήρας (Σεπτέμβριος 2006, Μάιος 2007, Σεπτέμβριος 2008) και μια επίσκεψη στο Εθνικό Αρχαιολογικό Μουσείο της Αθήνας (ΕΑΜ) (Νοέμβριος 2006), όπου φωτογραφήσαμε ένα αντιπροσωπευτικό αριθμό συντηρημένων τοιχογραφιών προερχόμενων από το σύμπλεγμα δωματίων της Ξεστής 3 και του Κτιρίου Β. Στο Σχ. 2.5 δείχνουμε μια συνολική άποψη της



Σχήμα 2.4: Τμήμα από το τοπογραφικό διάγραμμα του αρχαιολογικού χώρου στο Ακρωτήρι Θήρας. Περιλαμβάνονται τα συμπλέγματα δωματίων της Ξεστής 3 και του Κτιρίου Β από τα οποία προέρχονται οι τοιχογραφίες που έχουμε φωτογραφήσει. Σχέδιο των Αξιώτη-Σάλη και Κριτωτάκη.

κάθε τοιχογραφίας. Επίσης, στον Πίν. 2.1 παραθέτουμε πλήρη κατάλογο με τις τοιχογραφίες που φωτογραφήσαμε, περιλαμβάνοντας πληροφορίες για τη θέση της σύνθεσης εντός του κτιρίου με αναφορά στο τοπογραφικό διάγραμμα 2.4, τις διαστάσεις του, καθώς και από μια αρχαιολογική παραπομπή όπου ο ενδιαφερόμενος αναγνώστης μπορεί να αναζητήσει περαιτέρω πληροφορίες για την κάθε τοιχογραφία. Αντίστοιχος κατάλογος με τις τοιχογραφίες που έχουν αποτυπωθεί βρίσκεται και στην ιστοσελίδα <http://cvsp.cs.ntua.gr/projects/AkrotiriThera/DataSets>. Οι φωτογραφίες αυτές έγιναν σε συνεργασία με την κ. Φραγκούλα Γεώργια, το Δρ. Ανδρέα Βλαχόπουλο, τον κ. Μανώλη Χαμαού (ανασκαφή Ακρωτηρίου Θήρας) και τον κ. Ματθαίο Αληφραγκή (ΕΜΠ). Η φωτογράφιση στο ΕΑΜ έγινε με την άδεια της κ. Λένιας Παπάζογλου, υπεύθυνης της προϊστορικής συλλογής του μουσείου.

2.2.2 Στόχοι και Διαδικασία Φωτογράφισης

Η φωτογράφιση των τοιχογραφιών έχει γίνει με τρόπο που να εξασφαλίζει την όσο το δυνατόν λεπτομερέστερη και πιστότερη ψηφιακή τους αποτύπωση. Αξίζει να σημειωθεί εδώ ότι η πιστή φωτογράφιση έργων τέχνης αποτελεί από μόνη της ξεχωριστή ερευνητική περιοχή, στην οποία έχουν αναπτυχθεί ιδιαίτερες τεχνικές και χρησιμοποιείται ειδικευμένος και πολύ ακριβός εξοπλισμός (π.χ.δες το [GBT⁺02]). Στην περίπτωση μας περιοριζόμαστε σε χρήση σχετικά φθηνού ημι-επαγγελματικού εξοπλισμού φωτογράφισης. Πιο συγκεκριμένα, σε όλες τις φωτογραφίες έχουμε χρησιμοποιήσει την ψηφιακή SLR μηχανή Nikon D70s με αισθητήρα 6 MPixels και φακό Nikon Nikkor μεταβλητής εστιακής απόστασης 18-70mm. Οι ψηφιακές φωτογραφίες έχουν αποθηκευθεί τόσο σε ασυμπίεστη μορφή RAW (γραμμική απόκριση με περιγραφή 12 bits/χρωματικό κανάλι) για βέλτιστη ποιότητα, όσο και σε συμπιεσμένη μορφή JPEG για ευχρηστία. Προσπαθούμε να υπερβούμε τους περιορισμούς του φωτογραφικού εξοπλισμού μας με λήψη πολλαπλών φωτογραφικών στιγμιότυπων από κάθε τοιχογραφία, τα οποία και συνενώνουμε σε μωσαϊκά υπερ-υψηλής ανάλυσης με τεχνικές υπολογιστικής φωτογραφίας που περιγράφουμε στην Ενότητα 2.3.

Όσο αφορά τον φωτισμό, επιλέξαμε να φωτογραφίσουμε τα θέματα με διάχυτο φως προερχόμενο από λαμπτήρες φθορισμού των αιθουσών έκθεσης των τοιχογραφιών. Τούτο έγινε επειδή δεν είχαμε στη διάθεσή μας ειδικά επαγγελματικά φωτιστικά στοιχεία, ενώ απλά φωτογραφικά φλας θα δημιουργούσαν έντονες ανεπιθύμητες ανακλάσεις. Για την ελαχιστοποίηση του θολώματος της εικόνας εξαιτίας των μακρών χρόνων έκθεσης, όλες οι φωτογραφίες έγιναν με χρήση τρίποδα. Για βέλτιστη χρωματική πιστότητα των γάμμα-διορθωμένων JPEG φωτογραφιών, σε κάθε χώρο έγινε *αντιστάθμιση λευκού* (white balancing) πριν τη λήψη των φωτογραφιών. Τέλος, κάθε πλάνο λήφθηκε σε τρία επίπεδα φωτεινότητας (κανονικό, υποφωτισμένο, υπερφωτισμένο), ώστε να είναι εφικτή η χρήση τεχνικών ψηφιακής φωτογραφίας *υψηλού δυναμικού πεδίου φωτεινότητας* (high dynamic range).

Καθότι δεν θα ήταν σε καμία περίπτωση εφικτό να αποτυπωθεί σε ένα μόνο φωτογραφικό στιγμιότυπο και σε επαρκή λεπτομέρεια η κάθε τοιχογραφία, ακολουθήσαμε μια πολυκλιμακωτή τακτική φωτογράφισης. Συγκεκριμένα, φωτογράψαμε τις τοιχογραφίες σε τρία επίπεδα:

1. Σε ένα πρώτο αδρό επίπεδο αποτυπώθηκε ολόκληρη η σύνθεση ώστε να γίνεται άμεσα αντιληπτό το βασικό της μοτίβο (1 πλάνο/τοιχογραφία). Η χωρική ανάλυση που



(α) Κροκοσυλλέκτρια & Πότνια



(β) Ακολουθία Ανδρών



(γ) Ακολουθία Γυναικών



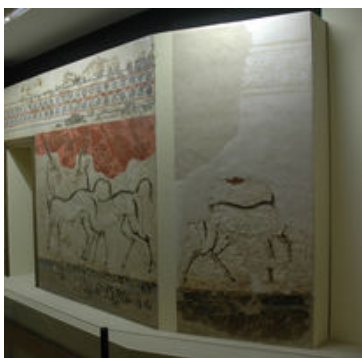
(δ) Καλαμιώνας



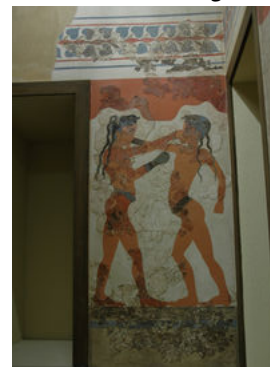
(ε) Σπείρες



(στ) Αντιλόπες-Α



(ζ) Αντιλόπες-Δ και Β



(η) Πυγμαχία

Σχήμα 2.5: Γενική άποψη των φωτογραφημένων τοιχογραφιών, από την Ξεστή 3 (α-ε) και από το Κτίριο Β (στ-η). Οι τοιχογραφίες (α-στ) βρίσκονται στο εργαστήριο τοιχογραφιών στο Ακρωτήρι Θήρας και οι (ζ-η) στο Εθνικό Αρχαιολογικό Μουσείο. Περισσότερα στοιχεία για κάθε τοιχογραφία περιλαμβάνονται στον Πίν. 2.1.

Πίνακας 2.1: Λίστα με στοιχεία των φωτογραφημένων τοιχογραφιών προερχόμενων από την ανασκαφή Ακρωτηρίου Θήρας. Βρίσκονται στο εργαστήριο τοιχογραφιών της ανασκαφής, εκτός από αυτές που στη στήλη Κωδικός υποδηλώνεται ΕΑΜ, οι οποίες εκτίθενται στο Εθνικό Αρχαιολογικό Μουσείο της Αθήνας.

Όνομα	Φωτο Σχ. 2.5	Κωδ.	Θέση	Διαστάσεις (Υ./Π.)(cm)	Παραπομπή	Χρόνος Φωτογρά- φησης
Σύμπλεγμα δωματίων Ξεστής 3						
Κροκοσυλ- λέκτρια & Πότνια	(α)	2038	Όροφος 1, Δωμ. 3α, Β. τοίχος	230/322	[Dou92], Σχ.122- 128, σελ. 158- 165	Μάιος-07
Ακολουθία Ανδρών	(β)	2041	Ισόγειο, Δωμ. 3β	192/193	[Dou92], Σχ.109, 112, 113, σελ. 146-149	Σεπ.-06
Ακολουθία Γυναικών	(γ)	2764	Όροφος 1, Δωμ. 3β, Ν. διάδρο- μος	250/220	[Vla08], Σχ.34, 36, σελ. 461	Σεπ.-06
Καλαμιώνας	(δ)	2046	Όροφος 1, Δωμ. 3β, Δ. τοίχος	248-275	[Vla08], Σχ.22, σελ. 459	Σεπ.-06
Σπείρες	(ε)	-	Όροφος 2, Ν. τοίχος	305/355	[Vla08], Σχ.47, σελ. 464	Σεπ.-08
Σύμπλεγμα δωματίων Κτιρίου Β						
Αντιλόπες-Α	(στ)	2807	Όροφος 1, Δωμ. Β1, Α. τοίχος	275/180	[Γεώ09]	Σεπ.-08
Αντιλόπες-Δ	(ζ)	ΕΑΜ	Όροφος 1, Δωμ. Β1, Δ. τοίχος	275/200	[Dou92], Σχ.83, 84, σελ. 117-119	Νοέμ.-06
Αντιλόπες-Β	(ζ)	ΕΑΜ	Όροφος 1, Δωμ. Β1, Β. τοίχος	141/115	[Dou92], Σχ.82, σελ. 116	Νοέμ.-06
Πυγμαχοί	(η)	ΕΑΜ	Όροφος 1, Δωμ. Β1, Ν. τοίχος	275/94	[Dou92], Σχ.78- 81, σελ. 111-115	Νοέμ.-06

πετυχαίνουμε με αυτόν τον τρόπο είναι της τάξης των 30 σημείων/ίντσα (dots/inch (dpi)).

2. Σε ένα ενδιάμεσο επίπεδο η κάθε σύνθεση χωρίστηκε σε ένα ιδεατό κάρναβο μέσης ανάλυσης, κάθε στοιχείο του οποίου κάλυπτε χονδρικά ένα μέρος της σύνθεσης και ελήφθησαν αντίστοιχες φωτογραφίες (περίπου 10 πλάνα/τοιχογραφία) με επαρκείς αλληλεπικαλύψεις μεταξύ των πλάνων. Ανάλυση της τάξης των 70 dpi.
3. Τέλος, στο πιο λεπτομερές επίπεδο η κάθε σύνθεση χωρίστηκε σε ένα ιδεατό κάρναβο λεπτομερούς ανάλυσης, ώστε να καταγραφεί όσο το δυνατόν με μεγαλύτερη πιστότητα η μικροδομή κάθε τοιχογραφίας (περίπου 30-40 πλάνα με κάποιο βαθμό αλληλεπικάλυψης ανά τοιχογραφία). Με αυτόν τον τρόπο αυξάνουμε περαιτέρω τη χωρική ανάλυση στα 150 dpi.

Κατά τη λήψη φωτογραφιών στο ενδιάμεσο και στο λεπτομερές επίπεδο ανάλυσης μετακινούσαμε τη φωτογραφική μηχανή (κατά μήκος και ύψος) ώστε η φωτογράφιση κάθε τμήματος να γίνεται με τον οπτικό άξονα της μηχανής να προβάλλεται υπό σχεδόν ορθή γωνία στην τοιχογραφία και ρυθμίζαμε την εστιακή απόσταση ώστε το πλάνο να καλύπτει το επιθυμητό τμήμα της παράστασης. Παράδειγμα αυτής της πολυκλιμακωτής τεχνικής φωτογράφισης (στιγμιότυπα από την τοιχογραφία *‘Κροκοσυλλήκτρια & Πότινια’*, συλλογή Ακρωτηρίου Θήρας) φαίνεται στο Σχ. 2.6. Το σύνολο των ληφθέντων φωτογραφιών (γύρω στις 700) έχει αποθηκευθεί σε οπτικό δίσκο (DVD) και είναι διαθέσιμο σε ερευνητές που συνεργάζονται με την Ανασκαφή. Κάθε ψηφιακό αρχείο φωτογραφίας συνοδεύεται από επιπλέον πληροφορίες (EXIF μεταδεδομένα) που αφορούν τη λήψη της, όπως ακριβή ημερομηνία και ώρα λήψης, τεχνικές πληροφορίες για τη λήψη (για παράδειγμα, εστιακή απόσταση, άνοιγμα διαφράγματος, ταχύτητα κλείστρου, ευαισθησία αισθητήρα).

2.3 Σύνθεση Τοιχογραφικών Μωσαϊκών Υπερ-Υψηλής Ανάλυσης

Η πολυκλιμακωτή τεχνική φωτογράφισης που μόλις περιγράψαμε μας εξασφαλίζει ότι αφενός ολόκληρη η τοιχογραφική σύνθεση περιλαμβάνεται σε μια φωτογραφία αδρής κλίμακας και αφετέρου κάθε σημείο της τοιχογραφίας έχει αποτυπωθεί σε υψηλή ανάλυση σε τουλάχιστον μία από τις φωτογραφίες λεπτομερούς κλίμακας που περιλαμβάνει τον περίγυρό του. Έχοντας στη διάθεσή μας το σύνολο αλληλεπικαλυπτόμενων φωτογραφικών λήψεων από κάθε τοιχογραφία, προχωρήσαμε στην δημιουργία ενός ενιαίου ψηφιακού τοιχογραφικού μωσαϊκού ανά τοιχογραφία, που καλύπτει ολόκληρη την έκταση της σύνθεσης και περιγράφει κάθε σημείο της στη μέγιστη χωρική ανάλυση που περιέχεται στις φωτογραφίες λεπτομερούς κλίμακας.

Ίσως η πιο γνωστή εφαρμογή όπου συνδυάζει κανείς πολλαπλές επιμέρους φωτογραφίες για τη δημιουργία μιας συνθετικής εικόνας είναι η *πανοραμική φωτογραφία*. Τυπικό παράδειγμα πανοραμικής φωτογράφισης είναι η δημιουργία κυλινδρικού πανοράματος, Σχ. 2.7(α,γ). Για αυτό λαμβάνουμε πολλαπλές φωτογραφίες μιας σκηνής περιστρέφοντας σταδιακά την κάμερα γύρω από άξονα που διέρχεται από το οπτικό της κέντρο, με αποτέλεσμα η συνθετική εικόνα να διαθέτει οριζόντιο εύρος οπτικού πεδίου 360°. Μια άλλη έκφανση της πανοραμικής φωτογραφίας είναι το γραμμικό πανόγραμμα, Σχ. 2.7(β,δ).

(α) Αδρό επίπεδο



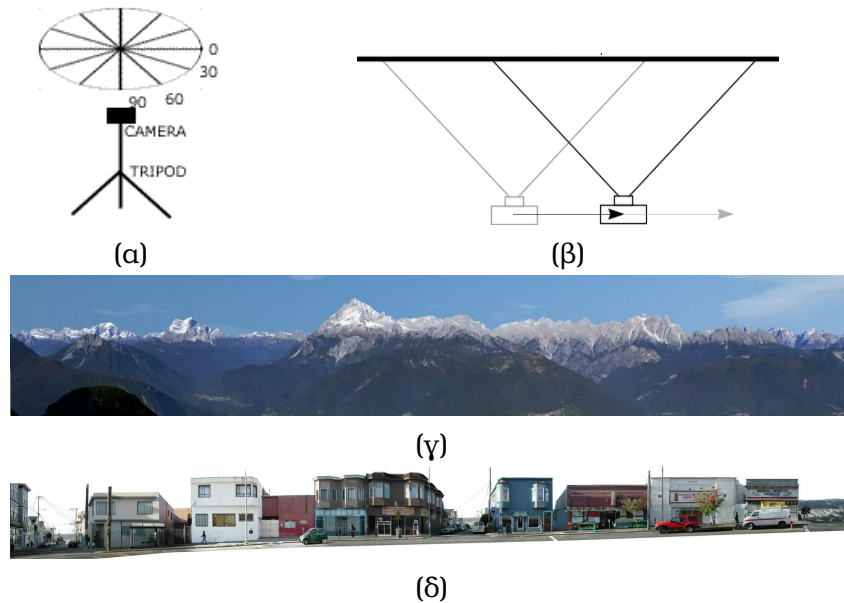
(β) Μέσο επίπεδο



(γ) Λεπτομερές επίπεδο



Σχήμα 2.6: Παράδειγμα πολυκλιμακωτής φωτογράφισης (τοιχογραφία 'Κροκοσυληέκτρια & Πότνια', συλλογή Ακρωτηρίου Θήρας): η λεπτομέρεια αυξάνει από πάνω (αδρό επίπεδο) προς τα κάτω (λεπτομερές επίπεδο). Το γαλάζιο περίγραμμα δηλώνει το τμήμα της τοιχογραφίας στο οποίο επικεντρωνόμαστε στο αμέσως πιο λεπτομερές πλάνο.

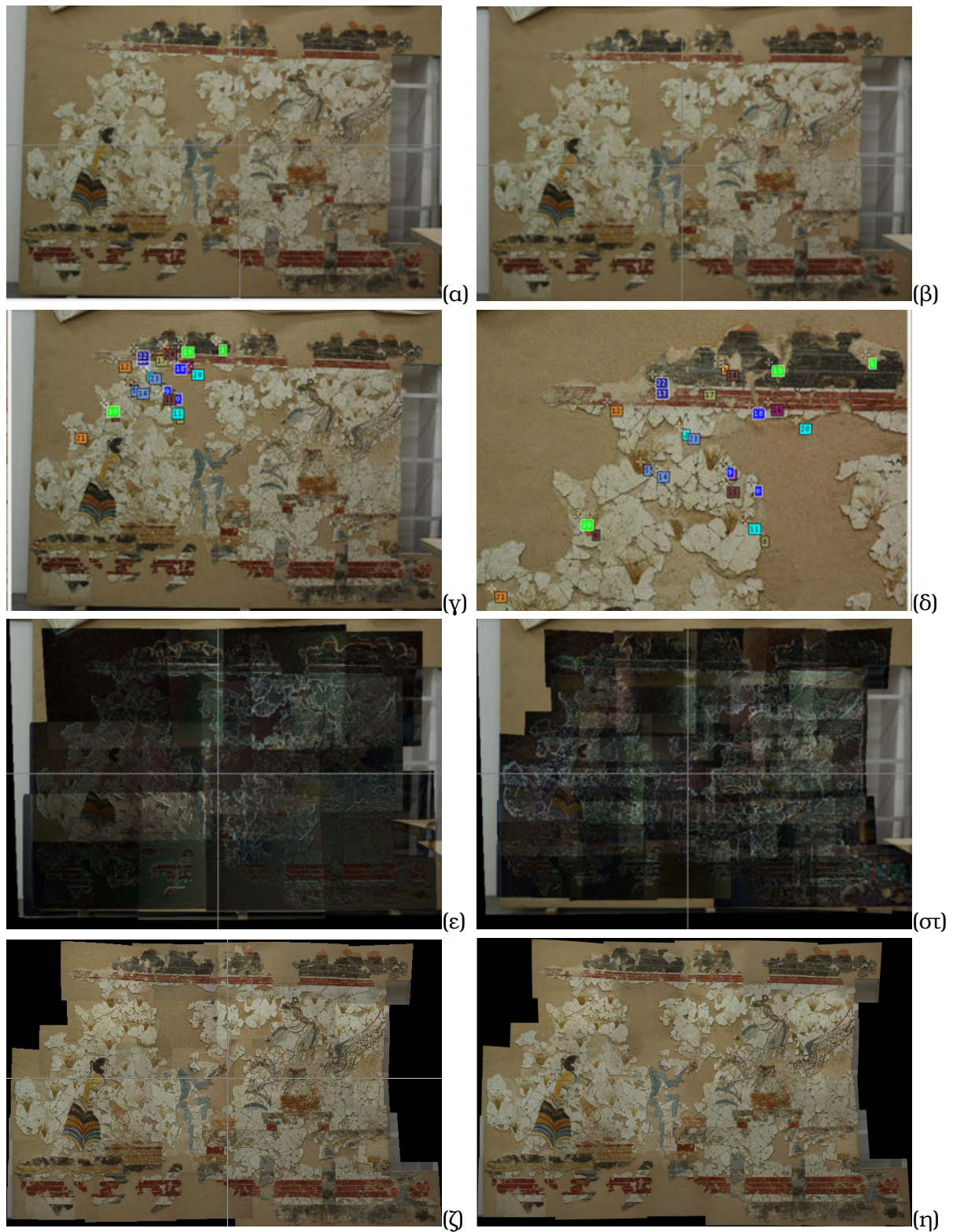


Σχήμα 2.7: Παραδείγματα πανοραμικής φωτογράφισης. Διάταξη φωτογράφισης και αποτέλεσμα σύνθεσης για κυλινδρικό (α,γ) και γραμμικό πανόραμα (β,δ). Το (δ) απεικονίζει τα κτίρια σε μια οδό του San Fransisco και είναι μέρος του πονήματος Seamless City του καλλιτέχνη Michael Koller.

Σε αντιδιαστολή με το κυλινδρικό πανόραμα, στο γραμμικό πανόραμα η κάμερα κινείται μεταφορικά με τον οπτικό της άξονα να διατηρεί σταθερή διεύθυνση. Η συνθετική εικόνα μπορεί να καλύπτει σκηνές πολύ μεγάλου μήκους. Από τις πιο εντυπωσιακές εφαρμογές της τεχνικής γραμμικού πανοράματος είναι το *πόνημα* (project) Seamless City του καλλιτέχνη Michael Koller, που σκοπό έχει να αποτυπώσει της προσόψεις όλων των κτιρίων του San Fransisco σε ένα γιγαντιαίο πανόραμα, <http://www.seamlesscity.com>. Κατά την φωτογράφιση των τοιχογραφιών, η γεωμετρική διάταξη υπό την οποία έχουμε λάβει τα πολλαπλά στιγμιότυπα κάθε σύνθεσης είναι παρόμοια με αυτήν του γραμμικού πανοράματος, με τη διαφοροποίηση ότι επιτρέπουμε και κατακόρυφες μετατοπίσεις της κάμερας και επίσης μεταξύ των επιμέρους φωτογραφιών επιτρέπουμε αλλαγές στην απόσταση κάμερας-σκηνής, στα γεωμετρικά και στα φωτομετρικά χαρακτηριστικά της έκθεσης.

Αν και πανοραμικές τεχνικές με χρήση ειδικών φωτογραφικών μηχανών ή τεχνικών επεξεργασίας φιλμ είναι γνωστές σχεδόν από τις απαρχές της φωτογραφίας το 19ο αιώνα, η δημιουργία φωτογραφικών μωσαϊκών έχει φτάσει σε νέα επίπεδα ευχρηστίας και ποιότητας τα τελευταία μόνο χρόνια. Η καταγραφή ψηφιακών φωτογραφικών δεδομένων και η δυνατότητα ψηφιακής επεξεργασίας τους επιτρέπουν την υπολογιστική σύνθεση πανοραμικών φωτογραφιών υψηλής ποιότητας, όπως αναλύεται διεξοδικά στο [Sze06]. Σε ένα ευρύτερο πλαίσιο, υπολογιστικές τεχνικές έχουν παρεισφρήσει σε όλα τα στάδια της φωτογράφισης. Ελέγχοντας ψηφιακά το φωτισμό, τα οπτικά χαρακτηριστικά της κάμερας και την επεξεργασία της εικόνας μπορεί κανείς να πετύχει εντυπωσιακά φωτογραφικά αποτελέσματα. Τέτοιες μέθοδοι συνιστούν την αναδυόμενη περιοχή της *υπολογιστικής φωτογραφίας* (computational photography) [DS07].

Στο συγκεκριμένο πλαίσιο της εφαρμογής μας, προκειμένου να συρράψουμε ψηφιακά

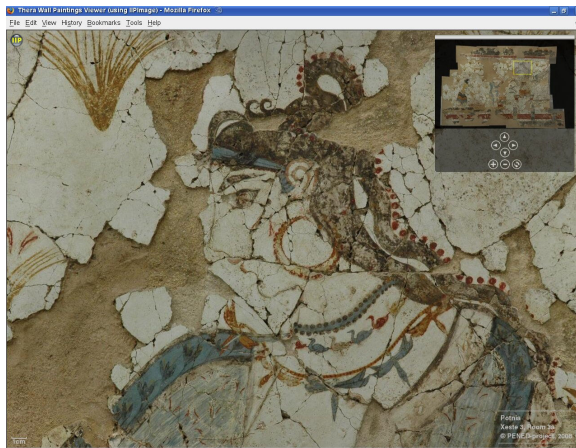


Σχήμα 2.8: Κύρια βήματα κατά τη σύνθεση φωτογραφικών μωσαϊκών υπερ-υψηλής ανάλυσης. (α & β) Αρχική αδρή φωτογραφία και ανορθωμένη εκδοχή της. (γ & δ) Αντίστοιχα σημεία-κλειδιά σε φωτογραφίες αδρής και μέσης κλίμακας αυτόματα εντοπισμένα ύστερα από ταίριασμα SIFT περιγραφέων. (ε & στ) Φωτογραφίες μέσης και λεπτομερούς κλίμακας αυτόματα τοποθετημένες επί της αρχικής αδρής φωτογραφίας. (ζ & η) Μωσαϊκό υψηλής ανάλυσης πριν και μετά τη διαδικασία ομαλής μείξης.

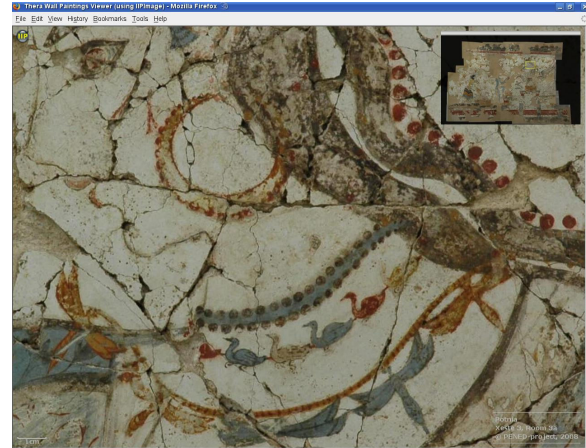
το σύνολο των φωτογραφικών στιγμιοτύπων κάθε τοιχογραφίας σε ένα γιγαντιαίο ψηφιακό μωσαϊκό υπερ-υψηλής ανάλυσης ανά τοιχογραφία, ακολουθήσαμε μια σειρά από βήματα που περιγράφουμε στη συνέχεια αναφερόμενοι στο Σχ. 2.8:

1. Διόρθωση οπτικών ατελειών του συγκεκριμένου φακού που χρησιμοποιούμε με βάση κατάλληλο παραμετρικό μοντέλο της οπτικής παραμόρφωσης που εισάγει (οι παράμετροι που χρησιμοποιήσαμε προέρχονται από τη βάση βαθμονόμησης *lensfun*). Κατόπιν ανόρθωση φωτογραφίας αδρής κλίμακας με διόρθωση της προοπτικής παραμόρφωσης, καθορίζοντας μια οριζόντια και μια κατακόρυφη γραμμή στη φωτογραφία.
2. Εύρεση αντίστοιχων σημείων-κλειδιών μεταξύ των στιγμιοτύπων, που επιτρέπουν τη συστοίχιση όλων των στιγμιοτύπων σε ένα ενιαίο σύνολο. Η διαδικασία αυτή που παραδοσιακά ήταν ιδιαίτερα χρονοβόρα γιατί απαιτούσε εμπλοκή του χρήστη έχει πλέον σχεδόν αυτοματοποιηθεί. Με χρήση μοντέρνων τεχνικών της όρασης υπολογιστών είναι πλέον δυνατόν να εντοπίζονται σημεία ενδιαφέροντος της εικόνας και να περιγράφονται με χρήση χαμηλοδιάστατων διανυσμάτων χαρακτηριστικών. Κατόπιν είναι εφικτή η αξιόπιστη αναζήτηση αντίστοιχων σημείων σε δύο εικόνες με εύρεση γειτόνων στο χώρο των χαρακτηριστικών. Στην προκειμένη περίπτωση κάνουμε χρήση των σημείων ενδιαφέροντος τύπου *κηλίδας* (blob) του [Lin98] και του SIFT περιγραφητή [Low04], συνδυασμός που έχει αποδειχθεί ιδιαίτερα αποδοτικός στην πανοραμική φωτογράφιση [BL07].
3. Βελτιστοποίηση των ελεύθερων παραμέτρων του συστήματος, όπως θέση και γωνίες σκόπευσης της κάμερας κατά τη λήψη κάθε φωτογραφικού στιγμιοτύπου ώστε να ευθυγραμμίζονται όσο το δυνατόν καλύτερα τα σημεία ενδιαφέροντος που έχουν ήδη αντιστοιχηθεί. Στην πράξη εμφανίζονται κάποιες φορές δυσκολίες εξαιτίας είτε λόγω *έκτοπων* (outlier) αντιστοιχίσεων, είτε λόγω απουσίας επαρκούς αριθμού αντιστοιχίσεων μεταξύ ελαφρά αλληλεπικαλυπτόμενων πλάνων, είτε λόγω έναρξης της διαδικασίας επαναληπτικής βελτιστοποίησης από κακή αρχική συνθήκη. Έχουμε βρει ότι βελτιστοποιώντας τις παραμέτρους του μοντέλου αντιστοίχισης περιλαμβάνοντας πρώτα τις λήψεις μέσης κλίμακας και ύστερα προσθέτοντας τις λεπτομερείς λήψεις καθιστά τη διαδικασία πολύ πιο εύρωστη. Επίσης, η ύπαρξη αδρών λήψεων αποτρέπει τη σταδιακή συσσώρευση μικρών σφαλμάτων αντιστοίχισης που συναντάται συχνά όταν χρησιμοποιούνται αποκλειστικά λεπτομερείς λήψεις.
4. Ρύθμιση φωτομετρικών παραμέτρων ώστε να αντισταθμίζονται πιθανές διαφορές στο χρόνο έκθεσης ή/και στο άνοιγμα διαφράγματος μεταξύ γειτονικών λήψεων.
5. Ακολουθεί τέλος η συγκόλληση των αντιστοιχισμένων φωτογραφιών ώστε να δημιουργηθεί ένα ενιαίο μωσαϊκό δίχως εμφανείς ραφές. Αυτό επιτυγχάνεται με σταδιακή ανάμειξη γειτονικών εικόνων στο πεδίο Λαπλασιανών πυραμίδων [BA83b].

Η όλη διαδικασία διενεργήθηκε με χρήση ελεύθερου λογισμικού, όπως το Hugin, το Enblend, και το Panorama Tools. Το τελικό αποτέλεσμα είναι ένα ενιαίο μωσαϊκό για κάθε τοιχογραφία, με επιφάνεια της τάξης των 250 MPixels. Η χωρική ανάλυση που πετυχαίνουμε είναι της τάξης των 150 dpi, που μας επιτρέπει να εστιάσουμε σε μικρές λεπτομέρειες επί των τοιχογραφιών. Το μωσαϊκό αποθηκεύεται σε TIFF αρχείο με μέγεθος της τάξης των 500 MB.



(α)



(β)

Σχήμα 2.9: Παράδειγμα πλοήγησης εντός της τοιχογραφίας 'Κροκοσυλλέκτρια & Πότνια'.

Το ιδιαίτερα μεγάλο μέγεθος αρχείου δυσχεραίνει τη διαχείριση, το μοίρασμα, και την πλοήγηση εντός του προκύπτοντος τοιχογραφικού μωσαϊκού. Προκειμένου να διευκολύνουμε την πρόσβαση των συνεργαζόμενων αρχαιολόγων και μελετητών των τοιχογραφιών σε αυτόν τον χρήσιμο πόρο, έχουμε αναπτύξει στον ιστότοπο του εργαστηρίου μια εφαρμογή πελάτη-εξυπηρετητή βασισμένη στο ελεύθερο λογισμικό IPImage που διευκολύνει την πλοήγηση μέσω διαδικτύου σε τέτοιες εικόνες υπερ-υψηλής ανάλυσης. Ένα παράδειγμα πλοήγησης εντός της τοιχογραφίας 'Κροκοσυλλέκτρια & Πότνια' όπου ο χρήστης επικεντρώνεται στη μελέτη του περιδέραιου της γυναικείας θεϊκής μορφής απεικονίζεται στο Σχ. 2.9. Ο αναγνώστης ενθαρρύνεται να επισκεφθεί την ιστοσελίδα <http://cvsp.cs.ntua.gr/projects/AkrotiriThera/DataSets> όπου μπορεί να περιηγηθεί σε μωσαϊκά υπερ-υψηλής ανάλυσης κάποιων από τις τοιχογραφίες που έχουμε αποτυπώσει και να αποκτήσει αίσθηση της πιστότητας με την οποία έχουν αποτυπωθεί οι συνθέσεις (username: pened-thera, password: pe20ned03).

Η προσέγγιση που έχουμε ακολουθήσει στην εργασία μας για αποτύπωση έργων τέχνης σε υπερ-υψηλή ανάλυση με χρήση απλού φωτογραφικού εξοπλισμού και ύστερη συρραφή των στιγμιότυπων σε ένα ενιαίο μωσαϊκό φαίνεται να κερδίζει ευρύτερα έδαφος ως εναλλακτική μεθόδων που απαιτούν ειδικευμένο εξοπλισμό υψηλού κόστους. Είναι αξιοσημείωτο πως πριν από μερικούς μήνες (Ιανουάριος 2009) ανακοινώθηκε από τη Google ότι αποτυπώθηκαν με παρόμοιες τεχνικές κάποιον από τους γνωστότερους πίνακες ζωγραφικής που εκτίθενται στο Μουσείο Prado της Μαδρίτης, <http://www.google.com/intl/en/landing/prado>. τα προκύπτοντα μωσαϊκά είναι προσβάσιμα μέσω της υπηρεσίας Google Earth. Περαιτέρω, η ευκολία με την οποία μπορεί κανείς να δημιουργήσει ψηφιακά πανοραμικές εικόνες άριστης ποιότητας έχει οδηγήσει στη δημιουργία μιας δυναμικής κοινότητας φωτογράφων που μοιράζονται διαδικτυακά τις πανοραμικές φωτογραφίες που δημιουργούν, <http://gigapan.org>. Η κοινότητα αυτή αποτελεί τον πυρήνα μιας αναδυόμενης αγοράς εξειδικευμένων προϊόντων υλικού/λογισμικού χαμηλού κόστους ειδικά σχεδιασμένων ώστε να διευκολύνουν τη διαδικασία λήψης και συρραφής φωτογραφικών πανοραμάτων.

Κεφάλαιο 3

Πολυπλεγματικοί Αλγόριθμοι για Κατάτμηση και Ενδοσυμπλήρωση με Μοντέλα Μερικών Διαφορικών Εξισώσεων

Οι *Μερικές Διαφορικές εξισώσεις - ΜΔΕ* (Partial Differential Equations - PDE) επινοήθηκαν αρχικά για να περιγράψουν θεμελιώδη φυσικά φαινόμενα και αποτελούν περιοχή στενής επαφής των μαθηματικών με τη φυσική [CH53]. Το γεγονός ότι θεμελιώδεις ΜΔΕ περιγράφουν φυσικά φαινόμενα σημαίνει ότι μπορεί κανείς να αποκτήσει φυσική διαίσθηση για τη συμπεριφορά της λύσης τους, καθοδηγώντας έτσι την ανάπτυξη μοντέλων ΜΔΕ με επιθυμητές ιδιότητες για την περιγραφή και μη-φυσικών διεργασιών. Μία ακόμη εξήγηση της ευρείας χρήσης των ΜΔΕ έγκειται στο γεγονός ότι ανακύπτουν κατά την ελαχιστοποίηση γενικευμένων συναρτησιακών ενέργειας με τεχνικές *μεταβολικού λογισμού* (variational calculus) [Wei74].

Η χρήση μοντέλων ΜΔΕ και τεχνικών μεταβολικού λογισμού στην όραση υπολογιστών είναι ευρεία. Ίσως η πρώτη πετυχημένη εφαρμογή ΜΔΕ στην όραση υπολογιστών, ήδη από τη δεκαετία του '70, είναι η μέθοδος του Horn για εκτίμηση του σχήματος από τη σκίαση [HB89], η οποία μάλιστα βασίζεται σε ένα απλοποιημένο φυσικό μοντέλο της αλληλεπίδρασης φωτισμού και επιφάνειας αντικειμένων. Μεταβολικές τεχνικές εισήχθησαν στην όραση υπολογιστών κατά τη δεκαετία του '80. Παραδείγματα είναι το μοντέλο των Horn και Schunck για υπολογισμό της οπτικής ροής [HS81], η σύνδεση του γκαουσιανού χώρου κλίμακας με την εξίσωση διάχυσης από τον Koenderink [Koe84], το μοντέλο των Mumford και Shah για συνδυασμένη κατάτμηση/απλοποίηση εικόνων [MS89b], καθώς και μια σειρά μεθόδων για ανακατασκευή 3-Δ επιφανειών από οπτική πληροφορία [Gri83, PTK85, Ter88, BZ87].

Αποφασιστική για την περαιτέρω καθιέρωση των τεχνικών ΜΔΕ και μεταβολικού λογισμού στην όραση υπολογιστών ήταν η μεγάλη εξάπλωση ορισμένων ιδιαίτερα σημαντικών μεθόδων εκφρασμένων σε φορμαλισμό ΜΔΕ, που είναι αφενός οι *χώροι κλίμακας* (scale-spaces) και αφετέρου τα *επιπεδοσύνολα* (level-sets) και οι μη-γραμμικές γεωμετρικές ΜΔΕ.

Ο χώρος κλίμακας είναι μια μονοπαραμετρική οικογένεια που περιλαμβάνει μια αρχική εικόνα και διαδοχικά απλοποιημένες εκδοχές της, που αντιστοιχούν σε μια αυξανό-

μενη συνεχή παράμετρο κλίμακας. Η ενσωμάτωση ιδεών από χώρους κλίμακας επιτρέπει τη σχεδίαση συστημάτων όρασης υπολογιστών που διέπονται από την αρχή του *αναβλησιμότητας κλίμακας* (scale-invariance) και για αυτό αποτελεί συστατικό στοιχείο σύγχρονων τεχνικών ανίχνευσης αντικειμένων [Lin98, Low04]. Οι χώροι κλίμακας αποτελούν επίσης το φυσικό πλαίσιο για την ιεραρχική επεξεργασία εικόνων, αφού σε αδρές κλίμακες μόνο τα πιο κυρίαρχα αντικείμενα παραμένουν εμφανή. Αν και ο χώρος κλίμακας όπως πρωτοεισήχθηκε από τον Witkin παράγεται με βαθυπερατό φιλτράρισμα της εικόνας με ισοτροπικούς πυρήνες Gauss αυξανόμενου χωρικού εύρους [Wit83], γρήγορα έγινε φανερό ότι ισοδύναμα ο ίδιος χώρος κλίμακας μπορεί να θεωρηθεί ως η εξέλιξη της λύσης της ισοτροπικής και ομογενούς ΜΔΕ διάχυσης θερμότητας [Koe84]. Σημαντικές επακόλουθες επεκτάσεις ήταν οι χώροι κλίμακας παραγόμενοι από μη-γραμμικές [PM90, CLMC92, ALM92, BM94] και ανισοτροπικές [Wei98] ΜΔΕ διάχυσης. Περισσότερες λεπτομέρειες για μη-γραμμικές και ανισοτροπικές ΜΔΕ διάχυσης θα εξεταστούν στην Ενότητα 3.2.

Τα επιπεδοσύνολα [OS88] συνιστούν έναν έμμεσο τρόπο αναπαράστασης περιγραμμάτων, ως σύνολο τομής επιφάνειας και επιπέδου. Κύρια τους πλεονεκτήματα είναι η ευκολία χειρισμού τοπολογικών αλλαγών του περιγράμματος και ο γεωμετρικός τους χαρακτήρας, με την έννοια ότι υπερβαίνουν την εξάρτηση από την παραμετροποίηση της καμπύλης που είναι εγγενής σε προγενέστερα παραμετρικά μοντέλα καμπυλών. Για αυτούς τους λόγους τα επιπεδοσύνολα χρησιμοποιούνται ως αναπαράσταση στα μοντέλα γεωμετρικών ενεργών περιγραμμάτων, μιας ιδιαίτερα σημαντικής κατηγορίας τεχνικών για κατάτμηση εικόνων που παρουσιάζουμε αναλυτικά στην Ενότητα 3.1. Η έννοια του επιπεδοσυνόλου είναι πάντως προγενέστερη της χρήσης της στο [OS88], παίζοντας σημαντικό ρόλο στη μαθηματική μορφολογία ήδη από τη δεκαετία του '70 [Ser82]. Συγκεκριμένα, η έννοια του επιπεδοσυνόλου είναι θεμελιώδης κατά τη μετάβαση από δυαδικούς (δηλ. που δέχονται ως είσοδο σύνολα) σε γκριζούς (δηλ. που δέχονται ως είσοδο συνεχείς συναρτήσεις) επίπεδους μορφολογικούς τελεστές [Mar02]. Σε αυτό το πλαίσιο ιδεών που αναλύεται διεξοδικά στο [Mar05] εντάσσεται και η περιγραφή βασικών γεωμετρικών μορφολογικών πράξεων όπως είναι η συστολή/διαστολή 2-Δ σχημάτων με χρήση μη-γραμμικών ΜΔΕ που επιτρέπουν χρήση συνεχούς κλίμακας [BM94], καθώς και ο *μετασχηματισμός απόστασης* (distance transform).

Η ερευνητική μας συνεισφορά στις περιοχές αυτές, όπως θα παρουσιαστεί στην Ενότητα 3.3, έγκειται στην ανάπτυξη νέων *πολυπλεγματοειδών* (multigrid) αλγορίθμων για την ταχεία αριθμητική επίλυση τόσο μοντέλων γεωμετρικών ενεργών περιγραμμάτων, με εφαρμογή στην κατάτμηση εικόνων, όσο και μοντέλων ανισοτροπικών μη-γραμμικών ΜΔΕ, με εφαρμογή στην ενδοσυμπλήρωση εικόνων. Σχετικές δημοσιεύσεις μας είναι οι [PM04, PM07].

3.1 Κατάτμηση με Μερικές Διαφορικές Εξισώσεις: Γεωμετρικά Μοντέλα Ενεργών Περιγραμμάτων

Τα *ενεργά περιγράμματα* (active contours/snakes) είναι από τα πιο σημαντικά εργαλεία της όρασης υπολογιστών. Εισήχθησαν στο [KWT88] και έχουν χρησιμοποιηθεί στην ανάλυση εικόνων και βίντεο σε εφαρμογές όπως ανίχνευση περιγραμμάτων και παρακολούθηση αντικειμένων [BI98].

Παρά την ευρεία του αποδοχή, το αρχικό παραμετρικό μοντέλο ενεργών περιγραμμάτων έχει σημαντικά μειονεκτήματα. Αφενός, οι ιδιότητες του μοντέλου εξαρτώνται όχι μόνο από τη γεωμετρία της καμπύλης, αλλά και από την παραμετροποίησή της, καθιστώντας το μη-γεωμετρικό. Αφετέρου, είναι δύσκολο να χειριστεί κανείς αριθμητικά τοπολογικές αλλαγές στη γεωμετρία της εξελισσόμενης καμπύλης· το ζήτημα αυτό αντιμετώπιστηκε ικανοποιητικά μόνο σχετικά πρόσφατα [MT00]. Τα μοντέλα *γεωμετρικών ενεργών περιγραμμάτων* (geometric active contours), αρχικές εκδοχές των οποίων παρουσιάστηκαν στα [CCCD93, MSV95] και περαιτέρω εξελίχθηκαν με τη μορφή των *γεωδαιτικών ενεργών περιγραμμάτων* (geodesic active contours) [CKS97, YKK⁺97], υπερβαίνουν τις προαναφερθείσες αδυναμίες των παραμετρικών μοντέλων. Στα γεωδαιτικά ενεργά περιγράμματα η εξέλιξη της καμπύλης καθορίζεται με ελαχιστοποίηση του γεωδαιτικού μήκους της καμπύλης, το οποίο υπολογίζεται σε κατάλληλο χώρο εφοδιασμένο με μια μετρική εξαρτώμενη από χαρακτηριστικά της εικόνας. Καθότι δεν υπεισέρχεται κάποια αυθαίρετη παραμετροποίηση της καμπύλης το συγκεκριμένο μοντέλο είναι καθαρά γεωμετρικό. Πέρα από *ενδείξεις ακμών* (edge cues), εμπλουτίζοντας το μοντέλο με *χωρικές ενδείξεις* (region cues) έχει οδηγήσει σε μια σειρά από ιδιαίτερα επιτυχημένα σύγχρονα μοντέλα γεωμετρικών ενεργών περιγραμμάτων. Κάποια από αυτά, π.χ. τα [CV01, TYW01], έλκουν την καταγωγή τους από το καρτούν μοντέλο εικόνας των Mumford και Shah [MS89b]. Μια άλλη κατηγορία μοντέλων βασισμένων σε χωρικές ενδείξεις αποτελούν μετεξελίξεις του στατιστικού μοντέλου των Zhu και Yuille [ZY96]· ενδεικτικά αναφέρουμε τα [PD02, RBD03, JBBA03, KEM09], που πετυχαίνουν επιδόσεις *αιχμής* (state-of-the-art) σε κατάτμηση βασισμένη στην υφή ή την κίνηση αντικειμένων. Πρότερη πληροφορία για το σχήμα αντικειμένων μπορεί επίσης να ενσωματωθεί στο φορμαλισμό αυτό και να καθοδηγήσει κατάλληλα την εξέλιξη της καμπύλης, οδηγώντας σε μεθόδους κατάτμησης ανθεκτικές σε σημαντικές αλλοιώσεις της εικόνας ή αλληλεπικαλύψεις αντικειμένων [LGF00, CTWS02, RP02, KM05].

Οι περισσότερες αριθμητικές υλοποιήσεις γεωμετρικών ενεργών περιγραμμάτων βασίζονται στη μέθοδο των επιπεδοσυνόλων των Osher και Sethian [OS88] στην οποία το ενεργό περίγραμμα καθορίζεται έμμεσα ως το μηδενικό επιπεδοσύνολο μιας βαθμωτής συνάρτησης *εμβύθισης* (embedding) με πεδίο ορισμού το επίπεδο της εικόνας. Τούτο επιτρέπει αλλαγές στην τοπολογία της καμπύλης με πολύ φυσικότερο τρόπο σε σχέση με τα παραμετρικά ενεργά περιγράμματα. Για συγκεκριμένες ειδικές περιπτώσεις ενεργών περιγραμμάτων, μπορεί εναλλακτικά να χρησιμοποιήσει κανείς τεχνικές επίλυσης βασισμένες σε μεθόδους *τομής γραφών* (graph cuts) [KB05] ή *συντομότερου μονοπατιού* (minimal paths) [CK97]. Καθότι όμως οι αντίστοιχες μέθοδοι δεν είναι τόσο γενικά εφαρμόσιμες όσο τα επιπεδοσύνολα δεν θα μας απασχολήσουν περαιτέρω εδώ.

Παρόλα τα πλεονεκτήματα της μεθόδου των επιπεδοσυνόλων, το υπολογιστικό της κόστος μπορεί να είναι μεγάλο, καθιστώντας την προβληματική για εφαρμογές πραγματικού χρόνου. Διάφορες προσεγγίσεις έχουν προταθεί που βελτιώνουν τις υπολογιστικές ιδιότητες της μεθόδου. Με χρήση τεχνικών *στενής ζώνης* (narrow band), οι υπολογισμοί περιορίζονται στη γειτονιά του εξελισσόμενου περιγράμματος· το αντίτιμο είναι ότι απαιτείται κάποιος μηχανισμός επανακαθορισμού της υπολογιστικής ζώνης κατά τη μετακίνηση του περιγράμματος [Set99, SSO94, PMO⁺99]. Εξελίσσοντας την καμπύλη διαδοχικά ξεκινώντας από αδρές και προχωρώντας σε πιο λεπτομερείς κλίμακες με χρήση πυραμιδικών δομών επίσης οδηγεί σε ταχύτερη υλοποίηση [GGCV95]. Διαχωρισμός αριθμητικά σχήματα όπως ο *Προσθετικός Διαχωρισμός Τελεστών* (Additive Operator Splitting

- AOS) [LNT91, WtHRV98] έχουν προσαρμοστεί στο πρόβλημα των ενεργών γεωδαιτικών περιγραμμάτων για να χαλαρώσουν τον περιορισμό, εγγενή στα *άμεσα* (explicit) αριθμητικά σχήματα, ως προς το μέγιστο χρονικό βήμα εξέλιξης της καμπύλης που εξασφαλίζει αριθμητική ευστάθεια [GKRR01, WK03]. Αν και το AOS σχήμα έχει καλές ιδιότητες ευστάθειας, όταν συνδυάζεται με μεγάλο χρονικό βήμα μπορεί να οδηγήσει σε *διαχωριστικά ψεύδεργα* (splitting artifacts) εξαιτίας έλλειψης περιστροφικής συμμετρίας. Συνεπώς, εκτός και αν θυσιάσουμε την πιστότητα της αριθμητικής προσέγγισης, ο αριθμός των επαναλήψεων που απαιτούνται για τη σύγκλιση του περιγράμματος παραμένει σχετικά μεγάλος.

Η εργασία μας προτείνει αποδοτικούς πολυπλεγματούς αλγορίθμους για την επίλυση μιας ευρείας κατηγορίας μοντέλων γεωμετρικών ενεργών περιγραμμάτων. Οι προτεινόμενοι αλγόριθμοι, επειδή συνδυάζονται με εγγενώς 2-Δ (μη-διαχωρίσιμα) σχήματα χωρικής διακριτοποίησης, διατηρούν την ακρίβειά τους και επιδεικνύουν πολύ καλές ιδιότητες ευστάθειας και περιστροφικής συμμετρίας ακόμη και όταν εφαρμόζονται με ιδιαίτερα μεγάλο χρονικό βήμα στην αριθμητική εξέλιξη της καμπύλης. Τούτο επιτρέπει την ταχεία εξέλιξη και σύγκλιση της καμπύλης μετά από πολύ λίγες επαναλήψεις.

Στα προτεινόμενα αριθμητικά σχήματα κάνουμε σαφή διάκριση μεταξύ των σχετιζόμενων με την καμπυλότητα του περιγράμματος *εσωτερικών δυνάμεων*, σχεδιασμένων να διατηρούν την ομαλότητα της καμπύλης, και των υπόλοιπων ειδικών για κάθε μοντέλο *εξωτερικών δυνάμεων*, σχεδιασμένων να καθοδηγούν την καμπύλη ανάλογα με το περιεχόμενο της εικόνας. Η διάκριση αυτή επιβάλλεται επειδή οι εσωτερικές δυνάμεις είναι αριθμητικά ιδιαίτερα *δύσκαμπτες* (stiff), δηλαδή απαιτούν ειδικό χειρισμό για να προσομοιωθούν με ευστάθεια. Έτσι, χειριζόμαστε τις εσωτερικές δυνάμεις με έμμεσα (πεπλεγμένα) αριθμητικά σχήματα, ενώ τις εξωτερικές με άμεσα. Το υβριδικό *έμμεσο-άμεσο* [ARW95] σχήμα που προκύπτει συνδυάζει σε ένα ισχυρό μείγμα την ευστάθεια των έμμεσων σχημάτων και την ευελιξία των άμεσων σχημάτων.

Κλειδί για την επιτυχία των προτεινόμενων αλγορίθμων είναι η αποδοτική επίλυση ενός μεγάλου αλλά αραιού γραμμικού συστήματος εξισώσεων που ανακύπτει σε κάθε χρονικό βήμα εξέλιξης του περιγράμματος. Για αυτόν τον σκοπό εφαρμόζουμε πολυπλεγματούς τεχνικές [TOS01], επεκτείνοντας το πλαίσιο εφαρμογής αναλόγων μεθόδων από άλλους συγγραφείς [Act98, KKY04]. Το υπολογιστικό κόστος της πολυπλεγματούς λύσης ανά βήμα εξέλιξης του περιγράμματος είναι ανάλογο με το αντίστοιχο κόστος που συνοδεύει τη χρήση διαχωρίσιμων σχημάτων. Παρόλα αυτά το συνολικό υπολογιστικό κόστος ως τη σύγκλιση του περιγράμματος είναι τυπικά αισθητά μικρότερο λόγω του μειωμένου αριθμού των απαιτούμενων επαναλήψεων, απόρροια του μεγαλύτερου ανεκτού χρονικού βήματος. Συμπληρωματικά, δείχνουμε ότι οι προτεινόμενες πολυπλεγματούς τεχνικές συνδυάζονται ευνοϊκά με αποδοτικούς αλγορίθμους [FH04] για υπολογισμό του μετασχηματισμού απόστασης, που επίσης απαιτείται σε κάθε βήμα εξέλιξης του περιγράμματος. Αξίζει να σημειωθεί ότι οι προτεινόμενοι αλγόριθμοι είναι συμβατοί με τις προαναφερθείσες τεχνικές στενού εύρους ζώνης και πυραμιδικής επιτάχυνσης της εξέλιξης της καμπύλης.

3.1.1 Ενοποιημένη Θεώρηση Γεωμετρικών Μοντέλων Ενεργών Περιγραμμάτων

Θεωρούμε μια καμπύλη C που κείται στο επίπεδο της εικόνας και έχει μήκος $L(C)$ και έστω $\vec{C}(s) \equiv (x(s), y(s)) : [0, L(C)] \rightarrow \mathbb{R}^2$ η φυσική της παραμετροποίηση, με χρήση του μήκους τόξου s ως παραμέτρου. Θεωρούμε επίσης τη συνάρτηση I που αντιστοιχεί στην ένταση φωτεινότητας της εικόνας. Το κριτήριο που ελαχιστοποιούμε στα γεωδαιτικά ενεργά περιγράμματα είναι το γεωδαιτικό μήκος της καμπύλης [CKS97, YKK⁺97]

$$E(C) = \int_C g(\|\nabla I(\vec{C}(s))\|) ds, \quad (3.1)$$

όπου η ενδείκτρια ακμών $g : [0, +\infty) \rightarrow \mathbb{R}^+$ είναι μια φθίνουσα συνάρτηση του $\|\nabla I\|$ (ή άλλου παρόμοιου μέτρου έντασης ακμών), έτσι ώστε $g(0) = 1$, $g(r) \rightarrow 0$ καθώς $r \rightarrow +\infty$. Ελαχιστοποιώντας το συναρτησιακό της Εξ. (3.1) με μεταβολικές τεχνικές οδηγεί σε μια ΜΔΕ τύπου Euler-Lagrange· προκειμένου να οδηγηθούμε σε (τοπικό) ελάχιστο έχοντας ως αφειτηρία μια αρχική καμπύλη, την εξελίσσουμε προς την κατεύθυνση μέγιστης καθόδου, εισάγοντας μια μεταβλητή χρόνου t .

Η αριθμητική υλοποίηση του μοντέλου γεωδαιτικών περιγραμμάτων με χρήση επιπεδοσυνόλων επιτρέπει με φυσικό τρόπο αλλαγές στην τοπολογία της καμπύλης, δηλαδή διαχωρισμό και συγχώνευση περιγραμμάτων. Στο πλαίσιο των επιπεδοσυνόλων, η εξελισσόμενη καμπύλη $C(t)$ ορίζεται έμμεσα ως το μηδενικό επιπεδοσύνολο μιας συνάρτησης εμπύθισης u με πεδίο ορισμού όλο το επίπεδο της εικόνας, δηλαδή $C(t) = \{(x, y) : u(x, y, t) = 0\}$, όπου $u(x, y, t) : \mathbb{R}^2 \times [0, +\infty) \rightarrow \mathbb{R}$. Κατά σύμβαση αποδίδουμε στη u αρνητικές τιμές στο εσωτερικό και θετικές τιμές στο εξωτερικό του περιγράμματος. Συχνά μάλιστα η προσημασμένη συνάρτηση απόστασης από την καμπύλη (μετασχηματισμός απόστασης) επιλέγεται για την επέκταση της u μακριά από την καμπύλη, εξαιτίας των ευνοϊκών αριθμητικών ιδιοτήτων της. Μεταβαίνοντας στην αναπαράσταση επιπεδοσυνόλων, οι νόμοι κίνησης της καμπύλης εύκολα μεταφράζονται σε αντίστοιχους νόμους για εξέλιξη της συνάρτησης εμπύθισης u [OS88]. Για την περίπτωση των γεωδαιτικών ενεργών περιγραμμάτων (3.1) ο νόμος εξέλιξης της συνάρτησης εμπύθισης γράφεται $\frac{\partial u}{\partial t} = \|\nabla u\| \operatorname{div} \left(g(\|\nabla I\|) \frac{\nabla u}{\|\nabla u\|} \right)$ [CKS97]. Συνήθως χρησιμοποιείται η γενικότερη μορφή

$$\frac{\partial u}{\partial t} = \|\nabla u\| \left\{ \operatorname{div} \left(g(\|\nabla I\|) \frac{\nabla u}{\|\nabla u\|} \right) + cg(\|\nabla I\|) \right\}, \quad (3.2)$$

όπου έχουμε επιπλέον έναν όρο που αντιστοιχεί στην επονομαζόμενη 'δύναμη μπαλονιού', που δρα 'φουσκώνοντας' (για $c < 0$) ή 'ζεφουσκώνοντας' (για $c > 0$) την καμπύλη, επιταχύνοντας έτσι την εξέλιξή της [Coh91]. Βελτιωμένη ευθυγράμμιση της καμπύλης με ακμές της εικόνας μπορεί να επιτευχθεί με προσθήκη ενός όρου σχετιζόμενου με τον Λαπλασιανό τελεστή ανίχνευσης ακμών, όπως προτείνεται στο [KB03].

Το βασισμένο σε ενδείξεις ακμών μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων μπορεί να εμπλουτιστεί σημαντικά ενσωματώνοντας στην εξέλιξη της καμπύλης πληροφορία από το σύνολο της εικόνας. Τέτοια μοντέλα που ενσωματώνουν ενδείξεις περιοχής εμφανίζουν σημαντικά προτερήματα, όπως μειωμένη εξάρτηση από την αρχικοποίηση και αντοχή στο θόρυβο. Πολλά τέτοια μοντέλα έχουν εμπνευστεί από την καρτούν θέρωση της εικόνας των Mumford και Shah [MS89b], υπό την οποία το περίγραμμα C

διαχωρίζει το πεδίο Ω σε προσκήνιο R και παρασκήνιο R' και η παρατηρούμενη εικόνα I είναι μια αλλοιωμένη εκδοχή της ιδανικής εικόνας f . η f θεωρείται ότι είναι ομαλή εντός των τμημάτων R και R' , αλλά όχι κατά μήκος του συνόρου C . Ο νόμος εξέλιξης στην αναπαράσταση επιπεδοσυνόλων για το αντίστοιχο γεωμετρικό ενεργό περίγραμμα είναι [TYW01, VC02]

$$\frac{\partial u}{\partial t} = \|\nabla u\| \left\{ \operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right) + \alpha/2((f_R - I)^2 - (f_{R'} - I)^2) + \beta/2(\|\nabla f_R\|^2 - \|\nabla f_{R'}\|^2) \right\}, \quad (3.3)$$

όπου τα α και β είναι θετικές σταθερές. Καθότι η f θεωρείται άγνωστη, πρέπει να επανεκτιμάται καθώς το περίγραμμα εξελίσσεται [CTWS02]. Μια απλοποιημένη εκδοχή του προηγούμενου σχήματος είναι το Chan-Vese μοντέλο [CV01] υπό το οποίο η εικόνα f θεωρείται τμηματικά σταθερή ($c_R/c_{R'}$ εντός/εκτός της καμπύλης) και το περίγραμμα εξελίσσεται ως

$$\frac{\partial u}{\partial t} = \|\nabla u\| \left\{ \operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right) + \alpha/2((c_R - I)^2 - (c_{R'} - I)^2) \right\}. \quad (3.4)$$

Τα $c_R/c_{R'}$ τυπικά ανανεώνονται σε κάθε βήμα εξέλιξης, λαμβάνοντας ως τιμή τη μέση φωτεινότητα της εικόνας εντός/εκτός της καμπύλης, αντίστοιχα.

Πέρα από τη θεώρηση της εικόνας ως τμηματικά ομαλής συνάρτησης, μια σειρά τεχνικών που εμπνεύστηκαν από το μοντέλο *Ανταγωνισμού Περιοχών* (Region Competition) [ZY96] θεωρεί την εικόνα ως τμηματικά στατιστικά ομοιογενή. Έτσι, αντί για τη φωτεινότητα, κριτήριο συνεκτικότητας των περιοχών είναι κατάλληλες μετρήσεις που μπορεί να χαρακτηρίζουν το χρώμα, την υφή, την κίνηση ή άλλη ιδιότητα που διαφοροποιεί τις προς κατάτμηση περιοχές και οι οποίες μοντελοποιούνται πιθανοτικά [PD02, YTW99, RBD03, JBBA03, KEM09]. Στην περίπτωση ενός μοντέλου για το προσκήνιο R και ενός για το παρασκήνιο R' , καταλήγουμε στο νόμο

$$\frac{\partial u}{\partial t} = \|\nabla u\| \left\{ \operatorname{div} \left(g(\|\nabla I\|) \frac{\nabla u}{\|\nabla u\|} \right) + \alpha \log \frac{Pr(x|R)}{Pr(x|R')} \right\}, \quad (3.5)$$

όπου $Pr(x|R)$ συμβολίζει την πιθανοφάνεια του χαρακτηριστικού στη θέση x υπό το μοντέλο R . Σε προσεγγίσεις καθαρά βασισμένες σε χωρικές ενδείξεις θέτουμε $g \equiv 1$ στην Εξ. (3.5). ενδείξεις ακμών μπορούν να ενσωματωθούν με χρήση μιας ενδείκτριας συνάρτησης ακμών g , παρόμοια με την Εξ. (3.1) [PD02]. Εφαρμογή τέτοιου είδους μοντέλων σε κατάτμηση εικόνων ή παρακολούθηση αντικειμένων έχει αποδώσει πολύ καλά αποτελέσματα.

Αξίζει να σημειωθεί ότι όλες οι παραπάνω εκδοχές μοντέλων γεωμετρικών ενεργών περιγραμμάτων μπορούν να εκφραστούν στην ενοποιημένη μορφή

$$\frac{\partial u}{\partial t} = \|\nabla u\| \left\{ \operatorname{div} \left(g(\|\nabla I\|) \frac{\nabla u}{\|\nabla u\|} \right) + F(u) \right\}, \quad (3.6)$$

όπου ο πρώτος όρος αντιστοιχεί στην εσωτερική δύναμη ομαλότητας της καμπύλης, ενώ ο δεύτερος όρος $F(u)$ αντιστοιχεί στην εξωτερική δύναμη που καθοδηγεί την εξέλιξη του περιγράμματος ανάλογα με τα χαρακτηριστικά της εικόνας. Αν και δεν θα αναφερθούμε με

λεπτομέρεια σε αυτές, τεχνικές που ενσωματώνουν πρότερη πληροφορία συγκεκριμένου σχήματος μπορούν και αυτές να τεθούν στην ίδια μορφή [LGF00, CTWS02, RP02, KM05]. Η Εξ. (3.6) θα αποτελέσει έτσι την αφετηρία για την ενοποιημένη αριθμητική αντιμετώπιση μοντέλων γεωμετρικών ενεργών περιγραμμάτων με τους πολυπλεγματικούς αλγορίθμους της Ενότητας 3.3.

3.2 Ενδοσυμπλήρωση με Μερικές Διαφορικές Εξισώσεις

Μια σειρά από τεχνικές ενδοσυμπλήρωσης ελλιπούσας πληροφορίας σε εικόνες βασίζεται σε μοντέλα ΜΔΕ. Ένα από τα πρώτα τέτοια μοντέλα ΜΔΕ είναι αυτό που ανέπτυξε ο S. Carlsson τη δεκαετία του '80 [Car88] στην εργασία του σε κωδικοποίηση εικόνων. Το μοντέλο αυτό αφορά το κλασικό πρόβλημα ανακατασκευής εικόνας από αραυή πληροφορία-σκίτσο ακμών (θέση ακμών και τιμή της φωτεινότητας σε αυτές) και η προτεινόμενη λύση για ανακατασκευή της φωτεινότητας στις ομαλές περιοχές που περιβάλλονται από τις ακμές βασίζεται σε επίλυση της ελλειπτικής ΜΔΕ Laplace στα ομαλά χωρία, με συνοριακή συνθήκη Poisson στις ακμές. Είναι αξιοσημείωτο ότι η αριθμητική επίλυση της εξίσωσης Laplace στο μοντέλο του [Car88] γίνεται με πολυπλεγματικές μεθόδους, τις οποίες εμείς εφαρμόζουμε στη συνέχεια για την επίλυση πιο γενικών μη-γραμμικών και ανισοτροπικών ΜΔΕ.

Η κλασική θεώρηση ότι η εικόνα αποτελείται από ομαλά χωρία που διαχωρίζονται από έντονες ακμές, το λεγόμενο *καρτούν μοντέλο εικόνας* των Mumford και Shah [MS89b] που αναπτύχθηκε επίσης τη δεκαετία του '80, είναι εγγενής στη μέθοδο του Carlsson, όπως και σχεδόν σε όλες τις τεχνικές ενδοσυμπλήρωσης βασισμένες σε ΜΔΕ. Συνεπώς οι μέθοδοι ενδοσυμπλήρωσης με ΜΔΕ αποδίδουν ικανοποιητικά σε εικόνες που προσεγγίζονται από το καρτούν μοντέλο, αλλά όχι τόσο ικανοποιητικά σε εικόνες με έντονη υφή. Για εικόνες με παρουσία τόσο συνιστώσας καρτούν u , όσο και συνιστώσας υφής v , είναι συνηθισμένο να διαχωρίζουμε πρώτα τις δύο συνιστώσες, σε μια διαδικασία γνωστή ως *διαχωρισμός $u+v$* ($u+v$ decomposition). Ύστερα η ενδοσυμπλήρωση στη συνιστώσα καρτούν μπορεί να γίνει με μεθόδους ΜΔΕ, ενώ στη συνιστώσα υφής με εναλλακτικές μεθόδους, για παράδειγμα τεχνικές κυματιδίων. Αν και η ορολογία 'διαχωρισμός $u + v$ ' είναι σχετικά πρόσφατη [VO03], η βασική ιδέα πίσω από αυτή εφαρμόζεται από πιο παλιά, π.χ. στο [Car88].

Βέβαια η ανακατασκευή εικόνας από το σκίτσο των ακμών της που αντιμετωπίζει η μέθοδος του Carlsson διαφέρει από το πρόβλημα της ενδοσυμπλήρωσης, καθότι στο πρόβλημα της ενδοσυμπλήρωσης έχουμε δυνητικά έλλειψη πληροφορίας οπουδήποτε στην εικόνα και όχι μόνο στις ομαλές περιοχές. Συνεπώς κάθε μέθοδος ενδοσυμπλήρωσης πρέπει, άμεσα ή έμμεσα, να ανιχνεύσει τις ελλείπουσες ακμές και να αποκαταστήσει ένα σκίτσο τους, το οποίο θα δρα ως οδηγός στην αποκατάσταση των υπόλοιπων ομαλών ελλειπουσών περιοχών. Οι αντίστοιχες ΜΔΕ για ταυτόχρονη ανίχνευση και ανακατασκευή περιοχών είναι συνήθως ισχυρά μη-γραμμικές.

Από τα πρώτα μη-γραμμικά μοντέλα ΜΔΕ για ενδοσυμπλήρωση είναι αυτό των Masnou και Morel [MM98a, Mas02], όπου ανιχνεύονται οι ακμές στο γνωστό τμήμα της εικόνας και διαδίδονται στο εσωτερικό των κενών με τρόπο που να ελαχιστοποιεί το συναρτησιακό *elastica* του Euler [Mum94]. Για να οδηγηθούν από αυτή τη διαισθητική ιδέα σε συγκεκριμένα μοντέλα ΜΔΕ, οι συγγραφείς κάνουν ευρεία χρήση της έννοιας της

ισοϋψούς της συνάρτησης φωτεινότητας, που είναι κεντρική στη θεωρία επιπεδοσυνόλων [OS88]. Οι ιδέες αυτές επεκτείνονται σημαντικά από τους Bertalmio, Sapiro, Caselles και Ballester στην ιδιαίτερα δημοφιλή εργασία [BSCB00] που μάλιστα εισήγαγε και τον όρο 'inpainting' (ενδοσυμπλήρωση). Μια ενοποιημένη θεώρηση μεθόδων ενδοσυμπλήρωσης με ΜΔΕ δίνεται από τους Shen και Chan στα [CS01a, CS05], ενώ στο [CS01b] εξετάζεται πώς μπορούμε να πετύχουμε καλύτερη αντιληπτικά αποκατάσταση ακμών με διάδοση πληροφορίας με τρόπο που να λαμβάνεται υπόψη η καμπυλότητα των ακμών.

3.2.1 Μοντέλα Ανισοτροπικής Διάχυσης

Πολλά από τα πιο πετυχημένα μοντέλα ΜΔΕ για ενδοσυμπλήρωση που αναφέραμε μπορούν να γραφούν ως διαδικασίες *ανισοτροπικής*, μη-γραμμικής διάχυσης. Η βασική ιδέα είναι ότι αναλύεται η κατευθυντικότητα σε κάθε τοπική γειτονιά της εικόνας και κατόπιν ευνοείται η διάχυση παράλληλα στις ακμές της εικόνας έναντι της διάχυσης εγκάρσια σε αυτές. Με αυτόν τον τρόπο διαδίδονται οι ακμές από τις γνωστές περιοχές προς τα κενά της εικόνας.

Για να γίνουμε πιο συγκεκριμένοι, θεωρούμε την πολυκαναλική εικόνα u με N κανάλια $u_i, i = 1 \dots N$. Σε έγχρωμες εικόνες έχουμε 3 κανάλια, ενώ σε γκριζες μόνο 1. Για να αναλύσουμε την τοπική κατευθυντικότητα της εικόνας, θεωρούμε σε κάθε σημείο x (το x δηλώνει το διδιάστατο διάνυσμα θέσης στο επίπεδο της εικόνας) το συμμετρικό *δομικό τανυστή* (structure tensor)

$$J = G_\rho * \sum_{i=1}^N (\nabla u_{i,\sigma}) (\nabla u_{i,\sigma})^T, \quad (3.7)$$

ο οποίος γενικεύει σε πολυκαναλικές εικόνες την έννοια της κατεύθυνσης. Η συνέλιξη με γκαουσιανές μικρού εύρους ρ, σ κανονικοποιεί τη μέτρηση. Με ιδιο-ανάλυση του δομικού τανυστή παίρνουμε

$$J = \mu_1 \hat{\mathbf{n}}_1 \hat{\mathbf{n}}_1^T + \mu_2 \hat{\mathbf{n}}_2 \hat{\mathbf{n}}_2^T, \quad (3.8)$$

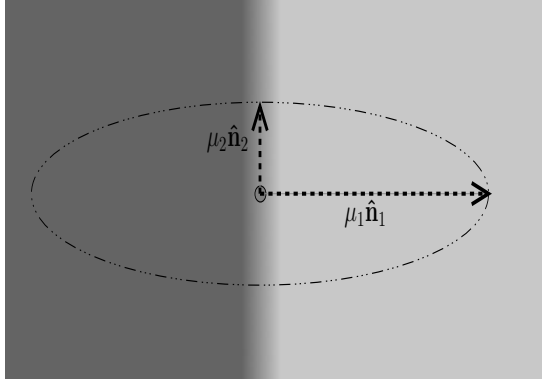
με $\mu_1 \geq \mu_2 \geq 0$ και $\hat{\mathbf{n}}_i^T \hat{\mathbf{n}}_j = \delta_{i,j}$, που γεωμετρικά αντιστοιχεί σε ελλειψοειδές με κύριο άξονα στη διεύθυνση του $\hat{\mathbf{n}}_1$ (εγκάρσια στις ακμές, μεγάλη ιδιοτιμή μ_1) και δευτερεύοντα άξονα στη διεύθυνση του $\hat{\mathbf{n}}_2$ (παράλληλα στις ακμές, μικρή ιδιοτιμή μ_2), όπως φαίνεται στο Σχ. 3.1(a). Στις ομαλές περιοχές της εικόνας έχουμε $\mu_1 \cong \mu_2 \cong 0$, κοντά σε έντονες ακμές $\mu_1 \gg \mu_2 \cong 0$, και κοντά σε γωνίες $\mu_1 \cong \mu_2 \gg 0$.

Έχοντας στη διάθεσή μας για κάθε σημείο της εικόνας πληροφορία κατευθυντικότητας, μπορούμε να διατυπώσουμε διαδικασίες ανισοτροπικής διάχυσης με δύο εναλλακτικούς φορμαλισμούς, όπως παρουσιάζεται στην ενοποιημένη θεώρηση του [Rou05]. Η πρώτη επιλογή είναι να εκφράσουμε τη διαδικασία ανισοτροπικής διάχυσης με την εξίσωση

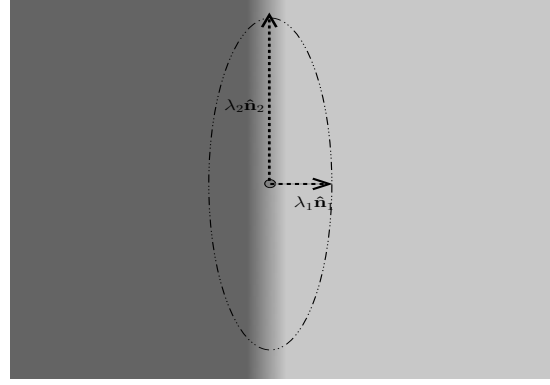
$$\frac{\partial u_i(x, t)}{\partial t} = \text{div} (D \nabla u_i), \quad i = 1, \dots, N, \quad (3.9)$$

με αρχική συνθήκη $u(x, 0) = u(x)$, κάνοντας χρήση του κοινού για όλα τα κανάλια συμμετρικού *τανυστή διάχυσης*

$$D = \begin{pmatrix} a & b \\ b & c \end{pmatrix} = D(x, t) = \lambda_1 \hat{\mathbf{n}}_1 \hat{\mathbf{n}}_1^T + \lambda_2 \hat{\mathbf{n}}_2 \hat{\mathbf{n}}_2^T. \quad (3.10)$$



(α) Δομικός τανυστής



(β) Τανυστής διάχυσης

Σχήμα 3.1: Δομικός τανυστής (αριστερά) και τανυστής διάχυσης (δεξιά) κοντά σε ακμή. Ο δομικός τανυστής είναι επιμήκης εγκάρσια στην ακμή (κατεύθυνση μέγιστης μεταβολής της φωτεινότητας), ενώ ο τανυστής διάχυσης επιλέγεται να είναι επιμήκης παράλληλα στην ακμή, ώστε η διάχυση να την αλλοιώνει ελάχιστα.

Οι ιδιοτιμές $\lambda_1 \leq \lambda_2$ του τανυστή διάχυσης $D(x, t)$ επιλέγονται με τρόπο που να ευνοείται η διάχυση παράλληλα με τις ακμές, όπως φαίνεται στο Σχ. 3.1(β): στο [Wei99] προτείνεται η επιλογή

$$\lambda_1 = \alpha \quad (3.11)$$

$$\lambda_2 = \begin{cases} \alpha, & \text{εάν } \mu_1 = \mu_2 \\ \alpha + (1 - \alpha) \exp\left(\frac{-C}{(\mu_1 - \mu_2)^2}\right), & \text{διαφορετικά} \end{cases} \quad (3.12)$$

Παραπάνω το α είναι μικρή θετική σταθερά που εξασφαλίζει ανελλιπή διάχυση, το C είναι θετική σταθερά κλίμακας, και το λ_2 είναι αύξουσα συνάρτηση της *συνεκτικότητας κατεύθυνσης* (directional coherence) $(\mu_1 - \mu_2)^2$ [Wei99, KW87]. Συνεπώς έχουμε ισχυρή διάχυση μόνο παράλληλα σε ισχυρές ακμές, με αποτέλεσμα να αυξάνεται η συνεκτικότητα κατεύθυνσης της εικόνας.

Εναλλακτικά, μπορούμε να εκφράσουμε τη διαδικασία ανιστροπικής εξομάλυνσης με φορμαλισμό κατευθυνόμενων παραγώγων, όπως οι Witkin και Kass [WK91] και αργότερα οι Tschumperle και Deriche [TD05]. Η διαδικασία ΜΔΕ διάχυσης που προτείνουν είναι η

$$\frac{\partial u_i(x, t)}{\partial t} = \lambda_1 \frac{\partial^2 u_i}{\partial \hat{n}_1^2} + \lambda_2 \frac{\partial^2 u_i}{\partial \hat{n}_2^2}, \quad i = 1, \dots, N, \quad (3.13)$$

που μπορεί εναλλακτικά να γραφεί στην πιο συμπυκνωμένη μορφή

$$\frac{\partial u_i(x, t)}{\partial t} = \text{trace}(TH_i), \quad i = 1, \dots, N, \quad (3.14)$$

όπου H_i είναι ο πίνακας μερικών παραγώγων δεύτερης τάξης του $u_i(x, t)$ και ο συμμετρικός *τανυστής εξομάλυνσης*

$$T = \begin{pmatrix} a & b \\ b & c \end{pmatrix} = T(x, t) = \lambda_1 \hat{n}_1 \hat{n}_1^T + \lambda_2 \hat{n}_2 \hat{n}_2^T \quad (3.15)$$

είναι και πάλι κοινός για όλα τα κανάλια. Οι ιδιοτιμές $\lambda_1 \leq \lambda_2$ επιλέγονται με τρόπο που να ευνοεί τη διάδοση ακμών, ακολουθώντας παρόμοια λογική με την επιλογή των αντίστοιχων ιδιοτιμών του τανυστή διάχυσης D παραπάνω·δες τα [TD05, Rou05] για περισσότερες λεπτομέρειες. Ένα σημαντικό πλεονέκτημα του φορμαλισμού κατευθυνόμενων παραγώγων είναι ότι η αντίστοιχη ΜΔΕ (3.14) μπορεί να εξελιχθεί σε μικρά βήματα Δt με επαναληπτικό προσαρμοσμένο φιλτράρισμα με γκαουσιανά φίλτρα των οποίων ο πίνακας συμμεταβλητότητας δίνεται απλά από τον κλιμακωμένο τανυστή $(4\Delta t)T$ [TD05]. Το γεγονός αυτό αναδεικνύει τη γεωμετρική σημασία του τανυστή εξομάλυνσης T και δείχνει τη στενή σχέση του φορμαλισμού ΜΔΕ ανισοτροπικής διάχυσης με τεχνικές επαναληπτικού προσαρμοστικού φιλτραρίσματος [YBFU96].

Ένα σημαντικό σημείο στο φορμαλισμό ΜΔΕ για ενδοσυμπλήρωση εικόνων αφορά στις συνοριακές συνθήκες που επιβάλλουμε στις παραπάνω εξισώσεις. Με αναφορά στο Σχ. 2.1, η βασική ιδέα είναι ότι εξελίσσουμε την ΜΔΕ στο κενό Ω , επιβάλλοντας συνοριακές συνθήκες Dirichlet (δηλαδή κρατώντας σταθερές τις γνωστές τιμές της εικόνας) στο σύνορο $\partial\Omega$ μεταξύ γνωστής περιοχής και κενού. Επίσης, στα σημεία που η οπή συνορεύει με τα όρια της εικόνας επιβάλλουμε συνοριακή συνθήκη Neuman (δηλαδή ανακλαστική συνοριακή συνθήκη), που είναι και η φυσική συνοριακή συνθήκη για το πρόβλημα.

3.3 Αριθμητικά Σχήματα και Πολυπλεγματικοί Αλγόριθμοι

Η υπολογιστική υλοποίηση των μεθόδων γεωμετρικών ενεργών περιγραμμάτων και ανισοτροπικής διάχυσης που συζητήσαμε κατά σειρά στις Ενότητες 3.1 και 3.2 απαιτεί κατάλληλη διακριτοποίηση των αντίστοιχων ΜΔΕ. Ακολουθούμε την τεχνική των πεπερασμένων διαφορών, και προχωρούμε στα επόμενα βήματα :

1. Διακριτοποίηση των χωρικών παραγώγων στο δεξί μέλος των εξισώσεων. Αφού πρόκειται για εξισώσεις παραβολικού τύπου, χρησιμοποιούμε κεντρικές διαφορές, με κατάλληλο χειρισμό των συνοριακών συνθηκών.
2. Για τη χρονική εξέλιξη της λύσης ως την τελική κατάσταση διακριτοποιούμε το χρόνο· συζητάμε τα προβλήματα ευστάθειας του απλού εμπρόσθιου σχήματος και πώς αυτά υπερβαίνονται από το οπίσθιο σχήμα χρονικής ολοκλήρωσης σε συνδυασμό με πολυπλεγματικές μεθόδους για επίλυση του προκύπτοντος γραμμικού συστήματος.

Η προσέγγιση που παρουσιάζουμε εδώ γενικεύει τις εργασίες μας [PM04, PM07], όπου η συζήτηση αφορούσε αποκλειστικά αλγόριθμους για την ταχεία εξέλιξη γεωμετρικών ενεργών περιγραμμάτων.

3.3.1 Διακριτοποίηση Χωρικών Παραγώγων

Για τη διακριτοποίηση του παραβολικού χωρικού όρου ακολουθούμε παρόμοια βήματα τόσο στην περίπτωση των γεωμετρικών ενεργών περιγραμμάτων, όσο και στην περίπτωση της ενδοσυμπλήρωσης με τεχνικές διάχυσης.

Γεωμετρικά Ενεργά Περιγράμματα

Αναφερόμενοι στην Εξ. (3.6), ο όρος $\|\nabla u\|(\operatorname{div}(g(\|\nabla I\|)\nabla u/\|\nabla u\|))$ περιγράφει κίνηση υπό γεωδαιτική καμπυλότητα. Όπως σημειώνεται στο [GKRR01], μια ιδιαίτερα επιθυμητή απλοποίηση επιτυγχάνεται αν επαναρχικοποιούμε σε κάθε επανάληψη τη συνάρτηση εμπύθισης u σε μετασχηματισμό απόστασης, έτσι ώστε $\|\nabla u\| = 1$. Τότε η εξέλιξη της καμπύλης όσο αφορά τον όρο γεωδαιτικής καμπυλότητας μπορεί να περιγραφεί από την απλοποιημένη γραμμική μη-ομογενή εξίσωση διάχυσης

$$\frac{\partial u}{\partial t} = \operatorname{div}(g(x, y)\nabla u), \quad (3.16)$$

όπου ο συντελεστής διάχυσης $g(x, y) = g(\|\nabla I\|)$ εξαρτάται από την εικόνα I αλλά όχι από την εξελισσόμενη συνάρτηση u . Υιοθετούμε αυτήν την τεχνική επαναρχικοποίησης γιατί οδηγεί στο χρονικά ανεξάρτητο τελεστή $\operatorname{div}(g(x, y)\nabla \cdot)$: αυτό είναι ιδιαίτερα επιθυμητό στο πλαίσιο των προτεινόμενων πολυπλεγματικών αλγορίθμων, καθότι οι πίνακες των προκυπτόντων γραμμικών συστημάτων σε κάθε επίπεδο ανάλυσης χρειάζεται να υπολογιστούν μόνο μια φορά σε μια φάση αρχικοποίησης και όχι σε κάθε επανάληψη. Επιπλέον η επαναρχικοποίηση της συνάρτησης εμπύθισης βελτιώνει την αριθμητική ευρωστία των αλγορίθμων επιπεδοσυνόλων [Set99]. Για την επαναρχικοποίηση χρησιμοποιούμε τον ταχύ αλγόριθμο μετασχηματισμού απόστασης που προτείνεται στο [FH04]· δες το [PM07] για συγκρίσεις με εναλλακτικούς αλγορίθμους μετασχηματισμού απόστασης στο συζητούμενο πλαίσιο.

Ο πιο διαδεδομένος τρόπος για να διακριτοποιήσουμε τον όρο διάχυσης της Εξ. (3.16) είναι ο εξής [WtHRV98]

$$\operatorname{div}(g\nabla u) = (gu_x)_x + (gu_y)_y = D_x(gD_x(u)) + D_y(gD_y(u)), \quad (3.17)$$

όπου ο τελεστής $D_x(w) = (w_{i+1/2,j} - w_{i-1/2,j})/h$ δηλώνει κεντρικές διαφορές μεταξύ στοιχείων που απέχουν κατά h . Αναπτύσσοντας την παραπάνω σχέση, μπορούμε να εκφράσουμε το χωρικό όρο ως τοπική συσχέτιση (τη συμβολίζουμε με $*$)

$$\operatorname{div}(g\nabla u)\Big|_{i,j} = u(i, j) *' m_{i,j} \triangleq \sum_{i'=-1}^1 \sum_{j'=-1}^1 u(i+i', j+j') m_{i,j}(i', j') \quad (3.18)$$

με μια 3×3 χωρικά μεταβαλλόμενη μάσκα $m_{i,j}$ της οποίας τα 5 μη-μηδενικά στοιχεία δίνονται στο Σχ. 3.2. Η μάσκα $m_{i,j}$ καλείται *ίχνος* (stencil) της διακριτοποίησης στο σημείο (i, j) [Wes92].

Ανισοτροπική Διάχυση – Φορμαλισμός Τανυστή Διάχυσης

Αναφορικά με τη ΜΔΕ ανισοτροπικής διάχυσης με τανυστή, Εξ. (3.9), διακριτοποιούμε το χωρικό όρο ως εξής

$$\begin{aligned} \operatorname{div}(D\nabla u) &= \operatorname{div}\left(\begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}\right) = (au_x)_x + (bu_y)_x + (bu_x)_y + (cu_y)_y \\ &= D_x(aD_x(u)) + D_{2x}(bD_{2y}(u)) + D_{2y}(bD_{2x}(u)) + D_y(cD_y(u)), \end{aligned} \quad (3.19)$$

		i'			
		-1	0	1	
j'	-1	0	$\frac{g(i,j-1/2)}{h_y^2}$	0	
	0	$\frac{g(i-1/2,j)}{h_x^2}$	$-\frac{g(i-1/2,j)+g(i+1/2,j)}{h_x^2}$	$-\frac{g(i,j-1/2)+g(i,j+1/2)}{h_y^2}$	$\frac{g(i+1/2,j)}{h_x^2}$
	1	0	$\frac{g(i,j+1/2)}{h_y^2}$	0	

Σχήμα 3.2: Το ίχνος 5-σημείων για υπολογισμό του χωρικού όρου $\text{div}(g\nabla u)$. Προσεγγίζουμε τις εκτός πλέγματος διακριτοποίησης απαιτούμενες τιμές με γραμμική παρεμβολή, για παράδειγμα $g(i+1/2, j) \cong (g(i+1, j) + g(i, j))/2$.

όπου ο τελεστής $D_{2x}(w) = (w_{i+1,j} - w_{i-1,j})/(2h)$ δηλώνει κεντρικές διαφορές μεταξύ στοιχείων που απέχουν κατά $2h$ (επιλέγουμε το συγκεκριμένο συνδυασμό D_x και D_{2x} ώστε τελικά να οδηγηθούμε σε έκφραση που να εξαρτάται από τις τιμές της u μόνο στα σημεία του πλέγματος). Αναπτύσσοντας την παραπάνω σχέση, μπορούμε παρόμοια με πριν να εκφράσουμε τον διακριτοποιημένο χωρικό όρο με τη βοήθεια του 3×3 ίχνους που δίνεται στο Σχ. 3.3. Ένα εναλλακτικό σχήμα διακριτοποίησης που αντιστοιχεί σε 5×5 ίχνος και προσφέρει αυξημένη περιστροφική συμμετρία δίνεται στο [WS02].

		i'			
		-1	0	1	
j'	-1	$\frac{b(i-1,j)+b(i,j-1)}{4h_x h_y}$	$\frac{c(i,j-1/2)}{h_y^2}$	$-\frac{b(i+1,j)+b(i,j-1)}{4h_x h_y}$	
	0	$\frac{a(i-1/2,j)}{h_x^2}$	$-\frac{a(i-1/2,j)+a(i+1/2,j)}{h_x^2}$	$-\frac{c(i,j-1/2)+c(i,j+1/2)}{h_y^2}$	$\frac{a(i+1/2,j)}{h_x^2}$
	1	$-\frac{b(i-1,j)+b(i,j+1)}{4h_x h_y}$	$\frac{c(i,j+1/2)}{h_y^2}$	$\frac{b(i+1,j)+b(i,j+1)}{4h_x h_y}$	

Σχήμα 3.3: Το ίχνος 9-σημείων για υπολογισμό του χωρικού όρου $\text{div}(D\nabla u)$. Προσεγγίζουμε τις εκτός πλέγματος διακριτοποίησης απαιτούμενες τιμές με γραμμική παρεμβολή, για παράδειγμα $a(i+1/2, j) \cong (a(i+1, j) + a(i, j))/2$.

Ανισοτροπική Διάχυση – Φορμαλισμός Κατευθυνόμενων Παραγώγων

Παρόμοια ισχύουν στην περίπτωση της ΜΔΕ διάχυσης με φορμαλισμό κατευθυνόμενων παραγώγων (3.14), όπου διακριτοποιούμε τον χωρικό όρο ως εξής

$$\begin{aligned} \text{trace}(TH) &= \text{trace} \left(\begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{pmatrix} \right) = au_{xx} + 2bu_{xy} + cu_{yy} \\ &= aD_x(D_x(u)) + 2bD_{2x}(D_{2y}(u)) + cD_y(D_y(u)). \quad (3.20) \end{aligned}$$

Αναπτύσσοντας την παραπάνω σχέση, μπορούμε και πάλι να εκφράσουμε τον διακριτοποιημένο χωρικό όρο με τη βοήθεια του 3×3 ίχνους που δίνεται στο Σχ. 3.4. Σημειωτέον ότι στο εξής θα θεωρούμε για απλότητα κοινό βήμα χωρικής διακριτοποίησης $h_x = h_y = h$.

Είδαμε λοιπόν ότι και στην περίπτωση των γεωμετρικών ενεργών περιγραμμάτων και στην περίπτωση ενδοσυμπλήρωσης διάχυσης (ανεξάρτητα από το χρησιμοποιούμενο φορμαλισμό), μπορούμε να εκφράσουμε το χωρικό όρο στο δεξί μέλος των σχετικών εξισώσεων

		i'		
		-1	0	1
j'	-1	$\frac{b(i,j)}{2h_x h_y}$	$\frac{c(i,j)}{h_y^2}$	$-\frac{b(i,j)}{2h_x h_y}$
	0	$\frac{a(i,j)}{h_x^2}$	$-\frac{2a(i,j)}{h_x^2} - \frac{2c(i,j)}{h_y^2}$	$\frac{a(i,j)}{h_x^2}$
	1	$-\frac{b(i,j)}{2h_x h_y}$	$\frac{c(i,j)}{h_y^2}$	$\frac{b(i,j)}{2h_x h_y}$

Σχήμα 3.4: Το 3×3 ίχνος για υπολογισμό του χωρικού όρου $\text{trace}(TH)$.

με τη βοήθεια ενός ίχνους m , δηλαδή ως τοπική συσχέτιση με μια χωρικά μεταβαλλόμενη 3×3 μάσκα $m_{i,j}$. Προχωρούμε ένα βήμα παραπέρα διατάσσοντας τα εικονοστοιχεία με λεξικογραφική σειρά και θεωρώντας την εικόνα u ως διάνυσμα-στήλη με $M = M_x M_y$ στοιχεία, όσα και το πλήθος των εικονοστοιχείων. Τότε μπορούμε να γράψουμε τους αντίστοιχους χωρικούς όρους στη συμπυκνωμένη μορφή

$$\text{div}(g\nabla u) \quad \text{ή} \quad \text{div}(D\nabla u) \quad \text{ή} \quad \text{trace}(TH) = Au. \quad (3.21)$$

Ο $M \times M$ πίνακας A είναι ιδιαίτερα αραιός, καθότι η i -γραμμή του περιέχει μόνο 9 το πολύ μη-μηδενικά στοιχεία, που δίνονται ακριβώς από τα στοιχεία της μάσκας συσχέτισης που αντιστοιχεί στο i -εικονοστοιχείο σύμφωνα με τη λεξικογραφική διάταξη. Παρόλα αυτά το εύρος ζώνης του πίνακα A δεν μπορεί να είναι μικρότερο από $2 \min(M_x, M_y) + 3$ και κατά συνέπεια πράξεις πάνω στον A (π.χ. επίλυση συστήματος ή αντιστροφή) δεν μπορούν να επιταχυνθούν σημαντικά με τεχνικές για πίνακες στενής ζώνης.

3.3.2 Σχήματα Χρονικής Ολοκλήρωσης

Χρησιμοποιώντας τη χωρικά διακριτή έκφραση (3.21) για τις χωρικές παραγώγους και κρατώντας το χρόνο σε συνεχή μορφή [Wei98], οι εξισώσεις διάχυσης (3.9) και (3.14) γράφονται ως το ακόλουθο σύστημα M Συνήθων Διαφορικών Εξισώσεων (ΣΔΕ) στο χρόνο:

$$\frac{\partial u}{\partial t} = Au + F(u). \quad (3.22)$$

Στην εξίσωση (3.22) ο όρος Au περιέχει τους διακριτοποιημένους όρους χωρικής παραγώγισης που συζητήσαμε παραπάνω, ενώ έχει απορροφήσει και οποιουδήποτε άλλους γραμμικούς όρους ενδεχόμενα υπάρχουν, π.χ. όρους δύναμης επαναφοράς ελατηρίου της μορφής $-\alpha(u - u_0)$. Έχουμε επίσης προσθέσει και τον γενικά μη-γραμμικό όρο αντίδρασης (reaction term) F , που δεν είχαμε περιλάβει στην προηγούμενη συζήτησή μας. Τέτοιες διαδικασίες συνδυασμένης διάχυσης-αντίδρασης, που εισήχθησαν από τον A. Turing ως μοντέλα μορφογένεσης σε ζωντανούς οργανισμούς [Tur52], μπορούν να συνθέσουν μορφές (patterns) και έχουν χρησιμοποιηθεί με επιτυχία για σύνθεση υψής [WK91, Tur91]· δεξ και το [ZM97] για περαιτέρω πληροφορίες. Για την αριθμητική επίλυση της παραπάνω γενικευμένης εξίσωσης χρειάζεται να εφαρμόσουμε κάποιο αριθμητικό σχήμα χρονικής ολοκλήρωσης.

Το εμπρόσδιο (forward) σχήμα $(u^{n+1} - u^n)/\tau = Au^n + F(u^n)$ έχει χρησιμοποιηθεί ευρύτατα στην επίλυση ΜΔΕ για ανάλυση εικόνων γιατί είναι το πιο άμεσο και εύκολο να υλοποιηθεί, καθότι και οι δύο όροι διάχυσης/αντίδρασης στο δεξί μέλος της εξίσωσης

υπολογίζονται με χρήση γνωστών τιμών της εικόνας από το τρέχον βήμα n . Δυστυχώς, εξαιτίας του αριθμητικά *δύσκαμπτου* (stiff) παραβολικού όρου Au , το εμπρόσθιο σχήμα είναι ευσταθές μόνο για πολύ μικρό *χρονικό βήμα* (time-step) $\tau \lesssim h^2$ [WtHRV98] (σημειωτέον ότι για διάχυση με χωρικά μεταβαλλόμενο ανισοτροπικό τανυστή δεν υπάρχει ακριβώς γνωστό φράγμα ευστάθειας, αντίστοιχο του $\tau \leq 0.25h^2$ που ισχύει για ισοτροπική διάχυση· στο [WS02] προτείνουν $\tau \leq h^2$). Αυτή η συνθήκη ευστάθειας είναι πολύ περιοριστική στην πράξη [PTVF92]: Εάν για να λάβουμε το επιθυμητό αποτέλεσμα χρειάζεται να εξελίξουμε την εξίσωση για χρόνο T , απαιτείται να κάνουμε αριθμό επαναλήψεων ίσο με $\frac{T}{\tau} \propto \frac{T}{h^2} = TM = \mathcal{O}(M)$, όπου M είναι ο αριθμός των εικονοστοιχείων. Συνεπώς δυνητικά χρειαζόμαστε χιλιάδες επαναλήψεις για να επεξεργαστούμε μια εικόνα συνηθισμένων διαστάσεων. Καθότι το κόστος ανά επανάληψη είναι επίσης $\mathcal{O}(M)$, το συνολικό κόστος του εμπρόσθιου αλγορίθμου κλιμακώνεται ως $\mathcal{O}(M^2)$. Έτσι, για παράδειγμα, το κόστος επεξεργασίας για μια εικόνα με διαστάσεις 200×200 εικονοστοιχεία είναι περίπου 16 φορές μεγαλύτερο από το κόστος επεξεργασίας της ίδιας εικόνας στη μισή ανάλυση 100×100 .

Για να βελτιώσουμε την επίδοση, η βασική ιδέα του [WtHRV98] είναι ότι η *στενωπός επίδοσης* (performance bottleneck) στο εμπρόσθιο σχήμα είναι η αναγκαστική επιλογή πολύ μικρών χρονικών βημάτων. Αυτή η επιλογή δεν επιβάλλεται τόσο από περιορισμούς ακρίβειας, καθότι και με μεγαλύτερα χρονικά βήματα ολοκλήρωσης πετυχαίνουμε ικανοποιητική ακρίβεια, όσο από τον πολύ ισχυρό περιορισμό ευστάθειας που περιορίζει ισχυρά το εμπρόσθιο σχήμα. Για αυτόν το λόγο προτιμούμε τη χρήση *οπισθίου* (backward) σχημάτων, που είναι ιδιαίτερα κατάλληλα για χρονική ολοκλήρωση άκαμπτων διαφορικών εξισώσεων με χρήση μεγάλων χρονικών βημάτων [Ame92, PTVF92].

Το πλήρως οπίσθιο σχήμα για την Εξ. (3.22) γράφεται $(u^{n+1} - u^n)/\tau = Au^{n+1} + F(u^{n+1})$. Αυτό το σχήμα μπορεί με ευστάθεια να χειριστεί τον άκαμπτο παραβολικό όρο Au , ακόμα και για πολύ μεγάλο χρονικό βήμα [WtHRV98]. Όμως, για μη-γραμμική μορφή του όρου F , η επίλυση του προκύπτοντος (συνήθως πυκνού και μη-γραμμικού) αλγεβρικού συστήματος ως προς το διάνυσμα u^{n+1} μπορεί να είναι ιδιαίτερα δύσκολη. Αυτό καθιστά τη χρήση του πλήρως οπίσθιου σχήματος προβληματική για μοντέλα με μη-τετριμμένο όρο αντίδρασης F .

Παρόλα αυτά, η χρήση οπίσθιου σχήματος για τη χρονική ολοκλήρωση του όρου αντίδρασης $F(u)$ δεν είναι απαραίτητη. Αντίθετα με τον άκαμπτο όρο διάχυσης Au , ο όρος αντίδρασης συνήθως υπόκειται σε ευσταθή ολοκλήρωση και με το απλούστερο εμπρόσθιο σχήμα για αρκετά μεγάλη τιμή του χρονικού βήματος. Τα πλεονεκτήματα του ευσταθούς οπίσθιου σχήματος για τον άκαμπτο παραβολικό όρο και του απλού εμπρόσθιου σχήματος για το μη-γραμμικό όρο αντίδρασης συνδυάζονται στην υβριδική κλάση των *εμπρόσθιων-οπίσθιων* (Implicit-Explicit - IMEX) σχημάτων [ARW95], ο απλούστερος εκπρόσωπος των οποίων είναι το σχήμα

$$\frac{u^{n+1} - u^n}{\tau} = Au^{n+1} + F(u^n). \quad (3.23)$$

Προφανώς, στην περίπτωση απουσίας όρου αντίδρασης, το πλήρες οπίσθιο και το εμπρόσθιο-οπίσθιο σχήμα ταυτίζονται. Εμπρόσθια-οπίσθια σχήματα έχουν χρησιμοποιηθεί με επιτυχία σε διάφορες εφαρμογές·δες π.χ. το [KWT88].

Υιοθετώντας το εμπρόσθιο-οπίσθιο σχήμα (3.23) οδηγεί σε εξέλιξη της λύσης με εξαιρετική ευστάθεια, ακόμη και για μεγάλα χρονικά βήματα. Σε αντάλλαγμα απαιτεί την

επίλυση σε κάθε βήμα του γραμμικού συστήματος

$$\left[\frac{1}{\tau}I - A\right]u^{n+1} = \frac{1}{\tau}u^n + F(u^n). \quad (3.24)$$

Ενώ στη μονοδιάστατη περίπτωση ο πίνακας του συστήματος είναι στενής ζώνης και μπορούμε να επιλύσουμε την παραπάνω εξίσωση ιδιαίτερα αποδοτικά, η διδιάστατη περίπτωση που μας ενδιαφέρει εδώ αντιμετωπίζεται πιο δύσκολα. Ο πίνακας του συστήματος $L \equiv \frac{1}{\tau}I - A$ είναι πολύ μεγάλος και κληρονομεί την αραιή δομή του A που περιγράψαμε νωρίτερα. Όπως έχουμε ήδη συζητήσει, η συγκεκριμένη αραιή δομή δεν επιτρέπει τη χρήση αποδοτικών μεθόδων για απευθείας επίλυση του συστήματος. Απλές επαναληπτικές μέθοδοι, όπως οι Jacobi και Gauss-Seidel συγκλίνουν αργά για τόσο μεγάλα συστήματα. Η σύγκλιση επιβραδύνεται περαιτέρω όσο το χρονικό βήμα μεγαλώνει, καθώς τότε ο κανονικοποιητικός θετικός διαγώνιος όρος $1/\tau$ μικραίνει. Με αυτές τις απλές επαναληπτικές μεθόδους για επίλυση του συστήματος μέχρι κάποια επιθυμητή ακρίβεια ο αριθμός των επαναλήψεων που απαιτείται είναι ανάλογος του αριθμού των εικονοστοιχείων M [PTVF92]. Καθότι το υπολογιστικό κόστος ανά επανάληψη είναι επίσης $\mathcal{O}(M)$, οδηγούμαστε σε συνολικό κόστος $\mathcal{O}(M^2)$ για την εξέλιξη κατά ένα βήμα της Εξ. (3.24) – καμία βελτίωση έναντι του απλού εμπρόσθιου αλγορίθμου.

Στην περίπτωση ισοτροπικής διάχυσης, έχουμε μηδενικό μη-διαγώνιο στοιχείο $b = 0$ στους τανυστές διάχυσης και οι αντίστοιχες μάσκες συνέλιξης $m_{i,j}$ που δείχνουμε στα Σχ. 3.3 και 3.4 έχουν μηδενικά στοιχεία στις 4 γωνίες. Τότε μπορούμε να χρησιμοποιήσουμε *διαχωρισμένα* (split) σχήματα, όπως το AOS που προτείνεται στο [WtHRV98]. Αν και τα διαχωρισμένα σχήματα είναι ιδιαίτερα αποδοτικά, έχουν και κάποια σημαντικά μειονεκτήματα, κυρίως εμφάνιση *αλλοιώσεων* (artifacts) από έλλειψη περιστροφικής συμμετρίας, όπως δείχνουμε στο [PM07]. Στην περίπτωσή μας όμως, όπου έχουμε πλήρως ανισοτροπική διάχυση, το σχήμα AOS δεν είναι εφαρμόσιμο. Στο [Wei98] προτείνεται ένας υβριδικός συνδυασμός του AOS για τα διαγώνια στοιχεία a και του τανυστή διάχυσης με το απλό εμπρόσθιο σχήμα για τα μη-διαγώνια στοιχεία b , αλλά αυτό δεν εξαλείφει τελείως το πρόβλημα ευστάθειας για μεγάλα χρονικά βήματα.

Στον αλγόριθμο που προτείνουμε, υιοθετούμε το διδιάστατο IMEX σχήμα στην πλήρη του μορφή (3.23). Για την αποδοτική επίλυση του αντίστοιχου συστήματος (3.24) σε βέλτιστο χρόνο $\mathcal{O}(M)$ καταφεύγουμε σε *πολυπλεγματικές* (multigrid) τεχνικές, όπως περιγράφουμε παρακάτω. Σχεδιάζοντας προσεκτικά τα συνιστώσα μέρη της πολυπλεγματικής μεθόδου, η εξέλιξη της λύσης κατά ένα χρονικό βήμα μπορεί να γίνει με κόστος συγκρίσιμο του κόστους ανά βήμα της εμπρόσθιας μεθόδου ή της μεθόδου AOS. Εκμεταλλευόμενοι την άριστη ευστάθεια του οπίσθιου σχήματος και επιλέγοντας χρονικό βήμα πολύ μεγαλύτερο από αυτό που επιτρέπουν οι εναλλακτικές τεχνικές, μπορούμε να πετύχουμε με τους προτεινόμενους πολυπλεγματικούς αλγορίθμους πολύ καλύτερη απόδοση.

3.3.3 Πολυπλεγματική Επίλυση Ανισοτροπικών Προβλημάτων

Οι πολυπλεγματικές αριθμητικές μέθοδοι χρησιμοποιούν μια ιεραρχία πλεγμάτων διακριτοποίησης για να επιλύσουν αποδοτικά μια πληθώρα σημαντικών προβλημάτων, με προεξέχοντα αυτά που ανακύπτουν από τη διακριτοποίηση μοντέλων ΜΔΕ [Bra77]. Καλές πρόσφατες γενικές αναφορές για πολυπλεγματικές μεθόδους είναι τα βιβλία [BHM00, TOS01, Wes92]. Στο πεδίο της Όρασης Υπολογιστών πολυπλεγματικές τεχνικές

Αλγόριθμος 1 Γενικευμένος πολυπλεγματοειδής αλγόριθμος σε δύο πλέγματα

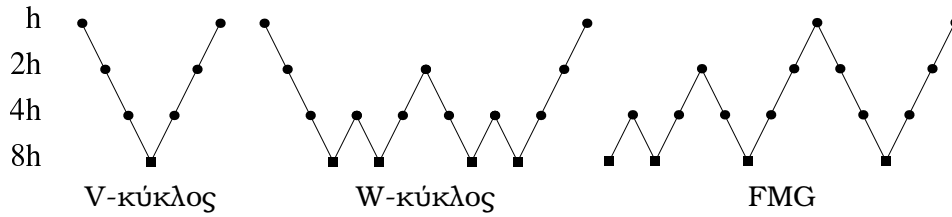
- 1: Κάνε n_1 -επαναλήψεις χαλάρωσης στο $L_h u_h = f_h$.
 - 2: Υπολόγισε το υπόλοιπο $r_h = f_h - L_h u_h$ στο λεπτομερές πλέγμα.
 - 3: Περιορίσε το υπόλοιπο $r_{2h} = Rr_h$ στο αδρό πλέγμα.
 - 4: Λύσε απευθείας το γραμμικό σύστημα του αδρού πλέγματος $L_{2h} e_{2h} = r_{2h}$.
 - 5: Επέκτεινε το σφάλμα στο λεπτομερές πλέγμα $e_h = P e_{2h}$.
 - 6: Διόρθωσε τη λύση λεπτομερούς πλέγματος $u_h \leftarrow u_h + e_h$.
 - 7: Κάνε n_2 -επαναλήψεις χαλάρωσης στο $L_h u_h = f_h$.
-

πρωτοχρησιμοποιήθηκαν συστηματικά από το [Ter86]. Για πολυπλεγματοειδείς αλγορίθμους εφαρμοσμένους στο πρόβλημα της οπτικής ροής, δες το [BWF⁺05]. Άλλες σχετικές αναφορές είναι τα [Act98, SBB00, KY03, CTWS02, KKY04, PM04].

Οι πολυπλεγματοειδείς τεχνικές υπερβαίνουν την τυπικά αργή σύγκλιση των συνηθισμένων μονοπλεγματοειδών τεχνικών χαλάρωσης για επίλυση γραμμικών συστημάτων, όπως οι μέθοδοι Jacobi, Gauss-Seidel και συζυγών κλίσεων [PTVF92]. Αν και οι μονοπλεγματοειδείς μέθοδοι εξαλείφουν αποδοτικά την υψίσυχη συνιστώσα του σφάλματος της λύσης, μειώνουν πολύ αργά τη χαμηλόσυχη συνιστώσα του. Τούτη η συμπεριφορά προέρχεται από τον τοπικό χαρακτήρα των υπολογισμών στις μονοπλεγματοειδείς μεθόδους, που επιτρέπει τη γρήγορη εξάλειψη των συνιστωσών σφάλματος που μεταβάλλονται σε κλίμακες συγκρίσιμες με το βήμα χωρικής διακριτοποίησης πλέγματος h , αλλά δυσχεραίνει την εξασθένιση των σφαλμάτων μεγαλύτερης κλίμακας. Καθώς το πλέγμα διακριτοποίησης γίνεται λεπτομερέστερο, η πληροφορία διαδίδεται ακόμα πιο αργά και ο μυωπικός χαρακτήρας των μονοπλεγματοειδών επαναληπτικών μεθόδων γίνεται ακόμα πιο ισχυρός. Οι πολυπλεγματοειδείς μέθοδοι υπερβαίνουν αυτές τις δυσκολίες χρησιμοποιώντας τέτοιες συμβατικές μεθόδους χαλάρωσης, καλούμενες ως *εξομαλυντές* (smoothers) στην πολυπλεγματοειδή ορολογία, σε μια *ιεραρχία* πλεγμάτων. Καθότι οι επαναλήψεις χαλάρωσης που πραγματοποιούνται σε κάθε επίπεδο του πλέγματος ομαλοποιούν τις συνιστώσες σφάλματος με κλίμακα συγκρίσιμη στο βήμα πλέγματος του συγκεκριμένου επιπέδου, ένας πολυπλεγματοειδής κύκλος μπορεί να εξαλείψει αποδοτικά τις συνιστώσες σφάλματος σε όλο το εύρος χωρικών συχνοτήτων. Επίσης, το κόστος εφαρμογής μεθόδων χαλάρωσης σε αδρές κλίμακες είναι πολύ μικρότερο από αυτό σε λεπτομερείς κλίμακες, αφού οι εμπλεκόμενες μεταβλητές είναι λιγότερες. Τέτοιου είδους πολυπλεγματοειδείς τεχνικές είναι βέλτιστες για ένα ευρύ φάσμα ενδιαφέροντων προβλημάτων, γραμμικών και μη-γραμμικών, με την έννοια ότι μπορούμε να τα λύσουμε σε οποιαδήποτε επιθυμητή ακρίβεια σε αριθμό πολυπλεγματοειδών κύκλων που είναι ανεξάρτητος από το μέγεθος του προβλήματος, και συνεπώς με γραμμικό κόστος $\mathcal{O}(M)$.

Για να γίνουμε πιο συγκεκριμένοι, δίνουμε το γενικευμένο πολυπλεγματοειδή Αλγόριθμο 1 για την επίλυση του γραμμικού συστήματος $L_h u_h = f_h$, όπως αυτό της Εξ. (3.24), στην ειδική περίπτωση δύο πλεγμάτων [BHM00]. Οι υποδείξεις δηλώνουν το βήμα διακριτοποίησης πλέγματος. Συγκεκριμένες μέθοδοι ανακύπτουν από τον Αλγόριθμο 1 αφού καθορίσουμε τις ακόλουθες *πολυπλεγματοειδείς συνιστώσες* (multigrid components):

1. Τον *εξομαλυντή* και τον *αριθμό επαναλήψεων πρότερης και ύστερης χαλάρωσης* n_1/n_2 : συνήθως επαρκούν $n_1 + n_2 \leq 3$ επαναλήψεις Gauss-Seidel εξομάλυνσης.
2. Τους διαπλεγματοειδείς πίνακες μεταφοράς, συγκεκριμένα τον $N_h \times N_{2h}$ πίνακα



Σχήμα 3.5: Συνηθισμένες πολυπλεγματικές τακτικές. Το βήμα πλέγματος δίνεται στην αριστερή στήλη. Διάγραμμα για την περίπτωση χρήσης 4 επιπέδων.

επέκτασης (παρεμβολής) P και τον $N_{2h} \times N_h$ πίνακα περιορισμού R , όπου $N_{(\cdot)}$ είναι ο αριθμός των αγνώστων σε ένα συγκεκριμένο επίπεδο, τυπικά $N_{2h} \approx 1/4N_h$ για 2-Δ προβλήματα· τυπικές επιλογές είναι δι-γραμμική παρεμβολή για το P και απλή υπο-δειγματοληψία για το R .

3. Τον πίνακα συστήματος για το αδρό επίπεδο L_{2h} : σε απλές περιπτώσεις αρκεί απλά να επαναδιακριτοποιήσουμε το συνεχές πρόβλημα στο αδρό πλέγμα.

Παρόλα αυτά οι παραπάνω τυπικές επιλογές πολυπλεγματικών συστατικών στοιχείων δεν είναι πάντα αποδοτικές. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο [PM07], όπου περιγράφουμε με λεπτομέρεια τις επιλογές που είναι πιο κατάλληλες για την τάξη των ανισοτροπικών προβλημάτων που αντιμετωπίζουμε εδώ.

Η επέκταση του δι-επίπεδου Αλγορίθμου 1 σε πολλαπλά επίπεδα γίνεται συνήθως με αναδρομικό τρόπο [TOS01]. Υπάρχουν πολλές εναλλακτικές προσεγγίσεις, που καθορίζουν και την ακριβή σειρά επίσκεψης στα διαφορετικά πλέγματα. Στο Σχ. 3.5 δείχνουμε παραδείγματα με τις πιο συνηθισμένες πολυπλεγματικές τακτικές. Χρησιμοποιώντας αρκετά επίπεδα διασφαλίζουμε ότι οι αγνώστοι στο πιο αδρό επίπεδο είναι πολύ λίγοι και συνεπώς είναι υπολογιστικά αμελητέα η απευθείας επίλυση του αντίστοιχου προβλήματος. Αξίζει εδώ να σημειώσουμε (δες και το Σχ. 3.5) ότι έχουμε δύο κατηγορίες πολυπλεγματικών αλγορίθμων: Οι αλγόριθμοι *διόρθωσης* (correction), όπως οι V- και W- κύκλοι, ξεκινούν από το πιο λεπτομερές επίπεδο και χρησιμοποιούν τα αδρότερα επίπεδα για να εξαλείψουν το χαμηλόσυχο σφάλμα. Από την άλλη, οι αλγόριθμοι *φωθιασμένης επανάληψης* (nested iteration), όπως ο FMG, ξεκινούν από τη λύση στα πιο αδρά επίπεδα και τη χρησιμοποιούν για να παράγουν μια αρχική εκτίμηση της λύσης στα πιο λεπτομερή επίπεδα. Ένας καλός ευρετικός κανόνας είναι να χρησιμοποιούμε αλγορίθμους διόρθωσης σε περιπτώσεις που έχουμε μια καλή αρχική εκτίμηση της λύσης, όπως ισχύει στην περίπτωση μας.

Η προηγούμενη συζήτηση αναδεικνύει μια σημαντική διαφοροποίηση μεταξύ πολυπλεγματικών τεχνικών και άλλων πολυκλιμακωτών αλγορίθμων εκλέπτυνσης της λύσης, όπως οι [HPB94, GGCV95, LHC96]. Οι πολυπλεγματικοί αλγόριθμοι περιλαμβάνουν τόσο ενδοπλεγματικές μεταβάσεις επέκτασης (από αδρά σε λεπτομερή επίπεδα διακριτοποίησης), όσο και ενδοπλεγματικές μεταβάσεις περιορισμού (από λεπτομερή σε αδρά επίπεδα διακριτοποίησης). Μόνο με αυτόν τον τρόπο επιτυγχάνεται ο βέλτιστος ρυθμός σύγκλισης που χαρακτηρίζει τις πολυπλεγματικές μεθόδους. Αντίθετα, οι παραδοσιακοί αλγόριθμοι πυραμιδικής ανάλυσης εικόνων περιλαμβάνουν μόνο μεταβάσεις επέκτασης· σε αυτούς ο ρόλος των αδρών πλεγμάτων εξαντλείται στο να παράσχουν μια εκτίμηση

Πίνακας 3.1: Υπολογιστικό κόστος κύριων αλγοριθμικών συνιστωσών κατά την κατάτμηση εικόνας διαστάσεων 256×256 εικονοστοιχείων με πολυπλεγματικούς αλγορίθμους.

Λειτουργία		Αλγόριθμος	Χρόνος (msec)	
			Αρχικοποίηση	Ανά βήμα
Σχήμα	2-Δ IMEX	Πολυπλεγματικός (Γεωμ. - MDMG)	35 ± 1	23 ± 1
		Πολυπλεγματικός (Αλγεβ. - AMG)	300 ± 20	40 ± 5
	AOS	Τριδιαγώνιος Επιλύτης (TRI)	13 ± 1	5 ± 1
Μετασχ. Απόστασης		Διαχωρίσιμος (SDT)	-	27 ± 1
		Ταχεία Προέλαση (FMDT)	-	38 ± 2

για καλή αρχικοποίηση της λύσης στο λεπτομερές πλέγμα. Για περαιτέρω λεπτομέρειες παραπέμπουμε στα [Wil02, BWF⁺05].

3.4 Πειράματα Κατάτμησης με Γεωμετρικά Ενεργά Περιγράμματα και Πολυπλεγματικούς Αλγορίθμους

Έχουμε αξιολογήσει την επίδοση των προτεινόμενων πολυπλεγματικών αλγορίθμων σε πειράματα κατάτμησης με τη χρήση διάφορων μοντέλων της οικογένειας των γεωμετρικών ενεργών περιγραμμάτων που συζητήσαμε στην Ενότητα 3.1. Έχουμε πειραματιστεί με δύο δημοφιλείς και δημόσια διαθέσιμους πολυπλεγματικούς *επιλύτες* (solvers) ανοιχτού κώδικα: 1) Τον επιλύτη *mgd9n*, που υλοποιεί προσαρμοστικές γεωμετρικές πολυπλεγματικές τεχνικές [dZ90] (MDMG). 2) Τον επιλύτη *amg1r5*, που υλοποιεί αλγεβρικές πολυπλεγματικές τεχνικές [RS87] (AMG). Παραπέμπουμε στο [PM07] για συγκριτική παρουσίαση γεωμετρικών και αλγεβρικών πολυπλεγματικών τεχνικών. Για υπολογισμό του μετασχηματισμού απόστασης (DT) έχουμε υλοποιήσει το διαχωρίσιμο αλγόριθμο του [FH04] (SDT) και τη μέθοδο *ταχείας προέλασης* (fast marching) [Set99] (FMDT). Συγκρίνουμε τις προτεινόμενες τεχνικές επίλυσης με εναλλακτικά διαχωρίσιμα σχήματα βασισμένα στην AOS τεχνική [WtHRV98, GKRR01, WK03]: η σχετική υλοποίησή μας χρησιμοποιεί κώδικα από τη συλλογή LAPACK για την επίλυση των τριδιαγώνιων συστημάτων εξισώσεων που ανακύπτουν (TRI). Στα πειράματα που ακολουθούν η σύγκλιση της εξελισσόμενης καμπύλης σηματοδοτείται όταν η Hausdorff απόσταση [Mar02] μεταξύ δύο διαδοχικών στιγμιότυπων του περιγράμματος πέσει κάτω από ένα κατώφλι H_T : συνήθως $H_T = 1 - 5$ εικονοστοιχεία. Οι χρόνοι εκτέλεσης αναφέρονται στην υλοποίησή μας σε C++¹. Οι μετρήσεις έχουν γίνει με χρήση φορητού υπολογιστή που διαθέτει επεξεργαστή Pentium M710 1.4 GHz και 512 MB κύρια μνήμη.

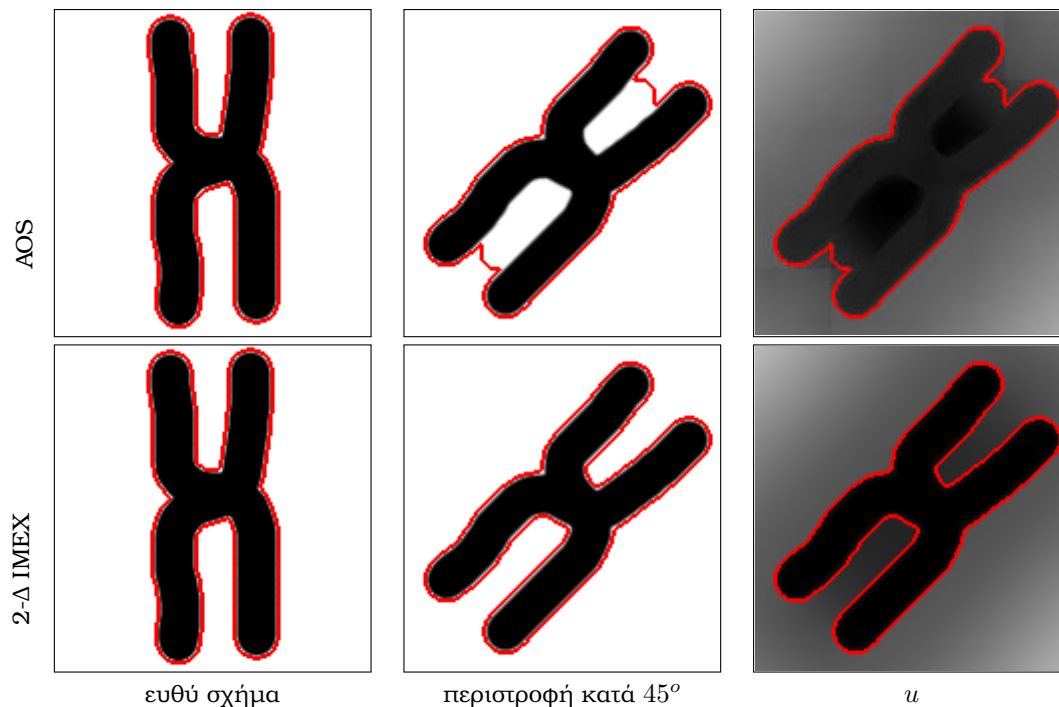
Παραθέτουμε στον Πίν. 3.1 το χρόνο εκτέλεσης των βασικών αλγοριθμικών συνιστωσών σε τυπικές εικόνες ανάλυσης 256×256 εικονοστοιχείων (χωρίς να γίνεται χρήση τεχνικών στενής ζώνης). Πιο συγκεκριμένα, δίνουμε πειραματικές μετρήσεις για το κόστος των εξής λειτουργιών: 1) *βήμα εξέλιξης περιγράμματος*, χρησιμοποιώντας είτε το πλήρως 2-Δ IMEX σχήμα (3.24) σε συνδυασμό με τους προτεινόμενους πολυπλεγματικούς αλγορίθμους (δοκιμάσαμε και τους δύο επιλύτες MDMG και AMG), είτε το διαχωρίσιμο

¹Για να διευκολύνουμε πειραματισμό με τους προτεινόμενους αλγορίθμους, διαθέτουμε ελεύθερα τον σχετικό πηγαίο κώδικα που έχουμε αναπτύξει. Δείτε το <http://cvsp.cs.ntua.gr/software>.

AOS σχήμα σε συνδυασμό με τριδιαγώνιους επιλύτες· 2) επαναρχικοποίηση συνάρτησης εμπύθισης με αλγορίθμους μετασχηματισμού απόστασης. Για τους πολυπλεγματούς και τριδιαγώνιους επιλύτες δίνουμε χωριστά το κόστος αρχικοποίησης και το κόστος ανά βήμα: επειδή στην προτεινόμενη τεχνική επαναρχικοποιούμε τη συνάρτηση εμπύθισης σε κάθε βήμα, οι πίνακες των αντίστοιχων γραμμικών συστημάτων μένουν αμετάβλητοι (εκτός κι αν χρησιμοποιούμε τεχνικές στενής ζώνης)· έτσι και στις δύο περιπτώσεις το κόστος αρχικοποίησης του επιλύτη (πολυπλεγματοί αλγόριθμοι: υπολογισμός τελεστών επέκτασης/περιορισμού και πινάκων συστήματος στα αδρά πλέγματα· τριδιαγώνιος επιλύτης: τριγωνοποίηση LU) το επωμιζόμαστε μόνο μια φορά.

Όπως μπορεί να ιδωθεί από τον Πίν. 3.1, το κόστος αρχικοποίησης είναι πολλαπλάσιο για τον αλγεβρικό πολυπλεγματού αλγόριθμο AMG σε σχέση με τον προσαρμοστικό γεωμετρικό πολυπλεγματού αλγόριθμο MDMG. Αυτό καθιστά τον επιλύτη MDMG πολύ πιο αποδοτικό σε σχέση με τον AMG, ακόμα και εάν συνδυάσουμε τον AMG με τεχνικές στενής ζώνης, όπως κάναμε στο [PM04]: αν και σε κάθε βήμα εξέλιξης της λύσης ο αριθμός των μεταβλητών στη στενή ζώνη είναι μικρότερος, το κόστος επαναρχικοποίησης λόγω συνεχούς μετακίνησης της στενής ζώνης υπερβαίνει το κέρδος. Το κόστος και των δύο πολυπλεγματού αλγορίθμων (MDMG και AMG) ανά βήμα εξέλιξης του περιγράμματος είναι μικρό καθώς ένας μόνος $V(0,1)$ πολυπλεγματού κύκλος αρκεί για να καταπιέσει επαρκώς το σφάλμα. Ειδικά για τον MDMG επιλύτη, το κόστος ανά βήμα εξέλιξης του μετώπου είναι της ίδιας τάξης με το κόστος του διαχωρίσιμου AOS σχήματος, διατηρώντας παράλληλα τα πλεονεκτήματα του πλήρως 2-Δ IMEX σχήματος. Για τα μοντέλα γεωμετρικών ενεργών περιγραμμάτων που βασίζονται αποκλειστικά σε χωρικές ενδείξεις, δηλαδή $g \equiv 1$, είναι δυνατόν να πετύχουμε ακόμα μεγαλύτερη αποδοτικότητα επειδή απλούστερες πολυπλεγματού συνιστώσες είναι επαρκείς. Σχετικά με το μετασχηματισμό απόστασης, ο διαχωρίσιμος αλγόριθμος SDT φαίνεται να είναι πιο αποδοτικός από τον αλγόριθμο ταχείας προέλασης FMĐT. Συνολικά λοιπόν, ο συνδυασμός του προσαρμοστικού πολυπλεγματού αλγορίθμου και του διαχωρίσιμου μετασχηματισμού απόστασης (MDMG+SDT) κοστίζει ≈ 50 msec για κάθε βήμα εξέλιξης του περιγράμματος, επιτρέποντας έτσι 20 ανανεώσεις του μετώπου ανά δευτερόλεπτο· καθότι με τα μεγάλα χρονικά βήματα που επιτρέπει το 2-Δ IMEX σχήμα πολύ λίγες επαναλήψεις αρκούν για τη σύγκλιση του περιγράμματος, ο τυπικός χρόνος εντός του οποίου μπορούμε να καταστήσουμε εικόνες 256×256 ιχνοστοιχείων είναι κάτω από ένα δευτερόλεπτο. Επίσης το υπολογιστικό κόστος των εμπλεκόμενων αλγορίθμων κλιμακώνεται γραμμικά με τον αριθμό των εικονοστοιχείων N . Έτσι στα υπόλοιπα πειράματα που περιγράφουμε έχουμε χρησιμοποιήσει αυτές τις δύο MDMG+SDT βασικές αλγοριθμικές συνιστώσες.

Το επόμενο πείραμα αναδεικνύει την αυξημένη ακρίβεια και περιστροφική συμμετρία του πλήρως 2-Δ IMEX σχήματος (3.23) σε σχέση με το διαχωρισμένο AOS σχήμα. Στο Σχ. 3.6 δείχνουμε σε μια συνθετική εικόνα μια ανισοτροπική στρέβλωση του AOS σχήματος την οποία καλούμε φαινόμενο σκίασης. Καθότι η διάχυση επιτελείται ανεξάρτητα στις διευθύνσεις των x - και y - αξόνων, κοιλότητες οι οποίες δεν είναι απολύτως ορατές από παρατηρητή ο οποίος σαρώνει την εικόνα σε αυτές τις διευθύνσεις δεν 'φωτίζονται' επαρκώς και η καμπύλη δεν μπορεί να εισχωρήσει επαρκώς σε αυτές. Το πλήρως 2-Δ IMEX σχήμα δεν αντιμετωπίζει τέτοια ζητήματα καθότι επιτρέπει διάχυση σε κάθε κατεύθυνση. Τα καλύτερα ποιοτικά χαρακτηριστικά του IMEX σχήματος αναδεικνύονται περαιτέρω στο μεγεθυμένο αποτέλεσμα κατάτμησης του Σχ. 3.7· παρατηρούμε ότι ακόμα και για σχετικά μικρά χρονικά βήματα, εδώ $\tau = 2$, το AOS σχήμα μπορεί να οδηγήσει σε



Σχήμα 3.6: Ανισοτροπικά ψεύδεργα του AOS σχήματος. Πάνω: Το AOS σχήμα δίνει διαφορετικά αποτελέσματα μετά από περιστροφή της εικόνας κατά 45° . Αυτό οφείλεται στο φαινόμενο σκίασης του AOS σχήματος, που οπτικοποιείται καλύτερα μέσω της γραφικής παράστασης της συνάρτησης εμβύθισης u (τρίτη στήλη). Κάτω: Η περιστροφή δεν επηρεάζει το αποτέλεσμα του πλήρως 2- Δ IMEX σχήματος. Πειράματα με χρήση του μοντέλου γεωδαιτικών ενεργών περιγραμμάτων, Εξ. (3.2), με $c = 0$ και $\tau = 1000$.

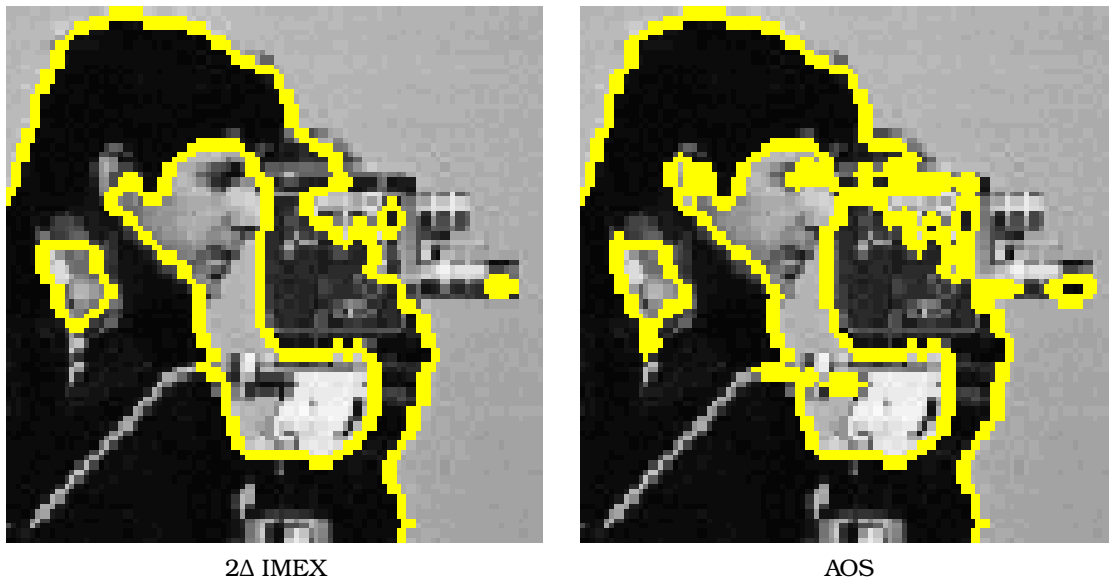
ορατά ψεύδεργα λόγω απώλειας περιστροφικής συμμετρίας.

Μια επιπλέον σειρά από αποτελέσματα παρουσιάζονται στη συνέχεια. Κατάτμηση βασισμένη στο χωρικό μοντέλο Chan-Vese δείχνεται στο Σχ. 3.8. Το περίγραμμα συγκλίνει ύστερα από 5 μόλις επαναλήψεις εντός 0.55 sec. Αποτέλεσμα σε κατάτμηση βασισμένη στο περιεχόμενο υψής μιας 436×286 εικόνας με χρήση του περιγραφητή υψής που προτείνεται στο [RBD03] παρατίθεται στο Σχ. 3.9. Το μέτωπο συγκλίνει ύστερα από 7 επαναλήψεις εντός 1.8 sec, συλλαμβάνοντας ικανοποιητικά το περίγραμμα της ζέβρας.

3.5 Πειράματα Ενδοσυμπλήρωσης με Μερικές Διαφορικές Εξισώσεις Ανισοτροπικής Διάχυσης και Πολυπλεγματικούς Αλγορίθμους

Έχουμε προχωρήσει σε πειράματα ενδοσυμπλήρωσης, συνδυάζοντας τα μοντέλα ΜΔΕ ανισοτροπικής διάχυσης με αποδοτική πολυπλεγματική υλοποίηση. Ο χρόνος αποκατάστασης για μια εικόνα 256×256 είναι της τάξης του 1 δευτερολέπτου.

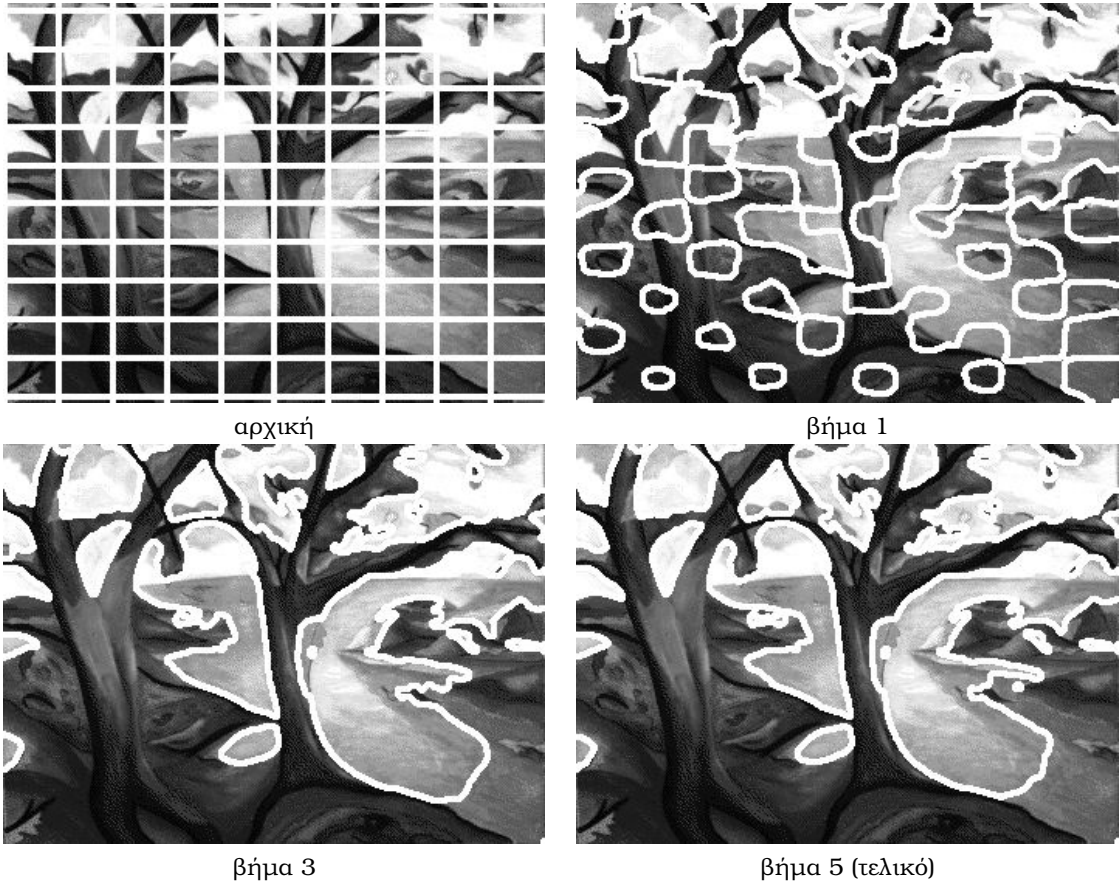
Στο παράδειγμα που δίνουμε στο Σχ. 3.10, με αποκατάσταση μέτρων βλαβών Τύπου I και II στην τοιχογραφία 'Κροκοσυλλέκτρια & Πότνια', φαίνονται τα βήματα που ακολου-



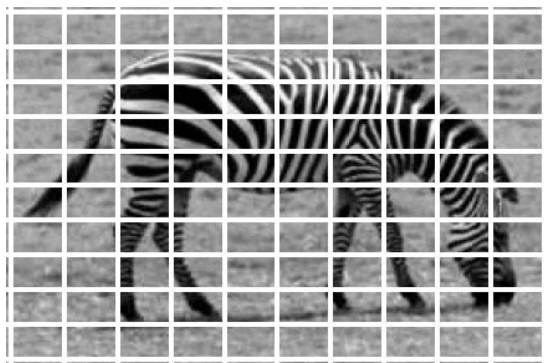
Σχήμα 3.7: Πλήρως 2-Δ IMEX σε σύγκριση με τη διαχωρίσιμη AOS διακριτοποίηση. Το AOS σχήμα δεν κανονικοποιεί επαρκώς την καμπύλη και ευνοεί το σχηματισμό γραμμών παράλληλα προς τους άξονες. Κατάτμηση με το μοντέλο Chan-Vese, Εξ. (3.4), κατάτμηση της εικόνας 'Cameraman' (λεπτομερής άποψη), $\tau = 2$.

Θούμε κατά την ενδοσυμπλήρωση με μη-γραμμικές ΜΔΕ. Ξεκινάμε με την αρχική εικόνα και επισημαίνουμε τα κενά που πρέπει να αποκατασταθούν, Σχ. 3.10(α). Στη συνέχεια προχωράμε σε μια πρώτη ομαλή αποκατάσταση της εικόνας, λύνοντας τη γραμμική ΜΔΕ Laplace, Σχ. 3.10(β). Έχοντας ως αφετηρία τη λύση της εξίσωσης Laplace, εξελίσσουμε μια μη-γραμμική ανισοτροπική ΜΔΕ, στο συγκεκριμένο παράδειγμα αυτήν της Εξ. (3.9)· αυτή, σε διαδοχικές επαναλήψεις, Σχ. 3.10(γ-δ), βελτιώνει την πρώτη εκτίμηση και δίνει το τελικό αποτέλεσμα. Η λεπτομερής σύγκριση των λύσεων Laplace, Σχ. 3.10(ε) και μη-γραμμικής ανισοτροπικής ΜΔΕ, Σχ. 3.10(ζ) αναδεικνύει την ικανότητα της μη-γραμμικής μεθοδολογίας να επεκτείνει τις ακμές της εικόνας από τις γνωστές περιοχές προς τα κενά· για παράδειγμα η γραμμή στη ράχη της κροκοσυλλέκτριας αποκαθίσταται επιτυχώς από τη ΜΔΕ ανισοτροπικής διάχυσης. Βέβαια σε ορισμένες περιπτώσεις αυτό μπορεί να οδηγήσει σε ψεύδεργα· για παράδειγμα, η μη-γραμμική ανισοτροπική ΜΔΕ έχει συνενώσει εσφαλμένα τις ρίγες στο φόρεμα της κροκοσυλλέκτριας, όπως φαίνεται στο Σχ. 3.10(η), ενώ η γραμμική λύση Laplace είναι συντηρητικότερη και απλά προκρίνει μια θολή συμπλήρωση για τη συγκεκριμένη περιοχή, όπως φαίνεται στο Σχ. 3.10(στ).

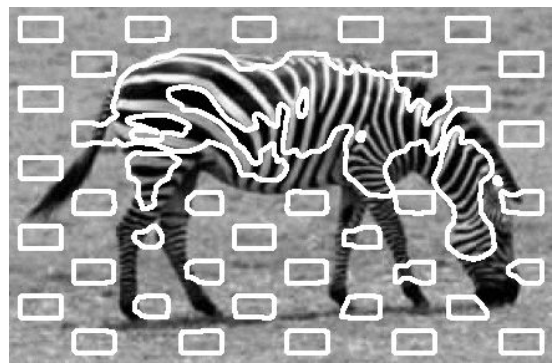
Το παράδειγμα αυτό αναδεικνύει βασικά χαρακτηριστικά των μεθόδων ΜΔΕ για ενδοσυμπλήρωση που συζητήσαμε νωρίτερα. Συγκεκριμένα φαίνεται ότι οι μέθοδοι ΜΔΕ αποδίδουν καλά σε αποκατάσταση μικρών κενών, αλλά το αποτέλεσμα υστερεί στην ενδοσυμπλήρωση μεγάλων κενών, κυρίως λόγω υπερβολικού θολώματος. Επίσης η αποκατάσταση επίπεδων καρτούν περιοχών είναι ικανοποιητική, όχι όμως και η αποκατάσταση περιοχών υψής, όπου η προτεινόμενη συμπλήρωση συχνά είναι υπερβολικά ομαλή. Κάποιοι από αυτούς τους περιορισμούς μπορεί να αρθούν όταν κάνουμε χρήση εναλλακτικών μεθόδων ενδοσυμπλήρωσης, όπως αυτές που συζητάμε στο Κεφάλαιο 6.



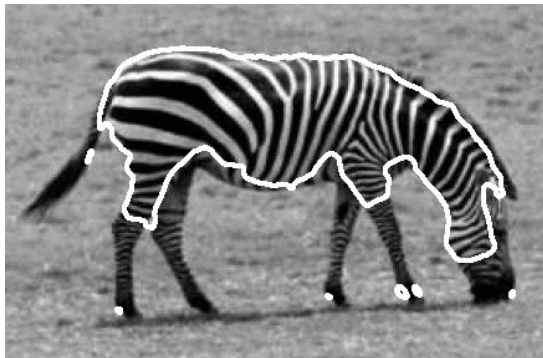
Σχήμα 3.8: Κατάτμηση μιας 350×258 εικόνας με τη μέθοδο Chan-Vese. Σύγκλιση μετά από 5 επαναλήψεις ($\tau = 50$, χρόνος εκτέλεσης 0.55 sec).



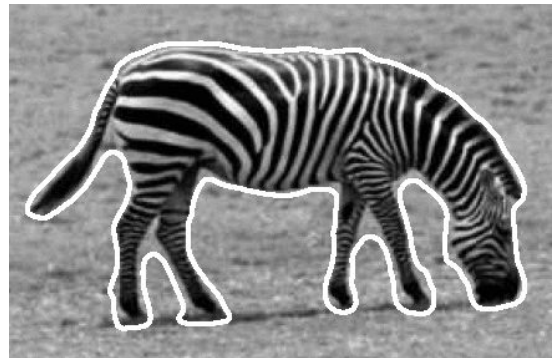
αρχική



βήμα 1



βήμα 2



βήμα 7 (τελικό)

Σχήμα 3.9: Κατάτμηση υφής με τη μέθοδο του [RBD03] σε εικόνα 436×286 . Σύγκλιση σε 7 επαναλήψεις ($\tau = 25$, χρόνος εκτέλεσης 1.8 sec).



(α) επισημειωμένη εικόνα



(β) αποκατάσταση με εξίσωση Laplace



(γ) πρώτη επανάληψη ανισοτρ. διάχυσης



(δ) τελική κατάσταση ανισοτρ. διάχυσης



(ε,στ) λεπτομέρειες από τη λύση Laplace



(ζ,η) λεπτομέρειες από την τελική λύση ανισοτροπικής διάχυσης



Σχήμα 3.10: Ενδοσυμπλήρωση με ΜΔΕ Laplace και ανισοτροπικής διάχυσης σε τμήμα της τοιχογραφίας 'Κροκοσυλλέκτρια & Πότνια'.

Κεφάλαιο 4

Αυτόματη Επιλογή Κλίμακας Μέσω Στατιστικής Διεπικύρωσης

Σημαντικό πρόβλημα σε πολλές εφαρμογές ανάλυσης εικόνας και όρασης υπολογιστών είναι η παρουσία θορύβου, που εισάγεται κατά τη διάρκεια λήψης ή κωδικοποίησης της εικόνας και δυσκολεύει τη μετέπειτα επεξεργασία της. Πολλά συστήματα επεξεργασίας εικόνας έχουν στο στάδιο προ-επεξεργασίας ένα υποσύστημα αποθορυβοποίησης προκειμένου να απαλλαγούν από αυτόν τον ανεπιθύμητο θόρυβο.

Η απλοποίηση μιας θορυβώδους εικόνας με τεχνικές μη-γραμμικών *χώρων κλίμακας* (scale-spaces) είναι από τις πιο επιτυχημένες μεθόδους αποθορυβοποίησης [RM09]. Τέτοιοι μη-γραμμικοί χώροι κλίμακας, βασισμένοι είτε σε μη-γραμμικές διαδικασίες διάχυσης, είτε σε επεξεργασία της εικόνας με μορφολογικούς τελεστές έχουν την επιθυμητή ιδιότητα να απλοποιούν και να αποθορυβοποιούν την εικόνα, ενώ παράλληλα διατηρούν σε καλό βαθμό τη θέση και την αντίθεση των ακμών. Σε ένα ευρύτερο πλαίσιο, η πολυκλιμακωτή ανάλυση εικόνας αποτελεί θεμελιώδες εργαλείο για πολλές εφαρμογές της Όρασης Υπολογιστών. Σημαντικό ζήτημα στις πολυ-κλιμακωτές τεχνικές είναι η αυτόματη επιλογή της πιο σημαντικής κλίμακας σε κάθε εικόνα, πάντα σε σχέση με τις ανάγκες κάθε συγκεκριμένης εφαρμογής.

Στο παρόν Κεφάλαιο μελετούμε τη βέλτιστη επιλογή κλίμακας σε μη-γραμμικούς χώρους κλίμακας, όταν αυτοί χρησιμοποιούνται στο πρόβλημα της αποθορυβοποίησης. Προσεγγίζουμε το πρόβλημα ως ζήτημα *στατιστικής επιλογής μοντέλου* (statistical model selection) και αναζητούμε τη βέλτιστη λύση του με τη μέθοδο της στατιστικής *διεπικύρωσης* (cross-validation). Οι προτεινόμενοι αλγόριθμοι είναι πλήρως αυτόματοι, είναι συμβατοί με διαφορετικές κατανομές λευκού θορύβου, και δεν απαιτούν πρότερη γνώση της μεταβλητότητας του θορύβου. Είναι ενδιαφέρον ότι η προσέγγισή μας μπορεί να συνδυαστεί με ετερόκλητες τεχνικές δημιουργίας χώρων κλίμακας, σχηματισμένων με διάφορες γραμμικές και μη-γραμμικές εξισώσεις διάχυσης, καθώς και μη-γραμμικά φίλτρα. Έτσι η προτεινόμενη μέθοδος αποτελεί ένα ισχυρό γενικό εργαλείο που εμπλουτίζει υπάρχουσες τεχνικές αποθορυβοποίησης με δυνατότητα αυτόματης επιλογής κλίμακας. Στα πειράματά μας έχουμε επικεντρωθεί σε αποθορυβοποίηση με δύο είδη χώρων κλίμακας, παραγόμενων από διαδικασίες μη-γραμμικής διάχυσης και μορφολογικό φιλτράρισμα, αντίστοιχα, τους οποίους και θα εξετάσουμε στη συνέχεια.

4.1 Αποθορυβοποίηση με Διαδικασίες Μη-Γραμμικής Διάχυσης

Οι μη-γραμμικές *Μερικές Διαφορικές Εξισώσεις - ΜΔΕ* (Partial Differential Equations - PDE) που χρησιμοποιούνται ευρέως για αποθορυβοποίηση είναι γενικεύσεις της απλής γραμμικής και ομογενούς εξίσωσης διάχυσης θερμότητας, που αντιστοιχεί σε γραμμική συνέλιξη με ομογενή γκαουσιανά φίλτρα [Koe84]. Αντιπροσωπευτικά παραδείγματα μη-γραμμικών ΜΔΕ διάχυσης είναι το μοντέλο των Perona & Malik [PM90] και παραλλαγές του [CLMC92], καθώς και το μοντέλο ανισοτροπικής διάχυσης [Wei98]. Στα πειράματά μας χρησιμοποιούμε το μοντέλο του [CLMC92]

$$\frac{\partial u(x, t)}{\partial t} = \operatorname{div} (g(\|\nabla u_\sigma\|) \nabla u). \quad (4.1)$$

Στην παραπάνω εξίσωση το x δηλώνει το διδιάστατο διάνυσμα θέσης στο επίπεδο της εικόνας και η αρχική συνθήκη είναι $u(x, 0) = f_0(x)$, όπου $f_0(x)$ δηλώνει τη θορυβώδη εικόνα. Αυτή η διαδικασία δημιουργεί το χώρο κλίμακας $\{f_t : t \geq 0\}$, με $f_t(x) \triangleq u(x, t)$. Στην Εξ. (4.1) ο συντελεστής διάχυσης $g : [0, +\infty) \rightarrow \mathbb{R}^+$ είναι φθίνουσα συνάρτηση του μέτρου της παραγώγου της εικόνας (με $g(0) = 1$, $g(r) \rightarrow 0$ καθώς $r \rightarrow +\infty$) και έτσι ευνοεί την ομαλοποίηση εντός επιπέδων περιοχών της εικόνας, χωρίς όμως να επιτρέπει την αλλοίωση των ακμών. Τέλος, με u_σ δηλώνουμε συνέλιξη της u με πυρήνα Gauss μικρής μεταβλητότητας σ , που κανονικοποιεί το μοντέλο και το καθιστά μαθηματικά 'καλώς ορισμένο'.

4.2 Αποθορυβοποίηση με Μορφολογικά Φίλτρα

Μια ευρεία κατηγορία μορφολογικών φίλτρων επίσης παράγει μη-γραμμικούς χώρους κλίμακας με καλές ιδιότητες αποθορυβοποίησης. Το απλούστερο παράδειγμα τέτοιων φίλτρων είναι τα *ανοίγματα* (openings) και *κλεισίματα* (closings) τύπου Minkowski, που είναι απλές σειριακές εναλλαγές επιπέδων *συστολών* (erosions) και *διαστολών* (dilations) με δίσκους ως δομικά στοιχεία. Εναλλάσσοντας ανοίγματα και κλεισίματα σταδιακά μεγαλύτερης κλίμακας (δηλαδή αυξάνοντας την ακτίνα του δομικού στοιχείου) παίρνουμε το *εναλλασσόμενο σειριακό φίλτρο* (alternating sequential filter - ASF) [Ser88]. Περαιτέρω, τα ανοίγματα/κλεισίματα τύπου Minkowski στο ASF μπορούν να αντικατασταθούν από γενικευμένα ανοίγματα/κλεισίματα, για παράδειγμα τα *φίλτρα ανακατασκευής* (reconstruction filters) [SS95]. Οι περισσότεροι από τους προαναφερόμενους χώρους κλίμακας μπορούν να παραχθούν και με ΜΔΕ, κάτι που παρουσιάζει αρκετά πλεονεκτήματα σε σχέση με υλοποίηση βασισμένη σε ψηφιακά φίλτρα [Mar02]. Ένα σημαντικό μειονέκτημα των φίλτρων ανακατασκευής είναι όμως ότι δεν αποτελούν *αυτο-δυϊκούς* (self-dual) τελεστές, με αποτέλεσμα να χειρίζονται ασυμμετρικά το προσκήνιο και το παρασκήνιο (δηλαδή παίρνουμε διαφορετική έξοδο αν εφαρμόσουμε τον τελεστή στο συμπλήρωμα της εικόνας και ύστερα συμπληρώσουμε το αποτέλεσμα).

Ένας πιο μοντέρνος μορφολογικός τελεστής που ενοποιεί και γενικεύει τα ανοίγματα/κλεισίματα ανακατασκευής και παράλληλα έχει την ιδιότητα της αυτο-δυϊκότητας είναι το *φίλτρο επιπέδωσης* (leveling) [MM00]. Τα φίλτρα επιπέδωσης είναι μη-γραμμικά αντικειμενοστρεφή φίλτρα που απλοποιούν μια εικόνα αναφοράς f μέσω ταυτόχρονης

επέκτασης και συρρίκνωσης μίας αρχικής εικόνας-σπόρου που αποκαλείται σημαδευτής m , με τρόπο που να περιορίζεται από την εικόνα αναφοράς. Πιο συγκεκριμένα, έστω $u(x, t)$ η εξέλιξη του σημαδευτή, ξεκινώντας από $u(x, 0) = m(x)$. Ορίζουμε το u ως την ασθενή λύση της ΜΔΕ [MM00, Mar03]

$$\partial u(x, t)/\partial t = -\text{sgn}(u - f)\|\nabla u\|. \quad (4.2)$$

Αυτή η ΜΔΕ έχει μια μη-τετριμμένη τελική κατάσταση $\Lambda(m|f) \triangleq \lim_{t \rightarrow \infty} u(x, t)$, που καλείται επιπέδωση της f ως προς το m [Mar03].

Ας θεωρήσουμε τώρα διάφορους σημαδευτές m_t , $t = 1, 2, \dots$, που εξαρτώνται από μια αύξουσα παράμετρο κλίμακας t και ας κατασκευάσουμε μια ακολουθία από επιπεδώσεις $f_t = \Lambda(m_t|f_{t-1})$, $t > 1$ της εικόνας ξεκινώντας από τη θορυβώδη εκδοχή της $f_0(x)$. Οι εικόνες $\{f_t : t \geq 0\}$ αποτελούν μια ιεραρχία από πολυ-κλιμακωτές επιπεδώσεις της αρχικής εικόνας f_0 που ικανοποιούν τη σχέση αιτιότητας ότι το f_j είναι επιπέδωση της f_i για $j > i$. Ένας απλός τρόπος να κατασκευάσουμε στην πράξη τέτοιες πολυ-κλιμακωτές επιπεδώσεις είναι να χρησιμοποιήσουμε ως ακολουθία πολυ-κλιμακωτών σημαδευτών εικόνες προερχόμενες από δειγματοληψία του γραμμικού χώρου κλίμακας της f_0 .

4.3 Αυτόματη Επιλογή Κλίμακας με Στατιστική Διεπικύρωση

Όλες οι παραπάνω πολυ-κλιμακωτές μέθοδοι αποθορυβοποίησης ξεκινούν από μία θορυβώδη εικόνα f_0 και παράγουν μια ιεραρχία από απλοποιημένες εκδοχές της $\{f_t : t \geq 0\}$ με δείκτη μια αύξουσα παράμετρο κλίμακας t . Ένα σημαντικό ζήτημα που ανακύπτει είναι η βέλτιστη επιλογή της κλίμακας όπου θα σταματήσουμε την απλοποίηση και θα λάβουμε την καλύτερη δυνατή εκτίμηση για την καθαρή εικόνα. Στο ένα άκρο, σε πολύ μικρές κλίμακες ο θόρυβος δεν έχει ακόμα καταπιεστεί επαρκώς· με στατιστικούς όρους η εκτίμηση χαρακτηρίζεται από μεγάλη *μεταβλητότητα* (variance). Στο άλλο άκρο, σε πολύ μεγάλες κλίμακες ο θόρυβος έχει εξαλειφθεί, αλλά όμως δομικά στοιχεία της εικόνας έχουν αλλοιωθεί σημαντικά· η εκτίμηση χαρακτηρίζεται από ισχυρή *πόθωση* (bias).

Από τη φύση τους οι περισσότερες πολυ-κλιμακωτές αναπαραστάσεις δεν διαθέτουν κάποια εγγενή χαρακτηριστική κλίμακα κατάλληλη για το πρόβλημα της αποθορυβοποίησης. Για παράδειγμα, η εξέλιξη της μη-γραμμικής εξίσωσης διάχυσης (4.1) οδηγεί τελικά στην τετριμμένη τελική κατάσταση σταθερής εικόνας. Ακόμα και όταν ο συντελεστής διάχυσης g της Εξ. (4.1) επιλεγεί έτσι ώστε η τελική κατάσταση f_∞ να είναι μη-τετριμμένη [BSMH98, GZS01], δεν υπάρχει καμία εγγύηση ότι η προκύπτουσα εικόνα f_∞ είναι βέλτιστη για το πρόβλημα της αποθορυβοποίησης για κάποιο αποδεκτό κριτήριο. Είναι λοιπόν ξεκάθαρο ότι απαιτείται μια αυτόματη μέθοδος που θα επιλέγει το βέλτιστο στιγμιότυπο του χώρου κλίμακας f_t ως καλύτερο εκτιμητή της καθαρής εικόνας, επιτρέποντας την εξέλιξη του χώρου κλίμακας για επαρκή χρόνο ώστε να καταπιεστεί αρκετά ο θόρυβος, χωρίς όμως να αλλοιωθεί ιδιαίτερα η εικόνα. Αξίζει να σημειωθεί εδώ ότι η ανάγκη για ρητούς μηχανισμούς επιλογής κλίμακας έχει τονιστεί από διάφορους ερευνητές (δες π.χ. [Lin98, EZ98]), αν και οι προηγούμενες μελέτες δεν αφορούν το πρόβλημα της αποθορυβοποίησης.

Στην προσέγγισή μας θεωρούμε το πρόβλημα επιλογής βέλτιστης κλίμακας για αποθορυβοποίηση ως πρόβλημα *στατιστικής επιλογής μοντέλου* (statistical model selection

problem). Επιχειρηματολογούμε ότι μια επιθυμητή στρατηγική επιλογής κλίμακας θα πρέπει να προτιμά εκείνο το στιγμιότυπο από το χώρο κλίμακας που ελαχιστοποιεί την αναμενόμενη απόσταση από την (γενικά άγνωστη) καθαρή εικόνα, με την απόσταση να ορίζεται με βάση κάποια *συνάρτηση απωλειών* (loss-function) κατάλληλη για το συγκεκριμένο πρόβλημα. Οι προτεινόμενοι αλγόριθμοι προσεγγίζουν αυτή την επιθυμητή στρατηγική επιλογής κλίμακας με χρήση της τεχνικής της στατιστικής διεπικύρωσης και αποδεικνύονται ιδιαίτερα αποδοτικοί και εύρωστοι στην πράξη. Η προσέγγισή μας είναι στο ίδιο μήκος κύματος με τη μέθοδο που προτείνεται στο [Nas96], όπου παρόμοιες τεχνικές διεπικύρωσης χρησιμοποιούνται για τη βέλτιστη επιλογή κατωφλίου σε μεθόδους αποθορυβοποίησης με κατωφλιοποίηση κυματιδίων.

4.3.1 Ορισμός Προβλήματος

Θεωρούμε μια θορυβώδη γκριζα εικόνα f_0 αποτελούμενη από $M = M_x M_y$ εικονοστοιχεία. Ορίζουμε μια λεξικογραφική διάταξη των εικονοστοιχείων και συμβολίζουμε με $f_0(x_i)$ τη φωτεινότητα στο i -στό εικονοστοιχείο, με $1 \leq i \leq M$. Θεωρούμε ότι το f_0 είναι δείγμα μιας στοχαστικής διαδικασίας F_0 , που μοντελοποιούμε ως εξής:

$$F_0(x_i) = f^*(x_i) + N(x_i), \quad i = 1, \dots, M, \quad (4.3)$$

όπου το f^* δηλώνει τη (συνήθως άγνωστη) ‘καθαρή’ εικόνα και N είναι μια διαδικασία λευκού θορύβου, με μηδενική μέση τιμή και μεταβλητότητα σ^2 . Έχοντας τη θορυβώδη εικόνα f_0 ως αρχική συνθήκη, τα στιγμιότυπα του χώρου κλίμακας $\{f_t : t \geq 0\}$, με δείκτη την αύξουσα παράμετρο κλίμακας t , είναι υποψήφια ως εκτιμητές της f^* .

Για να βρούμε το βέλτιστο εκτιμητή της f^* ανάμεσα στα υποψήφια στιγμιότυπα f_t , πρέπει να ορίσουμε ένα κατάλληλο κριτήριο βελτιστότητας. Ας υποθέσουμε ότι παράγουμε ένα νέο δείγμα $f_0^{new} = f^* + n^{new}$ της θορυβώδους εικόνας, με $n^{new}(x_i)$ ανεξάρτητα αλλά ομοίως κατανομημένα με τα $n(x_i)$. Ορίζουμε το (ευδο-δειγματικό) σφάλμα πρόβλεψης ((in-sample) prediction error) του μοντέλου f_t ως:

$$PE(t) \equiv PE(f_t) = E_N\{L(f_0^{new}, f_t)\}, \quad (4.4)$$

παίρνοντας αναμενόμενη τιμή ως προς τη διαδικασία θορύβου $\{n^{new}(x_i)\}$. το L είναι μια (κανονικοποιημένη ανά εικονοστοιχείο) συνάρτηση απωλειών που τιμωρεί την απόκλιση μεταξύ f_0^{new} και f_t . Τυπικά παραδείγματα για τη συνάρτηση απωλειών είναι οι νόρμες $L_p(f, f_0) = \left(\frac{1}{M} \sum_{i=1}^M |f(x_i) - f_0(x_i)|^p\right)^{1/p}$, για $p \in \{1, 2\}$, αν και για κάποιες εφαρμογές άλλες αντιληπτικά πιο εύλογες επιλογές για τη συνάρτηση απωλειών [WBSS04] μπορεί να είναι καταλληλότερες. Είναι λογικό να ορίσει κανείς το βέλτιστο χρόνο διακοπής (optimal stopping time) t^* ως την κλίμακα για την οποία το $PE(t)$ ελαχιστοποιείται, δηλαδή $t^* = \operatorname{argmin}_{t \geq 0} PE(t)$.

Μια άλλη ποσότητα χρήσιμη στην ανάλυσή μας είναι το σφάλμα μοντέλου (model error), οριζόμενο ως $ME(t) \equiv ME(f_t) = L(f^*, f_t)$. Για πολύ ασθενή θόρυβο ($\sigma \rightarrow 0$), τα $ME(t)$ και $PE(t)$ συμπίπτουν. Διαφορετικά, η επιπλέον αβεβαιότητα εξαιτίας του θορύβου συνεπάγεται ότι $PE(t) > ME(t)$. Για παράδειγμα, στην περίπτωση τετραγωνικής συνάρτησης απωλειών L_2 και λευκού θορύβου μεταβλητότητας σ^2 , μπορούμε εύκολα να δείξουμε ότι $PE(t) = ME(t) + \sigma^2$. Η χρησιμότητα του $ME(t)$ είναι στην πράξη περιορισμένη, καθώς η καθαρή εικόνα f^* είναι συνήθως άγνωστη. Παρόλα αυτά, μπορούμε

να χρησιμοποιήσουμε το $ME(t)$ και την κλίμακα $t_{ME}^* = \operatorname{argmin}_{t \geq 0} ME(t)$ που αντιστοιχεί στο ελάχιστο ως αναφορά στην περίπτωση διαθέσιμης καθαρής εικόνας f^* , όπως στα παραδείγματα που θα δώσουμε στη συνέχεια, στα οποία προσθέτουμε σε γνωστές καθαρές εικόνες f^* συνθετικό θόρυβο.

Το (εξω-δειγματικό) σφάλμα πρόβλεψης ((extra-sample) prediction error), οριζόμενο από τη σχέση

$$PE_+(t) \equiv PE_+(f_t) = E_{N,X} \{L(f_0^{new}, f_t)\}, \quad (4.5)$$

διαφοροποιείται από το $PE(t)$ γιατί θεωρεί τις θέσεις X των εικονοστοιχείων στο νέο δείγμα εικόνας f_0^{new} ως τυχαίες μεταβλητές [Efr86, Bre92]. Τούτο σημαίνει ότι για το $PE_+(t)$ μπορεί να χρειαστεί να παρεμβάλουμε τις τιμές τις εικόνας σε ενδιάμεσα σημεία έξω από το πλέγμα δειγματοληψίας $\{x_i : 1 \leq i \leq M\}$ της αρχικής εικόνας, όπως κάνουμε για παράδειγμα στο πρόβλημα της μεγέθυνσης εικόνας. Η επιπλέον αβεβαιότητα παρεμβολής σε αυτό το σενάριο σημαίνει ότι $PE_+(t) \geq PE(t)$, καθότι στο $PE_+(t)$ συναθροίζονται η αβεβαιότητα λόγω θορύβου και σφάλματος μοντέλου (που περιλαμβάνονται στο $PE(t)$) με το επιπλέον σφάλμα παρεμβολής. Όμως για την εφαρμογή μας δεν μας ενδιαφέρει αυτό καθαυτό το $PE(t)$, αλλά η κλίμακα t^* που το $PE(t)$ ελαχιστοποιείται. Υποθέτοντας ότι οι όροι σφάλματος παρεμβολής στο $PE_+(t)$ για τα διαφορετικά μοντέλα $\{f_t : t \geq 0\}$ αλληλοακυρώνονται, είναι θεμιτό να θεωρήσουμε ότι $t^* \approx t_+^*$, όπου $t_+^* = \operatorname{argmin}_{t \geq 0} PE_+(t)$ είναι η κλίμακα που ελαχιστοποιείται το $PE_+(t)$. Το όφελος αυτής της παραδοχής είναι ότι το $PE_+(t)$ μπορεί να εκτιμηθεί απευθείας από το μοναδικό διαθέσιμο δείγμα της θορυβώδους εικόνας f_0 με τεχνικές διεπικύρωσης, όπως θα δούμε στη συνέχεια.

Έχοντας εισάγει τον αναγκαίο συμβολισμό, δίνουμε μια σύντομη ανασκόπηση ορισμένων εναλλακτικών προσεγγίσεων στο πρόβλημα επιλογής κλίμακας, επικεντρωνόμενοι στη βιβλιογραφία πάνω σε χώρους κλίμακας. Στο [Wei99] ο Weickert επιλέγει την κλίμακα t_{snr}^* που ικανοποιεί τη σχέση $\frac{\operatorname{var}(f_{t_{\text{snr}}^*})}{\operatorname{var}(f_0)} = \frac{1}{1+1/\text{snr}}$, υποθέτοντας γνωστό σηματοθορυβικό λόγο snr και θεωρώντας ότι το φίλτρο διάχυσης είναι τόσο αποδοτικό που πρακτικά το $f_{t_{\text{snr}}^*}$ είναι άριστη προσέγγιση του f^* . Ο Mrazek στο [MN03] προτείνει ένα κριτήριο αποσυσχέτισης, επιλέγοντας την κλίμακα t_{dec}^* που ελαχιστοποιεί το συντελεστή συσχέτισης $t_{\text{dec}}^* = \operatorname{argmin}_{t \geq 0} \frac{\operatorname{cov}(f_0 - f_t, f_t)}{\sqrt{\operatorname{var}(f_0 - f_t)\operatorname{var}(f_t)}}$, θεωρώντας το $f_0 - f_t$ ως 'θόρυβο' και $f_{t_{\text{dec}}^*} \approx f^*$. Όμως το κριτήριο αποσυσχέτισης δε φαίνεται να συνδέεται με οποιοδήποτε κριτήριο στατιστικής βελτιστότητας, κάτι που εξάλλου παραδέχονται οι συγγραφείς του [MN03]. Τέλος, ο Solo στο [Sol01], υποθέτοντας λευκό προσθετικό θόρυβο Gauss γνωστής ισχύος και τετραγωνική ποινή, φτάνει σε ένα κριτήριο τύπου SURE [Ste81] για την επιλογή της κλίμακας που ελαχιστοποιεί το $PE(t)$. Από τις εναλλακτικές προσεγγίσεις που μόλις περιγράψαμε, καμιά δεν μπορεί να χειριστεί αυθαίρετες συναρτήσεις απωλειών, ενώ οι δύο πρώτες δεν θεμελιώνονται σε στατιστικά επιχειρήματα. Οι αλγόριθμοι διεπικύρωσης που προτείνουμε στη συνέχεια ξεπερνούν και τους δύο αυτούς περιορισμούς.

4.3.2 Προτεινόμενος Αλγόριθμος Επιλογής Κλίμακας

Η διεπικύρωση είναι μια μη-παραμετρική μέθοδος για την απευθείας εκτίμηση από τα δεδομένα του εξω-δειγματικού σφάλματος πρόβλεψης $PE_+(t)$ της Εξ. (4.5) [Sto74]. Οι τεχνικές διεπικύρωσης για επιλογή μοντέλου είναι πολύ γενικές, με την έννοια ότι μπορούν να εφαρμοστούν με επιτυχία σε συνδυασμό με οποιαδήποτε συνάρτηση απωλειών L ή μη-γραμμικό μοντέλο εκτίμησης, σε αντίθεση με εναλλακτικές μεθόδους για επιλο-

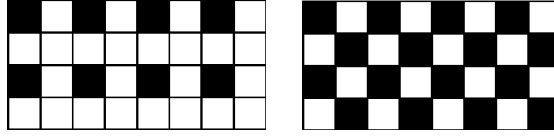
γή μοντέλου, όπως αυτές που βασίζονται στο *Κριτήριο Πληροφορίας του Akaike* (Akaike Information Criterion - AIC), τη *Cp Στατιστική του Mallows* (Mallows' Cp Statistic), ή το *Μπεϋζιανό Κριτήριο Πληροφορίας* (Bayesian Information Criterion - BIC), που τυπικά είναι εφαρμόσιμα μόνο στα πλαίσια του γραμμικού μοντέλου εκτίμησης, τετραγωνικής συνάρτησης απωλειών, και γνωστής ισχύος θορύβου [HTF01]. Για αυτούς τους λόγους, μέθοδοι διεπικύρωσης είναι ιδιαίτερα ελκυστικές στο πλαίσιο του προβλήματος της επιλογής κλίμακας για αποθορυβοποίηση με μη-γραμμικούς χώρους κλίμακας, χωρίς πρότερη γνώση για την ισχύ του θορύβου. Εδώ πρέπει να υπογραμμίσουμε ένα σημαντικό περιορισμό των αλγορίθμων διεπικύρωσης που προτείνουμε: η εφαρμογή τους προϋποθέτει ότι ο θόρυβος σε γειτονικά εικονοστοιχεία είναι ασυσχέτιστος. Διαφορετικά, έχει δείχτει στο [Alt90], στο πλαίσιο της επιλογής κλίμακας σε μεθόδους ομαλοποίησης με πυρήνα kernel smoothing, ότι στην περίπτωση που ο θόρυβος σε γειτονικά εικονοστοιχεία είναι θετικά - αρνητικά - συσχετισμένος, τότε τα μοντέλα που επιλέγονται από μη-αντισταθμισμένη διεπικύρωση τείνουν να *υπερειδικεύονται* (overfit) - *υποειδικεύονται* (underfit) - στα δεδομένα μας (δες και το [Nas96]).

Για να εκτιμήσουμε το $PE_+(t)$ με διεπικύρωση, επαναδειγματοληπούμε τα θορυβώδη δεδομένα μας (τιμές της εικόνας) $\mathcal{D} = \{(x_i, f_0(x_i)) : 1 \leq i \leq M\}$. Συγκεκριμένα, παρακρατούμε μέρος των δεδομένων μας, έστω $\mathcal{T}_1 \subset \mathcal{D}$. Χρησιμοποιούμε τα εναπομείναντα δεδομένα $\mathcal{D} - \mathcal{T}_1$ για τη δημιουργία ενός μη-γραμμικού χώρου κλίμακας $f_t^{-\mathcal{T}_1}$. Η ικανότητα γενίκευσης του $f_t^{-\mathcal{T}_1}$ μπορεί να αξιολογηθεί στο παρακρατηθέν σύνολο ελέγχου \mathcal{T}_1 μέσω της $\widehat{PE}_+(f_t^{-\mathcal{T}_1}) = L(f_0^{\mathcal{T}_1}, f_t^{-\mathcal{T}_1})$, καθότι το \mathcal{T}_1 είναι *ανεξάρτητο* από τα δεδομένα με τα οποία χτίστηκε το μοντέλο. Επαναλαμβάνοντας K φορές για διαφορετικές διαμερίσεις σε υποσύνολα $\mathcal{T}_1, \dots, \mathcal{T}_K$ με παρόμοιο πλήθος στοιχείων και σταθμίζοντας τα μερικά αποτελέσματα προκύπτει η εκτίμηση

$$\widehat{PE}_+(t) = PE^{CV}(t) = \frac{1}{K} \sum_{k=1}^K \widehat{PE}_+(f_t^{-\mathcal{T}_k}). \quad (4.6)$$

Για να οδηγηθούμε σε πρακτικούς αλγορίθμους για την εφαρμογή μας, πρέπει πρώτα να καθορίσουμε τα σύνολα επαναδειγματοληψίας $\{\mathcal{T}_i : 1 \leq i \leq K\}$. Αυτό το ζήτημα έχει απασχολήσει αρκετά τη σχετική βιβλιογραφία, δες π.χ. το [Koh95]. Για κάποια προβλήματα επιλογής κλίμακας, όπως η επιλογή σταθεράς κανονικοποίησης σε *σπλίνες ομαλοποίησης* (smoothing splines), η ειδική περίπτωση της διεπικύρωσης *με-εξάιρεση-ενός* (leave-one-out) (όπου $K = M$ και το \mathcal{T}_i περιέχει μόνο το $(x_i, f(x_i))$) μπορεί να προσεγγιστεί αναλυτικά, οδηγώντας σε λύση που δίνεται σε κλειστό τύπο [HTF01], κάτι που δυστυχώς δεν είναι εφικτό για την περίπτωση των μη-γραμμικών χώρων κλίμακας που εξετάζουμε. Προτείνουμε δύο κατάλληλες, εναλλακτικές μεταξύ τους, στρατηγικές επαναδειγματοληψίας, που στο εφεξής τις καλούμε *quadruple-cv* και *double-cv*. Στο Σχ. 4.1 δείχνουμε για καθεμιά από αυτές τα εικονοστοιχεία $\mathcal{D} - \mathcal{T}_i$ που χρησιμοποιούμε για τη δημιουργία του χώρου κλίμακας (μαύρο χρώμα), καθώς και τα εικονοστοιχεία \mathcal{T}_i στα οποία αξιολογούμε τη γενικευτική ικανότητα του μοντέλου (λευκό χρώμα) για την περίπτωση $i = 1$.

Στην εκδοχή *quadruple-cv*, επιλέγοντας μία από τις τέσσερις εκδοχές $\{(0,0), (0,1), (1,0), (1,1)\}$ για το διάνυσμα μετατόπισης (s_i, s_j) , προκύπτουν τέσσερις υποδειγματοληπτημένες, κατά ένα παράγοντα 2 ανά κατεύθυνση, εκδοχές της θορυβώδους εικόνας f_0 . Η καθεμιά αποτελείται από περίπου $M/4$ εικονοστοιχεία με συντεταγμένες



Σχήμα 4.1: Εναλλακτικά σχέδια επαναδειγματοληψίας. Το \mathcal{T}_1 δίνεται με λευκό. *Αριστερά*: Quadruple-CV. *Δεξιά*: Double-CV.

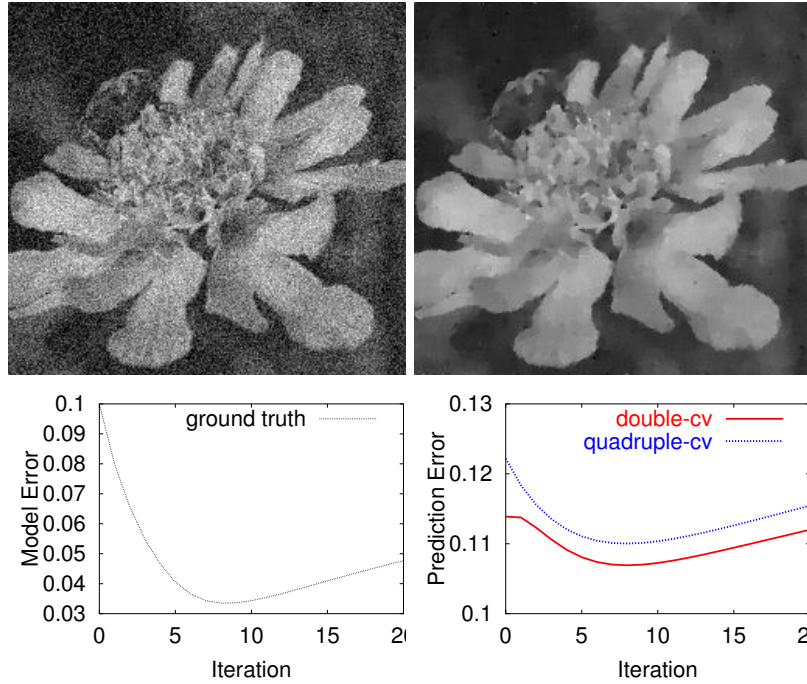
$\{(2i + s_i, 2j + s_j) : 0 \leq i < M_x/2, 0 \leq j < M_y/2\}$. Κατόπιν χτίζουμε το χώρο κλίμακας για καθεμιά από τις τέσσερις υποδειγματοληπτημένες εικόνες, κατάλληλα κλιμακώνοντας τις ΜΔΕ ή τα δομικά στοιχεία των μορφολογικών φίλτρων. Στη συνέχεια παίρνουμε το $\text{PE}_{\text{quadruple-cv}}^{CV}(t)$ από την Εξ. (4.6), υπολογίζοντας καθένα από τους τέσσερις όρους $\{\widehat{\text{PE}}_+(f_t^{-\mathcal{T}_k}) : 1 \leq k \leq 4\}$ ως εξής: Παίρνουμε το αντίστοιχο στιγμιότυπο του βοηθητικού χώρου κλίμακας, διαστάσεων $M_x/2$ επί $M_y/2$, και στη συνέχεια παρεμβάλλουμε τις τιμές στα υπόλοιπα $3M/4$ εικονοστοιχεία $p_i \in \mathcal{T}_k$ (με απλή γραμμική παρεμβολή στην υλοποίησή μας). Κατόπιν τιμωρούμε (με τη συνάρτηση απωλειών L) τη διαφορά μεταξύ των παρεμβαλλόμενων τιμών και της αρχικής θορυβώδους εικόνας $f_0(x_i)$ στα ίδια εικονοστοιχεία p_i και παίρνουμε το μέσο όρο σε όλα τα $3M/4$ εικονοστοιχεία του \mathcal{T}_k για να παράγουμε το $\widehat{\text{PE}}_+(f_t^{-\mathcal{T}_k})$. Το επιπλέον υπολογιστικό φορτίο είναι περίπου ίσο με το κόστος παραγωγής του κύριου χώρου κλίμακας, αφού χρειάζεται να παραχθούν 4 βοηθητικοί χώροι κλίμακας μεγέθους $M/4$ εικονοστοιχείων ο καθένας.

Η κύρια διαφορά στην περίπτωση του *double-cv* είναι ότι τα εικονοστοιχεία που τροφοδοτούν τους δύο βοηθητικούς χώρους κλίμακας δεν βρίσκονται τοποθετημένα σε τετραγωνικό πλέγμα (δες Σχ. 4.1, δεξιά). Για αυτό, είναι βολικό πρώτα να παρεμβάλλουμε τις τιμές στα $M/2$ ‘λευκά’ εικονοστοιχεία του \mathcal{T}_k από τις τιμές των υπολοίπων $M/2$ ‘μαύρων’ εικονοστοιχείων της θορυβώδους εικόνας f_0 και στη συνέχεια να δημιουργήσουμε τους βοηθητικούς χώρους κλίμακας. Το επιπλέον υπολογιστικό κόστος είναι τώρα διπλάσιο της *quadruple-cv* εκδοχής.

4.3.3 Πειράματα και Συγκρίσεις

Ένα παράδειγμα αποθορυβοποίησης εικόνας με αυτόματη επιλογή κλίμακας μέσω διεπικύρωσης φαίνεται στο Σχ. 4.2. Η πρώτη γραμμή δείχνει τη θορυβώδη εικόνα και την αποθορυβοποιημένη εκδοχή της σε κλίμακα $t_{\text{double-cv}}^*$ επιλεγμένη από τον αλγόριθμο *double-cv*. Οι αντίστοιχες γραφικές παραστάσεις του $\text{ME}(t)$ (με γνώση αληθούς καθαρής εικόνας) και $\text{PE}_+(t)$ (όπως εκτιμήθηκε από τους δύο αλγόριθμους διεπικύρωσης) φαίνονται στη δεύτερη γραμμή. Αξίζει να παρατηρήσουμε ότι αν και το $\text{ME}(t)$ είναι μικρότερο $\text{PE}_+(t)$, και τα δύο παρουσιάζουν ελάχιστο περίπου στην ίδια κλίμακα (μετά από 8 επαναλήψεις). Περαιτέρω παραδείγματα αποθορυβοποίησης μιας εικόνας MRI με μη-γραμμική διάχυση και μιας αεροφωτογραφίας με φίλτρο επιπέδωσης, δίνονται στο Σχ. 4.3. Επιπλέον παραδείγματα εφαρμογής της μεθόδου σε φωτογραφίες τοιχογραφιών από το Ακρωτήρι παρουσιάζονται στο Σχ. 4.4.

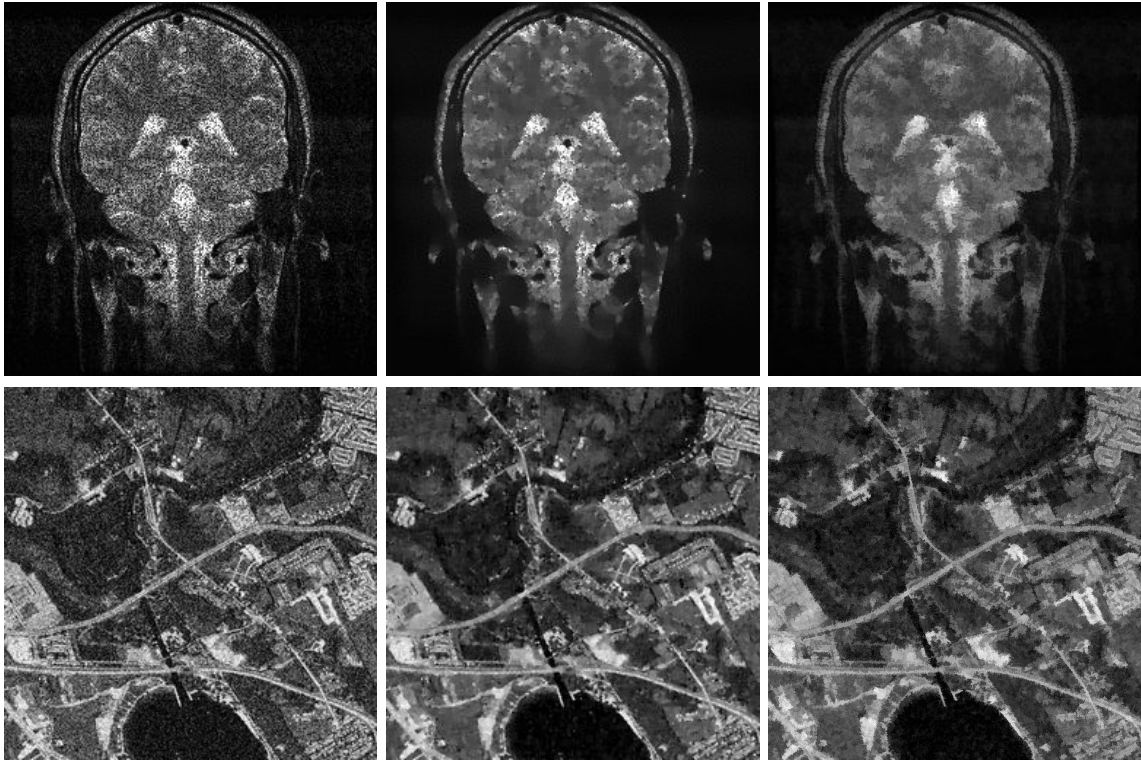
Για τη συστηματική συγκριτική αξιολόγηση της απόδοσης των προτεινόμενων και παλαιότερων αλγορίθμων, τρέξαμε μια σειρά από πειράματα αποθορυβοποίησης σε μία βάση 39 γκριζων φυσικών εικόνων (συλλογές *kodak*, *aerial* και *misc1* από το <http://www.cipr.rpi.edu/resource/stills/>), αλλοιωμένων με συνθετικό θόρυβο έ-



Σχήμα 4.2: Παράδειγμα αποθορυβοποίησης με αυτόματη επιλογή κλίμακας. Πάνω: Η θορυβώδης εικόνα και η καθαρισμένη εκδοχή της σε αυτόματα επιλεγμένη κλίμακα. Κάτω: σφάλμα μοντέλου $ME(t)$ και εκτίμηση του σφάλματος πρόβλεψης $PE_+(t)$ με τους δύο αλγορίθμους διεπικύρωσης.

τσι ώστε να έχουμε διαθέσιμη την αληθινή καθαρή εικόνα f^* . Εκτός από τους δύο προτεινόμενους διεπικυρωτικούς αλγορίθμους (*double-cv* και *quadruple-cv*), έχουμε υλοποιήσει και τη βασισμένη στο σηματοθορυβικό λόγο μέθοδο του Weickert και τη μέθοδο αποσυσχέτισης του Mrazek, που τις καλούμε στη συνέχεια *snr* και *dec*, αντίστοιχα. Ο χώρος κλίμακας που χρησιμοποιείται στο πείραμα αυτό παράγεται από τη μη-γραμμική ΜΔΕ διάχυσης (4.1), με $g(r) = 1/[1 + (r/\lambda)^2]$ [PM90], $\lambda = 0.01$ (οι τιμές της εικόνας ήταν μεταξύ 0 και 1), $\sigma = 2$ εικονοστοιχεία, και διακριτοποίηση βασισμένη στο AOS αριθμητικό σχήμα [WtHRV98].

Πειραματιστήκαμε με τρία διαφορετικά είδη θορύβου $nt \in \{ \text{gaussian, salt\&pepper, speckle} \}$. Οι αλλοιωμένες εικόνες δίνονται αντίστοιχα από τις σχέσεις $f_0(x_i) = f^*(x_i) + \epsilon_i$ (gaussian), $f_0(x_i) = (1 + \epsilon_i)f^*(x_i)$ (speckle), και $f_0(x_i) = f^*(x_i)$, με πιθανότητα $1 - p$ και $f_0(x_i) = 0$ ή 1 με πιθανότητα $p/2$ το καθένα (salt&pepper). Στις περιπτώσεις θορύβου gaussian και speckle, τα ϵ_i είναι i.i.d. δείγματα από την $N(0, \sigma^2)$. Κάναμε πειράματα για διάφορα επίπεδα του θορύβου $nl \in \{0.05, 0.1, 0.15, 0.2, 0.3, 0.4\}$, με $nl = \sigma$ στις περιπτώσεις θορύβου gaussian και speckle και $nl = p$ για θόρυβο salt&pepper και για δύο επιλογές της συνάρτησης απωλειών $L \in \{L_1, L_2\}$. Για την αξιολόγηση της επίδοσης των αλγορίθμων τρέξαμε πειράματα στις 39 εικόνες της βάσης και για τους $2 \cdot 3 \cdot 6 = 36$ συνδυασμούς των (L, nt, nl) , και στη συνέχεια υπολογίσαμε το μέσο όρο της *σχετικής αύξησης του σφάλματος μοντελοποίησης*, που ορίζεται από $\frac{\Delta(ME)}{ME} = \frac{ME(t_{\text{alg}}^*) - ME(t_{\text{ME}}^*)}{ME(t_{\text{ME}}^*)}$ και δηλώνει πόσο χειρότερα αποδίδει ο αλγόριθμος από το ground-truth (μικρό $\frac{\Delta(ME)}{ME}$ είναι επιθυμητό).

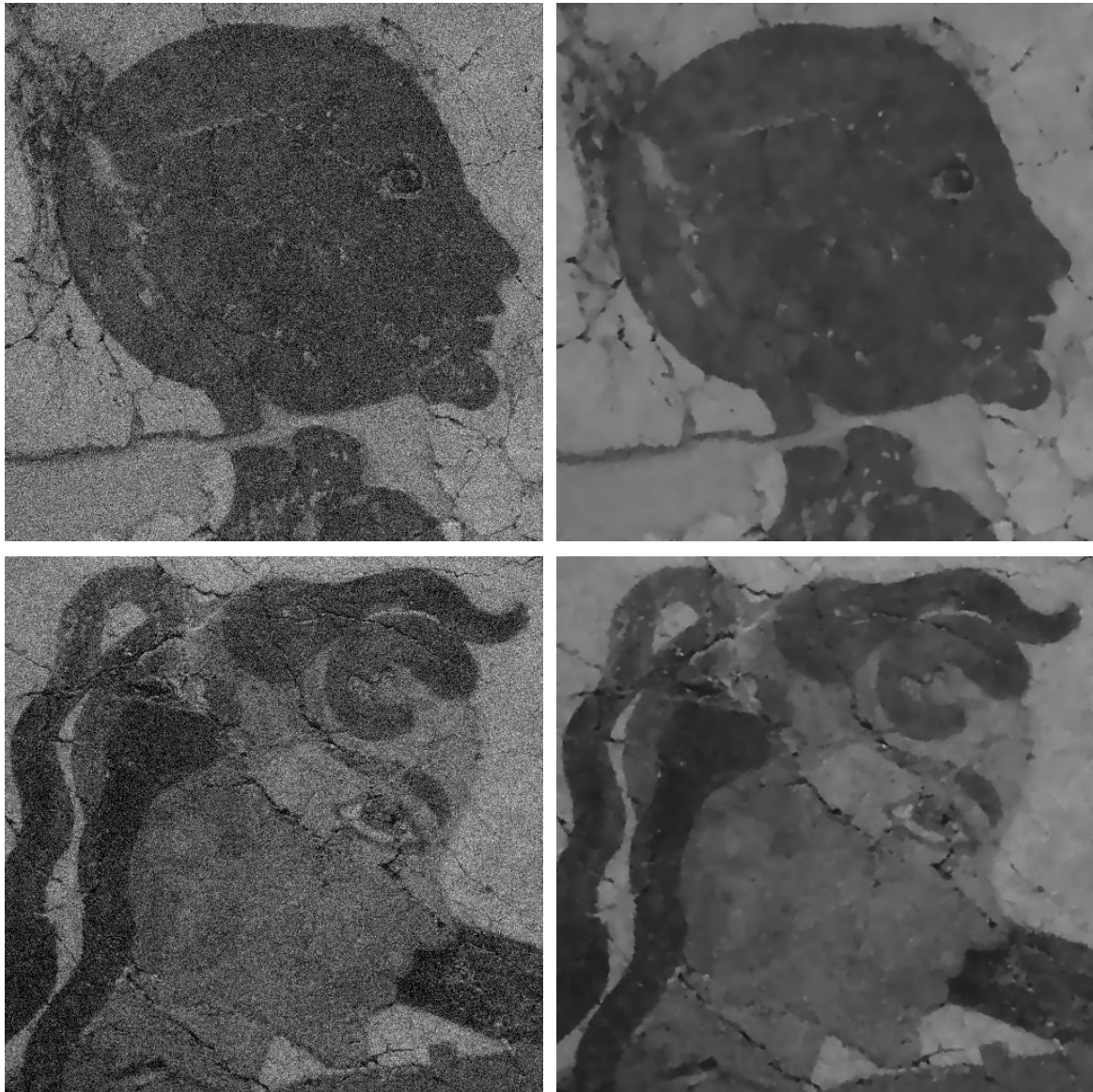


Σχήμα 4.3: Θορυβώδης εικόνα (αριστερά), καθαρισμένη με μη-γραμμική διάχυση (κέντρο) και φίλτρο επιπέδωσης (δεξιά) σε αυτόματα επιλεγμένες κλίμακες. Πάνω: Εικόνα MRI, πολλαπλασιαστικός θόρυβος τύπου speckle. Κάτω: Αεροφωτογραφία, γκαουσιανός θόρυβος.

Στο Σχ. 4.5 παρουσιάζουμε μέσες τιμές του $\frac{\Delta(ME)}{ME}$. Παρατηρούμε ότι οι *double-cv* και *quadruple-cv* δίνουν αποτελέσματα ορισμένες φορές μια τάξη μεγέθους καλύτερα από τους εναλλακτικούς *snr* και *dec* αλγορίθμους. Η επίδοση των διεπικυρωτικών αλγορίθμων είναι ιδιαίτερα εντυπωσιακή στην περίπτωση γκαουσιανού θορύβου, όπου το $\frac{\Delta(ME)}{ME}$ τους είναι σχεδόν πάντα μικρότερο από 1%. Η ευρωστία των αλγορίθμων διεπικύρωσης, ανεξάρτητα από το είδος του θορύβου ή τη συνάρτηση απωλειών είναι επίσης αξιοσημείωτη. Μεταξύ των άλλων δύο αλγορίθμων, ο *snr* φαίνεται να αποδίδει καλύτερα από τον *dec*. Ο αλγόριθμος αποσυσχέτισης *dec* δεν απέδωσε καλά στα πειράματά μας, ενώ και στο [MN03] παρατηρείται ότι δεν αποδίδει καλά σε συνδυασμό με τη συγκεκριμένη μη-γραμμική ΜΔΕ διάχυσης (4.1).

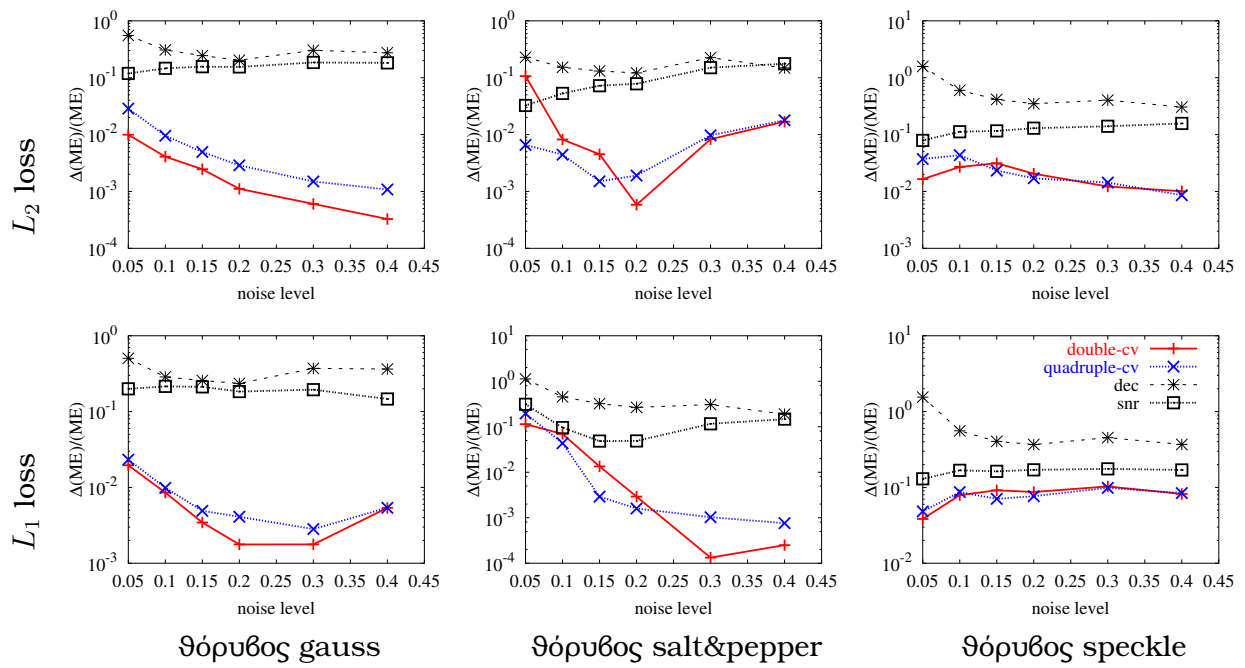
4.3.4 Συμπεράσματα

Η μελέτη μας έχει δείξει την αξία θεώρησης του προβλήματος της βέλτιστης επιλογής κλίμακας για αποθορυβοποίηση με μη-γραμμικούς χώρους κλίμακας από την οπτική γωνία της στατιστικής επιλογής μοντέλου. Μέσα από αυτήν τη θεώρηση έχουμε οδηγηθεί σε ιδιαίτερα αποδοτικούς αλγορίθμους επιλογής κλίμακας μέσω στατιστικής διεπικύρωσης. Είναι αξιοσημείωτο ότι αυτοί οι αλγόριθμοι είναι πλήρως αυτόματοι, είναι συμβατοί με διαφορετικές κατανομές λευκού θορύβου, και δεν απαιτούν πρότερη γνώση της με-



Σχήμα 4.4: Θορυβώδης εικόνα (συνθετικός προσθετικός θόρυβος Gauss) (αριστερά), καθαρισμένη με μη-γραμμική διάχυση (δεξιά) σε αυτόματα επιλεγμένες κλίμακες. Πάνω: Λεπτομέρεια από την τοιχογραφία 'Ακολουθία Ανδρών', Σχ. 2.5(β). Κάτω: Λεπτομέρεια από την τοιχογραφία 'Πυγμαχοί', Σχ. 2.5(στ).

ταβλητότητας του θορύβου. Η προτεινόμενη μεθοδολογία οδηγεί σε τεχνικές επιλογής κλίμακας που μπορούν εύκολα να συνδυαστούν ουσιαστικά με κάθε μέθοδο αποθορυβοποίησης που βασίζεται σε ανάλυση σε πολλαπλές κλίμακες.



Σχήμα 4.5: Σχετική αύξηση του σφάλματος μοντελοποίησης $\Delta(\text{ME})/\text{ME}$ ως συνάρτηση του επιπέδου του θορύβου nl για τους εξεταζόμενους αλγορίθμους επιλογής κλίμακας. Μέσα αποτελέσματα σε 39 εικόνες. (Λογαριθμική κλίμακα στον κατακόρυφο άξονα, υπόμνημα στη γραφική παράσταση κάτω-δεξιά.)

Κεφάλαιο 5

Αναπαράσταση και Στατιστικές Ιδιότητες Φυσικών Εικόνων στο Πεδίο των Κυματιδίων

Οι μέθοδοι μερικών διαφορικών εξισώσεων για ενδοσυμπλήρωση που συζητήσαμε στο Κεφάλαιο 3 βασίζονται σε γεωμετρική μοντελοποίηση της εικόνας στο χωρικό πεδίο. Μια άλλη κατηγορία τεχνικών ενδοσυμπλήρωσης βασίζεται σε στατιστική μοντελοποίηση στο πεδίο γραμμικών μετασχηματισμών της εικόνας. Στο παρόν Κεφάλαιο θα συζητήσουμε εναλλακτικές αναπαραστάσεις εικόνας τόσο στο πεδίο του μετασχηματισμού Fourier, όσο και στο πεδίο χωρικά εντοπισμένων κυματιδιακών μετασχηματισμών, που πλεονεκτούν για την περιγραφή μη στάσιμων φυσικών εικόνων και έχουν αποδειχθεί ιδιαίτερα αποδοτικοί σε εφαρμογές συμπίεσης εικόνων. Θα αναφερθούμε τόσο σε κυματίδια χώρου-κλίμακας, κατάλληλα για αναπαράσταση τμηματικά ομαλών εικόνων, όσο και σε κυματίδια χώρου-συχνότητας, κατάλληλα για αναπαράσταση τοπικά περιοδικών παραστάσεων υφής. Ειδικότερα, θα εστιάσουμε σε υπερπλήρεις μιγαδικές κυματιδιακές αναπαραστάσεις πλαισίου, συγκεκριμένα στο δίδεντρο κυματιδιακό μετασχηματισμό και στο διαμορφωμένο μιγαδικό επικαλυπτόμενο μετασχηματισμό, που εμφανίζουν σημαντικά βελτιωμένες ιδιότητες μεταφορικής και περιστροφικής συμμετρίας σε σχέση με τους κρίσιμα δειγματοληπτημένους ορθοκανονικούς μετασχηματισμούς και είναι για αυτό προτιμότεροι για την εφαρμογή μας. Μελετώντας τα ποιοτικά χαρακτηριστικά των συντελεστών κυματιδίων φυσικών εικόνων σε αυτές τις αναπαραστάσεις αναδεικνύονται κανονικότητες που θα αποτελέσουν τη βάση για την ανάπτυξη των κυματιδιακών στατιστικών μοντέλων ενδοσυμπλήρωσης του Κεφαλαίου 6.

5.1 Κλασική Μοντελοποίηση Φυσικών Εικόνων ως Στάσιμων Στοχαστικών Διαδικασιών και Αναπαράσταση στο Πεδίο Μετασχηματισμού Fourier

Η απλούστερη στατιστική περιγραφή εικόνων μπορεί να γίνει αν τις θεωρήσουμε ως 2-Δ αυστηρά στάσιμες στοχαστικές διαδικασίες, δηλαδή θεωρήσουμε ότι τα σήματα εικόνων υπόκεινται σε στοχαστική *συμμετρία μετατόπισης* [Sim05]. Τότε οι ροπές δεύ-

τερης τάξης της στοχαστικής διαδικασίας μπορούν να περιγραφούν από τη συνάρτηση αυτοσυσχέτισης $R(\mathbf{d}) = E\{I(\mathbf{x})I(\mathbf{x} + \mathbf{d})\}$, ή, ισοδύναμα στο πεδίο του μετασχηματισμού Fourier (ΜΦ), από την αντίστοιχη πυκνότητα φάσματος ισχύος $S(\omega) = \mathcal{F}\{R(\mathbf{d})\}$, όπου $F(\omega) = \mathcal{F}\{f(\mathbf{x})\} = \sum_{\mathbf{x}} f(\mathbf{x})e^{-j(\omega,\mathbf{x})}$ δηλώνει το 2-Δ ΜΦ διακριτού χώρου/συνεχούς συχνότητας [OS98]. Στην ειδική περίπτωση περιοδικών συνοριακών συνθηκών η R είναι κυκλική (circulant) και συνεπώς διαγωνοποιήσιμη από το ΜΦ, άρα οι όροι $S(\omega)$ είναι ανεξάρτητοι. Δυο ακόμα υποθέσεις αρκούν για να καταλήξουμε σε συγκεκριμένη μορφή για τη συνάρτηση πυκνότητας φάσματος ισχύος S : 1) *συμμετρία περιστροφής*, που συνεπάγεται $S(\omega) = S(\omega)$, όπου $\omega \equiv |\omega|$, και 2) *συμμετρία κλίμακας*, δηλαδή φωτογραφίες αντικειμένων σε πολύ διαφορετικές κλίμακες έχουν παρόμοιες ιδιότητες· πιο συγκεκριμένα απαιτούμε το φάσμα ισχύος να είναι *συμμεταβαλλόμενο* (covariant) με την κλίμακα, δηλαδή $S(s\omega) = h(s)S(\omega)$. Τότε:

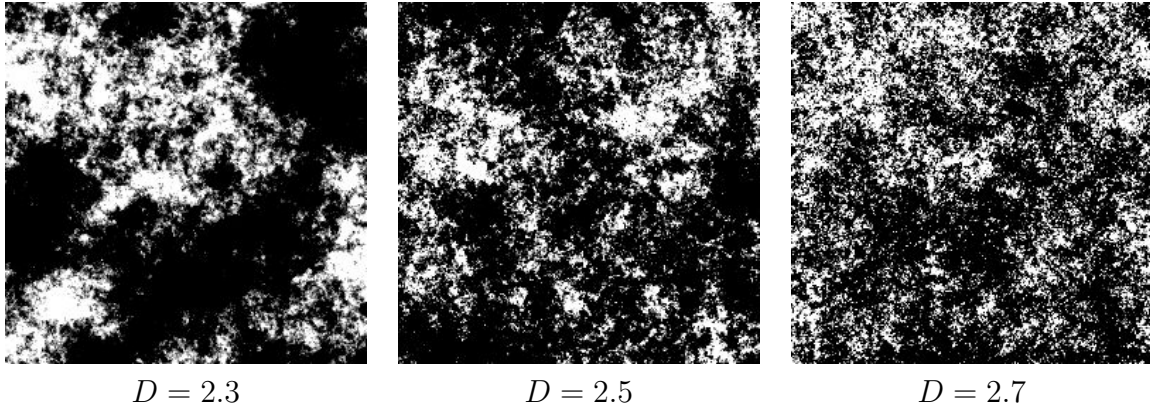
$$S(\omega) = \frac{c}{\omega^\beta} \quad (5.1)$$

Απόδειξη. Από τη σχέση $S(s\omega) = h(s)S(\omega)$ για $\omega = 1$ προκύπτει $h(s) = S(s)/S(1)$, οπότε $S(s\omega) = c_1 S(s)S(\omega)$, με $c_1 = 1/S(1)$. Για να λύσουμε αυτή τη συναρτησιακή εξίσωση, παραγωγίζουμε ως προς s και θέτουμε στη συνέχεια $s = 1$, οπότε καταλήγουμε στη διαφορική εξίσωση $\omega S'(\omega) = -\beta S(\omega)$, με $\beta = -c_1 S'(1) = -S'(1)/S(1)$, της οποίας η λύση είναι η $S(\omega) = c/\omega^\beta$, με αντίστοιχο παράγοντα συμμεταβολής $h(s) = 1/s^\beta$. \square

Αυτός ο εκθετικός νόμος για φάσματα ισχύος εικόνων είχε παρατηρηθεί εμπειρικά σε τηλεοπτικά σήματα ήδη από την δεκαετία του '50 (δες τα [SLSZ03, Rud94] για επισκόπηση της βιβλιογραφίας) και έγινε ευρύτερα γνωστός τη δεκαετία του '80 από μια μελέτη του Field που συνέδεε τις στατιστικές ιδιότητες των φυσικών εικόνων με τις ιδιότητες του ανθρώπινου συστήματος όρασης [Fie87]. Η τιμή της παραμέτρου-εκθέτη β μπορεί να εκτιμηθεί εμπειρικά με ταίριασμα του μοντέλου σε εμπειρικές παρατηρήσεις από φυσικές εικόνες. Για τις περισσότερες κατηγορίες εικόνων έχουμε $\beta \approx 2$. Η φράκταλ διάσταση δειγμάτων τέτοιων 2-Δ στοχαστικών διαδικασιών με συμμετρία κλίμακας δίνεται από τη σχέση $D = 3.5 - \beta/2$ [Vos88]. Για $\beta = 2$ έχουμε $D = 2.5$, που αντιστοιχεί στην κλασική 2-Δ κίνηση Brown, και η φασματική ενέργεια του σήματος εικόνας είναι ομοιόμορφα κατανομημένη σε κάθε φασματική οκτάβα στο επίπεδο του 2-Δ ΜΦ, καθότι, για $\beta = 2$, το ολοκλήρωμα $\int_{\omega_0}^{\alpha\omega_0} S(\omega)(2\pi\omega)d\omega$ είναι ανεξάρτητο του ω_0 .

Βέβαια, η γνώση ροπών ως και δεύτερης τάξης (μέσω της συνάρτησης αυτοσυσχέτισης ή του φάσματος ισχύος) δεν αρκεί γενικά για τον πλήρη καθορισμό της στοχαστικής διαδικασίας, εκτός κι αν πρόκειται για στοχαστική διαδικασία Gauss. Η στοχαστική διαδικασία Gauss με εκθετικό φάσμα που περιγράφεται από την εξίσωση (5.1) είναι γνωστή ως *κλασματική κίνηση Brown* (fractional Brownian motion - fBm) και αποτελεί γενίκευση της κλασικής κίνησης Brown. Διδιάστατα δείγματα αυτής για διαφορετικές τιμές της φράκταλ διάστασης D απεικονίζονται στο Σχ. 5.1.

Τα δείγματα fBm μοιάζουν οπτικά με σύννεφα (ή ως βραχώδη όρη εάν οπτικοποιηθούν ως 2-Δ επιφάνειες) αλλά, πέρα από αυτό, δεν φαίνεται να μπορούν να προσεγγίσουν ιδιαίτερα την αίσθηση που μας δίνουν οι φυσικές εικόνες. Ειδικότερα, βασικό μειονέκτημα των fBm δειγμάτων είναι ότι δεν περιέχουν ακμές, οι οποίες αποτελούν κυρίαρχη συνιστώσα των φυσικών εικόνων. Αιτία αυτής της αποτυχίας είναι ότι το μοντέλο Gauss περιορίζεται σε μοντελοποίηση με ροπές μέχρι και δεύτερης τάξης, ενώ η σύνθεση ακμών φαίνεται να απαιτεί μοντελοποίηση ροπών υψηλότερης τάξης. Συνεπώς η μοντελοποίηση



Σχήμα 5.1: Δείγματα κλασματικής κίνησης Brown για διάφορες τιμές της φράκταλ διάστασης $D = 3.5 - \beta/2$. Η σύνθεση έγινε με μεθόδους στο πεδίο του ΜΦ: Σύνθεση λευκού θορύβου στο πεδίο των συντελεστών Fourier, πολλαπλασιασμός με $S(\omega)^{1/2}$, αντίστροφος ΜΦ [Vos88].

με ροπές ως και δεύτερης τάξης και το συνακόλουθο στατικό μοντέλο Gauss, αν και μπορούν να συλλάβουν κάποιες ιδιότητες των φυσικών εικόνων, π.χ. τη συμμετρία κλίμακας, δεν συνιστούν επαρκή απάντηση στο πρόβλημα της μοντελοποίησης φυσικών εικόνων.

5.2 Χωρικά Εντοπισμένοι Μετασχηματισμοί Εικόνων

A representation is a formal system for making explicit certain entities or types of information, together with a specification of how the system does this ... any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover. Από το [Mar82, σελ. 20].

Η υπόθεση ότι οι φυσικές εικόνες είναι ομοιογενείς και μπορούν να μοντελοποιηθούν ως καθολικά στάσιμες στοχαστικές διαδικασίες είναι υπερβολικά απλουστευτική. Στις εικόνες του φυσικού κόσμου συνήθως συνυπάρχουν ετερογενείς περιοχές με διαφορετικές ιδιότητες χρώματος ή υφής, διαχωριζόμενες μεταξύ τους από ακμές. Για την αναπαράσταση τέτοιων μη στάσιμων σημάτων είναι προσφορότερη η χρήση τοπικών μετασχηματισμών που, σε αντίθεση με τον καθολικό μετασχηματισμό Fourier, εμφανίζουν καλές ιδιότητες χωρικού εντοπισμού.

Θα συζητήσουμε δύο κατηγορίες χωρικά εντοπισμένων μετασχηματισμών εικόνας [Mey93]: (α) το μετασχηματισμό κυματιδίων, που αντιστοιχεί σε ανάλυση της εικόνας σε συντεταγμένες *χώρου-κλίμακας* (space-scale) και προσφέρεται ιδιαίτερα για την περιγραφή τμηματικά ομαλών εικόνων και (β) τον τοπικά αρμονικό μετασχηματισμό εικόνας με παραθυρική ανάλυση που αντιστοιχεί σε ανάλυση της εικόνας σε συντεταγμένες *χώρου-συχνότητας* (space-frequency) και προσφέρεται ιδιαίτερα για την περιγραφή ταλαντούμενων κυματομορφών σε εικόνες υφής. Θα εξετάσουμε επίσης υπερπλήρεις μιγαδικές εκδοχές αυτών των μετασχηματισμών που χρησιμοποιούμε στο Κεφάλαιο 6 γιατί παρουσιάζουν βελτιωμένα χαρακτηριστικά μεταφορικής ή περιστροφικής συμμετρίας και είναι καταλληλότεροι για την ανάλυση εικόνων.

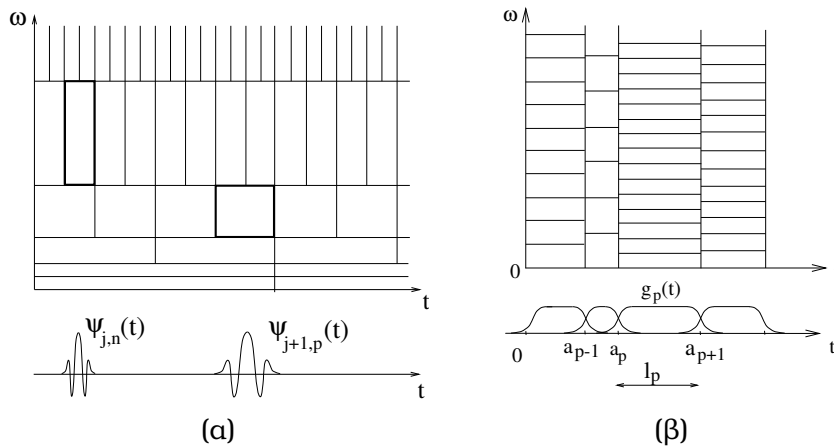
5.2.1 Κυματιδιακή Αναπαράσταση Χώρου-Κλίμακας για Σήματα και Εικόνες

Ο μετασχηματισμός κυματιδίων αποτελεί σημαντικό εργαλείο για την αναπαράσταση μη στάσιμων σημάτων και έχει βρει πλήθος εφαρμογών στην ανάλυση και κωδικοποίηση εικόνων [Mal89, Dau92, Mey93, Vai93, VK95, Mal99a]. Προέκυψε από σύνθεση ιδεών προερχόμενων από διάφορες επιστημονικές περιοχές (επεξεργασία σήματος, όραση υπολογιστών, αρμονική ανάλυση, φυσική, στατιστική) [HW06]. Εγγενής στο μετασχηματισμό κυματιδίων είναι η ιδέα της *πολυκλιμακωτής προσέγγισης* (multiresolution approximation - MRA) σημάτων των Mallat και Meyer [Mal89, Mey92], η οποία έχει τις ρίζες της στη Λαπλασιανή πυραμίδα των Burt και Adelson [BA83a].

Θεωρούμε ένα 1-Δ αναλογικό σήμα πεπερασμένης ενέργειας $x(t) \in L^2(R)$. Η θεωρία της πολυκλιμακωτής προσέγγισης [Mal99a, Κεφ. 7] μελετά τη δυνατότητα αναπαράστασης του $x(t)$ σε διαδοχικά επίπεδα ανάλυσης $-\infty < j \leq j_c$, όπου το j -επίπεδο αντιστοιχεί σε μήκος ανάλυσης (κλίμακα) 2^j και το j_c είναι συνεπώς το αδρότερο επίπεδο, ως εξής:

$$x(t) = \sum_{n=-\infty}^{\infty} \alpha_{j_c}[n] \phi_{j_c,n}(t) + \sum_{j=-\infty}^{j_c} \sum_{n=-\infty}^{\infty} d_j[n] \psi_{j,n}(t), \quad (5.2)$$

όπου τα $\phi_{j,n}(t) = 2^{-j/2} \phi(2^{-j}(t - 2^j n))$ και $\psi_{j,n}(t) = 2^{-j/2} \psi(2^{-j}(t - 2^j n))$ είναι κλιμακωμένες και μετατοπισμένες εκδοχές συναρτήσεων πατρικού $\phi(t)$ (βαθυπερατή) και μητρικού $\psi(t)$ (ζωνοπερατή) κυματιδίου, αντίστοιχα. Τα $\alpha_{j_c}[n]$ και $d_j[n]$ ονομάζονται συντελεστές κλιμακωσης και λεπτομέρειας, αντίστοιχα. Σημειωτέον ότι η απόσταση μεταξύ γειτονικών κυματιδίων $\psi_{j,n}$ και $\psi_{j,n+1}$ ισούται με $\Delta t = 2^j$ και συνεπώς υποδιπλασιάζεται καθώς μεταβαίνουμε στην αμέσως λεπτομερέστερη κλίμακα. Αυτή η χαρακτηριστική ιδιότητα συγκέντρωσης αντικατοπτρίζεται στο διάγραμμα χρόνου-συχνότητας 5.2(a).



Σχήμα 5.2: Ιδιότητες χρονικού/φασματικού εντοπισμού της κυματιδιακής ανάλυσης με τη βοήθεια διαγραμμάτων ορθογωνίων Heisenberg, επίσης γνωστών και ως διαγραμμάτων κουτιών χρόνου-συχνότητας. (α) Κυματίδια χρόνου-κλίμακας, όπου σε διαδοχικά επίπεδα ανάλυσης υποδιπλασιάζεται η χρονική διάρκεια του φίλτρου με παράλληλο διπλασιασμό του εύρους φασματικής απόκρισης. (β) Κυματίδια χρόνου-συχνότητας, με πιθανά μεταβλητή διάρκεια l_p : τα άτομα εντός κάθε χρονικού παραθύρου καλύπτουν ομοιόμορφα τον άξονα συχνοτήτων. Σχήματα δανεισμένα από το [Mal99a, Ενót. 1.3].

Επιλέγοντας κατάλληλα τις συναρτήσεις ϕ και ψ , τα $\{\phi_{j_c,n}, \psi_{j_c,n}, j \leq j_c\}$ αποτελούν πολυκλιμακωτή ορθοκανονική βάση του $L^2(R)$, και συνεπώς $\alpha_{j_c}[n] = \langle x, \phi_{j_c,n} \rangle$, $d_j[n] = \langle x, \psi_{j,n} \rangle$. Μια από τις σχετικές συνθήκες πολυκλιμακωτής ανάλυσης [Mal99a] επιβάλλει στα κυματίδια να ικανοποιούν τις συναρτησιακές σχέσεις

$$\phi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} h_0[n] \phi(2t - n) \quad (\text{εξ. πατρικού κυματιδίου}) \quad (5.3\alpha)$$

$$\psi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} h_1[n] \phi(2t - n) \quad (\text{εξ. μητρικού κυματιδίου}), \quad (5.3\beta)$$

όπου για τέλεια ανακατασκευή απαιτείται από τα φίλτρα h_0 και h_1 να ικανοποιούν τις σχέσεις [SB86, Vai93, VK95]

$$|H_0(\omega)|^2 + |H_0(\omega + \pi)|^2 = 2 \quad (\text{συνθήκη CMF}) \quad (5.4\alpha)$$

$$h_1[n] = (-1)^{1-n} h_0[1 - n]. \quad (5.4\beta)$$

Η πρώτη από τις παραπάνω συνθήκες χαρακτηρίζει το h_0 ως *συζυγές ανακλαστικό φίλτρο* (conjugate mirror filter - CMF). Βλέπουμε λοιπόν ότι το πρόβλημα της σχεδίασης των αναλογικών συναρτήσεων ϕ και ψ που απαιτούνται για την κυματιδιακή ανάλυση ανάγεται σε πρόβλημα σχεδίασης του διακριτού φίλτρου $h_0[n]$. Έχουν αναπτυχθεί τεχνικές σχεδίασης κατάλληλων FIR συζυγών ανακλαστικών φίλτρων που οδηγούν σε ορθοκανονικές κυματιδιακές βάσεις με επιθυμητές ιδιότητες (π.χ. αυξημένη ομαλότητα/πεπερασμένο μήκος για το μητρικό κυματίδιο ϕ ή μικρό μήκος για το φίλτρο h_0) [Dau92, Mal99a].

Υπάρχει στενή σχέση ανάμεσα στο μετασχηματισμό κυματιδίων και σε τεχνικές ανάλυσης σημάτων σε υποζώνες με χρήση συστοιχιών φίλτρων [Mal89]. Αν στην Εξ. (5.2) αυξήσουμε κατά 1 το j_0 , τότε οι συντελεστές κλιμάκωσης/λεπτομέρειας που αντιστοιχούν στη νέα αδρότερη κλίμακα ανάλυσης μπορούν να υπολογιστούν με την αναδρομή

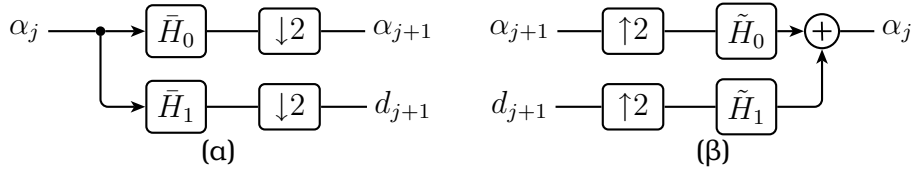
$$\alpha_{j+1}[p] = \sum_{n=-\infty}^{\infty} h_0[n - 2p] \alpha_j[n] = (\alpha_j * \bar{h}_0)[2p] \quad (5.5\alpha)$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{\infty} h_1[n - 2p] \alpha_j[n] = (\alpha_j * \bar{h}_1)[2p], \quad (5.5\beta)$$

όπου $\bar{x}[n] \triangleq x[-n]$, $\bar{X}(\omega) = X(-\omega)$, που περιλαμβάνει φιλτράρισμα των συντελεστών κλιμάκωσης α_j με ανεστραμένες εκδοχές των φίλτρων h_0, h_1 ακολουθούμενο από υποδειγματοληψία, όπως φαίνεται στο Σχ. 5.3(α). Αντίστροφα, μπορούμε να μειώσουμε το j_0 και να προχωρήσουμε σε σύνθεση των συντελεστών $\alpha_j[n]$ από τους κυματιδιακούς συντελεστές της αδρότερης κλίμακας

$$\begin{aligned} \alpha_j[p] &= \sum_{n=-\infty}^{\infty} h_0[p - 2n] \alpha_{j+1}[n] + \sum_{n=-\infty}^{\infty} h_1[p - 2n] d_{j+1}[n] \\ &= (\check{\alpha}_{j+1} * h_0)[p] + (\check{d}_{j+1} * h_1)[p], \end{aligned} \quad (5.6)$$

όπου $\check{x}[n] \triangleq x[p]$, εάν $n = 2p$ και 0 διαφορετικά, $\check{X}(\omega) = X(2\omega)$, δηλαδή η διαδικασία σύνθεσης (5.6) περιλαμβάνει υπερδειγματοληψία με ένθεση μηδενικών ακολουθούμενη από φιλτράρισμα με τα h_0, h_1 , όπως φαίνεται στο Σχ.5.3(β).



Σχήμα 5.3: Ένα βήμα κυματιδιακής ανάλυσης (α) και σύνθεσης (β) με τον αναδρομικό DWT αλγόριθμο του Mallat βασισμένο σε συστοιχίες φίλτρων. Στην περίπτωση χρήσης ορθογώνιων κυματιδίων το δυϊκό ζεύγος φίλτρων $\{\tilde{h}_0, \tilde{h}_1\}$ που εμφανίζεται στο βήμα σύνθεσης (β) συμπίπτει με το πρωτεύον ζεύγος φίλτρων $\{h_0, h_1\}$.

Αν στις παραπάνω αναδρομικές διαδικασίες επιτρέψουμε χρήση διαφορετικών ζευγών φίλτρων στα στάδια ανάλυσης και σύνθεσης, το *πρωτεύον* (primal) $\{h_0, h_1\}$ και το *δυϊκό* (dual) $\{\tilde{h}_0, \tilde{h}_1\}$, αντίστοιχα, οδηγούμαστε στα λεγόμενα *διορθογώνια* (biorthogonal) κυματίδια που δημιουργούν διορθογώνιες (μη ορθοκανονικές) Riesz βάσεις του $L^2(R)$ [CDF92, Ma199a]. Συνθήκες που γενικεύουν τις Εξ. (5.4) και εξασφαλίζουν τέλεια ανακατασκευή μπορούν να βρεθούν στο [VK95]. Ανάλογα με την ορθογώνια περίπτωση (5.2), εγγενής στα διορθογώνια κυματίδια είναι η πολυκλιμακωτή ανάλυση

$$x(t) = \sum_{n=-\infty}^{\infty} \alpha_{j_c}[n] \tilde{\phi}_{j_c, n}(t) + \sum_{j=-\infty}^{j_c} \sum_{n=-\infty}^{\infty} d_j[n] \tilde{\psi}_{j, n}(t). \quad (5.7)$$

Το δυϊκό ζεύγος κυματιδίων $\{\tilde{\phi}, \tilde{\psi}\}$ χαρακτηρίζεται από το δυϊκό ζεύγος φίλτρων $\{\tilde{h}_0, \tilde{h}_1\}$ από σχέσεις ανάλογες της (5.3) και υπεισέρχεται στον κλάδο σύνθεσης της συστοιχίας φίλτρων, Σχ. 5.3(β). Το πρωτεύον ζεύγος κυματιδίων $\{\phi, \psi\}$ συνεχίζει να υπεισέρχεται στον κλάδο ανάλυσης της συστοιχίας φίλτρων, Σχ. 5.3(α), οπότε ισχύει $\alpha_{j_c}[n] = \langle x, \phi_{j_c, n} \rangle$, $d_j[n] = \langle x, \psi_{j, n} \rangle$. Αξίζει να σημειωθεί ότι τα δύο ζεύγη φίλτρων παίζουν συμμετρικούς ρόλους (το ένα είναι δυϊκό του άλλου), οπότε μπορεί κανείς να ορίσει το δυϊκό μετασχηματισμό κυματιδίων που χρησιμοποιεί τα φίλτρα $\{\tilde{h}_0, \tilde{h}_1\}$ κατά την ανάλυση και τα φίλτρα $\{h_0, h_1\}$ κατά τη σύνθεση. Η πρόσθετη ελευθερία κατά τη σχεδίαση διορθογώνιων κυματιδίων επιτρέπει χρήση συμμετρικών βαθυπερατών φίλτρων h_0, \tilde{h}_0 *γραμμικής φάσης* (linear phase) [OS98] που παράγουν συμμετρικά $\phi, \tilde{\phi}$ και συμμετρικά (αντισυμμετρικά) $\psi, \tilde{\psi}$ για περιττό (αντ. άρτιο) μήκος φίλτρου h_0 . Τέτοια διορθογώνια κυματίδια χρησιμοποιούνται συχνά για ανάλυση και κωδικοποίηση εικόνων.

Οι αναδρομικές σχέσεις (5.5) και (5.6) οδηγούν στο μετασχηματισμό κυματιδίων για διακριτά σήματα [VK95], όπου κανείς θεωρεί ότι το διακριτό σήμα $x[n]$ αντιστοιχεί στους συντελεστές κλιμάκωσης $\alpha_{j_f}[n]$ ενός υποκείμενου συνεχούς σήματος $x(t)$ και η κυματιδιακή ανάλυση εκτείνεται σε πεπερασμένο εύρος κλιμάκων $j_f < j \leq j_c$ ($L = j_c - j_f$ επίπεδα). Τότε η Εξ. (5.5) ορίζει αναδρομικά το *διακριτό μετασχηματισμό κυματιδίων* (discrete wavelet transform - DWT), ενώ η Εξ. (5.6) τον αντίστροφό του, που ανακατασκευάζει επακριβώς το αρχικό διακριτό σήμα $x[n]$. Η ανάλυση ενός διακριτού σήματος με τον DWT μπορεί να ερμηνευθεί ως ανάλυση σε βάση του χώρου $l^2(Z)$. Τα διανύσματα βάσης που εμπλέκονται στο στάδιο ανάλυσης είναι πλέον σήματα διακριτού χρόνου $\{\phi_{j_c}[p - 2^{j_c - j_f} n], \psi_j[p - 2^{j - j_f} n], j_f < j \leq j_c\}$ και καθορίζονται από τη δομή της συστοιχίας φίλτρου του Σχ. 5.3. Η (δι)ορθογωνιότητα των συνεχών συναρτήσεων βάσης μεταφέρεται και στις αντίστοιχες διακριτού χρόνου εκδοχές τους. Με προσεκτικό χειρισμό των συνωριακών συνθηκών μπορεί να οριστεί ο DWT και για διακριτά σήματα πεπερασμένου

μήκους N , που παράγει ίσο αριθμό συντελεστών κυματιδίων [Mal99a] – ιδιότητα *κρίσιμης δειγματοληψίας* (critical sampling). Το υπολογιστικό κόστος τόσο για τον ευθύ όσο και για τον αντίστροφο DWT είναι ανάλογο του N (για FIR φίλτρα h_0, h_1). Ιδιαίτερα χρήσιμη είναι η αναπαράσταση του DWT με συμβολισμό πινάκων. Εάν θεωρήσουμε ως διανύσματα το σύνολο των διακριτών δειγμάτων του σήματος, $\mathbf{x} = (x[1], \dots, x[N])^T$, και του διακριτού κυματιδιακού μετασχηματισμού του, $\mathbf{y} = (y[1], \dots, y[N])^T$ (παράθεση συντελεστών κλιμάκωσης και λεπτομέρειας), τότε ο DWT μπορεί να εκφραστεί ως $\mathbf{y} = F_a \mathbf{x}$ (ευθύς) και $\mathbf{x} = F_s \mathbf{y}$ (αντίστροφος), $F_a^{-1} = F_s$. Στη διορθογώνια περίπτωση και αντιστρέφοντας τους ρόλους πρωτευόντων και δυϊκών φίλτρων έχουμε επιπρόσθετα τους δυϊκούς μετασχηματισμούς $\tilde{\mathbf{y}} = \tilde{F}_a \mathbf{x}$ και $\mathbf{x} = \tilde{F}_s \tilde{\mathbf{y}}$, $\tilde{F}_a^{-1} = \tilde{F}_s$. Οι DWT πίνακες συνδέονται με τις σχέσεις [VK95]

$$F_s^H = F_a \quad (\text{ορθογώνια κυματίδια}) \quad (5.8\alpha')$$

$$\tilde{F}_s^H = F_a, \quad F_s^H = \tilde{F}_a, \quad (\text{διορθογώνια κυματίδια}) \quad (5.8\beta')$$

όπου A^H συμβολίζει τον ανάστροφο-συζυγή του πίνακα A (ταυτίζεται με τον ανάστροφο για πραγματικούς πίνακες). Αυτές οι ιδιότητες μας επιτρέπουν να υπολογίζουμε γινόμενα πίνακα διανύματος τύπου $F_s^H \mathbf{x}$ και $F_a^H \mathbf{y}$ με τους αποδοτικούς αλγορίθμους για ευθύ και αντίστροφο DWT, αντίστοιχα.

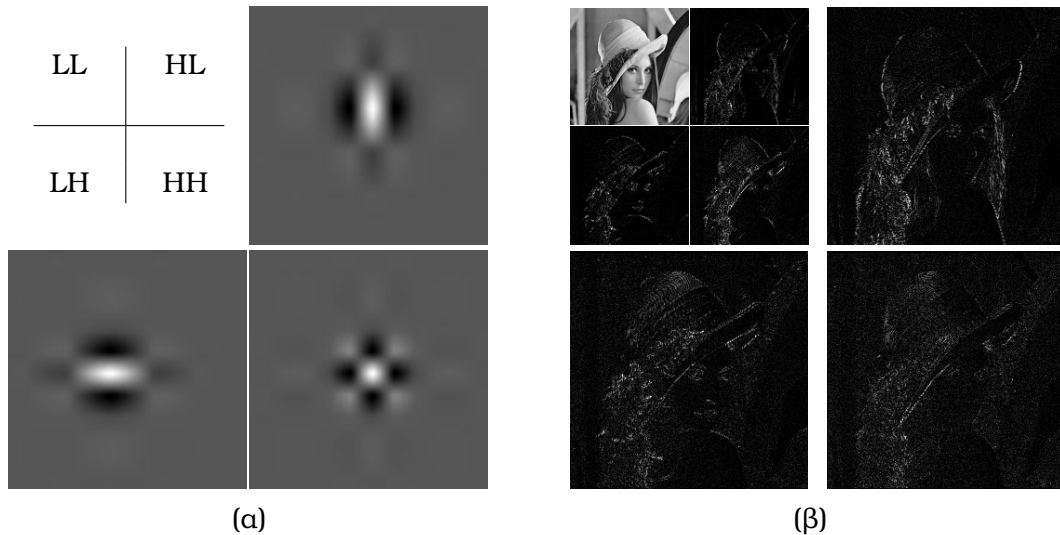
Όπως και στην περίπτωση του μετασχηματισμού Fourier, η μετάβαση της κυματιδιακής ανάλυσης από 1-Δ σήματα σε 2-Δ εικόνες συνήθως βασίζεται σε διαχωρίσιμη επεξεργασία, κατά την οποία ο 1-Δ μετασχηματισμός εφαρμόζεται διαδοχικά στην οριζόντια και στην κάθετη διεύθυνση. Η διαδικασία αυτή οδηγεί σε κατασκευή Riesz βάσεων του $L^2(R^2)$ με χρήση μετατοπισμένων/κλιμακωμένων εκδοχών ενός πατρικού και τριών μητρικών κυματιδίων

$$\begin{aligned} \phi(x, y) = \phi(x)\phi(y) \quad (\text{LL κυματίδιο}) \quad \psi_2(x, y) = \psi(x)\phi(y) \quad (\text{HL κυματίδιο}) \\ \psi_1(x, y) = \phi(x)\psi(y) \quad (\text{LH κυματίδιο}) \quad \psi_3(x, y) = \psi(x)\psi(y) \quad (\text{HH κυματίδιο}) \end{aligned} \quad (5.9)$$

Η παρουσία τριών μητρικών κυματιδίων αντιστοιχεί σε ανάλυση της εικόνας σε τρία κατευθυντικά κανάλια, οριζόντιο (LH, ακμές στη διεύθυνση 0°), κάθετο (HL, ακμές στη διεύθυνση 90°) και διαγώνιο (HH, ακμές στις διευθύνσεις $\pm 45^\circ$). Οι διπλοί λωβοί του ψ_3 κυματιδίου σημαίνουν ότι το HH κανάλι αποκρίνεται εξίσου σε ακμές στις διευθύνσεις $+45^\circ$ και -45° και συνεπώς αδυνατεί να τις διακρίνει. Στο Σχ. 5.4 δείχνουμε τυπικά παραδείγματα κυματιδίων που αντιστοιχούν σε 2-Δ διαχωρίσιμη ανάλυση, καθώς και τους 2-Δ συντελεστές κυματιδίων που αντιστοιχούν στο 2-Δ DWT μιας τυπικής εικόνας.

5.2.2 Αναπαράσταση Σημάτων σε Πλαίσια και ο Δίδεντρος Μιγαδικός Κυματιδιακός Μετασχηματισμός

Η αναπαράσταση σημάτων σε βάσεις κυματιδίων χαρακτηρίζεται από έλλειψη *μεταφορικής συμμετρίας* (translation invariance) [SFAH92]. Συγκεκριμένα, μετατόπιση του σήματος στο πεδίο του χρόνου (1-Δ) ή του χώρου (2-Δ) δεν μεταφράζεται απλά σε αντίστοιχη μετατόπιση των συντελεστών κυματιδίων, ενώ επιπλέον συνοδεύεται από ανταλλαγή ενέργειας μεταξύ υποζωνών του κυματιδιακού μετασχηματισμού. Το φαινόμενο αυτό προκαλείται λόγω *φασματικής αναδίπλωσης* (aliasing) κατά το βήμα υποδειγματοληψίας των συντελεστών κυματιδίων στον υπολογισμό του DWT με τη συστοιχία φίλτρου του Σχ. 5.3.



Σχήμα 5.4: (α) Τα τρία κυματίδια που αντιστοιχούν στον διαχωρισμό 2-Δ (DWT). (β) Απόλυτη τιμή των συντελεστών κυματιδίων που αντιστοιχούν στο 2-Δ DWT (2 επίπεδα ανάλυσης) της 512×512 εικόνας Lena.

Αν και λόγω αλληλοεξουδετέρωσης της φασματικής αναδίπλωσης μεταξύ διαφορετικών υποζωνών επιτυγχάνεται τέλεια ανακατασκευή με τον αντίστροφο DWT, σε εφαρμογές ανάλυσης εικόνας η έλλειψη μεταφορικής συμμετρίας αποτελεί σημαντικό μειονέκτημα για τον κρίσιμα δειγματοληπτημένο DWT.

Επιπλέον μειονέκτημα του διαχωρισμού 2-Δ DWT είναι η μειωμένη κατευθυντικότητα του και η παράλληλη έλλειψη *περιστροφικής συμμετρίας* (rotational invariance) [SFAH92]. Ο διαχωρισμός 2-Δ DWT αναλύει την εικόνα σε τρία μόνο κατευθυντικά κανάλια και μάλιστα αδυνατεί να διακρίνει μεταξύ διαγώνιων ακμών στις διευθύνσεις $+45^\circ$ και -45° . Περιστρέφοντας την εικόνα κατά θ° οδηγεί σε αλλαγές στις κυματιδιακές αποκρίσεις και αλληλεπιδράσεις μεταξύ των διαφορετικών καναλιών που δεν μπορούν να εκφραστούν με απλό τρόπο στο DWT πεδίο.

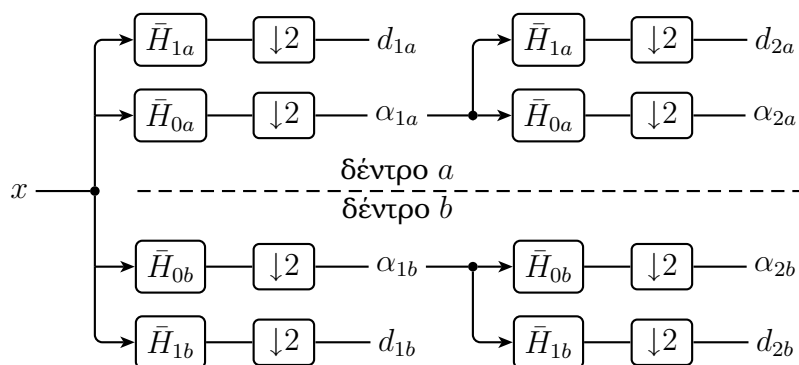
Μπορεί κανείς να σχεδιάσει γραμμικούς τοπικά εντοπισμένους μετασχηματισμούς που διαθέτουν (προσεγγιστικά ή επακριβώς) τις επιθυμητές ιδιότητες μεταφορικής ή/και περιστροφικής συμμετρίας, ενώ παράλληλα διατηρούν τη φιλοσοφία και την πολυκλιμακωτή υπολογιστική δομή του κυματιδιακού μετασχηματισμού, παραγόμενοι με επαναληπτικό φιλτράρισμα με ζωνοπερατά κατευθυντικά φίλτρα. Για να επιτευχθεί αυτό θα πρέπει κανείς να θυσιάσει την ιδιότητα της κρίσιμης δειγματοληψίας, οδηγούμενος έτσι σε *υπερπλήρεις/πλεονάζουσες αναπαραστάσεις* (overcomplete/redundant representations), όπου ένα σήμα x σε έναν κατάλληλο χώρο Hilbert \mathbf{H} εκφράζεται ως άθροισμα γραμμικά εξαρτημένων *ατόμων* (atoms) $\{\phi_i\}$ (π.χ., για πεπερασμένα σήματα μήκους N έχουμε αναπαράσταση με χρήση $M > N$ ατόμων). Από μαθηματική άποψη τα $\{\phi_i\}$ δεν αποτελούν πλέον βάση αλλά *πλαίσιο* (frame) του \mathbf{H} . Η θεωρία των πλαισίων [DS52, Dau90, Grö93] έχει αποδειχτεί ιδιαίτερα χρήσιμη για την περιγραφή αυτών των υπερπλήρων αναπαραστάσεων· ο αναγνώστης παραπέμπεται στα [KC07a, KC07b] για μια πρόσφατη επισκόπησή της από τη σκοπιά της επεξεργασίας σήματος.

Η πιο απλή υπερπλήρης αναπαράσταση που εξασφαλίζει μεταφορική συμμετρία, όχι όμως και περιστροφική συμμετρία, είναι ο μη-δειγματοληπτημένος DWT [Mal99a]. Ο

υπολογισμός του βασίζεται σε παραλλαγή της συστοιχίας φίλτρων 5.3 όπου παραλείπεται το στάδιο υποδειγματοληψίας κατά την ανάλυση και παράλληλα παρεμβάλλονται μηδενικά μεταξύ των δειγμάτων των φίλτρων (αλγόριθμος *à trous*). Για ανάλυση σήματος μήκους N σε L επίπεδα οδηγούμαστε έτσι σε $M = (L + 1)N$ κυματιδιακούς συντελεστές (βαθμός πλεονασμού $L + 1$).

Μια δημοφιλής αναπαράσταση εικόνων που εξασφαλίζει περιστροφική συμμετρία είναι η *περιστρεφόμενη πυραμίδα* (steerable pyramid) [FA91, SFAH92, SF95, Per95, PS00b]. Η σχεδίαση της περιστρεφόμενης πυραμίδας επιτρέπει τον υπολογισμό της κατευθυντικής απόκρισης σε οποιαδήποτε συνεχή διεύθυνση θ° ως γραμμικού συνδυασμού αποκρίσεων ενός πεπερασμένου αριθμού φίλτρων προσανατολισμένων σε προκαθορισμένες διακριτές διευθύνσεις. Τα φίλτρα της περιστρεφόμενης πυραμίδας αντιστοιχούν σε πολυκλιμακωτές διακριτές προσεγγίσεις τελεστών μερικής παραγωγίσης και ο βαθμός πλεονασμού της αναπαράστασης εξαρτάται από την τάξη παραγωγίσης. Παρόμοιες τεχνικές έχουν χρησιμοποιηθεί για τη σχεδίαση φίλτρων προσαρμοζόμενης κλίμακας [SFAH92, Per95].

Μια αναπαράσταση που συνδυάζει πολύ καλές ιδιότητες μεταφορικής συμμετρίας και κατευθυντικότητας και χρησιμοποιούμε ευρέως στην εργασία μας είναι ο *δίδεντρος μιγαδικός κυματιδιακός μετασχηματισμός* (dual-tree complex wavelet transform - DT-CWT) [Kin99, Kin01, SBK05]. Ονομάζεται έτσι γιατί βασίζεται σε ανάπτυγμα με όρους μιγαδικά κυματίδια $\psi_c(t) = \psi_a(t) + j\psi_b(t)$ (ακόμη και για αναπαράσταση πραγματικών σημάτων), με την ανάλυση 1-Δ σημάτων να υλοποιείται με δύο παράλληλες συστοιχίες φίλτρων (δέντρα a και b), όπως φαίνεται στο Σχ. 5.5. Οι συντελεστές λεπτομέρειας d_{ja} και d_{jb} παράγουν αντίστοιχα το πραγματικό και φανταστικό μέρος των μιγαδικών συντελεστών λεπτομέρειας $d_{jc} = d_{ja} + jd_{jb}$ του DT-CWT. Η ανάλυση πραγματικών 1-Δ σημάτων μήκους N με τον DT-CWT παράγει N μιγαδικούς συντελεστές κυματιδίων (από N πραγματικούς συντελεστές σε κάθε κλάδο) και συνιστά αναπαράσταση σε πλαίσιο με βαθμό πλεονασμού 2, ανεξάρτητα από τον αριθμό επιπέδων κυματιδιακής ανάλυσης.



Σχήμα 5.5: Ο 1-Δ δίδεντρος μιγαδικός κυματιδιακός μετασχηματισμός (DT-CWT) αποτελείται από δύο δέντρα ανάλυσης, a και b (το σχήμα υποθέτει ανάλυση σε δύο επίπεδα). Η συστοιχία φίλτρων σε καθένα από τα δύο δέντρα υπολογίζει από έναν πραγματικό DWT με εφαρμογή διαφορετικών ζευγών φίλτρων, $\{h_{0a}, h_{1a}\}$ και $\{h_{0b}, h_{1b}\}$, αντίστοιχα. Οι συντελεστές λεπτομέρειας d_{ja} και d_{jb} παράγουν αντίστοιχα το πραγματικό και φανταστικό μέρος των μιγαδικών συντελεστών λεπτομέρειας του DT-CWT.

Η συστοιχία φίλτρων σε καθένα από τα δύο δέντρα υπολογίζει από έναν πραγματικό DWT με εφαρμογή διαφορετικών ζευγών φίλτρων, $h_a = \{h_{0a}, h_{1a}\}$ και $h_b = \{h_{0b}, h_{1b}\}$, αντίστοιχα, καθένα από τα οποία ικανοποιεί τις συνθήκες τέλει ανακατασκευής της

Ενότητας 5.2.1. Οι πολύ καλές ιδιότητες μεταφορικής συμμετρίας του DT-CWT οφείλονται σε δύο παράγοντες: *Πρώτον*, η παρουσία δύο συστοιχιών φίλτρων έχει ως έμμεσο αποτέλεσμα ότι διπλασιάζεται ο ρυθμός δειγματοληψίας σε κάθε επίπεδο ανάλυσης και συνεπώς μειώνεται ο βαθμός φασματικής αναδίπλωσης· παρόμοια συμπεριφορά έχουν και άλλοι υπερπλήρεις μετασχηματισμοί [Sel01b, SED04]. *Δεύτερο*, τα δύο ζεύγη φίλτρων στα δέντρα a και b σχεδιάζονται από κοινού έτσι ώστε τα αντίστοιχα κυματίδια $\psi_a(t)$ και $\psi_b(t)$ να αποτελούν προσεγγιστικά ζεύγος μετασχηματισμών Hilbert. Για να ισχύει αυτό τα διακριτά φίλτρα στα δέντρα a και b πρέπει να έχουν διαφορά φάσης μισό δείγμα $h_{0a}[n] \approx h_{0b}[n - 0.5]$, ισοδύναμα $H_{0a}(\omega) \approx H_{0b}(\omega)e^{-j0.5\omega}$ [Sel01a, SBK05], δηλαδή το βαθυπερατό φίλτρο στο δέντρο b δειγματοληπτεί σε χωρικά σημεία στο μέσο μεταξύ διαδοχικών δειγμάτων του βαθυπερατού φίλτρου στο δέντρο a . Το προκύπτον μιγαδικό κυματίδιο $\psi_c(t) = \psi_a(t) + j\psi_b(t)$ είναι σε καλή προσέγγιση αναλυτικό σήμα [Pap77] με αμελητέο φασματικό περιεχόμενο στις αρνητικές συχνότητες, γεγονός που μειώνει σημαντικά τη φασματική αναδίπλωση κατά την υποδειγματοληψία. Η κρίσιμη αυτή ιδιότητα χαρακτηρίζει τον DT-CWT και του προσδίδει σχεδόν πλήρη μεταφορική συμμετρία. Για να καλυφθούν οι παραπάνω ιδιαίτερες ανάγκες απαιτείται σχεδίαση φίλτρων κατάλληλων για ανάλυση με τον DT-CWT, με τον DWT στα επιμέρους δέντρα να προκύπτει είτε διορθογώνιος είτε ορθογώνιος· στην ορθογώνια περίπτωση η συνολική αναπαράσταση DT-CWT αποτελεί *σφιχτό πλαίσιο* (tight frame) του αντίστοιχου χώρου. Με συμβολισμό πινάκων και χρησιμοποιώντας τους πίνακες DWT/IDWT από τα επιμέρους δέντρα, ο 1- Δ DT-CWT και ο αντίστροφός του για διακριτά σήματα πεπερασμένου μήκους μπορούν να εκφραστούν ως

$$F_a^{C-1D} = \begin{bmatrix} F_a^a \\ F_a^b \end{bmatrix} \quad (\text{ευθύς}) \quad (5.10\alpha')$$

$$F_s^{C-1D} = \frac{1}{2} \begin{bmatrix} F_s^a & F_s^b \end{bmatrix}, \quad (\text{αντίστροφος}) \quad (5.10\beta')$$

με το πραγματικό και το φανταστικό μέρος του 1- Δ DT-CWT να αντιστοιχούν στο άνω και κάτω μισό του πίνακα F_a^{C-1D} . Περισσότερες πληροφορίες μπορεί να αναζητηθούν στα [Kin01, SBK05] και στις αναφορές αυτών.

Το γεγονός ότι βασίζεται σε ανάπτυγμα με χρήση προσεγγιστικά αναλυτικών κυματιδίων $\psi_c(t) = \psi_a(t) + j\psi_b(t)$ με το πραγματικό και το φανταστικό μέρος να αποτελούν ένα *ορθογωνικό ζεύγος φίλτρων* (quadrature filter pair - QFP) που βρίσκονται σε μεταξύ τους διαφορά φάσης 90° επιτρέπει ως ένα βαθμό να θεωρούμε τον DT-CWT ως μια εντοπισμένη στο χώρο και την κλίμακα εκδοχή του μετασχηματισμού Fourier, ο οποίος επίσης βασίζεται σε ανάπτυγμα σε ορθογωνικές συνιστώσες $e^{j\Omega t} = \cos(\Omega t) + j\sin(\Omega t)$. Για χαρακτηριστικά που βρίσκονται εντός της *περιοχής υποστήριξης* (support) του φίλτρου, το μέτρο των συντελεστών λεπτομέρειας d_{jc} του DT-CWT κωδικοποιεί την έντασή τους ενώ η φάση τη θέση τους. Για παράδειγμα, το αποτέλεσμα της μετατόπισης του σήματος εισόδου κατά ένα μικρό χωρικό βήμα (σε σχέση με το μήκος των φίλτρων ανάλυσης) μεταφράζεται με πολύ καλή προσέγγιση σε απλή γραμμική μεταβολή της φάσης του σήματος εισόδου, γεγονός που έχει χρησιμοποιηθεί με επιτυχία για τον υπολογισμό του πεδίου οπτικής ροής σε ακολουθίες εικόνων [MK98]. Η παραπάνω συζήτηση αναδεικνύει αρκετές ομοιότητες μεταξύ του DT-CWT και της ανάλυσης με άτομα Gabor [SBK05], τα οποία συζητάμε με μεγαλύτερη λεπτομέρεια στην Ενότη. 5.2.3. Η ουσιαστικότερη διαφορά μεταξύ των δύο προσεγγίσεων έγκειται στη διαφορετική στρατηγική τοποθέτησης των φίλτρων

στο διάγραμμα χρόνου-συχνότητας, με τον DT-CWT να υιοθετεί τη δειγματοληψία του Σχ. 5.2(α) και τη Gabor ανάλυση να υιοθετεί τη δειγματοληψία του Σχ. 5.2(β).

Ο DT-CWT επεκτείνεται κατάλληλα για ανάλυση εικόνων και γενικότερα d -διάστατων σημάτων, με την προκύπτουσα αναπαράσταση να έχει βαθμό πλεονασμού 2^d (4 στην περίπτωση εικόνων) [Kin99, SBK05]. Συγκεκριμένα, για ανάλυση εικόνων προχωρούμε σε διαχωρίσιμη ανάλυση με τον κρίσιμα δειγματοληπτημένο DWT με τέσσερις διαφορετικούς τρόπους, χρησιμοποιώντας όλους τους δυνατούς συνδυασμούς των φίλτρων h_a και h_b για ανάλυση κατά στήλες και γραμμές της εικόνας. Συμβολίζοντας π.χ. με F_a^{ab} τον ευθύ DWT μετασχηματισμό που αντιστοιχεί σε ανάλυση κατά στήλες με τα h_a φίλτρα και κατά γραμμές με τα h_b φίλτρα και παρόμοια για τους υπόλοιπους τρεις συνδυασμούς, προκύπτουν οι τέσσερις ευθείς DWT μετασχηματισμοί $\{F_a^{aa}, F_a^{bb}, F_a^{ba}, F_a^{ab}\}$ και οι αντίστοιχοι αντίστροφοί τους $\{F_s^{aa}, F_s^{bb}, F_s^{ba}, F_s^{ab}\}$, τους οποίους συνδυάζουμε για να προκύψει ο 2-Δ DT-CWT ως [SBK05]

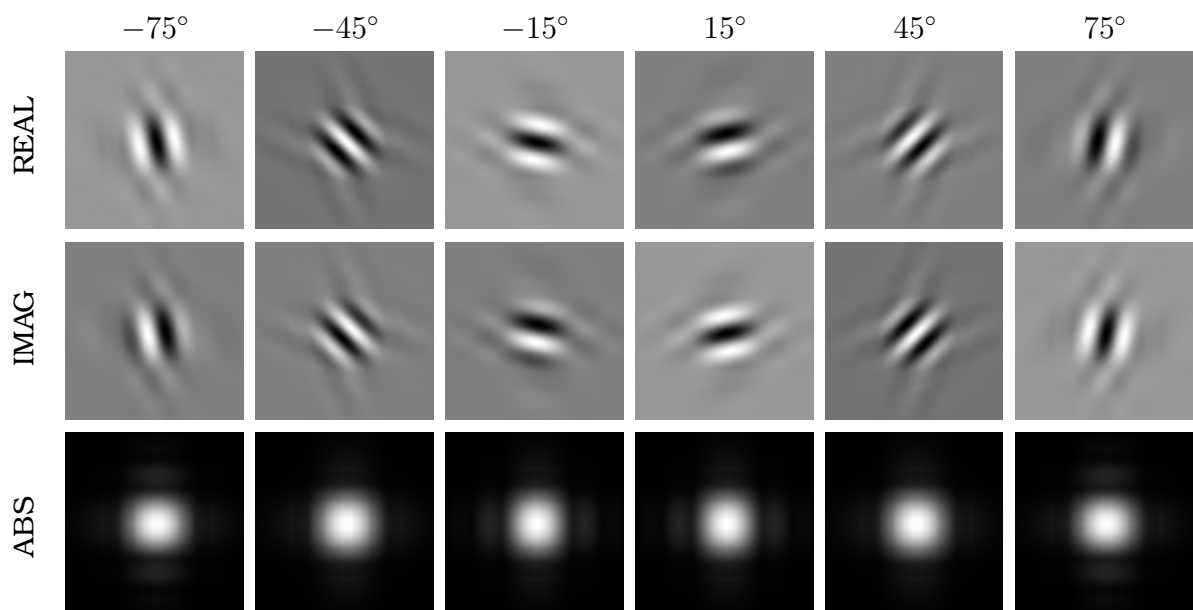
$$F_a^{C-2D} = \frac{1}{\sqrt{8}} \begin{bmatrix} I & -I & & \\ I & I & & \\ & & I & I \\ & & I & -I \end{bmatrix} \begin{bmatrix} F_a^{aa} \\ F_a^{bb} \\ F_a^{ba} \\ F_a^{ab} \end{bmatrix} \quad (\text{ευθύς}) \quad (5.11\alpha)$$

$$F_s^{C-2D} = \frac{1}{\sqrt{8}} \begin{bmatrix} F_s^{aa} & F_s^{bb} & F_s^{ba} & F_s^{ab} \end{bmatrix} \begin{bmatrix} I & I & & \\ -I & I & & \\ & & I & I \\ & & I & -I \end{bmatrix} \quad (\text{αντίστροφος}) \quad (5.11\beta)$$

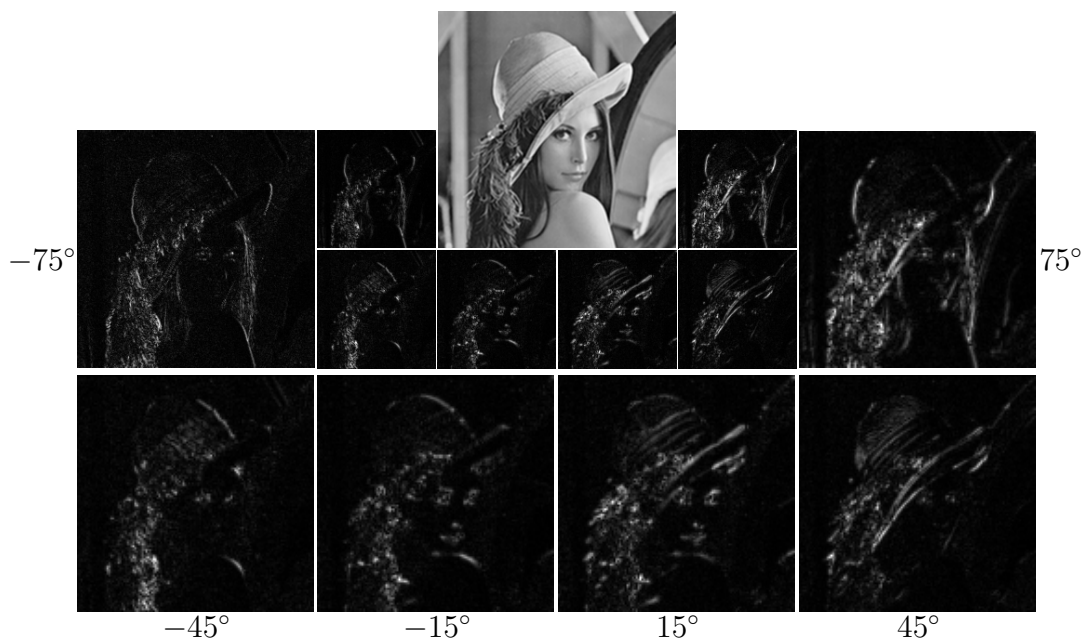
όπου έχουμε ακολουθήσει τη σύμβαση ότι το πραγματικό και το φανταστικό μέρος του DT-CWT αντιστοιχούν στο άνω και κάτω μισό του πίνακα F_a^{C-2D} . Πέρα από μεταφορική συμμετρία, ο 2-Δ DT-CWT χαρακτηρίζεται από καλή περιστροφική συμμετρία. Συγκεκριμένα, κάθε κλίμακα αποτελείται από έξι μιγαδικές ισχυρά κατευθυντικές υποζώνες που καλύπτουν ομοιόμορφα το εύρος των δυνατών διευθύνσεων με βήμα 30° , έχοντας προσανατολισμό $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$, όπως φαίνεται στο Σχ. 5.6, όπου επίσης δείχνουμε ένα παράδειγμα ανάλυσης της εικόνας Lena με τον 2-Δ DT-CWT.

5.2.3 Τοπικά Αρμονική Αναπαράσταση Χώρου-Συχνότητας και ο Διαμορφωμένος Μιγαδικός Επικαλυπτόμενος Μετασχηματισμός

Όπως έχουμε συζητήσει, οι εικόνες του φυσικού κόσμου δεν μπορεί να θεωρηθούν ως καθολικά στάσιμες στοχαστικές διαδικασίες, συνεπώς ο καθολικός μετασχηματισμός Fourier της Ενότητας 5.1 αποτελεί φτωχή αναπαράσταση για αυτές. Η κυματιδιακή αναπαράσταση χώρου-κλίμακας με χρήση του DWT (Ενότητα 5.2.1) ή του DT-CWT (Ενότητα 5.2.2) είναι κατάλληλη για μη στάσιμες, τμηματικά ομαλές εικόνες, τις οποίες και μπορεί να περιγράψει με μικρό αριθμό μη μηδενικών συντελεστών συγκεντρωμένων γύρω από τις ακμές τους. Μια εναλλακτική προσέγγιση που συζητούμε εδώ βασίζεται στην υπόθεση ότι οι εικόνες είναι *τοπικά στάσιμες* στοχαστικές διαδικασίες, που οδηγεί σε τοπικά αρμονικές αναπαραστάσεις, επίσης επονομαζόμενες αναπαραστάσεις χώρου-συχνότητας, που είναι ιδιαίτερα κατάλληλες για περιγραφή ταλαντούμενων κυματομορφών σε εικόνες υφής. Αυτή η προσέγγιση φέρει αναλογίες με τη χρήση του *βραχύχρονου μετασχηματισμού Fourier* (short-time Fourier transform – STFT) στην ανάλυση ακουστικών σημάτων [RS78].



(α)



(β)

Σχήμα 5.6: (α) Τα έξι μιγαδικά κατευθυντικά κυματίδια που αντιστοιχούν στον 2-Δ DT-CWT. Δείχνουμε τόσο το πραγματικό όσο και το φανταστικό τους μέρος, καθώς και το μέτρο τους. (β) Ανάλυση της 512×512 εικόνας Lena σε δύο επίπεδα με τον 2-Δ DT-CWT. Η διάσταση καθεμιάς από τις υποζώνες είναι 256×256 και 128×128 εικονοστοιχεία για το πρώτο και δεύτερο επίπεδο ανάλυσης, αντίστοιχα, ενώ η διάσταση των πραγματικών συντελεστών κλιμάκωσης είναι 256×256 εικονοστοιχεία. Σύγκρινε με το αντίστοιχο Σχ. 5.4 για τον πραγματικό DWT.

Αναπαράσταση εικόνων με χωρο-συχνотικά άτομα Gabor

Για ένα 1-Δ σήμα $x(t)$, η αναπαράσταση χρόνου-συχνότητας βασίζεται σε ανάπτυγμα σε πολλαπλά χρονικά εντοπισμένα και ημιτονοειδώς διαμορφωμένα άτομα της μορφής

$$g_{\omega_0, t_0}(t) = e^{j\omega_0 t} g(t - t_0), \quad (5.12)$$

που αντιστοιχεί σε τοπικά εντοπισμένη Fourier ανάλυση

$$X_{t_0}(\omega) = \langle x, g_{\omega, t_0} \rangle = \int x(t) g(t - t_0) e^{-j\omega t} dt, \quad (5.13)$$

όπου χρησιμοποιούμε τον συνηθισμένο ορισμό $\langle f, g \rangle = \int f(t) g^*(t) dt$ για το εσωτερικό γινόμενο δύο μιγαδικών συναρτήσεων του $L^2(R)$. Η επιλογή γκαουσιανού παραθύρου ανάλυσης $g(t) = (2\pi\sigma^2)^{-1/2} e^{-t^2/(2\sigma^2)}$ οδηγεί στα κλασικά φίλτρα Gabor, τα οποία και επιτυγχάνουν το μέγιστο επιτρεπόμενο από την 1-Δ σχέση αβεβαιότητας εντοπισμό από κοινού στο χρόνο και στη συχνότητα [Gab46]. Μπορεί κανείς να θεωρήσει τη διακριτή οικογένεια ατόμων $\{g_{k\omega_0, nt_0}\}_{(k,n) \in \mathbb{Z}^2}$ που προκύπτει με ομοιόμορφη δειγματοληψία του επιπέδου χρόνου-συχνότητας με βήμα t_0, ω_0 , αντίστοιχα, όπως φαίνεται στο Σχ. 5.2(β). Αναγκαία συνθήκη για αυτήν να αποτελεί πλαίσιο του $L^2(R)$ είναι η πυκνότητα δειγματοληψίας να υπερβαίνει το κατώφλι $(2\pi)/(t_0\omega_0) \geq 1$, που σχετίζεται με τον ελάχιστο ρυθμό δειγματοληψίας που προβλέπει το κριτήριο του Nyquist [Dau90]. Για $(2\pi)/(t_0\omega_0) = 1$ έχουμε κρίσιμη δειγματοληψία που επιτρέπει δημιουργία Riesz βάσεων. Δυστυχώς όμως σε αυτήν την περίπτωση θα πρέπει η αντίστοιχη συνάρτηση παραθύρωσης $g(t)$ να είναι είτε μη ομαλή, είτε να έχει άπειρη υποστήριξη (θεώρημα Balian-Low)· σε αντίθετη περίπτωση, π.χ. για φίλτρα Gabor, η αναπαράσταση είναι μη ευσταθής (το κάτω φράγμα πλαισίου είναι 0). Για $(2\pi)/(t_0\omega_0) > 1$ έχουμε υπερδειγματοληψία και μπορούμε να κατασκευάσουμε πλαίσια. Στην περίπτωση που τα πλαίσια αυτά είναι δεν είναι σφιχτά (άνω και κάτω φράγμα πλαισίου διαφέρουν), η ανακατασκευή παρουσιάζει δυσκολίες, καθώς το δυϊκό πλαίσιο δεν είναι εν γένει γνωστό· για φίλτρα Gabor το πλαίσιο πλησιάζει να γίνει σφιχτό, οπότε πρωτεύον και δυϊκό πλαίσιο τείνουν να ταυτιστούν, καθώς ο βαθμός πλεονασμού $(2\pi)/(t_0\omega_0)$ αυξάνει, π.χ. υπερβαίνει το 4. Ο αναγνώστης παραπέμπεται στα [Dau90, Mal99a, QC99] για περισσότερες πληροφορίες, αποδοτικούς αλγορίθμους για 1-Δ σήματα, και επισκόπηση της σχετικής βιβλιογραφίας.

Η 2-Δ επέκταση των φίλτρων Gabor, όπου το γκαουσιανό παράθυρο ανάλυσης έχει πλέον ελλειπτικό σχήμα και η διαμόρφωση γίνεται με 2-Δ ημιτονοειδές σήμα, έχει αποδειχθεί ιδιαίτερα χρήσιμη για ανάλυση εικόνων [Dau85, Dau88, PZ88, BCG90, Lee96]. Σε δύο διαστάσεις, μπορεί κανείς να σχεδιάσει συστοιχίες φίλτρων Gabor κατανεμημένων στον 2-Δ χώρο/ 2-Δ χωρική συχνότητα, οι οποίες αποτελούνται από φίλτρα των οποίων παράμετροι όπως χωρικό κέντρο, προσανατολισμός και κεντρική χωρική συχνότητα έχουν επιλεγεί κατάλληλα για τις ανάγκες συγκεκριμένων εφαρμογών. Όπως και οι 1-Δ εκδοχές τους, τα 2-Δ μιγαδικά φίλτρα Gabor συνεχίζουν να επιτυγχάνουν το μέγιστο δυνατό από κοινού εντοπισμό στο χώρο και στη συχνότητα που επιτρέπει η 2-Δ αρχή απροσδιοριστίας [Dau85]. Επίσης στοιχειώδη άτομα τύπου Gabor ανακλύπουν αν προσπαθήσει κανείς να μοντελοποιήσει φυσικές εικόνες με αθροιστικά αναγεννητικά μοντέλα με αραιές ή ανεξάρτητες αποκρίσεις [OF96, BS97, HH00]. Επιπλέον σημαντικό επιχείρημα που ενθαρρύνει τη χρήση μιγαδικών φίλτρων Gabor στην ανάλυση εικόνων είναι το γεγονός ότι προσομοιάζουν το προφίλ απόκρισης στο υποδεκτικό πεδίο (receptive field) των

σύνθετων κυττάρων (complex cells) στον *πρωτοταγή οπτικό φλοιό - περιοχή V1* (primary visual cortex - V1 area) θηλαστικών, με το αντίστοιχο πραγματικό ή φανταστικό τους μέρος να ταιριάζει με το προφίλ απόκρισης του υποκείμενου ζεύγους *απλών κυττάρων* (simple cells) [Dau80, Dau85, AB85] - για μια πρόσφατη επισκόπηση της λειτουργίας του πρώιμου σταδίου όρασης σε βιολογικά συστήματα παραπέμπουμε στο [CDM⁺05].

Παρόλα τα σημαντικά του πλεονεκτήματα για *ανάλυση* εικόνων, ο Gabor μετασχηματισμός παρουσιάζει μειωμένη ευελιξία για *σύνθεση* εικόνων. Με κατάλληλη σχεδίαση της συστοιχίας φίλτρων, ο 2-Δ Gabor μετασχηματισμός συνιστά πλαίσιο του $L^2(R^2)$ και είναι συνεπώς αναστρέψιμος, αν και στην πράξη το δυϊκό πλαίσιο δεν είναι εν γένει διαθέσιμο [Lee96]. Μπορεί κανείς να χρησιμοποιήσει μια αναγεννητική εκδοχή του Gabor μετασχηματισμού, όπου μια εικόνα σχηματίζεται ως υπέρθεση σταθμισμένων 2-Δ Gabor κυματιδίων, και να βρει τους συντελεστές που ελαχιστοποιούν το σφάλμα ανακατασκευής με τη μέθοδο των κλίσεων όπως στο [Dau88], αν και η αντίστοιχη επαναληπτική διαδικασία συγκλίνει αργά. Μια εναλλακτική προσέγγιση είναι να αυξήσει κανείς το βαθμό πλεονασμού του μετασχηματισμού. Με κατάλληλη σχεδίαση η προκύπτουσα αναπαράσταση τείνει να γίνει σφικτό πλαίσιο, οπότε δυϊκό και πρωτεύον πλαίσιο τείνουν να ταυτιστούν [Lee96]. Πάντως, σε κάθε περίπτωση, ο 2-Δ Gabor μετασχηματισμός είναι δυσκολότερα αντιστρέψιμος από τον 2-Δ κυματιδιακό μετασχηματισμό, και αυτό τον καθιστά λιγότερο εύχρηστο σε εφαρμογές όπου απαιτείται όχι μόνο ανάλυση αλλά και σύνθεση οπτικών σημάτων.

Αναπαράσταση εικόνων σε ανεξάρτητα μπλοκ

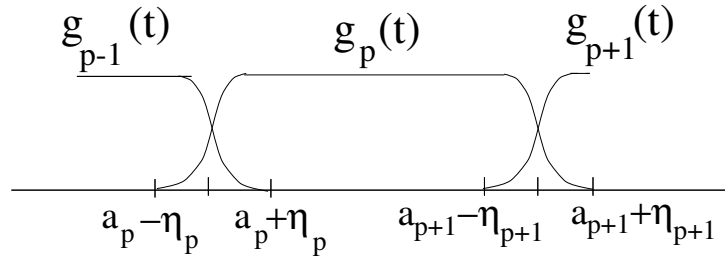
Μια εναλλακτική προσέγγιση στη μοντελοποίηση μη στάσιμων σημάτων που είναι ιδιαίτερα δημοφιλής για κωδικοποίηση εικόνων βασίζεται στη χρήση μπλοκ αναπαραστάσεων. Έστω το 1-Δ σήμα $x(t)$ · διαμερίζουμε τον χρονικό άξονα σε διαδοχικά διαστήματα $[a_p, a_{p+1}]$ μήκους $l_p = a_{p+1} - a_p$ και θεωρούμε το αντίστοιχο τετραγωνικό παράθυρο ανάλυσης $g_p(t) = 1$, εάν $a_p \leq t \leq a_{p+1}$, 0 διαφορετικά. Εάν $\{e_k\}_{k \in Z}$ είναι ορθοκανονική βάση στο διάστημα $L^2([0, 1])$, τότε η οικογένεια $\{g_{p,k}\}_{(p,k) \in Z^2}$, με

$$g_{p,k}(t) = g_p(t) \frac{1}{\sqrt{l_p}} e_k \left(\frac{t - a_p}{l_p} \right) \quad (5.14)$$

αποτελεί ορθοκανονική βάση του $L^2(R)$ [Mal99a]. Μπορεί κανείς αναλόγως να κατασκευάσει βάσεις για διακριτά σήματα καθώς και για 2-Δ εικόνες χωρισμένες σε μη επικαλυπτόμενα ορθογώνια μπλοκ, καθένα από τα οποία μετασχηματίζεται ανεξάρτητα από τα άλλα. Με χρήση Fourier μετασχηματισμού εντός κάθε μπλοκ οδηγείται κανείς σε αναπαραστάσεις χρόνου-συχνότητας (1-Δ) ή χώρου-συχνότητας (2-Δ). Συνήθως προτιμάται αντί του Fourier μετασχηματισμού η χρήση συνημιτονοειδών αναπαραστάσεων και συγκεκριμένα κάποιος εκδοχής του *διακριτού συνημιτονοειδούς μετασχηματισμού* (discrete cosine transform - DCT) που απαλύνει τον αντίκτυπο του μη ομαλού τετραγωνικού παραθύρου κατά την φασματική ανάλυση [Str99], αν και η ύπαρξη χαρακτηριστικών *μπλοκ ψευδέργων* (block artifacts) παραμένει βασικό μειονέκτημα των μπλοκ αναπαραστάσεων. Ο μπλοκ DCT μετασχηματισμός αποτελεί τη βάση αναπαράστασης σε πολλές τεχνικές κωδικοποίησης εικόνας και βίντεο, όπως για παράδειγμα το διεθνές πρότυπο JPEG για συμπίεση εικόνων [Wal91, Bov05, Str99].

Αναπαράσταση εικόνων σε αλληλεπικαλυπτόμενα μπλοκ

Προκειμένου να εξαλειφθούν τα ψεύδεργα που είναι εγγενή σε μπλοκ μετασχηματισμούς, έχουν αναπτυχθεί μέθοδοι για ανάλυση σημάτων σε αλληλεπικαλυπτόμενα μπλοκ [PJB87, MS89a, Mal90, Mal92, Mey93, VK95, Mal99a]. Οι προκύπτοντες μπλοκ επικαλυπτόμενοι μετασχηματισμοί χρησιμοποιούν ομαλά φίλτρα ανάλυσης ενώ ταυτόχρονα επιτρέπουν τέλεια ανακατασκευή με χρήση ταχέων αλγορίθμων, συνδυάζοντας έτσι βασικά πλεονεκτήματα των Gabor και μπλοκ αναπαραστάσεων.



Σχήμα 5.7: Αλληλεπικαλυπτόμενα παράθυρα ανάλυσης μεταξύ διαδοχικών μπλοκ. Σχήμα δανεισμένο από το [Mal99a, Σχ. 8.17].

Πιο συγκεκριμένα, μπορεί κανείς και πάλι να κατασκευάσει ορθοκανονική βάση του $L^2(R)$ ξεκινώντας από μια ορθοκανονική βάση $\{e_k\}_{k \in Z}$ στο διάστημα $L^2([0, 1])$. Το παράθυρο ανάλυσης g_p είναι ομαλά αποσβενόμενο και μη μηδενικό στο διάστημα $[a_p - \eta_p, a_{p+1} + \eta_{p+1}]$, που σημαίνει ότι διαδοχικά μπλοκ αλληλεπικαλύπτονται, όπως φαίνεται στο Σχ. 5.7. Το αντίστοιχο διάγραμμα Heisenberg απεικονίζεται στο Σχ. 5.2(β). Προκειμένου να εξασφαλίζεται τέλεια ανακατασκευή, θα πρέπει η συνάρτηση παραθύρωσης g_p στο εσωτερικό διάστημα $[a_p + \eta_p, a_{p+1} - \eta_{p+1}]$ να ισούται με μονάδα ενώ στα διαστήματα αλληλεπικάλυψης με το αριστερό και δεξί μπλοκ να ικανοποιεί τις σχέσεις

$$g_p(t) = \begin{cases} \beta((t - a_p)/\eta_p) & t \in [a_p - \eta_p, a_p + \eta_p] \\ \beta((a_{p+1} - t)/\eta_{p+1}) & t \in [a_{p+1} - \eta_{p+1}, a_{p+1} + \eta_{p+1}], \end{cases} \quad (5.15)$$

με το προφίλ αλληλεπικάλυψης β να ικανοποιεί τη σχέση συμπληρωματικότητας ισχύος $\beta^2(t) + \beta^2(-t) = 1$ στην περιοχή αλληλεπικάλυψης $t \in [-1, 1]$, ενώ $\beta(t) = 0, t < -1$ και $\beta(t) = 1, t > 1$. Για παράδειγμα, ένα κατάλληλο ομαλό παράθυρο παράγεται από το ημιτονοειδές προφίλ $\beta(t) = \sin\left(\frac{\pi}{4}(1+t)\right), t \in [-1, 1]$. Υπό αυτές τις συνθήκες, μπορεί να δειχθεί ότι η οικογένεια $\{g_{p,k}\}_{(p,k) \in Z^2}$, με

$$g_{p,k}(t) = g_p(t) \frac{1}{\sqrt{l_p}} \tilde{e}_k\left(\frac{t - a_p}{l_p}\right) \quad (5.16)$$

αποτελεί ορθοκανονική βάση του $L^2(R)$. Στην παραπάνω σχέση τα $\tilde{e}_k(t)$ αποτελούν επεκτάσεις των $e_k(t)$ στο R με περίοδο 4, συμμετρικές ως προς το 0 και αντισυμμετρικές ως προς το 1, που ορίζονται από τη σχέση

$$\tilde{e}_k(t) = \begin{cases} e_k(t) & t \in [0, 1] \\ e_k(-t) & t \in (-1, 0) \\ -e_k(2-t) & t \in [1, 2) \\ -e_k(2+t) & t \in [-1, -2). \end{cases} \quad (5.17)$$

Μια ενδιαφέρουσα ειδική περίπτωση με πρακτικό ενδιαφέρον προκύπτει θεωρώντας τη DCT-IV βάση $e_k(t) = \sqrt{2} \cos[(k + 0.5)\pi t]$, $t \in [0, 1]$, που λόγω της συμμετρίας της παράγει ομαλή επέκταση $\tilde{e}_k(t) = \sqrt{2} \cos[(k + 0.5)\pi t]$, $t \in R$, με αποτέλεσμα η προκύπτουσα βάση να αποτελείται από ομαλές συναρτήσεις $g_{p,k}(t) = g_p(t) \sqrt{\frac{2}{l_p}} \cos \left[\frac{\pi}{l_p} (k + 0.5)(t - a_p) \right]$.

Ανάλογα με τη συνεχή περίπτωση, μπορούν να οριστούν αλληλεπικαλυπτόμενοι μπλοκ μετασχηματισμοί για διακριτά σήματα $x[n]$. Αντίστοιχα, θεωρούμε την αύξουσα ακολουθία $\{a_p\}_{p \in Z}$, με $a_p + 0.5 \in Z$, που ορίζει διαδοχικά διαστήματα $[a_p, a_{p+1}]$ που το καθένα περιέχει $l_p = a_{p+1} - a_p$ δείγματα. Θεωρούμε επίσης ομαλά αποσβεσμένα παράθυρα που προκύπτουν με ομοιόμορφη δειγματοληψία των αντίστοιχων συνεχών παραθύρων που ικανοποιούν την Εξ. (5.15), δηλαδή $g_p[n] = g_p(n)$. Μπορεί πάλι να δειχθεί ότι με αφετηρία μια ορθοκανονική γεννήτρια βάση $\{e_{l_p,k}[n]\}_{0 \leq k < l_p}$ με διανύσματα ορισμένα για $0 \leq n < l_p$ μπορεί κανείς να κατασκευάσει επικαλυπτόμενη ορθοκανονική βάση $\{g_{p,k}[n]\}$ του $l^2(Z)$ με παρόμοιο τρόπο με τη συνεχή περίπτωση (5.16), επεκτείνοντας τα διανύσματα βάσης με διακριτή εκδοχή της (5.17). Επίσης, ο υπολογισμός του διακριτού επικαλυπτόμενου μετασχηματισμού $X_p[k] = \langle x, g_{p,k} \rangle$, $0 \leq k < l$, μπορεί να γίνει αποδοτικά με εσωτερικά γινόμενα κατάλληλα αναδιπλωμένης εκδοχής του σήματος \tilde{x} και των διανυσμάτων της γεννήτριας βάσης $e_{l_p,k}$, $X_p[k] = \langle \tilde{x}, e_{l_p,k} \rangle$. Η διαδικασία ανάλυσης περιλαμβάνει διαχωρισμό του σήματος σε αλληλεπικαλυπτόμενα μπλοκ, παραθύρωση, συστολική αναδίπλωση, μετασχηματισμό μπλοκ l_p δειγμάτων. Η διαδικασία σύνθεσης περιλαμβάνει αντιστροφή του μπλοκ μετασχηματισμού, διαστολική αναδίπλωση, παραθύρωση, παράθεση και άθροιση των αλληλεπικαλυπτόμενων μπλοκ. Για περισσότερες λεπτομέρειες παραπέμπουμε στο [Mal99a, Ενότητα 8.4.4]. Μια ειδική περίπτωση αποτελεί η επικαλυπτόμενη βάση προερχόμενη από τη διακριτή DCT-IV βάση $e_{l,k}[n] = \sqrt{\frac{2}{l}} \cos \left[\frac{\pi}{l} (k + 0.5)(n + 0.5) \right]$, $0 \leq k < l$, με ομαλά διανύσματα βάσης που δίνονται από τη σχέση $g_{p,k}[n] = g_p[n] \sqrt{\frac{2}{l_p}} \cos \left[\frac{\pi}{l_p} (k + 0.5)(n - a_p) \right]$, με $0 \leq k < l_p$. Μια άλλη ιδιαίτερα ενδιαφέρουσα περίπτωση αποτελεί ο *διαμορφωμένος επικαλυπτόμενος μετασχηματισμός* (modulated lapped transform - MLT) [PJB87, Mal90, Mal92, Shl97], γνωστός και ως *παραλληλαγμένος διακριτός συνημιτονικός μετασχηματισμός* (modified discrete cosine transform - MDCT), ο οποίος χρησιμοποιείται ευρύτατα για συμπίεση ακουστικών σημάτων [PS00a]. Στον MLT όλα τα μπλοκ έχουν το ίδιο μήκος $l = l_p = a_{p+1} - a_p$, ενώ τα διανύσματα βάσης (l ανά μπλοκ) έχουν μήκος $2l$ (50% βαθμός επικάλυψης μεταξύ διαδοχικών παραθύρων) και ορίζονται από τη σχέση

$$g_{p,k}[n] = g_p[n] \sqrt{\frac{2}{l}} \cos \left[\frac{\pi}{l} (k + 0.5) \left(n - a_p + \frac{l}{2} \right) \right], \quad 0 \leq k < l, \quad 0 \leq n - a_p - 0.5 < 2l, \quad (5.18)$$

με ημιτονοειδές παράθυρο ανάλυσης $g_p[n] = \sin \left(\frac{\pi}{2l} (n - a_p) \right)$, $0 \leq n - a_p - 0.5 < 2l$. Σημαντικό πλεονέκτημα του MLT και άλλων ημιτονικά διαμορφωμένων μετασχηματισμών είναι ότι οι ίδιοι και οι αντιστροφοί τους μπορούν να υπολογιστούν αποδοτικά με πολυπλοκότητα $\mathcal{O}(l \log l)$ ανά μπλοκ, αφού το στάδιο διαμόρφωσης μπορεί να αναχθεί σε υπολογισμό DCT/DFT [Mal90, Vai93, Mal99a]· το γεγονός αυτό έχει δώσει σημαντική ώθηση στην υιοθέτησή τους σε πρακτικές εφαρμογές.

Μεταφορικά και περιστροφικά συμμετρική αναπαράσταση εικόνων με το διαμορφωμένο μιγαδικό αλληλεπικαλυπτόμενο μετασχηματισμό

Οι ορθοκανονικοί αλληλεπικαλυπτόμενοι μπλοκ μετασχηματισμοί που έχουμε συζητήσει ως τώρα αποτελούν κρίσιμα δειγματοληπτημένες αναπαραστάσεις για 1-Δ σήματα και 2-Δ εικόνες. Παρόμοια με τον κρίσιμα δειγματοληπτημένο κυματιδιακό μετασχηματισμό, ο αλληλεπικαλυπτόμενος μπλοκ μετασχηματισμός χαρακτηρίζεται από έλλειψη μεταφορικής συμμετρίας, όπως έχουμε συζητήσει στην Ενότητα 5.2.2. Για παράδειγμα [Mal99a, Ενότητα 8.4.3], αν στον MLT δωθεί ως είσοδος ένα από τα διανύσματα βάσης $x[n] = g_{p,k}[n]$, βλ. Εξ. (5.18), μόνο ο αντίστοιχος συντελεστής $X_p[k]$ θα είναι μη μηδενικός· αν καθυστερήσουμε το σήμα εισόδου κατά $\Delta n = l/(2k+1)$ δείγματα, το ολισθημένο σήμα θα έχει διαφορά φάσης 90° και άρα θα είναι σχεδόν ορθογώνιο του $g_{p,k}[n]$, με αποτέλεσμα η ενέργεια του σήματος να κατανεμηθεί σε μεγάλο αριθμό συντελεστών $X_p[k]$.

Παρόμοια με την περίπτωση του κυματιδιακού μετασχηματισμού της Ενότητας 5.2.2, μπορεί κανείς να κατασκευάσει υπερπλήρεις μιγαδικούς αλληλεπικαλυπτόμενους μετασχηματισμούς με αισθητά βελτιωμένα χαρακτηριστικά μεταφορικής συμμετρίας [YK93, Mal99b]. Ιδιαίτερο ενδιαφέρον παρουσιάζει ο *διαμορφωμένος μιγαδικός αλληλεπικαλυπτόμενος μετασχηματισμός* (modulated complex lapped transform - MCLT), ο οποίος αποτελεί μιγαδική επέκταση του MLT [Mal99b, Mal03]. Οι MCLT συντελεστές υπολογίζονται ως $X_p[k] = \langle x, g_{p,k} \rangle$, όπου τα διανύσματα $g_{p,k}$ έχουν μήκος $2l$ και είναι μιγαδικές επεκτάσεις αυτών του MLT, δηλαδή

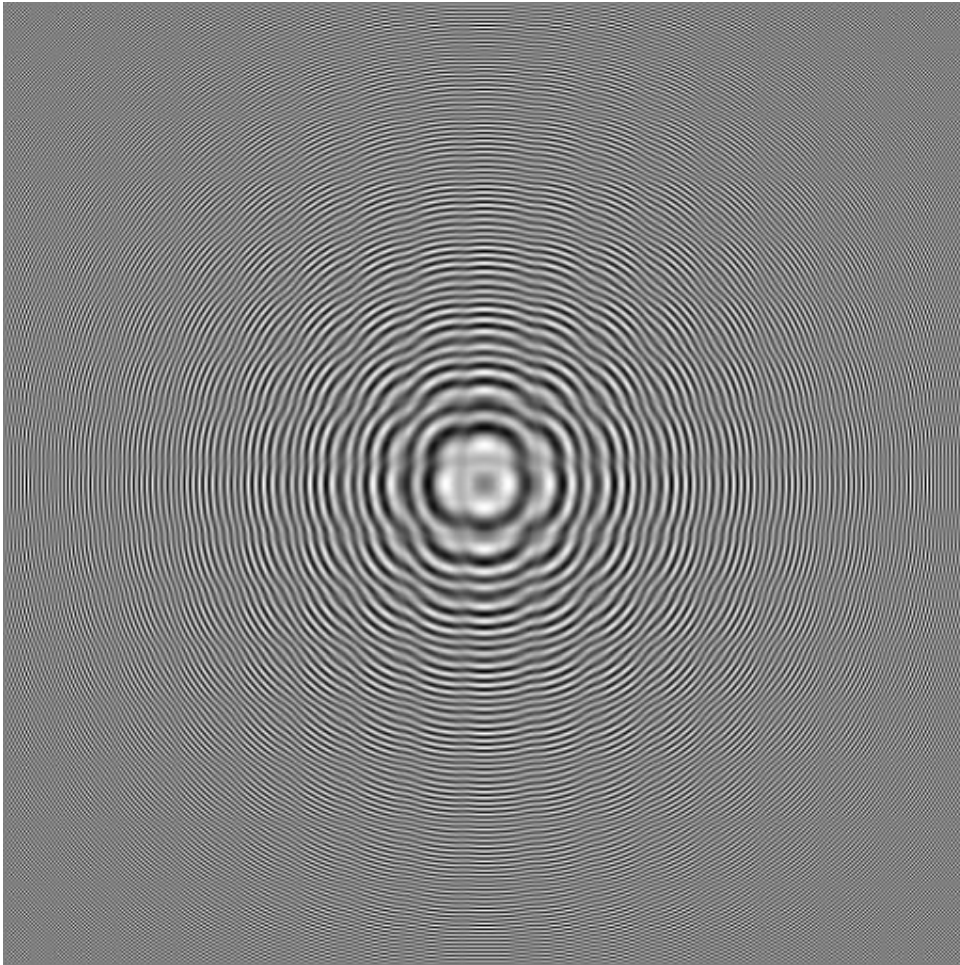
$$g_{p,k}[n] = g_p[n] \sqrt{\frac{2}{l}} \exp \left[\frac{j\pi}{l} (k+0.5) \left(n - a_p + \frac{l}{2} \right) \right], \quad 0 \leq n - a_p - 0.5 < 2l. \quad (5.19)$$

όπου και πάλι χρησιμοποιούμε ημιτονοειδές παράθυρο ανάλυσης. Όταν η είσοδος $x[n]$ είναι πραγματικό 1-Δ σήμα, υπολογίζουμε l μιγαδικούς συντελεστές ανά μπλοκ, $X_p[k], 0 \leq k < l$, το πραγματικό μέρος των οποίων ταυτίζεται με τους αντίστοιχους συντελεστές MLT. Ο MCLT ορίζεται και για μιγαδικά 1-Δ σήματα εισόδου, οπότε υπολογίζουμε $2l$ μιγαδικούς συντελεστές ανά μπλοκ, δηλαδή $X_p[k], 0 \leq k < 2l$. Και στις δύο περιπτώσεις ο MCLT είναι δύο φορές υπερπλήρης. Ο υπολογισμός του MCLT και του αντίστροφου του μπορεί να γίνει σε κάθε μπλοκ με ταχύ αλγόριθμο βασισμένο σε FFT μήκους $2l$ δειγμάτων [Mal03], τον οποίο έχουμε κατάλληλα επεκτείνει ώστε να καλύπτει και την περίπτωση μιγαδικών σημάτων εισόδου.

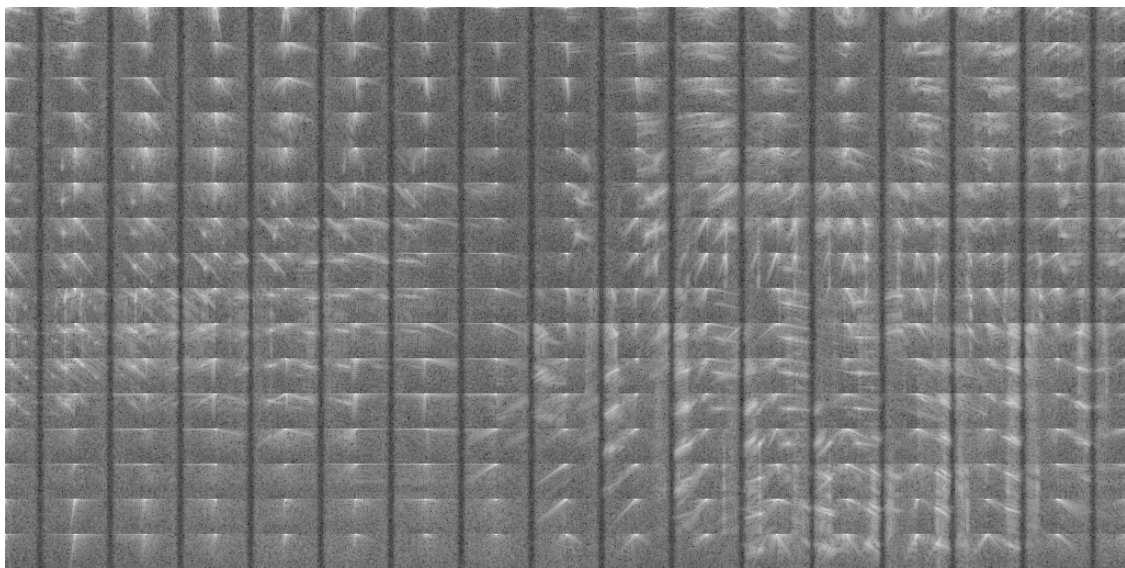
Για ανάλυση εικόνων χρησιμοποιούμε διαχωρίσιμη 2-Δ εκδοχή του MCLT, παρόμοια με την προσέγγιση στο [YK93]. Για μια εικόνα διάστασης $M \times N$ εφαρμόζουμε πρώτα τον 1-Δ πραγματικό MCLT στις στήλες της εικόνας, παίρνοντας μια μιγαδική $M \times N$ διάταξη συντελεστών· κατόπιν εφαρμόζουμε κατά γραμμές τον 1-Δ μιγαδικό MCLT, παίρνοντας την τελική μιγαδική $M \times 2N$ διάταξη συντελεστών MCLT της εικόνας.

Η 2-Δ MCLT αναπαράσταση έχει καλές ιδιότητες περιστροφικής συμμετρίας. Στο Σχ. 5.8 παραθέτουμε τα MCLT άτομα που αντιστοιχούν σε ανάλυση σε 16×16 μπλοκ και παρατηρούμε ότι συνθέτουν ένα μωσαϊκό σχεδόν ακτινικά συμμετρικό. Σε αντίθεση με τον DT-CWT όπου τα άτομα είναι ευθυγραμμισμένα σε μία από τις έξι διακριτές διευθύνσεις $\pm 15^\circ, \pm 45^\circ, \pm 75^\circ$, η διεύθυνση μέγιστης απόκρισης των MCLT ατόμων μεταβάλλεται με πιο ομαλό τρόπο. Παρόλα αυτά μπορεί κανείς ακόμα να χρησιμοποιήσει τις παραπάνω έξι διακριτές διευθύνσεις για να ομαδοποιήσει τα MCLT άτομα παρόμοια με αυτά του DT-CWT, ώστε οι δύο αναπαραστάσεις να έρθουν σε συγκρίσιμη μορφή. Στο Σχ. 5.9 δείχνουμε

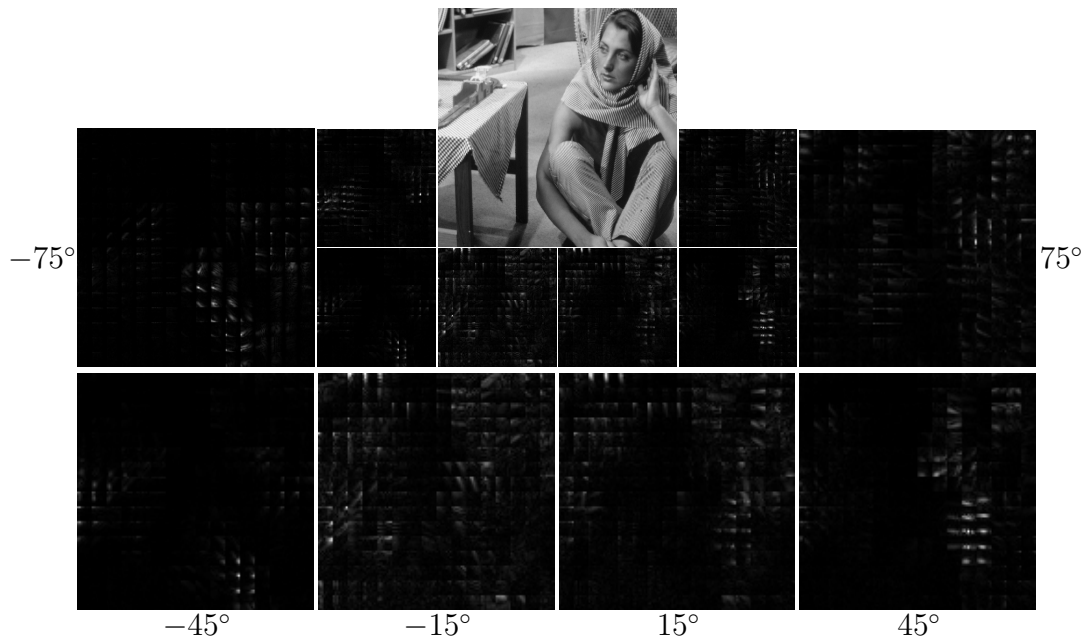
με δύο εναλλακτικούς τρόπους απεικόνισης το μέτρο του 2-Δ MCLT που αντιστοιχεί στην εικόνα Barbara: στο (α) μέρος απεικονίζουμε τους συντελεστές όπως έχουν υπολογιστεί μετά τη διαχωρίσιμη επεξεργασία, ενώ στο (β) τους απεικονίζουμε αναδιατεταγμένους έτσι ώστε οι συντελεστές MCLT να είναι οργανωμένοι σε διαδοχικά επίπεδα ανάλυσης και έξι κατευθυντικές υποζώνες σε ευθεία αναλογία με τους αντίστοιχους DT-CWT συντελεστές του Σχ. 5.6(β).



Σχήμα 5.8: Παραθέτουμε τα MCLT άτομα που αντιστοιχούν σε ανάλυση με μήκος μπλοκ $l = 16$. Για να παράγουμε την εικόνα αυτή ξεκινάμε από συντελεστές MCLT διατεταγμένους όπως φαίνεται στο Σχ. 5.9(α). Σε κάθε μπλοκ μόνο ένας MCLT συντελεστής αφήνεται ενεργός, έχοντας μοναδιαίο μέτρο και κατάλληλα επιλεγμένη φάση ώστε να εξασφαλίζεται η συνέχεια των δακτυλίων, ενώ όλοι οι υπόλοιποι μηδενίζονται. Κατόπιν υπολογίζουμε τον αντίστροφο MCLT για να μεταβούμε στο χωρικό πεδίο, και το αποτέλεσμα είναι αυτό που δείχνεται στο σχήμα μας. Παρατηρούμε ότι τα MCLT άτομα είναι κατευθυντικές ταλαντούμενες συναρτήσεις και συνθέτουν ένα σχεδόν ακτινικά συμμετρικό μωσαϊκό.



(α)



(β)

Σχήμα 5.9: Μέτρο των μιγαδικών συντελεστών του 2-Δ MCLT που αντιστοιχεί στην 512×512 εικόνα Barbara, για ανάλυση σε μπλοκ των 16×16 εικονοστοιχείων. (α) Οι MCLT συντελεστές (σε λογαριθμική κλίμακα) όπως προκύπτουν από τη διαχωρίσιμη ανάλυση· κάθε MCLT μπλοκ έχει διάσταση 16×32 στοιχεία, οπότε η μετασχηματισμένη εικόνα αποτελείται από 512×1024 μιγαδικά στοιχεία. (β) Οι ίδιοι MCLT συντελεστές (σε γραμμική κλίμακα) επαναδιαταγμένοι παρόμοια με τους DT-CWT του Σχ. 5.6(β) σε έξι κατευθυντικές υποζώνες, όπως θα συζητήσουμε περαιτέρω στην Ενότητα 6.2.2, Σχ. 6.2. Το κάθε επίπεδο αντιστοιχεί σε μια φασματική οκτάβα ανάλυσης. Στο ανώτερο επίπεδο δείχνουμε μια κλιμακωμένη εκδοχή της αρχικής εικόνας (αντί για τους συντελεστές κλιμάκωσης στην περίπτωση του DT-CWT).

5.3 Στατιστικές Ιδιότητες Φυσικών Εικόνων στο Πεδίο των Κυματιδίων

Κυματιδιακές αναπαραστάσεις όπως αυτές που έχουμε συζητήσει ως τώρα στο παρόν Κεφάλαιο ενδείκνυνται για την περιγραφή και μοντελοποίηση φυσικών εικόνων. Θα χρησιμοποιήσουμε κυρίως τις υπερπλήρεις μιγαδικές κυματιδιακές αναπαραστάσεις του DT-CWT (ανάλυση χώρου-κλίμακας) και του MCLT (ανάλυση χώρου-συχνότητας), καθώς αυτές διαθέτουν σε μεγάλο βαθμό τις επιθυμητές ιδιότητες συμμετρίας μετατόπισης και περιστροφής, σε αντίθεση με τον κλασικό κρίσιμα δειγματοληπτημένο μετασχηματισμό κυματιδίων [SFAH92].

Ορισμένα από τα επιχειρήματα που συνηγορούν στη χρήση κατάλληλων τοπικών, πολυ-κλιμακωτών, και πολυ-κατευθυντικών γραμμικών μετασχηματισμών όπως της κυματιδιακής αναπαραστάσης για τη μοντελοποίηση εικόνων είναι, μεταξύ άλλων, τα εξής:

- Ο μετασχηματισμός κυματιδίων είναι, σε καλή προσέγγιση, μετασχηματισμός *αποσυσχέτισης* για μια ευρεία κατηγορία στοχαστικών διαδικασιών, περιλαμβανομένης της κλασματικής κίνησης Brown που μελετήσαμε νωρίτερα [Wor93]. Όπως έχουμε ήδη συζητήσει, ένα βασικό πλεονέκτημα του μετασχηματισμού κυματιδίων έναντι καθολικών αναπαραστάσεων όπως ο μετασχηματισμός Fourier είναι ότι μπορεί να αποσυσχετίσει σε ικανοποιητικό βαθμό και μη-στατικές διαδικασίες, χάρη στην τοπικότητά του. Η ιδιότητα αποσυσχέτισης επιτρέπει συχνά την επεξεργασία εικόνων με απλούς και αποδοτικούς αλγορίθμους που χειρίζονται καθένα συντελεστή κυματιδίου ανεξάρτητα από τους υπολοίπους.
- Σειρά ψυχο-φυσιολογικών πειραμάτων σε ανθρώπους και άλλα θηλαστικά έχουν δείξει ότι το πρώτο στάδιο επεξεργασίας σε βιολογικά συστήματα όρασης περιλαμβάνει εξαγωγή χαρακτηριστικών με φίλτρα που η απόκρισή τους προσομοιάζει μιγαδικά φίλτρα Gabor [Dau85]. Αυτό φαίνεται ότι είναι αποτέλεσμα εξελικτικής διαδικασίας και σχετίζεται με τις στατιστικές ιδιότητες φυσικών εικόνων [Fie87].
- Ο μετασχηματισμός κυματιδίων, ειδικότερα υπερ-πλήρεις εκδοχές του, επιτρέπουν την *αραιή* αναπαράσταση σημάτων εικόνων, δηλαδή πιστή ανακατασκευή με σχετικά λίγους μη-μηδενικούς συντελεστές. Τούτο διευκολύνει την κωδικοποίηση, επεξεργασία, και αποκατάσταση σημάτων εικόνων [Fie94]. Αντίστροφα, μαθαίνοντας από δεδομένα φυσικών εικόνων αραιούς κώδικες οδηγούμεστε σε φίλτρα με αποκρίσεις ποιοτικά παρόμοιες με φίλτρα Gabor [OF96, HH00]. Προσαρμόζοντας την αναπαράσταση στα δεδομένα της εικόνας μπορεί μάλιστα να οδηγήσει σε βελτιωμένες επιδόσεις σε σχέση με τους παραμετρικούς κυματιδιακούς μετασχηματισμούς που κυρίως χρησιμοποιούμε στην εργασία μας [AEB06, BCM05, EL99], όμως με σημαντικά αυξημένο κόστος. Ποιοτικά παρόμοια αποτελέσματα προκύπτουν και με εφαρμογή αλγορίθμων *ανάλυσης σε ανεξάρτητες συνιστώσες* (Independent Component Analysis – ICA) σε μεγάλες βάσεις φυσικών εικόνων [BS97, vHvdS98].

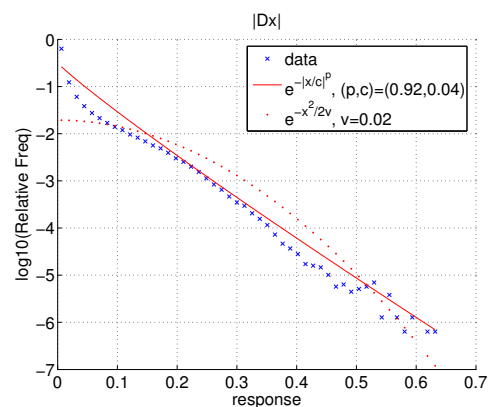
Στη συνέχεια περιγράφουμε ποιοτικά βασικές ιδιότητες των συντελεστών κυματιδίων φυσικών εικόνων που έχουν καθοδηγήσει τη σχεδίαση των μοντέλων ενδοσυμπλήρωσης που παρουσιάζουμε στο Κεφάλαιο 6. Μια και χρησιμοποιούμε κατά κύριο λόγο τις μιγαδικές κυματιδιακές αναπαραστάσεις, συζητάμε ξεχωριστά τη συμπεριφορά του μέτρου και της φάσης των συντελεστών κυματιδίων φυσικών εικόνων.

5.3.1 Ιδιότητες Μέτρου Συντελεστών Κυματιδίων: Κυρτωτικά Περιθώρια Ιστογράμματα και Αλληλεξαρτήσεις Συντελεστών

Η μη-γκαουσιανή κατανομή των αποκρίσεων των συντελεστών κυματιδίων, αφενός, και η αραιή αναπαράσταση φυσικών εικόνων σε βάσεις κυματιδίων, αφετέρου, συνεπάγονται τον *λεπτόκυρτο* χαρακτήρα των *περιθώριων ιστογραμμάτων* (marginal histograms) συντελεστών για κάθε *υποζώνη* (sub-band) του μετασχηματισμού κυματιδίων [Fie87]. Η *κύρτωση* μιας κατανομής ορίζεται από τη σχέση $E\{(X - \mu)^4\}/E\{(X - \mu)^2\}^2$. Η κατανομή Gauss έχει κύρτωση 3, ενώ ως λεπτόκυρτες (αντ. πλατύκυρτες) χαρακτηρίζονται οι κατανομές με κύρτωση μεγαλύτερη (αντ. μικρότερη) του 3. Οι λεπτόκυρτες κατανομές έχουν πιο ισχυρή κορυφή στο 0 και ταυτόχρονα πιο *έντονες ουρές* (heavy-tails) σε σχέση με κατανομές Gauss ίδιας μεταβλητότητας. Για τις εικόνες η ισχυρή κορυφή των ιστογραμμάτων στο 0 οφείλεται στο γεγονός ότι η συντριπτική πλειοψηφία των συντελεστών κυματιδίων είναι σχεδόν μηδέν (οι ομαλές περιοχές καταλαμβάνουν τη μεγαλύτερη έκταση στις φυσικές εικόνες): οι έντονες ουρές αντιστοιχούν σε κυματίδια με έντονη απόκριση, τοποθετημένα κοντά σε ακμές (οι ακμές καταλαμβάνουν σχετικά μικρή έκταση στις εικόνες). Ένα παράδειγμα περιθώριου ιστογράμματος των συντελεστών κυματιδίων Haar (λεπτομερής κλίμακα, x - κατεύθυνση) στο οποίο φαίνεται ο κυρτωτικός του χαρακτήρας δίνεται στο Σχ. 5.10.



εικόνα I



ιστόγραμμα I_x

Σχήμα 5.10: *Αριστερά*: Παράδειγμα εικόνας από τη βάση του van Hateren [vHvdS98]. *Δεξιά*: Λογαριθμικό ιστόγραμμα μερικής παραγώγου ως προς x (συντελεστές Haar) $\partial I/\partial x$, μαζί με ταιριασμένες γκαουσιανή και γενικευμένη γκαουσιανή κατανομές.

Υπάρχουν διάφορες προσεγγίσεις για μαθηματική μοντελοποίηση των λεπτόκυρτων ιστογραμμάτων των συντελεστών κυματιδίων. Μια επιλογή είναι το ταιρίασμα των ιστογραμμάτων με γενικευμένες κατανομές Gauss, που έχουν τη μορφή $p(x) \propto e^{-(x/c)^p}$ [Mal89], που γενικεύουν την κλασική κατανομή Gauss ($p = 2$). Οι κατανομές αυτής της οικογένειας είναι λεπτόκυρτες για $p < 2$, με χαρακτηριστικότερο παράδειγμα την κατανομή Laplace ($p = 1$). Άλλες παραμετρικές κατανομές που ταιριάζουν ικανοποιητικά στα εμπειρικά ιστογράμματα συντελεστών κυματιδίων είναι η t -κατανομή του Student [HM99], το μοντέλο Bessel K-forms [GS01] και άλλα παρεμφερή μοντέλα, συχνά εμπνευσμένα από αντίστοιχες προσεγγίσεις από την περιοχή της εύρωσης στατιστικής.

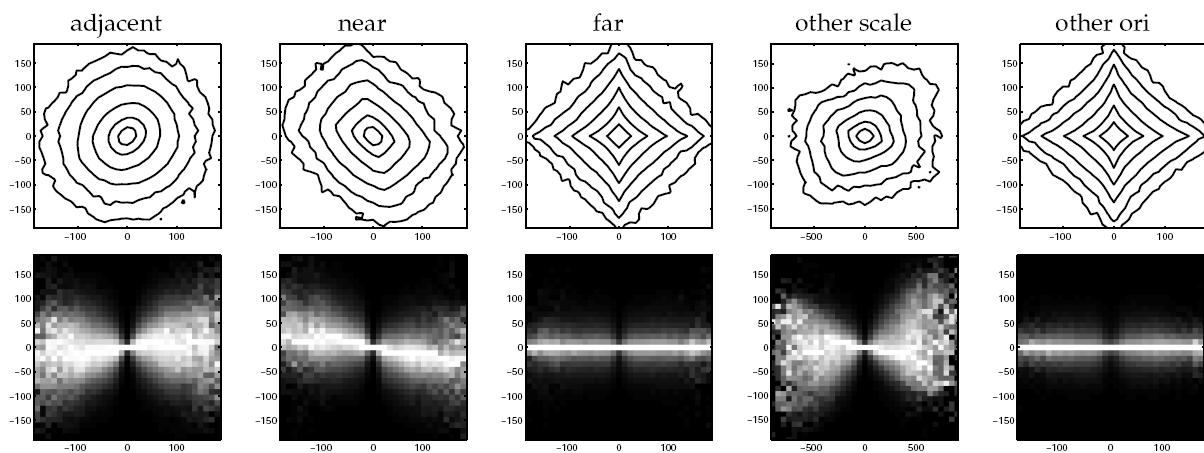
Μια εναλλακτική προσέγγιση στη μοντελοποίηση των κυρτωτικών ιστογραμμάτων φυσικών εικόνων είναι μέσω (πεπερασμένου ή άπειρου) μείγματος γκαουσιανών κατανομών. Χρησιμοποιώντας πεπερασμένο μείγμα γκαουσιανών μπορούμε να προσεγγίσουμε ικανοποιητικά τις κατανομές που μας ενδιαφέρουν. Ακόμα και με μείγμα δύο μόνο γκαουσιανών (μία μικρής μεταβλητότητας για μοντελοποίηση της ισχυρής κορυφής και μία μεγάλης μεταβλητότητας για μοντελοποίηση των έντονων ουρών του ιστογράμματος) μπορούμε να μοντελοποιήσουμε επαρκώς για κάποιες εφαρμογές ιστογράμματα φυσικών εικόνων [CNB98]. Η προσέγγιση των ιστογραμμάτων με άπειρο μείγμα γκαουσιανών [WSW01] είναι μια άλλη εναλλακτική μέθοδος, που όμως δεν θα συζητήσουμε περαιτέρω εδώ.

Πέρα από τα περιθώρια ιστογράμματα, ο μη-γκαουσιανός χαρακτήρας της κατανομής συντελεστών κυματιδίων εμφανίζεται έντονα και στα από κοινού ιστογράμματα [PSWS03] και τις υπό συνθήκη κατανομές. Αν και ο μετασχηματισμός κυματιδίων εξαλείφει τις συσχετίσεις δεύτερης τάξης, οι συντελεστές κυματιδίων που αντιστοιχούν σε φίλτρα γειτονικά μεταξύ τους (είτε χωρικά, είτε στην κλίμακα, είτε στην κατεύθυνση) δεν είναι ανεξάρτητοι. Ειδικότερα, η υπό συνθήκη μεταβλητότητα κάθε συντελεστή εξαρτάται από τις τιμές των γειτόνων, με αποτέλεσμα οι υπό-συνθήκη κατανομές να έχουν τη μορφή *πεταλιούδας* (bow-tie), όπως φαίνεται στο Σχ. 5.11. Αυτό σημαίνει ότι οι ενεργοί συντελεστές κυματιδίων σχηματίζουν *συστάδες* (clusters) και εντοπίζονται κυρίως κοντά στις ακμές της εικόνας, με την ύπαρξη ενός ενεργού συντελεστή να αποτελεί σημαντική ένδειξη για την παρουσία και άλλων ενεργών συντελεστών στην ίδια γειτονιά, χωρική ή κλίμακας. Μια άλλη ένδειξη της εξάρτησης μεταξύ αποκρίσεων γειτονικών φίλτρων είναι η θετική αμοιβαία τους πληροφορία [LM01]. Οι εξαρτήσεις αναβιώνουν και ως συσχετίσεις δεύτερης τάξης αν περάσουμε τους συντελεστές κυματιδίων από ένα απλό μη-γραμμικό σύστημα, για παράδειγμα ένα σύστημα που ως έξοδο δίνει την απόλυτη τιμή της εισόδου.

5.3.2 Ιδιότητες Φάσης Συντελεστών Κυματιδίων

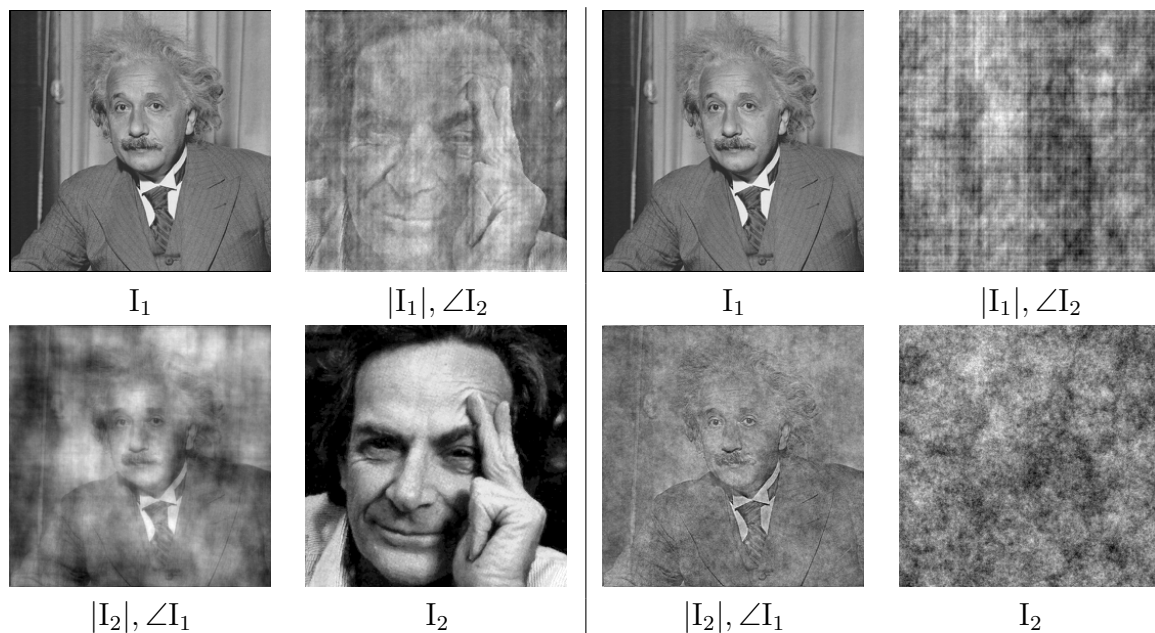
Ως τώρα η συζήτηση σχετικά με τις στατιστικές ιδιότητες εικόνων του φυσικού κόσμου έχει περιοριστεί στην απόκριση μέτρου των συντελεστών μιγαδικών κυματιδίων. Παρόλα αυτά, η απόκριση φάσης των συντελεστών μιγαδικών κυματιδίων παίζει εξίσου σημαντικό, αν όχι μεγαλύτερο, ρόλο στην ανθρώπινη οπτική αντίληψη.

Για την περίπτωση ανάλυσης εικόνων με το μετασχηματισμό Fourier (ΜΦ), η σημασία της Fourier μιγαδικής φάσης έχει δείχτει με εντυπωσιακό τρόπο στο κλασικό πείραμα που περιγράφεται στο [OL81] και παράλληλα στο [PC82]. Σε αυτό το πείραμα παίρνουμε το ΜΦ δύο εικόνων, I_1 και I_2 , και συνθέτουμε καινούργιες εικόνες με αντίστροφο ΜΦ, χρησιμοποιώντας το μέτρο Fourier από τη μια εικόνα και τη φάση Fourier από την άλλη, και κρατώντας το πραγματικό μέρος του αποτελέσματος. Τα αποτελέσματα αυτού του πειράματος φαίνονται στο Σχ. 5.12. Η συνθετική εικόνα αντιληπτικά μοιάζει εντυπωσιακά με την εικόνα που μοιράζονται την ίδια απόκριση φάσης. Στο Σχ. 5.13 επαναλαμβάνουμε το πείραμα, χρησιμοποιώντας όμως μιγαδικά κυματίδια στη θέση του ΜΦ. Η ανάμειξη μέτρου και φάσης του μιγαδικού μετασχηματισμού κυματιδίων έχει τώρα λιγότερο εντυπωσιακά αποτελέσματα, κυρίως επειδή ο μετασχηματισμός κυματιδίων έχει αρκετά καλές ιδιότητες εντοπισμού, οπότε η ανταλλαγή της φάσης δεν αλλοιώνει τελείως τη γεωμετρική θέση των αντικειμένων. Παρόλα αυτά, και στην περίπτωση των μιγαδικών κυματιδίων η αντιληπτική σημασία της φάσης παραμένει σημαντική.



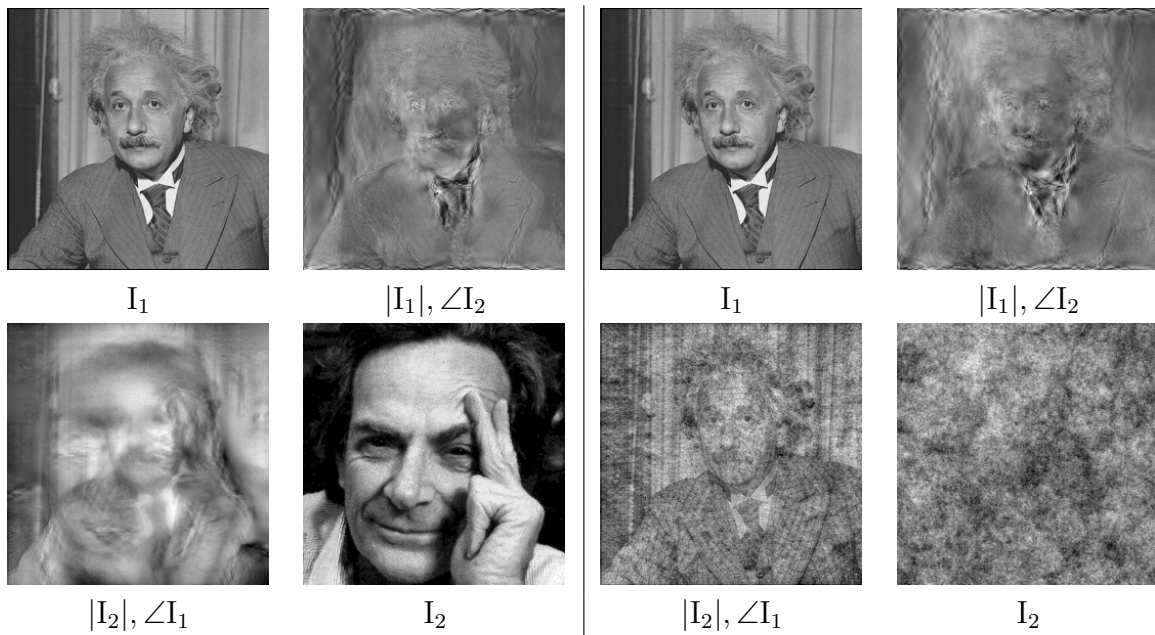
Σχήμα 5.11: Ιστογράμματα συντελεστών κυματιδίων για διάφορα ζευγάρια φίλτρων εφαρμοσμένα σε μία εικόνα. Οι τρεις πρώτες στήλες αντιστοιχούν σε φίλτρα ίδιας κλίμακας και κατεύθυνσης αλλά χωρικά μετατοπισμένα. Η τέταρτη αφορά ίδια θέση και κατεύθυνση αλλά διαφορετική κλίμακα. Η πέμπτη αφορά ίδια θέση και κλίμακα αλλά διαφορετική κατεύθυνση. *Πάνω γραμμή:* Το κοινό ιστόγραμμα ως contour log-plot. Θα αναμέναμε σχήματα ελλείψεων στη γκαουσιανή περίπτωση. *Κάτω γραμμή:* Αντίστοιχα υπό-συνθήκη ιστογράμματα· για ανεξάρτητες μεταβλητές θα αναμέναμε τα ιστογράμματα να είναι ορθογώνια αντί να έχουν σχήμα πεταλούδας. Παρατηρούμε ιδιαίτερα έντονες εξαρτήσεις για φίλτρα με μικρή χωρική μετατόπιση (πρώτη στήλη) και για φίλτρα σε διαδοχικές κλίμακες (τέταρτη στήλη). Από το [Sim05].

Τούτο εξηγείται από το γεγονός ότι η ανθρώπινη οπτική αντίληψη βασίζεται σε σημαντικό βαθμό στην ανίχνευση *πρωτευόντων* (primitive) χαρακτηριστικών, όπως είναι οι ακμές, ράβδοι, και κηλίδες. Όπως έχει αναδείξει το μοντέλο ανίχνευσης χαρακτηριστικών μέσω *σύμπτωσης φάσης* (phase congruency) [MB88, Κον99], η ύπαρξη, η ακριβής θέση, και άλλες ιδιότητες των ακμών (και άλλων χαρακτηριστικών) στην εικόνα καθορίζονται κατά κύριο λόγο από την απόκριση φάσης. Δεν είναι τυχαίο ότι στο επιτυχημένο σύστημα αναγνώρισης ανθρώπων με βάση φωτογραφία της ίριδας του ματιού τους, που περιγράφεται στο [Dau04], γίνεται εξαγωγή χαρακτηριστικών με βάση συστοιχία φίλτρων Gabor, από τα οποία διατηρείται μόνο πληροφορία φάσης, και μάλιστα σε αδρά κβαντισμένη μορφή (2 bits ανά φίλτρο)·δες επίσης το [VHHLO83].



Σχήμα 5.12: Η σημασία της Fourier φάσης στην αντίληψη εικόνων. Ανταλλάσσουμε τις αποκρίσεις μέτρου και φάσης σε δύο ζευγάρια εικόνων. *Αριστερά:* Αινστάιν (I_1) και Φάινμαν (I_2). *Δεξιά:* Αινστάιν (I_1) και fBm με $D = 2.5$ (I_2). Όσο αφορά την ανθρώπινη αντίληψη, η Fourier φάση κυριαρχεί του Fourier μέτρου!

Παρόλη τη σημασία της φάσης για την ανθρώπινη αντίληψη εικόνων, η μοντελοποίηση πληροφορίας φάσης είναι αρκετά πιο δύσκολη από τη μοντελοποίηση μέτρου. Επίσης η επεξεργασία της φάσης επηρεάζει το σήμα με μη-γραμμικό τρόπο. Σε προκαταρκτικά πειράματα, η προσέγγιση που ακολουθούμε βασίζεται στις ακόλουθες ιδιότητες *συνοχής* (coherence) του μιγαδικού μετασχηματισμού κυματιδίων με φίλτρα γραμμικής φάσης: (1) η τοπική φάση μεταβάλλεται γραμμικά με την απόσταση από μεμονωμένα γεωμετρικά χαρακτηριστικά (π.χ. γραμμές ή ακμές) και (2) ο αντίστοιχος ρυθμός μεταβολής της φάσης για φίλτρα σε μια συγκεκριμένη κλίμακα είναι διπλάσιος από τον ρυθμό μεταβολής για φίλτρα στην αμέσως πιο αδρή κλίμακα (για ζωνοπερατά φίλτρα με κεντρικές συχνότητες που διαφέρουν μια οκτάβα), υπό την προϋπόθεση ότι το γεωμετρικό χαρακτηριστικό είναι αναλλοίωτο στις δύο κλίμακες. Αυτές οι σχέσεις συνοχής διαταράσσονται από παρεμβολές μεταξύ πολύ κοντινών γεωμετρικών χαρακτηριστικών, όπως συμβαίνει σε εικόνες με έντονη μικροδομή υφής, ή όταν οι ακμές της εικόνας είναι θολωμένες, οπότε



Σχήμα 5.13: Η σημασία της κυματιδικής φάσης στην αντίληψη εικόνων. Παρόμοια με το Σχ. 5.12, ανταλλάσσουμε τις αποκρίσεις μέτρου και φάσης στα ίδια ζευγάρια εικό-νων, αυτή τη φορά για το μιγαδικό μετασχηματισμό κυματιδίων. Η φάση είναι κι εδώ καθοριστική της αντίληψης. Βέβαια, λόγω των καλών ιδιοτήτων εντοπισμού του μετασχη-ματισμού κυματιδίων, η ανταλλαγή της φάσης δεν αλλοιώνει τελείως τη γεωμετρική θέση των αντικειμένων.

δεν έχουν αναλλοίωτο προφίλ σε διαδοχικές κλίμακες. Αυτές οι ιδιότητες του μετασχηματισμού κυματιδίων έχουν χρησιμοποιηθεί στο [PS00b] σε πειράματα σύνθεσης υψής, στα [RWCB03, AKF05, MK08b, MK08a] για αποθορυβοποίηση, ανάλυση και εξαγωγή γεωμετρικών χαρακτηριστικών εικόνων, ενώ η εργασία [WS03] δείχνει πώς μπορεί να εκτιμηθεί η έκταση του θολώματος μιας εικόνας από το βαθμό στον οποίο ισχύουν οι παραπάνω σχέσεις. Πάντως η επιτυχής μοντελοποίηση πληροφορίας φάσης παραμένει σημαντικό ζήτημα για περαιτέρω μελλοντική διερεύνηση.

Σε αυτό το σημείο αξίζει να σημειώσουμε ότι το πρόβλημα της ανακατασκευής 1-Δ σήματος ή 2-Δ εικόνας με χρήση μόνο πληροφορίας μέτρου ή φάσης έχει μελετηθεί από πολλούς ερευνητές. Οι πρώτες σχετικές εργασίες εξέταζαν το πρόβλημα για τον καθολικό Fourier μετασχηματισμό. Συνθήκες που εξασφαλίζουν τέλεια ανακατασκευή και σχετικοί αλγόριθμοι παρουσιάζονται στα [HLO80, Hay82, VHHLO83]. Το σχετιζόμενο πρόβλημα ανακατασκευής από μερική πληροφορία μέτρου ή φάσης τοπικών Gabor μετασχηματισμών έχει επίσης μελετηθεί [BPZ92, UPC98], ενώ το [HO08, HO07] επικεντρώνεται στην ειδική περίπτωση του DT-CWT. Οι περισσότερες τεχνικές ανακατασκευής βασίζονται σε επαναληπτικούς αλγόριθμους, με τυπικό παράδειγμα τον αλγόριθμο εναλλασσόμενων *προβολών σε κυρτά σύνολα* (projection onto convex sets - POCS) [YW82], όπου κανείς μεταβαίνει διαρκώς μεταξύ του χωρικού και του μετασχηματισμένου πεδίου επιβάλλοντας σε κάθε βήμα τους διαθέσιμους περιορισμούς ωσότου ο αλγόριθμος συγκλίνει σε ένα πραγματικό σήμα/εικόνα με τις προδιαγεγραμμένες τιμές για το μέτρο ή τη φάση του στο μετασχηματισμένο πεδίο. Είναι χρήσιμο να αντιδιαστείλουμε αυτήν τη διαδικασία με τα πειράματα που περιγράψαμε νωρίτερα που απλά περιλαμβάνουν ανάμειξη μέτρου και φάσης προερχόμενων από δύο διαφορετικά σήματα, ακολουθούμενη από μία μετάβαση στο χωρικό πεδίο με εφαρμογή του αντίστοιχου αντίστροφου μετασχηματισμού.

Κεφάλαιο 6

Ενδοσυμπλήρωση με Κυματίδια Οργανωμένα σε Δενδρικό Στατιστικό Μοντέλο

6.1 Η Ενδοσυμπλήρωση ως Πρόβλημα Στατιστικής Συμπερασματολογίας

Συχνά θεωρεί κανείς ότι οι 2-Δ εικόνες που σχηματίζονται στο ανθρώπινο σύστημα όρασης ή στους αισθητήρες φωτογραφικών μηχανών είναι αυθαίρετες συναρτήσεις ορισμένες στο 2-Δ επίπεδο. Στην πραγματικότητα όμως, λόγω του μηχανισμού σχηματισμού τους, οι εικόνες παρουσιάζουν πολύ πιο συγκεκριμένη δομή, την οποία μπορούμε να εκμεταλλευτούμε κατά την ανάλυση και επεξεργασία τους. Ανάλογα με το είδος του φορμαλισμού που υιοθετεί κανείς, μπορεί να ακολουθήσει διαφορετικές προσεγγίσεις για να χαρακτηρίσει τους μαθηματικούς χώρους στους οποίους κινούνται οι φυσικές εικόνες. Για παράδειγμα, η προσέγγιση του Κεφαλαίου 3 με μεταβολικά μοντέλα και μερικές διαφορικές εξισώσεις επικεντρώνεται στις γεωμετρικές ιδιότητες της εικόνας, ενώ όσοι χρησιμοποιούν εργαλεία συναρτησιακής ανάλυσης συνηθίζουν να μοντελοποιούν τις φυσικές εικόνες ως μέλη αφηρημένων συναρτησιακών χώρων [Mal99a].

Σε αυτό το Κεφάλαιο υιοθετούμε μια θεώρηση του προβλήματος της ενδοσυμπλήρωσης που πηγάζει από τη στατιστική προσέγγιση στην ερμηνεία εικόνων, π.χ. [KR96]. Αυτή απαιτεί την ανάπτυξη στατιστικών μοντέλων για εικόνες του φυσικού κόσμου, αν είναι δυνατόν εξειδικεύσιμων για εικόνες συγκεκριμένης κατηγορίας, π.χ. μοντέλο τοιχογραφιών Θήρας για τις ανάγκες της εφαρμογής του Κεφαλαίου 2. Αν έχουμε διαθέσιμο ένα αξιόπιστο τέτοιο *πρότερο* πιθανοτικό μοντέλο για φυσικές εικόνες, έστω $P(\mathbf{x}|\boldsymbol{\theta})$, με $\boldsymbol{\theta}$ σύνολο παραμέτρων που επιτρέπουν την προσαρμογή του μοντέλου σε ορισμένη κατηγορία εικόνων, και ένα μοντέλο παρατήρησης $P(\mathbf{x}_o|\mathbf{x})$ που να σχετίζει την παρατηρούμενη αλλοιωμένη εικόνα \mathbf{x}_o και την πραγματική εικόνα \mathbf{x} , τότε, εφαρμόζοντας το θεώρημα του Bayes, μπορούμε, τουλάχιστον καταρχήν, να υπολογίσουμε την *ύστερη* (ή *υπό συνθήκη*) κατανομή για την πραγματική εικόνα, που περιέχει όλη την πληροφορία που μπορούμε δυνητικά να συνάγουμε από τα δεδομένα μας:

$$P(\mathbf{x}|\mathbf{x}_o, \boldsymbol{\theta}) = \frac{P(\mathbf{x}|\boldsymbol{\theta})P(\mathbf{x}_o|\mathbf{x})}{P(\mathbf{x}_o)}. \quad (6.1)$$

Είναι λοιπόν σαφές ότι η ανάπτυξη πιστών πρότερων πιθανοτικών μοντέλων $P(\mathbf{x})$ για φυσικές εικόνες παίζει κεντρικό ρόλο στη στατιστική αντιμετώπιση του προβλήματος της ενδοσυμπλήρωσης. Για την πρακτική χρησιμότητα αυτής της προσέγγισης είναι επιθυμητό το πιθανοτικό μοντέλο που θα υιοθετήσουμε να έχει τέτοια δομή που οι ακόλουθες λειτουργίες να είναι υπολογιστικά εφικτές:

- *Συμπερασματολογία.* Θα πρέπει το μοντέλο να επιτρέπει υπολογιστικά τη συναγωγή συμπερασμάτων για την πραγματική εικόνα \mathbf{x} με τον κανόνα του Bayes (6.1).
- *Μάθηση.* Με χρήση ενός συνόλου εικόνων μάθησης, θα πρέπει να μπορούμε να ρυθμίζουμε τις παραμέτρους θ , π.χ. μεγιστοποιώντας την πιθανοφάνεια $L(\theta) = P(\mathbf{x}|\theta)$, ώστε να πετυχαίνουμε πιστότερη προσαρμογή του μοντέλου στα δεδομένα.
- *Δειγματοληψία/Σύνθεση.* Η ευχέρεια στη λήψη τυχαίων δειγμάτων από την πρότερη $P(\mathbf{x}|\theta)$ ή την ύστερη $P(\mathbf{x}|\mathbf{x}_o, \theta)$ κατανομή φυσικών εικόνων είναι επιθυμητή γιατί επιτρέπει την άμεση οπτική αξιολόγηση των παραδοχών που έχει κάνει κανείς κατά την ανάπτυξη του μοντέλου. Ειδικά για την εφαρμογή μας, κάθε δείγμα από την υπό συνθήκη κατανομή $P(\mathbf{x}|\mathbf{x}_o, \theta)$ είναι μια πιθανή λύση του προβλήματος της ενδοσυμπλήρωσης, οπότε η δυνατότητα δειγματοληψίας από το μοντέλο είναι ιδιαίτερα σημαντική.

Η ικανοποίηση όλων αυτών των προδιαγραφών από ένα στατιστικό μοντέλο εικόνας, το οποίο ταυτόχρονα πρέπει να είναι αρκετά πλούσιο ώστε να συλλαμβάνει τα βασικά χαρακτηριστικά των φυσικών εικόνων, δεν είναι εύκολη υπόθεση.

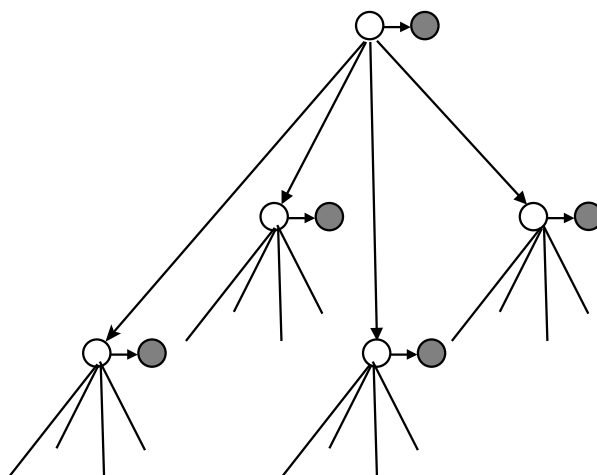
Σε αυτό το σημείο και προτού προχωρήσουμε στην ανάλυση συγκεκριμένων στοχαστικών μοντέλων εικόνας, είναι χρήσιμο να οριοθετήσουμε τις προσδοκίες που έχουμε από αυτά [MG01]. Σε αυτήν την εργασία θα επικεντρωθούμε σε μοντέλα *χαμηλού* και *μέσου* επιπέδου όρασης. Πιο συγκεκριμένα, τα μοντέλα μας πρέπει να μπορούν να περιγράψουν και να αναπαράγουν βασικές ιδιότητες χαμηλού επιπέδου των εικόνων του φυσικού κόσμου, αλλά δεν απαιτούμε από αυτά να μοντελοποιούν χαρακτηριστικά υψηλού επιπέδου, όπως την εμφάνιση ολόκληρων αντικειμένων. Με άλλα λόγια, δείγματα εξαγμένα από τα μοντέλα μας θα πρέπει να δίνουν σε μια πρώτη ματιά την *αίσθηση* εικόνων του φυσικού κόσμου, χωρίς όμως να περιμένουμε να δούμε σε αυτά συγκεκριμένα αντικείμενα υψηλού επιπέδου, για παράδειγμα πρόσωπα ή σπίτια. Η δημιουργία ειδικευμένων μοντέλων για περιγραφή αντικειμένων υψηλού επιπέδου βρίσκεται έξω από το βεληνεκές των μοντέλων που θα συζητήσουμε στη συνέχεια.

6.2 Μοντελοποίηση Συντελεστών Μιγαδικών Κυματιδίων Εικόνων με Κρυφό Δένδρο Markov

Συνοψίζοντας τη συζήτηση του Κεφαλαίου 5 σχετικά με τη στατιστική συμπεριφορά των συντελεστών κυματιδίων φυσικών εικόνων, αναδεικνύονται οι ακόλουθες βασικές ιδιότητες:

1. Τα περιθώρια ιστογράμματα έχουν λεπτόκυρτο προφίλ, με ισχυρή κορυφή στο 0 (μικρή απόκριση στις ομαλές περιοχές της εικόνας) και έντονες ουρές (μεγάλη απόκριση κοντά στις ακμές ή σε περιοχές με υφή).

2. Οι ισχυροί συντελεστές κυματιδίων εμφανίζονται κυρίως σε συστάδες στο χώρο ή στην κλίμακα, και ενεργοποιούνται από την παρουσία ακμών στην περιοχή.
3. Η παρουσία γεωμετρικών χαρακτηριστικών δημιουργεί ιδιαίτερη συμφωνία στην απόκριση φάσης μιγαδικών κυματιδίων.



Σχήμα 6.1: Η δομή του κρυφού δένδρου Markov (βαθμού 4) στη μοντελοποίηση συντελεστών κυματιδίων εικόνων. Οι λευκοί κόμβοι συμβολίζουν τις διακριτές κρυφές μεταβλητές q_t (π.χ. ‘ακμή’/‘μη-ακμή’), οι προσκείμενοι γκριζοί κόμβοι συμβολίζουν τις συνεχείς παρατηρήσεις y_t (μιγαδικοί συντελεστές κυματιδίων), ενώ τα βέλη δηλώνουν πιθανοτικές εξαρτήσεις.

6.2.1 Το HMT Μοντέλο για Περιγραφή 1-Δ και 2-Δ Πραγματικών Κυματιδίων

Μια δημοφιλής τεχνική που αναπαράγει σε σημαντικό βαθμό τις δύο πρώτες από τις παραπάνω ιδιότητες έχει προταθεί στο [CNB98] και βασίζεται σε μοντελοποίηση των συντελεστών κυματιδιακής λεπτομέρειας με *κρυφά δέντρα Markov* (hidden Markov tree - HMT), με διακριτές κρυφές μεταβλητές και συνεχείς παρατηρήσεις μοντελοποιημένες με μία γκαουσιανή κατανομή ανά κατάσταση. Στην πιο απλή περίπτωση έχουμε δύο μόνο κρυφές καταστάσεις, που αντιστοιχούν σε ‘ακμή’/‘μη-ακμή’. Η γκαουσιανή κατανομή που αντιστοιχεί σε ακμή θα έχει μεγάλη μεταβλητότητα (έντονες ουρές), ενώ αυτή που αντιστοιχεί σε μη-ακμή θα έχει μικρή μεταβλητότητα (ισχυρή κορυφή στο 0): έτσι αναπαράγουμε ποιοτικά το λεπτόκυρτο χαρακτήρα των περιθώριων ιστογραμμάτων. Επίσης, οι πίνακες μετάβασης στην μαρκοβιανή αλυσίδα ευνοούν ομοειδείς μεταβάσεις (ακμή→ακμή, μη-ακμή→μη-ακμή) έναντι των μη-ομοειδών (ακμή→μη-ακμή, μη-ακμή→ακμή). Συνήθως επιτρέπουμε μεγαλύτερο πλήθος κρυφών καταστάσεων (π.χ. τέσσερις) για να αυξήσουμε την εκφραστικότητα του μοντέλου πέρα από τον απλοϊκό διαχωρισμό ‘ακμή’/‘μη-ακμή’. Εάν μοντελοποιήσουμε τους συντελεστές κυματιδίων ως μείγμα γκαουσιανών κατανομών δίχως εξαρτήσεις μεταξύ των κόμβων του δέντρου, προκύπτει το *κρυφό*

ανεξάρτητο δέντρο (hidden independent tree – HIT). Τα HMT και HIT μπορεί να θεωρηθούν ως δενδρικές γενικεύσεις των ακολουθιακών κρυφών μοντέλων Markov (hidden Markov model – HMM) και μοντέλων γκαουσιανών μειγμάτων (Gaussian mixture model – GMM). Σημειώνουμε ότι τα HMT και HIT μοντέλα δεν περιγράφουν τους συντελεστές κλιμάκωσης του κυματιδιακού μετασχηματισμού, που συνεπώς πρέπει κανείς να χειριστεί ξεχωριστά.

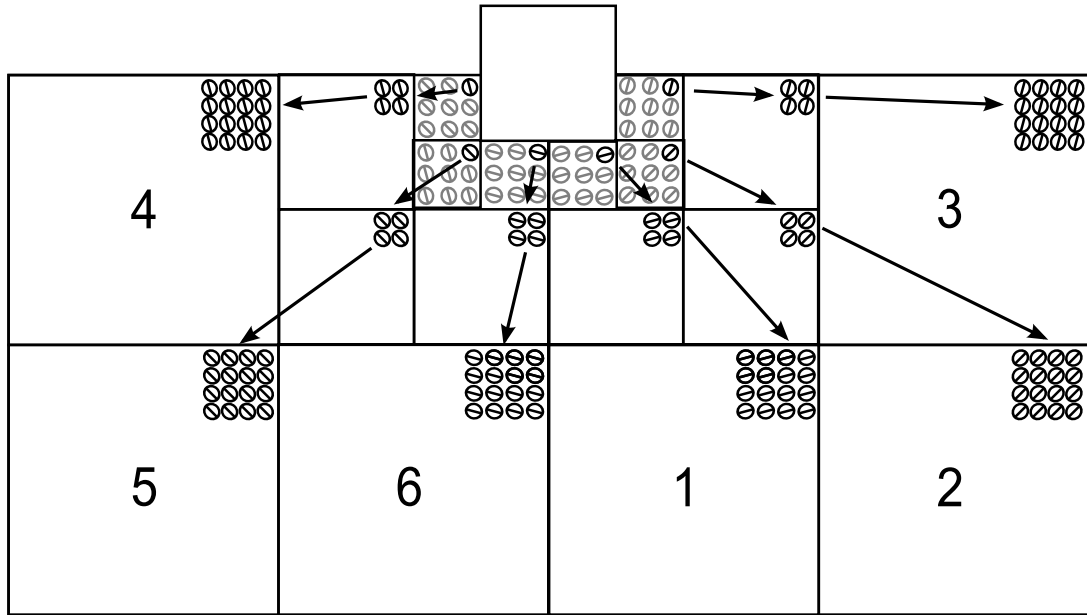
Το αρχικό HMT μοντέλο του [CNB98], σχεδιασμένο για 1-Δ σήματα και πραγματικό μετασχηματισμό κυματιδίων, έχει επεκταθεί για 2-Δ σήματα εικόνων στο [RCB01] με τον πιθανοτικό γράφο του HMT να έχει μορφή δένδρου βαθμού 4 (κάθε κόμβος έχει 4 παιδιά), όπως φαίνεται στο Σχ. 6.1 όπου χρησιμοποιούμε συμβολισμό κατευθυνόμενων γραφικών μοντέλων (graphical models) [Bis06] για να υποδηλώσουμε τις πιθανοτικές εξαρτήσεις. Ανάλογες δενδρικές μορφές έχουν χρησιμοποιηθεί με σημαντική επιτυχία για διευθυνσιοδότηση των ενεργών κυματιδιακών συντελεστών στο πρόβλημα της κωδικοποίησης εικόνων [Sha93, SP96]. Στην περίπτωση του πραγματικού DWT, προτείνεται στο [RCB01] η χρήση ενός ανεξάρτητου δέντρου μαρκοβιανών εξαρτήσεων για καθεμιά από τις τρεις κατευθυντικές υποζώνες (subbands). Εναλλακτικά, μπορεί κανείς να θεωρήσει το σύνολο των τριών κυματιδιακών συντελεστών ως ενιαίο διάλυμα παρατήρησης [DV02], οπότε έχουμε ένα μόνο δέντρο μαρκοβιανών εξαρτήσεων και για τις τρεις κυματιδιακές υποζώνες. Τέλος μπορεί κανείς να μοντελοποιήσει τις εξαρτήσεις μεταξύ των κρυφών μεταβλητών σε διαφορετικές υποζώνες, όπως προτείνεται στα [FX03, SPS06].

Αξίζει να αναφέρουμε ότι, εκτός του HMT, μοντέλα δενδρικής μορφής σε διάφορες παραλλαγές έχουν διερευνηθεί διεξοδικά στη βιβλιογραφία, τόσο επειδή η δομή τους προσφέρεται για την περιγραφή πολυκλιμακωτών φαινομένων, όσο και επειδή είναι επιδεκτικά ανάλυσης με ταχείς αλγόριθμους. Ενδεικτικά αναφέρουμε τις εργασίες [BS94] και [LPH00, PCL00] που αναπτύσσουν πολυκλιμακωτά δενδρικά μοντέλα ως εναλλακτικά των τυχαίων πεδίων Markov (Markov random fields – MRF) για το πρόβλημα της κατάτμησης εικόνων. Σε αυτές τις τεχνικές η κρυφή διακριτή κατάσταση αντιστοιχεί στην ετικέτα κατάτμησης για μια περιοχή και η συνεχής παρατήρηση αντιστοιχεί στις τιμές έντασης της εικόνας (είτε θεωρείται ότι έχουμε παρατηρήσεις μόνο στο πιο λεπτομερές επίπεδο, είτε θεωρείται ότι κατάλληλα στιγμιότυπα μιας Gaussian πυραμίδας αποτελούν ανεξάρτητες παρατηρήσεις στα αδρότερα επίπεδα του δέντρου), δηλαδή στα μοντέλα αυτά δεν γίνεται χρήση κυματιδιακών αναπαραστάσεων. Συχνά η δενδρική δομή αντιστοιχεί σε πολυκλιμακωτή περιγραφή περιοχών της εικόνας, με τη σχέση γονιού-παιδιού να δηλώνει υποδιαίρεση περιοχής [SG00]. Σε άλλες εργασίες, π.χ. [CWB94, WSW01], οι κρυφές μεταβλητές είναι κι αυτές συνεχείς, οπότε τα μοντέλα που προκύπτουν αποτελούν δενδρικές γενικεύσεις του δυναμικού μοντέλου Kalman. Σε πρόσφατη εργασία μας έχουμε αναπτύξει μια παραλλαγή του HMT μοντέλου ειδικά προσαρμοσμένη για αποθορυβοποίηση εικόνων αλλοιωμένων από Poisson θόρυβο [LMP09]. Για μια διεξοδική ανασκόπηση της βιβλιογραφίας σε πολυκλιμακωτές τεχνικές για ανάλυση σημάτων και εικόνας παραπέμπουμε στο [Wil02].

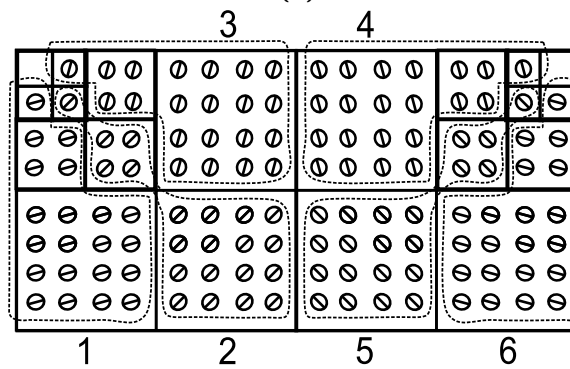
6.2.2 Προσαρμογή του HMT Μοντέλου σε 2-Δ Μιγαδικά Κυματίδια: Δενδρικές Εξαρτήσεις στις Αναπαραστάσεις DT-CWT και MCLT

Η προσαρμογή του HMT μοντέλου στο 2-Δ μιγαδικό DT-CWT με τις έξι κατευθυντικές υποζώνες συνήθως γίνεται με χρήση έξι ανεξάρτητων δέντρων [CRBK00]. Η προκύπτουσα δομή φαίνεται στο Σχ. 6.2(α), όπου κάθε μιγαδικός συντελεστής λεπτομέρειας συμβολίζεται με ένα μικρό κύκλο του οποίου η χορδή υποδηλώνει τη διεύθυνση μέγιστης απόκρισης. Κόμβοι που έχουν συγγένεια γονέα-παιδιού αντιστοιχούν σε συγκατευθυντικά κυματίδια με λόγο κλίμακας 2:1 και χωρική αλληλεπικάλυψη, με το κυματίδιο που αντιστοιχεί στο γονέα να υπερκαλύπτει τα κυματίδια που αντιστοιχούν στα τέσσερα παιδιά. Η παρουσία ακμής στην περιοχή αλληλεπικάλυψης των κυματιδίων προκαλεί ισχυρές κυματιδιακές αποκρίσεις σε γονέα και παιδί, φαινόμενο που εμπνέει την εισαγωγή βέλους Markov εξάρτησης μεταξύ τους. Παράπλευροι κόμβοι αντιστοιχούν σε συγκατευθυντικά κυματίδια της ίδιας κλίμακας που είναι μεταξύ τους χωρικά μετατοπισμένα. Αν και το δέντρο του Σχ. 6.1 δεν περιλαμβάνει απευθείας συνδέσεις μεταξύ παρακείμενων κυματιδίων, οι μεταξύ τους αλληλεξαρτήσεις συλλαμβάνονται έμμεσα λόγω της κοινής καταγωγής τους από προγονικό κόμβο υψηλότερα στο δέντρο [Bis06]. Στην περίπτωση που το πλήθος των επιπέδων κυματιδιακής ανάλυσης δεν είναι το μέγιστο δυνατό, στο πιο αδρό επίπεδο κάθε υποζώνης θα έχουμε περισσότερους του ενός συντελεστές κυματιδιακής λεπτομέρειας, που υποδηλώνονται με γκρίζο χρώμα στο Σχ. 6.2(α). Καθένας από αυτούς θα αποτελεί ρίζα ενός ξεχωριστού δέντρου, οπότε στο σύνολο της εικόνας και για κάθε κατευθυντική υποζώνη θα αντιστοιχεί ένα δάσος από HMT μοντέλα.

Ανάλογη είναι η δομή του HMT μοντέλου κατά την ανάλυση με 2-Δ μιγαδικά κυματίδια χώρου-συχνότητας μέσω του MCLT, οπότε έχουμε και πάλι έξι ανεξάρτητα δέντρα ανά μπλοκ ανάλυσης. Ακολουθώντας ανάλογη προσέγγιση με εργασίες που μελετούν κωδικοποίηση εικόνων με μπλοκ-DCT μετασχηματισμούς [XGO96, TN99], οργανώνουμε τους MCLT συντελεστές σε δενδρικές δομές, όπως φαίνεται στο Σχ. 6.2(β) το οποίο περιγράφει ένα μπλοκ της εικόνας με $l = 8$, που μεταφράζεται σε αλληλεπικαλυπτόμενα κυματίδια με αποτύπωμα 16×16 εικονοστοιχεία. Τα κυματίδια που βρίσκονται στο ίδιο δέντρο έχουν μεταξύ τους παρεμφερή (όχι όμως ταυτόσημη όπως στην περίπτωση του DT-CWT) διεύθυνση μέγιστης απόκρισης, που υποδηλώνεται κατάλληλα στο Σχ. 6.2(β). Κόμβοι που έχουν συγγένεια γονέα-παιδιού αντιστοιχούν σε σχεδόν συγκατευθυντικά κυματίδια με λόγο χωρικής συχνότητας περίπου 2:1, πρόκειται δηλαδή για αρμονικά μεταξύ τους κυματίδια. Εάν η περιοχή εντός του παραθύρου ανάλυσης χαρακτηρίζεται από πρότυπο υψής με περιοδική δομή, αναμένουμε να ενεργοποιηθεί τόσο το κυματίδιο που αντιστοιχεί στη βασική χωρική συχνότητα (γονιός), όσο και τα κυματίδια που αντιστοιχούν στις αρμονικές χωρικές συχνότητες (απόγονοι), φαινόμενο το οποίο δικαιολογεί τη συγκεκριμένη δομή πιθανοτικών εξαρτήσεων στο μοντέλο μας. Σε αντίθεση με την περίπτωση του αντίστοιχου DT-CWT μοντέλου, παράπλευροι κόμβοι αντιστοιχούν σε κυματίδια μετατοπισμένα στη χωρική συχνότητα (αντί για το χώρο). Για κάθε διεύθυνση, η ανάλυση ολόκληρης της εικόνας εμπλέκει ένα δάσος από HMT μοντέλα, ένα για κάθε μπλοκ της εικόνας. Τόσο στην περίπτωση του DT-CWT, όσο και στην περίπτωση του MCLT, το απλοποιημένο HMT μοντέλο προκύπτει από το αντίστοιχο HMT μοντέλο αφαιρώντας τα βέλη πιθανοτικής εξάρτησης μεταξύ κόμβων σε διαδοχικές κλίμακες.



(α)



(β)

Σχήμα 6.2: Η δομή του κατευθυνόμενου δενδρικού γραφικού μοντέλου στο προτεινόμενο στατιστικό μοντέλο εικόνας για τις δύο μιγαδικές κυματιδιακές αναπαραστάσεις. Και στις δύο περιπτώσεις μοντελοποιούμε με ανεξάρτητα δέντρα καθεμιά από τις έξι κατευθυντικές υποζώνες. (α) Ανάλυση χώρου-κλίμακας με τον 2-Δ DT-CWT σε 3 κλίμακες. (β) Ανάλυση χώρου-συχνότητας με τον 2-Δ MCLT για μπλοκ ανάλυσης με $l = 8$ εικονοστοιχεία.

6.2.3 Περιγραφή Εικόνων με το HMT Μοντέλο: Παραμετροποίηση και Βασικοί Αλγόριθμοι για Συμπερασματολογία και Μάθηση

Και στις δύο μιγαδικές κυματιδιακές αναπαραστάσεις DT-CWT και MCLT μοντελοποιούμε καθένα μιγαδικό συντελεστή λεπτομέρειας ως μείγμα αποτελούμενο από Q 2- Δ κατανομές Gauss, χρησιμοποιώντας ένα ανεξάρτητο δέντρο για καθεμιά από τις έξι κατευθυντικές υποζώνες του μετασχηματισμού. Έστω $\theta = \{\pi_i, \{A_{l,ji}\}_{l=1}^{L-1}, \{\mu_{l,i}, \Sigma_{l,i}\}_{l=1}^L\}$ το σύνολο των παραμέτρων του HMT μοντέλου για μία από τις κατευθυντικές υποζώνες. Υποθέτουμε ότι το HMT δέντρο αποτελείται από L επίπεδα, με $l = 1$ να αντιστοιχεί στο πιο λεπτομερές επίπεδο και $l = L$ στο πιο αδρό. Οι παράμετροι θ αντιστοιχούν σε (1) π_i , $1 \leq i \leq Q$: πρότερη πιθανότητα κατάληψης της i κατάστασης στη ρίζα του δέντρου· (2) $A_{l,ji}$, $1 \leq i, j \leq Q$, $1 \leq l < L$: πιθανότητα μετάβασης από γονιό που βρίσκεται στην j κατάσταση του $l + 1$ επιπέδου σε παιδί στην i κατάσταση του l επιπέδου· (3) $\{\mu_{l,i}, \Sigma_{l,i}\}$, $1 \leq i \leq Q$, $1 \leq l \leq L$: μέσος και συμμεταβλητότητα της γκαουσιανής που αντιστοιχεί στην i κατάσταση του l επιπέδου. Έστω επίσης ότι το t διατρέχει τους $T + 1$ κόμβους του δέντρου, με $t = 0$ να αντιστοιχεί στη ρίζα του δέντρου, και $t = T$ στον τελευταίο κόμβο στην πιο λεπτομερή κλίμακα, που σημαίνει ότι $p_t = \lfloor (t - 1)/4 \rfloor$ είναι ο γονιός του t . Συμβολίζουμε με T_l το σύνολο των δεικτών που αντιστοιχούν στο l επίπεδο ανάλυσης, δηλαδή $T_L = \{0\}$ και $T_l = \{t : p_t \in T_{l+1}\}$, $L < l \leq 1$. Τότε η πρότερη πιθανότητα κατάληψης της κρυφής διακριτής ακολουθίας καταστάσεων $\mathbf{q} = (q_0, \dots, q_T)^T$ είναι

$$P(\mathbf{q}|\theta) = \pi_{q_0} \prod_{l=1}^{L-1} \prod_{t \in T_l} A_{l,q_{p_t} q_t}. \quad (6.2)$$

Δεδομένου του διανύσματος κατάστασης \mathbf{q} , η πιθανότητα του διανύσματος συντελεστών κυματιδίων $\mathbf{y} = (y_0, \dots, y_T)^T$, όπου $\mathbf{y}_t = (y_{t,r}, y_{t,i})^T$ συγκεντρώνει το πραγματικό και φανταστικό μέρος του κάθε μιγαδικού συντελεστή, είναι

$$P(\mathbf{y}|\mathbf{q}, \theta) = N(\mathbf{y}; \boldsymbol{\mu}_{\mathbf{q}}, \boldsymbol{\Sigma}_{\mathbf{q}}) = \prod_{l=1}^{L-1} \prod_{t \in T_l} N(\mathbf{y}_t; \boldsymbol{\mu}_{l,q_t}, \boldsymbol{\Sigma}_{l,q_t}), \quad (6.3)$$

όπου $N(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ συμβολίζει τη συνάρτηση πυκνότητας πιθανότητας όταν το \mathbf{x} ακολουθεί πολυ-μεταβλητή γκαουσιανή κατανομή με μέσο διάνυσμα $\boldsymbol{\mu}$ και πίνακα συμμεταβλητότητας $\boldsymbol{\Sigma}$. Επιβάλλουμε στον πίνακα συμμεταβλητότητας $\boldsymbol{\Sigma}$ μπλοκ-διαγώνια δομή, έτσι ώστε, δεδομένου του \mathbf{q} , οι συντελεστές κυματιδίων \mathbf{y}_t να είναι ανεξάρτητοι μεταξύ τους. Αυτήν την ιδιότητα υπό συνθήκη ανεξαρτησίας (conditional independence), επιβάλλει η δομή του γράφου εξαρτήσεων του HMT μοντέλο στο Σχ. 6.1. Ειδικότερα, όπως και στο [CRBK00], μοντελοποιούμε τις κατανομές παρατηρήσεων ως 2- Δ σφαιρικές γκαουσιανές, $\boldsymbol{\Sigma}_{q_t} = \sigma_{q_t}^2 I$, δηλαδή θεωρούμε ότι η φάση $\phi_t = \angle \mathbf{y}_t$ κατανέμεται ομοιόμορφα στο $[0, 2\pi]$. Θα συζητήσουμε στην Ενότητα 6.6.3 πώς μπορούμε να βελτιώσουμε το πρότερο μοντέλο ως προς αυτό το σημείο. Στην περίπτωση έγχρωμων εικόνων, το διάνυσμα συντελεστών κυματιδίων \mathbf{y}_t θα έχει έξι στοιχεία, συγκεντρώνοντας τις πραγματικές και φανταστικές κυματιδιακές αποκρίσεις καθενός από τα τρία χρωματικά κανάλια στον RGB χρωματικό χώρο.

Συνδυάζοντας το πρότερο μοντέλο (6.2) για το κρυφό μονοπάτι \mathbf{q} με το υπό συνθήκη μοντέλο παραγωγής (6.3) για τους συντελεστές κυματιδίων \mathbf{y} , οδηγούμαστε στην από

κοινού κατανομή

$$P(\mathbf{y}, \mathbf{q}|\boldsymbol{\theta}) = P(\mathbf{y}|\mathbf{q}, \boldsymbol{\theta})P(\mathbf{q}|\boldsymbol{\theta}) = \prod_{l=1}^{L-1} \prod_{t \in T_l} A_{l, q_{p_t} q_t} N(\mathbf{y}_t; \boldsymbol{\mu}_{l, q_t}, \boldsymbol{\Sigma}_{l, q_t}), \quad (6.4)$$

η οποία, όπως και οι συνιστώσες της, αναλύεται σε γινόμενο παραγόντων ορισμένων στους T κόμβους του δέντρου.

Μεγάλο πλεονέκτημα της HMT τεχνικής και γενικότερα πιθανοτικών μοντέλων που μπορούν να εκφραστούν με δενδρικούς γράφους, είναι ότι επιτρέπουν συμπερασματολογία με χρήση ιδιαίτερα αποδοτικών αλγορίθμων *διάδοσης πίστης* (belief propagation) [Pea88, Bis06]. Οι αλγόριθμοι αυτοί, γραμμικής πολυπλοκότητας $\mathcal{O}(T)$ ως προς το πλήθος των κόμβων του δέντρου, γενικεύουν στην περίπτωση δενδρικού γράφου αντίστοιχους αλγορίθμους για γράφους αλυσίδας που είναι γνωστοί από τη βιβλιογραφία των HMM. Πιο συγκεκριμένα, στα HMT, παρόμοια με τα HMM [Rab89], οι ακόλουθες λειτουργίες επιτυγχάνονται με αποδοτικό τρόπο (παραπέμπουμε στα [CNB98, LPH00, PCL00] για περαιτέρω πληροφορίες):

- *Συμπερασματολογία*: Δοσμένης μιας εικόνας, εύρεση της πιθανότητας κάθε κρυφή μεταβλητή να είναι σε κάποια κατάσταση, $\gamma_t(i) = P(q_t = i | \mathbf{y}, \boldsymbol{\theta})$. Αυτό επιτυγχάνεται διατρέχοντας το δένδρο δύο φορές, πρώτα από τα φύλλα προς τη ρίζα και ύστερα από τη ρίζα προς τα φύλλα, υποχρεωτικά με αυτή τη σειρά. Ο αντίστοιχος *άνω-κάτω* (upward-downward) αλγόριθμος γενικεύει τον *μπρος-πίσω* (forward-backward) αλγόριθμο για συμπερασματολογία σε αλυσίδες Markov. Σημειωτέον ότι το άνω πέρασμα του αλγορίθμου αρκεί για τον *υπολογισμό πιθανοφάνειας* (scoring) της εικόνας με βάση το μοντέλο $P(\mathbf{y}|\boldsymbol{\theta})$.
- *Αποκωδικοποίηση*: Δοσμένης μιας εικόνας, εύρεση της πιο πιθανής ακολουθίας καταστάσεων στο δένδρο, $\hat{\mathbf{q}} = \operatorname{argmax}_{\mathbf{q}} P(\mathbf{q}|\mathbf{y}, \boldsymbol{\theta})$. Παρόμοια με τη συμπερασματολογία απαιτεί ένα προς τα άνω και ένα προς τα κάτω πέρασμα, και ο σχετικός αλγόριθμος δυναμικού προγραμματισμού γενικεύει τον αλγόριθμο Viterbi για αποκωδικοποίηση σε αλυσίδες Markov.
- *Μάθηση*: Δοσμένου ενός συνόλου εικόνων εκπαίδευσης, εύρεση εκείνων των παραμέτρων (πιθανότητα αρχικής κατάστασης, πίνακας μετάβασης καταστάσεων, μέσος και συμμεταβλητότητα κατανομών παρατήρησης) που μεγιστοποιούν την πιθανοφάνεια του μοντέλου, $\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} P(\mathbf{y}|\boldsymbol{\theta})$. Τούτο γίνεται με τον επαναληπτικό αλγόριθμο *Μεγιστοποίησης Προσδοκίας* (Expectation Maximization - EM) [DLR77], χρησιμοποιώντας τον άνω-κάτω αλγόριθμο ως υπορουτίνα, εντελώς αντίστοιχα με την περίπτωση των HMM. Σημειωτέον ότι ο EM αλγόριθμος εγγυάται την εύρεση ενός τοπικού μόνο μεγίστου της συνάρτησης πιθανοφάνειας. Επίσης, τεχνικές για εκπαίδευση HMM στην περίπτωση λίγων δεδομένων εκπαίδευσης, όπως, για παράδειγμα, *δέσιμο παραμέτρων* (parameter tying), μεταφέρονται εύκολα και στην περίπτωση του HMT. Για παράδειγμα, δένουμε τις παραμέτρους που αντιστοιχούν στην ίδια κατεύθυνση κυματιδιακής ανάλυσης, έτσι ώστε να χρησιμοποιούμε όλα τα δέντρα του συγκεκριμένου δάσους ως σύνολο μάθησης και οι αντίστοιχες παράμετροι να εκτιμούνται αξιόπιστα. Τεχνικές κανονικοποίησης που αποτρέπουν την *υπερ-ειδίκευση* (over-fitting) της αναπαράστασης μπορεί να φανούν χρήσιμες στην περίπτωση που αυξηθεί υπερβολικά η πολυπλοκότητα του μοντέλου [SPS06, KSJ07a, KSJ07b].

Παρόμοιες τεχνικές για συμπερασματολογία, αποκωδικοποίηση και μάθηση μπορούν να αναπτυχθούν και για το απλούστερο HIT μοντέλο, γενικεύοντας αντίστοιχες τεχνικές από το ακολουθιακό GMM μοντέλο. Καθότι σε αυτήν την περίπτωση δεν υπάρχουν εξαρτήσεις μεταξύ διαδοχικών επιπέδων του δέντρου, μπορεί κανείς να τα χειριστεί ανεξάρτητα με απλούστερους αλγορίθμους.

Ως τώρα, το HMT μοντέλο κυματιδίων έχει χρησιμοποιηθεί με επιτυχία σε πλήθος εφαρμογών επεξεργασίας και ανάλυσης εικόνας. Στις Ενότητες 6.3 και 6.4 θα περιγράψουμε την εφαρμογή του στην αποθορυβοποίηση και σύνθεση εικόνων, αντίστοιχα, ενώ στο Ενότητα 6.5 θα περιγράψουμε με λεπτομέρεια την εφαρμογή του στο πρόβλημα της ενδοσυμπλήρωσης εικόνων, που αποτελεί και τη βασική συμβολή της εργασίας μας. Επιπρόσθετα, αναφέρουμε ενδεικτικά την εφαρμογή HMT μοντέλου σε προβλήματα κατηγοριοποίησης, κατάτμησης και ανάκτησης εικόνας [CB01, DV02, SPS06, KSJ07b]. Σε αυτές τις μεθόδους εναλλακτικά μοντέλα HMT, εκπαιδευμένα σε διαφορετικές εικόνες υψής, ανταγωνίζονται για να ερμηνεύσουν μια εικόνα ελέγχου· η εικόνα ελέγχου αποδίδεται στο μοντέλο που μεγιστοποιεί την πιθανοφάνεια παρατήρησης. Η επιτυχία του HMT μοντέλου σε αυτές τις εφαρμογές αναδεικνύει την *διακριτική* (discriminating) του ικανότητα ως προς το χαρακτηρισμό της υψής εικόνων.

6.3 Αποθορυβοποίηση με Κρυφό Δένδρο Markov στο Χώρο των Μιγαδικών Κυματιδίων

Στο κλασικό πρόβλημα της αποθορυβοποίησης προσπαθούμε να εκτιμήσουμε ένα σήμα ή εικόνα από θορυβώδη εκδοχή του. Μια από τις πρώτες επιτυχημένες εφαρμογές του HMT μοντέλου ήταν σε αποθορυβοποίηση σημάτων και εικόνων αλλοιωμένων από λευκό προσθετικό θόρυβο Gauss. Το αρχικό μοντέλο αφορούσε 1-Δ σήματα και πραγματικά κυματίδια [CNB98] και αργότερα επεκτάθηκε σε 2-Δ εικόνες [RCB01], με χρήση και του DT-CWT [CRBK00]. Αναφορικά με το πρόβλημα της αποθορυβοποίησης, οι πρωτότυπες συνεισφορές μας συνίστανται στη βελτίωση του HMT μοντέλου ως προς τα ακόλουθα σημεία:

1. Ενώ το HMT μοντέλο ως τώρα έχει χρησιμοποιηθεί με κυματίδια χώρου-κλίμακας όπως ο DT-CWT, δείχνουμε ότι μπορεί να συνδυαστεί και με κυματίδια χώρου-συχνότητας και συγκεκριμένα με τον 2-Δ MCLT, δίνοντας αρκετά βελτιωμένες επιδόσεις αποθορυβοποίησης για ορισμένες κατηγορίες εικόνων.
2. Ενώ σε προηγούμενες σχετικές εργασίες η εκτίμηση των παραμέτρων θ του HMT μοντέλου από τα θορυβώδη δεδομένα της εικόνας γινόταν με ευρετικές τεχνικές, έχουμε αναπτύξει EM επαναληπτικούς αλγορίθμους για μάθηση με το κρυφό μοντέλο Markov υπό καθεστώς θορυβωδών παρατηρήσεων, οι οποίοι αποδεδειγμένα αυξάνουν μονοτονικά την πιθανοφάνεια των θορυβωδών παρατηρήσεων. Ανάλογους EM αλγορίθμους μάθησης έχουμε αναπτύξει για κρυφές μαρκοβιανές αλυσίδες [PKPM09] και για Poisson στατιστικά θορύβου [LMP09] σε άλλες μας εργασίες.

Υποθέτοντας για απλότητα ότι απεικονίζουμε τη θορυβώδη εικόνα σε ορθοκανονική βάση (υπόθεση που ισχύει επακριβώς μόνο για ορθογώνιους κρίσιμα δειγματοληπτημένους μετασχηματισμούς), ο i.i.d. προσθετικός θόρυβος Gauss με μέση τιμή 0 και



Σχήμα 6.3: Οι επτά γκριζες εικόνες που χρησιμοποιήθηκαν στα πειράματα αποθρομβοποίησης. Από αριστερά προς τα δεξιά: 'Lena', 'Barbara', 'boat', 'cameraman', 'fingerprint', 'house', 'peppers'. Η διάστασή τους είναι 512×512 ('Lena', 'Barbara', 'boat', 'fingerprint') ή 256×256 ('cameraman', 'house', 'peppers') εικονοστοιχεία. Οι εικόνες (εκτός του 'cameraman') είναι ταυτόσημες με αυτές του [PSWS03].

μεταβλητότητα σ_n^2 στο χωρικό πεδίο απεικονίζεται σε i.i.d. προσθετικό θόρυβο Gauss με τα ίδια χαρακτηριστικά στο πεδίο των κυματιδίων. Έτσι, εάν y είναι το ζητούμενο διάλυμα κυματιδιακών συντελεστών της καθαρής εικόνας και y_n η θορυβώδης εκδοχή του, τότε $y_n \sim N(y, \Sigma_n)$, όπου $\Sigma_n = \sigma_n^2 I$. Υποθέτοντας ότι οι συντελεστές y της καθαρής εικόνας ακολουθούν το HMT μοντέλο, Εξς. (6.2) και (6.3), οδηγούμαστε σε θορυβώδεις συντελεστές κυματιδίων y_n , των οποίων η πιθανότητα δεδομένου του μονοπατιού q είναι

$$P(y_n | q, \theta) = N(y_n; \mu_q, \Sigma_q + \Sigma_n) = \prod_{t=0}^T N(y_{n,t}; \mu_{q_t}, \Sigma_{q_t} + \Sigma_{n,t}). \quad (6.5)$$

Συνδυάζοντας την Εξ. (6.5) με την πρότερη κατανομή Εξ. (6.2) για το κρυφό μονοπάτι $P(q|\theta)$, προκύπτει ότι η από κοινού κατανομή $P(y_n, q|\theta) = P(y_n|q, \theta)P(q|\theta)$ αποδομείται σε γινόμενο όρων ανά κόμβο του δέντρου του Σχ. 6.1, όπως και η αντίστοιχη έκφραση $P(y_n, q|\theta)$ της Εξ. (6.4) για την καθαρή εικόνα. Πιο συγκεκριμένα, η βασική διαφορά της $P(y_n|q, \theta)$ σε σχέση με τη μη θορυβώδη περίπτωση της Εξ. (6.3) είναι ότι η συμμεταβλητότητα του θορύβου $\Sigma_{n,t}$ επαυξάνει τη συμμεταβλητότητα Σ_{q_t} κάθε κατάστασης του HMT μοντέλου.

Οι αλγόριθμοι για χειρισμό του HMT μοντέλου που συζητήσαμε στην Ενότητα 6.2 δεν είναι άμεσα εφαρμόσιμοι στην περίπτωση θορυβωδών δεδομένων. Οι τεχνικές που έχουν προταθεί στη σχετική βιβλιογραφία για προσαρμογή των σχετικών αλγορίθμων [CNB98] έχουν ευρετικό χαρακτήρα και έτσι δεν εξασφαλίζουν την καλύτερη δυνατή επίδοση. Μια ενδιαφέρουσα συνεισφορά της εργασίας μας είναι ότι έχουμε αναπτύξει βέλτιστους αλγορίθμους για συμπερασματολογία, αποκωδικοποίηση, και μάθηση με το κρυφό μοντέλο Markov υπό καθεστώς θορυβωδών παρατηρήσεων, τους οποίους και παρουσιάζουμε για την περίπτωση του HMM με λεπτομέρεια στο Παράρτημα, Ενότητα Α.3. Στα πειράματα που παρουσιάζουμε στη συνέχεια έχουμε χρησιμοποιήσει αυτούς τους αλγορίθμους, κατάλληλα προσαρμοσμένους στο HMT μοντέλο.

Σε αυτό το πλαίσιο, ο αλγόριθμος για αποθρομβοποίηση με το HMT μοντέλο αποτελεί επέκταση των τεχνικών στο [CNB98] και συνίσταται στα ακόλουθα βήματα:

1. *Μάθηση*: Εκτίμηση των παραμέτρων θ του HMT μοντέλου με μεγιστοποίηση της πιθανοφάνειας πάνω στη θορυβώδη εικόνα, $P(y_n|\theta)$, με τη βοήθεια του προτεινόμενου επαναληπτικού αλγορίθμου EM.
2. *Συμπερασματολογία*: Εκτίμηση της πιθανότητας κατάληψης για τις κρυφές καταστάσεις $\gamma_t(i) = P(q_t = i | y_n, \theta)$ στους κρυφούς κόμβους του δέντρου με τη βοήθεια του άνω-κάτω αλγορίθμου.

3. *Προσαρμοστική Wiener αποθορυβοποίηση*: Εκτίμηση των συντελεστών κυματιδίων που ελαχιστοποιούν το αναμενόμενο τετραγωνικό σφάλμα. Ο σχετικός υπολογισμός αποδομείται ανά συντελεστή κυματιδίων, δίνοντας $\hat{y}_t = \sum_i \gamma_t(i) \Sigma_{i,n} (\Sigma_i^{-1} \mu_i + \Sigma_n^{-1} y_{n,t})$, όπου $\Sigma_{i,n}^{-1} = \Sigma_i^{-1} + \Sigma_n^{-1}$.
4. *Λύση στο χωρικό πεδίο*: Η αποθορυβοποιημένη εικόνα προκύπτει με αντίστροφο μετασχηματισμό κυματιδίων, $\hat{x} = \text{IDWT}(\hat{y})$.

Παρουσιάζουμε στη συνέχεια κάποια ενδεικτικά πειράματα αποθορυβοποίησης εικόνων με τη βοήθεια του HMT μοντέλου και χρήση των κυματιδιακών αναπαραστάσεων DT-CWT και MCLT, ενώ συγκρίνουμε και με το αντίστοιχο HIT μοντέλο που χειρίζεται ξεχωριστά κάθε κυματιδιακό συντελεστή. Για αναφορά, περιλάβαμε στα πειράματά μας και την κλασική χωρικά προσαρμοστική μέθοδο Wiener [Lee80], όπως υλοποιείται στη συνάρτηση *wiener2* του *Matlab* (χρησιμοποιήσαμε παράθυρο ανάλυσης 3×3 εικονοστοιχεία). Επίσης περιλάβαμε δύο state-of-the-art τεχνικές αποθορυβοποίησης, την επίσης κυματιδιακή BLS-GSM τεχνική του [PSWS03] και τη μη-παραμετρική BM3D τεχνική του [DFKE07] που αυτή τη στιγμή θεωρείται ίσως η πιο αποδοτική (με κριτήριο το μέσο τετραγωνικό σφάλμα) δημοσιευμένη μέθοδος αποθορυβοποίησης.

Χρησιμοποιήσαμε επτά τυπικές γκριζες εικόνες, Σχ. 6.3, με μέγιστη τιμή φωτεινότητας 255 τις οποίες αλλοιώσαμε με λευκό προσθετικό θόρυβο Gauss σε διάφορα επίπεδα έντασης (τυπική απόκλιση θορύβου σ_n 5, 10, 20, 50). Στη αναπαράσταση χώρου-κλίμακας με τον DT-CWT, Ενότ. 5.2.2, προχωρήσαμε σε τέσσερα επίπεδα ανάλυσης και χρησιμοποιήσαμε τα (9,7) διορθογώνια φίλτρα 'Antonini' (επίπεδο 1) και τα (6,6) ορθογώνια φίλτρα 'q-shift' (επίπεδα 2-4) [Kin01, σελ. 246]. Στη αναπαράσταση χώρου-συχνότητας με τον MCLT, Ενότ. 5.2.3, χρησιμοποιήσαμε παράθυρα συνημιτονοειδούς προφίλ με $l = 8$ (άρα βάθος δέντρου 3 επίπεδα) και 50% αλληλεπικάλυψη που αντιστοιχεί σε κυματίδια με αποτύπωμα 16×16 εικονοστοιχεία. Η κατανομή των κυματιδίων στους κόμβους των HMT και HIT μοντέλων που χρησιμοποιήσαμε μοντελοποιήθηκε με μείγμα $Q = 4$ σφαιρικών συναρτήσεων Gauss. Εκπαιδεύσαμε τα σχετικά μοντέλα πάνω στους συντελεστές κυματιδίων της θορυβώδους εικόνας, αυξάνοντας σταδιακά το πλήθος των μειγμάτων από $Q = 1$ σε $Q = 4$ και βελτιώνοντας την πιθανοφάνεια των παρατηρήσεων ρυθμίζοντας τις παραμέτρους θ με τον EM αλγόριθμο της Ενότ. 6.2 (50 επαναλήψεις για κάθε Q).

Συνοψίζουμε τα αποτελέσματα του πειράματος στον Πίν. 6.1, όπου η απόδοση του κάθε αλγορίθμου αξιολογείται με βάση το σηματοθορυβικό λόγο (σε dB) του αποθορυβοποιημένου αποτελέσματος, ορισμένου ως $\text{PSNR} = 10 \log_{10} (255^2 / \text{MSE})$, όπου MSE είναι το μέσο τετραγωνικό σφάλμα εκτίμησης. Οι δείκτες επίδοσης των BLS-GSM και BM3D τεχνικών είναι αυτοί που έχουν αναφερθεί στις αντίστοιχες δημοσιεύσεις. Στον πίνακα αυτό δείχνουμε και τη μέση βελτίωση του σηματοθορυβικού λόγου σε σχέση με τη θορυβώδη εικόνα IPSNR για καθεμιά από τις συγκρινόμενες μεθόδους, τόσο ξεχωριστά για κάθε εικόνα παίρνοντας μέσο όρο ως προς όλα τα επίπεδα θορύβου, όσο και ξεχωριστά για κάθε επίπεδο θορύβου παίρνοντας μέσο όρο ως προς όλες τις εικόνες, όσο, τέλος, παίρνοντας μέσο όρο ως προς όλες τις εικόνες και όλα τα επίπεδα θορύβου. Βλέπουμε καταρχήν ότι τα κυματιδιακά μοντέλα αποδίδουν αισθητά καλύτερα (γύρω στα 2.5 dB) από το απλό Wiener χωρικό μοντέλο. Επίσης, κατά μέσο όρο, οι DT-CWT και MCLT αναπαραστάσεις είναι περίπου το ίδιο αποδοτικές. Σε συγκεκριμένες εικόνες όμως οι δύο εναλλακτικές μιγαδικές αναπαραστάσεις εμφανίζουν διαφοροποιήσεις στην επίδοση· για παράδειγμα, στις εικόνες 'Lena', 'cameraman', 'peppers' (που χαρακτηρίζονται από τμηματικά ομα-

λό χαρακτήρα) προηγείται ελαφρά η DT-CWT αναπαράσταση (γύρω στα 0.2 dB), ενώ στην εικόνα 'Barbara' (που χαρακτηρίζεται από σημαντική κάλυψη σε περιοδικές δομές) προηγείται η MCLT αναπαράσταση (0.75 dB κατά μέσο όρο). Συγκρίνοντας τα μοντέλα με δενδρικές εξαρτήσεις (HMT-DTCWT, HMT-MCLT) με τα αντίστοιχα μοντέλα όπου οι κυματιδιακοί συντελεστές μοντελοποιούνται ανεξάρτητα (HMT-DTCWT, HMT-MCLT), παρατηρούμε βελτίωση που είναι συστηματική για όλες τις εικόνες και επίπεδα θορύβου και ανέρχεται στα 0.3-0.4 dB κατά μέσο όρο. Ποιοτικά, η βελτίωση αυτή φαίνεται για παράδειγμα στις επίπεδες περιοχές της εικόνας 'cameraman' στο Σχ. 6.4, όπου το HMT αποτέλεσμα είναι πιο ομαλό σε σχέση με το HIT.

Είναι ενθαρρυντικό ότι η επίδοση των κυματιδιακών HMT μοντέλων είναι πολύ κοντά στις πιο αποδοτικές δημοσιευμένες τεχνικές αποθορυβοποίησης. Συγκεκριμένα, τα εξεταζόμενα HMT μοντέλα υπολείπονται λίγο της επίσης παραμετρικής BLS-GSM τεχνικής (0.2-0.3 dB κατά μέσο όρο), ενώ σε εικόνες με έντονες περιοδικές δομές 'Barbara', 'fingerprint' οι BLS-GSM και HMT-MCLT τεχνικές πετυχαίνουν περίπου το ίδιο μέσο τετραγωνικό σφάλμα. Η μη παραμετρική BM3D τεχνική, που είναι καθοδηγούμενη από τα δεδομένα και βασίζεται στη συστηματική αξιοποίηση παραθύρων ανάλυσης από όλη την εικόνα παρακάμπτοντας έτσι το πρόβλημα της μοντελοποίησης της εικόνας, υπερτερεί περισσότερο (περίπου 0.9-1.0 dB κατά μέσο όρο). Σημειώνουμε όμως εδώ ότι καμία από τις BLS-GSM και BM3D τεχνικές δεν αποτελεί αναγεννητικό μοντέλο εικόνας, και συνεπώς, αντίθετα από το κυματιδιακό HMT μοντέλο δεν προσφέρεται για να προσαρμοστεί στο πρόβλημα της ενδοσυμπλήρωσης εικόνων που μας ενδιαφέρει κυρίως εδώ.

6.4 Σύνθεση Υφής με Κρυφό Δένδρο Markov στο Χώρο των Μιγαδικών Κυματιδίων

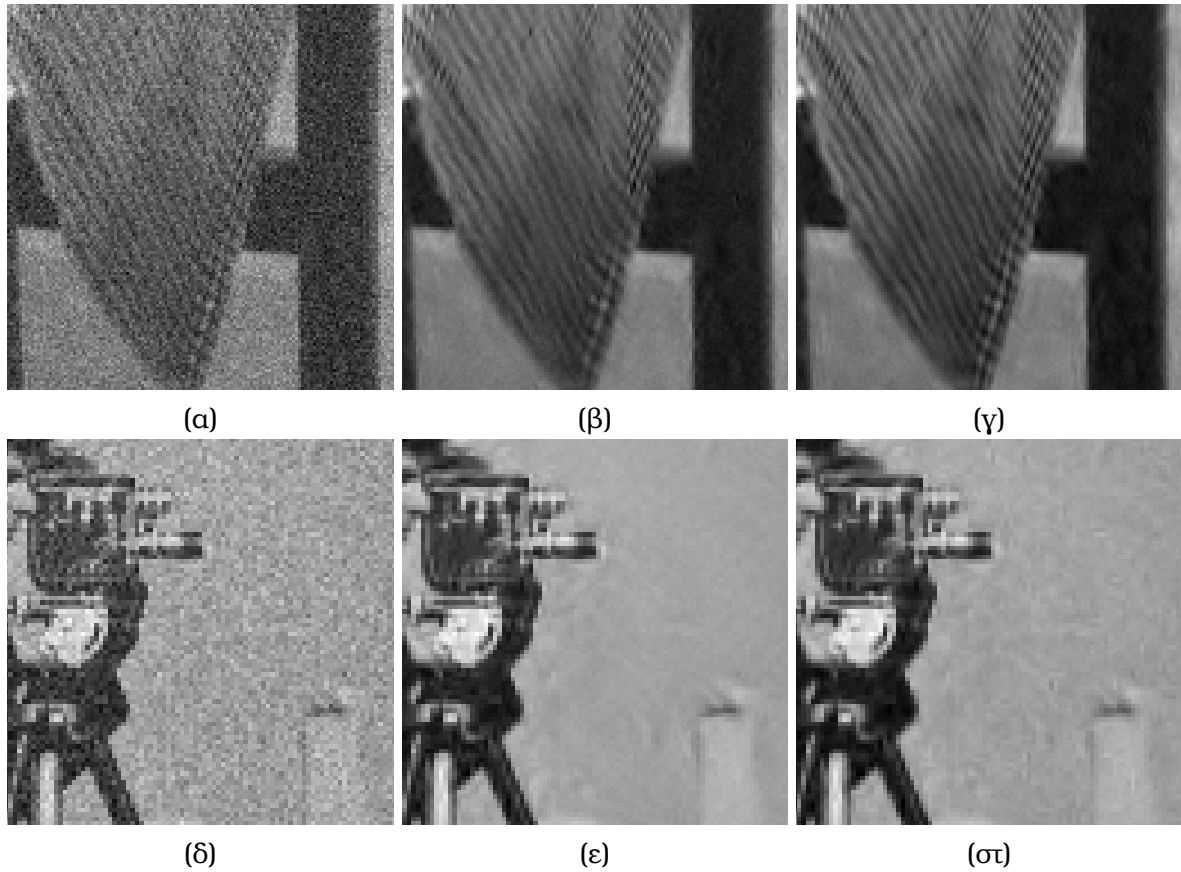
Καθότι το κυματιδιακό HMT μοντέλο είναι *αναγεννητικό* (generative), ένας άμεσος τρόπος για να μελετήσουμε τις ιδιότητές του είναι να εξετάσουμε δείγματα εικόνων που συντίθενται από αυτό [SPS06]. Η σημασία αυτής της διαδικασίας ελέγχου μοντέλων εικόνας έχει ιδιαίτερα τονιστεί από τη σχολή της θεωρίας προτύπων του Grenander [Gre93, Mum96, GM07]. Επίσης, η διαδικασία παραγωγής δειγμάτων από το μοντέλο εικόνας μας προετοιμάζει για το πρόβλημα της ενδοσυμπλήρωσης της Ενót. 6.5, το οποίο μπορεί επίσης να θεωρηθεί ως πρόβλημα σύνθεσης δειγμάτων εικόνας, υπό τη συνθήκη αυτά να λαμβάνουν συγκεκριμένες τιμές στις γνωστές περιοχές της εικόνας. Για αυτόν το σκοπό έχουμε πραγματοποιήσει πειράματα σύνθεσης υφής, όπου πρώτα εκπαιδεύουμε ένα κυματιδιακό HMT μοντέλο σε μια εικόνα συγκεκριμένης υφής με τη μέθοδο μέγιστης πιθανοφάνειας που συζητήσαμε στην Ενót. 6.2, και ύστερα λαμβάνουμε δείγματα από το εκπαιδευμένο μοντέλο.

Η σχετική διαδικασία (*δίχως όρους*) *δειγματοληψίας* ((unconditional) sampling) από το HMT μοντέλο κυματιδίων είναι ιδιαίτερα απλή χάρη στον αναγεννητικό του χαρακτήρα. Χρησιμοποιώντας το συμβολισμό της Ενót. 6.2.3 για τις παραμέτρους του μοντέλου, η σύνθεση εικόνας λαμβάνει χώρα σε τρία βήματα:

1. *Δειγματοληψία κρυφών μεταβλητών.* Λαμβάνουμε δείγμα των κρυφών μεταβλητών του δέντρου, ως εξής: (1) Λαμβάνουμε δείγμα της διακριτής κρυφής μεταβλητής που αντιστοιχεί στη ρίζα του δένδρου, δηλαδή $P(q_0 = i) = \pi_i$. (2) Διατρέχοντας τα

Πίνακας 6.1: Πείραμα αποθορυβοποίησης για λευκό προσθετικό θόρυβο Gauss μεταβλητότητας σ_n^2 . Εξετάζουμε την επίδοση, μετρημένη με όρους PSNR (dB), του δενδρικού μοντέλου HMT συνδυασμένου με τον DT-CWT ή τον MCLT σε σχέση με αντίστοιχες εκδοχές του ανεξάρτητου μοντέλου HIT δίχως δενδρικές εξαρτήσεις. Επίσης περιλαμβάνουμε αποτελέσματα εναλλακτικών τεχνικών (Wiener, BLS-GSM, BM3D). Για κάθε μέθοδο δείχνουμε επίσης μέσο κέρδος αποθορυβοποίησης για κάθε εικόνα ως προς όλα τα επίπεδα θορύβου και για κάθε επίπεδο θορύβου ως προς όλες τις εικόνες.

Image/ σ_n	PSNR (dB) / Methods							
	NOISY	WIE- NER	BLS- GSM	BM3D	HMT- DTCWT	HMT- MCLT	HIT- DTCWT	HIT- MCLT
lenna/5	34.16	37.18	38.49	38.72	38.54	38.59	38.26	38.36
lenna/10	28.14	33.00	35.61	35.93	35.37	35.39	34.91	35.07
lenna/20	22.12	28.50	32.66	33.05	32.17	32.06	31.73	31.80
lenna/50	14.16	21.69	28.61	28.86	27.77	27.49	27.35	27.35
lenna/IPSNR	0.00	5.44	9.19	9.49	8.81	8.74	8.41	8.49
barbara/5	34.16	36.17	37.39	38.31	37.58	38.10	37.27	37.97
barbara/10	28.14	31.33	34.03	34.98	33.59	34.36	33.02	34.13
barbara/20	22.12	26.66	30.32	31.78	29.50	30.35	28.86	30.28
barbara/50	14.16	20.64	25.48	27.17	24.71	25.38	24.04	25.31
barbar/IPSNR	0.00	4.05	7.16	8.41	6.70	7.40	6.15	7.27
boat/5	34.16	36.01	36.97	37.28	37.07	37.18	36.88	37.03
boat/10	28.14	31.82	33.58	33.92	33.43	33.49	33.10	33.23
boat/20	22.12	27.65	30.38	30.88	30.10	30.15	29.64	29.86
boat/50	14.16	21.35	26.38	26.64	25.82	25.73	25.46	25.58
boat/IPSNR	0.00	4.56	7.18	7.53	6.96	6.99	6.62	6.78
cameraman/5	34.14	37.01	-	38.29	37.75	37.63	37.24	37.11
cameraman/10	28.12	31.94	-	34.18	33.11	32.97	32.61	32.43
cameraman/20	22.10	27.12	-	30.48	29.04	28.98	28.50	28.53
cameraman/50	14.14	20.75	-	25.84	24.65	24.47	24.01	24.01
camera/IPSNR	0.00	4.58	-	7.57	6.51	6.39	5.97	5.89
fingerprint/5	34.16	34.86	36.68	36.51	36.43	36.57	36.37	36.52
fingerprint/10	28.14	30.05	32.45	32.46	32.30	32.39	32.24	32.34
fingerprint/20	22.12	26.18	28.60	28.81	28.59	28.60	28.53	28.57
fingerprint/50	14.16	20.93	24.16	24.36	24.00	23.98	23.86	23.98
finger/IPSNR	0.00	3.35	5.82	5.89	5.68	5.74	5.60	5.70
house/5	34.14	36.98	38.65	39.83	38.69	39.14	38.26	38.89
house/10	28.12	32.67	35.35	36.71	35.06	35.38	34.56	35.09
house/20	22.10	28.14	32.39	33.77	31.80	32.00	31.26	31.57
house/50	14.14	21.44	28.26	29.37	27.48	27.16	26.95	27.03
house/IPSNR	0.00	5.18	9.04	10.30	8.64	8.79	8.13	8.52
peppers/5	34.14	36.85	37.31	38.12	37.73	37.46	37.50	37.14
peppers/10	28.12	32.38	33.77	34.68	33.87	33.64	33.39	33.10
peppers/20	22.10	27.75	30.31	31.29	30.18	30.13	29.58	29.58
peppers/50	14.14	21.17	25.90	26.41	24.97	24.92	24.63	24.72
pepper/IPSNR	0.00	4.91	7.20	8.00	7.07	6.91	6.65	6.51
IPSNR/5	0.00	2.28	-	4.00	3.53	3.66	3.24	3.42
IPSNR/10	0.00	3.75	-	6.56	5.69	5.81	5.27	5.49
IPSNR/20	0.00	5.32	-	9.32	8.09	8.21	7.62	7.91
IPSNR/50	0.00	6.98	-	12.80	11.48	11.44	11.03	11.27
IPSNR/avg	0.00	4.58	-	8.17	7.19	7.28	6.79	7.02



Σχήμα 6.4: Λεπτομέρειες αποτελεσμάτων αποθορυβοποίησης για λευκό θόρυβο Gauss με $\sigma_n = 20$. Πάνω: απόσπασμα από την εικόνα 'Barbara'· θορυβώδης εικόνα (α) και αποτέλεσμα αποθορυβοποίησης με τις τεχνικές HMT-DTCWT (β) και HMT-MCLT (γ). Κάτω: απόσπασμα από την εικόνα 'cameraman'· θορυβώδης εικόνα (δ) και αποτέλεσμα αποθορυβοποίησης με τις τεχνικές HMT-DTCWT (ε) και HIT-DTCWT (στ).

επίπεδα από τη ρίζα προς τα φύλλα, λαμβάνουμε δείγμα για κάθε κρυφή μεταβλητή q_t στο επίπεδο l , δεδομένου ότι στο γονιό της q_{p_t} στο επίπεδο $l + 1$ έχει αποδοθεί η κατάσταση j , με $P(q_t = i | q_{p_t} = j) = A_{l,j,i}$. Η διαδικασία που περιγράψαμε αποτελεί άμεση γενίκευση του αλγορίθμου δειγματοληψίας σε διακριτές αλυσίδες Markov.

2. *Δειγματοληψία συντελεστών κυματιδιακής λεπτομέρειας.* Δεδομένης της κρυφής μεταβλητής q_t στον κόμβο t /επίπεδο l του δέντρου, ο αντίστοιχος μιγαδικός συντελεστής κυματιδίου y_t είναι γκαουσιανά κατανομημένος με μέσο $\mu_{l,i}$ και συμμεταβλητότητα $\Sigma_{l,i}$, οπότε αρκεί να λάβουμε δείγμα του $y_t \sim N(\mu_{l,i}, \Sigma_{l,i})$.
3. *Σύνθεση της εικόνας στο χωρικό πεδίο.* Η συνθετική εικόνα x παράγεται από τους συνθετικούς συντελεστές κυματιδίων y μέσω του αντίστροφου κυματιδιακού μετασχηματισμού. Καθότι το μοντέλο μας περιγράφει μόνο τους συντελεστές λεπτομέρειας, είτε αντιγράφουμε τους συντελεστές κλιμάκωσης από την αρχική εικόνα εκπαίδευσης, είτε τους αποδίδουμε μηδενικές τιμές.

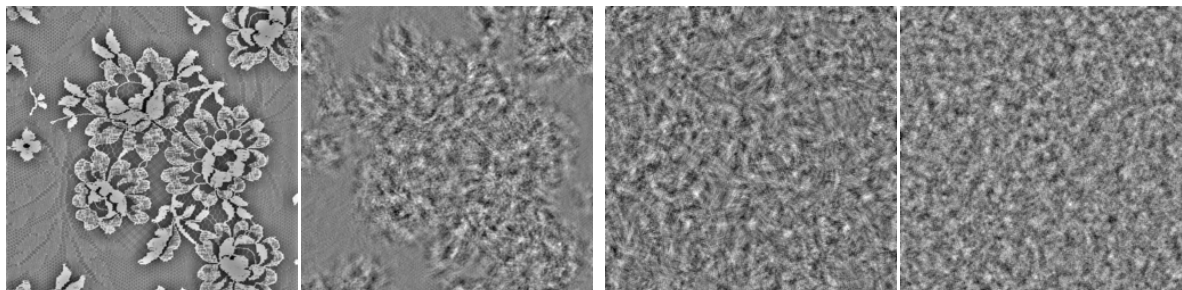
Η παραπάνω διαδικασία είναι πανομοιότυπη ανεξάρτητα αν χρησιμοποιούμε την DT-CWT ή την MCLT μιγαδική κυματιδιακή αναπαράσταση.

Ακολουθώντας την παραπάνω τακτική, πραγματοποιήσαμε πειράματα σύνθεσης υφής με μοντέλα εκπαιδευμένα τόσο στο γνωστό λεύκωμα εικόνων υφής του P. Brodatz [Bro66], όσο και σε τμήματα τοιχογραφιών από το Ακρωτήρι. Δείχνουμε αντιπροσωπευτικά αποτελέσματα στα Σχ. 6.5 και 6.6, αντίστοιχα. Δύο παρατηρήσεις: (α) Καθότι η HMT τεχνική μοντελοποιεί τους συντελεστές λεπτομέρειας και όχι τους συντελεστές κλιμάκωσης του μετασχηματισμού κυματιδίων, δείχνουμε για όλες τις εικόνες μόνο το υπεραυτό τμήμα τους, που προκύπτει μηδενίζοντας τους συντελεστές κλιμάκωσης. Εναλλακτικά, θα μπορούσαμε να αντιγράψουμε αυτούσια τους συντελεστές κλιμάκωσης από την πηγαία εικόνα. (β) Στο μοντέλο μας ενσωματώνουμε μόνο πληροφορία σχετική με το μέτρο των μιγαδικών συντελεστών κυματιδίων, αγνοώντας την πληροφορία φάσης· με άλλα λόγια, η εκπαίδευση δύο μοντέλων HMT, ένα σε εικόνα υφής με την αρχική φάση (*σύμφωνη* αρχική εικόνα), και ένα στην ίδια εικόνα υφής αλλά με τυχαιοποιημένη φάση (*ασύμφωνη* αρχική εικόνα) οδηγεί στο ίδιο ακριβώς μοντέλο· καθότι η ασύμφωνη αρχική εικόνα αποτελεί άνω φράγμα για την επίδοση μοντέλων βασισμένων αποκλειστικά σε πληροφορία μέτρου, τη δείχνουμε και αυτή στο Σχ. 6.5.

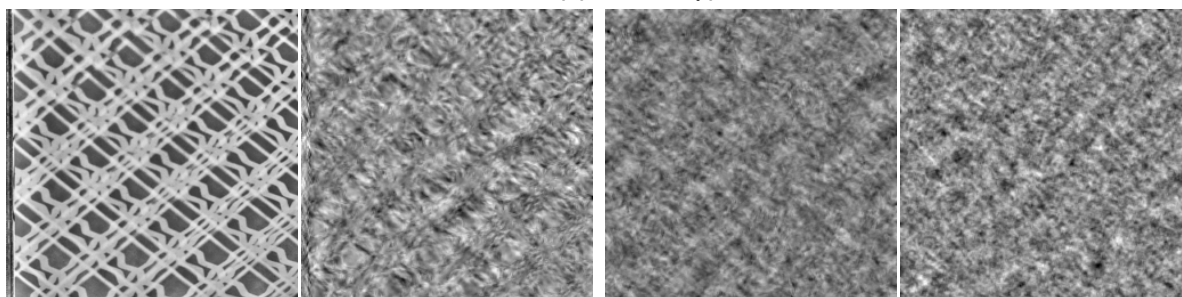
Παρατηρώντας τα Σχ. 6.5 και 6.6, αξίζει να σχολιάσουμε τα ακόλουθα σημεία:

1. Το HMT μοντέλο πετυχαίνει σε ικανοποιητικό βαθμό να αναπαράγει ιδιότητες *χαμηλού επιπέδου* των εικόνων υφής. Για παράδειγμα, στην υφή δαντέλας D42, Σχ. 6.5, η συνθετική εικόνα έχει σε σημαντικό βαθμό την έντονη μικροδομή που συναντάμε στην αρχική εικόνα και οφείλεται στο δίκτυο της δαντέλας. Οπτικά, οι εικόνες συνθετικής υφής μοιάζουν με φράκταλς, και θα μπορούσαμε να πούμε ότι φράκταλ περιγραφές της αρχικής εικόνας (ακόμα περισσότερο, της ασύμφωνης αρχικής εικόνας), όπως είναι η φράκταλ διάσταση, αναπαράγονται στην εικόνα υφής· τούτο μπορεί να έχει και θεωρητική βάση δεδομένης της καταλληλότητας του μετασχηματισμού κυματιδίων για περιγραφή φράκταλ σημάτων [Wor93]. Βέβαια, το HMT μοντέλο συνθέτει κυματίδια με μη γκαουσιανές κατανομές και κατά συνέπεια οι αντίστοιχες εικόνες έχουν πλουσιότερη δομή σε σχέση με τις γκαουσιανές φράκταλ fBm 2-Δ στοχαστικές διαδικασίες του Σχ. 5.1.
2. Το μοντέλο HMT δεν πετυχαίνει να ανασυνθέσει ρεαλιστικές μακροδομές που υπάρχουν στις αρχικές εικόνες, ιδιαίτερα αυτές που έχουν σαφείς ακμές, όπως για παράδειγμα τα βότσαλα D31 ή ο πλίνθινος τοίχος D96. Με άλλα λόγια, το μοντέλο HMT δεν μπορεί να συλλάβει ιδιότητες υψηλού επιπέδου στις εικόνες. Προς αυτήν την κατεύθυνση θα ήταν επιθυμητή η ανάπτυξη ιεραρχικών μοντέλων αντικειμένων ικανών να καθοδηγούν τη διαδικασία σύνθεσης οργανώνοντας τους ενεργούς κυματιδιακούς συντελεστές σε σύμφωνες δομές. Δύο σχετικές παρατηρήσεις, (α) οι συνθετικές εικόνες μοιάζουν πολύ με τις ασύμφωνες αρχικές εικόνες, άρα η μοντελοποίηση του μέτρου των συντελεστών κυματιδίων με το μοντέλο HMT μπορεί να θεωρηθεί αρκετά ικανοποιητική και (β) οι ίδιες οι ασύμφωνες αρχικές εικόνες αντιληπτικά μοιάζουν λίγο με τις σύμφωνες αρχικές εικόνες. Αυτές οι παρατηρήσεις φαίνεται να συντείνουν στην υπόθεση ότι υπάρχουν αρκετά περιθώρια βελτίωσης του μοντέλου HMT στην κατεύθυνση μοντελοποίησης της φάσης των μιγαδικών συντελεστών κυματιδίων, ως προνομιακού πεδίου για την περιγραφή σύμφωνων δομών της εικόνας με μοντέλα οργάνωσης υψηλότερου επιπέδου. Η κατεύθυνση αυτή θα αποτελέσει πεδίο διερεύνησης της εργασίας μας για το άμεσο μέλλον.

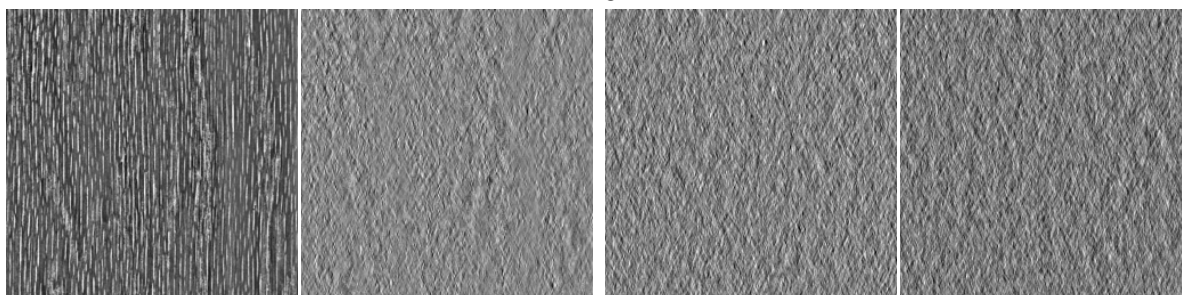
D42-δαντέλα



D46-συρματόπλεγμα



D68-ξύλο



(α)

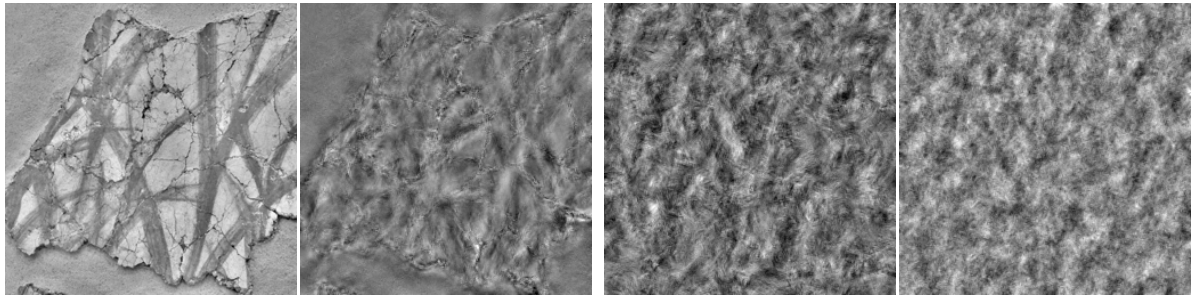
(β)

(γ)

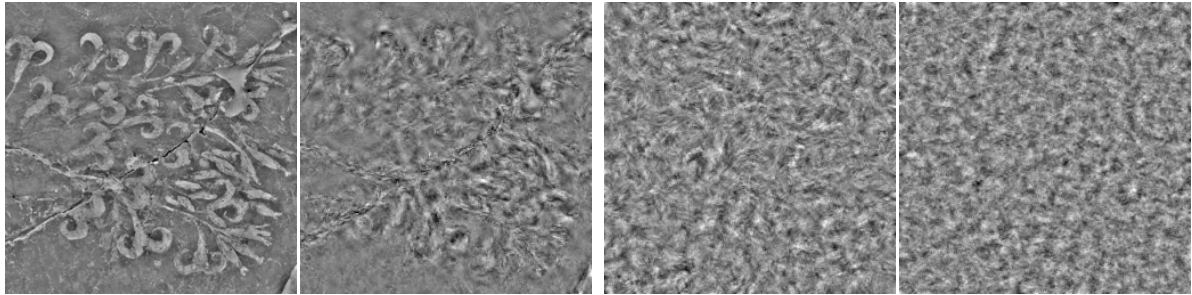
(δ)

Σχήμα 6.5: Πειράματα ανασύνθεσης υφής από εικόνες της συλλογής Brodatz με μιγαδικά κυματίδια και τα HMT/HIT μοντέλα. (α) Υπερβατική αρχική εικόνα, που παράγεται μηδενίζοντας τους συντελεστές κλιμακώσης. (β) Υπερβατική και ασύμφωνη αρχική εικόνα, όπου επιπλέον έχουμε τυχαίοποιησει τη φάση των μιγαδικών συντελεστών κυματιδίων. (γ) Το υπερβατικό δείγμα σύνθεσης με το HMT μοντέλο. (δ) Αντίστοιχο αποτέλεσμα για το HIT μοντέλο. Οι αρχικές εικόνες κλιμακώθηκαν σε μέγεθος 256×256 πριν την επεξεργασία. Σε όλες τις περιπτώσεις έχουμε μείγμα $Q = 4$ κατανομών Gauss ανά επίπεδο· άλλες παράμετροι D42: MCLT, $L = 4$. D46: DTCWT, $L = 5$. D68: DTCWT, $L = 4$.

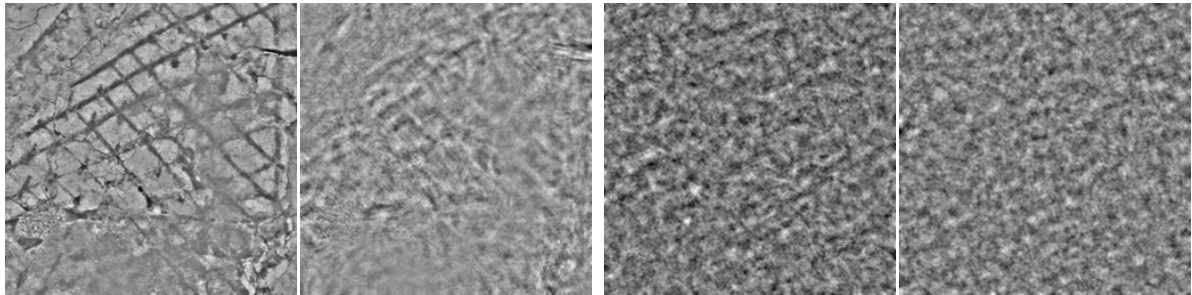
καλάμια - λεπτομέρεια Σχ. 2.5(α)



κρίνα - λεπτομέρεια Σχ. 2.5(γ)



φόρεμα - λεπτομέρεια Σχ. 2.5(γ)



(α)

(β)

(γ)

(δ)

Σχήμα 6.6: Συνέχεια του Σχ. 6.5. Εφαρμογή σε εικόνες τοιχογραφιών από το Ακρωτήρι. Πάλι έχουμε μείγμα $Q = 4$ κατανομών Gauss ανά επίπεδο· άλλες παράμετροι καλάμια: DT-CWT, $L = 5$ · κρίνα: DT-CWT, $L = 4$ · φόρεμα: MCLT, $L = 4$.

6.5 Ενδοσυμπλήρωση με Κρυφό Δένδρο Markov στο Χώρο των Μιγαδικών Κυματιδίων

Για το πρόβλημα της ενδοσυμπλήρωσης που κυρίως μας ενδιαφέρει, χρειάζεται να δουλέψουμε με την υπό συνθήκη κατανομή των συντελεστών κυματιδίων, δεδομένων των τιμών φωτεινότητας της εικόνας στα γνωστά εικονοστοιχεία. Το γεγονός όμως ότι έχουμε παρατηρήσεις σε ορισμένα μόνο σημεία της εικόνας εισάγει αλληλεξαρτήσεις μεταξύ των συντελεστών κυματιδίων και δυσχεραίνει τους πιθανοτικούς χειρισμούς με το HMT μοντέλο. Η κύρια πρωτότυπη ερευνητική μας συνεισφορά σε αυτήν την περιοχή συνίσταται στην ανάπτυξη τεχνικών που υπερβαίνουν αυτές τις δυσκολίες και καθιστούν εφικτό το χειρισμό ελλειπών χωρικών δεδομένων στο πλαίσιο του HMT μοντέλου. Οδηγούμαστε έτσι σε νέες μεθόδους και αλγορίθμους ενδοσυμπλήρωσης με έντονο πιθανοτικό χαρακτήρα: παρουσιάσαμε μια πρώτη εκδοχή αυτής της προσέγγισης στο [PMK08].

Για να γίνουμε πιο συγκεκριμένοι, ας γράψουμε $\mathbf{x} = (\mathbf{x}_o^T, \mathbf{x}_h^T)^T$, διαχωρίζοντας το ορισμένο στα M εικονοστοιχεία της εικόνας διάνυσμα φωτεινότητας \mathbf{x} στο παρατηρήσιμο τμήμα του \mathbf{x}_o ορισμένο σε M_o εικονοστοιχεία και στο κρυφό τμήμα του \mathbf{x}_m ορισμένο στα απωλεσθέντα M_m εικονοστοιχεία ($M_m = M - M_o$). Τότε, υποθέτοντας λευκό θόρυβο μέτρησης, οδηγούμαστε στο μοντέλο παρατήρησης

$$P(\mathbf{x}_o|\mathbf{y}) = N(\mathbf{x}_o; A\mathbf{y}, \Sigma_n), \quad \text{με} \quad A = WF_s, \Sigma_n = \sigma_n^2 I, \quad (6.6)$$

όπου $W = [I_{M_o \times M_o} \quad O_{M_o \times M_m}]$ είναι ένας $M_o \times M$ πίνακας προβολής και F_s συμβολίζει τον αντίστροφο μετασχηματισμό κυματιδίων. Με άλλα λόγια, έχουμε στη διάθεσή μας μόνο έμμεση μέτρηση των συντελεστών κυματιδίων \mathbf{y} μέσω της ακόλουθης διαδικασίας: (1) αντίστροφος μετασχηματισμός κυματιδίων F_s , (2) απόκρυψη ορισμένων εικονοστοιχείων από τη μάσκα ενδοσυμπλήρωσης W , και (3) προσθήκη στο χωρικό πεδίο γκαουσιανού θορύβου με πίνακα συμμεταβλητότητας Σ_n . Ο μη-ορθογώνιος και μη-αντιστρέψιμος χαρακτήρας του μετασχηματισμού $A = WF_s$, ακόμα κι αν κανείς χρησιμοποιήσει ορθογώνια οικογένεια κυματιδίων, περιπλέκει το πρόβλημα της εκτίμησης των συντελεστών κυματιδίων \mathbf{y} από τις μετρήσεις \mathbf{x}_o .

Στο πιθανοτικό πλαίσιο που ακολουθούμε με χρήση του μοντέλου HMT, ενδιαφερόμαστε να εκτιμήσουμε την υπό συνθήκη κατανομή των συντελεστών κυματιδίων \mathbf{y} , δεδομένων των παρατηρήσεων \mathbf{x}_o :

$$P(\mathbf{y}|\mathbf{x}_o, \boldsymbol{\theta}) \propto P(\mathbf{x}_o|\mathbf{y})P(\mathbf{y}) = P(\mathbf{x}_o|\mathbf{y}) \sum_{\mathbf{q} \in \mathcal{Q}} P(\mathbf{y}|\mathbf{q})P(\mathbf{q}), \quad (6.7)$$

όπου \mathcal{Q} είναι το σύνολο όλων των δυνατών κρυφών μονοπατιών \mathbf{q} . Δυστυχώς, καθότι οι μετρήσεις \mathbf{x}_o εισάγουν αλληλεξαρτήσεις μεταξύ των συντελεστών κυματιδίων \mathbf{y} , η Εξ. (6.7) δεν μπορεί να αποδομηθεί ως γινόμενο ανεξάρτητων όρων ανά κόμβο του δέντρου κυματιδίων όπως γίνεται στην περίπτωση της αποθορυβοποίησης της Ενότητας 6.3. Παρουσιάζουμε στη συνέχεια δύο νέες προσεγγίσεις, μία ντετερμινιστική και μία στοχαστική, για την αποδοτική εκτίμηση των συντελεστών κυματιδίων \mathbf{y} . Και οι δύο βασίζονται στο γεγονός ότι η γνώση του \mathbf{q} διευκολύνει την εκτίμηση του \mathbf{y} , και αντίστροφα. Ειδικότερα, δεδομένου του κρυφού μονοπατιού \mathbf{q} , οι συντελεστές κυματιδίων \mathbf{y} ακολουθούν κατανομή Gauss,

$$P(\mathbf{y}|\mathbf{q}, \mathbf{x}_o) \propto P(\mathbf{x}_o|\mathbf{y})P(\mathbf{y}|\mathbf{q}) = N(\mathbf{y}; \boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}}, \boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{q}}), \quad (6.8)$$

με $\boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}} = \boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{q}}(\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\boldsymbol{\mu}_{\mathbf{q}} + A^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_o)$ και $\boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{q}}^{-1} = \boldsymbol{\Sigma}_{\mathbf{q}}^{-1} + A^T\boldsymbol{\Sigma}_n^{-1}A$. Για να υπολογίσουμε αποδοτικά το $\boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}}$, λύνουμε το γραμμικό σύστημα

$$\boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{q}}^{-1}\boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}} = \boldsymbol{\Sigma}_{\mathbf{q}}^{-1}\boldsymbol{\mu}_{\mathbf{q}} + A^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_o \quad (6.9)$$

επαναληπτικά με τη μέθοδο των *συζυγών κλίσεων* (conjugate gradients – CG), χρησιμοποιώντας το $\boldsymbol{\Sigma}_{\mathbf{q}}^{-1}$ ως διαγώνιο *προορυθμιστή* (preconditioner), όπως περιγράφουμε αναλυτικά στην Ενót. 6.Α'. Αυτό είναι εφικτό επειδή ο πίνακας συστήματος $\boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{q}}^{-1}$ είναι θετικά ορισμένος. Με τη μέθοδο συζυγών κλίσεων αποφεύγουμε τον άμεσο σχηματισμό του πίνακα συστήματος, γιατί απαιτείται μόνο η επανειλημμένη αποτίμηση του γινομένου πίνακα-διανύσματος $\boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{q}}^{-1}\mathbf{y}$: αυτό περιλαμβάνει υπολογισμό των $F_s\mathbf{y}$ και $F_s^T\mathbf{x}$, το οποίο μπορεί να γίνει αποδοτικά με τη βοήθεια του αντίστροφου (χρησιμοποιώντας το πρωτεύον σύνολο φίλτρων) και του ευθύ (χρησιμοποιώντας το δυϊκό σύνολο φίλτρων), αντίστοιχα, διακριτού μετασχηματισμού κυματιδίων στη γενική περίπτωση που χρησιμοποιούμε διορθογώνιους μετασχηματισμούς· βλέπε σχετικά τις Εξς. (5.8) και (5.11) του Κεφαλαίου 5.

6.5.1 Ντετερμινιστική Επίλυση με Εναλλασσόμενη Βελτιστοποίηση

Αφού δεν είναι εφικτό να βελτιστοποιήσουμε από κοινού τις κρυφές καταστάσεις \mathbf{q} και τους συντελεστές κυματιδίων \mathbf{y} , μια ντετερμινιστική προσέγγιση για εύρεση της λύσης είναι να βελτιστοποιήσουμε εναλλακτικά τις κρυφές καταστάσεις, δεδομένων των συντελεστών κυματιδίων, και ύστερα τους συντελεστές κυματιδίων, δεδομένων των κρυφών καταστάσεων, ωστόσο να πετύχουμε σύγκλιση σε μια *αυτοσυμβατή* (self-consistent) λύση.

Πιο συγκεκριμένα, επαναλαμβάνουμε ως τη σύγκλιση τα επόμενα δύο βήματα:

1. Δεδομένης της εκτίμησης \mathbf{y}_{i-1} για τους συντελεστές κυματιδίων, υπολογίζουμε το πιο πιθανό διάνυσμα κρυφών καταστάσεων $\mathbf{q}_i = \operatorname{argmax} P(\mathbf{q}|\mathbf{y}_{i-1})$ με χρήση του αλγορίθμου Viterbi που σκιαγραφήσαμε στην Ενót. 6.2.
2. Δεδομένης της εκτίμησης \mathbf{q}_i για το διάνυσμα κρυφών καταστάσεων, θέτουμε ως \mathbf{y}_i το διάνυσμα $\boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}_i}$ που μεγιστοποιεί τη γκαουσιανή $P(\mathbf{y}|\mathbf{q}_i, \mathbf{x}_o)$ της Εξ. (6.8).

Αυτή η μέθοδος είναι εγγυημένο ότι αυξάνει μονοτονικά την από κοινού πιθανοφάνεια του μοντέλου, αφού $P(\mathbf{y}, \mathbf{q}|\mathbf{x}_o) = P(\mathbf{y}|\mathbf{q}, \mathbf{x}_o)P(\mathbf{q}|\mathbf{x}_o) = P(\mathbf{q}|\mathbf{y}, \mathbf{x}_o)P(\mathbf{y}|\mathbf{x}_o)$, οπότε

$$\begin{aligned} P(\mathbf{y}_i, \mathbf{q}_i|\mathbf{x}_o) &= P(\mathbf{y}_i|\mathbf{q}_i, \mathbf{x}_o)P(\mathbf{q}_i|\mathbf{x}_o) \geq P(\mathbf{y}_{i-1}|\mathbf{q}_i, \mathbf{x}_o)P(\mathbf{q}_i|\mathbf{x}_o) \\ &= P(\mathbf{q}_i|\mathbf{y}_{i-1}, \mathbf{x}_o)P(\mathbf{y}_{i-1}|\mathbf{x}_o) \geq P(\mathbf{q}_{i-1}|\mathbf{y}_{i-1}, \mathbf{x}_o)P(\mathbf{y}_{i-1}|\mathbf{x}_o) = P(\mathbf{y}_{i-1}, \mathbf{q}_{i-1}|\mathbf{x}_o). \end{aligned} \quad (6.10)$$

Η τελική σύγκλιση είναι σε ένα τοπικό μέγιστο της $P(\mathbf{y}, \mathbf{q}|\mathbf{x}_o)$.

Παρόμοια στρατηγική ακολουθείται από τον αλγόριθμο *Επαναληπτικών υπό Συνθήκη Μεγίστων* (Iterated Conditional Modes – ICM) για υπολογισμό της πιθανότερης κατάστασης σε τυχαία πεδία Markov [Bes86]. Βέβαια εκεί ανανεώνουμε την κρυφή κατάσταση ξεχωριστά σε καθένα εικονοστοιχείο της εικόνας, ενώ εδώ ανανεώνουμε το σύνολο των κρυφών καταστάσεων από κοινού. Συνεπώς ο προτεινόμενος μπλοκ αλγόριθμος είναι λιγότερο επιρρεπής στην εύρεση φτωχών τοπικών ελαχίστων σε σχέση με τον κλασικό μεμονωμένο ICM αλγόριθμο. Επίσης, η προτεινόμενη τεχνική σχετίζεται με αλγορίθμους τύπου EM που έχουν πρόσφατα προταθεί για το πρόβλημα της ενδοσυμπλήρωσης εικόνων [LM05, FS05].

6.5.2 Στοχαστική Επίλυση με Δειγματοληψία Gibbs

Αν και η ντετερμινιστική τεχνική της Ενότητας 6.5.1 συνήθως δίνει ικανοποιητικά πρακτικά αποτελέσματα ενδοσυμπλήρωσης, έχει τον εγγενή περιορισμό ότι μπορεί στην καλύτερη περίπτωση να αποκαλύψει το σημείο μεγίστου της από κοινού κατανομής $P(\mathbf{y}, \mathbf{q}|\mathbf{x}_o)$. Η πλήρης πιθανοτική περιγραφή της λύσης όμως προϋποθέτει ότι κανείς έχει επαρκή εικόνα όλης της κατανομής [GCSR03, Bis06]. Αφού τούτο δεν είναι δυνατόν με αναλυτικά μέσα για το προς επίλυση πρόβλημα, καταφεύγουμε σε στοχαστικές τεχνικές Μόντε-Κάρλο δειγματοληψίας και συγκεκριμένα στο *δειγματολήπτη* (Gibbs sampler) [GG84, TW87, GS90], που ανήκει στην κατηγορία τεχνικών δειγματοληψίας *Μόντε-Κάρλο Markov αλυσίδας* (Markov-chain Monte-Carlo - MCMC) [GRS95].

Πιο αναλυτικά, στόχος μας είναι η δειγματοληψία της από κοινού κατανομής συντελεστών κυματιδίων και κρυφών καταστάσεων

$$P(\mathbf{y}, \mathbf{q}|\mathbf{x}_o) \propto P(\mathbf{x}_o|\mathbf{y}, \mathbf{q})P(\mathbf{y}, \mathbf{q}) = P(\mathbf{x}_o|\mathbf{y})P(\mathbf{y}|\mathbf{q})P(\mathbf{q}). \quad (6.11)$$

Η απευθείας δειγματοληψία από την Εξ. (6.11) δεν είναι εφικτή. Παρόλα αυτά, όπως εξηγούμε με λεπτομέρεια στη συνέχεια, μπορούμε να σχεδιάσουμε έναν μπλοκ Gibbs δειγματολήπτη που εναλλάσσεται μεταξύ δειγματοληψίας από τις δύο υπό συνθήκη κατανομές $\mathbf{y}|\mathbf{q}, \mathbf{x}_o$ και $\mathbf{q}|\mathbf{y}, \mathbf{x}_o$, καταλήγοντας σε μια πρακτική MCMC τεχνική που διερευνά ικανοποιητικά την κατανομή της Εξ. (6.11).

Δειγματοληψία Συντελεστών Κυματιδίων Δεδομένης της Κρυφής Κατάστασης: $\mathbf{y}|\mathbf{q}, \mathbf{x}_o$

Με δοσμένη την κρυφή κατάσταση \mathbf{q} , η παρακάτω διαδικασία αρκεί για να λάβουμε δείγμα \mathbf{y}_s από την γκαουσιανή $P(\mathbf{y}|\mathbf{q}, \mathbf{x}_o)$ της Εξ. (6.8):

1. Λήψη δείγματος από την πρότερη κατανομή $\mathbf{y}_{u,s} \sim N(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q)$.
2. Λήψη θορυβώδους δείγματος των παρατηρήσεων $\mathbf{x}_{o,s} \sim N(\mathbf{x}_o, \boldsymbol{\Sigma}_n)$.
3. Επίλυση της $\mathbf{y}_s = \boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{q}}(\boldsymbol{\Sigma}_q^{-1}\mathbf{y}_{u,s} + \mathbf{A}^T\boldsymbol{\Sigma}_n^{-1}\mathbf{x}_{o,s})$.

Πράγματι, η \mathbf{y}_s είναι γκαουσιανή μεταβλητή, ως γραμμική συνάρτηση των γκαουσιανών $\mathbf{y}_{u,s}$ και $\mathbf{x}_{o,s}$, και μπορεί απευθείας να ελεγχθεί ότι έχει το σωστό μέσο $E\{\mathbf{y}_s\} = \boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}}$ και τον σωστό πίνακα συμμεταβλητότητας $E\{(\mathbf{y}_s - \boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}})(\mathbf{y}_s - \boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}})^T\} = \boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{q}}$. Αξίζει να σημειωθεί ότι ο πίνακας κατά την επίλυση του συστήματος στο 3ο βήμα του παραπάνω αλγορίθμου ταυτίζεται με τον πίνακα συστήματος στην Εξ. (6.9), οπότε χρησιμοποιούμε ακριβώς τις ίδιες επαναληπτικές τεχνικές για την επίλυσή του. Καθότι το κόστος λήψης των δειγμάτων $\mathbf{y}_{u,s}$ και $\mathbf{x}_{o,s}$ στα πρώτα δύο βήματα του αλγορίθμου είναι μικρό σε σχέση με το κόστος επίλυσης του γραμμικού συστήματος στο τρίτο βήμα του αλγορίθμου, βλέπουμε ότι η εύρεση του μεγίστου και η λήψη δείγματος από την δεσμευμένη γκαουσιανή κατανομή $P(\mathbf{y}|\mathbf{q}, \mathbf{x}_o)$ έχουν στην πράξη ανάλογο κόστος για το πρόβλημά μας.

Δειγματοληψία Κρυφής Κατάστασης Δεδομένων των Συντελεστών Κυματιδίων: $q|y, x_o$

Με δεδομένη μια εκτίμηση για τους συντελεστές κυματιδίων, η κατανομή των κρυφών μεταβλητών είναι ανεξάρτητη των παρατηρήσεων x_o :

$$P(q|y, x_o) = P(q|y). \quad (6.12)$$

Σε αντιδιαστολή με τη Viterbi αποκωδικοποίηση, όπου στόχος είναι να βρεθεί πιο πιθανό κρυφό μονοπάτι, στόχος μας εδώ είναι να ληφθεί δείγμα του μονοπατιού q δεδομένου του διανύσματος μιγαδικών κυματιδίων y , το οποίο να ακολουθεί τη δεσμευμένη κατανομή Εξ. (6.12). Η συνηθισμένη προσέγγιση για δειγματοληψία από την Εξ. (6.12), π.χ. σε τυχαία πεδία Markov [GG84] ή σε εφαρμογές ανάλυσης ακουστικών σημάτων [FTDG08], είναι να ανανεώνεται ξεχωριστά ο κάθε κρυφός κόμβος του δέντρου, δηλαδή ακολουθιακά να λαμβάνονται δείγματα από την $P(q_t|y, q_{-t})$, όπου με q_{-t} συμβολίζουμε το σύνολο των κρυφών μεταβλητών πλην αυτής που αντιστοιχεί στον t κόμβο. Όμως αυτή η μεμονωμένη ανανέωση των κόμβων επιβραδύνει τη μίξη της Markov αλυσίδας στο Μόντε-Κάρλο αλγόριθμο [LWK94]. Μια εναλλακτική προσέγγιση είναι να δειγματοληψούμε μονομιάς ως μπλοκ ολόκληρο το μονοπάτι q . Χάρη στη δενδρική δομή του μοντέλου μας, τούτο μπορεί να επιτευχθεί με έναν αποδοτικό αναδρομικό ως προς την κλίμακα αλγόριθμο, που συνίσταται στα εξής:

1. Άνω πέρασμα στο δέντρο, ίδιο με αυτό του άνω-κάτω αλγορίθμου για συμπερασματολογία από την Ενót. 6.2, κατά το οποίο υπολογίζεται η ποσότητα $\beta_t(i) = P(q_t = i|y_{T_t})$ (y_{T_t} , όπου y_{T_t} συμβολίζει της τιμές των συντελεστών κυματιδίων στο υποδέντρο με ρίζα τον t κόμβο [CNB98]).
2. Στοχαστικό κάτω πέρασμα στο δέντρο, παρόμοιο με αυτό του Viterbi αλγορίθμου για αποκωδικοποίηση από την Ενót. 6.2, κατά το οποίο: (α) αναθέτουμε την κρυφή μεταβλητή της ρίζας με πιθανότητα $P(q_0 = i|y) \propto \pi_i \beta_0(i)$ και (β) αναδρομικά, δεδομένης της ήδη ανατεθειμένης κατάστασης του γονιού, αναθέτουμε την κρυφή μεταβλητή του t κόμβου με πιθανότητα $P(q_t = i|q_{p_t} = j, y) \propto A_{ji} \beta_t(i)$.

Αν και ο παραπάνω αλγόριθμος μπορεί να θεωρηθεί ως ειδική περίπτωση του γενικού αλγορίθμου διάδοσης πίστης [Pea88], συναντάται ρητά αρκετά αργότερα, και συγκεκριμένα για την απλούστερη περίπτωση των HMM στα [Chi96, Sco02] και για δενδρικούς γράφους στα [LPH00, PCL00]. Ανεξάρτητα από την εργασία μας [PMK08], ο ίδιος αλγόριθμος για μπλοκ δειγματοληψία του διακριτού κρυφού μονοπατιού q χρησιμοποιήθηκε στο πλαίσιο του HMT μοντέλου κυματιδίων και στο [KSJ07b].

Σύννοψη Gibbs Δειγμάτων

Εναλλάσσοντας τα βήματα δειγματοληψίας που περιγράφονται στις προηγούμενες παραγράφους, και πιθανόν απορρίπτοντας τα πρώτα λίγα δείγματα 'ζεστάματος' ('burn-in'), καταλήγουμε με μια ακολουθία $\{\{q_1, y_1\}, \{q_2, y_2\}, \dots, \{q_S, y_S\}\}$ δειγμάτων Gibbs που στο όριο προέρχονται από την από κοινού κατανομή (6.11). Τυπικά επιθυμούμε να συνοψίσουμε την ακολουθία δειγμάτων σε μια εκτίμηση για τους συντελεστές κυματιδίων \hat{y} που αντιστοιχεί στην τελική εκτίμηση-εικόνα για το αποτέλεσμα ενδοσυμπλήρωσης $\hat{x} = F_s \hat{y}$.

Δύο πιθανά κριτήρια για επιλογή της εκτιμήτριας \hat{y} είναι:

1. Ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος MMSE, που είναι το πιο συχνά χρησιμοποιούμενο κριτήριο. Το αποτέλεσμα είναι ικανοποιητικό στις επίπεδες περιοχές της εικόνας, αλλά σε κάποιες περιπτώσεις υπερβολικά ομαλό στις περιοχές με έντονη υφή.
2. Η δεύτερη λύση προτείνει ως αποτέλεσμα ενδοσυμπλήρωσης ένα από τα τελευταία δείγματα \mathbf{y}_s της ακολουθίας Gibbs, πιθανόν με κάποια μικρή επεξεργασία. Αυτή η λύση οδηγεί σε μεγαλύτερο τετραγωνικό σφάλμα, αλλά συχνά είναι πιο ικανοποιητική αντιληπτικά, τουλάχιστον σε περιοχές με έντονη υφή, γιατί οδηγεί σε εικόνες με πιο τυπική και ζωντανή εμφάνιση. Δες και τα [ÓRF94, Kok98] για περαιτέρω συζήτηση των διαφορών μεταξύ εκτιμητριών MMSE και δειγματοληψίας σε εφαρμογές ενδοσυμπλήρωσης ακουστικών σημάτων και εικόνων, αντίστοιχα.

Αξίζει να σημειώσουμε ότι στην περίπτωση που έχουμε πολύ μεγάλα κενά στην εικόνα, η πρώτη MMSE εκτιμήτρια θα τείνει να μηδενίσει τους κυματιδιακούς συντελεστές λεπτομέρειας, ενώ η δεύτερη εκτιμήτρια θα τείνει να προτείνει για τη συμπλήρωση των κενών δείγματα από την πρότερη κατανομή σαν κι αυτά που συναντήσαμε στα πειράματα σύνθεσης υφής της Ενότητας 6.4.

Στην περίπτωση που επιλέξουμε η λύση μας να ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα, μπορεί κανείς να χρησιμοποιήσει το συνήθη MMSE εκτιμητή για Μόντε-Κάρλο προσομοιώσεις, που είναι ο δειγματικός μέσος $\hat{\mathbf{y}}_S = \frac{1}{S} \sum_{s=1}^S \mathbf{y}_s$. Εδώ όμως προτιμούμε τον *Rao-Blackwell (RB)* εκτιμητή $\hat{\mathbf{y}}_{RB} = \frac{1}{S} \sum_{s=1}^S E[\mathbf{y}|\mathbf{q}_s, \mathbf{x}_o]$ που έχει σημαντικά ελαττωμένη μεταβλητότητα [LWK94, RC05]. Διαισθητικά, τούτο επιτυγχάνεται επειδή ο RB εκτιμητής για κάθε δείγμα από το κρυφό μονοπάτι \mathbf{q}_s κάνει χρήση της ακριβούς αναλυτικής εκτίμησης $\boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}_s} = E[\mathbf{y}|\mathbf{q}_s, \mathbf{x}_o]$ αντί να βασίζεται στο μοναδικό δείγμα $\mathbf{y}_s \sim \mathbf{y}|\mathbf{q}_s, \mathbf{x}_o$. Το τίμημα είναι ότι σε κάθε βήμα του αλγορίθμου Gibbs καλούμαστε να επιλύσουμε ένα ακόμα γραμμικό σύστημα για τον υπολογισμό του $\boldsymbol{\mu}_{\mathbf{y}|\mathbf{q}_s}$, σχεδόν διπλασιάζοντας έτσι το υπολογιστικό κόστος. Όμως, χάρη στη μειωμένη μεταβλητότητα του RB εκτιμητή και στην ταχύτατη μίξη της Μόντε-Κάρλο Markov αλυσίδας λόγω του μπλοκ χαρακτήρα του δειγματοληπτή Gibbs, μπορούμε να λάβουμε ικανοποιητικά αποτελέσματα ενδοσυμπλήρωσης εικόνας ύστερα από πολύ λίγα βήματα δειγματοληψίας S , που συνήθως αντισταθμίζει το επιπλέον υπολογιστικό κόστος.

6.5.3 Αρχικοποίηση με μεθόδους Μερικών Διαφορικών Εξισώσεων

Τόσο ο ντετερμινιστικός αλγόριθμος εναλλασσόμενης βελτιστοποίησης, όσο και ο στοχαστικός αλγόριθμος Gibbs δειγματοληψίας, είναι επαναληπτικού χαρακτήρα και για αυτό επωφελούνται μιας καλής αρχικοποίησης. Στην υλοποίησή μας χρησιμοποιούμε ως τέτοια αρχικοποίηση τη λύση της ΜΔΕ Laplace $\nabla^2 u = 0$ στα κενά της εικόνας, με Dirichlet συνοριακές συνθήκες στο όριο με τις γνωστές περιοχές. Αυτό μπορεί να θεωρηθεί ως η απλούστερη και υπολογιστικά αποδοτικότερη πρόταση ενδοσυμπλήρωσης βασισμένη στα μοντέλα ΜΔΕ του Κεφαλαίου 3, η οποία έχει αποδειχθεί επαρκής στα πειράματά μας. Πιο προηγμένες τεχνικές ενδοσυμπλήρωσης με ΜΔΕ, όπως αυτές της Ενότητας 3.2 μπορούν εναλλακτικά να χρησιμοποιηθούν, αν και με μεγαλύτερο υπολογιστικό κόστος.

6.6 Επεκτάσεις του Βασικού Μοντέλου

6.6.1 Αποκατάσταση Συνιστωσών Καρτούν και Υφής με Χρήση Πολλαπλών Λεξικών

Όπως έχουμε συζητήσει στο Κεφάλαιο 5, στις εικόνες του φυσικού κόσμου συνυπάρχουν ετερογενείς περιοχές με διαφορετικές δομικές ιδιότητες. Είναι δύσκολο να σχεδιάσει κανείς μια μοναδική αναπαράσταση αρκετά εκφραστική ώστε να περιγράψει με αραιό τρόπο το εύρος των μορφών που συναντώνται σε εικόνες του φυσικού κόσμου. Από τις αναπαραστάσεις που συζητήσαμε στο Κεφάλαιο 5, μετασχηματισμοί κυματιδίων χώρου-κλίμακας όπως ο DT-CWT ενδείκνυνται για την περιγραφή τμηματικά ομαλών εικόνων, ενώ τοπικά αρμονικές αναπαραστάσεις χώρου-συχνότητας όπως ο MCLT ενδείκνυνται για την περιγραφή ταλαντούμενων κυματομορφών σε εικόνες υφής.

Ο πιο απλός τρόπος για να αξιοποιήσει κανείς τα πλεονεκτήματα διαφορετικών αναπαραστάσεων είναι να επιλέξει το καταλληλότερο από ένα σύνολο διαθέσιμων λεξικών, ανάλογα με το περιεχόμενο της προς επεξεργασία εικόνας. Μια πιο σύνθετη προσέγγιση συνίσταται στον προσθετικό διαχωρισμό του περιεχομένου της εικόνας σε πολλαπλά κανάλια [Eua07], π.χ. διαχωρισμός σε συνιστώσες καρτούν και υφής, και κατόπιν ανάλυση του κάθε καναλιού με την οικεία του αναπαράσταση, όπως γίνεται για παράδειγμα στα [ESQD05, FTDG08].

6.6.2 Σύζευξη Μοντέλων Κυματιδίων και Τεχνικών ΜΔΕ στο Χωρικό Πεδίο

Οι μέθοδοι μερικών διαφορικών εξισώσεων για ενδοσυμπλήρωση που έχουμε παρουσιάσει στο Κεφάλαιο 3 βασίζονται σε γεωμετρική περιγραφή στο χωρικό πεδίο, ενώ οι μέθοδοι του παρόντος Κεφαλαίου βασίζονται σε αναπαράσταση της εικόνας στο πεδίο των κυματιδίων. Είναι συχνά επιθυμητό να συνδυάσει κανείς τα πλεονεκτήματα των δύο προσεγγίσεων.

Ίσως ο απλούστερος τρόπος για να συζεύξει κανείς τις δύο κατηγορίες τεχνικών είναι να θεωρήσει ένα ενεργειακό συναρτησιακό που περιλαμβάνει όρους προερχόμενους τόσο από μεταβολικές μεθόδους ΜΔΕ όσο και από την πιθανοτική κυματιδιακή περιγραφή, όπως προτείνεται π.χ. στο [ESQD05]. Παράδειγμα κατάλληλου συναρτησιακού ενέργειας είναι το

$$E(\mathbf{y}) = c_1 \text{TV}(F_s \mathbf{y}) + c_2 \log P(\mathbf{y} | \mathbf{x}_o), \quad (6.13)$$

όπου σταθμίζονται αφενός ο γεωμετρικός όρος ενέργειας που αντιστοιχεί στην *ολική μεταβολή* (total variation) της εικόνας στο χωρικό πεδίο [ROF92], ορισμένος ως $\text{TV}(u) = \int \|\nabla u\|$, και αφετέρου ο πιθανοτικός όρος ενέργειας προερχόμενος από το κυματιδιακό HMT μοντέλο, Εξ. (6.7). Με ρύθμιση του σχετικού βάρους των δύο όρων, μπορεί κανείς να μεταβεί από αμιγώς κυματιδιακά μοντέλα ($c_1 \ll c_2$) σε αμιγώς χωρικά μοντέλα ($c_1 \gg c_2$). Γνωστό παράδειγμα της δεύτερης κατηγορίας είναι η μέθοδος ενδοσυμπλήρωσης του [CSZ06], όπου ανακτώνται ελλείποντες κυματιδιακοί συντελεστές με κανονικοποίηση ολικής μεταβολής στο χωρικό πεδίο. Στην πράξη, η ανάκτηση της εικόνας με το υβριδικό μοντέλο της Εξ. (6.13) επιτυγχάνεται προσεγγιστικά με παρεμβολή μικρού αριθμού βημάτων κανονικοποίησης ολικής μεταβολής στους αλγορίθμους ενδοσυμπλήρωσης της Ενότητας 6.5. Πάντως, μειονέκτημα του υβριδικού χωρικού/κυματιδιακού

μοντέλου είναι ότι παύει να διατηρεί τον αναγεννητικό χαρακτήρα και την ξεκάθαρη πιθανοτική ερμηνεία του αμιγώς κυματιδιακού HMT μοντέλου.

6.6.3 Τεχνικές για Ενσωμάτωση Πληροφορίας Μιγαδικής Φάσης Κατά την Αποκατάσταση

Τα πρότερα μοντέλα που έχουμε ως τώρα συζητήσει επικεντρώνονται στην περιγραφή του μέτρου $r_t = \sqrt{y_{t,r}^2 + y_{t,i}^2}$ των μιγαδικών συντελεστών κυματιδίων. Για παράδειγμα, στο μοντέλο της Ενότητας 6.2, Εξ. (6.3), η φάση $\phi_t = \angle y_{t,r} + jy_{t,i}$ των μιγαδικών συντελεστών κυματιδίων περιγράφεται από τη μη πληροφοριακή ομοιόμορφη κατανομή στο διάστημα $[0, 2\pi]$. Αυτό έχει ως αποτέλεσμα η φάση στο αποτέλεσμα ενδοσυμπλήρωσης να καθορίζεται αποκλειστικά από τα ελλιπή δεδομένα μέτρησης και να μην κανονικοποιείται επαρκώς.

Όπως έχουμε συζητήσει στην Ενότητα 5.3.2, εάν γνωρίζουμε επακριβώς από ένα *μαντείο* (oracle) τη φάση μιγαδικών κυματιδιακών μετασχηματισμών όπως ο DT-CWT ή MCLT, μπορούμε να ανακτήσουμε την αρχική εικόνα στο χωρικό πεδίο. Ας υποθέσουμε προς στιγμήν ότι η μιγαδική φάση $\phi_t = \angle y_{t,r} + jy_{t,i}$ των συντελεστών κυματιδίων είναι κατά προσέγγιση γνωστή. Θα εξετάσουμε πώς αυτή η πληροφορία μπορεί να ενσωματωθεί στο υπό συζήτηση μοντέλο ενδοσυμπλήρωσης. Εάν αποφασίσουμε να εισάγουμε τη γνώση φάσης ως *άκαμπτο περιορισμό* (hard constraint) στο πρόβλημά μας, είναι χρήσιμο να επαναπαραμετροποιήσουμε το πρόβλημα βελτιστοποίησης σε πολικές συντεταγμένες, όπου εμφανίζεται πλέον ως άγνωστο το μέτρο $r_t = \sqrt{y_{t,r}^2 + y_{t,i}^2}$ των μιγαδικών συντελεστών κυματιδίων, που κατόπιν καθορίζει το πραγματικό και φανταστικό μέρος των μιγαδικών συντελεστών ως $y_{t,r} = r_t \cos(\phi_t)$ και $y_{t,i} = r_t \sin(\phi_t)$. Συγκεκριμένα οδηγούμαστε στο ακόλουθο πρόβλημα τετραγωνικού προγραμματισμού με περιορισμούς ανισότητας

$$\text{minimize } f(\mathbf{r}) = \frac{1}{2} \mathbf{r}^T B \mathbf{r} - \mathbf{b}^T \mathbf{r} \quad (6.14a)$$

$$\text{subject to } r_t \geq 0, \quad 1 \leq t \leq N \quad (6.14b)$$

Όπως έχουμε συζητήσει στην Ενότητα 6.6.3, ο πίνακας B δίνεται από τη σχέση $B = \begin{pmatrix} C & S^T \\ C^T & S \end{pmatrix} (\Sigma_q^{-1} + A^T \Sigma_n^{-1} A) \begin{pmatrix} C \\ S \end{pmatrix}$, όπου $C = \text{diag}(\cos(\phi_t))$ και $S = \text{diag}(\sin(\phi_t))$. Ο παραπάνω περιορισμός μη αρνητικότητας για το μέτρο περιγράφει ένα κυρτό επιτρεπτό σύνολο, ενώ και η αντικειμενική συνάρτηση $f(\mathbf{y})$ είναι κυρτή αφού ο πίνακας B είναι θετικά ορισμένος. Συνεπώς οι Εξς. (6.14) ορίζουν ένα *κυρτό* (convex) πρόβλημα βελτιστοποίησης που επιδέχεται μοναδική λύση [BV04]. Το γενικό πλαίσιο τεχνικών κυρτής ανάλυσης και κυρτού προγραμματισμού έχει αποδειχθεί ιδιαίτερα κατάλληλο για προβλήματα στην επεξεργασία σήματος και στην όραση υπολογιστών, π.χ. [Sch94, Mar95, EDGL07, KT07]. Συζητάμε στο Παράρτημα 6.Β' τεχνικές βελτιστοποίησης για την αριθμητική επίλυση του προβλήματος (6.14).

6.7 Πειράματα Ενδοσυμπλήρωσης

Έχουμε εφαρμόσει τις προτεινόμενες τεχνικές ενδοσυμπλήρωσης στο πρόβλημα της ψηφιακής συμπλήρωσης ελλιπόντων τμημάτων στις προϊστορικές τοιχογραφίες του Ακρωτηρίου Θήρας - δεξ Κεφάλαιο 2. Στα Σχ. 6.7 και Σχ. 6.8 δείχνουμε το αποτέλεσμα

εφαρμογής της προτεινόμενης τεχνικής για ενδοσυμπλήρωση με μιγαδικά κυματίδια DT-CWT και το μοντέλο HMT σε τμήματα από τις τοιχογραφίες 'Γυναίκες' και 'Ακολουθία Ανδρών' του Ακρωτηρίου. Η περιοχή ενδοσυμπλήρωσης έχει επισημειωθεί με το χέρι και δίνεται ως είσοδος στον αλγόριθμο.

Παρά τη μεγάλη έκταση των κενών, τα αποτελέσματα είναι αρκετά ικανοποιητικά, ειδικά στις ομαλές περιοχές της εικόνας. Παρόλα αυτά περαιτέρω βελτιώσεις είναι επιθυμητές, όπως π.χ. πιο ξεκάθαρη συνέχιση των ακμών εντός μεγάλων κενών.

Για να μελετήσουμε περαιτέρω την προτεινόμενη προσέγγιση, έχουμε πραγματοποιήσει πειράματα ενδοσυμπλήρωσης σε τεχνητά επικαλυμμένες εικόνες για τις οποίες συνεπώς μπορούμε να αξιολογήσουμε ποσοτικά το αποτέλεσμα. Στο Σχ. 6.9 συγκρίνουμε το αποτέλεσμα της εξεταζόμενης μεθόδου ενδοσυμπλήρωσης με το αντίστοιχο αποτέλεσμα της ΜΔΕ μεθόδου Laplace πάνω στην εικόνα 'Lena'. Το εξεταζόμενο HMT μοντέλο σε συνδυασμό με τον DT-CWT έχει επιλυθεί και με τη ντετερμινιστική και με τη στοχαστική τεχνική. Τα αποτελέσματα έχουν υπολογιστεί με 10 βήματα εναλλασσόμενης βελτιστοποίησης/δειγματοληψίας, αντίστοιχα, με χρήση 3 κρυφών καταστάσεων στο δέντρο, 6 επιπέδων κυματιδιακής ανάλυσης, και με 50 επαναλήψεις της μεθόδου συζυγών κλίσεων για επίλυση των προκυπτόντων γραμμικών συστημάτων. Η εκπαίδευση του μοντέλου HMT έγινε με τον EM αλγόριθμο [CNB98] πάνω στην εικόνα από τη Laplace αποκατάσταση. Παρατηρούμε ότι το προτεινόμενο μοντέλο HMT παράγει ποιοτικά καλύτερα και λιγότερο θολά αποτελέσματα σε σχέση με τη μέθοδο Laplace, ενώ και οι αντίστοιχες ποσοτικές τιμές σηματοθορυβικού λόγου PSNR (απόκλιση μεταξύ αποτελέσματος ενδοσυμπλήρωσης και γνωστής εικόνας υπολογισμένη στα κενά).

Πάντως, ανεξάρτητα ποιό από τα δύο εναλλακτικά κριτήρια ενδοσυμπλήρωσης επιλέξουμε, η απουσία μοντελοποίησης της φάσης των συντελεστών κυματιδίων στο μοντέλο HMT οδηγεί σε αντίστοιχα προβλήματα με αυτά που έχουμε ήδη συζητήσει στα πειράματα σύνθεσης υφής. Συγκεκριμένα, το μοντέλο δεν συμπληρώνει ικανοποιητικά *σύμφωνες* (coherent) δομές της εικόνας, με τα μη-ικανοποιητικά αποτελέσματα να είναι πιο ορατά στα μεγάλα κενά.

6.8 Σύνοψη και Προοπτικές Βελτίωσης του Δενδρικού Μοντέλου Κυματιδίων

Στο Κεφάλαιο αυτό δείξαμε πώς μπορούμε να αναπαραστήσουμε εικόνες με το δενδρικό κυματιδιακό HMT μοντέλο, το οποίο αναπαράγει βασικές κανονικότητες φυσικών εικόνων. Ενώ σε προηγούμενες εργασίες το HMT μοντέλο χρησιμοποιήθηκε κυρίως σε συνδυασμό με κυματίδια χώρου-κλίμακας, δείξαμε ότι είναι επίσης κατάλληλο για μοντελοποίηση κυματιδίων χώρου-συχνότητας όπως ο MCLT. Συζητήσαμε κατόπιν πώς μπορούμε να χρησιμοποιήσουμε την κυματιδιακή HMT αναπαράσταση σε διάφορες εφαρμογές ανάλυσης εικόνας και όρασης υπολογιστών. Μέσω μιας σειράς πειραμάτων σύνθεσης υφής, δείξαμε το είδος της οπτικής πληροφορίας που μπορεί να συλλάβει το μοντέλο και αναδείξαμε πλεονεκτήματα και μειονεκτήματα της αναπαράστασης. Σε εκτεταμένα πειράματα αποθορυβοποίησης δείξαμε ότι το αναγεννητικό κυματιδιακό HMT μοντέλο παρουσιάζει επιδόσεις συγκρίσιμες με τις καλύτερες δημοσιευμένες τεχνικές στη βιβλιογραφία. Πρωτότυπες ερευνητικές συνεισφορές μας σε σχέση με το πρόβλημα της αποθορυβοποίησης είναι, αφενός, η σύζευξη της HMT δενδρικής δομής με τον MCLT, συνδυασμός που για

ορισμένες εικόνες αποδίδει καλύτερα σε σχέση με το HMT μοντέλο ορισμένο στον MCLT, και, αφετέρου, η ανάπτυξη τεχνικών που επιτρέπουν την εκτίμηση των παραμέτρων του HMT μοντέλου κατευθείαν από τα θορυβώδη δεδομένα της εικόνας. Η κύρια ερευνητική μας συνεισφορά σε αυτό το Κεφάλαιο συνίσταται στην επέκταση του HMT κυματιδιακού μοντέλου για εφαρμογή στο πρόβλημα της ενδοσυμπλήρωσης εικόνων. Συγκεκριμένα, δείχνουμε πώς μπορούμε να χειριστούμε ελλιπή δεδομένα στο πλαίσιο του HMT μοντέλου, οδηγούμενοι σε μια σειρά ντετερμινιστικών και στοχαστικών αλγορίθμων. Οι προκύπτουσες τεχνικές ενδοσυμπλήρωσης αποδίδουν ικανοποιητικά στην αποκατάσταση κενών μικρής ή μεσαίας έκτασης.

Υπάρχουν πολλές προοπτικές για την περαιτέρω βελτίωση των τεχνικών του παρόντος Κεφαλαίου. Με αναφορά το πρόβλημα της ενδοσυμπλήρωσης εικόνων και ειδικότερα κενών ευρείας έκτασης, είναι φανερό ότι απαιτείται ο εμπλουτισμός του μοντέλου με ενσωμάτωση λειτουργιών υψηλότερου επιπέδου όρασης. Σε ένα πρώτο επίπεδο, σημαντικές βελτιώσεις θα μπορούσαν να επιτευχθούν με ενσωμάτωση στο μοντέλο λειτουργιών *αντιληπτικής οργάνωσης* (perceptual organization) (π.χ. καλή συνέχιση ακμών), συνδυάζοντας την κυματιδιακή περιγραφή της εικόνας με έννοιες πρωταρχικού σκίτσου, παρόμοια με την προσέγγιση του [GZW03]. Το πεδίο κυματιδίων, ιδιαίτερα των μιγαδικών κυματιδίων, είναι αναπαράσταση ιδιαίτερα πρόσφορη για εξαγωγή, επεξεργασία, και ενδοσυμπλήρωση του πρωταρχικού σκίτσου της εικόνας. Επίσης, υπάρχει πλούσια βιβλιογραφία σε μεθόδους για συνέχιση του πρωταρχικού σκίτσου, για παράδειγμα τα [Kan79, WJ96, GM96], ενώ και οι μέθοδοι ΜΔΕ ανισοτροπικής διάχυσης του Κεφαλαίου 3 με τις πολύ καλές ιδιότητες συνέχισης ακμών μπορεί να φανούν χρήσιμες προς αυτήν την κατεύθυνση. Θα είναι ενδιαφέρον να προσπαθήσουμε να ενσωματώσουμε ιδέες από αυτές τις τεχνικές στο σύστημα ενδοσυμπλήρωσης που αναπτύσσουμε.

Δεδομένης της σημασίας της κυματιδιακής φάσης στο σχηματισμό και στην πρόσληψη δομικών χαρακτηριστικών της εικόνας όπως οι ακμές, η ενσωμάτωση λειτουργιών αντιληπτικής οργάνωσης μπορεί να ιδωθεί ως σημαντική βελτίωση του μοντέλου HMT στον τομέα της μοντελοποίησης φάσης. Όπως έχουμε ήδη συζητήσει, μια από τις σημαντικότερες αδυναμίες του παρόντος μοντέλου είναι ότι αγνοεί την πληροφορία φάσης, γεγονός που μειώνει την πιστότητα των αποτελεσμάτων σύνθεσης υψής και ενδοσυμπλήρωσης σε περιοχές που εμφανίζονται σύμφωνες δομές. Πρόσφατες εργασίες έχουν δείξει ότι προσεκτικότερη μοντελοποίηση της κυματιδιακής φάσης με αξιοποίηση των συσχετίσεων μεταξύ διαδοχικών κλιμάκων μπορεί να αποφέρει σημαντικές βελτιώσεις [MK08b].

Σε ένα περαιτέρω επίπεδο, θα είχε ιδιαίτερο ενδιαφέρον η ανάπτυξη πλήρως ιεραρχικών αναγεννητικών μοντέλων όρασης [ZM06], ικανών να ενσωματώνουν σημασιολογική πληροφορία υψηλού επιπέδου για αντικείμενα παρόντα στην εικόνα. Κυματιδιακές αναπαραστάσεις παρόμοιες με αυτές που χρησιμοποιούμε στην εργασία μας πιθανόν να είναι κατάλληλες για αυτόν το σκοπό, όπως φαίνεται για παράδειγμα στο [WSFZ07]. Τέτοιου είδους μοντέλα θα μπορούσαν επίσης να χρησιμοποιηθούν για οπτική ανίχνευση αντικειμένων, ενώ θα επέτρεπαν την αποκατάσταση εικόνας λαμβάνοντας υπόψη τα αντικείμενα που αναπαρίστανται σε αυτή και επιτρέποντας ανθρώπινο έλεγχο της διαδικασίας σε σημασιολογικό επίπεδο.

6.Α' Επαναληπτική Επίλυση Συστημάτων Γραμμικών Εξισώσεων: Μέθοδοι Καθόδου Μέγιστης Κλίσης και Συζυγών Κλίσεων

Για την ανάκτηση των συντελεστών κυματιδίων \mathbf{y} με τα μοντέλα του παρόντος Κεφαλαίου ανακύπτει η ανάγκη για επίλυση του γραμμικού συστήματος $B\mathbf{y} = \mathbf{b}$, όπου ο $N \times N$ πίνακας $B = \Sigma_{\mathbf{y}|\mathbf{q}}^{-1} = \Sigma_{\mathbf{q}}^{-1} + A^T \Sigma_n^{-1} A$ είναι συμμετρικός, θετικά ορισμένος και πυκνός. Όταν χρησιμοποιούμε μιγαδικά κυματίδια, το πραγματικό και φανταστικό τους μέρος θεωρούνται ως ξεχωριστές διαδοχικές πραγματικές μεταβλητές στο διάνυσμα \mathbf{y} . Το σύστημα αυτό έχει μοναδική λύση, που μπορεί ισοδύναμα να ιδωθεί ως το διάνυσμα που ελαχιστοποιεί τη θετικά ορισμένη τετραγωνική μορφή $f(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T B \mathbf{y} - \mathbf{b}^T \mathbf{y}$. Η μεγάλη διάσταση του συστήματος (το πλήθος των αγνώστων είναι της τάξης του 10^6) μας επιβάλλει να καταφύγουμε σε επαναληπτικές μεθόδους για την επίλυσή του [GVL96, BBC⁺94, Saa96, She94, Grö93, SW01]. Ως επιπλέον ιδιαιτερότητα του προβλήματος που αντιμετωπίζουμε, είναι υπολογιστικά ιδιαίτερα ακριβό να σχηματίσουμε ρητά και να αποθηκεύσουμε στην κύρια μνήμη του υπολογιστή τον πίνακα συστήματος B : έχουμε όμως τη δυνατότητα να υπολογίζουμε γινόμενα πίνακα διανύσματος της μορφής $B\mathbf{y}$ με υπολογιστικά αποδοτικούς αλγορίθμους πολυπλοκότητας $\mathcal{O}(N)$ βασισμένους στον ευθύ/αντίστροφο διακριτό κυματιδιακό μετασχηματισμό.

Η απλούστερη επαναληπτική τεχνική για εύρεση του ελαχίστου της τετραγωνικής μορφής $f(\mathbf{y})$ είναι η μέθοδος *καθόδου μέγιστης κλίσης* (gradient descent - GD). Σε αυτήν βελτιώνουμε την παρούσα εκτίμηση \mathbf{y}^i κινούμενοι στην κατεύθυνση μέγιστης κλίσης $\mathbf{d}^i = -\nabla f(\mathbf{y}^i) = \mathbf{b} - B\mathbf{y}^i = \mathbf{r}^i$, δηλαδή $\mathbf{y}^{i+1} = \mathbf{y}^i + \alpha^i \mathbf{d}^i$. Το μέγεθος βήματος επιλέγεται έτσι ώστε να ελαχιστοποιείται η $f(\mathbf{y}^i + \alpha^i \mathbf{d}^i)$, που επιτυγχάνεται για $\alpha^i = \langle \mathbf{r}^i, \mathbf{r}^i \rangle / \langle \mathbf{r}^i, B\mathbf{r}^i \rangle$. Το υπολογιστικό κόστος κυριαρχείται από τον υπολογισμό του γινομένου πίνακα-διανύσματος $B\mathbf{r}^i$ ανά βήμα του αλγορίθμου. Αν και η μέθοδος καθόδου μέγιστης κλίσης μειώνει σε κάθε βήμα την αντικειμενική συνάρτηση και εξασφαλίζει σύγκλιση προς τη λύση, η σύγκλιση τείνει να επιβραδύνεται στην περίπτωση που οι ιδιοτιμές του πίνακα B , που αντιστοιχούν στους άξονες του ελλειψοειδούς που ορίζεται από τη σχέση $f(\mathbf{y}) = (\text{σταθ.})$, είναι κατά πολύ διαφορετικές μεταξύ τους. Τότε, εξαιτίας του μυωπικού χαρακτήρα της μεθόδου (η νέα κατεύθυνση αναζήτησης \mathbf{d}^i δεν εξαρτάται από τις προηγούμενες), συχνά οδηγούμαστε να κάνουμε βήματα στην ίδια κατεύθυνση με προηγούμενες επαναλήψεις και η τροχιά \mathbf{y}^i αναλώνεται σε μικρά βήματα/ταλαντώσεις που μειώνουν αργά την f .

Βελτιωμένη εκδοχή του αλγορίθμου μέγιστης κλίσης αποτελεί η μέθοδος *συζυγών κλίσεων* (conjugate gradients - CG). Αν και η μέθοδος συζυγών κλίσεων έχει παρόμοια χαμηλό υπολογιστικό κόστος με τη μέθοδο καθόδου μέγιστης κλίσης, καταφέρνει να διερευνά το χώρο με πιο καθολικό τρόπο, λαμβάνοντας υπόψη τα προηγούμενα βήματα αναζήτησης. Συγκεκριμένα, στο πρώτο βήμα επιλέγουμε το υπόλοιπο (κατεύθυνση μέγιστης μεταβολής) ως κατεύθυνση αναζήτησης, $\mathbf{d}^0 = \mathbf{r}^0$. Στα επόμενα όμως βήματα διορθώνουμε κατάλληλα το υπόλοιπο, ώστε η νέα κατεύθυνση αναζήτησης $\mathbf{d}^i = \mathbf{r}^i + \beta^i \mathbf{d}^{i-1}$ να είναι *συζυγής* στην προηγούμενη, $\langle \mathbf{d}^i, B\mathbf{d}^j \rangle = 0$, $j = i - 1$. Μπορεί να δειχθεί ότι η νέα κατεύθυνση αναζήτησης θα είναι τότε συζυγής και συνεπώς αρκούντως διαφορετική από όλες τις προηγούμενες της, \mathbf{d}^j , $j < i$, επιλέγοντας $\beta^i = \langle \mathbf{r}^i, \mathbf{r}^i \rangle / \langle \mathbf{r}^{i-1}, \mathbf{r}^{i-1} \rangle$. Το σύνολο των κατευθύνσεων αναζήτησης διαγράφει κατάλληλο

Krylov υπόχωρο $K_i = \text{span}(\{\mathbf{d}^j, 0 \leq j \leq i\}) = \text{span}(\{B^j \mathbf{r}^0, 0 \leq j \leq i\})$ που επεκτείνεται σταδιακά ωσότου καλύψει όλο το χώρο $K_{N-1} = \mathbb{R}^N$. Έτσι, αγνοώντας σφάλματα πεπερασμένης αριθμητικής, η μέθοδος συζυγών κλίσεων μπορεί να βρει την ακριβή λύση του γραμμικού συστήματος ύστερα από N το πολύ βήματα, αν και για το πλήθος των αγνώστων που μας απασχολεί στα προβλήματά μας αυτή η θεωρητική εγγύηση δεν έχει πρακτική αξία. Στην πράξη μερικές δεκάδες ή εκατοντάδες επαναλήψεις αρκούν για να προσεγγίσουμε τη λύση με ικανοποιητική ακρίβεια στα προβλήματα που αντιμετωπίζουμε.

Αλγόριθμος 2 Προρυθμισμένες μέθοδοι καθόδου μέγιστης κλίσης (PGD) και συζυγών κλίσεων (PCG).

function PGD(B, \mathbf{y}, M, n)

\mathbf{r} : residual

\mathbf{p} : search direction

\mathbf{q} : residual change direction

$\mathbf{r} \leftarrow \mathbf{b} - B\mathbf{y}$

for $i \leftarrow 1, n$ **do**

$\mathbf{p} \leftarrow M^{-1}\mathbf{r}$

$\delta \leftarrow \langle \mathbf{r}, \mathbf{p} \rangle$

$\mathbf{q} \leftarrow B\mathbf{p}$

$\alpha \leftarrow \delta / \langle \mathbf{p}, \mathbf{q} \rangle$

$\mathbf{y} \leftarrow \mathbf{y} + \alpha\mathbf{p}$

$\mathbf{r} \leftarrow \mathbf{r} - \alpha\mathbf{q}$

end for

return \mathbf{y}

end function

function PCG(B, \mathbf{y}, M, n)

\mathbf{r} : residual

\mathbf{d} : search direction

\mathbf{q} : residual change direction

$\mathbf{r} \leftarrow \mathbf{b} - B\mathbf{y}$

$\mathbf{d} \leftarrow M^{-1}\mathbf{r}$

$\delta_1 \leftarrow \langle \mathbf{r}, \mathbf{d} \rangle$

for $i \leftarrow 1, n$ **do**

$\mathbf{q} \leftarrow B\mathbf{d}$

$\alpha \leftarrow \delta_1 / \langle \mathbf{d}, \mathbf{q} \rangle$

$\mathbf{y} \leftarrow \mathbf{y} + \alpha\mathbf{d}$

$\mathbf{r} \leftarrow \mathbf{r} - \alpha\mathbf{q}$

$\mathbf{s} \leftarrow M^{-1}\mathbf{r}$

$\delta_0 \leftarrow \delta_1$

$\delta_1 \leftarrow \langle \mathbf{r}, \mathbf{s} \rangle$

$\beta \leftarrow \delta_1 / \delta_0$

$\mathbf{d} \leftarrow \mathbf{s} + \beta\mathbf{d}$

end for

return \mathbf{y}

end function

Τα χαρακτηριστικά σύγκλισης και των δύο προαναφερθεισών επαναληπτικών τεχνικών μπορούν να βελτιωθούν εάν *προρυθμίσουμε* (precondition) κατάλληλα το γραμμικό σύστημα, δηλαδή εάν εφαρμόσουμε τις τεχνικές αυτές στο ισοδύναμο προρυθμισμένο γραμμικό σύστημα $E^{-1}BE^{-T}\tilde{\mathbf{y}} = \mathbf{b}$, $\tilde{\mathbf{y}} = E^T\mathbf{y}$. Με τη βοήθεια του συμμετρικού θετικά ορισμένου προρυθμιστή $M = EE^T$ ο δείκτης κατάστασης (condition number) του μετασχηματισμένου πίνακα $E^{-1}BE^{-T}$ μπορεί να βελτιωθεί και αντίστοιχα να επιταχυνθεί η σύγκλιση των μεθόδων καθόδου μέγιστης κλίσης και συζυγών κλίσεων. Η σχεδίαση προρυθμιστή M που συνδυάζει απλότητα και αποδοτικότητα είναι γενικά δύσκολη και εξαρτάται από τις ιδιαιτερότητες κάθε εφαρμογής. Συνηθισμένη είναι η χρήση του διαγώνιου προρυθμιστή $\text{diag}(B)$, καθώς και προρυθμιστών βασισμένων σε ελλiptή παραγοντοποίηση πινάκων. Τέτοιοι γενικού τύπου προρυθμιστές δυστυχώς είναι δύσκολο να προσαρμοστούν στο πρόβλημά μας γιατί ο πίνακας συστήματος $B = \Sigma_{y|q}^{-1} = \Sigma_q^{-1} + A^T\Sigma_n^{-1}A$ δεν είναι ρητά διαθέσιμος. Στα πειράματά μας χρησιμοποιούμε την προερχόμενη από το πρότερο μοντέλο διαγώνια συνιστώσα του $M = \Sigma_q^{-1}$ ως προρυθμιστή, που περίπου διπλασιάζει την ταχύτητα σύγκλισης της μεθόδου συζυγών κλίσεων. Παραθέτουμε τις προρυθμισμένες

εκδοχές των μεθόδων καθόδου μέγιστης κλίσης (PGD) και συζυγών κλίσεων (PCG) στον Αλγόριθμο 2.

6.Β' Αριθμητική Επίλυση Προβλημάτων Κυρτού Τετραγωνικού Προγραμματισμού με Περιορισμούς Ανισότητας

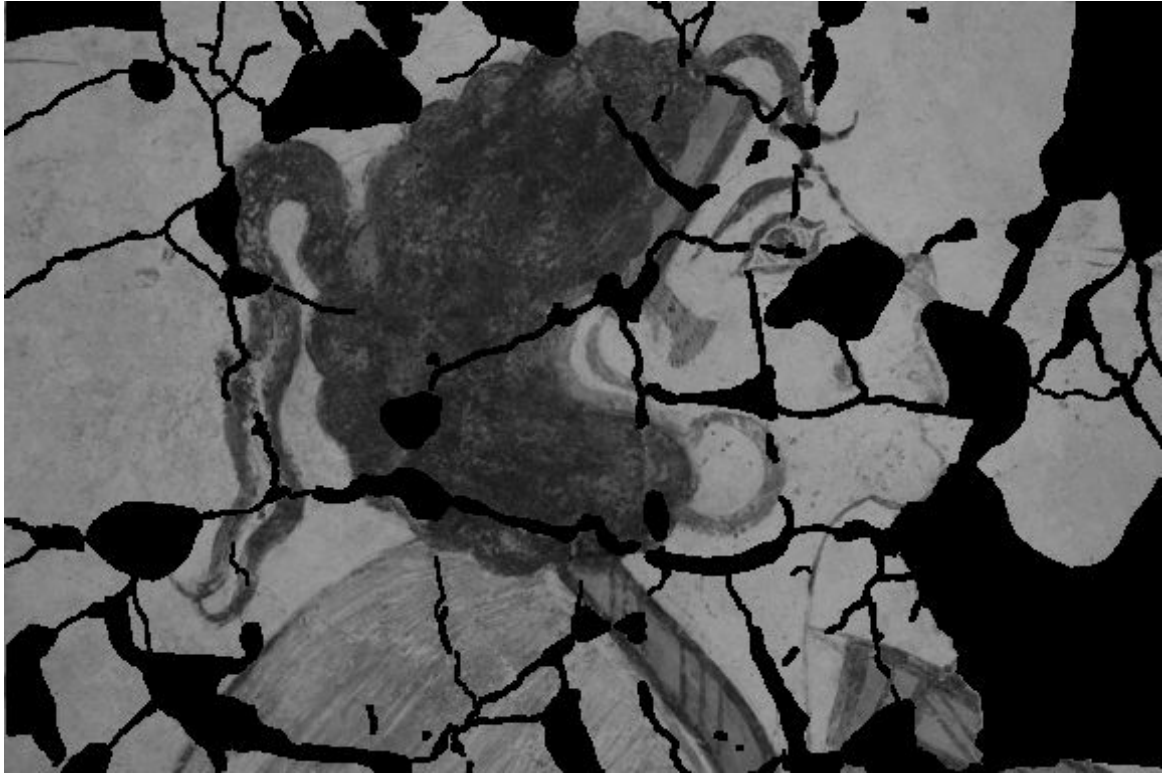
Όταν η μιγαδική φάση $\phi_t = \angle y_{t,r} + j y_{t,i}$ των συντελεστών κυματιδίων θεωρείται κατά προσέγγιση γνωστή, τα μοντέλα του παρόντος Κεφαλαίου οδηγούν σε προβλήματα βελτιστοποίησης ως προς το μέτρο $r_t = \sqrt{y_{t,r}^2 + y_{t,i}^2}$ των μιγαδικών συντελεστών κυματιδίων. Εδώ συζητάμε τεχνικές για την αριθμητική επίλυση του προκύπτοντος προβλήματος βελτιστοποίησης (6.14).

Το πρόβλημα (6.14) αποτελεί περίπτωση κυρτού τετραγωνικού προγράμματος με γραμμικούς ανισοτικούς περιορισμούς. Υπάρχουν διάφορες γενικές τεχνικές για την αριθμητική επίλυση αυτής της κλάσης προβλημάτων [BV04, NW06]. Οι μέθοδοι *εσωτερικού σημείου* (interior point) συνιστούν μία από τις πιο επιτυχημένες κατηγορίες σύγχρονων τεχνικών για επίλυση προβλημάτων κυρτής βελτιστοποίησης. Ονομάζονται έτσι γιατί, σε αντιδιαστολή με κλασικές τεχνικές γραμμικού προγραμματισμού (π.χ. μέθοδος simplex), οι ενδιάμεσες εκτιμήσεις της λύσης βρίσκονται στο εσωτερικό και όχι στο σύνορο του εφικτού κυρτού συνόλου. Μια κλάση αλγορίθμων εσωτερικού σημείου κατάλληλη για το πρόβλημά μας αποτελούν οι τεχνικές *φραγμού* (barrier method), οι οποίες αντικαθιστούν τους περιορισμούς ανισότητας με επιπλέον όρους ποινής στην αντικειμενική συνάρτηση και ανάγουν τη λύση του προβλήματος στην επίλυση μιας ακολουθίας προβλημάτων βελτιστοποίησης δίχως ανισοτικούς περιορισμούς [BV04]. Μια εναλλακτική κλάση αλγορίθμων εσωτερικού σημείου βασίζονται σε μεθόδους ανάλυσης σε *πρωτεύουσες-δυσικές* (primal-dual) μεταβλητές, δες π.χ. το [NW06]. Παραπέμπουμε στο [KKL⁺07] για παράδειγμα επίλυσης προβλήματος αντίστοιχης κλίμακας με αυτό του παρόντος Κεφαλαίου με τεχνικές εσωτερικού σημείου.

Οι τεχνικές *ενεργού συνόλου* (active set) αποτελούν μια άλλη κατηγορία μεθόδων για επίλυση προβλημάτων τετραγωνικού προγραμματισμού με ανισοτικούς περιορισμούς. Στο βέλτιστο σημείο-λύση του προβλήματος, ένα υποσύνολο \mathcal{A} των περιορισμών ανισότητας είναι *ενεργοί*, δηλαδή ικανοποιούνται ως ισότητες, ενώ οι υπόλοιποι είναι ανενεργοί. Εάν γνωρίζουμε το σύνολο των ενεργών περιορισμών, μπορούμε να αντικαταστήσουμε το τετραγωνικό πρόβλημα με περιορισμούς ανισότητας (6.14) με αντίστοιχο τετραγωνικό πρόβλημα με περιορισμούς ισότητας, που είναι πιο εύκολο να λυθεί. Στην περίπτωση μας μηδενίζουμε τους συντελεστές κυματιδίων που αντιστοιχούν στο ενεργό σύνολο $r_t = 0, t \in \mathcal{A}$, και λύνουμε το ευκολότερο πρόβλημα δίχως όρους ελαχιστοποίησης της τετραγωνικής μορφής που αντιστοιχεί στους υπόλοιπους συντελεστές, που η λύση του θα αποτελείται εξ ορισμού από θετικούς συντελεστές. Οι μέθοδοι ενεργού συνόλου προσπαθούν να ταυτοποιήσουν το ενεργό σύνολο για κάποιο συγκεκριμένο πρόβλημα. Συγκεκριμένα, έστω \mathbf{r}^i η παρούσα εκτίμηση για τη λύση (υποθέτουμε ότι είναι επιτρεπτή, δηλαδή ικανοποιεί τους περιορισμούς) και \mathcal{W}_i η αντίστοιχη εκτίμηση για το ενεργό σύνολο. Ως κατεύθυνση αναζήτησης \mathbf{d}^i χρησιμοποιούμε τη λύση του προβλήματος $\min f(\mathbf{r}^i + \mathbf{d}^i)$, υπό τους περιορισμούς ισότητας $r_t = 0, t \in \mathcal{W}_i$. Κατόπιν θέτουμε $\mathbf{r}^{i+1} = \mathbf{r}^i + \alpha^i \mathbf{d}^i$, όπου $\alpha^i \leq 1$ είναι

το μέγιστο βήμα που διατηρεί το \mathbf{r}^{i+1} στο επιτρεπτό σύνολο και ενημερώνουμε κατάλληλα το σύνολο των ενεργών περιορισμών ώστε να πάρουμε το \mathcal{W}_{i+1} , που συνήθως διαφέρει κατά ένα στοιχείο σε σχέση με το \mathcal{W}_i . Για περισσότερες πληροφορίες παραπέμπουμε στο [NW06].

Καθότι η εκτίμηση για το ενεργό σύνολο ανανεώνεται αργά, οι τεχνικές ενεργού συνόλου είναι πρακτικά εφαρμόσιμες σε προβλήματα μικρής και μεσαίας κλίμακας. Για προβλήματα μεγάλης κλίμακας όπως το (6.14), όπου επιπλέον οι ανισοτικοί περιορισμοί εμφανίζονται υπό την απλή μορφή επιτρεπτού διαστήματος για καθεμιά από τις εμπλεκόμενες μεταβλητές (αντί για τη γενική μορφή γραμμικών ανισοτικών περιορισμών τύπου $\mathbf{a}_t^T \mathbf{r} \geq 0, 1 \leq t \leq N$), είναι πιο αποδοτική η χρήση τεχνικών *προβολής κλίσης* (gradient projection). Αυτές αποτελούν παραλλαγή των μεθόδων ενεργού συνόλου με το πλεονέκτημα ότι η εκτίμηση για το ενεργό σύνολο αλλάζει ταχύτερα, δηλαδή τα \mathcal{W}_i και \mathcal{W}_{i+1} επιτρέπεται να διαφέρουν σε περισσότερα του ενός στοιχεία. Μια τυπική μορφή της ανανέωσης της λύσης με τη μέθοδο προβολής κλίσης είναι $\mathbf{r}^{i+1} = [\mathbf{r}^i - \alpha^i \nabla f(\mathbf{r}^i)]_+$, όπου $[\cdot]_+$ δηλώνει προβολή στο εφικτό σύνολο. Το βήμα α^i επιλέγεται έτσι ώστε να εξασφαλίζει επαρκή μείωση της αντικειμενικής συνάρτησης, $f(\mathbf{r}^{i+1}) - f(\mathbf{r}^i) \leq \mu \langle \nabla f(\mathbf{r}^i), \mathbf{r}^{i+1} - \mathbf{r}^i \rangle$, με μ σταθερά στο διάστημα $(0, 1)$. συγκεκριμένα ξεκινάμε με μια συγκεκριμένη τιμή για το βήμα α^i την οποία υποδιπλασιάζουμε ωσότου ικανοποιηθεί η συνθήκη επαρκούς μείωσης. Αφού καταλήξουμε στο ενημερωμένο \mathbf{r}^{i+1} , ανανεώνουμε κατάλληλα το σύνολο των ενεργών περιορισμών \mathcal{W}_{i+1} και προχωρούμε. Για περισσότερες λεπτομέρειες παραπέμπουμε στα [Ber99, NW06, FNW07]. Ενδιαφέρον έχει και η υβριδική τεχνική *προβολής κλίσης/συζυγών κλίσεων* (gradient projection conjugate gradients - GPCG) όπου εναλλάσσονται η φάση προβολής κλίσης, που αποσκοπεί στην ενημέρωση του ενεργού συνόλου \mathcal{W}_i , και η φάση συζυγών κλίσεων, όπου με βήματα του αλγορίθμου συζυγών κλίσεων βελτιώνεται η λύση στον υποχώρο που ορίζεται από το παρόν ενεργό σύνολο \mathcal{W}_i [MT91, BV03]. Για τα πειράματα της Ενότητας 6.6.3 έχουμε υλοποιήσει και χρησιμοποιήσει τις τεχνικές προβολής κλίσης και συνδυασμένης προβολής κλίσης/συζυγών κλίσεων που μόλις περιγράψαμε.

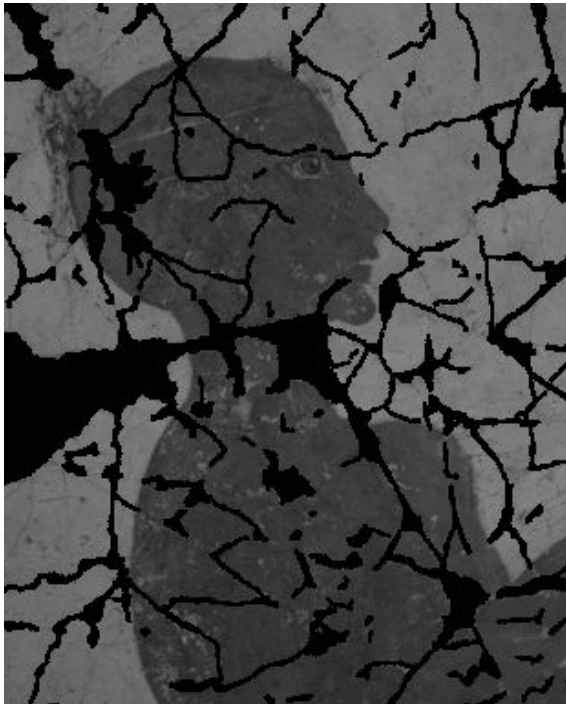


(α)



(β)

Σχήμα 6.7: Αποτέλεσμα ενδοσυμπλήρωσης σε τμήμα της τοιχογραφίας 'Γυναίκες' από το Ακρωτήρι. (α) Θηραϊκή τοιχογραφία με επισημειωμένα κενά. (β) Ενδοσυμπλήρωση με το προτεινόμενο μοντέλο. Χρησιμοποιούμε το μοντέλο HMT-DTCWT στο πεδίο των μιγαδικών κυματιδίων, επιλυμένο με την τεχνική δειγματοληψίας Gibbs. Rao-Blackwell εκτιμητής βασισμένος στο κριτήριο MMSE.



(α)



(β)

Σχήμα 6.8: Αποτέλεσμα ενδοσυμπλήρωσης σε τμήμα της θηραϊκής τοιχογραφίας Άκολουθία Ανδρών. (α) Θηραϊκή τοιχογραφία με επισημειωμένα κενά. (β) Ενδοσυμπλήρωση με το προτεινόμενο μοντέλο. Μοντέλο HMT-DTCWT στο πεδίο των μιγαδικών κυματιδίων, επιλυμένο με την τεχνική εναλλασσόμενης βελτιστοποίησης ICM.



(α)

(β)



(γ)

(δ)

Σχήμα 6.9: Συγκριτικά αποτελέσματα ενδοσυμπλήρωσης σε τεχνητά επικαλυμμένη εικόνα. (α) Επικαλυμμένη εικόνα. (β) Αποκατάσταση με ΜΔΕ Laplace (PSNR: 20.62 dB). (γ) HMT / εναλλασσόμενη βελτιστοποίηση (PSNR: 21.95 dB). (δ) HMT / δειγματοληψία Gibbs (PSNR: 22.00 dB).

Κεφάλαιο 7

Σύνοψη Διατριβής και Μελλοντική Έρευνα

7.1 Συνεισφορές Διατριβής

Η εργασία μας κινείται στην περιοχή της ανάλυσης εικόνων και όρασης υπολογιστών και έχει ως βασικό πεδίο εφαρμογής την ψηφιακή αποκατάσταση των τοιχογραφιών του Ακρωτηρίου Θήρας. Οι συνεισφορές της διατριβής κινούνται τόσο στο πρακτικό επίπεδο των αναγκών της συγκεκριμένης εφαρμογής, όσο και στο επίπεδο θεωρητικών προβλημάτων ή θεμάτων βασικής έρευνας και μπορούν να συνοψισθούν στα ακόλουθα σημεία :

- Προχωρήσαμε στη λεπτομερή φωτογραφική αποτύπωση σημαντικού αριθμού από τις τοιχογραφίες του Ακρωτηρίου Θήρας. Η λύση που δώσαμε στο πρόβλημα της αποτύπωσης μιας ολόκληρης τοιχογραφίας σε μία ενοποιημένη σύνθεση υπερυψηλής ανάλυσης συνίσταται στη λήψη μεγάλου πλήθους φωτογραφιών, με καθεμία από αυτές να απεικονίζει μέρος του τοιχογραφικού συνόλου, με μετέπειτα συνένωση των επιμέρους στιγμιότυπων σε ένα γιγαντιαίο φωτογραφικό μωσαϊκό ανά σύνθεση στο οποίο μπορεί εύκολα να περιηγηθεί ο μελετητής των τοιχογραφιών. Η προσέγγιση αυτή, αυτοματοποιημένη σε μεγάλο βαθμό χάρη στη χρήση σύγχρονων τεχνικών όρασης υπολογιστών, αποδείχθηκε ιδιαίτερα επιτυχημένη και αναδεικνύεται ως μια πρακτική και ιδιαίτερα αποδοτική μέθοδος φωτογραφικής αποτύπωσης έργων τέχνης, δίχως την ανάγκη ύπαρξης ιδιαίτερα ακριβού εξειδικευμένου εξοπλισμού.
- Με αφορμή την ανάγκη αποδοτικών αλγορίθμων για το πρόβλημα της ενδοσυμπλήρωσης εικόνων με μοντέλα μη γραμμικής ανισοτροπικής διάχυσης, αναπτύξαμε κατάλληλους πολυπλεγματικούς αλγορίθμους, που είναι επίσης εφαρμόσιμοι για την ταχεία αριθμητική επίλυση γεωμετρικών μοντέλων κατάτμησης και άλλων τεχνικών της όρασης υπολογιστών εκφρασμένων σε φορμαλισμό μερικών διαφορικών εξισώσεων.
- Η περιγραφή της εικόνας σε διαδοχικές κλίμακες συνιστά το *χώρο κλίμακας* (scale space). Αν και η πλήρης περιγραφή της δομής της εικόνας απαιτεί ολόκληρο το χώρο κλίμακας, για ορισμένες εφαρμογές αρκεί να επιλέξει κανείς ένα μόνο αντιπροσωπευτικό του στιγμιότυπο. Για την εφαρμογή της αποθορυβοποίησης εικόνων,

δείχνουμε πως η επιλογή του κατάλληλου στιγμιότυπου από το χώρο κλίμακας μπορεί να ιδωθεί ως ένα πρόβλημα επιλογής στατιστικού μοντέλου και να αντιμετωπιστεί αποδοτικά με μεθόδους στατιστικής *διεπικύρωσης* (cross-validation).

- Μοντέλα περιγραφής εικόνων στο πεδίο των κυματιδίων βρίσκονται στο επίκεντρο της σύγχρονης έρευνας στις περιοχές ανάλυσης σήματος και εικόνας. Σε αυτό το πλαίσιο, προτείνουμε ένα νέο μοντέλο ενδοσυμπλήρωσης βασισμένο σε στοχαστικά μοντέλα φυσικών εικόνων εκφρασμένα στο πεδίο των μιγαδικών κυματιδίων. Το μοντέλο αυτό είναι πιθανοτικό και χρησιμοποιεί αναδρομικές ως προς την κλίμακα μαρκοβιανές δομές που επιτρέπουν με αποδοτικό τρόπο στατιστική συμπερασματολογία για τις ελλείπουσες περιοχές της εικόνας.
- Εκτός από το πρόβλημα της ενδοσυμπλήρωσης, η ενασχόλησή μας με το πρόβλημα της οπτικής-ακουστικής ανάλυσης ομιλίας έχει οδηγήσει σε ενδιαφέροντα νέα αποτελέσματα σε δύο κύριους άξονες. Αφενός, έχουμε προτείνει στα πλαίσια μιας συνεργατικής προσπάθειας μια θεωρία για βέλτιστη προσαρμοστική σύμμιξη πολυαισθητηριακών δεδομένων βασισμένη στη διαχείριση της αβεβαιότητας των επιμέρους θορυβωδών μετρήσεων, που έχει αποδειχθεί ιδιαίτερα αποδοτική στο πρόβλημα της αυτόματης αναγνώρισης ομιλίας με συνδυασμό οπτικής και ακουστικής πληροφορίας. Αφετέρου, έχουμε αναπτύξει αποδοτικές τεχνικές για οπτική μοντελοποίηση και εξαγωγή χαρακτηριστικών ομιλίας από το ανθρώπινο πρόσωπο.

Κατάλογος με τις σχετικές μας δημοσιεύσεις παρατίθεται στο Παράρτημα Β'.

7.2 Κατευθύνσεις Μελλοντικής Έρευνας

Μια σειρά ζητημάτων σχετιζόμενα με την ερευνητική μας προσπάθεια στα πλαίσια του διδακτορικού παραμένουν ανοιχτά και θα απασχολήσουν την έρευνά μας στο μέλλον. Αυτά αφορούν τόσο βελτιώσεις των προτεινόμενων τεχνικών που αποτελούν αντικείμενο μελέτης για το άμεσο μέλλον, όσο και ένα σύνολο μακροπρόθεσμων στόχων που σχετίζονται με τις ευρύτερες προοπτικές της προσέγγισης που ακολουθήσαμε στο πρόβλημα της μοντελοποίησης εικόνων.

Πιο συγκεκριμένα, αξίζει να αναφερθούμε στα ακόλουθα ζητήματα:

- Σχετικά με το πρόβλημα της ψηφιακής αποκατάστασης των τοιχογραφιών στο Ακρωτήρι Θήρας με συμπλήρωση των ελλειπουσών περιοχών, παραμένει ανοιχτό το ζήτημα της αυτοματοποιημένης ανίχνευσης των κενών στις εικόνες, που θα επιτρέψει την εφαρμογή των προτεινόμενων τεχνικών ενδοσυμπλήρωσης σε μεγάλη κλίμακα. Προς αυτήν την κατεύθυνση, αξίζει να διερευνηθεί η δυνατότητα εφαρμογής τεχνικών υπολογιστικής φωτογραφίας που πιθανώς να διευκολύνουν σημαντικά την ανίχνευση των κενών, όπως συζητάμε στην Ενότητα 2.1, που όμως απαιτούν πολυπλοκότερη διάταξη για τη λήψη φωτογραφικού υλικού.
- Ενώ ως τώρα έχουμε εστιάσει την προσοχή μας στο πρόβλημα της αισθητικής αποκατάστασης των τοιχογραφιών, οι προτεινόμενες τεχνικές συμπλήρωσης οπτικής πληροφορίας μπορούν να αποδειχθούν ιδιαίτερα χρήσιμες και στο ιδιαίτερα σημαντικό πρόβλημα της αυτοματοποιημένης εύρεσης επαφών μεταξύ τοιχογραφικών σπαραγμάτων σε πρωιμότερο στάδιο της διαδικασίας συντήρησης. Οι ως τώρα προσεγγίσεις

στο πρόβλημα αξιολογούν την ποιότητα ταιριάσματος μεταξύ δύο υποψηφίων γειτόνων με κριτήρια βασισμένα κυρίως στο σχήμα των κομματιών. Χρησιμοποιώντας κάποιες από τις προτεινόμενες τεχνικές, μπορούμε να επεκτείνουμε περιμετρικά το απεικονιστικό περιεχόμενο των σπαραγμάτων σε ένα στενό δακτύλιο, επιτρέποντας έτσι την αξιολόγηση του ταιριάσματος δύο πιθανά γειτονικών σπαραγμάτων και με βάση τις ζωγραφικές παραστάσεις στην επιφάνειά τους.

- Στα πλαίσια του μοντέλου κυματιδίων για ενδοσυμπλήρωση που μελετήσαμε στο Κεφάλαιο 6, θα διερευνήσουμε περαιτέρω τεχνικές για τη μοντελοποίηση της φάσης των μιγαδικών συντελεστών κυματιδίων με χρήση ιδεών αντιληπτικής οργάνωσης και καλής συνέχισης των ακμών της εικόνας. Καθότι η αντιληπτική σημασία της φάσης είναι ιδιαίτερα σημαντική, αναμένουμε σημαντικά οφέλη από μια τέτοια επαύξηση του μοντέλου μας, ιδιαίτερα όσο αφορά την επιτυχία αποκατάστασης σύμφωνων γραμμικών δομών.
- Οι τεχνικές ενδοσυμπλήρωσης που έχουμε ως τώρα εξετάσει είναι πλήρως αυτοματοποιημένες και συνεπώς δεν αφήνουν περιθώρια ουσιαστικής παρέμβασης του χρήστη στη διαδικασία αποκατάστασης. Έχει ενδιαφέρον να κινηθεί κανείς στην κατεύθυνση ημιαυτόματων μεθόδων ενδοσυμπλήρωσης, όπου ψηφιακό σύστημα και ειδικός αλληλεπιδρούν για την αποκατάσταση της βλάβης μέσω κατάλληλης εύχρηστης γραφικής διεπαφής. Τέτοιου είδους διαδραστικές τεχνικές θα ήταν ιδιαίτερα χρήσιμες για την αποκατάσταση εικόνων με κενά μεγάλης έκτασης, όπου η υψηλού επιπέδου ανθρώπινη γνώση και κατανόηση της εικόνας είναι πρακτικά αδύνατο να ενσωματωθεί στο μοντέλο της εικόνας.
- Ως πιο μακροπρόθεσμο στόχο, θα μπορούσε κανείς να θέσει τη μετεξέλιξη των δενδρικών κυματιδιακών μοντέλων της εικόνας που μελετάμε στο Κεφάλαιο 6 σε ιεραρχικά αναγεννητικά μοντέλα περιγραφής αντικειμένων σε σημασιολογικό επίπεδο. Σε μια τέτοια αναπαράσταση θα πρέπει οι κρυφές μεταβλητές/κόμβοι του δέντρου ανάλυσης να κωδικοποιούν χαρακτηριστικά υψηλότερου επιπέδου σε σχέση με την ένταση απόκρισης ενός γραμμικού φίλτρου, όπως συμβαίνει με το υπάρχον HMT μοντέλο. Αυτό θα καθιστά εφικτή την αντικειμενοκεντρική αποκατάσταση μεγάλων κενών σε εικόνες, επιτρέποντας παράλληλα την αλληλεπίδραση με το χρήστη σε υψηλότερο επίπεδο.

Τα θέματα αυτά θα αποτελέσουν αντικείμενο της μελλοντικής ερευνητικής προσπάθειας του γράφοντα.

Παράρτημα Α΄

Προσαρμοστική Σύμμειξη Πληροφορίας για Συνδυασμένη Οπτική-Ακουστική Αναγνώριση Ομιλίας

Ενώ¹ η ακρίβεια μέτρησης χαρακτηριστικών εξαρτάται ισχυρά από τις μεταβαλλόμενες περιβαλλοντικές συνθήκες, η μελέτη των επιπτώσεων αυτού του γεγονότος έχει τύχει σχετικά λίγης προσοχής στη βιβλιογραφία. Στην εργασία μας λαμβάνουμε ρητά υπόψη μας την αβεβαιότητα στη μέτρηση χαρακτηριστικών και δείχνουμε πώς πρέπει να προσαρμόσουμε τους πολυτροπικούς κανόνες ταξινόμησης και μάθησης ώστε να αντισταθμίσουμε τις επιπτώσεις της. Η προσέγγισή μας είναι ιδιαίτερα εποικοδομητική σε εφαρμογές πολυτροπικής σύμμειξης πληροφορίας όπου συνδυάζονται πολλαπλές χρονορές συμπληρωματικών χαρακτηριστικών, όπως είναι για παράδειγμα η οπτική-ακουστική αναγνώριση ομιλίας. Σε τέτοιες εφαρμογές, έχοντας διαθέσιμη μια εκτίμηση του θορύβου μέτρησης για κάθε ροή χαρακτηριστικών, το προτεινόμενο πλαίσιο οδηγεί σε ισχυρά προσαρμοστικούς κανόνες σύμμειξης που είναι εύκολοι και αποδοτικοί στην υλοποίησή τους. Η τεχνική μας είναι ευρέως εφαρμόσιμη και μπορεί διάφανα να ενσωματωθεί σε σύγχρονες ή ασύγχρονες αρχιτεκτονικές περιγραφής ακολουθιακών δεδομένων. Δείχνουμε πώς αυτές οι ιδέες μπορούν να ενσωματωθούν σε ένα πρακτικό σύστημα οπτικής-ακουστικής αναγνώρισης ομιλίας. Σε αυτό το πλαίσιο προτείνουμε βελτιωμένες τεχνικές όρασης υπολογιστών βασισμένες σε ενεργά μοντέλα όψης για εξαγωγή οπτικών χαρακτηριστικών ομιλίας ανεξάρτητων του ομιλητή και εκτίμηση της αβεβαιότητάς τους. Επίσης συζητάμε πώς μπορούμε σε περιβάλλον ακουστικού θορύβου να εξάγουμε εμπλουτισμένα ακουστικά χαρακτηριστικά ομιλίας κατάλληλα για εφαρμογή με τη μέθοδό μας. Η αποδοτικότητα της προτεινόμενης προσέγγισης επαληθεύεται με πειράματα οπτικής-ακουστικής αναγνώρισης ομιλίας στη βάση CUAVE σε συνδυασμό τόσο με σύγχρονα όσο και με ασύγχρονες

¹Το Κεφάλαιο αυτό αποτελεί σύνοψη του άρθρου [PKPM09] και βασίζεται σε συνεργασία με τους Α. Κατσαμάνη, Β. Πιτσικάλη, και Π. Μαραγκό. Περιλαμβάνεται στην παρούσα αναφορά με κύριο σκοπό να δώσει μια διάσταση της διδακτορικής έρευνας του γράφοντος σε μοντελοποίηση ομιλίας με συνδυασμό οπτικής και ακουστικής πληροφορίας. Επίσης ορισμένα τεχνικά αποτελέσματα αναφορικά με το χειρισμό θορυβωδών δεδομένων σε κρυφά μαρκοβιανά μοντέλα που αναλύονται στο παρόν Κεφάλαιο αποτελούν τη βάση για το χειρισμό του δενδρικού κυματιδιακού HMT μοντέλου στο πλαίσιο της εφαρμογής αποθορυβοποίησης της Ενότητας 6.3.

αρχιτεκτονικές πολυτροπικής σύμμειξης.

Α.1 Εισαγωγή

Εμπνευσμένοι από τον πολυτροπικό χαρακτήρα της ανθρώπινης πρόσληψης ομιλίας [MM76], συμπληρωματικές πηγές πληροφορίας έχουν αξιοποιηθεί σε πολλές εφαρμογές. Χαρακτηριστικό παράδειγμα είναι τα συστήματα *οπτικής-ακουστικής αναγνώρισης ομιλίας* (audio-visual speech recognition (AV-ASR)) [SH96, PNGG03], όπου συνδυάζοντας οπτικές και ακουστικές ενδείξεις μπορεί να αυξήσει σημαντικά την επίδοση σε σχέση με αντίστοιχο σύστημα αποκλειστικά ακουστικής αναγνώρισης, ιδιαίτερα σε ακουστικά επιβαρυμένο περιβάλλον.

Παρόλα αυτά, η επιτυχής σύνθεση ετερογενών πηγών πληροφορίας αποτελεί δύσκολο πρόβλημα [CY90, MM98b, KHDM98, JDM00]. Η επινόηση εύρωστων μηχανισμών σύνθεσης είναι μη προφανής κυρίως επειδή αυτοί θα πρέπει να λειτουργούν σε ένα δυναμικό περιβάλλον που αλλοιώνει ανομοιόμορφα την αξιοπιστία των διαφορετικών τροπικοτήτων. Για παράδειγμα, το οπτικό κανάλι στην εφαρμογή AV-ASR θα πρέπει να παραβλέπεται όταν η κάμερα χάνει την οπτική επαφή με το πρόσωπο του ομιλητή. Άλλοι επιβαρυντικοί παράγοντες όπως η έλλειψη συγχρονισμού μεταξύ τροπικοτήτων περιπλέκουν περαιτέρω το πρόβλημα.

Η τεχνική που προτείνεται στην εργασία μας είναι ειδικά σχεδιασμένη ώστε να επιτρέπει τη δυναμική προσαρμογή των κανόνων πολυτροπικής σύμμειξης σε δυναμικά μεταβαλλόμενες περιβαλλοντικές συνθήκες. Προσεγγίζουμε το πρόβλημα της προσαρμοστικής πολυτροπικής σύμμειξης λαμβάνοντας ρητά υπόψη μας την αβεβαιότητα μέτρησης των επιμέρους χαρακτηριστικών. Η προσέγγιση μοντελοποίησης του θορύβου ακουστικών χαρακτηριστικών έχει αποδειχθεί χρήσιμη σε θορυβοσθεναρά συστήματα αποκλειστικά ακουστικής αναγνώρισης ομιλίας [DRO93, RHR94, YMJ97, DDA05] και σε εφαρμογές *επιβεβαίωσης ταυτότητας ομιλητή* (speaker verification) [YV02]. Επεκτείνουμε αυτές τις ιδέες στο πολυτροπικό περιβάλλον και δείχνουμε στην Ενότητα Α.2 πώς πρέπει να προσαρμόσουμε πολυροϊκούς κανόνες κατηγοριοποίησης για να εξισορροπήσουμε την αβεβαιότητα στη μέτρηση των χαρακτηριστικών. Συζητάμε με λεπτομέρεια και παράγουμε μετατρεμένους αλγορίθμους κατηγοριοποίησης για *μοντέλα μείγματος γκαουσιανών* (Gaussian mixture models (GMMs)) και *κρυφά μαρκοβιανά μοντέλα* (hidden Markov models (HMMs)), αλλά η τεχνική μπορεί δίχως πρόβλημα να ενσωματωθεί με μοντέλα ασύγχρονων χρονοροών όπως τα Product-HMMs [DL00, LPN01, GPN02, NLP⁺02]. Το προτεινόμενο σχήμα οδηγεί σε πολυτροπικούς κανόνες σύμμειξης που προσαρμόζονται χρονικά στο κατώτερο δυνατό επίπεδο του πλαισίου, είναι ευρέως εφαρμόσιμοι και εύκολο να υλοποιηθούν και να ενσωματωθούν σε υπάρχοντα συστήματα αναγνώρισης. Καλύπτουμε επίσης το πρόβλημα της εκπαίδευσης πολυτροπικών μοντέλων από θορυβώδη δεδομένα, παρουσιάζοντας στην Ενότητα Α.3 κατάλληλα προσαρμοσμένους EM αλγορίθμους για εκπαίδευση GMM και HMM μοντέλων.

Η πρακτική χρησιμότητα της προτεινόμενης προσέγγισης αναδεικνύεται στο πλαίσιο της εφαρμογής οπτικής-ακουστικής αναγνώρισης ομιλίας, όπως περιγράφεται στην Ενότητα Α.4. Παρόμοια με την προσέγγιση του [MCB⁺02], η εξαγωγή οπτικών χαρακτηριστικών ομιλίας βασίζεται σε αυτόματη ανάλυση του προσώπου του ομιλητή με την τεχνική των *ενεργών μοντέλων όψης* (active appearance models (AAM)) [CET01]. Μια σημαντική

καινοτομία στο δικό μας σύστημα εξαγωγής οπτικών χαρακτηριστικών είναι ένας μηχανισμός προσαρμογής στον ομιλητή ικανός να αγνοεί την εγγενή διαφορετικότητα στην όψη μεταξύ των προσώπων διαφορετικών ομιλητών που βρίσκονται σε ηρεμία, η οποία είναι άσχετη με τα λεγόμενα. Με αυτόν τον τρόπο το AAM μπορεί να συγκεντρώσει την περιγραφική του ισχύ στην μεταβλητότητα της όψης που οφείλεται στις εκφράσεις του προσώπου κατά την εκφορά του λόγου. Στο πλαίσιο της ανάλυσης ομιλίας καλούμε το συγκεκριμένο μοντέλο *φωνητικό ενεργό μοντέλο όψης* (visemic AAM). Δείχνουμε επίσης πώς μπορεί να υπολογιστεί η αβεβαιότητα στη μέτρηση των οπτικών χαρακτηριστικών AAM παράλληλα με τη διαδικασία ταιριάσματος του μοντέλου στο πρόσωπο. Για την εξαγωγή των ακουστικών χαρακτηριστικών βασιζόμαστε στην πρόσφατα προταθείσα τεχνική του [DDA05] που οδηγεί σε εκτίμηση τόσο των *εμπλουτισμένων* (enhanced) ακουστικών χαρακτηριστικών, όσο και της συνοδού αβεβαιότητας σε ένα ενιαίο πλαίσιο. Εντάσσουμε την τεχνική αυτή στο θεωρούμενο πολυτροπικό σενάριο που μας ενδιαφέρει, δείχνοντας ότι είναι και εκεί εφαρμόσιμη. Αξιολογούμε τις προτεινόμενες τεχνικές σε πειράματα AV-ASR με χρήση πολυροϊκών HMM, όπου παίρνουμε βελτιωμένες επιδόσεις αναγνώρισης. Συνδυάζοντας την τεχνική μας με ασύγχρονες αρχιτεκτονικές σύμμειξης παίρνουμε επιπλέον βελτιώσεις.

Α'.2 Αβεβαιότητα Χαρακτηριστικών και Πολυτροπική Σύμμειξη

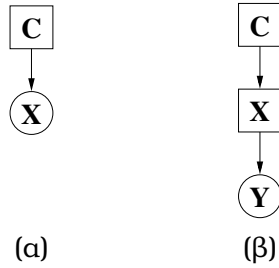
Ας θεωρήσουμε το κλασικό σενάριο κατηγοριοποίησης προτύπων υπό το οποίο μετράμε μια ιδιότητα (χαρακτηριστικό) ενός συγκεκριμένου δείγματος από το πρότυπο και επιθυμούμε να το κατηγοριοποιήσουμε σε μία από τις διαθέσιμες κλάσεις $c \in \{1, \dots, N\}$. Η μέτρηση μοντελοποιείται ως μία πραγμάτωση x του τυχαίου διανύσματος X , οι στατιστικές ιδιότητες του οποίου διαφέρουν για τις N τάξεις. Στην τυπική περίπτωση έχουμε εκπαιδεύσει ένα μοντέλο για καθεμιά τάξη που συλλαμβάνει τις στατιστικές της ιδιότητες στην υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας $p(x|c)$. Η απόφαση κατηγοριοποίησης λαμβάνεται στη συνέχεια με βάση κατάλληλα επιλεγμένο κανόνα, π.χ., το maximum a posteriori (MAP) κριτήριο $\hat{c} = \operatorname{argmax} p(c|x) = \operatorname{argmax} p(x|c)p(c)$.

Α'.2.1 Αβεβαιότητα Παρατήρησης Χαρακτηριστικών και η Αντιστάθμισή της Κατά την Κατηγοριοποίηση

Μπορούμε να μοντελοποιήσουμε την αβεβαιότητα στη μέτρηση των χαρακτηριστικών θεωρώντας ότι οι διαθέσιμες μετρήσεις y δεν είναι παρά μια θορυβώδης/αλλοιωμένη εκδοχή του θεωρούμενου ως καθαρού χαρακτηριστικού x το οποίο δεν είναι προσβάσιμο. Πιο συγκεκριμένα υιοθετούμε το μοντέλο μέτρησης

$$Y = X + E \tag{A'.1}$$

και υποθέτουμε ότι η κατανομή του θορύβου $p(e)$ είναι γνωστή· το θεωρούμενο σενάριο απεικονίζεται γραφικά στο Σχ. Α'.1 και αντιστοιχεί στα μοντέλα *σφάλματος μέτρησης* (measurement error) στην στατιστική [Ful87]. Υπό αυτό το μοντέλο παρατήρησης, οι αποφάσεις κατηγοριοποίησης θα πρέπει να βασίζονται στην $p(c|y) \propto p(y|c)p(c)$, και έτσι πρέπει να υπολογιστεί η $p(y|c)$.



Σχήμα Α'.1: Γραφική αναπαράσταση εναλλακτικών σεναρίων μέτρησης χαρακτηριστικών. Οι κρυφές μεταβλητές περικλείονται από τετράγωνα και οι παρατηρούμενες από κύκλους. (α) Συμβατική θεώρηση όπου τα χαρακτηριστικά x παρατηρούνται απευθείας. (β) Εναλλακτική θεώρηση όπου παρατηρούμε θορυβώδη χαρακτηριστικά y .

Για τον καθορισμό της ζητούμενης κατανομής θορυβωδών χαρακτηριστικών $p(y|c)$, απαιτείται να απαλείψουμε με ολοκλήρωση την κρυφή μεταβλητή καθαρών χαρακτηριστικών x

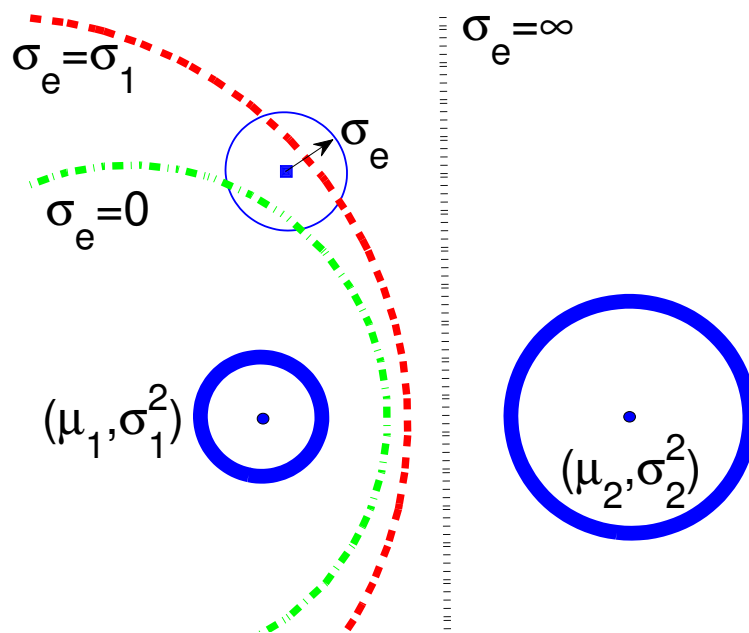
$$p(y|c) = \int p(y|x)p(x|c) dx. \quad (A'.2)$$

Αν και το ολοκλήρωμα στην Εξ. (Α'.2) είναι στη γενική περίπτωση δυσεπίλυτο, έχουμε στη διάθεσή μας τη λύση του σε κλειστή μορφή για τη σημαντική ειδική περίπτωση που ο μηχανισμός παραγωγής χαρακτηριστικών και ο μηχανισμός αλλοίωσης περιγράφονται από κατανομές Gauss, $p(x|c) = N(x; \mu_c, \Sigma_c)$ και $p(y|x) = N(y; x + \mu_e, \Sigma_e)$, αντίστοιχα, όπου $N(x; \mu, \Sigma)$ συμβολίζει την πολυμεταβλητή κανονική συνάρτηση πυκνότητας πιθανότητας για το x με μέσο μ και πίνακα συμμεταβλητότητας Σ . Μπορεί να δείξει κανείς τότε ότι το $p(y|c)$ δίνεται από

$$p(y|c) = N(y; \mu_c + \mu_e, \Sigma_c + \Sigma_e) \quad (A'.3)$$

που αντιστοιχεί στο να θεωρήσει κανείς τα χαρακτηριστικά y καθαρά, υπό την προϋπόθεση ότι μετακινούμε τους μέσους των μοντέλων μ_c κατά μ_e (βήμα *εμπλουτισμού*) και προσαιξάνουμε τους πίνακες συμμεταβλητότητας των μοντέλων Σ_c κατά Σ_e (βήμα *αντιστάθμισης αβεβαιότητας*). Παρόμοια προσέγγιση έχει ακολουθεί παλαιότερα σε μονοτροπικές εφαρμογές με εμπλοκή αποκλειστικά ακουστικών μετρήσεων [RHR94, YV02, DDA05].

Για να αναδείξουμε την Εξ. (Α'.3), συζητάμε με αναφορά στο Σχ. Α'.2 πώς η αβεβαιότητα παρατήρησης επηρεάζει τις αποφάσεις στο πλαίσιο ενός απλού παραδείγματος κατηγοριοποίησης σε δύο τάξεις. Οι δύο κλάσεις μοντελοποιούνται με 2-Δ σφαιρικές γκαουσιανές κατανομές, $N(\mu_1, \sigma_1^2 \mathbf{I})$, $N(\mu_2, \sigma_2^2 \mathbf{I})$ και θεωρούμε ότι είναι εκ των προτέρων ισοπίθανες. Εάν οι παρατηρήσεις y περιέχουν ιστροπικό θόρυβο Gauss με σφαιρικό πίνακα συμμεταβλητότητας $\sigma_e^2 \mathbf{I}$, τότε το παραλλαγμένο σύνορο απόφασης αποτελείται από το γεωμετρικό τόπο των σημείων y για τα οποία $N(y; \mu_1, \sigma_1^2 \mathbf{I} + \sigma_e^2 \mathbf{I}) = N(y; \mu_2, \sigma_2^2 \mathbf{I} + \sigma_e^2 \mathbf{I})$. Για μηδενικό σ_e^2 , έχουμε το αρχικό σύνορο απόφασης. Εάν το σ_e^2 είναι συγκρίσιμο της μεταβλητότητας των μοντέλων, το παραλλαγμένο σύνορο απόφασης διαφέρει σημαντικά από το αρχικό και έτσι αγνοώντας την αβεβαιότητα παρατήρησης κατά τη διαδικασία απόφασης αυξάνει τα λάθη κατηγοριοποίησης.



Σχήμα Α.2: Σύνορα απόφασης κατά την κατηγοριοποίηση μιας θορυβώδους παρατήρησης (τετράγωνο σημάδι) σε δύο τάξεις που δείχνονται ως κύκλοι, για διάφορες τιμές μεταβλητότητας θορύβου. Οι τάξεις μοντελοποιούνται ως σφαιρικές κανονικές κατανομές με μέσους μ_1 , μ_2 και μεταβλητότητα $\sigma_1^2 \mathbf{I}$, $\sigma_2^2 \mathbf{I}$, αντίστοιχα. Δείχνουμε το σύνορο απόφασης για τρεις τιμές της μεταβλητότητας θορύβου: (α) $\sigma_e = 0$ (δηλαδή απουσία θορύβου), (β) $\sigma_e = \sigma_1$, και (γ) $\sigma_e = \infty$. Με αυξανόμενη μεταβλητότητα θορύβου το σύνορο απόφασης απομακρύνεται από την αρχική του θέση.

Α'.2.2 Αβεβαιότητα Παρατήρησης και Πολυτροπική Σύμμειξη

Σε πολλές εφαρμογές μπορούμε να πετύχουμε αυξημένες επιδόσεις με αξιοποίηση συμπληρωματικών χαρακτηριστικών, εξαγμένων από μία ή περισσότερες τροπικότητες. Ας υποθέσουμε ότι θέλουμε να συνθέσουμε S ροές πληροφορίας που εκπέμπουν διανύσματα χαρακτηριστικών $\mathbf{x}_s, s = 1, \dots, S$. Με εφαρμογή του κανόνα του Bayes προκύπτει ότι η ύστερη πιθανότητα κλάσης δεδομένου του συνιστάμενου διανύσματος χαρακτηριστικών $\mathbf{x}_{1:S} \equiv (\mathbf{x}_1; \dots; \mathbf{x}_S)$ δίνεται από

$$p(c|\mathbf{x}_{1:S}) \propto p(c)p(\mathbf{x}_{1:S}|c). \quad (\text{A'.4})$$

Υποθέτοντας στατιστική ανεξαρτησία των διαφορετικών καναλιών χαρακτηριστικών δεδομένης της ετικέτας τάξης c , η υπο συνθήκη πιθανότητα του συνιστάμενου διανύσματος παρατηρήσεων $\mathbf{x}_{1:S}$ γίνεται διαχωρίσιμη και δίνεται από τον κανόνα γινομένου, που επιτρέπει στην Εξ. (A'.4) να γραφεί ως

$$p(c|\mathbf{x}_{1:S}) \propto p(c) \prod_{s=1}^S p(\mathbf{x}_s|c). \quad (\text{A'.5})$$

Αυτό το σενάριο καλείται από τους Clark & Yuille [CY90] *ελαφρά συζευγμένη* σύμμειξη.

Θα δείξουμε τώρα ότι λαμβάνοντας υπόψη την αβεβαιότητα μέτρησης χαρακτηριστικών οδηγεί αβίαστα σε έναν καινοτόμο μηχανισμό προσαρμοστικής σύμμειξης διαφορετικών πηγών πληροφορίας. Καθότι στο πλαίσιο των θορυβωδών μετρήσεων δεν έχουμε άμεση πρόσβαση στα χαρακτηριστικά \mathbf{x}_s , οι αποφάσεις λαμβάνονται με βάση τις θορυβώδεις εκδοχές τους $\mathbf{y}_s = \mathbf{x}_s + \mathbf{e}_s$. Υποθέτοντας ανεξαρτησία θορύβου στις διαφορετικές ροές, η πιθανότητα που μας ενδιαφέρει λαμβάνεται με απαλειφή της κρυφής μεταβλητής \mathbf{x}_s , δηλαδή,

$$p(c|\mathbf{y}_{1:S}) \propto p(c) \prod_{s=1}^S \int p(\mathbf{y}_s|\mathbf{x}_s)p(\mathbf{x}_s|c) d\mathbf{x}_s. \quad (\text{A'.6})$$

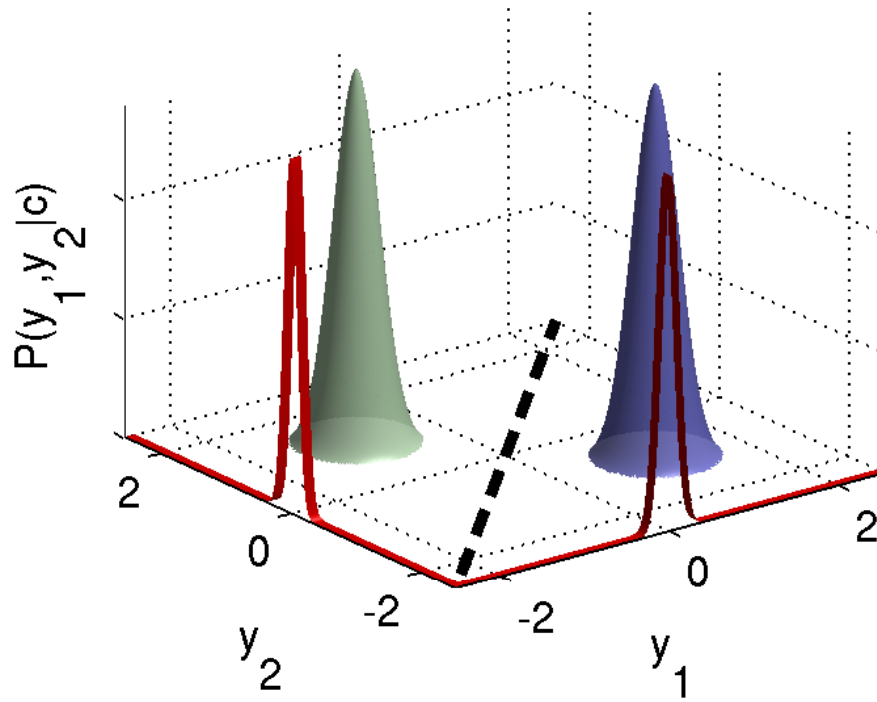
Για εκπομπή χαρακτηριστικών μοντελοποιημένη με GMM, *ι.ε.*, $p(\mathbf{x}_s|c) = \sum_{m=1}^{M_{s,c}} \rho_{s,c,m} N(\mathbf{x}_s; \boldsymbol{\mu}_{s,c,m}, \boldsymbol{\Sigma}_{s,c,m})$, και γκαουσιανό θόρυβο σε κάθε ροή, δηλαδή,

$$p(\mathbf{y}_s|\mathbf{x}_s) = N(\mathbf{y}_s; \mathbf{x}_s + \boldsymbol{\mu}_{e,s}, \boldsymbol{\Sigma}_{e,s}), \quad (\text{A'.7})$$

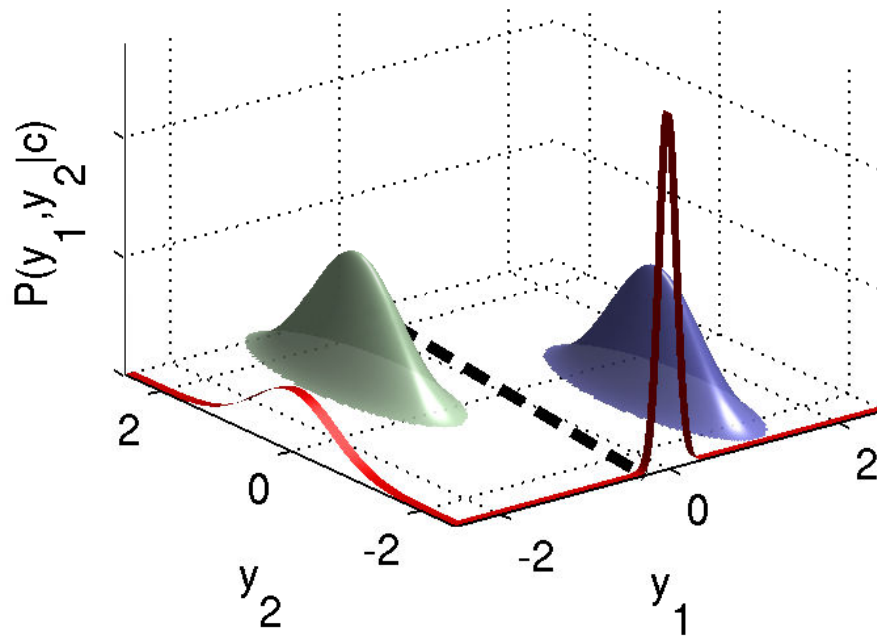
προκύπτει άμεσα ότι

$$p(c|\mathbf{y}_{1:S}) \propto p(c) \prod_{s=1}^S \sum_{m=1}^{M_{s,c}} \rho_{s,c,m} N(\mathbf{y}_s; \boldsymbol{\mu}_{s,c,m} + \boldsymbol{\mu}_{e,s}, \boldsymbol{\Sigma}_{s,c,m} + \boldsymbol{\Sigma}_{e,s}) \quad (\text{A'.8})$$

που, παρόμοια με την μονοτροπική περίπτωση (A'.3), μεταφράζεται σε θεώρηση των χαρακτηριστικών \mathbf{y}_s ως καθαρών, με μετατόπιση των μέσων του μοντέλου $\boldsymbol{\mu}_{s,c,m}$ κατά $\boldsymbol{\mu}_{e,s}$ και επαύξηση της συμμεταβλητότητας του μοντέλου $\boldsymbol{\Sigma}_{s,c,m}$ κατά $\boldsymbol{\Sigma}_{e,s}$. Σημειωτέον ότι ενώ ο θόρυβος μέτρησης είναι ο ίδιος για όλες τις κλάσεις c και επιμέρους συνιστώσες μείγματος m , επηρεάζονται περισσότερο οι πιο συγκεντρωμένες συνιστώσες για τις οποίες η συμμεταβλητότητα του θορύβου $\boldsymbol{\Sigma}_{e,s}$ είναι συγκρίσιμη με την αβεβαιότητα μοντελοποίησης $\boldsymbol{\Sigma}_{s,c,m}$. Ο προσαρμοστικός χαρακτήρας της σύμμειξης με αντιστάθμιση αβεβαιότητας επιδεικνύεται στο Σχ. Α'.3 στο πλαίσιο ενός σεναρίου ταξινόμησης σε δύο κλάσεις μέσω σύμμειξης πληροφορίας από δύο ροές.



(α)



(β)

Σχήμα Α.3: Η πολυτροπική αντιστάθμιση αβεβαιότητας οδηγεί σε προσαρμοστική σύμ-
 μειξη. Απεικονίζουμε ένα σενάριο ταξινόμησης σε δύο κλάσεις με χρήση δύο γκαουσιανών
 ροών χαρακτηριστικών, y_1 και y_2 , με ταυτόσημη μεταβλητότητα μοντέλου $\Sigma_{s,c} = \sigma^2 \mathbf{I}$. Η
 πυκνότητα πιθανότητας του θορύβου μέτρησης για κάθε κανάλι δείχνεται πάνω από τον
 αντίστοιχο άξονα, ενώ το σύνορο ταξινόμησης υποδηλώνεται με διακεκομμένη γραμμή. (α)
 Αμελητέος θόρυβος μέτρησης και στα δύο κανάλια: το σύνορο ταξινόμησης ταυτίζεται με
 τη διαγώνιο των αξόνων. (β) Έντονα αλλοιωμένες μετρήσεις στο y_2 κανάλι, $\sigma_{e,2} \gg \sigma_{e,1}$:
 το σύνορο ταξινόμησης μετακινείται με τρόπο που η απόφαση ταξινόμησης επηρεάζεται
 κατά κύριο λόγο από την τιμή του συγκριτικά πιο αξιόπιστου χαρακτηριστικού y_1 .

Α.3 Εκπαίδευση με Θορυβώδη Δεδομένα Χρησιμοποιώντας τον Αλγόριθμο EM

Σε πολλές περιπτώσεις έχουμε στη διάθεσή μας θορυβώδη δεδομένα εκπαίδευσης, που πρέπει να αξιοποιήσουμε για να εκπαιδεύσουμε τα μοντέλα μας. Συζητάμε πώς αυτό είναι εφικτό στα πλαίσια του Αλγορίθμου *μεγιστοποίησης αναμονής* (expectation maximization (EM)) για μοντέλα τύπου GMM και HMM.

Α.3.1 Θορυβώδης EM Εκπαίδευση GMM

Από την οπτική γωνία των αβέβαιων χαρακτηριστικών, παρατηρούμε μόνο μια θορυβώδη εκδοχή y των υποκείμενων πραγματικών ιδιοτήτων x . Η εκτίμηση των παραμέτρων θ του GMM μοντέλου με μεγιστοποίηση πιθανοφάνειας ενός συνόλου εκπαίδευσης $\mathcal{Y} = \{y_1, \dots, y_T\}$ μέσω του αλγορίθμου EM [DLR77] θα πρέπει συνεπώς να θεωρεί ως κρυφές μεταβλητές όχι μόνο τις ετικέτες μέλους κλάσης \mathcal{M} , αλλά επιπρόσθετα και τα καθαρά χαρακτηριστικά \mathcal{X} . Ο υπολογισμός κατά το **Ε-βήμα** της αναμενόμενης ολικής πιθανοφάνειας $Q(\theta, \theta') = E[\log p(\mathcal{Y}, \{\mathcal{X}, \mathcal{M}\} | \theta) | \mathcal{Y}, \theta']$ των παραμέτρων θ στην παρούσα επανάληψη του EM αλγορίθμου, δεδομένης της προηγούμενης εκτίμησης θ' , θα πρέπει συνεπώς να εμπεριέχει άθροιση την διακριτών και ολοκλήρωση των συνεχών κρυφών μεταβλητών. Στη μονοροϊκή περίπτωση αυτό μεταφράζεται σε

$$Q(\theta, \theta') = \sum_{t=1}^T \sum_{m=1}^M \log \pi_m p(m | y_t, \theta') + \sum_{t=1}^T \sum_{m=1}^M \int \log p(y_t | x_t) p(x_t, m | y_t, \theta') dx_t + \sum_{t=1}^T \sum_{m=1}^M \int \log p(x_t | m, \theta) p(x_t, m | y_t, \theta') dx_t. \quad (\text{A.9})$$

Λαμβάνουμε τις ανανεωμένες τιμές των παραμέτρων θ κατά το **Μ-βήμα** μεγιστοποιώντας την $Q(\theta, \theta')$ ως προς το θ , που δίνει

$$\pi_m = \frac{r_m}{T}, \quad \text{οπότε} \quad r_m = \sum_{t=1}^T r_{t,m}, \quad (\text{A.10})$$

$$\mu_m = \frac{1}{r_m} \sum_{t=1}^T r_{t,m} \hat{x}_{t,m} \quad (\text{A.11})$$

$$\Sigma_m = \frac{1}{r_m} \sum_{t=1}^T r_{t,m} (\Sigma_{x_{t,m}} + (\hat{x}_{t,m} - \mu_m)(\hat{x}_{t,m} - \mu_m)^T), \quad (\text{A.12})$$

όπου (με τόνο δηλώνουμε εκτιμήσεις από το προηγούμενο βήμα)

$$r_{t,m} = p(m | y_t, \theta') \propto \pi'_m N(y_t; \mu'_m + \mu_{e,t}, \Sigma'_m + \Sigma_{e,t}) \quad (\text{A.13})$$

$$\hat{x}_{t,m} = \Sigma_{x_{t,m}} ((\Sigma'_m)^{-1} \mu'_m + (\Sigma_{e,t})^{-1} (y_t - \mu_{e,t})) \quad (\text{A.14})$$

$$\Sigma_{x_{t,m}} = ((\Sigma'_m)^{-1} + (\Sigma_{e,t})^{-1})^{-1}. \quad (\text{A.15})$$

Ο προκύπτον EM αλγόριθμος έχει κάποιες αξιοσημείωτες διαφορές σε σχέση με την κλασική μη θορυβώδη εκδοχή του. Ειδικότερα, στον υπολογισμό των πιθανοτήτων ανάθεσης $r_{t,m}$ κατά το E -βήμα στην Εξ. (Α'.13), χρησιμοποιούμε αντισταθμισμένα σκορ. Επίσης, στην ανανέωση των μέσων και των πινάκων μεταβλητότητας κατά το M -βήμα στις Εξς. (Α'.11) και (Α'.12), θα πρέπει κανείς να αντικαταστήσει κάθε θορυβώδη μέτρηση y_t που χρησιμοποιείται κατά τη συμβατική εκπαίδευση του GMM μοντέλου με επαυξημένες της εκδοχές, που περιγράφονται από τις αναμενόμενες τιμές $\hat{x}_{t,m}$ και αντίστοιχες αβεβαιότητες $\Sigma_{x_{t,m}}$. Ειδικότερα, η αβεβαιότητα επαύξεσης $\Sigma_{x_{t,m}}$ υπεισέρχεται στην Εξ. (Α'.12) και κανονικοποιεί τον υπολογισμό του πίνακα συμμεταβλητότητας του μοντέλου Σ_m .

Α'.3.2 Θορυβώδης EM Εκπαίδευση HMM

Η παραπάνω ανάλυση επεκτείνεται εύκολα και στο HMM μοντέλο. Για HMM με συνεχείς παρατηρήσεις που μοντελοποιούνται με M γκαουσιανά μείγματα ανά κατάσταση, η ολική πιθανοφάνεια $Q(\theta, \theta') = E[\log p(\mathcal{Y}, \{Q, \mathcal{X}, \mathcal{M}\} | \theta) | \mathcal{Y}, \theta']$ των παραμέτρων θ στο παρόν βήμα του EM αλγορίθμου δεδομένης της προηγούμενης εκτίμησης θ' λαμβάνεται στο E -βήμα ως:

$$\begin{aligned}
Q(\theta, \theta') = & \sum_{q \in \mathcal{Q}} \sum_{t=1}^T \log a_{q_{t-1}q_t} P(\mathcal{Y}, q | \theta') + \\
& \sum_{q \in \mathcal{Q}} \sum_{t=1}^T \int \log p(\mathbf{y}_t | \mathbf{x}_t, q_t, \theta') P(\mathcal{Y}, q, \mathbf{x}_t | \theta') d\mathbf{x}_t + \\
& \sum_{q \in \mathcal{Q}} \sum_{t=1}^T \sum_{m=1}^M \int \log p(\mathbf{x}_t | m_t, q_t, \theta') P(\mathcal{Y}, q, m, \mathbf{x}_t | \theta') d\mathbf{x}_t + \\
& \sum_{q \in \mathcal{Q}} \sum_{t=1}^T \sum_{m=1}^M p(m | q_t, \theta') P(\mathcal{Y}, q, m | \theta') + \sum_{q \in \mathcal{Q}} \log \pi_{q_0} P(\mathcal{Y}, q | \theta'). \quad (\text{Α'.16})
\end{aligned}$$

Οι πιθανότητες ανάθεσης $\gamma_t(i, k) = p(q_t = i, m = k | \mathbf{y}_{1:t}) \propto a_t(i) \beta_t(i)$ εκτιμώνται με τον *μπρος-πίσω* (forward-backward) αλγόριθμο [Rab89]. προσαρμοσμένο ώστε να χρησιμοποιεί σκορ αντιστάθμιση αβεβαιότητας:

$$a_{t+1}(j) = P(\mathbf{y}_{1:t}, q_t = j | \theta') = \left[\sum_{i=1}^N \alpha_{ij} a_t(i) \right] b'_j(\mathbf{y}_{t+1}) \quad (\text{Α'.17})$$

$$\beta_t(i) = P(\mathbf{y}_{t+1:T} | q_t = i, \theta') = \sum_{j=1}^N \alpha_{ij} b'_j(\mathbf{y}_{t+1}) \beta_{t+1}(j) \quad (\text{Α'.18})$$

όπου $b'_j(\mathbf{y}_t) = \sum_{m=1}^M \rho_m N(\mathbf{y}_t; \boldsymbol{\mu}'_{j,m} + \boldsymbol{\mu}_{e_t}, \boldsymbol{\Sigma}'_{j,m} + \boldsymbol{\Sigma}_{e_t})$. Η αποτίμηση της πιθανότητας μιας ακολουθίας δίνεται ως συνήθως από $P(\mathbf{y}_{1:T} | \theta) = \sum_{i=1}^N a_T(i)$. Οι ανανεωμένες παράμετροι θ εκτιμώνται μέσα από σχέσεις παρόμοιες με την GMM περίπτωση της Ενότητας Α'.3.

Α.4 Συνδυασμένη Οπτική-Ακουστική Αναγνώριση Ομιλίας

Μια απαιτητική εφαρμογή για την αξιολόγηση τεχνικών σύμμειξης πολυτροπικής πληροφορίας είναι η αυτόματη αναγνώριση ομιλίας με συνδυασμό οπτικής και ακουστικής πληροφορίας (AV-ASR), καθότι απαιτεί επαρκή μοντελοποίηση τόσο της σχετικής αξιοπιστίας, όσο και του συγχρονισμού των δύο καναλιών πληροφορίας.

Α.4.1 Εξαγωγή και Εκτίμηση Αβεβαιότητας Οπτικών Χαρακτηριστικών Ομιλίας

Οπτική πληροφορία ομιλίας μπορεί να εξαχθεί από τους εμφανείς αρθρωτές του ομιλητή, κυρίως τα χείλη και τη σιαγόνα, που αποτελούν την οπτική *περιοχή ενδιαφέροντος* (region of interest (ROI)) γύρω από το στόμα [PNGG03]. Η οπτική πληροφορία μπορεί να διακριθεί σε γεωμετρικά χαρακτηριστικά σχήματος και σε πληροφορία υψής που αντιστοιχεί στους χρωματικούς τόνους που δειγματοληπτούμε στην περιοχή ενδιαφέροντος.

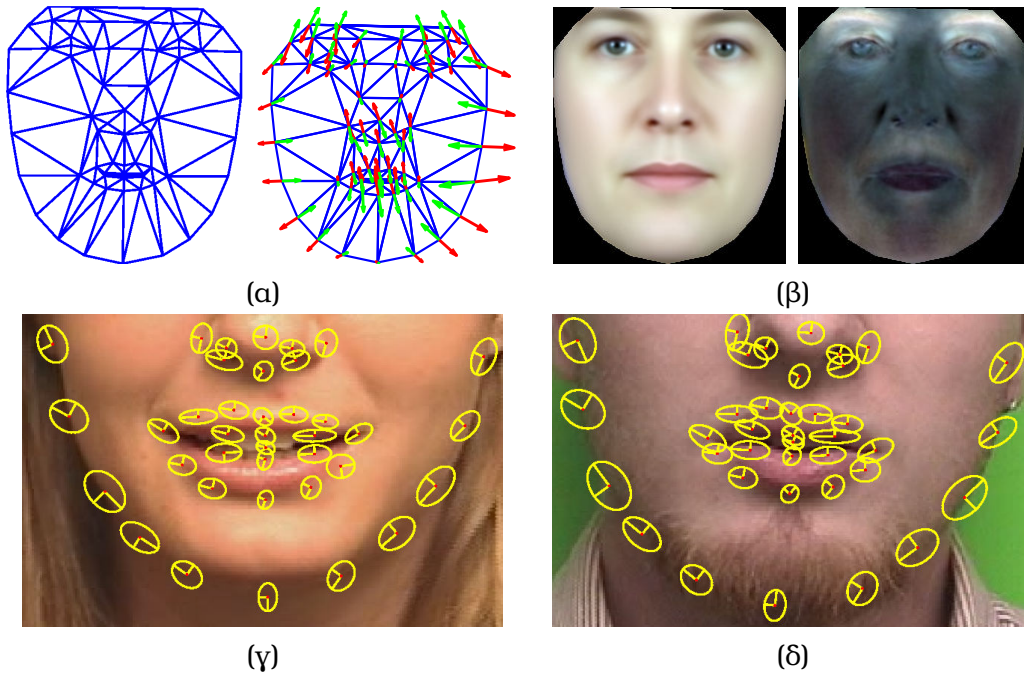
Χρησιμοποιούμε *ενεργά μοντέλα όψης* (active appearance models (AAMs)) [CET01] για να παρακολουθήσουμε με ακρίβεια το πρόσωπο του ομιλητή και να εξάγουμε οπτικά χαρακτηριστικά ομιλίας από αυτό. Τα ενεργά μοντέλα όψης, που πρωτοχρησιμοποιήθηκαν για AV-ASR στο [MCB⁺02], είναι αναγεννητικά μοντέλα της εμφάνισης αντικειμένων και έχουν αποδειχθεί ιδιαίτερα αποδοτικά στη μοντελοποίηση ανθρώπινων προσώπων σε εφαρμογές όπως αναγνώριση ή παρακολούθηση προσώπων. Η χαρακτηριστικότερη διαφορά τους σε σχέση με απλούστερες μεθόδους μετασχηματισμών τύπου DCT/PCA/DWT/ICA εφαρμοσμένων στις ανεπεξέργαστες τιμές της εικόνας έγκειται στο γεγονός ότι τα AAM μοντελοποιούν ρητά και ξεχωρίζουν τη μεταβλητότητα στο σχήμα και στην υφή του προσώπου [PNGG03]. Ειδικότερα, στα AAM το σχήμα του αντικειμένου μοντελοποιείται με μία μάσκα καθοριζόμενη από ένα σύνολο σημείων-κλειδιών $\{z_i, i = 1, \dots, L\}$, των οποίων οι συντεταγμένες ορίζουν ένα διάνυσμα σχήματος s μήκους $2L$. Επιτρέπουμε παρεκκλίσεις από το μέσο σχήμα s_0 επιτρέποντας στο s να κείται σε έναν n -Δ γραμμικό υπόχωρο 'ιδιοσχημάτων', δηλαδή

$$s = s_0 + \sum_{i=1}^n p_i s_i. \quad (\text{A'.19})$$

Η παραμόρφωση του σχήματος s από το μέσο σχήμα s_0 ορίζει μία απεικόνιση $z \mapsto \mathbf{W}(z; \mathbf{p})$, όπου z συμβολίζει τυχόν σημείο στο εσωτερικό του μέσου σχήματος, η οποία φέρνει το συγκεκριμένο παράδειγμα σχήματος στο παρόν πλαίσιο εικόνας I σε αντιστοίχιση με τη μέση τεμπλέτα προσώπου. Αφού αντισταθίσουμε την παραμόρφωση σχήματος, η υφή του αντιστοιχισμένου προσώπου μπορεί να μοντελοποιηθεί ως σταθμισμένο άθροισμα m 'ιδιοϋφών' $\{A_i\}$, δηλαδή,

$$I(\mathbf{W}(z; \mathbf{p})) = A_0(z) + \sum_{i=1}^m \lambda_i A_i(z) \quad (\text{A'.20})$$

όπου A_0 είναι το μέσο διάνυσμα υψής του προσώπου. Οι βάσεις ιδιοσχημάτων/ιδιοϋφών καθορίζονται μέσα από διαδικασία εκπαίδευσης με χρήση Το μέσο σχήμα/υφή και το



Σχήμα Α'.4: Υποσύστημα εξαγωγής οπτικών χαρακτηριστικών. (α) Μέσο σχήμα s_0 και το πρώτο ιδιοσχήμα s_1 , που υποδηλώνεται με τα βέλη να δείχνουν τη διαφοροποίηση σε σχέση με το μέσο σχήμα. (β) Μέσο διάνυσμα υφής A_0 και η πρώτη ιδιοϋφή A_1 . (γ & δ) Παραδείγματα ταιριάσματος της AAM μάσκας και υπολογισμού της αβεβαιότητας εντοπισμού χαρακτηριστικών στο πρόσωπο.

πρώτο ιδιοδιάνυσμα σχήματος/υφής εκπαιδευμένο με αυτή τη διαδικασία απεικονίζονται στο Σχ. Α'.4.

Έχοντας ένα εκπαιδευμένο AAM μοντέλο, το ταίριασμά του σε μια εικόνα προσώπου I έγκειται στην εύρεση εκείνων των παραμέτρων σχήματος και υφής $\mathbf{q} \equiv \{\mathbf{p}, \boldsymbol{\lambda}\}$ που ελαχιστοποιούν το κανονικοποιημένο κόστος ανακατασκευής

$$f(\mathbf{q}) = \frac{\kappa}{2\sigma^2} \|E(\mathbf{q})\|^2 + Q(\mathbf{q}) \quad (\text{Α'.21})$$

όπου $E(\mathbf{q}) \equiv I(\mathbf{W}(\mathbf{p})) - A_0 - \sum_{i=1}^m \lambda_i A_i$ είναι η εικόνα με το σφάλμα ανακατασκευής υφής του μοντέλου, σ^2 είναι η μεταβλητότητα του σφάλματος ανακατασκευής, $Q(\mathbf{q}) = \frac{1}{2}(\mathbf{q} - \mathbf{q}_0)^T \boldsymbol{\Sigma}_{\mathbf{q},0}^{-1}(\mathbf{q} - \mathbf{q}_0)$ είναι ένας τετραγωνικός όρος ποινής που αντιστοιχεί σε γκαουσιανή πρότερη κατανομή για τους συντελεστές με μέσο \mathbf{q}_0 και πίνακα συμμεταβλητότητας $\boldsymbol{\Sigma}_{\mathbf{q},0}$, και κ είναι θετική σταθερά που ρυθμίζει το μερίδιο των δύο όρων στο τελικό κριτήριο. Πραγματοποιούμε το ταίριασμα του μοντέλου με ιδιαίτερα αποδοτικούς αλγορίθμους που παρουσιάζουμε στο [PM08]. Ο πίνακας συμμεταβλητότητας $\boldsymbol{\Sigma}_{\mathbf{q}}$ στην εκτίμηση ελαχίστων τετραγώνων για το \mathbf{q} σχετίζεται με τον Hessian πίνακα της συνάρτησης σφάλματος $f(\mathbf{q})$, αποτιμημένο στο σημείο ελαχίστου [PTVF92, ζη. 15] και μπορεί να υπολογιστεί αποδοτικά κατά τη διάρκεια της διαδικασίας ταιριάσματος [PM08].

Μια ενδιαφέρουσα διάσταση του οπτικού συστήματος εξαγωγής πληροφοριών σε σχέση με προηγούμενα παρόμοια συστήματα βασισμένα σε AAM [MPNL01, MCB⁺02] είναι ότι χρησιμοποιούμε μια διάταξη δύο AAM. Το πρώτο full-face AAM εκτείνεται σε όλο

το πρόσωπο, καλύπτοντας την επιφάνεια που φαίνεται στα Σχ. Α'.4(α,β), και μπορεί να παρακολουθήσει (track) με αξιοπιστία τον ομιλητή σε μακρές ακολουθίες βίντεο. Όμως καθώς η σχετιζόμενη με την ομιλία οπτική πληροφορία περιορίζεται κατά κύριο λόγο στην περιοχή γύρω από το στόμα, για εξαγωγή οπτικών χαρακτηριστικών ομιλίας χρησιμοποιούμε ένα δεύτερο ROI-AAM που καλύπτει τη συγκεκριμένη περιοχή ενδιαφέροντος, όπως φαίνεται στο Σχ. Α'.4(γ,δ), και το οποίο αναλύει το σχήμα και την υφή της. Καθότι το ROI-AAM καλύπτει σχετικά μικρή περιοχή ώστε να επιτρέψει αξιόπιστη παρακολούθηση, το τοποθετούμε στην περιοχή ενδιαφέροντος με τη βοήθεια του full-face AAM. Ως οπτικό χαρακτηριστικό ομιλίας σε κάθε καρέ του βίντεο I_t χρησιμοποιούμε το διάνυσμα παραμέτρων ανάλυσης q_t του ROI-AAM, αξιοποιώντας και τις συνοδές εκτιμήσεις αβεβαιότητας. Παραδείγματα των αντίστοιχων θέσεων των σημείων κλειδιών και η αβεβαιότητα εντοπισμού τους δείχνονται στο Σχ. Α'.4.

Καθότι ενδιαφερόμαστε για αναγνώριση ομιλίας με μοντέλα ανεξάρτητα από την ταυτότητα του ομιλητή, η εξαγωγή οπτικών χαρακτηριστικών με καλές ιδιότητες αμεταβλητότητας ως προς τον ομιλητή αποτελεί βασικό ζητούμενο κατά το σχεδιασμό του συστήματός μας. Τα ενεργά μοντέλα όψης εκπαιδευμένα με την κλασική προσέγγιση του [CET01] σε επισημειωμένες βάσεις πολλαπλών προσώπων, όπως έχει γίνει στο [MCB⁺02], υστερούν σε αυτόν τον τομέα, καθότι η ερμηνευτική ισχύς τους αντί να αξιοποιείται για την περιγραφή των εκφράσεων προκαλούμενες από την ομιλία, αναλώνεται στην αναπαράσταση της εκτεταμένης μεταβλητότητας εμφάνισης που προκαλείται από τη διαφορετική ταυτότητα των ομιλητών. Η *αφαίρεση του μέσου διανύσματος χαρακτηριστικών* (feature mean subtraction) [PNGG03] μπορεί σε περιορισμένο μόνο βαθμό να απαλύνει αυτήν την αδυναμία γιατί δεν μπορεί να αναιρέσει το γεγονός ότι οι κύριοι ρυθμοί της PCA ανάλυσης που επιλέγονται κατά τη διάρκεια της εκπαίδευσης αφορούν κυρίως την ταυτότητα του ομιλητή και δευτερευόντως τα λεχθέντα. Για να αντιμετωπίσουμε αυτά τα ζητήματα, προβλέπουμε στο προτεινόμενο AAM μοντέλο το μέσο σχήμα s_0 και η μέση υφή A_0 να είναι εξαρτώμενα από τον ομιλητή. Στην πράξη, κατά την εκπαίδευση του ROI-AAM μοντέλου αφαιρούμε εξαρτώμενα από τον ομιλητή (αντί για καθολικά) μέσο σχήμα και υφή. Επίσης παραλλάσσουμε το στάδιο εξαγωγής των AAM χαρακτηριστικών αφαιρώντας μια εκτίμηση του μέσου σχήματος και υφής του προσώπου κάθε ομιλητή πριν την ανάλυση με το ROI-AAM. Στο πλαίσιο της εφαρμογής AV-ASR καλούμε το προκύπτον μοντέλο *φωνητικό ενεργό μοντέλο όψης* (visemic AAM), καθότι οι κυρίαρχοι ρυθμοί μεταβλητότητας σχήματος και υφής σχετίζονται άμεσα με οπτική ομιλία και έχουν συνεπώς σχετική ανοσία στη μεταβλητότητα μεταξύ ομιλητών. Μια παρόμοια προσέγγιση έχει υιοθετηθεί σε συνδυασμό με απλούστερες τεχνικές εξαγωγής οπτικών χαρακτηριστικών βασισμένες σε μετασχηματισμούς της εικόνας [Luc03], αλλά η απουσία άμεσου ελέγχου της παραμόρφωσης του σχήματος του προσώπου την καθιστά λιγότερο αποδοτική από τα AAM.

Α'.4.2 Εξαγωγή Ακουστικών Χαρακτηριστικών και Εκτίμηση της Αβεβαιότητάς τους

Με ορισμένες εξαιρέσεις, π.χ., [DL00], οι περισσότερες εργασίες σε AV-ASR μελετούν τη βελτίωση στην επίδοση σε σύγκριση με σχετικά απλά συστήματα βασισμένα μόνο σε ακουστική πληροφορία. Καθότι η AV-ASR εμπνέεται κυρίως από εφαρμογές σε θορυβώδη ακουστικά περιβάλλοντα, είναι σημαντικό να εξετάσουμε την απόδοσή τους σε σύγκριση με πιο προχωρημένα συστήματα αναγνώρισης που εμπεριέχουν μηχανισμούς ευρωστίας

έναντι του θορύβου.

Από την εκτεταμένη πρόσφατη βιβλιογραφία σε αναγνώριση φωνής με αποκλειστικά ακουστική πληροφορία έχουμε ενσωματώσει στο σύστημά μας την τεχνική του [DDA05]. Η προσέγγιση του [DDA05] ταιριάζει ιδιαίτερα στο πλαίσιο εργασίας μας καθώς περιλαμβάνει τόσο τον εμπλουτισμό των χαρακτηριστικών ομιλίας όσο και τον υπολογισμό της αβεβαιότητάς τους σε ένα ενιαίο πλαίσιο. Για περαιτέρω περιγραφή της τεχνικής παραπέμπουμε στη δημοσίευσή μας [PKPM09], όσο και στην εργασία [DDA05].

Α'4.3 Σύγχρονες και Ασύγχρονες Αρχιτεκτονικές Σύμμειξης Ακουθιακών Δεδομένων

Αν και η ως τώρα συζήτηση έχει επικεντρωθεί έχει επικεντρωθεί στη σύμμειξη με απλά μοντέλα GMM για περιγραφή στατικών δεδομένων και πολυροϊκά μαρκοβιανά μοντέλα σύγχρονης κατάστασης στατε-σφνσηρονους Μυλτιστρεαμ-HMM για περιγραφή δυναμικών δεδομένων, το προτεινόμενο σχήμα αντιστάθμισης αβεβαιότητας είναι πολύ πλατύτερα εφαρμόσιμο. Συγκεκριμένα, είναι συμβατό με πολύ γενικότερες αρχιτεκτονικές περιγραφής ασύγχρονων ροών δεδομένων που χρησιμοποιούνται στη μοντελοποίηση οπτικής-ακουστικής ομιλίας. Στα AV-ASR πειράματα της Ενότητας Α'5 περιλαμβάνουμε αξιολόγηση της προτεινόμενης τεχνικής σύμμειξης σε συνδυασμό με Product-HMM. Για περισσότερες πληροφορίες παραπέμπουμε στη δημοσίευσή μας [PKPM09].

Α'5 Πειραματική Αξιολόγηση

Η αποδοτικότητα του προτεινόμενου σχήματος σύμμειξης μέσω αντιστάθμισης της αβεβαιότητας έχει αποτιμηθεί σε σειρά πειραμάτων οπτικής-ακουστικής αναγνώρισης φωνής που παρουσιάζουμε στη συνέχεια.

Α'5.1 Βάση Δεδομένων και Μεθοδολογία Αξιολόγησης

Χρησιμοποιήσαμε την οπτικό-ακουστική βάση δεδομένων του Πανεπιστημίου του Clemson (CUAVE) [PGTG02] πάνω στη οποία πραγματοποιήσαμε πειράματα κατηγοριοποίησης ψηφίων. Τα πειράματα πραγματοποιήθηκαν στο 'Κανονικό' τμήμα της βάσης που αποτελείται από οπτικό-ακουστικές καταγραφές 36 (17 γυναικών και 19 ανδρών) ομιλητών με τον καθένα να έχει εκφωνήσει σπιακερς 50 απομονωμένα Αγγλικά ψηφία. Οι ομιλητές σε αυτό το τμήμα της βάσης δεδομένων κοιτούν την κάμερα και στέκονται σχετικά ακίνητοι. Οι καταγραφές του βίντεο έχουν πραγματοποιηθεί κάτω από καλές συνθήκες φωτισμού σε ανάλυση 720x480 εικονοστοιχείων και σε συχνότητα δειγματοληψίας 29.97 Hz· ένα αντιπροσωπευτικό καρέ εικόνας από κάθε ομιλητή φαίνεται στο Σχ. Α'5. Για τα πειράματα σε συνθήκες θορύβου, οι ηχητικές καταγραφές στο υποσύνολο που χρησιμοποιείται για δοκιμή έχουν αλλοιωθεί από προσθετικό babble θόρυβο από την βάση θορύβου NOISEX-92, σε διάφορα επίπεδα σηματοθορυβικού λόγου SNR.

Η απόδοση της αναγνώρισης εκτιμάται σε δεδομένα από 6 ομιλητές, ενώ οι καταγραφές των υπολοίπων 30 ομιλητών έχουν χρησιμοποιηθεί για την εκπαίδευση των μοντέλων για τα ψηφία. Επειδή η βάση CUAVE είναι σχετικά μικρή συγκρινόμενη με συλλογές από μόνο ακουστικές καταγραφές, έχουμε πραγματοποιήσει όλα τα πειράματα μας



Σχήμα Α'.5: Ενδεικτικά καρέ από όλα τα 36 υποκείμενα της βάσης δεδομένων CUAVE.

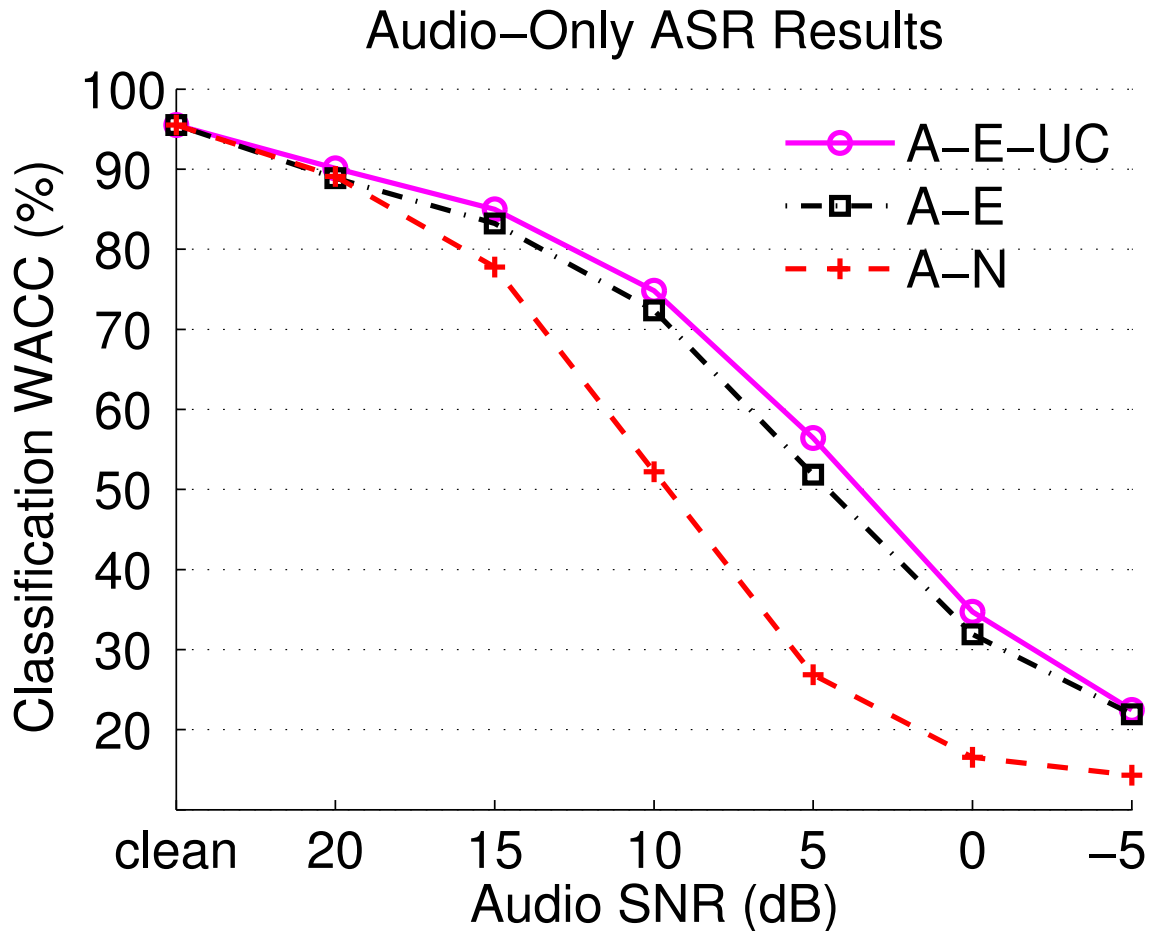
πολλαπλές φορές χρησιμοποιώντας κάθε φορά διαφορετικά τμήματα της βάσης για δοκιμή/εκπαίδευση ώστε να αυξήσουμε την στατιστική σημαντικότητα των αποτελεσμάτων μας. Πιο συγκεκριμένα χωρίσαμε την βάση μας σε 6 μη-επικαλυπτόμενα υποσύνολα, με το καθένα να αντιστοιχεί σε 6 ομιλητές μιας γραμμής του Σχήματος Α'.5. Στη συνέχεια χρησιμοποιήσαμε κάθε υποσύνολο ως το σύνολο της δοκιμής, εκπαιδεύοντας τα μοντέλα στα εναπομείναντα 5 υποσύνολα. Η διαδικασία αυτή οδηγεί σε 6 επαναλήψεις των πειραμάτων μας σε ανεξάρτητα σύνολα δοκιμής. Τα ακουστικά- και οπτικά-μόνο αποτελέσματα αναγνώρισης που αναφέρουμε στην Ενότητα Α'.5.2 αποτελούν τον μέσο όρο αυτών των 6 επαναλήψεων, ενώ για τα αποτελέσματα της οπτικό-ακουστικής αναγνώρισης της Ενότητας Α'.5.3 έχουμε διατηρήσει το πρώτο υποσύνολο για τον καθορισμό των καλύτερων βαρών για τις ροές (σφραγισ) και για αυτό το λόγο τα αναφερθέντα αποτελέσματα αποτελούν την μέση τιμή των υπολοίπων 5 επαναλήψεων.

Ως εξαγόμενα χαρακτηριστικά για τον ήχο, χρησιμοποιούμε τις λογαριθμικές ενέργειες που μετρώνται στο ηχητικό σήμα στην έξοδο του από μια συστοιχία φίλτρων τοποθετημένα με γραμμικό τρόπο στην συχνοτική κλίμακα Mel (αναπαράσταση τύπου FBANK). Συγκεκριμένα, εξάγουμε 26 συντελεστές FBANK από πλαίσια του ηχητικού σήματος μήκους 25 ms που έχουν παραθυρωθεί με παράθυρο Hamming και στα οποία έχει εφαρμοσθεί προέμφαση (συντελεστής: 0.97) σε συχνότητα 100 Hz. Ως οπτικά χαρακτηριστικά χρησιμοποιούμε τους AAM συντελεστές της ROI-περιοχής του στόματος των φωνητικών AAM, που υπολογίζονται όπως περιγράφεται στην Ενότητα Α'.4.1. Για να υπάρχει συμφωνία με το ρυθμό των ηχητικών πλαισίων, τα οπτικά χαρακτηριστικά έχουν υπερδειγματοληπτηθεί από το πραγματικό ρυθμό του βίντεο των 29.97 Hz στα 100 Hz μέσω απλής γραμμικής παρεμβολής. Σε όλα μας τα πειράματα, οι παράμετροι της παραγωγής και της επιτάχυνσης συνοδεύουν τόσο τα ηχητικά όσο και τα οπτικά χαρακτηριστικά. Επιπλέον σε όλες τις περιπτώσεις κάνουμε χρήση κρυφών Μαρκοβιανών μοντέλων πλήρους ψηφίου και τύπου δεξιά-προς-αριστερά, με το κάθε ένα να έχει 8 καταστάσεις και ένα μοναδικό διαγώνιο πίνακα συνμεταβλητότητας για την Γκαουσιανή κατανομή των παρατηρήσεων για κάθε ροή και κάθε κατάσταση. Όλα τα μοντέλα έχουν εκπαιδευθεί σε σήματα φωνής που δεν έχουν αλλοιωθεί από θόρυβο πριν αυτά δοκιμασθούν σε διαφορετικές συνθήκες θορύβου. Τα πειράματα μας πραγματοποιήθηκαν με τη χρήση του HMM Toolkit (HTK) [YEK⁺02], το οποίο τροποποιήσαμε προκειμένου να υλοποιήσουμε το σχήμα σύμμιξης μέσω της αντιστάθμισης της αβεβαιότητας.

Α'.5.2 Μονο-Τροπικά Πειράματα Αναγνώρισης Φωνής

Αρχικά παρουσιάζουμε πειράματα αναγνώρισης φωνής χρησιμοποιώντας είτε μόνο ακουστικά χαρακτηριστικά είτε μόνο οπτικά χαρακτηριστικά για να εξετάσουμε την σχετική απόδοση των διαφορετικών ηχητικών και οπτικών συστημάτων εξαγωγής χαρακτηριστικών.

Ξεκινούμε με ένα πείραμα κατηγοριοποίησης κάνοντας χρήση μόνο ακουστικών χαρακτηριστικών. Το πείραμα αυτό εξετάζει την απόδοση στη δοκιμασία της αναγνώρισης, της τεχνικής του εμπλουτισμού του σήματος φωνής και της αντιστάθμισης της αβεβαιότητας που περιγράφεται στην Ενότητα Α'.4.2. Εφαρμόζοντας τη μέθοδο του [DDA05], χρησιμοποιήσαμε καθαρό σήμα φωνής, μείγμα από 50 GMM για τα στατικά FBANK χαρακτηριστικά, που εκπαιδεύτηκε σε όλες τις καθαρές καταγραφές της βάσης CUAVE. Η σύγκριση αφορά στη χρήση των ανεπεξέργαστων θορυβωδών ηχητικών χαρακτηριστικών



Σχήμα Α.6: Αποτελέσματα κατηγοριοποίησης ψηφίων με μόνο ηχητικά χαρακτηριστικά για διάφορα επίπεδα babble θορύβου. Η σύγκριση γίνεται μεταξύ των ανεπεξέργαστων θορυβωδών ηχητικών χαρακτηριστικών (A-N), των εμπλουτισμένων ηχητικών χαρακτηριστικών (A-E) και των εμπλουτισμένων ηχητικών χαρακτηριστικών που αποκωδικοποιούνται με αντιστάθμιση της αβεβαιότητας (A-E-UC).

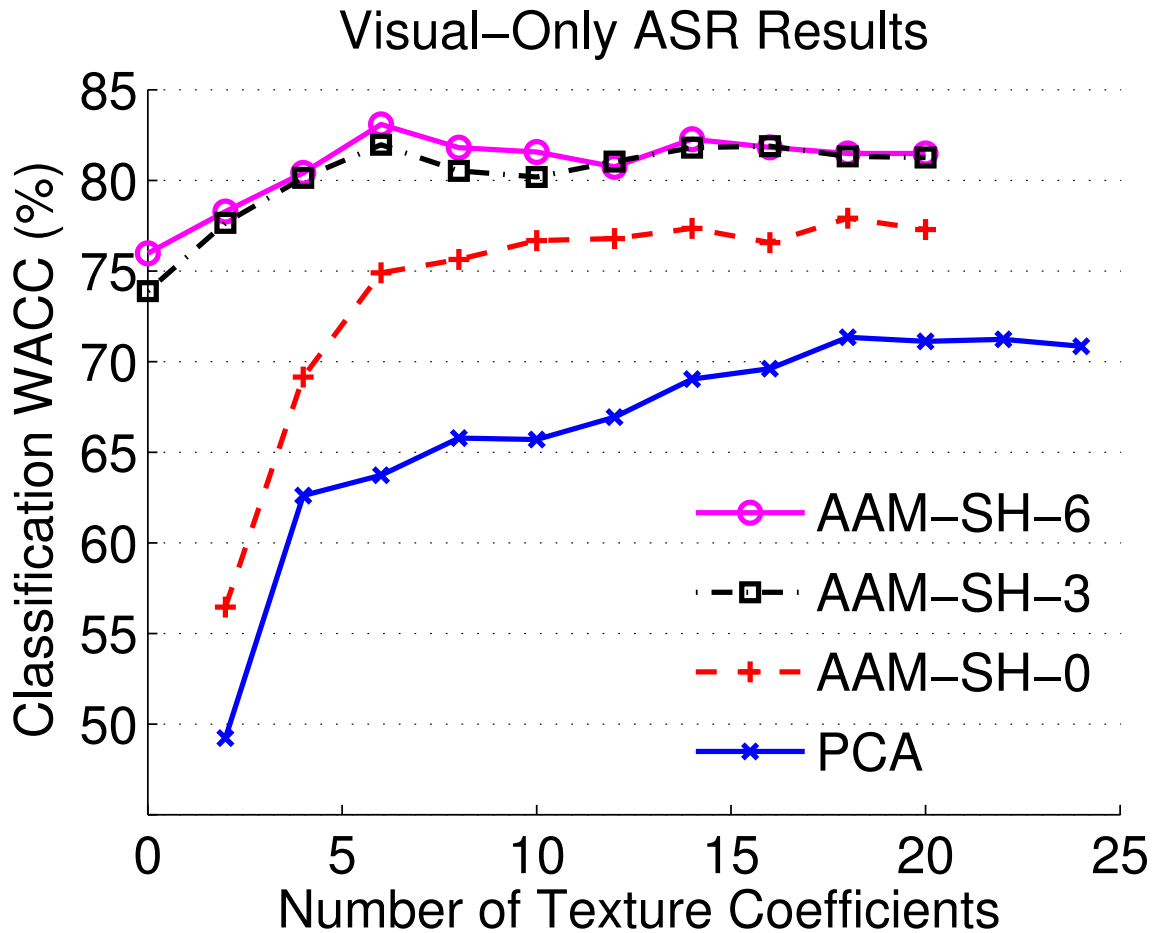
(A-N) και των εμπλουτισμένων ηχητικών χαρακτηριστικών (A-E και A-E-UC). Οι εκτιμήσεις της αβεβαιότητας που προκύπτουν από τη διαδικασία του εμπλουτισμού αγνοούνται στη συμβατική αποκωδικοποίηση (A-E), ενώ ενσωματώνονται στη διαδικασία απόφασης στην αποκωδικοποίηση με την αντιστάθμιση της αβεβαιότητας (A-E-UC). Η σύνοψη των αποτελεσμάτων στο Σχ. Α.6 καταδεικνύει ότι η χρήση των ανεπεξέργαστων θορυβωδών ηχητικών χαρακτηριστικών οδηγεί σε πολύ φτωχή απόδοση αναγνώρισης σε χαμηλά επίπεδα σηματοθορυβικού λόγου. Η χρήση των εμπλουτισμένων ηχητικών χαρακτηριστικών είναι συνεπώς αποφασιστική για την διατήρηση της καλής απόδοσης, ενώ η αντιστάθμιση της αβεβαιότητας προσφέρει μια επιπρόσθετη σημαντική βελτίωση. Για παράδειγμα στα 5 dB SNR, η ακρίβεια λέξης (WACC) μετά των εμπλουτισμό αυξάνεται περίπου κατά 25% σε απόλυτα μεγέθη, ενώ η αντιστάθμιση της αβεβαιότητας δίνει ένα περαιτέρω κέρδος της τάξης του 5%. Σε όλα μας τα οπτικό-ακουστικά πειράματα αυτόματης αναγνώρισης φωνής που αναφέρονται στη συνέχεια θα κάνουμε χρήση του συνόλου των εμπλουτισμένων ηχητικών χαρακτηριστικών.

Στη συνέχεια εξετάζουμε τη σχετική απόδοση για διαφορετικά συστήματα εξαγωγής οπτικών χαρακτηριστικών σε πειράματα στα οποία χρησιμοποιούμε μόνο οπτικά χαρακτηριστικά. Προκειμένου να συγκρίνουμε την προτεινόμενη τεχνική που βασίζεται στα φωνητικά AAM χαρακτηριστικά με εναλλακτικές μεθόδους εξαγωγής οπτικών χαρακτηριστικών, έχουμε επίσης εξάγει οπτικά PCA χαρακτηριστικά από την ίδια περιοχή ενδιαφέροντος του στόματος. Ο εντοπισμός για τις AAM και PCA μάσκες παρέχεται από τα AAM του πλήρους προσώπου. Το μέσο σχήμα και υφή του AAM, όπως και η μέση υφή της PCA τεχνικής εξαγωγής χαρακτηριστικών ανανεώνονται για κάθε ομιλητή όπως περιγράφεται στην Ενότητα Α.4.1, προκειμένου να αυξήσουμε την ανεξαρτησία των εξαγόμενων χαρακτηριστικών από τους ομιλητές. Στο Σχ. Α.7 συνοψίζουμε τα αποτελέσματα που λάβαμε από τις δύο εναλλακτικές μεθόδους για μεταβαλλόμενο αριθμό συντελεστών υφής που διατηρούμε. Για την περίπτωση των AAM παρέχουμε τρεις αναπαραστάσεις που αντιστοιχούν στη διατήρηση 0, 3, και 6 συντελεστών σχήματος. Τα προτεινόμενα φωνητικά AAM με 6 συντελεστές σχήματος και 6 συντελεστές υφής παρουσιάζουν συνολικά την καλύτερη απόδοση (83% WACC), ενώ η μέγιστη απόδοση της τεχνικής που βασίζεται στην εξαγωγή PCA χαρακτηριστικών είναι 71% και επιτεύχθηκε για 18 συντελεστές υφής. Αυτό που είναι ιδιαίτερα αξιοσημείωτο είναι η δυνατότητα αποδοτικής αναγνώρισης που παρουσιάζουν τα μοντέλα που βασίζονται στα φωνητικά AAM κάνοντας χρήση μόλις λίγων παραμέτρων. Για παράδειγμα, χρησιμοποιώντας μόνο 3 AAM συντελεστές σχήματος και κανέναν υφής λαμβάνουμε 74% WACC, το οποίο ξεπερνά την απόδοση του PCA μοντέλου των 18-συντελεστών· αυτό πρέπει να οφείλεται στην αυξημένη ακρίβεια του προτεινόμενου φωνητικού AAM αλγορίθμου προσαρμογής ομιλητή. Η δουλειά μας είναι η πρώτη που παρουσιάζει ανώτερη απόδοση οπτικό-ακουστικής αναγνώρισης για τα AAM χαρακτηριστικά. Στην προηγούμενη μελέτη του [MPNLO1], στην οποία AAM χαρακτηριστικά είχαν υποδεέστερη απόδοση από απλούστερα PCA χαρακτηριστικά, χρησιμοποιούνταν AAM πλήρους προσώπου και για την ανάλυση του προσώπου αλλά και για την παρακολούθηση του, ενώ δεν εφαρμοζόταν κανένας μηχανισμός για ανεξαρτησία ομιλητή. Το προτεινόμενο διαδοχικό ζεύγος των AAM (ένα για εύρωστη παρακολούθηση και ένα για ανάλυση της ROI περιοχής του στόματος) και ο προτεινόμενος μηχανισμός, που βασίζεται στα φωνητικά AAM, για ανεξαρτησία ομιλητή φαίνεται πως αντιμετωπίζουν αποδοτικά τα μειονεκτήματα των προηγούμενων AAM τεχνικών για οπτικό-ακουστική αναγνώριση και υποδεικνύουν ότι προσεγγίσεις με μοντέλα της όρασης υπολογιστών μπορεί να είναι ιδιαίτερα αποδοτικές για την εξαγωγή οπτικών από το πρόσωπο χαρακτηριστικών.

Τα οπτικό-ακουστικά πειράματα που αναφέρονται στη συνέχεια κάνουν χρήση του συνόλου των οπτικών φωνητικών AAM 6-σχήμα/6-υφή χαρακτηριστικών που παρουσιάζουν και την καλύτερη απόδοση.

Α.5.3 Πειράματα Συνδυασμένης Οπτικής-Ακουστικής Αναγνώρισης

Έχοντας μελετήσει την απόδοση κάθε τροπικότητας ξεχωριστά, παρουσιάζουμε στη συνέχεια το κύριο σύνολο οπτικό-ακουστικών πειραμάτων αναγνώρισης φωνής εξετάζοντας την απόδοση του σχήματος σύμμειξης με αντιστάθμιση της αβεβαιότητας, τόσο με αλλά και χωρίς χρήση βαρών στις ροές των χαρακτηριστικών. Σε όλα τα πειράματα χρησιμοποιούνται τα εμπλουτισμένα ηχητικά χαρακτηριστικά. Στο Σχ. Α.8 αποτυπώνουμε γραφικά την απόδοση των καλύτερων αποτελεσμάτων με μόνο ηχητικά χαρακτηριστικά



Σχήμα Α.7: Αποτελέσματα κατηγοριοποίησης ψηφίων με μόνο AAM και PCA οπτικά χαρακτηριστικά για μεταβαλλόμενο αριθμό συντελεστών υψής. Για τα AAM χαρακτηριστικά δείχνουμε επίσης πως η απόδοση της κατηγοριοποίησης εξαρτάται από τον αριθμό των συντελεστών σχήματος.

και κάνοντας χρήση της αντιστάθμισης της αβεβαιότητας (A-UC) (αντιστοιχούν στις ετικέτες A-E-UC του Σχ. Α.6) και στα καλύτερα αποτελέσματα με μόνο οπτικά χαρακτηριστικά (") και τα συγκρίνουμε με την απόδοση των τεσσάρων εκδοχών των multistream-HMM οπτικό-ακουστικών σύγχρονων-καταστάσεων μοντέλων σύμμιξης: οπτικό-ακουστικά με ίδια βάρη για τις δύο ροές και συμβατική αποκωδικοποίηση (AV)· οπτικό-ακουστικά με ίδια βάρη για τις δύο ροές και αποκωδικοποίηση με αντιστάθμιση αβεβαιότητας (AV-UC)· οπτικό-ακουστικά με βέλτιστα βάρη (AV-W)· και οπτικό-ακουστικά με βέλτιστα βάρη αποκωδικοποίηση με αντιστάθμιση αβεβαιότητας (AV-W-UC).

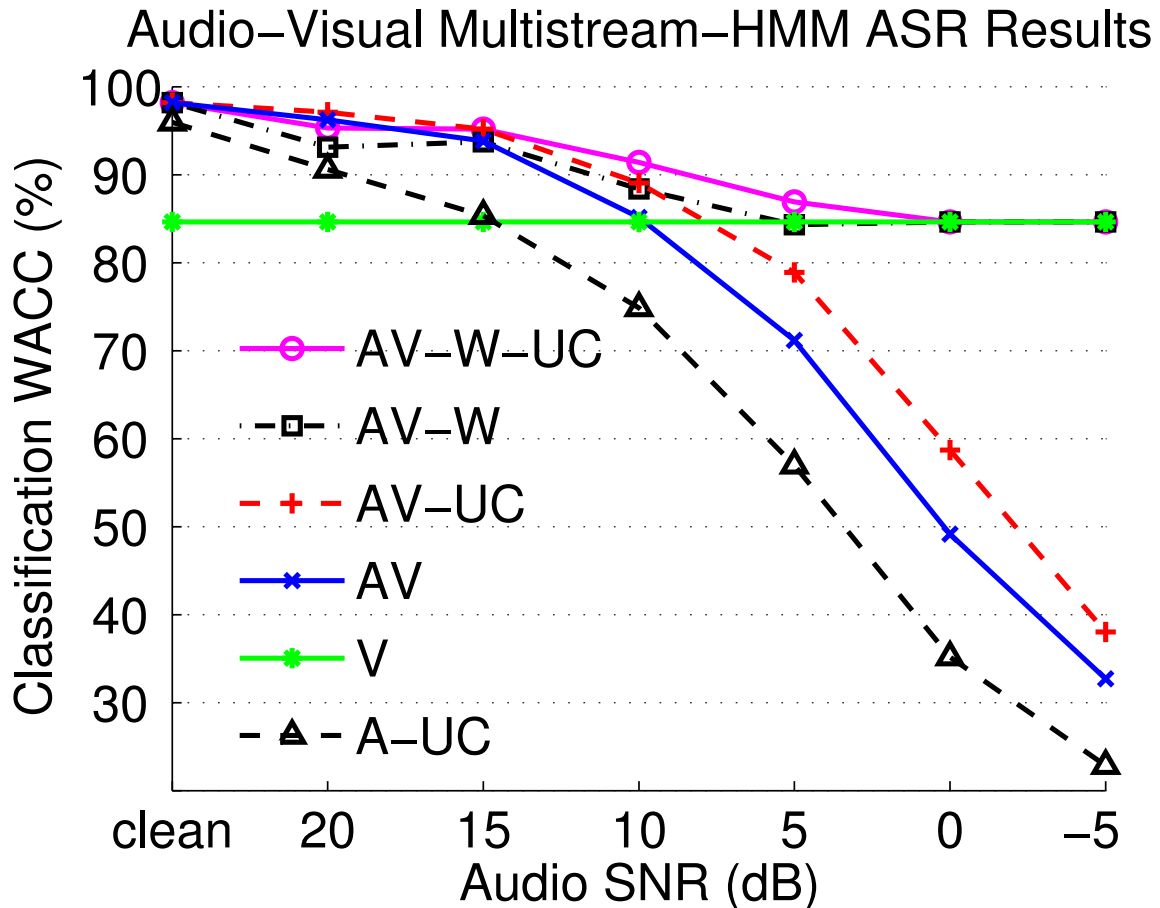
Προκειμένου να αναδείξουμε την βελτίωση της απόδοσης λόγω της αντιστάθμισης της αβεβαιότητας από μόνη της, παρουσιάζουμε στο Σχ. Α.9 τη σχετική μείωση στο ρυθμό λανθασμένων λέξεων (WER) όταν συγκρίνουμε το AV-UC με το AV (χωρίς βάρη στις ροές) και το AV-W-UC με το AV-W (με βάρη στις ροές). Η σχετική μείωση του WER δίνεται από $(WER_{AV} - WER_{AV-UC})/WER_{AV}$. Όπως περιγράφεται στην Ενότητα Α.5.1, όλα τα αποτελέσματα προκύπτουν ως μέσες τιμές για διαφορετικές επαναλήψεις των πειραμάτων με ανεξάρτητα υποσύνολα ελέγχου. Για τα πειράματα που περιλαμβάνουν βάρη

χρησιμοποιήσαμε εκθέτες στις ροές που αθροίζουν σε 1 και ψάχνουν εξαντλητικά σε κάθε επίπεδο θορύβου για το βάρος των ηχητικών χαρακτηριστικών μεταξύ του 0.0 και του 1.0 (με βήματα του 0.1) τα οποία οδήγησαν στα καλύτερα αποτελέσματα σε μια συγκεκριμένη επανάληψη του πειράματος (αποτελούμενη από τους έξι πρώτους ομιλητές ως σύνολο ελέγχου). Το καλύτερο βάρος για την ηχητική ροή για τα επίπεδα θορύβου 0 και -5 dB προέκυψε ότι ήταν το 0.0, που σημαίνει ότι οι αντίστοιχες AV-W και AV-W-UC τιμές στο Σχ. Α'8 ταυτίζονται με το αποτέλεσμα του οπτικού μόνο αποτελέσματος. Επειδή η βελτίωση εξαιτίας της αντιστάθμισης της αβεβαιότητας έχει μόνο νόημα όταν και οι δύο ροές είναι ενεργές, τα επίπεδα θορύβου 0 ανθ -5 dB στο Σχ. Α'9(β) έχουν προκύψει αφού θέσαμε πρώτα το βάρος της ηχητικής ροής ίσο με 0.1, *ι.ε.*, η μικρότερη θετική τιμή.

Συγκρίνοντας τα αποτελέσματα των AV και AV-UC, βλέπουμε ότι η σύμμειξη με αντιστάθμιση της αβεβαιότητας παρέχει μια συστηματική βελτίωση για όλες τις ακουστικές συνθήκες (4.8% μέση απόλυτη βελτίωση WACC ή 20.9% σχετική μείωση του WER που προκύπτει ως μέσος όρος για όλες τις καταστάσεις θορύβου) σε σχέση με την συμβατική αποκωδικοποίηση. Όμοια, συστηματική βελτίωση εμφανίζεται όταν συνδυάζουμε την αποκωδικοποίηση με την αβεβαιότητα μαζί με τη χρήση βαρών στις ροές (2.3% μέση απόλυτη βελτίωση WACC ή 19.4% σχετική μείωση του WER που προκύπτει ως μέσος όρος για όλες τις καταστάσεις θορύβου), όπως μπορεί να διαπιστωθεί συγκρίνοντας το AV-W με το AV-W-UC. Τα βάρη στις ροές είναι αναγκαία για να διατηρείται η απόδοση της οπτικό-ακουστικής αναγνώρισης υψηλότερα από την απόδοση της οπτικής αναγνώρισης σε συνθήκες χαμηλού σηματοθορυβικού λόγου· αυτό μπορεί να αποδοθεί σε μια υπερ-εκτίμηση της εμπιστοσύνης στην εκτίμηση των χαρακτηριστικών από την μέθοδο εμπλουτισμού του ηχητικού σήματος. Τα καλύτερα οπτικό-ακουστικά multistream-HMM αποτελέσματα του Σχ. Α'8 προκύπτουν από το AV-W-UC σχήμα που βελτιώνει το WACC σε σχέση με το καλύτερο αποτέλεσμα αναγνώρισης με μόνο ηχητικά χαρακτηριστικά (A-UC) κατά ένα απόλυτο ποσοστό 28.7% κατά μέσο όρο για όλα τα 6 επίπεδα θορύβου.

Προκειμένου να αυξήσουμε την εμπιστοσύνη μας για την στατιστική σημαντικότητα των βελτιωμένων λόγω της αποκωδικοποίησης με αντιστάθμιση της αβεβαιότητας οπτικό-ακουστικών αποτελεσμάτων, δείχνουμε στο Σχ. Α'9 όχι μόνο την μέση σχετική μείωση του WER, αλλά και τα ξεχωριστά αποτελέσματα για κάθε μια από τις 5 επαναλήψεις των πειραμάτων στα μη-επικαλυπτόμενα σύνολα ελέγχου. Τέτοιες συγκρίσεις για πολλές επαναλήψεις πειραμάτων επιτρέπουν την εξαγωγή ασφαλέστερων στατιστικών συμπερασμάτων σχετικά με την σχετική απόδοση των ανταγωνιζόμενων τεχνικών, επειδή η διακύμανση των αποτελεσμάτων εξαιτίας των διαφορών μεταξύ των ομιλητών μειώνεται [GC89, BN04, KBW06]. Παρατηρούμε ότι η βελτίωση στην πολυτροπική σύμμειξη λόγω της αποκωδικοποίησης με αντιστάθμιση της αβεβαιότητας είναι συστηματική σε όλες τις επαναλήψεις των πειραμάτων στα ανεξάρτητα σύνολα ελέγχου, τόσο όταν χρησιμοποιούμε βάρη στις ροές όσο και όταν δεν χρησιμοποιούμε. Το γεγονός αυτό ενδυναμώνει την στατιστική εγκυρότητα των ισχυρισμών μας.

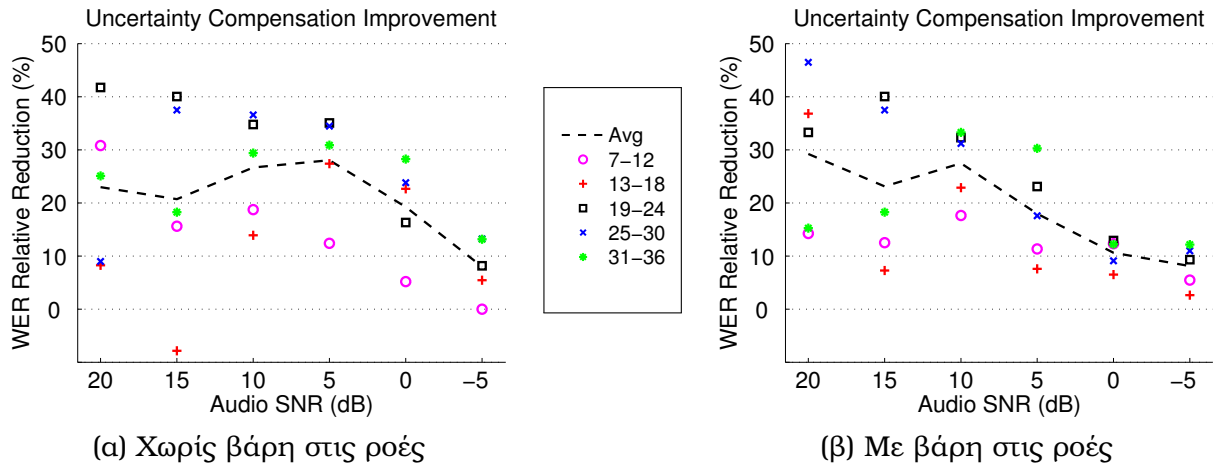
Το τελευταίο μας πείραμα ερευνά την απόδοση της αποκωδικοποίησης με την αβεβαιότητα σε σχέση με τα Product-HMMs, που, όπως συζητήθηκε στην Ενότητα Α'4.3, λαμβάνουν καλύτερα υπόψιν τους φαινόμενα ασυγχρονίας μεταξύ τους ηχητικού και οπτικού σήματος. Στο Σχ. Α'10 δείχνουμε αποτελέσματα του Product-HMM με συμβατική αποκωδικοποίηση (AV-P) και αποκωδικοποίηση με την αβεβαιότητα (AV-P-UC), τις εκδοχές τους χρησιμοποιώντας βάρη στις ροές των χαρακτηριστικών (AV-P-W και AV-P-W-UC), καθώς και το αποτέλεσμα που προέκυψε από το Multistream-HMM σύγχρονων-καταστάσεων με



Σχήμα Α.8: Οπτικό-ακουστική Multistream-HMM κατηγοριοποίηση ψηφίων για διάφορα επίπεδα babble θορύβου. Παρουσιάζουμε αποτελέσματα ακρίβειας λέξεων για τις παρακάτω μεθόδους: εμπλουτισμένα ηχητικά χαρακτηριστικά με αντιστάθμιση αβεβαιότητας (A-UC)· μόνο οπτικά (V)· μόνο ηχητικά (AV)· οπτικό-ακουστικά με αντιστάθμιση της αβεβαιότητας (AV-UC)· οπτικό-ακουστικά με βάρη (AV-W)· και οπτικό-ακουστικά με βάρη και με αντιστάθμιση της αβεβαιότητας (AV-W-UC). Σε όλα τα πειράματα που εμπειρέχουν ηχητικό σήμα έχουμε εμπλουτίσει τα ηχητικά χαρακτηριστικά. Τα μοντέλα ενεργών περιγραμμάτων χρησιμοποιήθηκαν για την οπτική τροπικότητα.

συμβατική αποκωδικοποίηση (AV) ως αποτέλεσμα αναφοράς. Η χρήση της αποκωδικοποίησης με την αβεβαιότητα δίνει ένα (μέση τιμή για όλα τα επίπεδα θορύβου) αβсолютο κέρδος $\epsilon\lambda\omega\alpha^{\circ}\Sigma$ σε απόλυτη τιμή της τάξης του 5.0% για την περίπτωση που χρησιμοποιούμε μοντέλα με τα ίδια βάρη (AV-P-UC vs. AV-P) και 0.6% όταν χρησιμοποιούμε βάρη στις ροές (AV-P-W-UC vs. AV-P-W). Η μέση απόλυτη WACC βελτίωση των Product-HMMs έναντι των Multistream-HMMs είναι 1.0% όταν χρησιμοποιούμε συμβατική αποκωδικοποίηση και 1.2% με αποκωδικοποίηση με την αβεβαιότητα. Συνολικά, τα καλύτερα αποτελέσματα οπτικό-ακουστικής αναγνώρισης τα λάβαμε με το AV-P-W-UC μοντέλο.

Όλα τα αναφερόμενα αποτελέσματα δείχνουν μια συστηματική βελτίωση των αποτελεσμάτων αναγνώρισης όταν γίνεται χρήση της αντιστάθμισης της αβεβαιότητας κατά τη διάρκεια της αποκωδικοποίησης. Συγκεκριμένα είναι άξιο αναφοράς το γεγονός ότι η προσαρμοστική σύμμειξη με αντιστάθμιση της αβεβαιότητας ενσωματώνεται αβίαστα με

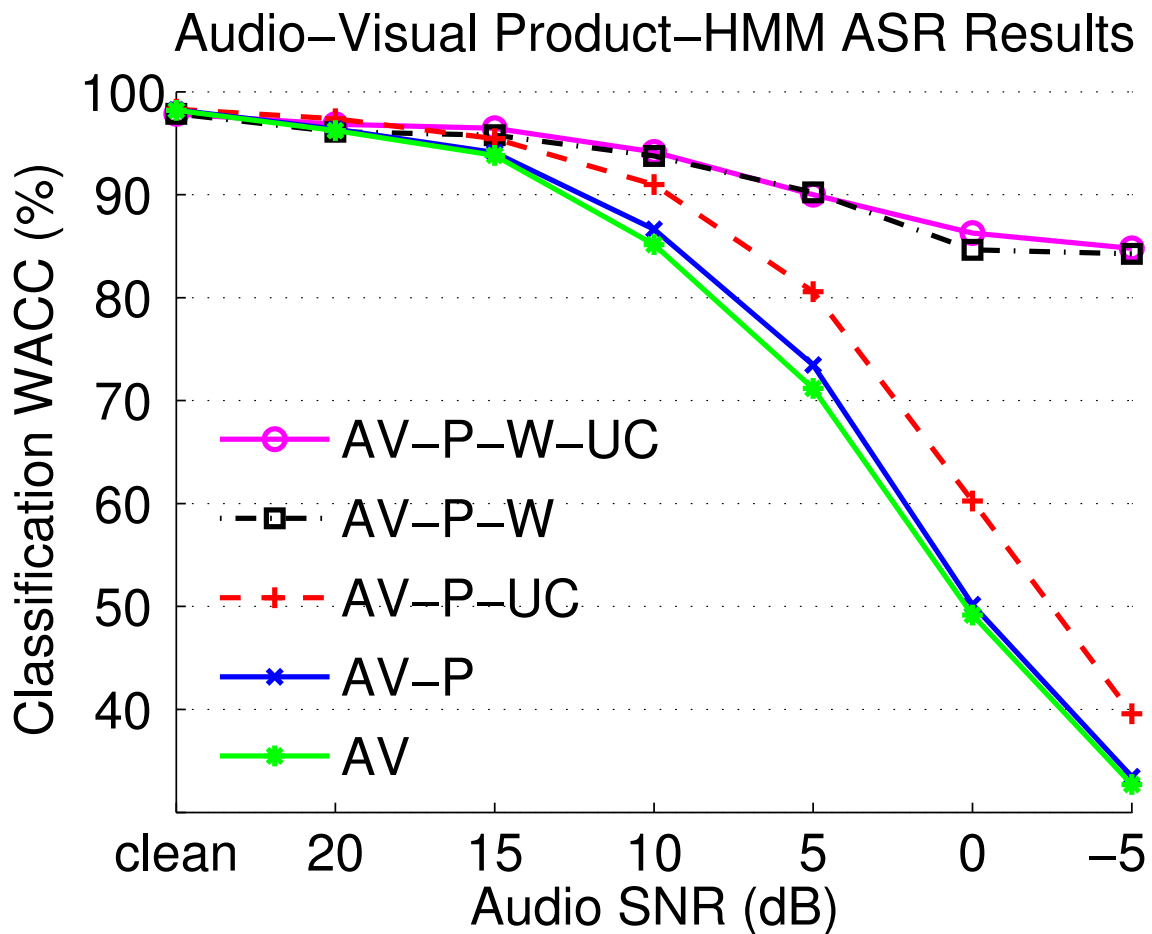


Σχήμα Α.9: Κέρδος στην απόδοση λόγω αντιστάθμισης της αβεβαιότητας για την οπτικό-ακουστική κατηγοριοποίηση ψηφίων για διάφορα επίπεδα babble θορύβου και τις 5 επαναλήψεις του πειράματος για διαφορετικά σύνολα ελέγχου. Δείχνουμε τη μείωση του ποσοστού του σχετικού σφάλματος λέξης όταν χρησιμοποιούμε αντιστάθμιση αβεβαιότητας. (α) Χωρίς βάρη στις ροές, *ι.ε.*, AV-UC έναντι του AV. (β) Με βάρη στις ροές, *ι.ε.*, AV-W-UC έναντι του AV-W. Σε όλες τις περιπτώσεις χρησιμοποιούνται τα εμπλουτισμένα ηχητικά χαρακτηριστικά.

τεχνικές πολυτροπικής ανάλυσης, όπως η χρήση βαρών στις ροές των χαρακτηριστικών ή τα Product-HMMs. Σε μια προγενέστερη δουλειά μας [PKPM07] έχουμε επίσης δείξει μια επιπλέον μικρή βελτίωση όταν λάβουμε υπόψιν εκτιμήσεις της αβεβαιότητας για τα οπτικά χαρακτηριστικά και κατά τη διάρκεια της εκπαίδευσης των μοντέλων. Η αντιστάθμιση της αβεβαιότητας αποδεικνύεται ένα ευέλικτο και αξιόπιστο εργαλείο για ένα ευρύ φάσμα από πλαίσια πολυτροπικής σύμμιξης.

Α.6 Συμπεράσματα

Σε αυτό το Κεφάλαιο παρουσιάσαμε ένα πρότυπο πλαίσιο εργασίας για πολυτροπική σύμμιξη μέσω αντιστάθμισης της αβεβαιότητας και παρουσιάσαμε την αποδοτικότητα του στη οπτικό-ακουστική αναγνώριση φωνής. Με δεδομένη μια εκτίμηση της αβεβαιότητας για κάθε ροή των χαρακτηριστικών, το προτεινόμενο πλαίσιο εργασίας μας οδηγεί με φυσικό τρόπο σε ιδιαίτερα προσαρμοστικούς κανόνες για πολυτροπική σύμμιξη που υλοποιούνται εύκολα και αποδοτικά. Η τεχνική μας μπορεί να χρησιμοποιηθεί ευρέως και να ενσωματωθεί είτε με σύγχρονες ή με ασύγχρονες πολυτροπικές ακολουθίες αρχιτεκτονικής ολοκλήρωσης που προκύπτουν σε τυπικές πολυτροπικές εφαρμογές. Επιπλέον δείξαμε ότι το σχήμα μας είναι συμβατό με την ευρέως χρησιμοποιούμενη διατύπωση των βαρών για τις ροές: ο συνδυασμός των δύο τεχνικών οδηγεί συστηματικά σε καλύτερα αποτελέσματα στα οπτικό-ακουστικά πειράματα που διεξήγαμε.



Σχήμα Α.10: Αποτελέσματα οπτικό-ακουστικής κατηγοριοποίησης ψηφίων με Product-HMM για διάφορα επίπεδα babble θορύβου. Δείχνουμε αποτελέσματα ακρίβειας αναγνώρισης λέξεων για τέσσερις (με βάρη) εκδοχές Product-HMM μοντέλων, δύο με την συμβατική αποκωδικοποίηση (AV-P και AV-P-W) και δύο με αποκωδικοποίηση λαμβάνοντας υπόψιν την αβεβαιότητα (AV-P-UC και AV-P-W-UC). Δίνονται επίσης τα αποτελέσματα της Multistream-HMM/συμβατικής αποκωδικοποίησης με σκοπό τη σύγκριση. Σε όλες τις περιπτώσεις έχουν χρησιμοποιηθεί τα εμπλουτισμένα ηχητικά χαρακτηριστικά και τα AAM οπτικά χαρακτηριστικά.

Παράρτημα Β΄

Δημοσιεύσεις Διατριβής

Αποτελέσματα της διδακτορικής μας έρευνας έχουν δημοσιευθεί σε διεθνώς αναγνωρισμένα περιοδικά και συνέδρια με κριτή.

Το κείμενο της διατριβής περιλαμβάνει ερευνητικά αποτελέσματα που έχουν δημοσιευθεί στις εξής εργασίες:

- Πολυπλεγματικοί αλγόριθμοι για μοντέλα μερικών διαφορικών εξισώσεων [J1, C1]. Σε συνεργασία με τον Π. Μαραγκό.
- Αυτόματη επιλογή κλίμακας μέσω στατιστικής διεπικύρωσης [C2, C3]. Σε συνεργασία με τον Π. Μαραγκό.
- Μοντελοποίηση φυσικών εικόνων στο πεδίο των κυματιδίων με εφαρμογή στο πρόβλημα της ενδοσυμπλήρωσης εικόνων [C8]. Σε συνεργασία με τους Π. Μαραγκό και Α. Kokaram.
- Τεχνικές για προσαρμοστική σύμμειξη οπτικής-ακουστικής πληροφορίας και εφαρμογή τους στην οπτική-ακουστική αναγνώριση ομιλίας [J2, C4, C5, C6, B1, B2]. Σε συνεργασία με τους Α. Κατσαμάνη, Β. Πιτσικάλη, και Π. Μαραγκό.

Πέρα από τα θέματα που παρουσιάζονται στην παρούσα αναφορά, η ερευνητική μας δραστηριότητα κινήθηκε και στις εξής περιοχές:

- Τεχνικές για ανάκτηση του σχήματος της φωνητικής οδού με χρήση οπτικής και ακουστικής πληροφορίας [J3, C7, C9, C11]. Σε συνεργασία με τους Α. Κατσαμάνη και Π. Μαραγκό.
- Αλγόριθμοι για παρακολούθηση προσώπου και εξαγωγή οπτικών χαρακτηριστικών από αυτό [C10]. Σε συνεργασία με τον Π. Μαραγκό.
- Πολυκλιμακωτές μέθοδοι για αποθορυβοποίηση εικόνας αλλοιωμένης από θόρυβο Poisson [J4, C12, C13]. Σε συνεργασία με τους Σ. Λευκιμμάτη και Π. Μαραγκό.

Ακολουθεί πλήρης κατάλογος των σχετικών δημοσιεύσεων. Ηλεκτρονικά ανάτυπα είναι διαθέσιμα από την ιστοσελίδα <http://cvsp.cs.ntua.gr/gpapan>.

Δημοσιεύσεις σε Περιοδικά

- J1. Papandreou, G. και P. Maragos: Multigrid geometric active contour models. *IEEE Transactions on Image Processing*, 16(1):229–240, Ιανουάριος 2007.
- J2. Papandreou, G., A. Katsamanis, V. Pitsikalis, και P. Maragos: Adaptive multimodal fusion by uncertainty compensation with application to audio-visual speech recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 17(3):423–435, Μάρτιος 2009.
- J3. Katsamanis, A., G. Papandreou, και P. Maragos: Face active appearance modeling and speech acoustic information to recover articulation. *IEEE Transactions on Audio, Speech and Language Processing*, 17(3):411–422, Μάρτιος 2009.
- J4. Lefkimmiatis, S., P. Maragos, και G. Papandreou: Bayesian Inference on Multiscale Models for Poisson Intensity Estimation: Applications to Photon-Limited Image Denoising. *IEEE Transactions on Image Processing*, 18(8):1724–1741, Αύγουστος 2009.

Δημοσιεύσεις σε Διεθνή Συνέδρια με Κριτή

- C1. Papandreou, G. και P. Maragos: A fast multigrid implicit algorithm for the evolution of geodesic active contours. Στο *Proc. IEEE Int. Conf. on Comp. Vision and Pat. Rec. (CVPR)*, Washington DC, June 2004, τόμος II, σελίδες 689–694.
- C2. Papandreou, G. και P. Maragos: A cross-validators statistical approach to scale selection for image denoising by nonlinear diffusion. Στο *Proc. IEEE Int. Conf. on Comp. Vision and Pat. Rec. (CVPR)*, San Diego, CA, June 2005, τόμος I, σελίδες 625–630.
- C3. Papandreou, G. και P. Maragos: Image denoising in nonlinear scale-spaces: Automatic scale selection via cross-validation. Στο *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, Genova, Italy, Sept. 2005, τόμος I, σελίδες 481–484.
- C4. Katsamanis, A., G. Papandreou, V. Pitsikalis, και P. Maragos: Multimodal fusion by adaptive compensation for feature uncertainty with application to audiovisual speech recognition. Στο *Proc. 14th European Signal Processing Conf. (EUSIPCO)*, Florence, Italy, Sept. 2006.
- C5. Pitsikalis, V., A. Katsamanis, G. Papandreou, και P. Maragos: Adaptive multimodal fusion by uncertainty compensation. Στο *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, Pittsburgh, PA, Sep. 2006, σελίδες 2458–2461.
- C6. Papandreou, G., A. Katsamanis, V. Pitsikalis, και P. Maragos: Multimodal fusion and learning with uncertain features applied to audiovisual speech recognition. Στο *Proc. IEEE Workshop on Multimedia Signal Processing (MMSP)*, Chania, Greece, Oct. 2007, σελίδες 264–267.

- C7. Katsamanis, A., G. Papandreou, και P. Maragos: Audiovisual-to-articulatory inversion using hidden markov models. Στο Proc. IEEE Workshop on Multimedia Signal Processing (MMSP), Chania, Greece, Oct. 2007, σελίδες 457–460.
- C8. Papandreou, G., P. Maragos, και A. Kokaram: Image inpainting with a wavelet domain hidden markov tree model. Στο Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, Nevada, U.S.A., Mar.-Apr. 2008, σελίδες 773–776.
- C9. Katsamanis, A., G. Papandreou, και P. Maragos: Audiovisual-to-articulatory speech inversion using active appearance models for the face and hidden markov models for the dynamics. Στο Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, Nevada, U.S.A., Mar.-Apr. 2008, σελίδες 2237–2240.
- C10. Papandreou, G. και P. Maragos: Adaptive and constrained algorithms for inverse compositional active appearance model fitting. Στο Proc. IEEE Int. Conf. on Comp. Vision and Pat. Rec. (CVPR), Anchorage, AK, June, 2008.
- C11. Katsamanis, A., G. Ananthakrishnan, G. Papandreou, P. Maragos, και O. Engwall: Audiovisual speech inversion by switching dynamical modeling governed by a hidden markov process. Στο Proc. 16th European Signal Processing Conf. (EUSIPCO), Lausanne, Switzerland, Aug. 2008.
- C12. Lefkimmiatis, S., G. Papandreou, και P. Maragos: Photon-limited image denoising by inference on multiscale models. Στο Proc. IEEE Int. Conf. on Image Processing (ICIP), San Diego, CA, Oct. 2008.
- C13. Lefkimmiatis, S., G. Papandreou, και P. Maragos: Poisson-Haar transform: A nonlinear multiscale representation for photon-limited image denoising. Στο Proc. IEEE Int. Conf. on Image Processing (ICIP), Cairo, Egypt, Nov. 2009.

Κεφάλαια σε Βιβλίο

- B1. Maragos, P., P. Gros, A. Katsamanis, και G. Papandreou: Cross-modal integration for performance improving in multimedia: A review. Στο Maragos, P., A. Potamianos, και P. Gros (επιμελητές): Multimodal Processing and Interaction: Audio, Video, Text, κεφάλαιο 1. Springer-Verlag, New York, 2008.
- B2. Papandreou, G., A. Katsamanis, V. Pitsikalis, και P. Maragos: Adaptive multimodal fusion by uncertainty compensation with application to audio-visual speech recognition. Στο Maragos, P., A. Potamianos, και P. Gros (επιμελητές): Multimodal Processing and Interaction: Audio, Video, Text, κεφάλαιο 4. Springer-Verlag, New York, 2008.

Βιβλιογραφία

- [AB85] Adelson, E.H. και J.R. Bergen: *Spatiotemporal energy models for the perception of motion*. J. of the Opt. Soc. Am. (A), 2(2):284–299, Φεβρουάριος 1985.
- [Act98] Acton, S.T.: *Multigrid anisotropic diffusion*. IEEE Trans. Image Process., 7(3):280–291, Μάρτιος 1998.
- [AEB06] Aharon, M., M. Elad, και A. Bruckstein: *K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation*. IEEE Trans. Signal Process., 54(11):4311–4322, Νοέμβριος 2006.
- [AKF05] Anderson, R., N. Kingsbury, και J. Fauqueur: *Coarse level object recognition using interlevel products of complex wavelets*. Στο Proc. IEEE Int. Conf. on Image Processing, τόμος 1, σελίδες 745–748, 2005.
- [ALM92] Alvarez, L., P. L. Lions, και J. M. Morel: *Image selective smoothing and edge detection by nonlinear diffusion, II*. SIAM J. Math. Anal., 29(3):845–866, 1992.
- [Alt90] Altman, N. S.: *Kernel smoothing of data with correlated errors*. J. of the Amer. Stat. Assoc., 85(411):749–759, 1990.
- [Ame92] Ames, W.: *Numerical Methods for Partial Differential Equations*. Academic Press, τρίτη έκδοση, 1992.
- [ARW95] Ascher, U.M., S.J. Ruuth, και B.T.R. Wetton: *Implicit-explicit methods for time-dependent partial differential equations*. SIAM J. Numer. Anal., 32(3):797–823, 1995.
- [BA83a] Burt, P.J. και E.H. Adelson: *The laplacian pyramid as a compact image code*. IEEE Trans. Commun., 31(4):532–540, Απρίλιος 1983.
- [BA83b] Burt, P.J. και E.H. Adelson: *A multiresolution spline with application to image mosaics*. ACM Tran. on Graphics, 2(4):217–236, 1983.
- [BBC⁺94] Barrett, R., M. Berry, T.F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, και H. Van der Vorst: *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, Philadelphia, PA, δεύτερη έκδοση, 1994.

- [BCG90] Bovik, A.C., M. Clark, και W.S. Geisler: *Multichannel texture analysis using localized spatial filters*. IEEE Trans. Pattern Anal. Mach. Intell., 12(1):55–73, Ιανουάριος 1990.
- [BCM05] Buades, A., B. Coll, και J.M. Morel: *A review of image denoising algorithms, with a new one*. Multiscale Model. Simul., 4(2):490–530, 2005.
- [Ber99] Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, Belmont, MA, USA, δεύτερη έκδοση, 1999.
- [Bes86] Besag, J.: *On the statistical analysis of dirty pictures*. J. of Royal Stat. Soc. (Series B), 48(3):259–302, 1986.
- [BHM00] Briggs, W. L., V. E. Henson, και S. F. McCormick: *A Multigrid Tutorial*. SIAM Books, Philadelphia, δεύτερη έκδοση, 2000.
- [BI98] Blake, A. και M. Isard: *Snakes: Active Contour Models*. Springer, 1998.
- [Bis06] Bishop, C.: *Pattern Recognition and Machine Learning*. Springer, 2006.
- [BL07] Brown, M. και D.G. Lowe: *Automatic panoramic image stitching using invariant features*. Int. J. of Comp. Vis., 74(1):59–73, 2007.
- [BM94] Brockett, R. και P. Maragos: *Evolution equations for continuous-scale morphological filtering*. IEEE Trans. Signal Process., 42:3377–3386, Δεκέμβριος 1994.
- [BN04] Bisani, M. και H. Ney: *Bootstrap estimates for confidence intervals in ASR performance evaluation*. Στο Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing, τόμος 1, σελίδες 409–412, 2004.
- [Bov05] Bovik, A.C. (επιμελήτης): *Handbook of Video and Image Processing*. Academic Press, δεύτερη έκδοση, 2005.
- [BPZ92] Behar, J., M. Porat, και Y.Y. Zeevi: *Image reconstruction from localized phase*. IEEE Trans. Signal Process., 40(4):736–743, Απρίλιος 1992.
- [Bra77] Brandt, A.: *Multi-level adaptive solutions to boundary-value problems*. Math. Comp., 31(138):333–390, 1977.
- [Bre92] Breiman, L.: *The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error*. J. of the Amer. Stat. Assoc., 87:738–754, 1992.
- [Bro66] Brodatz, P.: *Textures: a photographic album for artists and designers*. Dover, 1966.
- [BS94] Bouman, C.A. και M. Shapiro: *A multiscale random field model for Bayesian image segmentation*. IEEE Trans. Image Process., 3(2):162–177, Μάρτιος 1994.

- [BS97] Bell, A.J. και T.J. Sejnowski: *The ‘independent components’ of natural scenes are edge filters*. Vision Research, 37:3327–3338, 1997.
- [BSCB00] Bertalmio, M., G. Sapiro, V. Caselles, και C. Ballester: *Image inpainting*. Στο Proc. ACM Int. Conference on Computer Graphics and Interactive Techniques, σελίδες 417–424, 2000.
- [BSMH98] Black, M., G. Sapiro, D. H. Marimont, και D. Heeger: *Robust anisotropic diffusion*. IEEE Trans. Image Process., 7(3):421–432, Μάρτιος 1998.
- [BTFN⁺08] Brown, B.J., C. Toler-Franklin, D. Nehab, M. Bums, D. Dobkin, A. Vlachopoulos, C. Doumas, S. Rusinkiewicz, και T. Weyrich: *A system for high-volume acquisition and matching of fresco fragments: Reassembling Theran wall paintings*. ACM Tran. on Graphics (Proc. SIGGRAPH), 27(3), 2008.
- [BV03] Bardsley, J.M. και C.R. Vogel: *A nonnegatively constrained convex programming method for image reconstruction*. SIAM J. Scient. Comp., 25(4):1326–1343, 2003.
- [BV04] Boyd, S. και L. Vandenberghe: *Convex Optimization*. Cambridge Univ. Press, Cambridge, UK, 2004.
- [BWF⁺05] Bruhn, A., J. Weickert, C. Feddern, T. Kohlberger, και C. Schnörr: *Variational optical flow computation in real-time*. IEEE Trans. Image Process., 14(5):608–615, Μάιος 2005.
- [BZ87] Blake, A. και A. Zisserman: *Visual Reconstruction*. MIT Press, 1987.
- [Car88] Carlsson, S.: *Sketch based coding of grey level images*. Signal Processing, 15(1):57–83, 1988.
- [CB01] Choi, H. και R. G. Baraniuk: *Multiscale image segmentation using wavelet-domain hidden Markov models*. IEEE Trans. Image Process., 10(9):1309–1321, Σεπτέμβριος 2001.
- [CCCD93] Caselles, V., F. Catte, T. Coll, και F. Dibos: *A geometric model for active contours*. Numerische Mathematik, 66:1–31, 1993.
- [CDF92] Cohen, A., I. Daubechies, και J. C. Feauveau: *Biorthogonal bases of compactly supported wavelets*. Comm. on Pure and Applied Math., 45(5):485–560, 1992.
- [CDM⁺05] Carandini, M., J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, B.A. Olshausen, J.L. Gallant, και N.C. Rust: *Do we know what the early visual system does?* J. Neurosci., 25(46):10577–10597, 2005.
- [CET01] Cootes, T.F., G.J. Edwards, και C.J. Taylor: *Active appearance models*. IEEE Trans. Pattern Anal. Mach. Intell., 23(6):681–685, 2001.
- [CH53] Courant, R. και D. Hilbert: *Methods of Mathematical Physics*. Wiley, New York, 1953.

- [Chi96] Chib, S.: *Calculating posterior distributions and modal estimates in Markov mixture models*. J. of Econometrics, 75:79–97, 1996.
- [CK97] Cohen, L.D. και R. Kimmel: *Global minimum for active contour models: A minimal path approach*. Int. J. of Comp. Vis., 24(1):57–78, 1997.
- [CKS97] Caselles, V., R. Kimmel, και G. Sapiro: *Geodesic active contours*. Int. J. of Comp. Vis., 22(1):61–79, Φεβρουάριος 1997.
- [CLMC92] Catte, F., P.L. Lions, J.M. Morel, και T. Coll: *Image selective smoothing and edge detection by nonlinear diffusion*. SIAM J. Numer. Anal., 29(1):182–193, 1992.
- [CNB98] Crouse, M. S., R. D. Nowak, και R. G. Baraniuk: *Wavelet-based statistical signal processing using hidden Markov models*. IEEE Trans. Signal Process., 46(4):886–902, 1998.
- [Coh91] Cohen, Laurent D.: *On active contour models and balloons*. Comput. Vis. Graphics Image Proc., 53(2):211–218, 1991.
- [CRBK00] Choi, H., J. K. Romberg, R. G. Baraniuk, και N. Kingsbury: *Hidden Markov tree modeling of complex wavelet transforms*. Στο *Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing*, τόμος 1, σελίδες 133–136, 2000.
- [CS01a] Chan, T.F. και J. Shen: *Mathematical models for local nontexture inpaintings*. SIAM J. Applied Math., 62(3):1019–1043, 2001.
- [CS01b] Chan, T.F. και J. Shen: *Non-texture inpainting by curvature-driven diffusion (CDD)*. Journal of Visual Communication and Image Representation, 12(4):436–449, 2001.
- [CS05] Chan, T.F. και J. Shen: *Variational image inpainting*. Comm. on Pure and Applied Math., 58(5):579–619, 2005.
- [CSZ06] Chan, T.F., J. Shen, και H. M. Zhou: *Total variation wavelet inpainting*. J. Math. Imaging and Vision, 25:107–125, 2006.
- [CTWS02] Cremers, D., F. Tischhäuser, J. Weickert, και C. Schnörr: *Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional*. Int. J. of Comp. Vis., 50(3):295–313, 2002.
- [CV01] Chan, T.F. και L.A. Vese: *Active contours without edges*. IEEE Trans. Image Process., 10(2):266–277, 2001.
- [CWB94] Chou, K.C., A.S. Willsky, και A. Benveniste: *Multiscale recursive estimation, data fusion, and regularization*. IEEE Trans. Autom. Control, 39(3):464–478, Μάρτιος 1994.
- [CY90] Clark, J.J. και A.L. Yuille: *Data Fusion for Sensory Information Processing*. Kluwer Academic Publishers, 1990.

- [Dau80] Daugman, J.G.: *Two-dimensional spectral analysis of cortical receptive field profiles*. Vision Research, 20(10):847–856, 1980.
- [Dau85] Daugman, J.G.: *Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters*. J. of the Opt. Soc. Am. (A), 2(7):1160–1169, Ιούλιος 1985.
- [Dau88] Daugman, J.G.: *Complete discrete 2-D gabor transforms by neural networks for image analysis and compression*. IEEE Trans. Acoust., Speech, Signal Process., 36(7):1169–1179, Ιούλιος 1988.
- [Dau90] Daubechies, I.: *The wavelet transform, time-frequency localization and signal analysis*. IEEE Trans. Inf. Theory, 36(5):961–1005, Δεκέμβριος 1990.
- [Dau92] Daubechies, I.: *Ten Lectures on Wavelets*. SIAM, 1992.
- [Dau04] Daugman, J.G.: *How iris recognition works*. IEEE Trans. Circuits Syst. Video Technol., 14(1):21–30, Ιανουάριος 2004.
- [DDA05] Deng, L., J. Droppo, και A. Acero: *Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion*. IEEE Trans. Speech Audio Process., 13(3):412–421, 2005.
- [DFKE07] Dabov, K., A. Foi, V. Katkovnik, και K. Egiazarian: *Image denoising by sparse 3D transform-domain collaborative filtering*. IEEE Trans. Image Process., 16(8):2080–2095, Αύγουστος 2007.
- [DL00] Dupont, S. και J. Luettin: *Audio-visual speech modeling for continuous speech recognition*. IEEE Trans. Multimedia, 2(3):141–151, 2000.
- [DLR77] Dempster, A. P., N. M. Laird, και D. B. Rubin: *Maximum Likelihood from incomplete data via the EM algorithm*. J. of Royal Stat. Soc. (Series B), 39(1):1–38, 1977.
- [Dou92] Doumas, C.: *The Wall-paintings of Thera*. The Thera Foundation, 1992.
- [DRO93] Digalakis, V., J.R. Rohlicek, και M. Ostendorf: *ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition*. IEEE Trans. Speech Audio Process., σελίδες 431–442, 1993.
- [DS52] Duffin, R.J. και A.C. Schaeffer: *A class of nonharmonic Fourier series*. Trans. of the AMS, 72(2):341–366, Μάρτιος 1952.
- [DS07] Durand, F. και R. Szeliski: *Guest editors' introduction: Computational photography*. IEEE Comput. Graph. Appl., 27(2):21–22, 2007.
- [DV02] Do, M. N. και M. Vetterli: *Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models*. IEEE Trans. Multimedia, 4(4):517–527, Δεκέμβριος 2002.

- [dZ90] Zeeuw, P.M. de: *Matrix-dependent prolongations and restrictions in a black-box multigrid solver*. J. Comp. Appl. Math., 33:1–27, 1990.
- [EDGL07] Eldar, Y., T. Davidson, A.B. Gershman, και Z. Q. Luo: *Introduction to issue on convex optimization methods for signal processing*. IEEE J. Sel. Topics Signal Process., 1(4):537–539, Δεκέμβριος 2007.
- [Efr86] Efron, B.: *How biased is the apparent error rate of a prediction rule?* J. of the Amer. Stat. Assoc., 81:461–470, 1986.
- [EL99] Efros, A. A. και T. K. Leung: *Texture synthesis by non-parametric sampling*. Στο Proc. IEEE Int. Conf. on Computer Vision, τόμος 2, 1999.
- [ESQD05] Elad, M., J. L. Starck, P. Querre, και D. L. Donoho: *Simultaneous cartoon and texture image inpainting using morphological component analysis*. App. and Comp. Harm. Anal., 19:340–358, 2005.
- [EZ98] Elder, J. H. και S. W. Zucker: *Local scale control for edge detection and blur estimation*. IEEE Trans. Pattern Anal. Mach. Intell., 20(7):699–716, 1998.
- [FA91] Freeman, W.T. και E.H. Adelson: *The design and use of steerable filters*. IEEE Trans. Pattern Anal. Mach. Intell., 13(9):891–906, Σεπτέμβριος 1991.
- [FH04] Felzenszwalb, P. και D. Huttenlocher: *Distance transforms of sampled functions*. Τεχνική Αναφορά TR2004-1963, Cornell Computing and Information Science, 2004.
- [Fie87] Field, D.J.: *Relations between the statistics of natural images and the response properties of cortical cells*. J. of the Opt. Soc. Am. (A), 4(12):2379–2394, 1987.
- [Fie94] Field, D.J.: *What is the goal of sensory coding?* Neural Computation, 6:559–601, 1994.
- [FNW07] Figueiredo, M.A.T., R.D. Nowak, και S.J. Wright: *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*. IEEE J. Sel. Topics Signal Process., 1(4):586–597, Δεκέμβριος 2007.
- [FP03] Forsyth, D. A. και J. Ponce: *Computer Vision: A Modern Approach*. Prentice-Hall, 2003.
- [FS05] Fadili, M. J. και J. L. Starck: *EM algorithm for sparse representation-based image inpainting*. Στο Proc. IEEE Int. Conf. on Image Processing, 2005.
- [FTDG08] Févotte, C., B. Torrèsani, L. Daudet, και S. J. Godsill: *Sparse linear regression with structured priors and application to denoising of musical audio*. IEEE Trans. Audio, Speech, and Language Processing, 16(1):174–185, 2008.
- [Ful87] Fuller, W. A.: *Measurement Error Models*. Wiley, 1987.

- [FX03] Fan, G. και X. G. Xia: *Wavelet-based texture analysis and synthesis using hidden Markov models*. IEEE Trans. Circuits Syst. I, 50(1):106–120, Ιανουάριος 2003.
- [Gab46] Gabor, D.: *Theory of communication*. J. of the IEE, 93 (Part III)(26):429–457, Νοέμβριος 1946.
- [GBT⁺02] Godin, G., J. A. Beraldin, J. Taylor, L. Cournoyer, M. Rioux, S. El-Hakim, R. Baribeau, F. Blais, P. Boulanger, J. Domey, και M. Picard: *Active optical 3D imaging for heritage applications*. IEEE Comput. Graph. Appl., σελίδες 24–36, Sept./Oct. 2002.
- [GC89] Gillick, L. και S. Cox: *Some statistical issues in the comparison of speech recognition algorithms*. Στο *Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing*, τόμος 1, σελίδες 532–535, 1989.
- [GCSR03] Gelman, A., J.B. Carlin, H.S. Stern, και D.B. Rubin: *Bayesian data analysis*. Chapman & Hall/CRC, δεύτερη έκδοση, 2003.
- [GG84] Geman, S. και D. Geman: *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*. IEEE Trans. Pattern Anal. Mach. Intell., 6(6):721–741, 1984.
- [GGCV95] Geiger, D., A. Gupta, L.A. Costa, και J. Vlontzos: *Dynamic programming for detecting, tracking, and matching deformable contours*. IEEE Trans. Pattern Anal. Mach. Intell., 17(3):294–302, 1995.
- [GKRR01] Goldenberg, R., R. Kimmel, E. Rivlin, και M. Rudzsky: *Fast geodesic active contours*. IEEE Trans. Image Process., 10:1467–1475, 2001.
- [GM96] Guy, G. και G. Medioni: *Inferring global perceptual contours from local features*. Int. J. of Comp. Vis., 20(1/2):113–133, 1996.
- [GM07] Grenander, U. και M. Miller: *Pattern Theory: From Representation to Inference*. Oxford Univ. Press, 2007.
- [GNP06] Giakoumis, I., N. Nikolaidis, και I. Pitas: *Digital image processing techniques for the detection and removal of cracks in digitized paintings*. IEEE Trans. Image Process., 15(1):178–188, Ιανουάριος 2006.
- [GPN02] Gravier, G., G. Potamianos, και C. Neti: *Asynchrony modeling for audio-visual speech recognition*. Στο *Human Language Technology Conference*, 2002.
- [Gre93] Grenander, U.: *General Pattern Theory*. Oxford Univ. Press, 1993.
- [Gri83] Grimson, W.E.L.: *An implementation of a computational theory of visual surface interpolation*. Comput. Vis. Graphics Image Proc., 22(1):39–69, 1983.
- [Grö93] Gröchenig, K.: *Acceleration of the frame algorithm*. IEEE Trans. Signal Process., 41(12):3331–3340, Δεκέμβριος 1993.

- [GRS95] Gilks, W.R., S. Richardson, και D.J. Spiegelhalter (επιμελητές): *Markov chain Monte Carlo in practice*. Chapman & Hall, 1995.
- [GS90] Gelfand, A.E. και A.F.M. Smith: *Sampling-based approaches to calculating marginal densities*. J. of the Amer. Stat. Assoc., 85(410):398–409, 1990.
- [GS01] Grenander, U. και A. Srivastava: *Probability models for clutter in natural images*. IEEE Trans. Pattern Anal. Mach. Intell., 23(4):424–429, 2001.
- [GVL96] Golub, G.H. και C. Van Loan: *Matrix Computations*. John Hopkins Press, 1996.
- [GZS01] Gilboa, G., Y. Y. Zeevi, και N. Sochen: *Image enhancement, segmentation and denoising by time dependent nonlinear diffusion processes*. Στο *Proc. IEEE Int. Conf. on Image Processing*, 2001.
- [GZW03] Guo, C. E., S. C. Zhu, και Y. N. Wu: *Modeling visual patterns by integrating descriptive and generative methods*. Int. J. of Comp. Vis., 53(1):5–29, 2003.
- [Hay82] Hayes, M.H.: *The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform*. IEEE Trans. Acoust., Speech, Signal Process., 30(2):140–154, Απρίλιος 1982.
- [HB89] Horn, B.K.P. και M.J. Brooks: *Shape from Shading*. MIT Press, 1989.
- [HH00] Hyvärinen, A. και P. Hoyer: *Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces*. Neural Computation, 12(7):1705–1720, Ιούλιος 2000.
- [HLO80] Hayes, M.H., J.S. Lim, και A.V. Oppenheim: *Signal reconstruction from phase or magnitude*. IEEE Trans. Acoust., Speech, Signal Process., 28(6):672–680, Δεκέμβριος 1980.
- [HM99] Huang, J. και D. Mumford: *Statistics of natural images and models*. Στο *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, σελίδες 541–547, 1999.
- [HO07] Hua, G. και M.T. Orchard: *Image inpainting based on geometrical modeling of complex wavelet coefficients*. Στο *Proc. IEEE Int. Conf. on Image Processing*, τόμος 1, σελίδες 553–556, 2007.
- [HO08] Hua, G. και M.T. Orchard: *Image reconstruction from the phase or magnitude of its complex wavelet transform*. Στο *Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing*, σελίδες 3261–3264, 2008.
- [HPB94] Heitz, F., P. Pérez, και P. Bouthemy: *Multiscale minimization of global energy functions in some visual recovery problems*. Comput. Vis. Graphics Image Proc., 59(1):125–134, 1994.
- [HS81] Horn, B.K.P. και B.G. Schunck: *Determining optical flow*. Artificial Intelligence, 17(1-3):185–203, 1981.

- [HTF01] Hastie, T., R. Tibshirani, και J. Friedman: *The Elements of Statistical Learning*. Springer-Verlag, 2001.
- [HW06] Heil, C. και D.F. Walnut (επιμελητές): *Fundamental papers in wavelet theory*. Princeton University Press, 2006.
- [HZ03] Hartley, R. και A. Zisserman: *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, δεύτερη έκδοση, 2003.
- [JBBA03] Jehan-Besson, S., M. Barlaud, και G. Aubert: *DREAM²S: Deformable regions driven by an Eulerian accurate minimization method for image and video segmentation*. Int. J. of Comp. Vis., 53(1):45-70, 2003.
- [JDM00] Jain, A., R. Duin, και J. Mao: *Statistical pattern recognition: A review*. IEEE Trans. Pattern Anal. Mach. Intell., 22(1):4-37, Ιανουάριος 2000.
- [JM08] Jurafsky, D. και J.H. Martin: *Speech and language processing*. Prentice-Hall, δεύτερη έκδοση, 2008.
- [Kan79] Kanizsa, G.: *Organization in Vision: Essays on Gestalt Perception*. Praeger, New York, 1979.
- [KB03] Kimmel, R. και A. M. Bruckstein: *Regularized laplacian zero crossings as optimal edge integrators*. Int. J. of Comp. Vis., 53(3):225-243, 2003.
- [KB05] Kolmogorov, V. και Y. Boykov: *What metrics can be approximated by geodesics, or global optimization of length/area and flux*. Στο Proc. IEEE Int. Conf. on Computer Vision, σελίδες 564-571, 2005.
- [KBW06] Keller, M., S. Bengio, και S.Y. Wong: *Benchmarking non-parametric statistical tests*. Τόμος 18, σελίδες 651-658, 2006.
- [KC07a] Kovacevic, J. και A. Chebira: *Life beyond bases: The advent of frames (Part I)*. IEEE Signal Process. Mag., 24(4):86-104, Ιούλιος 2007.
- [KC07b] Kovacevic, J. και A. Chebira: *Life beyond bases: The advent of frames (Part II)*. IEEE Signal Process. Mag., 24(5):115-125, Σεπτέμβριος 2007.
- [KEM09] Kokkinos, I., G. Evangelopoulos, και P. Maragos: *Texture analysis and segmentation using modulation features, generative models, and weighted curve evolution*. IEEE Trans. Pattern Anal. Mach. Intell., 31(1):142-157, Ιανουάριος 2009.
- [KHDM98] Kittler, J., M. Hatef, R. Duin, και J. Matas: *On combining classifiers*. IEEE Trans. Pattern Anal. Mach. Intell., 20(3):226-239, Μάρτιος 1998.
- [Kin99] Kingsbury, N.: *Image processing with complex wavelets*. Phil. Trans. R. Soc. Lond. A (Math. Phys. Engin. Sci.), 357(1760):2543-2560, 1999.
- [Kin01] Kingsbury, N.: *Complex wavelets for shift invariant analysis and filtering of signals*. App. and Comp. Harm. Anal., 10(3):234-253, 2001.

- [KKL⁺07] Kim, S. J., K. Koh, M. Lustig, S. Boyd, και D. Gorinevsky: *An interior point method for large scale ℓ_1 -regularized least squares*. IEEE J. Sel. Topics Signal Process., 1(4):606–617, Δεκέμβριος 2007.
- [KKY04] Kenigsberg, A., R. Kimmel, και I. Yavneh: *A multigrid approach for fast geodesic active contours*. Τεχνική Αναφορά CIS-2004-06, Technion, 2004.
- [KM05] Kokkinos, I. και P. Maragos: *An expectation maximization approach to the synergy between image segmentation and object categorization*. Στο *Proc. IEEE Int. Conf. on Computer Vision*, τόμος I, σελίδες 617–624, 2005.
- [Koe84] Koenderink, J.J.: *The structure of images*. Biological Cybernetics, 50:363–370, 1984.
- [Kof35] Koffka, K.: *Principles of Gestalt Psychology*. Harcourt, Brace and Co, New York, 1935.
- [Koh95] Kohavi, R.: *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Στο *Proc. Int. Joint Conf. on Artificial Intel.*, σελίδες 1137–1143, 1995.
- [Kok98] Kokaram, A.C.: *Motion Picture Restoration*. Springer, 1998.
- [Kov99] Kovesei, P.: *Image features from phase congruency*. Videre: Journal of Computer Vision Research, 1(3):2–26, 1999.
- [KPM09] Katsamanis, A., G. Papandreou, και P. Maragos: *Face active appearance modeling and speech acoustic information to recover articulation*. IEEE Trans. on Audio, Speech and Language Process., 17(3):411–422, Μάρτιος 2009.
- [KR96] Knill, D. και W. Richards (επιμελητές): *Perception as Bayesian Inference*. Cambridge Univ. Press, 1996.
- [KSJ07a] Kivinen, J., E. Sudderth, και M. Jordan: *Image denoising with nonparametric hidden Markov trees*. Στο *Proc. IEEE Int. Conf. on Image Processing*, τόμος 3, σελίδες 121–124, 2007.
- [KSJ07b] Kivinen, J., E. Sudderth, και M. Jordan: *Learning multiscale representations of natural scenes using Dirichlet processes*. Στο *Proc. IEEE Int. Conf. on Computer Vision*, 2007.
- [KT07] Komodakis, N. και G. Tziritas: *Approximate labeling via graph cuts based on linear programming*. IEEE Trans. Pattern Anal. Mach. Intell., 29(8):1436–1453, Αύγουστος 2007.
- [KW87] Kass, M. και A. Witkin: *Analyzing oriented patterns*. Comput. Vis. Graphics Image Proc., 37:362–385, 1987.
- [KWT88] Kass, M., A.P. Witkin, και D. Terzopoulos: *Snakes: Active contour models*. Int. J. of Comp. Vis., 1(4):321–331, 1988.

- [KY03] Kimmel, R. και I. Yavneh: *An algebraic multigrid approach for image analysis*. SIAM J. Scient. Comp., 24(4):1218–1231, 2003.
- [Lee80] Lee, J. S.: *Digital image enhancement and noise filtering by use of local statistics*. IEEE Trans. Pattern Anal. Mach. Intell., 2:165–168, Μάρτιος 1980.
- [Lee96] Lee, T.S.: *Image representation using 2D gabor wavelets*. IEEE Trans. Pattern Anal. Mach. Intell., 18(10):959–971, Οκτώβριος 1996.
- [LGF00] Leventon, M., E. Grimson, και O. Faugeras: *Statistical shape influence in geodesic active contours*. Στο *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, τόμος 1, σελίδες 316–323, 2000.
- [LHC96] Leroy, B., I. L. Herlin, και L. Cohen: *Multiresolution algorithms for active contour models*. Στο *Proc. Int. Conf. Anal. Optimiz. Systems*, σελίδες 58–65, 1996.
- [Lin98] Lindeberg, T.: *Feature detection with automatic scale selection*. Int. J. of Comp. Vis., 30(2):79–116, 1998.
- [LM01] Liu, J. και P. Moulin: *Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients*. IEEE Trans. Image Process., 10(11):1647–1658, Νοέμβριος 2001.
- [LM05] Lee, T. και X. L. Meng: *A self-consistent wavelet method for denoising images with missing pixels*. Στο *Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing*, τόμος II, σελίδες 41–44, 2005.
- [LMP09] Lefkimmiatis, S., P. Maragos, και G. Papandreou: *Bayesian inference on multiscale models for Poisson intensity estimation: Applications to photon-limited image denoising*. IEEE Trans. on Image Process., 18(8):1724–1741, Αύγουστος 2009.
- [LNT91] Lu, T., P. Neittaanmäki, και X. C. Tai: *A parallel splitting up method and its application to Navier-Stokes equations*. Appl. Math. Lett., 4(2):25–29, 1991.
- [Low04] Lowe, D.G.: *Distinctive image features from scale-invariant keypoints*. Int. J. of Comp. Vis., 60(2):91–110, 2004.
- [LPH00] Laferté, J. M., P. Pérez, και F. Heitz: *Discrete Markov image modeling and inference on the quadtree*. IEEE Trans. Image Process., 9(3):390–404, Μάρτιος 2000.
- [LPN01] Luettin, J., G. Potamianos, και C. Neti: *Asynchronous stream modeling for large vocabulary audio-visual speech recognition*. Στο *Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing*, 2001.
- [Luc03] Lucey, S.: *An evaluation of visual speech features for the tasks of speech and speaker recognition*. Στο *Proc. of Audio-and Video-Based Biometric Person Authentication*, σελίδες 260–267, 2003.

- [LWK94] Liu, J.S., W.H. Wong, και A. Kong: *Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes*. *Biometrika*, 81(1):27–40, 1994.
- [Mal89] Mallat, S.G.: *A theory for multiresolution signal decomposition: The wavelet transform*. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):674–693, 1989.
- [Mal90] Malvar, H.S.: *Lapped transforms for efficient transform/subband coding*. *IEEE Trans. Acoust., Speech, Signal Process.*, 38(6):969–978, Ιούνιος 1990.
- [Mal92] Malvar, H.S.: *Signal Processing with Lapped Transforms*. Artech House, Boston, MA, 1992.
- [Mal99a] Mallat, S.: *A Wavelet Tour of Signal Processing*. Acad. Press, δεύτερη έκδοση, 1999.
- [Mal99b] Malvar, H.S.: *A modulated complex lapped transform and its applications to audio processing*. Στο *Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing*, σελίδες 1421–1424, 1999.
- [Mal03] Malvar, H.S.: *Fast algorithm for the modulated complex lapped transform*. *IEEE Signal Process. Lett.*, 10(1):8–10, Ιανουάριος 2003.
- [Mar82] Marr, D.: *Vision*. W.H. Freeman, 1982.
- [Mar95] Maragos, P.: *Slope transforms: Theory and application to nonlinear signal processing*. *IEEE Trans. Signal Process.*, 43(4):864–877, Απρίλιος 1995.
- [Mar99a] Marinatos, S.: *Excavations at Thera I-III, 1967-1969 Seasons*. Αριθμός 178 στο *BAE*. The Archaeological Society at Athens, δεύτερη έκδοση, 1999.
- [Mar99b] Marinatos, S.: *Excavations at Thera IV-V, 1970-1971 Seasons*. Αριθμός 179 στο *BAE*. The Archaeological Society at Athens, δεύτερη έκδοση, 1999.
- [Mar99c] Marinatos, S.: *Excavations at Thera VI-VII, 1972-1973 Seasons*. Αριθμός 180 στο *BAE*. The Archaeological Society at Athens, δεύτερη έκδοση, 1999.
- [Mar03] Maragos, P.: *Algebraic and pde approaches for lattice scale-spaces with global constraints*. *Int. J. of Comp. Vis.*, 52(2-3):121–137, 2003.
- [Mar05] Maragos, P.: *PDEs for morphological scale-spaces and eikonal applications*. Στο Bovik, A.C. [Bov05], κεφάλαιο 4.16.
- [Mas02] Masnou, S.: *Disocclusion: A variational approach using level lines*. *IEEE Trans. Image Process.*, 11(2):68–76, Φεβρουάριος 2002.
- [MB88] Morrone, M.C. και D.C. Burr: *Feature detection in human vision: a phase-dependent energy model*. *Proc. R. Soc. Lond. B*, 235:221–245, 1988.
- [MCB⁺02] Matthews, I., T. F. Cootes, J. A. Bangham, S. Cox, και R. Harvey: *Extraction of visual features for lipreading*. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(2):198–213, 2002.

- [Mey92] Meyer, Y.: *Wavelets and Operators*. Cambridge Univ. Press, 1992.
- [Mey93] Meyer, Y.: *Wavelets: Algorithms and Applications*. SIAM, 1993.
- [MG01] Mumford, D. και B. Gidas: *Stochastic models for generic images*. *Quart. of Appl. Math.*, 59(1):85–111, 2001.
- [MGKP08] Maragos, P., P. Gros, A. Katsamanis, και G. Papandreou: *Cross-modal integration for performance improving in multimedia: A review*. Στο Maragos, P., A. Potamianos, και P. Gros (επιμελητές): *Multimodal Processing and Interaction: Audio, Video, Text*, κεφάλαιο 1, σελίδες 3–48. Springer-Verlag, New York, 2008.
- [MK98] Magarey, J. και N. Kingsbury: *Motion estimation using a complex valued wavelet transform*. *IEEE Trans. Signal Process.*, 46(4):1069–1084, Απρίλιος 1998.
- [MK08a] Miller, M. και N. Kingsbury: *Image denoising using derotated complex wavelet coefficients*. *IEEE Trans. Image Process.*, 17(9):1500–1511, Σεπτέμβριος 2008.
- [MK08b] Miller, M. και N. Kingsbury: *Image modeling using interscale phase properties of complex wavelet coefficients*. *IEEE Trans. Image Process.*, 17(9):1491–1499, Σεπτέμβριος 2008.
- [MM76] McGurk, H. και J. MacDonald: *Hearing lips and seeing voices*. *Nature*, 264:746–748, 1976.
- [MM98a] Masnou, S. και J. M. Morel: *Level lines based disocclusion*. Στο *Proc. IEEE Int. Conf. on Image Processing*, τόμος 3, σελίδες 259–263, 1998.
- [MM98b] Movellan, J. R. και P. Mineiro: *Robust sensor fusion: Analysis and application to audio visual speech recognition*. 32:85–100, 1998.
- [MM00] Meyer, F. και P. Maragos: *Nonlinear scale-space representation with morphological levelings*. *Journal of Visual Communication and Image Representation*, 11:245–265, 2000.
- [MMP87] Marroquin, J., S. Mitter, και T. Poggio: *Probabilistic solution of ill-posed problems in computational vision*. *J. of the Amer. Stat. Assoc.*, 82(397):76–89, 1987.
- [MN03] Mrazek, P. και M. Navara: *Selection of optimal stopping time for nonlinear diffusion filtering*. *Int. J. of Comp. Vis.*, 52(2/3):189–203, 2003.
- [MPNL01] Matthews, I., G. Potamianos, C. Neti, και J. Luetin: *A comparison of model and transform-based visual features for audio-visual LVCSR*. Στο *Proc. of IEEE Int. Conference on Multimedia and Expo*, 2001.

- [MS89a] Malvar, H.S. και D.H. Staelin: *The LOT: Transform coding without blocking effects*. IEEE Trans. Acoust., Speech, Signal Process., 37(4):553–559, Απρίλιος 1989.
- [MS89b] Mumford, D. και J. Shah: *Optimal approximations by piecewise smooth functions and associated variational problems*. Comm. on Pure and Applied Math., 42(5):577–685, 1989.
- [MSV95] Malladi, R., J. A. Sethian, και B. C. Vemuri: *Shape modelling with front propagation: a level set approach*. IEEE Trans. Pattern Anal. Mach. Intell., 17(2):158–175, 1995.
- [MT91] Moré, J.J. και G. Toraldo: *On the solution of large quadratic programming problems with bound constraints*. SIAM J. Optimization, 1(1):93–113, 1991.
- [MT00] McInerney, T. και D. Terzopoulos: *T-snakes: Topology adaptive snakes*. Medical Image Analysis, 4:73–91, 2000.
- [Mum94] Mumford, D.: *Elastica and computer vision*. Στο Bajaj, C. L. (επιμελητής): *Algebraic Geometry and its Applications*, σελίδες 491–506. Springer-Verlang, New York, 1994.
- [Mum96] Mumford, D.: *Pattern theory: A unifying perspective*. Στο Knill, D. και W. Richards [KR96], κεφάλαιο 1.
- [Nas96] Nason, G. P.: *Wavelet shrinkage using cross-validation*. J. of Royal Stat. Soc. (Series B), 58(2):463–479, 1996.
- [NLP⁺02] Nefian, A.V., L. Liang, X. Pi, X. Liu, και K. Murphy: *Dynamic bayesian networks for audio-visual speech recognition*. EURASIP Journal on Applied Signal Processing, 11:1–15, 2002.
- [NW06] Nocedal, J. και S.J. Wright: *Numerical Optimization*. Springer, NY, USA, δεύτερη έκδοση, 2006.
- [OF96] Olshausen, B.A. και D.J. Field: *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*. Nature, 381:607–609, 1996.
- [OL81] Oppenheim, A.V. και J.S. Lim: *The importance of phase in signals*. Proc. IEEE, 69(5):529–541, Μάιος 1981.
- [ÓRF94] Ó Ruanaidh, J. J. K. και W. J. Fitzgerald: *Interpolation of missing samples for audio restoration*. Electronics Letters, 30(8):622–623, Απρίλιος 1994.
- [OS88] Osher, S. και J. A. Sethian: *Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations*. Journal of Computational Physics, 79:12–49, 1988.
- [OS98] Oppenheim, A.V. και R.W. Schaffer: *Discrete-Time Signal Processing*. Prentice Hall, δεύτερη έκδοση, 1998.

- [Pal03] Palmer, S.E.: *Visual perception of objects*. Στο Healy, A.F. και R.W. Proctor (επιμελητές): *Handbook of Psychology*, τόμος 4 (Experimental Psychology), κεφάλαιο 7, σελίδες 179–211. Wiley, 2003.
- [Pap77] Papoulis, A.: *Signal Analysis*. McGraw-Hill, 1977.
- [PBN03] Palmer, S.E., J.L. Brooks, και R. Nelson: *When does grouping happen?* *Acta Psychologica*, 114:311–330, 2003.
- [PC82] Piotrowski, L.N. και F.W. Campbell: *A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase*. *Perception*, 11:337–346, 1982.
- [PCL00] Pérez, P., A. Chardin, και J. M. Laferté: *Noniterative manipulation of discrete energy-based models for image analysis*. *Pattern Recognition*, 33:573–586, Μάρτιος 2000.
- [PD02] Paragios, Nikos και Rachid Deriche: *Geodesic active regions: A new framework to deal with frame partition problems in computer vision*. *Journal of Visual Communication and Image Representation*, 13(1-2):249–268, 2002.
- [Pea88] Pearl, J.: *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufman, 1988.
- [Per95] Perona, P.: *Deformable kernels for early vision*. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(5):488–499, Μάιος 1995.
- [PGTG02] Patterson, E. K., S. Gurbuz, Z. Tufekci, και J. N. Gowdy: *CUAVE: A new audio-visual database for multimodal human-computer interface research*. Στο *Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing*, 2002.
- [PJB87] Princen, J.P., A.W. Johnson, και A.B. Bradley: *Subband/transform coding using filter bank designs based on time domain aliasing cancellation*. Στο *Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing*, σελίδες 2161–2164, 1987.
- [PKPM07] Papandreou, G., A. Katsamanis, V. Pitsikalis, και P. Maragos: *Multimodal fusion and learning with uncertain features applied to audiovisual speech recognition*. Στο *Proc. IEEE Workshop on Multimedia Signal Processing (MMSP)*, σελίδες 264–267, Chania, Greece, Οκτώβριος 2007.
- [PKPM09] Papandreou, G., A. Katsamanis, V. Pitsikalis, και P. Maragos: *Adaptive multimodal fusion by uncertainty compensation with application to audiovisual speech recognition*. *IEEE Trans. on Audio, Speech and Language Process.*, 17(3):423–435, Μάρτιος 2009.
- [PKSC00] Perdikatsis, V., V. Kilikoglou, S. Sotiropoulou, και E. Chryssikopoulou: *Physicochemical characterization of pigments from thean wall paintings*. Στο Sherratt, S. [She00], σελίδες 103–118.

- [PM90] Perona, P. και J. Malik: *Scale space and edge detection using anisotropic diffusion*. IEEE Trans. Pattern Anal. Mach. Intell., 12(7):629–639, 1990.
- [PM04] Papandreou, G. και P. Maragos: *A fast multigrid implicit algorithm for the evolution of geodesic active contours*. Στο Proc. IEEE Int. Conf. on Comp. Vision and Pat. Rec. (CVPR), τόμος II, σελίδες 689–694, Washington, DC, Ιούνιος 2004.
- [PM05a] Papandreou, G. και P. Maragos: *A cross-validators statistical approach to scale selection for image denoising by nonlinear diffusion*. Στο Proc. IEEE Int. Conf. on Comp. Vision and Pat. Rec. (CVPR), τόμος I, σελίδες 625–630, San Diego, CA, Ιούνιος 2005.
- [PM05b] Papandreou, G. και P. Maragos: *Image denoising in nonlinear scale-spaces: Automatic scale selection via cross-validation*. Στο Proc. IEEE Int. Conf. on Image Processing (ICIP), τόμος I, σελίδες 481–484, Genova, Italy, Σεπτέμβριος 2005.
- [PM07] Papandreou, G. και P. Maragos: *Multigrid geometric active contour models*. IEEE Trans. on Image Process., 16(1):229–240, Ιανουάριος 2007.
- [PM08] Papandreou, G. και P. Maragos: *Adaptive and constrained algorithms for inverse compositional active appearance model fitting*. Στο Proc. IEEE Int. Conf. on Comp. Vision and Pat. Rec. (CVPR), Anchorage, AK, Ιούνιος 2008.
- [PMK08] Papandreou, G., P. Maragos, και A. Kokaram: *Image inpainting with a wavelet domain hidden markov tree model*. Στο Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), σελίδες 773–776, Las Vegas, NV, Απρίλιος 2008.
- [PMO⁺99] Peng, D., B. Merriman, S. Osher, H. Zhao, και M. Kang: *A PDE-based fast local level set method*. Journal of Computational Physics, 155:410–438, 1999.
- [PNGG03] Potamianos, G., C. Neti, G. Gravier, και A. Garg: *Automatic recognition of audio-visual speech: Recent progress and challenges*. Proc. IEEE, 91(9):1306–1326, 2003.
- [PPE⁺02] Papaodysseus, C., T. Panagopoulos, M. Exarhos, C. Triantafyllou, D. Fragoulis, και C. Doulmas: *Contour-shape based reconstruction of fragmented, 1600 B.C. wall paintings*. IEEE Trans. Signal Process., 50(6):1277–1288, Ιούνιος 2002.
- [PS00a] Painter, T. και A. Spanias: *Perceptual coding of digital audio*. Proc. IEEE, 88(4):451–515, Απρίλιος 2000.
- [PS00b] Portilla, J. και E. P. Simoncelli: *A parametric texture model based on joint statistics of complex wavelet coefficients*. Int. J. of Comp. Vis., 40(1):49–71, 2000.

- [PSWS03] Portilla, J., V. Strela, M.J. Wainwright, και E.P. Simoncelli: *Image denoising using scale mixtures of gaussians in the wavelet domain*. IEEE Trans. Image Process., 12(11):1338–1351, 2003.
- [PTK85] Poggio, T., V. Torre, και C. Koch: *Computational vision and regularization theory*. Nature, 317(26):314–319, Σεπτέμβριος 1985.
- [PTN98] Pessoa, L., E. Thompson, και A. Noë: *Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception (with discussion)*. Behavioral and Brain Sciences, 21:723–802, 1998.
- [PTVF92] Press, W., S. Teukolsky, W. Vetterling, και B. Flannery: *Numerical Recipes*. Cambridge Univ. Press, 1992.
- [PZ88] Porat, M. και Y.Y. Zeevi: *The generalized Gabor scheme of image representation in biological and machine vision*. IEEE Trans. Pattern Anal. Mach. Intell., 10(4):452–468, Ιούλιος 1988.
- [QC99] Qian, S. και D. Chen: *Joint time-frequency analysis*. IEEE Signal Process. Mag., 16(2):52–67, Μάρτιος 1999.
- [Rab89] Rabiner, L.: *A tutorial on hidden Markov models and selected applications in speech recognition*. Proc. IEEE, 77(2):257–286, 1989.
- [RBD03] Rousson, M., T. Brox, και R. Deriche: *Active unsupervised texture segmentation on a diffusion based feature space*. Στο Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, τόμος II, σελίδες 699–704, 2003.
- [RC05] Robert, C.P. και G. Casella: *Monte Carlo Statistical Methods*. Springer, δεύτερη έκδοση, 2005.
- [RCB01] Romberg, J.K., H. Choi, και R.G. Baraniuk: *Bayesian tree-structured image modeling using wavelet-domain hidden Markov models*. IEEE Trans. Image Process., 10(7):1056–1068, Ιούλιος 2001.
- [RHR94] Rose, R. C., E. M. Hofstetter, και D. A. Reynolds: *Integrated models of signal and background with application to speaker identification in noise*. IEEE Trans. Speech Audio Process., 2(2):245–257, 1994.
- [RM09] Roussos, A. και P. Maragos: *Reversible interpolation of vectorial images by an anisotropic diffusion-projection PDE*. Int. J. of Comp. Vis., 84(2):130–145, Αύγουστος 2009.
- [ROF92] Rudin, L.I., S. Osher, και E. Fatemi: *Nonlinear total variation based noise removal algorithms*. Physica D, 60:259–268, 1992.
- [Rou05] Roussos, A.: *Theory and applications of partial differential equations in computer vision*. Μεταπτυχιακή εργασία, Nat. Tech. Univ. of Athens, School of Elec. and Comp. Engin., Οκτώβριος 2005. in greek.

- [RP02] Rousson, M. και N. Paragios: *Shape priors for level set representations*. Στο *Proc. European Conf. on Computer Vision*, τόμος 2, σελίδες 78–92, 2002.
- [RS78] Rabiner, L. και R.W. Schafer: *Digital Processing of Speech Signals*. Prentice-Hall, NJ, USA, 1978.
- [RS87] Ruge, J. W. και K. Stüben: *Algebraic multigrid*. Στο Mc Cormick, S.F. (επιμελητής): *Multigrid Methods*. SIAM, 1987.
- [RTF⁺04] Raskar, R., K. H. Tan, R. Feris, J. Yu, και M. Turk: *Non-photorealistic camera: Depth edge detection and stylized rendering using multi-flash imaging*. *Proc. ACM Int. Conference on Computer Graphics and Interactive Techniques*, 23(3):679 – 688, Αύγουστος 2004.
- [Rud94] Ruderman, D.L.: *The statistics of natural images*. *Network: Computation in Neural Systems*, 5:517–548, 1994.
- [Rus09] Rusinkiewicz, S., 2009. personal communication.
- [RWCB03] Romberg, J.K., M.B. Wakin, H. Choi, και R.G. Baraniuk: *A geometric hidden Markov tree wavelet model*. Στο *Proc. of SPIE Wavelets X*, 2003.
- [Saa96] Saad, Y.: *Iterative Methods for Sparse Linear Systems*. PWS Publishing Comp., 1996.
- [SB86] Smith, M.J.T. και T.P. Barnwell: *Exact reconstruction techniques for tree-structured subband coders*. *IEEE Trans. Acoust., Speech, Signal Process.*, 34(3):434–441, Ιούνιος 1986.
- [SBB00] Sharon, E., A. Brandt, και R. Basri: *Fast multiscale image segmentation*. Στο *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, τόμος I, σελίδες 70–77, 2000.
- [SBK05] Selesnick, I.W., R.G. Baraniuk, και N. Kingsbury: *The dual-tree complex wavelet transform*. *IEEE Signal Process. Mag.*, 22(6):123–151, Νοέμβριος 2005.
- [Sch94] Schnörr, C.: *Unique reconstruction of piecewise smooth images by minimizing strictly convex non-quadratic functionals*. *J. Math. Imaging and Vision*, 4:189–198, 1994.
- [Sco02] Scott, S.L.: *Bayesian methods for hidden Markov models: Recursive computing in the 21st century*. *J. of the Amer. Stat. Assoc.*, 97(457):337–351, 2002.
- [SED04] Starck, J. L., M. Elad, και D. Donoho: *Redundant multiscale transforms and their application for morphological component separation*. *Adv. Imaging Electron. Phys.*, 132:287–348, 2004.
- [Sel01a] Selesnick, I.W.: *Hilbert transform pairs of wavelet bases*. *IEEE Signal Process. Lett.*, 8(6):170–173, Ιούνιος 2001.

- [Sel01b] Selesnick, I.W.: *Smooth wavelet tight frames with zero moments*. App. and Comp. Harm. Anal., 10:163–181, 2001.
- [Ser82] Serra, J.: *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [Ser88] Serra, J. (επιμελητής): *Image Analysis and Mathematical Morphology*, τόμος 2. Acad. Press, 1988.
- [Set99] Sethian, J. A.: *Level Set Methods and Fast Marching Methods*. Cambridge Univ. Press, δεύτερη έκδοση, 1999.
- [SF95] Simoncelli, E. P. και W. T. Freeman: *The steerable pyramid*. Στο Proc. IEEE Int. Conf. on Image Processing, τόμος 3, σελίδες 444–447, 1995.
- [SFAH92] Simoncelli, E. P., W. T. Freeman, E. H. Adelson, και D. J. Heeger: *Shiftable multiscale transforms*. IEEE Trans. Inf. Theory, 38(2):587–607, Μάρτιος 1992.
- [SG00] Salembier, P. και L. Garrido: *Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval*. IEEE Trans. Image Process., 9(4):561–576, Απρίλιος 2000.
- [SH96] Stork, D.G. και M.E. Hennecke (επιμελητές): *Speechreading by Humans and Machines*. Springer, Berlin, Germany, 1996.
- [Sha93] Shapiro, J. M.: *Embedded image coding using zerotrees of wavelet coefficients*. IEEE Trans. Signal Process., 41(12):3445–3462, Δεκέμβριος 1993.
- [She94] Shewchuk, J.R.: *An introduction to the conjugate gradient method without the agonizing pain*. unpublished manuscript, 1994.
- [She00] Sherratt, S. (επιμελητής): *The Wall Paintings of Thera: Proc. of the First Int. Symposium*. The Thera Foundation, 2000. Proc. of the 1997 Symposium, 2-vol. set.
- [Shl97] Shlien, S.: *The modulated lapped transform, its time-varying forms, and its applications to audio coding standards*. IEEE Trans. Speech Audio Process., 5(4):359–366, Ιούλιος 1997.
- [Sim05] Simoncelli, E. P.: *Statistical modeling of photographic images*. Στο Bovik, A.C. [Bov05], κεφάλαιο 4.7.
- [SLSZ03] Srivastava, A., A.B. Lee, E.P. Simoncelli, και S. C. Zhu: *On advances in statistical modeling of natural images*. J. Math. Imaging and Vision, 18:17–33, 2003.
- [Sol01] Solo, V.: *Automatic stopping criterion for anisotropic diffusion*. Στο Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing, τόμος 6, σελίδες 3929–3932, 2001.

- [SP96] Said, A. και W.A. Pearlman: *A new, fast, and efficient image codec based on set partitioning in hierarchical trees*. IEEE Trans. Circuits Syst. Video Technol., 6(3):243–250, Ιούνιος 1996.
- [SPS06] Spence, C., L.C. Parra, και P. Sajda: *Varying complexity in tree-structured image distribution models*. IEEE Trans. Image Process., 15(2):319–330, Φεβρουάριος 2006.
- [SS95] Salembier, P. και J. Serra: *Flat zones filtering, connected operators, and filters by reconstruction*. IEEE Trans. Image Process., 4:1153–1160, Αύγουστος 1995.
- [SSO94] Sussman, M., P. Smereka, και S. Osher: *A level set approach for computing solutions to incompressible two-phase flow*. Journal of Computational Physics, 114:146–159, 1994.
- [Ste81] Stein, C. M.: *Estimation of the mean of a multivariate normal distribution*. Ann. Stat., 9(6):1135–1151, 1981.
- [Sto74] Stone, M.: *Cross-validatory choice and assessment of statistical predictions (with discussion)*. J. of Royal Stat. Soc. (Series B), 36:111–147, 1974.
- [Str99] Strang, G.: *The discrete cosine transform*. SIAM Review, 41(1):135–147, 1999.
- [SW01] Schneider, M.K. και A.S. Willsky: *Krylov subspace estimation*. SIAM J. Scient. Comp., 22(5):1840–1864, 2001.
- [Sze06] Szeliski, R.: *Image alignment and stitching: a tutorial*. Found. and Trends in Comp. Graph. and Vision, 2(1):1–104, 2006.
- [TA77] Tikhonov, A.N. και V.Y. Arsenin: *Solutions of ill posed problems*. Winston, 1977.
- [TD05] Tschumperle, D. και R. Deriche: *Vector-valued image regularization with PDEs: A common framework for different applications*. IEEE Trans. Pattern Anal. Mach. Intell., 27(4):506–517, 2005.
- [Ter86] Terzopoulos, D.: *Image analysis using multigrid relaxation methods*. IEEE Trans. Pattern Anal. Mach. Intell., 8(2):129–139, 1986.
- [Ter88] Terzopoulos, D.: *The computation of visible-surface representations*. IEEE Trans. Pattern Anal. Mach. Intell., 10(4):417–438, 1988.
- [TN99] Tran, T.D. και T.Q. Nguyen: *A progressive transmission image coder using linear phase uniform filterbanks as block transforms*. IEEE Trans. Image Process., 8(11):1493–1507, Νοέμβριος 1999.
- [TOS01] Trottenberg, U., C. Oosterlee, και A. Schüller: *Multigrid*. Academic Press, 2001.

- [Tur52] Turing, A. M.: *The chemical basis of morphogenesis*. Phil. Trans. R. Soc. Lond. B (Biol. Sci.), 237(641):37–72, 1952.
- [Tur91] Turk, G.: *Generating textures on arbitrary surfaces using reaction-diffusion*. Στο *Proc. ACM Int. Conference on Computer Graphics and Interactive Techniques*, σελίδες 289–298, 1991.
- [TW87] Tanner, M.A. και W.H. Wong: *The calculation of posterior distributions by data augmentation*. J. of the Amer. Stat. Assoc., 82(398):528–540, 1987.
- [TYW01] Tsai, A., A. Yezzi, και A.S. Willsky: *Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification*. IEEE Trans. Image Process., 10(8):1169–1186, 2001.
- [UPC98] Urieli, S., M. Porat, και N. Cohen: *Optimal reconstruction of images from localized phase*. IEEE Trans. Image Process., 7(6):838–853, Ιούνιος 1998.
- [Vai93] Vaidyanathan, P.P.: *Multirate Systems and FilterBanks*. Prentice Hall, 1993.
- [VC02] Vese, Luminita A. και Tony F. Chan: *A multiphase level set framework for image segmentation using the Mumford and Shah model*. Int. J. of Comp. Vis., 50(3):271–293, 2002.
- [vH25] Helmholtz, H. von: *Physiological Optics*. Optical Soc. of America, New York, 1925. Transl. from the 3rd German Edition, 3-Vol. set.
- [VHLO83] Van Hove, P.L., M.H. Hayes, J.S. Lim, και A.V. Oppenheim: *Signal reconstruction from signed Fourier transform magnitude*. IEEE Trans. Acoust., Speech, Signal Process., 31(5):1286–1293, Οκτώβριος 1983.
- [vHvdS98] Hateren, J.H. van και A. van der Schaaf: *Independent component filters of natural images compared with simple cells in primary visual cortex*. Proc. R. Soc. Lond. B, 265:359–366, 1998.
- [VK95] Vetterli, M. και J. Kovacevic: *Wavelets and Subband Coding*. Prentice Hall, 1995.
- [Vla08] Vlachopoulos, A.G.: *The wall paintings from the Xeste 3 building at Akrotiri: Towards an interpretation of the iconographic programme*. Στο Brodie, N., J. Doole, G. Gavalas, και C. Renfrew (επιμελητές): *Horizons: A colloquium on the prehistory of the Cyclades*, κεφάλαιο 41, σελίδες 451–465. McDonald Institute Monographs, Cambridge, 2008.
- [VO03] Vese, L. A. και S. J. Osher: *Modeling textures with total variation minimization and oscillating patterns in image processing*. J. Scient. Comp., 19(1–3):553–572, 2003.
- [Vos88] Voss, R.: *Fractals in nature: From characterization to simulation*. Στο Peitgen, H. O. και D. Saupe (επιμελητές): *The Science of Fractal Images*, κεφάλαιο 1. Springer-Verlag, 1988.

- [Wal91] Wallace, G.K.: *The JPEG still picture compression standard*. Communications of the ACM, 34(4):30–44, Απρίλιος 1991.
- [WBSS04] Wang, Z., A.C. Bovik, H. Sheikh, και E.P. Simoncelli: *Image quality assessment: From error visibility to structural similarity*. IEEE Trans. Image Process., 13(4):600–612, 2004.
- [Wei74] Weinstock, R.: *Calculus of Variations*. Dover, 1974.
- [Wei98] Weickert, J.: *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, 1998.
- [Wei99] Weickert, J.: *Coherence-enhancing diffusion of colour images*. Image and Vision Comp., 17:201–212, 1999.
- [Wer23] Wertheimer, M.: *Untersuchungen zur Lehre von der Gestalt - II*. Psychologische Forschung, 4:301–350, 1923.
- [Wes92] Wesseling, P.: *An Introduction to Multigrid Methods*. Wiley, 1992. reprinted with corrections by R.T. Edwards, Inc, 2004.
- [Wil02] Willsky, A.: *Multiresolution Markov models for signal and image processing*. Proc. IEEE, 90(8):1396–1458, 2002.
- [Wit83] Witkin, A.P.: *Scale-space filtering*. Στο Proc. Int. Joint Conf. on Artificial Intel., σελίδες 1019–1022, 1983.
- [WJ96] Williams, L. R. και D. W. Jacobs: *Stochastic completion fields: A neural model of illusory contour shape and salience*. Neural Computation, 9:837–858, 1996.
- [WK91] Witkin, A. και M. Kass: *Reaction-diffusion textures*. Στο Proc. ACM Int. Conference on Computer Graphics and Interactive Techniques, σελίδες 299–308, 1991.
- [WK03] Weickert, J. και G. Kühne: *Fast methods for implicit active contour models*. Στο Osher, S. και N. Paragios (επιμελητές): *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer, 2003.
- [Wor93] Wornell, G. W.: *Wavelet-based representations for the $1/f$ family of fractal processes*. Proc. IEEE, 81(10):1428–1450, Οκτώβριος 1993.
- [WS02] Weickert, J. και H. Schar: *A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance*. Journal of Visual Communication and Image Representation, 13:103–118, 2002.
- [WS03] Wang, Z. και E.P. Simoncelli: *Local phase coherence and the perception of blur*. Στο Proc. Advances in Neural Information Processing Systems, 2003.
- [WSFZ07] Wu, Y.N., Z. Si, C. Fleming, και S. C. Zhu: *Deformable template as active basis*. Στο Proc. IEEE Int. Conf. on Computer Vision, 2007.

- [WSW01] Wainwright, M.J., E.P. Simoncelli, και A.S. Willsky: *Random cascades on wavelet trees and their use in modeling and analyzing natural imagery*. App. and Comp. Harm. Anal., 11(1):89–123, 2001.
- [WtHRV98] Weickert, J., B. ter Haar Romeny, και M. Viergever: *Efficient and reliable schemes for nonlinear diffusion filtering*. IEEE Trans. Image Process., 7(3):398–410, Μάρτιος 1998.
- [XGO96] Xiong, Z., O.G. Guleryuz, και M.T. Orchard: *A DCT-based embedded image coder*. IEEE Signal Process. Lett., 3(11):289–290, Νοέμβριος 1996.
- [YBFU96] Yang, G. Z., P. Burger, D. N. Firmin, και S. R. Underwood: *Structure adaptive anisotropic image filtering*. Image and Vision Comp., 14:135–145, 1996.
- [YEK⁺02] Young, S., G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, και P. Woodland: *The HTK book (for HTK version 3.2)*. Τεχνική αναφορά, Cambridge University Engineering Department, 2002.
- [YK93] Young, R.W. και N.G. Kingsbury: *Frequency-domain motion estimation using a complex lapped transform*. IEEE Trans. Image Process., 2(1):2–17, Ιανουάριος 1993.
- [YKK⁺97] Yezzi, A., S. Kichenassamy, A. Kumar, P. Olver, και A. Tannenbaum: *A geometric snake model for segmentation of medical imagery*. Medical Imaging, 16(2):199–209, 1997.
- [YMJ97] Yoma, N. B., F. McInnes, και M. Jack: *Weighted matching algorithms and reliability in noise canceling by spectral subtraction*. Στο Proc. IEEE Int. Conf. Acous., Speech, and Signal Processing, τόμος 2, σελίδες 1171–1174, 1997.
- [YTW99] Yezzi, A., A. Tsai, και A.S. Willsky: *A statistical approach to snakes for bimodal and trimodal imagery*. Στο Proc. IEEE Int. Conf. on Computer Vision, τόμος 2, σελίδες 898–903, 1999.
- [YV02] Yoma, N.B και M. Villar: *Speaker verification in noise using a stochastic version of the weighted viterbi algorithm*. IEEE Trans. Speech Audio Process., 10(3):158–166, 2002.
- [YW82] Youla, D.C. και H. Webb: *Image restoration by the method of convex projections: Part 1 - theory*. IEEE Trans. Med. Imag., 1(2):81–94, Οκτώβριος 1982.
- [ZM97] Zhu, S.C. και D. Mumford: *Prior learning and gibbs reaction-diffusion*. IEEE Trans. Pattern Anal. Mach. Intell., 19(11):1236–1250, Νοέμβριος 1997.
- [ZM06] Zhu, S.C. και D. Mumford: *A stochastic grammar for images*. Found. and Trends in Comp. Graph. and Vision, 2(4):259–362, 2006.
- [ZY96] Zhu, S. C. και A. Yuille: *Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation*. IEEE Trans. Pattern Anal. Mach. Intell., 18(9):884–900, Σεπτέμβριος 1996.

- [Βλα03] Βλαχόπουλος, Α.: *Βίρα-μάινα: Το χρονικό της συντήρησης μιας τοιχογραφίας από το Ακρωτήρι (1973-2003)*. Στο Βλαχόπουλος, Α. και Κ. Μπίρταχα (επιμελητές): *Αργοναύτης: Τιμητικός τόμος για τον Καθηγητή Χρίστο Γ. Ντούμα*, σελίδες 505–526. Καθημερινή, 2003.
- [ΓΜΣ92] Γέροντας, Α., Ι. Μιχαηλίδης, και Α. Στρατιάνης: *Η συντήρηση των τοιχογραφιών στο Ακρωτήρι*. Στο Ντούμας, Χ. [Ντο92], σελίδες 209–214.
- [Γεώ09] Γεώρμα, Φ.: *Οι Τοιχογραφίες του Κτηρίου Β από τον Προϊστορικό Οικισμό του Ακρωτηρίου Θήρας*. Διδακτορική διατριβή, Τμ. Ιστορίας-Αρχαιολογίας, Παν. Ιωαννίνων, 2009.
- [Ευα07] Ευαγγελόπουλος, Γ.: *Ανάλυση υφής και μικροδομών εικόνων με μοντέλα διαμόρφωσης, ενεργειακές και μεταβολικές τεχνικές: Ανίχνευση και διαχωρισμός*. Διδακτορική διατριβή, Σχ. Ηλ. Μηχ. & Μηχ. Υπολ., Ε.Μ.Π., 2007.
- [Μαρ02] Μαραγκός, Π.: *Όραση Υπολογιστών*. Ε.Μ.Π., 2002.
- [ΝΜΤ09] Ντούμας, Χ., Μ. Μαρθάρη, και Χ. Τελεβάντου: *Μουσείο Προϊστορικής Θήρας: Συνοπτικός Οδηγός*. Ταμείο Αρχ. Πόρων και Απαλλοτρ., τρίτη έκδοση, 2009.
- [Ντο92] Ντούμας, Χ. (επιμελητής): *Ακρωτήρι Θήρας: Είκοσι χρόνια έρευνας (1967-1987)*. Αριθμός 116 στο ΒΑΕ. Η εν Αθήναις Αρχαιολογική Εταιρεία, 1992.
- [Παλ99] Παλυβού, Κ.: *Ακρωτήρι Θήρας: Η Οικοδομική τέχνη*. Αριθμός 183 στο ΒΑΕ. Η εν Αθήναις Αρχαιολογική Εταιρεία, 1999.

Εκπαίδευση

- 2003–2009 **Διδακτορικό Δίπλωμα Ηλεκτρ. Μηχ. και Μηχ. Υπολ.**, Εθνικό Μετσόβιο Πολυτεχνείο.
Διατριβή: Ανάλυση Εικόνας και Όραση Υπολογιστών: Θεωρία και Εφαρμογές στην Αποκατάσταση Αρχαίων Τοιχογραφιών (υποστηρίχθηκε με επιτυχία τον Ιούλιο 2009)
Επιβλέπων: Καθ. Πέτρος Μαραγκός
- 1998–2003 **Δίπλωμα Ηλεκτρ. Μηχ. και Μηχ. Υπολ.**, Εθνικό Μετσόβιο Πολυτεχνείο.
Βαθμός: 9.54/10 (άριστα, κατετάγη στο άνω 1% της τάξης)
Διπλωματική Εργασία: Ταχείς Αλγόριθμοι για την Εξέλιξη Γεωδαιτικών Ενεργών Περιγραμμάτων με Εφαρμογές στην Όραση Υπολογιστών (*Επιβλέπων:* Καθ. Πέτρος Μαραγκός)

Ερευνητική Εμπειρία

- 2003–2009 **Διδακτορικός Ερευνητής**, Εθνικό Μετσόβιο Πολυτεχνείο.
Μέλος της ερευνητικής ομάδας Όρασης Υπολογιστών, Επικοινωνίας Λόγου & Επεξεργασίας Σήματος (CVSP). Συμμετοχή σε Ελληνικά και Ευρωπαϊκά ερευνητικά προγράμματα στις περιοχές της όρασης υπολογιστών και πολυτροπικής επεξεργασίας:
PENEΔ: Ανάπτυξη τεχνικών ενδοσυμπλήρωσης εικόνων με εφαρμογή στην ψηφιακή αποκατάσταση ελλειπών περιοχών στις αρχαίες τοιχογραφίες του προϊστορικού οικισμού Ακρωτηρίου Θήρας.
MUSCLE: Έρευνα σε αυτοματοποιημένη οπτική ανάλυση προσώπου και πολυτροπική σύμμιξη πληροφορίας για συνδυασμένη οπτική-ακουστική αναγνώριση ομιλίας. Ανάπτυξη ενός συστήματος πραγματικού χρόνου για οπτική-ακουστική αναγνώριση ομιλίας.
HIWIRE: Έρευνα σε αξιοποίηση οπτικών χαρακτηριστικών για εύρωστη αναγνώριση ομιλίας σε θορυβώδη ακουστικά περιβάλλοντα.
ASPI: Έρευνα σε αντιστροφή ομιλίας (ανάκτηση της γεωμετρίας της φωνητικής οδού) με συνδυασμό οπτικής και ακουστικής πληροφορίας.
- Καλοκαίρι 2006 **Επισκέπτης ερευνητής**, Trinity College Dublin, Ιρλανδία.
Αντικείμενο εργασίας: Ενδοσυμπλήρωση εικόνας με μιγαδικά κυματίδια.
Επιβλέπων: Καθ. Anil Kokaram.
- 2001–2003 **Προπτυχιακός Ερευνητής**, ΕΚΕΦΕ Δημόκριτος.
Μέλος του Ινστιτούτου Πληροφορικής και Τηλεπικοινωνιών, με συμμετοχή στο Ευρωπαϊκό ερευνητικό Έργο WIN. Αντικείμενο εργασίας η σχεδίαση και ανάπτυξη ασύρματου δικτύου ευρείας περιοχής (WAN) και συνοδών υπηρεσιών. Συνεισφορά στη συγγραφή ερευνητικών προτάσεων για επικείμενα Έργα.
Επιβλέπων: Δρ. Στέλιος Θωμόπουλος.

Υποτροφίες και Διακρίσεις

- 2006–2009 **Κοινωφελές Ίδρυμα Αλέξανδρος Σ. Ωνάσης**, Υποτροφία για διδακτορικές σπουδές.
- 2004–2007 **Ίδρυμα Κρατικών Υποτροφιών**, Υποτροφία για διδακτορικές σπουδές στην περιοχή της Τεχνητής Νοημοσύνης και Εφαρμογών.
- 1999–2003 **Κοινωφελές Ίδρυμα Ιωάννη Λάτση**, Υποτροφία για προπτυχιακές σπουδές.
- 2001 **Τεχνικό Επιμελητήριο Ελλάδας**, Βραβείο στους πρωτεύσαντες της Σχ. Ηλεκ. Μηχ. & Μηχ. Υπ. ΕΜΠ.
- 2000–2001 **Εθνικό Μετσόβιο Πολυτεχνείο**, Βραβεία Παπακυριακόπουλου (απενεμήθη δύο φορές, 2000 και 2001) και Κριτικού (2001) για διακεκριμένη επίδοση στα Μαθηματικά.

Επαγγελματικές Δραστηριότητες

Κριτής σε διεθνή επιστημονικά περιοδικά

- 2005–τώρα **IEEE Transactions on Image Processing.**
2006–τώρα **IEEE Transactions on Pattern Analysis and Machine Intelligence.**
2006 **International Journal of Computer Vision.**
2008 **Pattern Recognition.**

Μέλος επιστημονικών οργανώσεων

- 2003–τώρα **Institute of Electrical and Electronics Engineers (IEEE).**
2003–τώρα **Association for Computing Machinery (ACM).**
2008–τώρα **Society for Industrial and Applied Mathematics (SIAM).**
2004–τώρα **Τεχνικό Επιμελητήριο Ελλάδας.**

Γλώσσες

- Αγγλικά **Άριστα**
Γερμανικά **Καλά**
Ελληνικά **Μητρική**

Cambridge ESOL CPE (CEFR level C2)

Goethe-Institut ZMP (CEFR level B2)