Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Ηλεκτρολογων Μηχανικων
και Μηχανικων Υπολογιστων

Τομεας Σηματων, Ελεγχου και Ρομποτικης

Μέθοδοι Επεξεργασίας Πολυμεσικών Σημάτων για Βέλτιστη Αποθορυβοποίηση και Ανίχνευση με Εφαρμογές σε Δορυφορικά Δεδομένα

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΣΤΑΜΑΤΙΟΥ Σ. ΛΕΥΚΙΜΜΙΑΤΗ

Διπλωματούχου Μηχανικού Ηλεκτρονικών Υπολογιστών &
Πληροφορικής, Πολυτεχνικής Σχολής Πανεπιστημίου Πατρών.

Αθήνα, Δεκέμβριος 2009

Σχολη Ηλεκτρολογων Μηχανικων
και Μηχανικων Υπολογιστων

Τομεας Σηματων, Ελεγχου και Ρομποτικης

Μέθοδοι Επεξεργασίας Πολυμεσικών Σημάτων για Βέλτιστη Αποθορυβοποίηση και Ανίχνευση με Εφαρμογές σε Δορυφορικά Δεδομένα

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΣΤΑΜΑΤΙΟΥ Σ. ΛΕΥΚΙΜΜΙΑΤΗ

Διπλωματούχου Μηχανικού Ηλεκτρονικών Υπολογιστών &
Πληροφορικής, Πολυτεχνικής Σχολής Πανεπιστημίου Πατρών.

Συμβουλευτική Επιτροπή: Καθ. Πέτρος Μαραγκός (Επιβλέπων)
Καθ. Μιλτιάδης Αναγνώστου
Καθ. Γεώργιος Παπαβασιλόπουλος

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή στις 18 Δεκεμβρίου 2009:

.....	.....	.....
Π. Μαραγκός	Μ. Αναγνώστου	Γ. Παπαβασιλόπουλος
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.
.....	.....	
Σ. Κόλλιας	Κ. Τζαφέστας	
Καθηγητής Ε.Μ.Π.	Επ. Καθηγητής Ε.Μ.Π.	
.....	.....	
Κ. Μπερμπερίδης	Γ. Τσιροπούλα	
Καθηγητής Παν. Πατρών	Ερευν. Β΄ Αστεροσκοπείου Αθηνών	

Αθήνα, Δεκέμβριος 2009

Η παρούσα διδακτορική διατριβή πραγματοποιήθηκε στα πλαίσια του ερευνητικού 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 research 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%).

. . .

ΣΤΑΜΑΤΙΟΣ Σ. ΛΕΥΚΙΜΜΙΑΤΗΣ
Διπλωματούχος Μηχανικός Ηλεκτρονικών Υπολογιστών &
Πληροφορικής, Πολυτεχνικής Σχολής Πανεπιστημίου Πατρών.

Copyright © Σταμάτιος Λευκιμμιάτης, 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 η 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**

I Επεξεργασία Εικόνων σε Πολλαπλές Κλίμακες Ανάλυσης **7**

2 Αναπαράσταση Φυσικών Εικόνων με Τοπικά Εντοπισμένους Μετασχηματισμούς για Ανίχνευση Χαρακτηριστικών **9**

2.1 Αναπαράσταση Σημάτων και Εικόνων με Κυματίδια Εντοπισμένα στο Χώρο και την Κλίμακα . . . 10
2.1.1 Πολυκλιμακωτή Ανάλυση και Συστοιχίες Φίλτρων . . . 12
2.1.2 Συστοιχίες Φίλτρων Τέλειας Ανακατασκευής . . . 13
2.1.3 Διορθογώνιος Μετασχηματισμός Κυματιδίων . . . 14
2.1.4 Διακριτός Μετασχηματισμός Κυματιδίων . . . 15
2.1.5 Πολυκλιμακωτή Ανάλυση σε 2-Δ Σήματα . . . 16
2.1.6 Υπερπλήρεις Αναπαραστάσεις Σημάτων και Εικόνων . . . 18
2.2 Μη-Γραμμική Πολυκλιμακωτή Ανάλυση 1-Δ Σημάτων . . . 21
2.2.1 Haar Μετασχηματισμός . . . 22
2.2.2 Πολλαπλασιαστική Πολυκλιμακωτή Ανάλυση 1-Δ Σημάτων . . . 23
2.3 Μη-Γραμμική Πολυκλιμακωτή Ανάλυση Εικόνων . . . 26
2.3.1 Διαχωρίσιμος 2-Δ Πολλαπλασιαστικός Μετασχηματισμός . . . 26
2.3.2 Τετραδικός 2-Δ Πολλαπλασιαστικός Μετασχηματισμός . . . 29
2.3.3 Poisson-Haar Μετασχηματισμός . . . 30
2.3.4 Poisson-Haar Μετασχηματισμός και Συστοιχίες Φίλτρων . . . 34
2.4 Ιδιότητες Αναπαράστασης Εικόνων στο Πεδίο Τοπικά Εντοπισμένων Μετασχηματισμών 35
2.5 Σύνοψη και Συμπεράσματα . . . 37

3 Βέλτιστη Αποθορυβοποίηση Poisson Εικόνων με χρήση Μπεϋζιανής Συμπερασματολογίας και Δενδρικών Στατιστικών Μοντέλων **39**

3.1 Ανασκόπηση Βιβλιογραφίας και Κατηγοριοποίηση Μεθόδων . . . 40
3.2 Στατιστική Μοντελοποίηση Εικόνων . . . 42
3.3 Πολυκλιμακωτή Μοντελοποίηση Poisson Στοχαστικών Διαδικασιών . . . 44
3.3.1 Καθορισμός Προβλήματος . . . 44
3.3.2 Παραγοντοποίηση Πιθανοφάνειας 1-Δ Σημάτων και Στατιστική Μοντελοποίηση 45
3.3.3 Παραγοντοποίηση Πιθανοφάνειας Εικόνων και Στατιστική Μοντελοποίηση . . 47
3.4 Μπεϋζιανή Εκτίμηση Έντασης Φωτεινότητας . . . 51
3.4.1 Πολυκλιμακωτή Παραγοντοποίηση Ύστερης Πιθανότητας . . . 51
3.4.2 Μπεϋζιανή Εκτίμηση Μέσου Τετραγωνικού Σφάλματος . . . 53
3.4.3 Επαναληπτική Εκτίμηση και Κυκλικές Μετατοπίσεις . . . 53
3.5 Εκτίμηση Παραμέτρων με τη χρήση του ΕΜ αλγορίθμου . . . 54
3.6 Μοντελοποίηση Εξαρτήσεων Μεταξύ Κλιμάκων με Κρυφά Μαρκοβιανά Δένδρα . . . 58

3.7 Πειράματα και Συγκρίσεις 62
3.7.1 Συγκρίσεις με Μπεϋζιανές Μεθόδους Σχεδιασμένες Αποκλειστικά για Poisson Δεδομένα 63
3.7.2 Συγκρίσεις με VST Μεθόδους 65
3.7.3 Αποθορυβοποίηση Εικόνων με Πραγματικό Poisson Θόρυβο 71
3.8 Εφαρμογές σε Αστρονομικά Δορυφορικά Δεδομένα 71
3.8.1 Δορυφορικές Εικόνες με Συνθετικό Poisson Θόρυβο 73
3.8.2 Δορυφορικές Εικόνες με Πραγματικό Θόρυβο 76
3.8.3 Ανίχνευση σε Δορυφορικές Εικόνες 76
3.9 Σύνοψη και Συμπεράσματα 79
3.Α΄ Παραγοντοποίηση της Poisson Συνάρτησης Πιθανοφάνειας σε Τετραδικό Δένδρο . . 80
3.Β΄ Ανάπτυξη της Πολυωνυμικής Κατανομής σε Γινόμενο Δυωνυμικών Κατανομών . . . 80
3.Γ΄ Ισοδυναμία της MMSE Εκτίμησης στο Πεδίο Ορισμού της Εικόνας και στο Μετασχηματισμένο Πεδίο 81
3.Δ΄ Ποινικοποιημένος Αλγόριθμος Μεγιστοποίησης Προσδοκίας 82

II Πολυαισθητηριακή Επεξεργασία Ακουστικών Σημάτων 85

4 Βέλτιστη Αποθορυβοποίηση Ακουστικών Σημάτων με τη Χρήση Συστοιχίας Ψηφιακών Αισθητήρων 87
4.1 Καθορισμός Προβλήματος 88
4.1.1 Χωρική Δειγματοληψία 90
4.1.2 Πηγές Κοντινού και Μακρινού Πεδίου 91
4.2 Βασικές Τεχνικές Χωρικού Φιλτραρίσματος 91
4.2.1 Κέρδος Συστοιχίας 92
4.2.2 Υπερκατευθυντικό Φιλτράρισμα Διαμόρφωσης Δέσμης 94
4.3 Βέλτιστη Αποθορυβοποίηση Ακουστικών Σημάτων 94
4.3.1 Πολυκαναλικός MMSE Εκτιμητής 95
4.3.2 Βέλτιστοι Μη-Γραμμικοί Εκτιμητές 96
4.3.2.1 Πολυκαναλικός MMSE-STSA Εκτιμητής 97
4.3.2.2 Πολυκαναλικός MMSE log-STSA Εκτιμητής 98
4.4 Εκτίμηση Ύστερου-Φίλτρου 98
4.4.1 Ανάλυση του Πεδίου Θορύβου 99
4.4.2 Γενικευμένη Εκτίμηση Ύστερου-Φίλτρου 100
4.4.2.1 Εκτίμηση Φάσματος Πηγαίου Σήματος 100
4.4.2.2 Εκτίμηση Φάσματος Θορύβου 102
4.5 Πειραματικές Μετρήσεις και Αποτελέσματα 103
4.5.1 Συλλογή Ακουστικών Δεδομένων και Πρακτική Υλοποίηση του Συστήματος Αποθορυβοποίησης 103
4.5.2 Πειράματα Αποθορυβοποίησης Ακουστικών Σημάτων 104
4.6 Σύνοψη και Συμπεράσματα 107

5 Πολυαισθητηριακή Επεξεργασία για την Αυτόματη Ανίχνευση και Αναγνώριση Ακουστικών Σημάτων 109
5.1 Επεξεργασία Ακουστικών Σημάτων 110
5.1.1 Συστοιχίες Φίλτρων 110
5.1.1.1 Μη-Γραμμική Mel και Bark Κλίμακα Συχνοτήτων 111
5.1.2 Φίλτρα για Ακουστική Επεξεργασία 112
5.1.3 Ενεργειακές Μετρήσεις 113

5.1.3.1 Ενεργειακός Τελεστής Teager-Kaiser . . . 113
5.2 Ενοποίηση Ενεργειακών Μετρήσεων . . . 115
5.2.1 Ισοδυναμία Ενεργειακών Μετρήσεων . . . 115
5.2.2 Ερμηνεία στο Συχνοτικό Πεδίο . . . 118
5.2.3 Επέκταση Ενεργειακού Τελεστή για Μιγαδικά Σήματα . . . 118
5.2.4 Ανάλυση Ενεργειακών Μετρήσεων σε Θόρυβο . . . 119
5.2.5 Επέκταση Αποτελεσμάτων σε Σήματα Διακριτού Χρόνου . . . 120
5.3 Πολυαισθητηριακή-Πολυφασματική Αποδιαμόρφωση Ακουστικών Σημάτων . . . 121
5.3.1 Πολυαισθητηριακή - Πολυφασματική Αποδιαμόρφωση και Ενεργειακές Μετρήσεις . . . 122
5.3.2 Πολυαισθητηριακή Εξαγωγή Ενεργειακών Χαρακτηριστικών . . . 125
5.4 Πειραματικά Αποτελέσματα . . . 128
5.4.1 Περιγραφή Βάσης Δεδομένων και Στατιστικού Συστήματος . . . 129
5.4.2 Αποτελέσματα Αναγνώρισης Φωνής . . . 130
5.5 Σύνοψη και Συμπεράσματα . . . 131

6 Σύνοψη Διατριβής και Μελλοντική Έρευνα **133**
6.1 Ερευνητική Συνεισφορά . . . 133
6.2 Κατευθύνσεις Μελλοντικής Έρευνας . . . 134

Α΄ Δημοσιεύσεις Διατριβής **137**

Βιβλιογραφία **139**

Κατάλογος Σχημάτων

2.1 Ιδιότητες χρονικού/φασματικού εντοπισμού του βραχύχρονου μετασχηματισμού Fourier και της κυματιδιακής ανάλυσης με τη βοήθεια των Heisenberg διαγραμμάτων (α) Κυματίδια χρόνου-συχνότητας όπου τα άτομα εντός κάθε χρονικού παραθύρου καλύπτουν ομοιόμορφα τον άξονα συχνοτήτων. (β) Κυματίδια χρόνου-κλίμακας, σε διαδοχικές κλίμακες ανάλυσης υποδιπλασιάζεται η χρονική διάρκεια του φίλτρου (κάλυψη του χρονικού άξονα), ενώ παράλληλα διπλασιάζεται το εύρος ζώνης του (κάλυψη του άξονα συχνοτήτων). Ανατυπωμένο σχήμα από το [VK95]. 11

2.2 Ένα στάδιο κυματιδιακής ανάλυσης (α) και σύνθεσης (β) με τον αναδρομικό ταχύ μετασχηματισμό κυματιδίων που βασίζεται σε συστοιχίες φίλτρων. Στην περίπτωση συζυγών ανακλαστικών φίλτρων, το δυϊκό ζεύγος φίλτρων $\left\{\tilde{h}_0, \tilde{h}_1\right\}$ είναι το ίδιο με το ζεύγος φίλτρων $\{h_0, h_1\}$. 13

2.3 Τρία τυπικά μητρικά κυματίδια που αντιστοιχούν στο 2-Δ διαχωρίσιμο μετασχηματισμό DWT. (α) LH κυματίδιο, (β) HL κυματίδιο, (γ) HH κυματίδιο. 16

2.4 Απόλυτη τιμή των συντελεστών κυματιδίων που προκύπτουν από την ανάλυση της 512×512 εικόνας boat μέσω του 2-Δ DWT για 2 επίπεδα ανάλυσης. Πάνω δεξιά παρουσιάζονται οι συντελεστές που σχετίζονται με το LH κυματίδιο (κωδικοποίηση οριζόντιων ακμών), κάτω αριστερά οι συντελεστές που σχετίζονται με το HL κυματίδιο (κωδικοποίηση κατακόρυφων ακμών) και κάτω δεξιά οι συντελεστές που σχετίζονται με το HH κυματίδιο (κωδικοποίηση διαγώνιων ακμών). 17

2.5 (α) Ο μη-δειγματοληπτημένος DWT μετασχηματισμός (á trous αλγόριθμος) υλοποιείται μέσω της επαναληπτικής εφαρμογής της συστοιχίας φίλτρων, όπου σε κάθε στάδιο (κλίμακα ανάλυσης) γίνεται χρήση των αρχικών φίλτρων με κατάλληλη ένθεση μηδενικών. (β) Το αρχικό σήμα ανακατασκευάζεται μέσω συνελίξεων με τα δυϊκά φίλτρα. Ο πολλαπλασιασμός με τον παράγοντα $1/2$ είναι απαραίτητος για την τέλεια ανακατασκευή. 20

2.6 Πολυκλιμακωτή ανάλυση της εικόνας house μεγέθους 256×256 με τον UDWT μετασχηματισμό για δύο κλίμακες ανάλυσης. Από αριστερά προς τα δεξιά παρουσιάζονται οι συντελεστές κλιμάκωσης και οι απόλυτες τιμές των συντελεστών λεπτομέρειας για τρεις διαφορετικές κατευθύνσεις (οριζόντια, κατακόρυφη, διαγώνια). Τα επίπεδα ανάλυσης εμφανίζονται με σειρά από κάτω προς τα πάνω, δηλαδή το ευκρινέστερο επίπεδο βρίσκεται στην δεύτερη γραμμή και το αδρότερο στην πρώτη γραμμή. . . . 21

2.7 Δυαδικό δένδρο παραγόμενο από πολυκλιμακωτή (ευκρινή-αδρή) ανάλυση του $\boldsymbol{\lambda}$ σε 3 κλίμακες. Γνωρίζοντας τις τιμές των παραγόντων διαχωρισμού $\theta_j(k)$ και την συνολική τιμή του σήματος, $\lambda_J(0)$, μπορούμε να ανακτήσουμε επακριβώς την αρχική αναπαράσταση του $\boldsymbol{\lambda}$. 24

2.8 (α) Πολυκλιμακωτή αναπαράσταση ενός 1-Δ σήματος στο πεδίο του πολλαπλασιαστικού μετασχηματισμού μέσω επαναληπτικής ανάλυσης του σε υποζώνες με τη χρήση συστοιχίας φίλτρων. Σε κάθε στάδιο ανάλυσης αφού υποδειγματοληπτηθούν οι έξοδοι των φίλτρων h_0, h_1, η υποδειγματοληπτημένη έξοδος του φίλτρου h_1 κανονικοποιείται από την υποδειγματοληπτημένη έξοδο του φίλτρου h_0. (β) Τέλεια ανακατασκευή του σήματος ακολουθώντας αντίστροφη διαδικασία φιλτραρίσματος. 25

2.9 Απεικόνιση της γειτονιάς εικονοστοιχιων $C_{k,l}$ μεγέθους 2×2 σε μια κλίμακα $j-1$, η οποία συνδέεται με ένα εικονοστοιχείο της εικόνας με συντεταγμένες (k, ℓ) σε μια κλίμακα j. 26

2.10 Πολυκλιμακωτή ανάλυση 2-Δ εικόνας χρησιμοποιώντας το διαχωρίσιμο σχήμα παραγοντοποίησης των [TN99], [Now99a]. 27

2.11 Τιμή των συντελεστών που προκύπτουν από την ανάλυση της 512×512 εικόνας Lena μέσω του 2-Δ διαχωρίσιμου σχήματος για 2 επίπεδα ανάλυσης. Πάνω δεξιά αντιστοιχούν οι $\theta_j^{(1)}(k, \ell)$ συντελεστές, κάτω αριστερά οι $\theta_j^{(2)}(k, \ell)$ συντελεστές και κάτω δεξιά οι $\theta_j^{(3)}(k, \ell)$ συντελεστές. 28

2.12 Πολυκλιμακωτή ανάλυση 2-Δ εικόνας χρησιμοποιώντας το τετραδικό σχήμα ανάλυσης των [Now99a, ECKVD04]. 29

2.13 Τιμή των συντελεστών που προκύπτουν από την ανάλυση της 512×512 εικόνας Lena μέσω του 2-Δ τετραδικού μετασχηματισμού για 2 επίπεδα ανάλυσης. Πάνω δεξιά αντιστοιχούν οι $\theta_j^{(1)}(k, \ell)$ συντελεστές, κάτω αριστερά οι $\theta_j^{(2)}(k, \ell)$ συντελεστές και κάτω δεξιά οι $\theta_j^{(3)}(k, \ell)$ συντελεστές. Σύγκρινε με το αντίστοιχο Σχ. 2.11 για το διαχωρίσιμο 2-Δ μετασχηματισμό. 31

2.14 Πολυκλιμακωτή αναπαράσταση εικόνας μέσω του Poisson-Haar μετασχηματισμού. Και τα τρία δένδρα καταλήγουν στην ίδια τιμή του κόμβου-πατέρα $\lambda_j(k, \ell)$ στην j κλίμακα. Η διαφορά ανάμεσα τους έγκειται στον τρόπο που δημιουργείται η ενδιάμεση κλίμακα, σχεδιασμένη έτσι ώστε οι 3 παράγοντες διαχωρισμού $\theta^h, \theta^v, \theta^d$ της Εξ. (2.32) να παρουσιάζουν ευαισθησία σε ακμές εικόνων διαφορετικού προσανατολισμού (οριζόντιος, κάθετος, ή διαγώνιος). 32

2.15 Τιμή των συντελεστών που προκύπτουν από την ανάλυση της 512×512 εικόνας Lena μέσω του 2-Δ Poisson-Haar μετασχηματισμού για 2 επίπεδα ανάλυσης. Πάνω δεξιά αντιστοιχούν οι $\theta_j^{(h)}(k, \ell)$ συντελεστές, κάτω αριστερά οι $\theta_j^{(v)}(k, \ell)$ συντελεστές και κάτω δεξιά οι $\theta_j^{(d)}(k, \ell)$ συντελεστές. Σύγκρινε με τα αντίστοιχα Σχ. 2.11 και 2.13 για το διαχωρίσιμο και τετραδικό 2-Δ μετασχηματισμό. 33

2.16 Poisson-Haar πολυκλιμακωτή ανάλυση εικόνας μέσω της ανάλυσης της σε συχνοτικές υποζώνες με τη χρήση συστοιχιών φίλτρων. (α) Συστοιχία φίλτρων ανάλυσης, (β) Συστοιχία φίλτρων σύνθεσης. 34

2.17 Πολυκλιμακωτή ανάλυση της εικόνας house μεγέθους 256×256 με τον μη-κρίσιμα δειγματοληπτημένο Poisson-Haar μετασχηματισμό για δύο κλίμακες ανάλυσης. Από αριστερά προς τα δεξιά παρουσιάζονται οι συντελεστές κλιμάκωσης και οι συντελεστές διαχωρισμού για τρεις διαφορετικές κατευθύνσεις (οριζόντια, κατακόρυφη, διαγώνια). Τα επίπεδα ανάλυσης εμφανίζονται με σειρά από κάτω προς τα πάνω, δηλαδή το ευκρινέστερο επίπεδο βρίσκεται στην δεύτερη γραμμή και το αδρότερο στην πρώτη γραμμή. 36

3.1 Ιστόγραμμα των συντελεστών λόγου έντασης $\theta_1^{(2)}(k)$ της πηγαίας εικόνας Lena, δες Εξ. (2.30), μαζί με ταιριασμένες μια (διακεκομμένη γραμμή) και τρεις (συμπαγής γραμμή) συνιστώσες από ένα μείγμα συμμετρικών ($\alpha_{jm} = \beta_{jm}$) βήτα κατανομών. Οι παράμετροι του μοντέλου προκύπτουν μεγιστοποιώντας την πιθανοφάνεια· στην περίπτωση των πολλαπλών συνιστωσών γίνεται χρήση του EM αλγορίθμου. 47
3.2 Αντιστοίχιση των τιμών των συντελεστών $\lambda_j(k,\ell)$ και $x_j(k,\ell)$ στους κόμβους ενός τετραδικού δένδρου. Κάθε κόμβος του δένδρου, εκτός των κλαδιών, συνδέεται με τέσσερις κόμβους του αμέσως επόμενου επιπέδου. 49
3.3 Η μερική παράγωγος της λογαριθμικής συνάρτησης πιθανοφάνειας ως προς τη συμμετρική παράμετρο της βήτα κατανομής α_m και η ομαλοποιημένη εκδοχή της. Οι τρεις γραφικές αναπαραστάσεις αντιστοιχούν στη συνιστώσα με τη μικρή (αριστερά), μεσαία (κέντρο), και μεγάλη (δεξιά) τιμή της α_m παραμέτρου. Η επιβαλλόμενη ομαλοποίηση φράζει αποτελεσματικά τη ρίζα της συνιστώσας με την μεγαλύτερη παράμετρο, ενώ ουσιαστικά αφήνει ακέραιες τις δύο άλλες ρίζες. 57
3.4 Από κοινού ιστογράμματα και οι υπό συνθήκη κατανομές για δύο διαφορετικά σύνολα συντελεστών διαχωρισμού θ ("πατέρα"-"παιδιού"). (α)-(γ) Από κοινού ιστόγραμμα των συντελεστών του συνόλου "παιδιού" (λεπτομερής κλίμακα ανάλυσης) και του συνόλου "πατέρα" (αδρότερη κλίμακα ανάλυσης) για μια συνθετική εικόνα γκαουσιανού θορύβου και μια τυπική φυσική εικόνα, αντιστοίχως. (β)-(δ) Από κοινού ιστόγραμμα των συντελεστών του συνόλου "παιδιού" και του συνόλου "πατέρα" που προέκυψαν στην ίδια κλίμακα ανάλυσης αλλά για διαφορετικούς προσανατολισμούς για μια συνθετική εικόνα γκαουσιανού θορύβου και μια τυπική φυσική εικόνα, αντιστοίχως. 59
3.5 Η δομή του κρυφού μαρκοβιανού δένδρου (βαθμού 4) που χρησιμοποιείται για τη μοντελοποίηση δια-κλιμακωτών εξαρτήσεων των συντελεστών διαχωρισμού εικόνων. Οι λευκοί κόμβοι συμβολίζουν τις διακριτές κρυφές μεταβλητές $\mathbf{z}_n$, ενώ οι σκιασμένοι κόμβοι συμβολίζουν τις παρατηρήσεις. Τα βέλη δηλώνουν στατιστικές εξαρτήσεις με τη φορά του βέλους να καθορίζει ποια είναι η εξαρτημένη τυχαία μεταβλητή. Στην περίπτωση που εξαλειφθούν οι εξαρτήσεις μεταξύ των κρυφών καταστάσεων (λευκοί κόμβοι), τότε οδηγούμαστε στο ανεξάρτητο δένδρο (hidden independent tree -HIT) της Ενοτ. 3.5 60
3.6 Το σύνολο όλων των εικόνων που χρησιμοποιούνται για τις συγκρίσεις μεταξύ των αλγορίθμων. Οι εικόνες είναι, από αριστερά προς τα δεξιά, οι Lena, Barbara, boat, fingerprint και cameraman. 63
3.7 Αποτελέσματα αποθορυβοποίησης της εικόνας cameraman με μέγιστη τιμή έντασης 20 και συνθετικό Poisson θόρυβο. Ο μέσος αριθμός φωτονίων ανά εικονοστοιχείο στη θορυβώδη εικόνα είναι 9.4. (α) Θορυβώδης εικόνα (PSNR=16.29), (β) Quad-HMT (PSNR=26.23), (γ) TN [TN99] (PSNR=24.35), (δ) LKA [LKA04] (PSNR=24.67). . . . 66
3.8 Αποτελέσματα αποθορυβοποίησης της εικόνας boat με μέγιστη τιμή έντασης 10 και συνθετικό Poisson θόρυβο. Ο μέσος αριθμός φωτονίων ανά εικονοστοιχείο στη θορυβώδη εικόνα είναι 5. (α) Θορυβώδης εικόνα (PSNR=12.94), (β) PH-HMT (PSNR=25.57),(γ) TN [TN99] (PSNR=24.23), (δ) LKA [LKA04] (PSNR=24.51). 67
3.9 Αποτελέσματα αποθορυβοποίησης της εικόνας Lena με μέγιστη τιμή έντασης 40 και συνθετικό Poisson θόρυβο. Ο μέσος αριθμός φωτονίων ανά εικονοστοιχείο στη θορυβώδη εικόνα είναι 2.5. (α) Αλλοιωμένη εικόνα (PSNR=9.95), (β) PH-HMT(PSNR=25.78), (γ) BM3D [DFKE07] (PSNR=26.52), (δ) BLS-GSM [PSWS03] (PSNR=26.16). Σημειώνουμε ότι, παρόλο που η προτεινόμενη μέθοδος δίνει χαμηλότερο PSNR, εισάγει σαφώς λιγότερα οπτικά τεχνουργήματα συγκριτικά με τις άλλες δύο VST μεθόδους. 69

3.10 Αποθορυβοποίηση βιοϊατρικών εικόνων. (α) Αρχική εικόνα κοιλίας, (β) Αποτέλεσμα εκτίμησης με την προτεινόμενη Quad-HMT μέθοδο, (γ) Αρχική εικόνα θώρακα, (δ) Αποτέλεσμα εκτίμησης με την προτεινόμενη PH-HMT μέθοδο. 72
3.11 Αποθορυβοποίηση αστρονομικών δορυφορικών εικόνων με την PH-HMT μέθοδο. (α) Πηγαία εικόνα του αστρικού συστήματος Eta Carinae, (β) Θορυβώδης εικόνα (PSNR=22.10 dB), (γ) Αποτέλεσμα εκτίμησης (PSNR=30.62 dB). (δ) Πηγαία εικόνα του γαλαξία Whirlpool, (ε) Θορυβώδης εικόνα (PSNR=21.80 dB), (στ) Αποτέλεσμα εκτίμησης (PSNR=32.86 dB). 73
3.12 Αποθορυβοποίηση αστρονομικών δορυφορικών εικόνων με την Quad-HMT μέθοδο. (α) Πηγαία εικόνα του γαλαξία Sextans B, (β) Θορυβώδης εικόνα (PSNR=24.92 dB), (γ) Αποτέλεσμα εκτίμησης (PSNR=34.33 dB). (δ) Πηγαία εικόνα του γαλαξία Messier 87, (ε) Θορυβώδης εικόνα (PSNR=22.75 dB), (στ) Αποτέλεσμα εκτίμησης (PSNR=34.98 dB). 74
3.13 Αποθορυβοποίηση αστρονομικών δορυφορικών εικόνων με την Sep-HMT μέθοδο. (α) Πηγαία εικόνα του γαλαξία Pinwheel, (β) Θορυβώδης εικόνα (PSNR=21.54 dB), (γ) Αποτέλεσμα εκτίμησης (PSNR=30.39 dB). (δ) Πηγαία εικόνα του γαλαξία Messier 99, (ε) Θορυβώδης εικόνα (PSNR=24.54 dB), (στ) Αποτέλεσμα εκτίμησης (PSNR=34.06 dB). 75
3.14 Αποθορυβοποίηση αστρονομικών δορυφορικών εικόνων που καταγράφηκαν από το δορυφόρο XMM-Newton. (α)-(γ) Περιορισμένης φωτεινότητας αστρονομικές εικόνες. (δ)-(στ) Αποτέλεσμα εκτίμησης με τις προτεινόμενες Sep-HMT, Quad-HMT και PH-HMT μεθόδους, αντιστοίχως. 77
3.15 (α) Εικόνα με πολλαπλές αστρονομικές πηγές που καταγράφηκε από το δορυφόρο Chandra. (α) Αρχική καταγραφή, (β) Αποτέλεσμα επεξεργασίας των δεδομένων με τη μέθοδο PH-HMT. 78

4.1 Συστοιχία 5 αισθητήρων για τη σύλληψη ακουστικών σημάτων. 89
4.2 Φιλτράρισμα διαμόρφωσης δέσμης στο κοινό χρονο-συχνοτικό πεδίο. 92
4.3 Πολυαισθητηριακό σύστημα αποθορυβοποίησης ακουστικών σημάτων με χρήση ύστερου-φιλτραρίσματος. 101
4.4 Απεικόνιση του παράγοντα κατευθυντικότητας του MVDR διαμορφωτή δέσμης ο οποίος περιγράφει την ικανότητα του διαμορφωτή δέσμης να περιορίσει το θόρυβο. Η ικανότητά του αυτή παρουσιάζεται μικρή στην περιοχή των κλιμάκων βραχειών συχνοτήτων. 104
4.5 Φασματογραφήματα της εκφώνησης “r-e-w-y-8-56”. (α) Καθαρό ακουστικό σήμα. (β) Θορυβώδης παρατήρηση στον κεντρικό αισθητήρα της συστοιχίας (IS=1.44). (γ) Παρατηρήσιμο ακουστικό σήμα στην έξοδο του MVDR διαμορφωτή δέσμης (SSNRE=0.02 dB, IS=0.90). (δ) Γενικευμένο ύστερο-φίλτρο Zelinski (SSNRE=0.17 dB, IS=2.89). (ε) Ύστερο φίλτρο McCowan (SSNRE=3.95 dB, IS=2.08). (στ) MMSE (SSNRE=4.54 dB, IS=0.81). (ζ) STSA (SSNRE=4.46 dB, IS=0.82). (η) log-STSA (SSNRE=4.52 dB, IS=0.81). 106

5.1 Ομοιόμορφη τοποθέτηση των φίλτρων στη γραμμική, Mel και Bark κλίμακα συχνοτήτων. 111
5.2 (α) Κρουστική απόκριση και (β) απόκριση συχνότητας του Gabor φίλτρου με κεντρική συχνότητα $f_c = 2000$ Hz και ενεργό εύρος ζώνης 200 Hz. 112
5.3 (α) Κρουστική απόκριση και (β) απόκριση συχνότητας του Gammatone φίλτρου με κεντρική συχνότητα $f_c = 2000$ Hz και εύρος ζώνης $b = 200$ Hz. 113
5.4 Συστοιχία από (α) 11 τριγωνικά φίλτρα ομοιόμορφα διατεταγμένα στην Mel κλίμακα και (β) 11 Gabor φίλτρα ομοιόμορφα διατεταγμένα στη Bark κλίμακα συχνοτήτων. 114

5.5 Ανάλυση των πολλαπλών σημάτων εισόδου στη διάταξη των αισθητήρων σε φασματικές συνιστώσες. Οι φασματικές συνιστώσες προκύπτουν από το φιλτράρισμα των εισόδων με μια συστοιχία φίλτρων. 123
5.6 Ενεργόγραμμα διάταξης 8 αισθητήρων. Το γράφημα αυτό απεικονίζει σε ποιον από τους αισθητήρες της διάταξης υπολογίζεται η πιο αξιόπιστη ενεργειακή μέτρηση (η πλησιέστερη στην πραγματική μέτρηση) ανά συχνοτική υποζώνη και χρονικό πλαίσιο. 126
5.7 Πολυαισθητηριακή εξαγωγή ενεργειακών χαρακτηριστικών. Πολυκαναλικό φιλτράρισμα σε κάθε αισθητήρα της διάταξης, εντοπισμός των cross-ενεργειών (Teager ή τυπικής ενέργειας) και επιλογή της ελάχιστης μέσης τιμής ανά συχνοτική υποζώνη. Οι επιλεγόμενες ενεργειακές μετρήσεις χρησιμοποιούνται για την εξαγωγή του τελικού διανύσματος χαρακτηριστικών που θα αξιοποιηθεί στην αυτόματη αναγνώριση. Τα παράθυρα ανάλυσης είναι μήκους N δειγμάτων. 127
5.8 Σχεδιάγραμμα του δωματίου στο οποίο πραγματοποιήθηκαν οι καταγραφές των εκφωνήσεων. 129

Κατάλογος Πινάκων

3.1 Ορισμοί των κατανομών των πιθανοτήτων που θα χρησιμοποιηθούν στο Κεφ. 3. Η συνάρτηση γάμα ορίζεται ως $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\mathrm{d}t$, D είναι η διάσταση των τυχαίων διανυσμάτων, και $n = \sum_{k=1}^{D} x_k$. 45
3.2 Εκτίμηση της έντασης φωτεινότητας εικόνων αλλοιωμένων από συνθετικό Poisson θόρυβο χρησιμοποιώντας μείγμα κατανομών με 3 συμμετρικές συνιστώσες. Τα αποτελέσματα παρουσιάζονται σε όρους PSNR (dB) για διάφορα επίπεδα Poisson θορύβου. 64
3.3 Συγκριτικά αποτελέσματα της PH-HMT μεθόδου με δύο VST μεθόδους που χρησιμοποιούν τις BLS-GSM [PSWS03] και BM3D [DFKE07] τεχνικές ως ενδιάμεσες ρουτίνες για γκαουσιανή αποθορυβοποίηση. Τα αποτελέσματα παρουσιάζονται σε όρους PSNR (dB) για διάφορες τιμές εντάσεων που αντιστοιχούν σε διαφορετικά επίπεδα Poisson θορύβου. 70
3.4 Συγκριτικά αποτελέσματα της PH-HMT μεθόδου με τα καλύτερα αποτελέσματα της MSVST μεθόδου, τα οποία αναφέρονται στο [ZFS08], για τέσσερις εικόνες (Spots, Galaxy, Barbara, Cells). Τα αποτελέσματα παρουσιάζονται σε όρους NMISE (η μικρότερη τιμή υποδηλώνει καλύτερη απόδοση). 71

4.1 Ποιοτικές μετρήσεις στα δεδομένα που προέκυψαν από τα πειράματα στη βάση δεδομένων CMU 105

5.1 Αποτελέσματα αναγνώρισης (μέση ακρίβεια αναγνώρισης λέξεων (%) για διαφορετικές συνθήκες εκπαίδευσης/ελέγχου) με τη μέτρηση της τυπικής ενέργειας. 130
5.2 Αποτελέσματα αναγνώρισης (μέση ακρίβεια αναγνώρισης λέξεων (%) για διαφορετικές συνθήκες εκπαίδευσης/ελέγχου) με τη μέτρηση της Teager ενέργειας. 130

ΠΡΟΛΟΓΟΣ

Με την παρούσα διατριβή ολοκληρώνεται μια μακρόχρονη πορεία μελέτης και συνεχούς προσπάθειας στους γοητευτικούς δρόμους της επιστημονικής έρευνας. Η εκπόνησή της αποτελεί έναν σημαντικό όσο και καθοριστικό σταθμό, ορόσημο στην προσωπική μου ακαδημαϊκή διαδρομή. Σηματοδοτεί την ολοκλήρωση μιας δύσκολης ερευνητικής προσπάθειας, αλλά και τη χαρά και δικαίωση προσωπικών στόχων και επιθυμιών. Το τελικό περιεχόμενο της διατριβής διαμορφώθηκε ύστερα από επαναπροσδιορισμό των αρχικών αναζητήσεων οι οποίες αποκρυσταλλώθηκαν στην πορεία θέτοντας τα τελικά ερωτήματα και τους τελικούς στόχους. Πέρα όμως από το τελικό αποτέλεσμα μεγαλύτερη σημασία για μένα έχει η διαδρομή καθεαυτή χάρη στην οποία αποκόμισα γνώσεις και εμπειρίες που αποτελούν σημαντική παρακαταθήκη για τη συνέχιση του ταξιδιού σε νέους προορισμούς.

Το ταξίδι αυτό των 5 ετών μέχρι τη συγγραφή και παρουσίαση της διατριβής μου δεν υπήρξε ένα μοναχικό ταξίδι. Είχα την τιμή και τη χαρά να συνυπάρξω με πολύτιμους συνταξιδιώτες στο χώρο του εργαστηρίου, συναδέλφους που πέρα από τους ερευνητικούς τους στόχους, τις γνώσεις και την ακαδημαϊκή τους κατάρτιση υπήρξαν αρωγοί, συμπαραστάτες και φίλοι πολύτιμοι βοηθώντας στην επιστημονική συνεργασία και στην ανάπτυξη ομαδικότητας όσο και στην προώθηση άριστων ανθρώπινων σχέσεων.

Αρχικά, οφείλω πολλά στον επιβλέποντά μου καθηγητή και υπεύθυνο του εργαστηρίου, κ. Πέτρο Μαραγκό και πρώτα από όλα του οφείλω την ευκαιρία που μου έδωσε να αποτελέσω μέλος της ερευνητικής του ομάδας. Ήταν πάντα παρών σε όλη την προσπάθειά μου επιτρέποντάς μου παράλληλα να μπορώ να αναπτύξω τις δικές μου πρωτοβουλίες και ιδέες και να κινούμαι με ελευθερία στους δικούς μου ερευνητικούς δρόμους. Η πείρα του, οι παραινέσεις και η καθοδήγησή του υπήρξαν πολύτιμες. Ο χαρακτήρας, η συνεχής παρουσία και εργασία του στο πανεπιστήμιο και η αφοσίωσή του στην έρευνα και τη διδασκαλία συνιστούν πρότυπο και στόχο διαχρονικό που κάθε νέος ερευνητής φιλοδοξεί να επιτύχει.

Όλη η προσπάθεια στη διάρκεια εκπόνησης του διδακτορικού και η συνολική εμπειρία που αποκόμισα θα είχε πολύ μεγαλύτερα εμπόδια και θα ήταν σαφώς φτωχότερη, αν δεν υπήρχαν συνοδοιπόροι μου τα υπόλοιπα παιδιά του εργαστηρίου, ένας σπάνιος συνδυασμός διαφορετικών αλλά εξαιρετικών χαρακτήρων των οποίων οι γνώσεις, η υποστήριξη και η φιλία υπήρξαν γενναιόδωρες και στάθηκαν αφορμή για την απόκτηση νέων συνεργατών και φίλων. Η σχέση με τα παιδιά έχει ενισχυθεί μέσα από ατελείωτες ώρες συνύπαρξης στο εργαστήριο, εποικοδομητικές συζητήσεις, συνεργατική δραστηριότητα, αλληλεγγύη και υποστήριξη. Η παρέα και η συντροφιά τους επεκτάθηκε και εκτός του χώρου του εργαστηρίου ενδυναμώνοντας τη συναδελφικότητα και δίνοντας προοπτικές διαχρονικής φιλίας όπως εύχομαι να εξελιχθεί στο μέλλον. Ό,τι και να πω για αυτούς είναι λίγο. Ο Γιώργος Παπανδρέου εξαιρετικά μεθοδικός, δημιουργικός και πάντα διακριτικός. Έμαθα πολλά από αυτόν μέσα από τη συνεργασίας μας σε θέματα στατιστικής μοντελοποίησης. Η συνεργασία αυτή αποτέλεσε και την αφορμή για μια στενή φιλία. Ο Νάσος πάντα χαμογελαστός και ευχάριστος έδινε πάντα έναν τόνο αισιοδοξίας στο εργαστήριο. Ο Γιώργος Ευαγγελόπουλος, προσιτός και εμψυχωτής, μας έδινε κίνητρα να προβληματιστούμε. Ο Βασίλης επίμονος και ακούραστος, πάντα παρών στο εργαστήριο και πρόθυμος να βοηθήσει. Ο Τάσος,

πάντα καλοπροαίρετος και φιλότιμος. Η Νάνσυ, ανήσυχη και εργατική. Ο Δημήτρης, η Νατάσσα, ο Ιάσονας τα παλιότερα μέλη και ο Σταύρος ο νεώτερος του εργαστηρίου. Η Όλγα, η Βίκυ και η Δέσποινα σταθερές παρουσίες του εργαστηρίου. Τους ευχαριστώ όλους από καρδιάς.

Οφείλω να ευχαριστήσω θερμά τα μέλη της πενταμελούς επιτροπής του διδακτορικού μου, καθηγητές Μιλτιάδη Αναγνώστου, Γεώργιο Παπαβασιλόπουλο, Στέφανο Κόλλια και Κώστα Τζαφέστα για τα κριτικά τους σχόλια κατά την παρουσίαση της ενδιάμεσης προόδου· τον καθηγητή Κωνσταντίνο Μπερμπερίδη και την Δρ. Γεωργία Τσιροπούλα που συμπλήρωσαν την επταμελή επιτροπή εξέτασης. Όλοι διάβασαν προσεκτικά το κείμενο της διατριβής και συνεισέφεραν με τις εποικοδομητικές τους παρατηρήσεις κατά τη διάρκεια της εξέτασης. Ιδιαίτερος φόρος τιμής πρέπει να αποδοθεί στον αδικοχαμένο Αιμίλιο Χαρλαύτη, ο οποίος εμπνεύστηκε το αντικείμενο του ΠΕΝΕΔ στο οποίο εντάχθηκε η διδακτορική μου διατριβή. Εκφράζω επίσης της ευγνωμοσύνη μου στο Ίδρυμα Μποδοσάκη για την οικονομική και ψυχολογική υποστήριξη που μου παρείχε με την υποτροφία που μου χορήγησε κατά τη διάρκεια της διατριβής μου.

Σε όλη αυτή τη διάρκεια είχα στο πλευρό μου και πολλούς ανθρώπους εκτός του εργαστηρίου. Οι φίλοι μου από τα σχολικά και προπτυχιακά χρόνια. Η ευρύτερη οικογένεια μου: Οι γονείς μου, Στέλιος και Βιολέττα στους οποίους χρωστώ τα πάντα και τους είμαι ευγνώμων. Η αδελφή μου Έφη, ο γαμπρός μου Μάκης και ο ανηψιός μου Κωνσταντίνος. Η σύντροφος μου Ελίζα που αποτέλεσε διαρκή συμπαραστάτρια και συνοδοιπόρο μου και με βοήθησε αφάνταστα. Ο παιδικός μου φίλος Παναγιώτης. Όλοι τους με στήριξαν και είναι διαρκώς δίπλα μου. Τους ευχαριστώ.

Σταμάτης Λευκιμμιάτης

ΠΕΡΙΛΗΨΗ

Η εργασία μας εντάσσεται στην περιοχή της ανάλυσης και επεξεργασίας εικόνων και ακουστικών σημάτων. Βασικό ερευνητικό ζητούμενο είναι η ανάπτυξη νέων τεχνικών ψηφιακής επεξεργασίας σημάτων οι οποίες οδηγούν σε στατιστικά βέλτιστα αποτελέσματα αποθορυβοποίησης και ανίχνευσης. Στην παρούσα εργασία κινούμαστε σε δύο κατευθύνσεις. Η πρώτη κατεύθυνση αφορά στην ανάπτυξη μεθόδων για την αποθορυβοποίηση εικόνων πάνω στις οποίες επιδρά σηματο-εξαρτώμενος θόρυβος. Κίνητρό μας αποτέλεσε το πρόβλημα της ψηφιακής επεξεργασίας δορυφορικών εικόνων και η ενίσχυση της ποιότητάς τους η οποία υποβαθμίζεται από την επίδραση κβαντικού θορύβου. Η δεύτερη ερευνητική μας κατεύθυνση σχετίζεται με τη μελέτη και ανάπτυξη μεθόδων αποθορυβοποίησης και ανίχνευσης χαρακτηριστικών σε μονοδιάστατα σήματα τα οποία λαμβάνονται μέσω χωρικά διατεταγμένων ψηφιακών αισθητήρων.

Στο πρώτο μέρος της διατριβής μελετούμε το πρόβλημα της ψηφιακής αποκατάστασης εικόνων περιορισμένης φωτεινότητας. Στα σύγχρονα απεικονιστικά συστήματα ο σχηματισμός των εικόνων επιτυγχάνεται μέσω της ανίχνευσης του αριθμού των φωτονίων που προσπίπτουν στην επιφάνεια ενός αισθητήρα. Όταν μια εικόνα λαμβάνεται σε συνθήκες χαμηλού φωτισμού, λόγω της στοχαστικής φύσης της διαδικασίας ανίχνευσης, ο αριθμός των ανιχνευομένων φωτονίων σε κάθε σημείο του αισθητήρα δεν αντιστοιχεί στην πραγματική ένταση φωτεινότητας της φωτεινής πηγής, αλλά ακολουθεί μια Poisson κατανομή που έχει ως παράμετρο την πραγματική ένταση φωτεινότητας της πηγής. Κατά συνέπεια, η εικόνα που τελικά κανείς λαμβάνει είναι αλλοιωμένη από θόρυβο ο οποίος είναι άμεσα εξαρτημένος από το πηγαίο σήμα.

Η προσέγγιση που ακολουθούμε στη διατριβή για την αντιμετώπιση του συγκεκριμένου προβλήματος συνίσταται στην περιγραφή των εικόνων σε πολλαπλές κλίμακες ανάλυσης με τη χρήση μη-γραμμικών χωρικά εντοπισμένων μετασχηματισμών. Η περιγραφή των εικόνων στο νέο χώρο αναπαράστασης επιτρέπει την ανάπτυξη εκφραστικών και αποδοτικών μοντέλων για τη στατιστική περιγραφή τους και οδηγεί σε αποδοτική εκτίμηση της πραγματικής έντασης φωτεινότητάς τους. Σε αυτό το πλαίσιο συζητούμε τις στατιστικές ιδιότητες των φυσικών εικόνων στο νέο χώρο αναπαράστασης και μελετούμε διεξοδικά την οικογένεια των πολυκλιμακωτών δενδρικών μαρκοβιανών στατιστικών μοντέλων που μπορεί να τις περιγράψει σε ικανοποιητικό βαθμό. Μία από τις κύριες συνεισφορές μας σε αυτή την ερευνητική περιοχή είναι η ανάπτυξη ενός πρωτότυπου μη-γραμμικού τοπικά εντοπισμένου μετασχηματισμού που εμφανίζει τη δυνατότητα αποσύνδεσης του κβαντικού θορύβου από την πραγματική πληροφορία της εικόνας, ενώ, παράλληλα, διαθέτει μια στενή σχέση με το μετασχηματισμό κυματιδίων. Η σχέση αυτή αποδεικνύεται εξαιρετικά σημαντική, αφού αφενός υποδηλώνει ότι βασικές ιδιότητες του μετασχηματισμού κυματιδίων κληρονομούνται από τον προτεινόμενο μετασχηματισμό, αφετέρου επιτρέπει την αποδοτική υλοποίησή του. Επιπλέον, προτείνουμε νέους αλγορίθμους εκπαίδευσης των παραμέτρων των στατιστικών μοντέλων απευθείας από τα θορυβώδη δεδομένα. Η εύρεση των παραμέτρων αυτών είναι εξαιρετικής σημασίας, αφού οδηγεί σε καλύτερη στατιστική μοντελοποίηση των εικόνων και σε βελτιωμένα αποτελέσματα αποθορυβοποίησης.

Στο δεύτερο μέρος της διατριβής εξετάζουμε τεχνικές ψηφιακής επεξεργασίας που βασίζονται σε πολυαισθητηριακά συστήματα λήψης ακουστικών σημάτων. Τα σήματα που λαμβάνονται

ξεχωριστά από κάθε αισθητήρα της συστοιχίας μπορούν να αντιμετωπιστούν ως διαφορετικά στιγμιότυπα του ίδιου σήματος. Σκοπός της πολυαισθητηριακής επεξεργασίας είναι ο συνδυασμός των στιγμιοτύπων αυτών και η εκμετάλλευση των μεταξύ τους συσχετίσεων με τέτοιο τρόπο, ώστε το τελικό εξαγόμενο σήμα να είναι όσο το δυνατόν περισσότερο απαλλαγμένο από το θόρυβο και να παρουσιάζει καλύτερη ευκρίνεια και ποιότητα σε σχέση με τα αρχικά λαμβανόμενα δεδομένα.

Σε αυτή την ερευνητική περιοχή για το πρόβλημα της απαλοιφής των επιδράσεων του θορύβου μελετούμε στατιστικά βέλτιστες πολυαισθητηριακές μεθόδους αποθορυβοποίησης σημάτων. Οι μέθοδοι αυτές αντιστοιχούν στο φιλτράρισμα των λαμβανόμενων σημάτων από πολυκαναλικά φίλτρα και θεωρητικά οδηγούν σε βέλτιστα αποτελέσματα. Ωστόσο, η εφαρμογή τους δεν είναι δυνατή, αν προηγουμένως δεν εκτιμηθούν οι εκφράσεις των φίλτρων, οι οποίες εξαρτώνται από το μη-παρατηρήσιμο σήμα πηγής που ιδανικά θέλουμε να εκτιμήσουμε. Προκειμένου να είναι εφικτή η χρήση των μεθόδων αυτών, προτείνουμε μια νέα γενική τεχνική εκτίμησης των παραμέτρων μιας ευρείας κατηγορίας πολυκαναλικών φίλτρων των οποίων η έκφραση καθορίζεται από το επιλεγόμενο κριτήριο βελτιστοποίησης και το είδος του θορύβου.

Τέλος, ένα ακόμα πρόβλημα που εξετάζουμε είναι αυτό της αυτόματης ανίχνευσης και αναγνώρισης ακουστικών σημάτων σε συνθήκες θορύβου. Οι κύριες ερευνητικές μας συνεισφορές στην αντιμετώπιση αυτού του προβλήματος εντοπίζονται στη διαμόρφωση ενός ενοποιητικού πλαισίου δύο βασικών ενεργειακών τελεστών, οι οποίοι χρησιμοποιούνται στη λήψη ενεργειακών μετρήσεων, και στην ανάπτυξη ενός πρωτότυπου πολυαισθητηριακού σχήματος εξαγωγής εύρωστων στο θόρυβο διανυσμάτων χαρακτηριστικών τα οποία χρησιμοποιούνται στο στατιστικό σύστημα αναγνώρισης.

ABSTRACT

Our work lies in the area of statistical signal and image processing. Our main interest is the development of optimal signal processing methods for denoising and detection. In the current work we follow two research directions. The first one is focused on the problem of denoising images which suffer from signal-dependent noise and is motivated by the problem of enhancing the quality of satellite images degraded by quantum noise. The second direction is related to the study and development of multi-sensor signal denoising and detection methods.

In the first part of the thesis we study the problem of photon-limited image denoising. In many modern imaging systems image acquisition is accomplished by detecting photon counts at different spatial locations of a sensor. For low intensity levels, one of the dominant noise sources responsible for the degradation of the quality of the captured images is the so-called quantum or shot noise. This kind of noise is due to fluctuations on the number of detected photons, an inherent limitation of the stochastic nature of the detection process. Quantum noise obeys a Poisson law and, hence, is highly dependent on the underlying intensity pattern being imaged.

The approach we follow in the thesis to deal with the problem of estimating the underlying image intensity is based on a representation of the image over multiple scales, using a family of multiscale nonlinear transforms. The representation of the image in the new domain permits us to employ graphical models in order to capture its statistical properties and can lead to efficient intensity estimation schemes. In this framework, we discuss important statistical properties of natural images in the domain of the transform, captured to a good extent by the family of multiscale Markov tree models which we study in depth. One of our main contributions in this area is the development of a novel multiscale nonlinear transform which can decouple the Poisson noise and the image true content. In addition, the proposed transform is proved that is closely related to the wavelet transform. This relation is of significant importance since it implies that key wavelet properties are also inherited by the proposed transform. Moreover, we propose new algorithms for learning the statistical model parameters directly from noisy data. The accurate estimation of the model parameters is crucial since it can lead to improved denoising results.

In the second part of the thesis we study multi-sensor signal processing methods. Signals that are captured separately by a sensor of the array can be considered as different realizations of the same source signal. The goal in multi-sensor signal processing is to combine these realizations and exploit their correlation in a way that will lead to an output signal of superior quality compared to the input signals. In this research area we study statistical optimal multi-sensor denoising methods to deal with the problem of signal enhancement. The studied methods correspond to multichannel filtering of the captured signals. However, despite their theoretically optimal results, these filters are difficult to realize. This is due to the requirement for knowledge of second order statistics for both the signal and the corrupting noise that makes these filters signal-dependent. In order to address this issue, we propose a novel generalized estimation method for estimating the necessary parameters. This method is general,

appropriate for different noise conditions and various optimization criteria.

Finally, we also study the problem of automatic signal detection and recognition under noisy conditions. Our main research contributions in this area is the development of a theoretical unified framework of two basic energy operators, which are used for obtaining energy measurements and the development of a novel multi-sensor feature extraction scheme which can serve as the front-end of a statistical recognition system.

Κεφάλαιο 1

Εισαγωγή

Γενική Περιγραφή και Σημασία του Προβλήματος

Επιστήμες, όπως είναι η Αστρονομία και η Αστροφυσική, οι οποίες ερευνούν και εξετάζουν τα ουράνια σώματα καθώς και τις σχέσεις, τις κινήσεις και τη δυναμική αυτών, βασίζονται σε πολύ μεγάλο βαθμό στις παρατηρήσεις που λαμβάνονται από τον κόσμο που μας περιβάλλει. Οι επιστήμες αυτές ανέκαθεν αποτελούσαν ένα συναρπαστικό πεδίο μελέτης, δεδομένου ότι ο άνθρωπος ως φιλομαθές ον αισθάνεται έντονη την επιθυμία της έρευνας στην προσπάθεια του να κατανοήσει τις αρχές και τους κανόνες που διέπουν τη συμπεριφορά των ουράνιων σωμάτων. Πέρα, όμως, από την ανθρώπινη "περιέργεια" μια τέτοια μελέτη υπαγορεύεται και από πρακτικούς λόγους. Η αμερικάνικη Εθνική Υπηρεσία Αεροναυτικής και Διαστήματος (NASA) έχει ξεκινήσει από τις αρχές της δεκαετία του '60 μια προσπάθεια εξερεύνησης του διαστήματος με απώτερο στόχο την συλλογή πληροφοριών οι οποίες αποδεικνύονται εποικοδομητικές στην εξέλιξη άλλων επιστημών, όπως είναι η ιατρική, η βιολογία, η γεωλογία, κ.τ.λ . Ιδιαιτέρως τα τελευταία χρόνια, οι επιστήμες της Αστρονομίας και της Αστροφυσικής γνωρίζουν μεγάλη άνθηση και προσελκύουν το έντονο ενδιαφέρον της ερευνητικής κοινότητας αλλά και του ευρύτερου κοινού. Χαρακτηριστικό παράδειγμα είναι ότι το έτος 2009 ανακηρύχθηκε παγκόσμιο έτος Αστρονομίας. Η άνθηση αυτή αποδίδεται σε μεγάλο βαθμό στη ραγδαία τεχνολογική εξέλιξη που παρατηρείται τόσο σε επίπεδο συλλογής δεδομένων όσο και σε επίπεδο ανάλυσης πληροφοριών.

Από το 17ο αιώνα, οπότε και ανακαλύφθηκε το πρώτο τηλεσκόπιο συμβάλλοντας στην εξέλιξη της Αστρονομίας σε σύγχρονη επιστήμη, έως σήμερα που οι αστρονόμοι έχουν στην αποκλειστική τους διάθεση δορυφόρους για την παρατήρηση αστρονομικών φαινομένων, έχει σημειωθεί τεράστια πρόοδος στον τομέα κατασκευής τεχνολογικών συσκευών για τη συλλογή δεδομένων. Οι σύγχρονοι ψηφιακοί αισθητήρες οι οποίοι συλλέγουν δεδομένα με τη μορφή ψηφιακών σημάτων παρουσιάζουν πληθώρα πλεονεκτημάτων και δυνατοτήτων συγκριτικά με τους αισθητήρες που υπήρχαν στη διάθεσή μας τις προηγούμενες δεκαετίες. Παράλληλα, το κόστος τους έχει μειωθεί σημαντικά επιτρέποντας έτσι σε ένα ευρύτερο κοινό τη χρήση τους. Για παράδειγμα, οι σύγχρονες συσκευές λήψης επίγειων ή δορυφορικών αστρονομικών εικόνων βασίζονται στην τεχνολογία αισθητήρων συζευγμένου φορτίου (Charged Coupled Device - CCD) και χρησιμοποιούνται κατά κόρον τόσο από επαγγελματίες όσο και από ερασιτέχνες αστρονόμους λόγω του μειωμένου τους κόστους. Μάλιστα οι τελευταίοι χρησιμοποιώντας τηλεσκόπια που είναι εφοδιασμένα με τη συγκεκριμένη τεχνολογία έχουν συμβάλει σημαντικά σε επιστημονικές ανακαλύψεις ειδικά παροδικών φαινομένων. Παράλληλα με την ανάπτυξη ψηφιακών αισθητήρων για τη λήψη οπτικών δεδομένων (εικόνες), τα τελευταία 30 χρόνια έχει ξεκινήσει η μελέτη και ανάπτυξη χωρικά διατεταγμένων συστοιχιών από αισθητήρες για τη σύλληψη μονοδιάστατων σημάτων (χρονοσειρές). Τέτοιες διατάξεις χρησιμοποιούνται και στη ραδιοαστρονομία, υποκλάδο της αστρονομίας που ασχολείται με τη μελέτη ουράνιων σωμάτων σε ραδιοσυχνότητες. Χάρη στην ερευνητική και πειραματική ανάπτυξη

που συντελείται μέχρι σήμερα ο κλάδος που ασχολείται με την κατασκευή διατάξεων αισθητήρων έχει φτάσει σε ένα τέτοιο στάδιο ωρίμασης, ώστε οι τεχνολογίες που βασίζονται σε συστοιχίες αισθητήρων να είναι άμεσα εφαρμόσιμες σε ένα μεγάλο αριθμό σύγχρονων συστημάτων, ενώ επιπλέον υπάρχουν τεράστια περιθώρια και δυνατότητες για την ενσωμάτωσή τους σε μελλοντικές συσκευές.

Παρόλη τη συνεχή τεχνολογική εξέλιξη και τη διαρκή βελτίωση των σύγχρονων συσκευών συλλογής δεδομένων, σε πολλές περιπτώσεις τα λαμβανόμενα δεδομένα είναι πολύ δύσκολο να χρησιμοποιηθούν με την υπάρχουσα μορφή τους για ανάλυση και εξαγωγή συμπερασμάτων. Ένας βασικός λόγος που καθιστά τα δεδομένα ακατάλληλα για περαιτέρω χρήση είναι η έντονη υποβάθμιση την οποία συχνά παρουσιάζουν λόγω του θορύβου αλλά και άλλων παραμορφώσεων που έχουν υποστεί από το περιβάλλον έως ότου συλλεγούν από τους αισθητήρες. Επίσης, ακόμα και οι ίδιοι οι αισθητήρες εμφανίζουν εγγενείς αδυναμίες και εισάγουν και αυτοί με τη σειρά τους επιπλέον παραμορφώσεις. Ένα τέτοιο παράδειγμα αποτελεί η περίπτωση των αστρονομικών εικόνων η λήψη των οποίων (δορυφορική ή επίγεια) πραγματοποιείται από συσκευές που βασίζονται στην αναγνώριση του αριθμού των φωτονίων, τα οποία εκπέμπει η φωτεινή πηγή, και συνήθως πραγματοποιείται σε χαμηλές συνθήκες φωτισμού. Υπό αυτές τις συνθήκες και εξαιτίας των εγγενών περιορισμών που υπάρχουν, λόγω της διακριτής φύσης της διαδικασίας ανίχνευσης των εκπεμπόμενων φωτονίων, η ποιότητα του καταγεγραμμένου οπτικού σήματος δεν είναι πάντοτε ικανοποιητική. Στις δορυφορικές λήψεις η κύρια πηγή αλλοίωσης, η οποία καλείται κβαντικός θόρυβος, αποδίδεται κυρίως στη στοχαστικότητα που εμφανίζει η διαδικασία εκπομπής και ανίχνευσης φωτονίων, ενώ στις επίγειες λήψεις ο θόρυβος που επηρεάζει την ποιότητα των εικόνων οφείλεται και σε άλλους παράγοντες, όπως είναι οι παραμορφώσεις που υπεισέρχονται κατά τη διάδοση του οπτικού σήματος μέσα από την ατμόσφαιρα. Το φαινόμενο του κβαντικού θορύβου παρατηρείται και στις ψηφιακές κάμερες του εμπορίου, για παράδειγμα στην περίπτωση νυκτερινής λήψης φωτογραφιών, οπότε στη φωτογραφία παρατηρείται η ύπαρξη έντονων κουκκίδων τυχαία διεσπαρμένων.

Όσον αφορά στις συστοιχίες αισθητήρων, επειδή η πηγή του σήματος ενδιαφέροντος βρίσκεται συχνά σε μεγάλη απόσταση, το σήμα που τελικά φθάνει και συλλαμβάνεται είναι πολύ χαμηλής ισχύος και "καλυμμένο" από θόρυβο. Επίσης, τα πολλαπλά σήματα που φθάνουν σε μια χωρική συστοιχία ψηφιακών αισθητήρων συχνά χαρακτηρίζονται από διαφορετικά επίπεδα θορύβου, απόσβεσης, καθυστέρησης και άλλων χωρο-χρονικών παραμορφώσεων λόγω των διαφορετικών μονοπατιών που ακολουθούν μέχρι τελικά να καταγραφούν από τους αισθητήρες. Σε όλες τις περιπτώσεις που αναφέρθηκαν η παρουσία του θορύβου και άλλων παραμορφώσεων αποτελεί έναν ισχυρό παράγοντα που αποτρέπει την εξαγωγή χρήσιμης και βάσιμης πληροφορίας από τα λαμβανόμενα δεδομένα. Ένας τρόπος να αρθούν οι συγκεκριμένοι περιορισμοί και να καταστούν τα δεδομένα αξιόπιστα για χρήση είναι η ανάπτυξη τεχνικών ψηφιακής επεξεργασίας σημάτων με σκοπό την εξάλειψη των παραμορφώσεων και την ελάττωση του θορύβου που υπεισέρχεται στις μετρήσεις.

Αντικείμενο και Συνεισφορές Διατριβής

Βασικό ερευνητικό ζητούμενο της παρούσας διατριβής είναι η ανάπτυξη νέων τεχνικών ψηφιακής επεξεργασίας σημάτων οι οποίες οδηγούν σε στατιστικά βέλτιστα αποτελέσματα αποθορυβοποίησης και ανίχνευσης. Στην παρούσα εργασία κινούμαστε σε δύο κατευθύνσεις. Η πρώτη κατεύθυνση αφορά στην ανάπτυξη μεθόδων για την αποθορυβοποίηση οπτικών σημάτων που λαμβάνονται με χρήση αισθητήρων συζευγμένου φορτίου. Κίνητρό μας αποτέλεσε η επεξεργασία δορυφορικών εικόνων των οποίων η καταγραφή πραγματοποιείται με ψηφιακές συσκευές που βασίζονται στη συγκεκριμένη τεχνολογία. Ωστόσο, πέρα από τις δορυφορικές εικόνες τέτοιου είδους καταγραφές πραγματοποιούνται και σε εικόνες που απαντώνται και σε άλλους επιστημονικούς κλάδους, όπως για παράδειγμα η βιοϊατρική. Συνεπώς, οι μέθοδοι που μελετήθηκαν και αναπτύ-

χθηκαν με στόχο την επεξεργασία δορυφορικών εικόνων μπορούν να εφαρμοσθούν με επιτυχία και σε αυτά τα είδη δεδομένων και πιθανώς και σε άλλα. Η δεύτερη ερευνητική μας κατεύθυνση σχετίζεται με τη μελέτη και ανάπτυξη μεθόδων αποθορυβοποίησης και ανίχνευσης χαρακτηριστικών σε μονοδιάστατα σήματα τα οποία λαμβάνονται μέσω χωρικά διατεταγμένων ψηφιακών αισθητήρων. Η σύλληψη και ανάλυση χρονοσειρών με πολυαισθητηριακά συστήματα πραγματοποιείται, όπως αναφέραμε και προηγουμένως, στη ραδιοαστρονομία αλλά και σε κλάδους που ασχολούνται με την ανάλυση σημάτων φωνής.

Στο πρώτο μέρος της διατριβής μελετούμε το πρόβλημα της ψηφιακής αποκατάστασης εικόνων περιορισμένης φωτεινότητας. Συγκεκριμένα, ο σχηματισμός των εικόνων σε απεικονιστικά συστήματα που βασίζονται στην CCD τεχνολογία επιτυγχάνεται μέσω της ανίχνευσης του αριθμού των φωτονίων που προσπίπτουν στην επιφάνεια του αισθητήρα. Όταν μια εικόνα λαμβάνεται σε χαμηλές συνθήκες φωτισμού, φαινόμενο αρκετά συχνό στη λήψη αστρονομικών ή βιοϊατρικών εικόνων, ο αριθμός των φωτονίων που ο αισθητήρας ανιχνεύει τελικά σε κάθε σημείο του δεν αντιστοιχεί στην πραγματική ένταση φωτεινότητας της φωτεινής πηγής αλλά ακολουθεί μια Poisson κατανομή που έχει ως παράμετρο την πραγματική ένταση φωτεινότητας της πηγής. Κατά συνέπεια, η εικόνα που τελικά λαμβάνουμε είναι αλλοιωμένη με θόρυβο που είναι άμεσα εξαρτημένος από το πηγαίο σήμα. Το φαινόμενο αυτό απαιτεί μια ειδική αντιμετώπιση που είναι διαφορετική σε σχέση με αυτήν που κανείς θα ακολουθούσε για την αποκατάσταση εικόνων οι οποίες έχουν αλλοιωθεί από ομογενή προσθετικό θόρυβο, όπως είναι για παράδειγμα ο γκαουσιανός θόρυβος.

Στο δεύτερο μέρος της διατριβής μελετούμε πολυαισθητηριακά συστήματα λήψης ακουστικών σημάτων. Μέσω της ψηφιακής επεξεργασίας σημάτων με συστοιχίες αισθητήρων είναι δυνατόν να αντιμετωπίσουμε τα σήματα που λαμβάνονται ξεχωριστά από κάθε αισθητήρα ως στιγμιότυπα του ίδιου σήματος, όπου το κάθε ένα έχει υποστεί μια διαφορετική μετατόπιση (καθυστέρηση), λόγω των διαφορετικών μονοπατιών διάδοσης, και έχει αλλοιωθεί από εξωτερικό θόρυβο διαφορετικής ισχύος. Σκοπός της πολυαισθητηριακής επεξεργασίας είναι ο συνδυασμός των στιγμιοτύπων αυτών και η εκμετάλλευση των μεταξύ τους συσχετίσεων με τέτοιο τρόπο, ώστε το τελικό εξαγόμενο σήμα να είναι όσο το δυνατόν περισσότερο απαλλαγμένο από το θόρυβο και να παρουσιάζει καλύτερη ευκρίνεια και ποιότητα σε σχέση με τα αρχικά λαμβανόμενα δεδομένα. Το βασικό πλεονέκτημα που διαθέτουν τα συστήματα επεξεργασίας που βασίζονται σε συστοιχίες αισθητήρων έναντι αυτών που χρησιμοποιούν ένα μοναδικό αισθητήρα είναι ότι παρουσιάζουν τόσο χωρική όσο και συχνοτική διαχωρισιμότητα. Συνεπώς είναι δυνατόν εκτός από τη φασματική πληροφορία του σήματος να αξιοποιήσει κανείς και την χωρική διάσταση του προβλήματος για να επιτύχει περαιτέρω βελτίωση της ποιότητας των υπό επεξεργασία σημάτων.

Οι κύριες ερευνητικές μας συνεισφορές στο πρώτο μέρος της παρούσας εργασίας που αφορά στην ψηφιακή αποκατάσταση εικόνων μπορούν να συνοψισθούν στα εξής:

- Μελέτη και ανάπτυξη πολυκλιμακωτών στρατηγικών ανάλυσης που προορίζονται αποκλειστικά για την ιεραρχική αναπαράσταση εικόνων περιορισμένης φωτεινότητας. Η βασική ιδέα είναι ότι αντί κανείς να προσπαθεί να αποσυνδέσει το θόρυβο από την πραγματική πληροφορία στο αρχικό πεδίο ορισμού της εικόνας, είναι συχνά πιο αποτελεσματικό να οδηγείται πρώτα σε ένα νέο πεδίο μέσω ενός κατάλληλου μετασχηματισμού των δεδομένων. Η αναπαράσταση των φυσικών εικόνων σε ένα νέο χώρο μέσω των προτεινόμενων σχημάτων ανάλυσης επιτρέπει την ανάπτυξη εκφραστικών και αποδοτικών μοντέλων για τη στατιστική περιγραφή τους τα οποία δείχνουμε πως μπορούν να χρησιμοποιηθούν στο πρόβλημα της ψηφιακής αποκατάστασης των εικόνων.

- Βελτίωση των υπαρχόντων στατιστικών μοντέλων με την ενσωμάτωση αναδρομικών ως προς την κλίμακα ανάλυσης μαρκοβιανών πιθανοτικών δομών. Οι δομές αυτές επιτρέπουν την καλύτερη στατιστική περιγραφή των δεδομένων μοντελοποιώντας με έναν ακριβέστερο τρόπο τις στατιστικές εξαρτήσεις που επικρατούν.

- Ανάπτυξη τεχνικών για την εκπαίδευση μιας ευρείας κατηγορίας παραμετρικών δενδρικών μαρκοβιανών μοντέλων, τα οποία χρησιμοποιούνται σε εικόνες περιορισμένης φωτεινότητας, απευθείας από τις θορυβώδεις παρατηρήσεις. Η αξιόπιστη και εύρωστη εκτίμηση αυτών των παραμέτρων είναι καθοριστικής σημασίας για την αποδοτικότητα των μεθόδων οι οποίες υιοθετούν αυτά τα μοντέλα και εφαρμόζονται στο πρόβλημα της αποθορυβοποίησης εικόνων.

Στο δεύτερο μέρος που αφορά στην επεξεργασία ακουστικών σημάτων τα οποία λαμβάνονται από μια χωρικά διατεταγμένη συστοιχία αισθητήρων οι συνεισφορές μας μπορούν να περιγραφούν συνοπτικά ως εξής:

- Μελέτη στατιστικά βέλτιστων πολυαισθητηριακών μεθόδων αποθορυβοποίησης σημάτων. Οι μέθοδοι αυτές αντιστοιχούν στο φιλτράρισμα των σημάτων εισόδου και θεωρητικά οδηγούν σε βέλτιστα αποτελέσματα. Ωστόσο, η εφαρμογή τους δεν είναι δυνατή, αν προηγουμένως δεν εκτιμηθούν οι εκφράσεις των φίλτρων, οι οποίες εξαρτώνται από το μη-παρατηρήσιμο σήμα πηγής. Προκειμένου να είναι εφικτή η χρήση των μεθόδων αυτών αναπτύξαμε μια γενική τεχνική εκτίμησης των παραμέτρων μιας ευρείας κατηγορίας φίλτρων των οποίων η έκφραση καθορίζεται από το επιλεγόμενο κριτήριο βελτιστοποίησης και το είδος του θορύβου.

- Με αφετηρία το πρόβλημα της αυτόματης κατηγοριοποίησης και αναγνώρισης σημάτων υπό συνθήκες θορύβου μελετήσαμε ένα πολυαισθητηριακό σχήμα αποδιαμόρφωσης ακουστικών σημάτων σε φασματικές συνιστώσες. Με βάση αυτό το σχήμα και τη διαμόρφωση ενός ενοποιητικού πλαισίου για δύο βασικούς ενεργειακούς τελεστές, οι οποίοι χρησιμοποιούνται στη λήψη ενεργειακών μετρήσεων, αναπτύξαμε ένα πολυαισθητηριακό σχήμα εξαγωγής εύρωστων στο θόρυβο διανυσμάτων χαρακτηριστικών.

Οργάνωση της Διατριβής

Το υπόλοιπο της διατριβής είναι οργανωμένο ως εξής:

Στο **Κεφάλαιο 2** μελετούμε την αναπαράσταση εικόνων με τη χρήση τοπικά εντοπισμένων μετασχηματισμών. Αρχικά, παρουσιάζουμε αναπαραστάσεις εικόνων με κυματίδια χώρου-κλίμακας των οποίων οι κλιμακούμενες και μετατοπισμένες εκδοχές συνιστούν πολυκλιμακωτές ορθοκανονικές ή διορθογώνιες βάσεις στο χώρο $\mathbf{L}^2(\mathbb{R}^2)$. Στη συνέχεια παρουσιάζουμε μια συγγενή κατηγορία μη-γραμμικών πολλαπλασιαστικών μετασχηματισμών που επίσης επιτρέπουν την αναπαράσταση των εικόνων σε πολλαπλές κλίμακες ανάλυσης. Οι τελευταίοι (μετασχηματισμοί) παρουσιάζονται ως ιδανικότεροι υποψήφιοι για την αναπαράσταση εικόνων που εμφανίζουν αλλοιώσεις από σηματο-εξαρτώμενο θόρυβο, αφού διαθέτουν τη δυνατότητα αποσύνδεσης αυτού του είδους θορύβου από την πραγματική πληροφορία των εικόνων. Επιπλέον, δείχνουμε ότι οι δυο κατηγορίες μετασχηματισμών συνδέονται στενά με αποτέλεσμα πολλές από τις βασικές ιδιότητες του μετασχηματισμού κυματιδίων να κληρονομούνται από την οικογένεια των πολλαπλασιαστικών μετασχηματισμών. Τέλος, προτείνουμε ένα πρωτότυπο 2-Δ πολλαπλασιαστικό μετασχηματισμό που εμφανίζει επιπλέον πλεονεκτήματα σε σχέση με προηγούμενες 2-Δ εκδοχές, αφού καταφέρνει να μοντελοποιεί με μεγαλύτερη ακρίβεια πρωτεύουσες δομές των εικόνων. Σχετική δημοσίευση μας είναι η [LPM09].

Στο **Κεφάλαιο 3** μελετούμε το πρόβλημα εκτίμησης της γενεσιουργού έντασης Poisson στοχαστικών διαδικασιών το οποίο αντιστοιχεί στην αποθορυβοποίηση εικόνων με Poisson θόρυβο. Επικεντρωνόμαστε σε μια στατιστική αντιμετώπιση του προβλήματος υιοθετώντας και

βελτιώνοντας υπάρχοντα στατιστικά μοντέλα για την περιγραφή των συντελεστών που προκύπτουν κατά την πολυκλιμακωτή ανάλυση των εικόνων μέσω του πολλαπλασιαστικού μετασχηματισμού. Συζητούμε πως μπορούν να χρησιμοποιηθούν τα στατιστικά μοντέλα που περιγράφουμε στο πρόβλημα της εκτίμησης και παρουσιάζουμε νέες τεχνικές εκπαίδευσής τους μέγιστης πιθανοφάνειας για τις οποίες αρκεί η γνώση μόνο της θορυβώδους εκδοχής της εικόνας. Αξιολογούμε τις επιδόσεις των προτεινόμενων μεθόδων σε σχέση με εναλλακτικές μεθόδους και παραθέτουμε ενδεικτικά παραδείγματα εφαρμογής τους σε πραγματικά δεδομένα. Σχετικές δημοσιεύσεις μας είναι οι [LPM08, LMP09]

Στο **Κεφάλαιο 4** περιγράφουμε το πρόβλημα της αποθορυβοποίησης ακουστικών σημάτων με τη χρήση συστοιχίας ψηφιακών αισθητήρων. Επικεντρωνόμαστε στη υιοθέτηση στατιστικών κριτηρίων τα οποία οδηγούν σε μεθόδους αποθορυβοποίησης που αντιστοιχούν στο φιλτράρισμα των σημάτων εισόδου το οποίο μπορεί να αναλυθεί σε δύο διαδοχικά στάδια. Περιγράφουμε μια γενική μέθοδο που αναπτύξαμε για την εκτίμηση των παραμέτρων των φίλτρων, η οποία μπορεί να χρησιμοποιηθεί σε διαφορετικές συνθήκες θορύβου και για διαφορετικά κριτήρια βελτιστοποίησης που οδηγούν σε γραμμικά και μη-γραμμικά βέλτιστα φίλτρα. Σχετικές δημοσιεύσεις μας είναι οι [LM06, LDM06, LM07].

Στο **Κεφάλαιο 5** μελετούμε το πρόβλημα της πολυαισθητηριακής επεξεργασίας με σκοπό την αυτόματη αναγνώριση ακουστικών σημάτων υπό συνθήκες θορύβου. Προς αυτή την κατεύθυνση περιγράφουμε κάποιες από τις τεχνικές ακουστικής επεξεργασίας που προηγούνται της εξαγωγής διανυσμάτων χαρακτηριστικών αλλά και τους βασικούς ενεργειακούς τελεστές που χρησιμοποιούνται για τη λήψη ενεργειακών μετρήσεων. Παρέχουμε ένα ενοποιητικό πλαίσιο των ενεργειακών μετρήσεων που μας επιτρέπει την επιλογή της καταλληλότερης μέτρησης ανάλογα με τα χαρακτηριστικά του θόρύβου. Στη συνέχεια περιγράφουμε ένα πρωτότυπο πολυαισθητηριακό σχήμα αποδιαμόρφωσης των ακουστικών σημάτων το οποίο αποτελεί τη βάση για το τελικό σύστημα εξαγωγής χαρακτηριστικών εύρωστων στο θόρυβο. Τα εξαγόμενα χαρακτηριστικά χρησιμοποιούνται τελικά στο στατιστικό σύστημα αυτόματης αναγνώρισης. Σχετικές δημοσιεύσεις μας είναι οι [DML07, LMK08].

Στο **Κεφάλαιο 6** συνοψίζουμε τις συνεισφορές της διατριβής και σκιαγραφούμε κατευθύνσεις στις οποίες θα μπορούσε να επικεντρωθεί στο μέλλον η ερευνητική μας προσπάθεια. Κλείνοντας στο **Παράρτημα Α΄** παραθέτουμε ένα πλήρη κατάλογο με τις δημοσιεύσεις που προέκυψαν κατά την εκπόνηση της διδακτορικής διατριβής.

Μέρος Ι

Επεξεργασία Εικόνων σε Πολλαπλές Κλίμακες Ανάλυσης

Κεφάλαιο 2

Αναπαράσταση Φυσικών Εικόνων με Τοπικά Εντοπισμένους Μετασχηματισμούς για Ανίχνευση Χαρακτηριστικών

Τα σήματα με τα οποία ερχόμαστε καθημερινά σε επαφή, όπως είναι ο ήχος (1-Δ), οι εικόνες (2-Δ), το βίντεο (3-Δ) απαρτίζονται από τεράστιες ποσότητες δεδομένων. Μέσα σε αυτόν το φοβερά μεγάλο όγκο δεδομένων ο εντοπισμός της πιο σχετικής μόνο πληροφορίας αποδεικνύεται εξαιρετικά δύσκολος. Η επεξεργασία των σημάτων καθίσταται γρηγορότερη και απλούστερη, όταν η περιγραφή τους γίνει με τέτοιο τρόπο, ώστε η πληροφορία που μας ενδιαφέρει να αναπαρίσταται μόνο σε λίγους συντελεστές. Τέτοιες αναπαραστάσεις πληροφορίας καλούνται *αραιές* και μπορούν να επιτευχθούν μέσω της ανάλυσης των σημάτων σε στοιχειώδη δομικά στοιχεία τα οποία επιλέγονται από ένα *λεξικό*. Θα πρέπει ωστόσο να σημειωθεί ότι δεν υπάρχει μια καθολικά ιδανική αναπαράσταση κατάλληλη για όλα τα είδη σημάτων. Ανάλογα με το είδος των δεδομένων που επιθυμεί κανείς να επεξεργαστεί, αλλά και το πρόβλημα που καλείται να αντιμετωπίσει θα πρέπει να επιλέγει τη χρήση κατάλληλων μετασχηματισμών που θα μπορούν να προσαρμόζονται με επιτυχία στις ιδιότητες των υπό εξέταση σημάτων.

Στη διεθνή βιβλιογραφία υπάρχει μια πληθώρα από διαφορετικά σύνολα χαρακτηριστικών που μπορούν να χρησιμοποιηθούν για την περιγραφή των εικόνων. Κάποιες από τις κλασσικές προσεγγίσεις περιλαμβάνουν τους συντελεστές του μετασχηματισμού Fourier της εικόνας, τα χαρακτηριστικά που προκύπτουν από την εφαρμογή του διακριτού μετασχηματισμού συνημιτόνου (DCT) ή των Gabor συναρτήσεων, τους συντελεστές κυματιδίων και πολλά άλλα. Κοινό χαρακτηριστικό όλων αυτών των συνόλων είναι ότι προσπαθούν να αναπαραστήσουν τις εικόνες με ένα βέλτιστο τρόπο. Στο παρόν Κεφάλαιο θα συζητήσουμε για τις αναπαραστάσεις εικόνων στο πεδίο χωρικά εντοπισμένων κυματιδιακών μετασχηματισμών (*wavelets*) που υπερτερούν στην περιγραφή μη στάσιμων φυσικών εικόνων και κωδικοποιούν με αποδοτικό τρόπο τη σχετική πληροφορία. Οι συγκεκριμένες αναπαραστάσεις αποδεικνύονται ιδανικές για τη στατιστική περιγραφή των εικόνων και μας επιτρέπουν τη χρήση εξελιγμένων πιθανοτικών τεχνικών επεξεργασίας. Επιπλέον επειδή το ενδιαφέρον μας επικεντρώνεται σε μια ειδική κατηγορία φυσικών εικόνων, δηλαδή τις εικόνες εκείνες που λαμβάνονται μέσω της ανίχνευσης του αριθμού των φωτονίων που εκπέμπει η φωτεινή πηγή, θα μελετήσουμε διεξοδικά και μια συγγενή κατηγορία μετασχηματισμών οι οποίοι διατηρούν πολλά από τα πλεονεκτήματα των κυματιδιακών μετασχηματισμών ενώ παράλληλα εμφανίζονται καταλληλότεροι για την επεξεργασία του συγκεκριμένου είδους εικόνων.

2.1 Αναπαράσταση Σημάτων και Εικόνων με Κυματίδια Εντοπισμένα στο Χώρο και την Κλίμακα

Ο μετασχηματισμός κυματιδίων (wavelet) έχει αναδειχθεί τα τελευταία χρόνια ένα σημαντικό νέο εργαλείο για την αναπαράσταση μη-στάσιμων σημάτων και έχει αποδειχθεί εξαιρετικά προνομιακός στην ανάλυση και επεξεργασία εικόνων [Mal89b, Dau92, Mey93, Vai93, VK95, SN96, Mal99], γνωρίζοντας πλήθος εφαρμογών σε προβλήματα, όπως μεταξύ άλλων είναι η συμπίεση [Sha93], η κατάτμηση [CB01], η ενδοσυμπλήρωση [PM08] και η αποθορυβοποίηση [CB98, ML99, MKRM99, CYV00, SS02a, SS02b]. Προέκυψε μέσα από τη σύνθεση ιδεών προερχόμενων από διαφορετικά ερευνητικά πεδία, όπως είναι η επεξεργασία σήματος, η όραση υπολογιστών, η αρμονική ανάλυση, η φυσική και η στατιστική [HW06]. Ο μετασχηματισμός κυματιδίων, ομοίως με τον μετασχηματισμό Fourier, περιγράφει ένα σήμα ως διάνυσμα σε έναν κατάλληλο διανυσματικό (γραμμικό) χώρο. Η διαφορά τους έγκειται στο γεγονός ότι ο μετασχηματισμός κυματιδίων αντικαθιστά τις ημιτονοειδείς συναρτήσεις βάσης άπειρης ταλάντωσης, οι οποίες χρησιμοποιούνται στον μετασχηματισμό Fourier, με ένα σύνολο από τοπικά ταλαντούμενες συναρτήσεις βάσης που καλούνται κυματίδια (wavelets). Η δημιουργία της πρώτης οικογένειας ορθοκανονικών βάσεων κυματιδίων οφείλεται στο Meyer και έδωσε την ώθηση στην ερευνητική κοινότητα για την αναζήτηση νέων ορθοκανονικών βάσεων. Η συστηματική κατασκευή τους κατέστη εφικτή μέσω της θεωρίας που ανέπτυξαν οι Mallat [Mal89a] και Meyer [Mey92] και η οποία βασίζεται στην ιδέα της πολυκλιμακωτής προσέγγισης (Multiresolution Approximation - MRA) σημάτων. Η θεωρία αυτή παρουσιάζει πολλά κοινά σημεία με τη λαπλασιανή πυραμίδα των Burt και Adelson [BA83] που χρησιμοποιείται για την ανάλυση εικόνων σε πολλαπλές κλίμακες.

Θεωρούμε ένα 1-Δ σήμα πεπερασμένης ενέργειας $x(t) \in \mathbf{L}^2(\mathbb{R})$. Η θεωρία της πολυκλιμακωτής προσέγγισης μελετά τη δυνατότητα επιλογής των συναρτήσεων κυματιδίων, ώστε να συνιστούν μια ορθοκανονική βάση στο χώρο $\mathbf{L}^2(\mathbb{R})$. Οι συναρτήσεις αυτές αποτελούν κλιμακούμενες και μετατοπισμένες εκδοχές μιας βασικής πραγματικής συνάρτησης μητρικού $\psi(t)$ (ζωνοπερατή) κυματιδίου και μιας βασικής πραγματικής συνάρτησης πατρικού $\phi(t)$ (βαθυπερατή) κυματιδίου. Όταν οι συναρτήσεις $\psi(t)$ και $\phi(t)$ επιλεγούν προσεκτικά, τότε οι κλιμακούμενες και μετατοπισμένες εκδοχές τους

$$\psi_{j,n} = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j}\right) \tag{2.1α'}$$

$$\phi_{j,n} = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t - 2^j n}{2^j}\right), \tag{2.1β'}$$

συνιστούν πολυκλιμακωτή ορθοκανονική βάση και μπορούν να χρησιμοποιηθούν για την αναπαράσταση του σήματος $x(t)$ σε πολλαπλές κλίμακες ως [Dau92]:

$$x(t) = \sum_{n=-\infty}^{\infty} c_J(n)\phi_{J,n}(t-n) + \sum_{j=-\infty}^{J} \sum_{n=-\infty}^{\infty} d_j(n)\psi_{j,n}(t-n)\,. \tag{2.2}$$

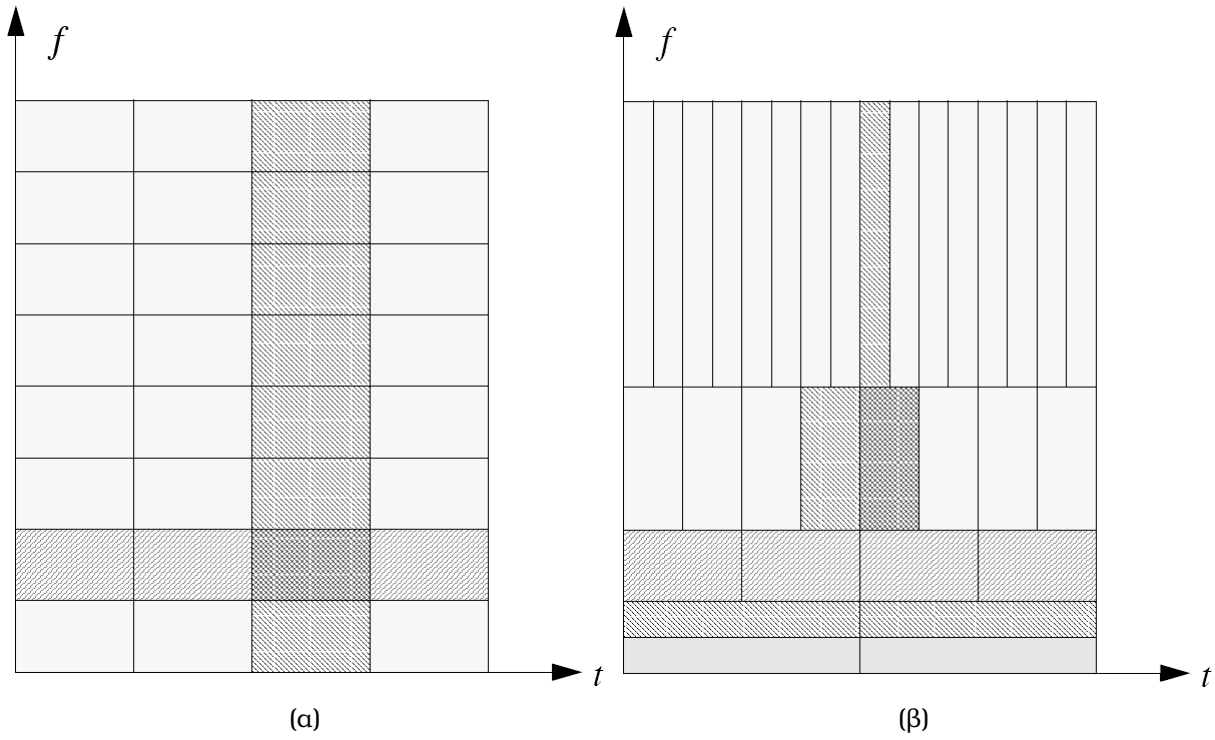
Σε αυτή την αναπαράσταση το j υποδηλώνει την κλίμακα της ανάλυσης με το μικρότερο j να αντιστοιχεί σε υψηλότερης ευκρίνειας κλίμακα (συνεπώς, η J είναι η πιο αδρή κλίμακα), ενώ το n ορίζει το χωρικό πλαίσιο της ανάλυσης. Οι συντελεστές $c(n)$ και $d(n)$ ονομάζονται συντελεστές κλιμάκωσης (scaling) και λεπτομέρειας (detail), αντιστοίχως και, επειδή το σύνολο των συναρτήσεων $\{\phi_{J,n},\, \psi_{j,n},\, j,n \in \mathbb{Z}\}$ αποτελεί πολυκλιμακωτή ορθοκανονική βάση, μπορούν να υπολογιστούν

μέσω των παρακάτω εσωτερικών γινομένων:

$$c_J(n) = \langle x, \phi_{J,n} \rangle = \int_{-\infty}^{\infty} x(t)\phi_{J,n}(t) \tag{2.3α'}$$

$$d_j(n) = \langle x, \psi_{j,n} \rangle = \int_{-\infty}^{\infty} x(t)\psi_{j,n}(t)\,. \tag{2.3β'}$$

Οι συντελεστές $c(n)$ και $d(n)$ παρέχουν μια χρονο-συχνοτική ανάλυση του σήματος μετρώντας το συχνοτικό του περιεχόμενο (καθορίζεται από την κλιμακα ανάλυσης j) σε διαφορετικές χρονικές στιγμές (καθορίζεται από τον παράγοντα μετατόπισης n). Η συγκεκριμένη χρονο-συχνοτική ανάλυση παρουσιάζει ομοιότητες με την ανάλυση του σήματος με τον βραχύχρονο μετασχηματισμό Fourier [RS78]. Ωστόσο, η βασική διαφορά είναι ότι, ενώ στο βραχύχρονο μετασχηματισμό Fourier το παράθυρο της συχνοτικής ανάλυσης διατηρείται σταθερό, στο μετασχηματισμό κυματιδίων η ευκρίνεια της φασματικής μέτρησης μεταβάλλεται. Αυτή η σημαντική διαφορά παρουσιάζεται στο Σχ. 2.1 με τη βοήθεια των Heisenberg διαγραμμάτων.



Σχήμα 2.1: Ιδιότητες χρονικού/φασματικού εντοπισμού του βραχύχρονου μετασχηματισμού Fourier και της κυματιδιακής ανάλυσης με τη βοήθεια των Heisenberg διαγραμμάτων (α) Κυματίδια χρόνου-συχνότητας όπου τα άτομα εντός κάθε χρονικού παραθύρου καλύπτουν ομοιόμορφα τον άξονα συχνοτήτων. (β) Κυματίδια χρόνου-κλίμακας, σε διαδοχικές κλίμακες ανάλυσης υποδιπλασιάζεται η χρονική διάρκεια του φίλτρου (κάλυψη του χρονικού άξονα), ενώ παράλληλα διπλασιάζεται το εύρος ζώνης του (κάλυψη του άξονα συχνοτήτων). Ανατυπωμένο σχήμα από το [VK95].

2.1.1 Πολυκλιμακωτή Ανάλυση και Συστοιχίες Φίλτρων

Σύμφωνα με τη θεωρία της πολυκλιμακωτής ανάλυσης οι συναρτήσεις του πατρικού και του μητρικού κυματιδίου μπορούν να προκύψουν ως λύσεις των παρακάτω αναδρομικών εξισώσεων:

$$\phi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} h_0(n)\phi(2t-n) \tag{2.4α'}$$

$$\psi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} h_1(n)\phi(2t-n)\,. \tag{2.4β'}$$

Η Εξ. (2.4α') καλείται εξίσωση διαστολής (*dilation*) και η επίλυσή της οδηγεί στη συνάρτηση του πατρικού κυματιδίου, ενώ η Εξ. (2.4β') καλείται εξίσωση κυματιδίου και παράγει τη συνάρτηση του μητρικού κυματιδίου. Οι ακολουθίες $h_0(n)$ και $h_1(n)$ μπορούν να θεωρηθούν ως οι συντελεστές ενός βαθυπερατού και ενός υψιπερατού διακριτού φίλτρου, αντιστοίχως. Βασιζόμενοι σε αυτή την παρατήρηση οι Mallat [Mal89a] και Meyer [Mey92] απέδειξαν ότι οι συναρτήσεις βάσης του πατρικού και μητρικού κυματιδίου μπορούν να χαρακτηριστούν από τους συντελεστές των δύο αυτών φίλτρων. Έτσι, αν τα φίλτρα επιλεγούν κατάλληλα τότε για το $h_0(n)$ αντιστοιχεί μια συνεχούς χρόνου συνάρτηση πατρικού κυματιδίου, ενώ για το $h_1(n)$ μια συνεχούς χρόνου συνάρτηση μητρικού κυματιδίου. Προκειμένου να είναι εφικτή η τέλεια ανακατασκευή του αρχικού σήματος, απαιτείται από τα φίλτρα h_0 και h_1 να ικανοποιούν τις ακόλουθες σχέσεις [Vai93, VK95]:

$$|H_0(\omega)|^2 + |H_0(\omega+\pi)|^2 = 2 \tag{2.5α'}$$

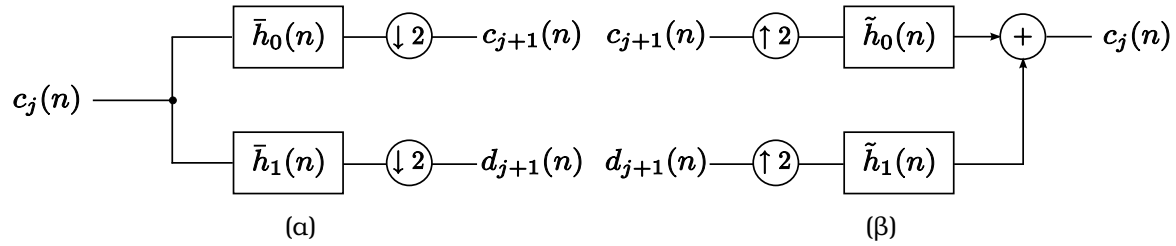
$$h_1(n) = (-1)^{1-n} h_0(1-n)\,. \tag{2.5β'}$$

Η πρώτη από τις δύο συνθήκες εξασφαλίζει ότι το $H_0(\omega)$ είναι ο μετασχηματισμός Fourier μιας συνάρτησης πατρικού κυματιδίου, ενώ η δεύτερη χαρακτηρίζει το φίλτρο h_0 ως συζυγές κατοπτρικό φίλτρο (Conjugate Mirror Filter - CMF). Με βάση τα παραπάνω διαπιστώνουμε ότι το πρόβλημα της εύρεσης κατάλληλων συναρτήσεων κυματιδίων μπορεί να αναχθεί σε πρόβλημα σχεδιασμού φίλτρων που επιτρέπουν την πλήρη ανακατασκευή του σήματος. Για το συγκεκριμένο πρόβλημα έχουν αναπτυχθεί εξελιγμένες τεχνικές σχεδίασης οι οποίες επιτρέπουν την κατασκευή FIR φίλτρων με διάφορες επιθυμητές ιδιότητες για τις συναρτήσεις κυματιδίων. Τέτοιες επιθυμητές ιδιότητες αποτελούν η πεπερασμένη χρονική διάρκεια και η γρήγορη συχνοτική εξασθένιση (προκειμένου να διασφαλιστεί ότι η ανάλυση είναι όσο το δυνατόν πιο τοπική στο κοινό χρονο-συχνοτικό πεδίο), καθώς και η αυξημένη ομαλότητα για το μητρικό κυματίδιο.

Η ερμηνεία των ακολουθιών $h_0(n)$ και $h_1(n)$ ως συντελεστών διακριτών φίλτρων είναι εξαιρετικής σημασίας, καθώς συνδέει την αναπαράσταση ενός σήματος σε μια ορθοκανονική πολυκλιμακωτή κυματιδιακή βάση με την επαναληπτική ανάλυση του σήματος σε υποζώνες με τη χρήση συστοιχίας φίλτρων. Η ανακάλυψη της στενής αυτής σχέσης επέτρεψε την κατασκευή ενός αναδρομικού αλγορίθμου γραμμικής πολυπλοκότητας για τον υπολογισμό των συντελεστών κλιμάκωσης, $c_j(n)$, και λεπτομέρειας, $d_j(n)$ [Mal89b]. Ο αλγόριθμος αυτός καλείται *Ταχύς Μετασχηματισμός Κυματιδίων* (Fast Wavelet Transform - FWT) και είναι εξίσου αποδοτικός και ευσταθής με τον αντίστοιχο Ταχύ Fourier Μετασχηματισμό για την ανάλυση σημάτων σε Fourier βάσεις. Με βάση τον ταχύ μετασχηματισμό κυματιδίων, αν γνωρίζουμε τους συντελεστές κλιμάκωσης και λεπτομέρειας του σήματος $x(t)$ σε μια κλίμακα j, τότε μπορούμε να βρούμε τους αντίστοιχους συντελεστές για την αμέσως αδρότερη κλίμακα μέσω των παρακάτω αναδρομικών τύπων:

$$c_{j+1}(p) = \sum_{n=-\infty}^{\infty} h_0(n-2p)c_j(n) = (c_j * \bar{h}_0)(2p) \tag{2.6α'}$$

$$d_{j+1}(p) = \sum_{n=-\infty}^{\infty} h_1(n-2p)c_j(n) = (c_j * \bar{h}_1)(2p)\,, \tag{2.6β'}$$



Σχήμα 2.2: Ένα στάδιο κυματιδιακής ανάλυσης (α) και σύνθεσης (β) με τον αναδρομικό ταχύ μετασχηματισμό κυματιδίων που βασίζεται σε συστοιχίες φίλτρων. Στην περίπτωση συζυγών ανακλαστικών φίλτρων, το δυϊκό ζεύγος φίλτρων $\left\{\tilde{h}_0, \tilde{h}_1\right\}$ είναι το ίδιο με το ζεύγος φίλτρων $\{h_0, h_1\}$.

όπου με $\bar{h}(n)$ συμβολίζουμε το ανεστραμμένο φίλτρο, δηλαδή $\bar{h}(n) = h(-n)$. Ουσιαστικά, για να υπολογίσουμε τους συντελεστές κλιμάκωσης ή λεπτομέρειας σε ένα επίπεδο ανάλυσης $j+1$, απαιτείται το φιλτράρισμα των αντίστοιχων συντελεστών της προηγούμενης κλίμακας με τις ανεστραμμένες εκδοχές των φίλτρων $h_0(n)$, $h_1(n)$ και η υποδειγματοληψία της εξόδου κατά ένα παράγοντα 2. Μπορούμε επίσης μέσω του αντίστροφου FWT να ακολουθήσουμε την αντίστροφη διαδικασία, δηλαδή από τους συντελεστές μιας αδρής κλίμακας να υπολογίσουμε τους συντελεστές κλιμάκωσης της αμέσως επόμενης ευκρινέστερης κλίμακας

$$\begin{aligned} c_j(p) &= \sum_{n=-\infty}^{\infty} h_0(p-2n)c_{j+1}(n) + \sum_{n=-\infty}^{\infty} h_1(p-2n)d_{j+1}(n) \\ &= \left(\check{c}_{j+1} * h_0\right)(p) + \left(\check{d}_{j+1} * h_1\right)(p)\,, \end{aligned} \tag{2.7}$$

όπου

$$\check{c}(n) = \begin{cases} c(p) & \text{εάν } n = 2p, \\ 0 & \text{εάν } n = 2p+1\,. \end{cases} \tag{2.8}$$

Η Εξ. (2.21) υποδηλώνει ότι η διαδικασία σύνθεσης περιλαμβάνει ένα στάδιο υπερδειγματοληψίας με ένθεση μηδενικών μεταξύ γειτονικών δειγμάτων ακολουθούμενη από ένα στάδιο φιλτραρίσματος με τα φίλτρα h_0 και h_1. Η διαδικασία ανάλυσης και η διαδικασία σύνθεσης με τη χρήση συστοιχίας φίλτρων παρουσιάζεται σε μορφή διαγραμμάτων στο Σχ. 2.2.

2.1.2 Συστοιχίες Φίλτρων Τέλειας Ανακατασκευής

Όπως ήδη αναφέραμε στην προηγούμενη ενότητα, ο ταχύς μετασχηματισμός κυματιδίων αναλύει ένα σήμα εισόδου σε δύο συχνοτικές συνιστώσες, μια βαθυπερατή και μια υψιπερατή, και τις υποδειγματοληπτεί κατά ένα παράγοντα 2. Η ανακατασκευή του σήματος πραγματοποιείται μέσω του αντίστροφου μετασχηματισμού. Αυτού του είδους η πολυφασματική ανάλυση αποτέλεσε ένα σημαντικό πεδίο έρευνας από τα μέσα της δεκαετίας του 1970, οπότε ανακαλύφθηκε ότι κάτι τέτοιο είναι εφικτό με τη χρήση ορθογώνιων κατοπτρικών φίλτρων (Quadrature Mirror Filters - QMF) [Vai93]. Ωστόσο, τα φίλτρα αυτά εκτός από μια ειδική περίπτωση, τα Haar φίλτρα, δεν μπορούν να έχουν πεπερασμένη κρουστική απόκριση και, επομένως, δεν παρουσιάζουν σημαντικό πρακτικό ενδιαφέρον. Στην παρούσα ενότητα εξετάζουμε τις συνθήκες που πρέπει να ικανοποιούνται προκειμένου να είναι εφικτή η τέλεια κατασκευή με τη χρήση FIR φίλτρων. Τα φίλτρα που προκύπτουν μας οδηγούν σε μια γενική κατηγορία μετασχηματισμών κυματιδίων.

Σε μια συστοιχία δύο φίλτρων ένα σήμα εισόδου $x(n)$ συνελίσσεται με ένα βαθυπερατό, $\bar{h}_0(n) = h_0(-n)$ και ένα υψιπερατό φίλτρο, $\bar{h}_1(n) = h_1(-n)$, και η έξοδος των φίλτρων υποδειγματοληπτείται κατά ένα παράγοντα 2. Το ανακατασκευασμένο σήμα $\hat{x}(n)$ προκύπτει ως το άθροισμα των υπερδειγματοληπτημένων και φιλτραρισμένων εξόδων με τα δυϊκά φίλτρα $\tilde{h}_0(n)$ $\tilde{h}_1(n)$, όπως φαίνεται στο Σχ. 2.2. Χρησιμοποιώντας τις γνωστές ιδιότητες υποδειγματοληπτημένων/υπερδειγματοληπτημένων ακολουθιών καταλήγουμε στις εξής σχέσεις μεταξύ εισόδου-εξόδου στο πεδίο του μετασχηματισμού Z:

$$\begin{aligned}\hat{X}(z) =& \frac{1}{2}\left[X(z)H_0(z^{-1})\tilde{H}_0(z) + X(-z)H_0(-z^{-1})\tilde{H}_0(z)\right] \\ &+ \frac{1}{2}\left[X(z)H_1(z^{-1})\tilde{H}_1(z) + X(-z)H_1(-z^{-1})\tilde{H}_1(z)\right] . \end{aligned} \tag{2.9}$$

Ο πρώτος όρος που εμφανίζεται στο δεξιό μέλος της Εξ. (2.9) προκύπτει ακολουθώντας την επεξεργασία στον άνω κλάδο του ζεύγους των συστοιχιών ανάλυσης και σύνθεσης, ενώ ο δεύτερος όρος προκύπτει από τον αντίστοιχο κάτω κλάδο. Λόγω των πράξεων υποδειγματοληψίας και υπερδειγματοληψίας που λαμβάνουν χώρα παρουσιάζονται φαινόμενα αναδίπλωσης συχνοτήτων (*aliasing*). Αυτό είναι φανερό από την εμφάνιση των όρων $X(-z)$. Αναδιατάσσοντας την Εξ. (2.9) θα έχουμε:

$$\begin{aligned}\hat{X}(z) =& \frac{1}{2}\left[H_0(z^{-1})\tilde{H}_0(z) + H_1(z^{-1})\tilde{H}_1(z)\right] X(z) \\ &+ \frac{1}{2}\left[H_0(-z^{-1})\tilde{H}_0(z) + H_1(-z^{-1})\tilde{H}_1(z)\right] X(-z) . \end{aligned} \tag{2.10}$$

Προκειμένου να αντιμετωπιστεί το πρόβλημα της αναδίπλωσης συχνοτήτων, θα πρέπει να επιλέξουμε τα φίλτρα ανάλυσης και σύνθεσης με τέτοιο τρόπο, ώστε η αναδίπλωση συχνοτήτων, η οποία εμφανίζεται στο στάδιο ανάλυσης, να ακυρώνεται από την αναδίπλωση συχνοτήτων που παρουσιάζεται κατά το στάδιο της σύνθεσης δηλαδή, θα πρέπει να ισχύει:

$$H_0(-z^{-1})\tilde{H}_0(z) + H_1(-z^{-1})\tilde{H}_1(z) = 0 . \tag{2.11}$$

Επιπλέον, για να έχουμε τέλεια ανακατασκευή, θα πρέπει στην έξοδο του συστήματος να λαμβάνει κανείς την αρχική είσοδο χωρίς καμία παραμόρφωση δηλαδή, θα πρέπει να ισχύει:

$$H_0(z^{-1})\tilde{H}_0(z) + H_1(z^{-1})\tilde{H}_1(z) = 2 \tag{2.12}$$

Σε αυτό το σημείο έχουμε να σχεδιάσουμε τέσσερα φίλτρα, h_0, h_1, $\tilde{h}_0$, $\tilde{h}_1$ τα οποία θα πρέπει να ικανοποιούν τις συνθήκες (2.11) και (2.12). Επιλύοντας το σύστημα των δύο εξισώσεων ως προς τα δυϊκά φίλτρα $\tilde{h}_0$ και $\tilde{h}_1$ καταλήγουμε στην

$$\begin{pmatrix} \tilde{H}_0(z) \\ \tilde{H}_1(z) \end{pmatrix} = \frac{2}{H_0(z^{-1})H_1(-z^{-1}) - H_0(-z^{-1})H_1(z^{-1})} \begin{pmatrix} H_1(-z^{-1}) \\ -H_0(-z^{-1}) \end{pmatrix} , \tag{2.13}$$

που είναι η σχέση που συνδέει τα φίλτρα σύνθεσης με τα φίλτρα ανάλυσης. Μπορούμε, επομένως, να επιλέξουμε τα φίλτρα ανάλυσης $h_0(n)$, $h_1(n)$ και στη συνέχεια μέσω της σχέσης (2.13) να υπολογίσουμε τα αντίστοιχα φίλτρα σύνθεσης, ώστε να εξασφαλίζεται η τέλεια ανακατασκευή. Βέβαια η επιλογή των φίλτρων θα πρέπει να γίνεται με τέτοιο τρόπο, ώστε τα φίλτρα που θα προκύπτουν να είναι ευσταθή.

2.1.3 Διορθογώνιος Μετασχηματισμός Κυματιδίων

Όταν όλα τα φίλτρα των συστοιχιών διαθέτουν πεπερασμένη κρουστική απόκριση, FIR φίλτρα, τότε προκύπτουν απλές σχέσεις που συνδέουν τα φίλτρα ανάλυσης και σύνθεσης [Mal99]. Σε μια

τέτοια περίπτωση ισχύουν οι ακόλουθες σχέσεις:

$$H_0^*(\omega)\tilde{H}_0(\omega) + H_0^*(\omega+\pi)\tilde{H}_0(\omega+\pi) = 2 \tag{2.14α'}$$

$$h_1(n) = (-1)^{1-n}\tilde{h}_0(1-n) \quad \text{και} \quad \tilde{h}_1(n) = (-1)^{1-n}h_0(1-n)\,, \tag{2.14β'}$$

οι οποίες αποτελούν γενικεύσεις των σχέσεων (2.5α') και (2.5β'). Πράγματι, αν επιβάλλουμε το βαθυπερατό φίλτρο ανάλυσης, $h_0(n)$, να είναι ίσο με το βαθυπερατό φίλτρο σύνθεσης, $\tilde{h}_0(n)$, τότε κανείς καταλήγει από τις σχέσεις (2.14α') και (2.14β') στις σχέσεις (2.5α') και (2.5β') οι οποίες χαρακτηρίζουν τα συζυγή κατοπτρικά φίλτρα και χρησιμοποιούνται για την κατασκευή ορθογώνιων συναρτήσεων κυματιδίων.

Όπως μπορεί κανείς να παρατηρήσει, η κατασκευή ορθογώνιων κυματιδίων δεν επιτρέπει μεγάλη ελευθερία στο σχεδιασμό των φίλτρων. Αν επιτρέψουμε τη χρήση διαφορετικών ζευγών φίλτρων κατά το στάδιο της ανάλυσης και σύνθεσης, του πρωτεύοντος (primal) $\{h_0, h_1\}$ και του δυϊκού (dual) $\left\{\tilde{h}_0, \tilde{h}_1\right\}$, αντιστοίχως, τότε καταλήγουμε στα διορθογώνια κυματίδια (biorthogonal) που συνιστούν διορθογώνιες (μη ορθοκανονικές) Riesz βάσεις του $\mathbf{L}^2(\mathbb{R})$. Αναλόγως, με την ορθογώνια περίπτωση το πρωτεύον ζεύγος πατρικού και μητρικού κυματιδίου, $\{\phi, \psi\}$, και το δυϊκό ζεύγος, $\left\{\tilde{\phi}, \tilde{\psi}\right\}$, χαρακτηρίζονται από το πρωτεύων και δυϊκό ζεύγος φίλτρων, αντιστοίχως

$$\phi(t) = \sqrt{2}\sum_{n=-\infty}^{\infty} h_0(n)\phi(2t-n), \quad \tilde{\phi}(t) = \sqrt{2}\sum_{n=-\infty}^{\infty} \tilde{h}_0(n)\tilde{\phi}(2t-n) \tag{2.15α'}$$

$$\psi(t) = \sqrt{2}\sum_{n=-\infty}^{\infty} h_1(n)\phi(2t-n), \quad \tilde{\psi}(t) = \sqrt{2}\sum_{n=-\infty}^{\infty} \tilde{h}_1(n)\tilde{\phi}(2t-n)\,. \tag{2.15β'}$$

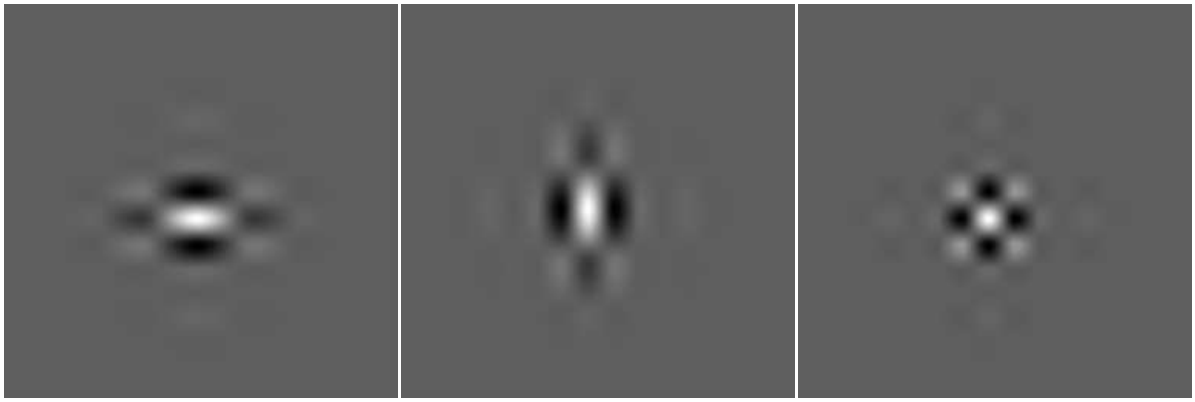
Επιπλέον, και στην περίπτωση των διορθογώνιων κυματιδίων είναι δυνατή η πολυκλιμακωτή ανάλυση ενός συνεχούς χρόνου σήματος $x(t)$ ως εξής:

$$x(t) = \sum_{n=-\infty}^{\infty} c_J(n)\tilde{\phi}_{J,n}(t-n) + \sum_{j=-\infty}^{J}\sum_{n=-\infty}^{\infty} d_j(n)\tilde{\psi}_{j,n}(t-n)\,, \tag{2.16}$$

όπου οι συντελεστές κλιμάκωσης $c_J(n)$ και λεπτομέρειας $d_j(n)$ υπολογίζονται μέσω των εσωτερικών γινομένων με το πρωτεύον πατρικό και μητρικό κυματίδιο, όπως δίνονται στις Εξ. (2.3α') και (2.3β'). Λόγω της δυϊκότητας που εμφανίζουν τα δύο ζεύγη φίλτρων, μπορεί κανείς να χρησιμοποιήσει το δυϊκό μετασχηματισμό κυματιδίων που προκύπτει, αν γίνει χρήση των φίλτρων $\left\{\tilde{h}_0, \tilde{h}_1\right\}$ κατά το στάδιο της ανάλυσης και των φίλτρων $\{h_0, h_1\}$ κατά το στάδιο της σύνθεσης. Η κατασκευή διορθογώνιων κυματιδίων παρέχει μεγαλύτερη ελευθερία στο σχεδιασμό των φίλτρων και επιτρέπει τη χρήση συμμετρικών βαθυπερατών φίλτρων h_0, $\tilde{h}_0$ γραμμικής φάσης τα οποία παράγουν συμμετρικά πατρικά κυματίδια ϕ, $\tilde{\phi}$ και συμμετρικά (αντισυμμετρικά) μητρικά κυματίδια ψ, $\tilde{\psi}$ για περιττό (άρτιο) μήκος φίλτρου h_0.

2.1.4 Διακριτός Μετασχηματισμός Κυματιδίων

Ο υπολογισμός του μετασχηματισμού κυματιδίων για διακριτά σήματα $x(n)$ μπορεί να πραγματοποιηθεί μέσω της ανάλυσης του σήματος από μια συστοιχία φίλτρων όμοια με αυτές που χρησιμοποιούνται στον αναδρομικό αλγόριθμο του FWT ο οποίος περιγράφεται στην Εξ. (2.6). Σε αυτή την περίπτωση θεωρεί κανείς ότι το διακριτό σήμα $x(n)$ αντιστοιχεί στους συντελεστές κλιμάκωσης $c_{j_f}(n)$ ενός συνεχούς σήματος $x(t)$ και η κυματιδιακή ανάλυση πραγματοποιείται μέσω της αναδρομικής χρήσης της συστοιχίας των φίλτρων για ένα πεπερασμένο εύρος κλιμάκων $j_f \leq j \leq J$, όπου $L = J - j_f$ είναι ο συνολικός αριθμός των σταδίων της αναδρομής.



(α) (β) (γ)

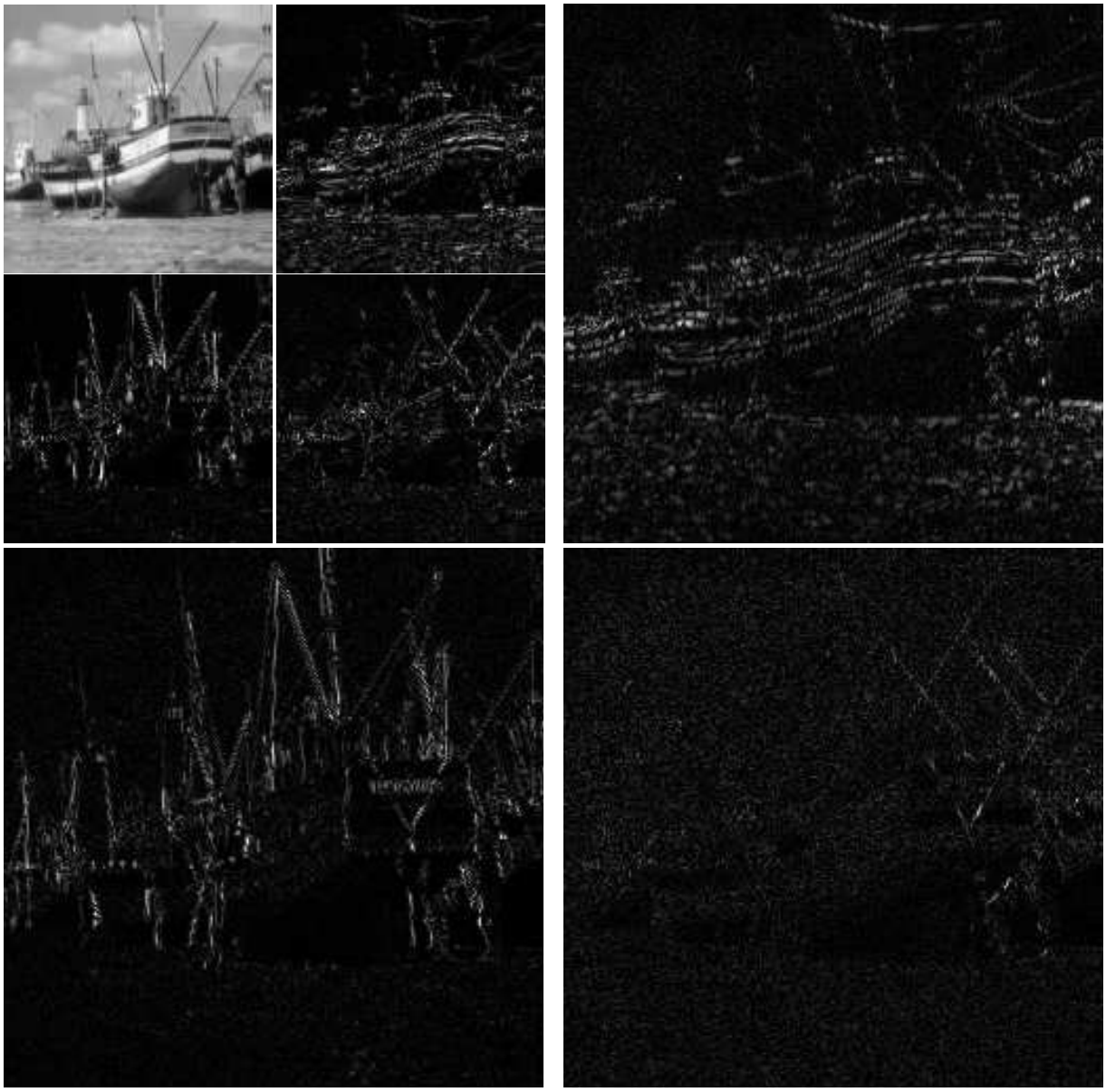
Σχήμα 2.3: Τρία τυπικά μητρικά κυματίδια που αντιστοιχούν στο 2-Δ διαχωρίσιμο μετασχηματισμό DWT. (α) LH κυματίδιο, (β) HL κυματίδιο, (γ) HH κυματίδιο.

Η επεξεργασία που υφίσταται το διακριτό σήμα μέσα από αυτή τη διαδικασία αντιστοιχεί στο διακριτό μετασχηματισμό κυματιδίων (Discrete Wavelet Transform - DWT) και μπορεί να ερμηνευθεί ως η ανάλυση του σήματος σε μια βάση του χώρου $\ell^2(\mathbb{Z})$. Για L στάδια εφαρμογής του αναδρομικού αλγορίθμου το διακριτό σήμα $x(n)$ αναλύεται σε ένα σύνολο διανυσμάτων $\left\{\phi_J(p-2^L n), \psi_j(p-2^L n),\, j_f \le j \le J,\, n \in \mathbb{Z}\right\}$ που αποδεικνύεται ότι συνιστούν βάση του χώρου $\ell^2(\mathbb{Z})$ [Mal99]. Μάλιστα για την περίπτωση χρήσης συζυγών ανακλαστικών φίλτρων στη συστοιχία η βάση που προκύπτει είναι ορθοκανονική. Στην αντίθετη περίπτωση, στην οποία κανείς χρησιμοποιεί διαφορετικά φίλτρα στο στάδιο ανάλυσης και σύνθεσης, η προκύπτουσα βάση είναι μη-ορθογώνια. Τέλος, ανάλογα με την περίπτωση των συνεχών σημάτων, έτσι και εδώ, αν επιτρέψουμε τη χρήση ενός πρωτεύοντος ζεύγους φίλτρων και του δυϊκού του ζεύγους στα στάδια ανάλυσης και σύνθεσης, αντιστοίχως, οδηγούμαστε σε διορθογώνιες διακριτές βάσεις.

2.1.5 Πολυκλιμακωτή Ανάλυση σε 2-Δ Σήματα

Η κυματιδιακή ανάλυση που συζητήσαμε στην Ενοτ. 2.1.1 για 1-Δ σήματα είναι δυνατόν να επεκταθεί και σε σήματα μεγαλύτερων διαστάσεων, όπως είναι οι εικόνες (2-Δ). Συνήθως μια τέτοια ανάλυση πραγματοποιείται για λόγους υπολογιστικής ευκολίας χρησιμοποιώντας διαχωρίσιμες στρατηγικές επεξεργασίας, όπως συμβαίνει και στην περίπτωση του μετασχηματισμού Fourier. Ωστόσο, θα πρέπει να αναφερθεί ότι είναι επίσης εφικτό η ανάλυση στο πεδίο του μετασχηματισμού κυματιδίων υψηλότερων διαστάσεων να πραγματοποιηθεί και με μη-διαχωρίσιμες τεχνικές [SA90, KV92, VK95, KS00, FVU05]. Κατά τη διαχωρίσιμη επεξεργασία ο 1-Δ μετασχηματισμός εφαρμόζεται διαδοχικά σε όλες τις διαστάσεις. Στην περίπτωση των εικόνων ο 1-Δ κυματιδιακός μετασχηματισμός εφαρμόζεται διαδοχικά στην οριζόντια και στην κάθετη διεύθυνση οδηγώντας στην κατασκευή μιας οικογένειας κυματιδίων που αποτελείται από ένα πατρικό και τρία μητρικά κυματίδια:

$$\begin{aligned}
\phi(x_1,x_2) &= \phi(x_1)\phi(x_2) \quad \text{(LL κυματίδιο)} \\
\psi^1(x_1,x_2) &= \phi(x_1)\psi(x_2) \quad \text{(LH κυματίδιο)} \\
\psi^2(x_1,x_2) &= \psi(x_1)\phi(x_2) \quad \text{(HL κυματίδιο)} \\
\psi^3(x_1,x_2) &= \psi(x_1)\psi(x_2) \quad \text{(HH κυματίδιο)}\,.
\end{aligned} \tag{2.17}$$



Σχήμα 2.4: Απόλυτη τιμή των συντελεστών κυματιδίων που προκύπτουν από την ανάλυση της 512×512 εικόνας boat μέσω του 2-Δ DWT για 2 επίπεδα ανάλυσης. Πάνω δεξιά παρουσιάζονται οι συντελεστές που σχετίζονται με το LH κυματίδιο (κωδικοποίηση οριζόντιων ακμών), κάτω αριστερά οι συντελεστές που σχετίζονται με το HL κυματίδιο (κωδικοποίηση κατακόρυφων ακμών) και κάτω δεξιά οι συντελεστές που σχετίζονται με το HH κυματίδιο (κωδικοποίηση διαγώνιων ακμών).

Το πατρικό κυματίδιο (LL) προκύπτει ως το γινόμενο δύο βαθυπερατών συναρτήσεων και, συνεπώς, αντιστοιχεί σε βαθυπερατή ανάλυση της εικόνας (διατήρηση μόνο των αδρών χαρακτηριστικών). Τα τρία μητρικά κυματίδια κωδικοποιούν λεπτομερή χαρακτηριστικά (ακμές) της εικόνας σε διαφορετικές κλίμακες και κατευθύνσεις. Διαισθητικά μια ακμή σε μια εικόνα είναι ένα σημείο στο οποίο παρουσιάζονται μεγάλες διακυμάνσεις στις τιμές των εικονοστοιχείων, δηλαδή έχουμε την εμφάνιση ακμής, όταν η τιμή ενός εικονοστοιχείου είναι σημαντικά διαφορετική από τις τιμές των γειτονικών του εικονοστοιχείων. Αν σε μια συγκεκριμένη χωρική θέση η διακύμανση των τιμών είναι μικρή προς την κατακόρυφη κατεύθυνση, αλλά μεγάλη προς την οριζόντια, τότε το εικονοστοιχείο αποτελεί σημείο κατακόρυφης ακμής στην εικόνα. Αναλόγως, αν η διακύμανση είναι μεγάλη προς την οριζόντια κατεύθυνση, αλλά μικρή προς την κατακόρυφη, τότε το εικονοστοιχείο αποτελεί σημείο οριζόντιας ακμής στην εικόνα. Αν, τέλος, η διακύμανση είναι μεγάλη και προς τις δύο κατευθύνσεις, τότε το εικονοστοιχείο αποτελεί σημείο διαγώνιας ακμής. Από τις διαχωρίσιμες εκφράσεις των μητρικών κυματιδίων στις Εξ. (2.17) προκύπτει ότι το $|\Psi_1(\omega_1, \omega_2)|$ έχει μεγάλες τιμές σε χαμηλές οριζόντιες συχνότητες ω_1 και σε υψηλές κατακόρυφες συχνότητες ω_2, το $|\Psi_2(\omega_1, \omega_2)|$ είναι μεγάλο για υψηλές οριζόντιες και χαμηλές κατακόρυφες συχνότητες, ενώ το $|\Psi_3(\omega_1, \omega_2)|$ έχει μεγάλες τιμές για υψηλές οριζόντιες και κατακόρυφες συχνότητες. Με βάση τα παραπάνω τα μητρικά κυματίδια μπορούν να αναλύσουν την εικόνα και να κωδικοποιήσουν τις ακμές της σε τρεις διαφορετικές κατευθύνσεις, οριζόντια (LH κυματίδιο, ακμές στη διεύθυνση $0°$), κάθετη (HL κυματίδιο, ακμές στη διεύθυνση $90°$) και διαγώνια (HH κυματίδιο, ακμές στη διεύθυνση $\pm 45°$). Στο Σχ. 2.3 παρουσιάζουμε τρία τυπικά μητρικά κυματίδια τα οποία αντιστοιχούν σε 2-Δ διαχωρίσιμη ανάλυση. Τα κυματίδια αυτά προκύπτουν κατά τη χρήση διορθογώνιων ζευγών φίλτρων. Επιπλέον, στο Σχ. 2.4 απεικονίζονται οι 2-Δ συντελεστές κυματιδίων που προκύπτουν από την ανάλυση μιας τυπικής εικόνας με τον 2-Δ DWT.

Όπως και στη μονοδιάστατη περίπτωση, έτσι και στις δύο διαστάσεις το σύνολο των κλιμακούμενων και μετατοπισμένων εκδοχών του πατρικού και των τριών μητρικών κυματιδίων $\left\{\phi_{Jn}, \psi^1_{jn}, \psi^2_{jn}, \psi^3_{jn}\right\}_{n\in\mathbb{Z}^2}$ συνιστούν Riesz βάση του χώρου $\mathbf{L}^2(\mathbb{R}^2)$, όπου

$$\begin{aligned} \psi^k_{jn} &= \frac{1}{2^j}\psi^k\left(\frac{x_1 - 2^j n_1}{2^j}, \frac{x_2 - 2^j n_2}{2^j}\right) \quad \text{για } 1 \le k \le 3 \\ \phi_{jn} &= \frac{1}{2^j}\phi\left(\frac{x_1 - 2^j n_1}{2^j}, \frac{x_2 - 2^j n_2}{2^j}\right). \end{aligned} \tag{2.18}$$

2.1.6 Υπερπλήρεις Αναπαραστάσεις Σημάτων και Εικόνων

Όπως ήδη αναφέραμε και στην εισαγωγή του κεφαλαίου ο μετασχηματισμός κυματιδίων έχει αποτελέσει ένα πολύ σημαντικό εργαλείο για την επεξεργασία σημάτων. Ωστόσο εκτός από τα πολλά πλεονεκτήματα που παρουσιάζει και καθιστούν προνομιακή τη χρήση του εμφανίζει και δύο σημαντικά μειονεκτήματα. Το πρώτο μειονέκτημα σχετίζεται με το γεγονός ότι η αναπαράσταση σημάτων σε διακριτές βάσεις κυματιδίων χαρακτηρίζεται από έλλειψη μεταφορικής συμμετρίας (translation invariance). Αυτό σημαίνει ότι μια μικρή μετατόπιση του σήματος στο πεδίο του χρόνου ή του χώρου (για την περίπτωση 2-Δ σημάτων) δεν αντιστοιχεί σε ανάλογες μικρές μετατοπίσεις των συντελεστών κυματιδίων. Το γεγονός αυτό περιπλέκει αρκετά την επεξεργασία στο πεδίο των κυματιδίων, αφού θα πρέπει να ληφθεί υπόψιν, από τους αλγορίθμους η ανταλλαγή ενέργειας μεταξύ υποζωνών του κυματιδιακού μετασχηματισμού. Το δεύτερο εξίσου σημαντικό μειονέκτημα που παρουσιάζει είναι το έντονο φαινόμενο της φασματικής αναδίπλωσης. Η φασματική αναδίπλωση, όπως ήδη έχουμε έχει αναφερθεί, οφείλεται στο γεγονός ότι οι κυματιδιακοί συντελεστές υπολογίζονται μέσα από μια αναδρομική διαδικασία η οποία περιλαμβάνει υποδειγματοληψία και φιλτράρισμα με μη ιδανικά βαθυπερατά και υψιπερατά φίλτρα, όπως φαίνεται και στο Σχ. 2.2. Αν και θα πρέπει να σημειώσουμε ότι η φασματική αναδίπλωση εξουδετερώνεται κατά το στάδιο του

αντίστροφου μετασχηματισμού, οπότε και επιτυγχάνεται τέλεια ανακατασκευή, κάτι τέτοιο είναι εφικτό μόνο, εάν οι συντελεστές κλιμάκωσης και λεπτομέρειας παραμείνουν οι ίδιοι. Οποιαδήποτε επεξεργασία των συντελεστών (κατωφλιοποίηση, φιλτράρισμα, κβαντισμός) διαταράσσει την ισορροπία μεταξύ του ευθύ και αντίστροφου μετασχηματισμού με αποτέλεσμα την εμφάνιση τεχνουργημάτων στο αποκατεστημένο σήμα.

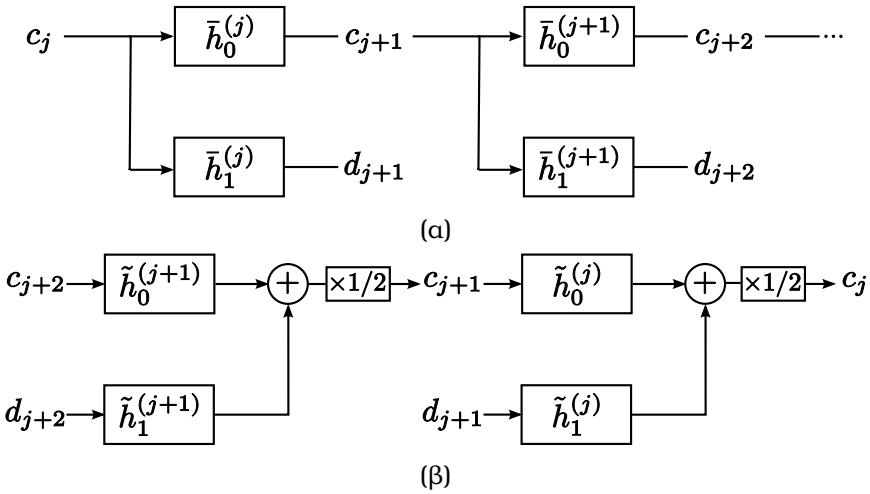
Προκειμένου να αντιμετωπιστούν τα συγκεκριμένα προβλήματα του μετασχηματσμού κυματιδίων, είναι δυνατόν κανείς να σχεδιάσει γραμμικούς τοπικά εντοπισμένους μετασχηματισμούς οι οποίοι εξακολουθούν να βασίζονται στην ιδέα την πολυκλιμακωτής προσέγγισης, ενώ παράλληλα η υλοποίησή τους είναι εφικτή μέσα από μια αναδρομική διαδικασία που περιλαμβάνει φιλτράρισμα με ζωνοπερατά φίλτρα. Αυτό που διαφοροποιεί τους συγκεκριμένους μετασχηματισμούς από το μετασχηματισμό κυματιδίων είναι ότι στην περίπτωσή τους παύει να ισχύει η κρίσιμη δειγματοληψία (όριο Nyquist) και οδηγούμαστε, έτσι, σε υπερπλήρεις αναπαραστάσεις (overcomplete representations), κατά τις οποίες, για παράδειγμα, στην περίπτωση σημάτων πεπερασμένου μήκους ο αριθμός των παραγόμενων συντελεστών του μετασχηματισμού είναι μεγαλύτερος από τον αριθμό δειγμάτων του σήματος. Αυτό σημαίνει ότι ένα σήμα x μπορεί να αναπαρασταθεί σε ένα κατάλληλο χώρο Hilbert $\mathbf{H}$ ως το άθροισμα ατόμων $\{\phi_n\}$ που είναι γραμμικά εξαρτημένα και τα οποία δε συνιστούν πλέον μια βάση, αλλά ένα πλαίσιο (frame) [Dau92, VK95, Mal99]. Όπως είναι φυσικό, η ανακατασκευή του αρχικού σήματος από μια τέτοια αναπαράσταση δεν είναι μοναδική. Η ανακατασκευή του σήματος είναι μοναδική μόνο υπό την έννοια ότι η λύση που επιλέγεται ελαχιστοποιεί τη νόρμα μεταξύ όλων των άλλων δυνατών λύσεων. Η θεωρία των πλαισίων και των υπερπλήρων αναπαραστάσεων βρίσκει ολοένα και μεγαλύτερη απήχηση και συνδέεται άμεσα με την επεξεργασία σημάτων μέσω συστοιχιών φίλτρων [KC07a, KC07b].

Μια αρκετά διαδεδομένη και σχετικά απλή τεχνική κατασκευής υπερπλήρων αναπαραστάσεων που διαθέτει την ιδιότητα της μεταφορικής συμμετρίας βασίζεται στο μη-δειγματοληπτημένο DWT μετασχηματισμό [Mal99]. Για την κατασκευή υπερπλήρων αναπαραστάσεων που διαθέτουν επιθυμητές ιδιότητες για την επεξεργασία εικόνων, όπως είναι η περιστροφική συμμετρία, η αυξημένη κατευθυντικότητα ή η καλύτερη κωδικοποίηση των δομικών στοιχείων μιας εικόνας (γραμμές, καμπύλες, ακμές), έχουν προταθεί και εναλλακτικές τεχνικές, όπως η καθοδηγήσιμη πυραμίδα (steerable pyramid) [FA91, SFAH92, SF95], ο μιγαδικός δίδενδρος μετασχηματισμός κυματιδίων (dual-tree complex wavelet) [Kin98, Kin00, Kin01, SBK05], οι ridglet [CD99a, CD99c], curvelet [CD99b, SCD02, CDDY05] και contourlet [DV05] μετασχηματισμοί. Στη συνέχεια της ενότητας θα ασχοληθούμε μόνο με την περίπτωση του μη-δειγματοληπτημένου DWT μετασχηματισμού (Undecimated DWT - UDWT), αφού, όπως θα δείξουμε σε επόμενη ενότητα, μπορεί να χρησιμοποιηθεί άμεσα και με τους μη-γραμμικούς μετασχηματισμούς που θα παρουσιάσουμε για την αναπαράσταση εικόνων περιορισμένης φωτεινότητας (photon-limited).

Ο UDWT μετασχηματισμός αποτελεί μια παραλλαγή του κρίσιμα δειγματοληπτημένου μετασχηματισμού και υλοποιείται μέσω της επαναληπτικής εφαρμογής μιας συστοιχίας φίλτρων, όπως παρουσιάζεται στο Σχ. 2.5. Η ιδέα πάνω στην οποία βασίζεται είναι η απαλοιφή των σταδίων δειγματοληψίας τα οποία προκαλούν τα φαινόμενα της φασματικής αναδίπλωσης των συχνοτήτων. Αν κανείς εφαρμόσει τον UDWT σε ένα 1-Δ σήμα $c_0(n)$ κάνοντας χρήση μιας συστοιχίας φίλτρων $\{h_0(n), h_1(n)\}$ για J στάδια ανάλυσης, θα οδηγηθεί σε ένα σύνολο συντελεστών $\{\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_J, \boldsymbol{c}_J\}$, όπου $\boldsymbol{d}_j$ είναι οι συντελεστές λεπτομέρειας της j κλίμακας και $\boldsymbol{c}_J$ οι συντελεστές κλιμάκωσης της αδρότερης κλίμακας ανάλυσης. Το πέρασμα από τη μια κλίμακα ανάλυσης (ευκρινέστερη) προς την επόμενη (αδρότερη) επιτυγχάνεται μέσω του αλγορίθμου á trous [HaJMT89, She92, SF07]:

$$c_{j+1}(p) = \sum_n h_0^{(j)}(n-p)c_j(n) = (c_j * \bar{h}_0^{(j)})(p) \tag{2.19α'}$$

$$d_{j+1}(p) = \sum_n h_1^{(j)}(n-p)c_j(n) = (c_j * \bar{h}_1^{(j)})(p)\,. \tag{2.19β'}$$



Σχήμα 2.5: (α) Ο μη-δειγματοληπτημένος DWT μετασχηματισμός (á trous αλγόριθμος) υλοποιείται μέσω της επαναληπτικής εφαρμογής της συστοιχίας φίλτρων, όπου σε κάθε στάδιο (κλίμακα ανάλυσης) γίνεται χρήση των αρχικών φίλτρων με κατάλληλη ένθεση μηδενικών. (β) Το αρχικό σήμα ανακατασκευάζεται μέσω συνελίξεων με τα δυϊκά φίλτρα. Ο πολλαπλασιασμός με τον παράγοντα $1/2$ είναι απαραίτητος για την τέλεια ανακατασκευή.

Τα φίλτρα $h_0^{(j)}$ και $h_1^{(j)}$ αποτελούν υπερδειγματοληπτημένες εκδοχές των φίλτρων h_0 και h_1, όπου

$$h^{(j)}(n) = \begin{cases} h(n) & \text{εάν } n/2^j \text{ ακέραιος,} \\ 0 & \text{αλλού.} \end{cases} \tag{2.20}$$

Ο αλγόριθμος αυτός ονομάστηκε á trous, δηλαδή με τρύπες, ακριβώς από το γεγονός ότι τα φίλτρα που χρησιμοποιούνται σε κάθε στάδιο ανάλυσης αποτελούν εκδοχές των αρχικών φίλτρων με έναν επιπλέον αριθμό παρεμβαλλόμενων μηδενικών μεταξύ των συντελεστών τους, ο οποίος καθορίζεται από την κλίμακα ανάλυσης. Η ανακατασκευή του αρχικού σήματος $c_0(n)$ παρέχεται από την αναδρομική εφαρμογή της παρακάτω εξίσωσης:

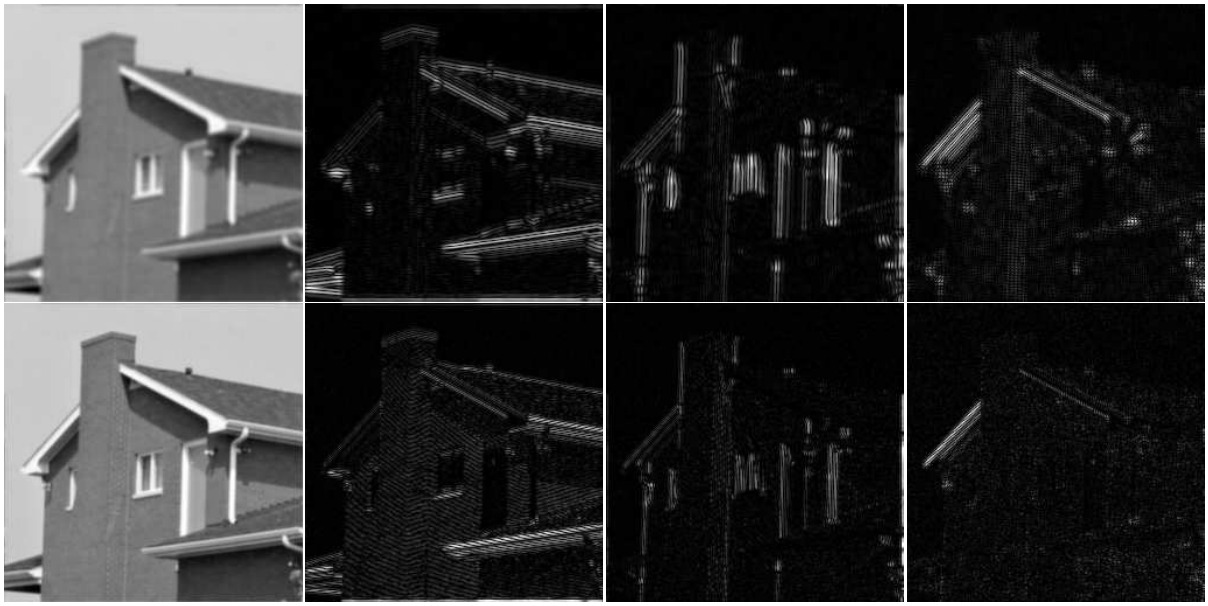
$$\begin{aligned} c_j(p) =& \sum_n \tilde{h}_0^{(j)}(p-n) c_{j+1}(n) + \sum_n \tilde{h}_1^{(j)}(p-n) d_{j+1}(n) \\ =& \frac{1}{2}\left[\left(c_{j+1} * \tilde{h}_0^{(j)}\right)(p) + \left(d_{j+1} * \tilde{h}_1^{(j)}\right)(p)\right] . \end{aligned} \tag{2.21}$$

Επειδή στην περίπτωση του UDWT δεν περιλαμβάνονται πράξεις δειγματοληψίας, το ζεύγος των συστοιχιών φίλτρων ανάλυσης και σύνθεσης του Σχ. 2.5 απαιτείται να ικανοποιεί μόνο τη συνθήκη της μη παραμόρφωσης κατά την ανακατασκευή. Η συνθήκη αυτή προκύπτει χρησιμοποιώντας τη σχέση εισόδου-εξόδου της συστοιχίας στο πεδίο του μετασχηματισμού Z και έχει τη μορφή:

$$H_0(z^{-1})\tilde{H}_0(z) + H_1(z^{-1})\tilde{H}_1(z) = 1 . \tag{2.22}$$

Το γεγονός αυτό παρέχει μεγαλύτερη ελευθερία στο σχεδιασμό των δυϊκών φίλτρων $\left\{\tilde{h}_0, \tilde{h}_1\right\}$.

Ο αλγόριθμος á trous που παρουσιάσαμε αφορά σε 1-Δ σήματα. Ωστόσο, είναι δυνατή η επέκταση του και σε σήματα μεγαλύτερων διαστάσεων μέσα από μια διαχωρίσιμη τεχνική, όπως αυτή που συζητήσαμε και στην Ενοτ. 2.1.5. Για την περίπτωση των εικόνων (2-Δ σήματα) η πολυκλιμακωτή ανάλυση μπορεί να επιτευχθεί μέσω των παρακάτω επαναληπτικών εξισώσεων:



Σχήμα 2.6: Πολυκλιμακωτή ανάλυση της εικόνας house μεγέθους 256×256 με τον UDWT μετασχηματισμό για δύο κλίμακες ανάλυσης. Από αριστερά προς τα δεξιά παρουσιάζονται οι συντελεστές κλιμάκωσης και οι απόλυτες τιμές των συντελεστών λεπτομέρειας για τρεις διαφορετικές κατευθύνσεις (οριζόντια, κατακόρυφη, διαγώνια). Τα επίπεδα ανάλυσης εμφανίζονται με σειρά από κάτω προς τα πάνω, δηλαδή το ευκρινέστερο επίπεδο βρίσκεται στην δεύτερη γραμμή και το αδρότερο στην πρώτη γραμμή.

$$
\begin{aligned}
c_{j+1}(n_1,n_2) &= \left[\bar{h}_0^{(j)}\bar{h}_0^{(j)} * c_j\right](n1,n2) \\
d^1_{j+1}(n_1,n_2) &= \left[\bar{h}_0^{(j)}\bar{h}_1^{(j)} * c_j\right](n1,n2) \\
d^2_{j+1}(n_1,n_2) &= \left[\bar{h}_1^{(j)}\bar{h}_0^{(j)} * c_j\right](n1,n2) \\
d^3_{j+1}(n_1,n_2) &= \left[\bar{h}_1^{(j)}\bar{h}_1^{(j)} * c_j\right](n1,n2)\,,
\end{aligned}
\tag{2.23}
$$

όπου το $h_0h_1 * c$ συμβολίζει τη συνέλιξη του 2-Δ σήματος c από το διαχωρίσιμο φίλτρο h_0h_1, δηλαδή πρώτα συνέλιξη όλων των στηλών του c με το φίλτρο h_0 και στη συνέχεια τη συνέλιξη των γραμμών του αποτελέσματος με το φίλτρο h_1. Ένα παράδειγμα ανάλυσης εικόνας με τον UDWT για δύο επίπεδα ανάλυσης παρουσιάζεται στο Σχ. 2.6.

Στην περίπτωση 1-Δ σημάτων η ανάλυση ενός σήματος μήκους N δειγμάτων για L στάδια ανάλυσης θα μας οδηγήσει σε $M = (L+1)N$ κυματιδιακούς συντελεστές με βαθμό πλεονασμού (redundancy factor) ίσο με $R = (L+1)$. Αντιστοίχως, για την ανάλυση μιας εικόνας μεγέθους $N_1 \times N_2$ ο βαθμός πλεονασμού θα ισούται με $R_2 = 3(L+1)$, αφού σε κάθε κλίμακα διατηρούμε τρία είδη κυματιδιακών συντελεστών.

2.2 Μη-Γραμμική Πολυκλιμακωτή Ανάλυση 1-Δ Σημάτων

Ο μετασχηματισμός κυματιδίων, αλλά και οι υπερπλήρεις εκδοχές τους έχουν γνωρίσει σημαντική απήχηση τις τελευταίες δεκαετίες αποτελώντας τη βάση για την ανάπτυξη μεθόδων και αλγορίθμων που εφαρμόζονται σε μια πληθώρα προβλημάτων και παρέχουν εξαιρετικά αποτελέσματα. Ωστόσο, δεδομένου ότι δεν υπάρχει μια καθολικά ιδανική αναπαράσταση για όλα τα είδη

δεδομένων και για όλα τα προβλήματα, στην παρούσα ενότητα θα παρουσιάσουμε μια κατηγορία μετασχηματισμών που καλούνται πολυκλιμακωτοί πολλαπλασιαστικοί μετασχηματισμοί και εμφανίζουν στενή σχέση με το Haar μετασχηματισμό κυματιδίων. Η συγκεκριμένη κατηγορία μετασχηματισμών, όπως θα δείξουμε και στη συνέχεια, αποτελεί ιδανικό υποψήφιο για την αναπαράσταση σημάτων τα οποία προκύπτουν από Poisson μετρήσεις, δηλαδή σημάτων που μπορούν να μοντελοποιηθούν ως στιγμιότυπα Poisson στοχαστικών διαδικασιών. Παράλληλα, η κατηγορία αυτή διατηρεί πολλές από τις ιδιότητες του μετασχηματισμού κυματιδίων που τον κατέστησαν εξαιρετικά δημοφιλή στην επεξεργασία σημάτων. Όπως θα συζητήσουμε αναλυτικά στο Κεφ. 3, δεδομένα που μοντελοποιούνται με αυτό τον τρόπο παρουσιάζονται σε προβλήματα επεξεργασίας εικόνων περιορισμένης φωτεινότητας (photon-limited imaging). Η αποκατάσταση των εικόνων αυτού του είδους αποτελεί ένα από τα βασικά προβλήματα που καλούμαστε να αντιμετωπίσουμε στην παρούσα διατριβή.

Λόγω της στενής σχέσης του πολυκλιμακωτού πολλαπλασιαστικού μετασχηματισμού με το Haar μετασχηματισμό κυματιδίων, αρχικά θα περιγράψουμε συνοπτικά το δεύτερο, ώστε στη συνέχεια να οδηγηθούμε στη σχέση που τους συνδέει.

2.2.1 Haar Μετασχηματισμός

Ο Haar μετασχηματισμός αποτελεί τον παλαιότερο και ταυτόχρονα τον πιο απλό εκπρόσωπο της οικογένειας των μετασχηματισμών κυματιδίων. Όπως όλοι οι μετασχηματισμοί κυματιδίων έτσι και Haar μετασχηματισμός αναλύει ένα μονοδιάστατο διακριτό σήμα $x(k)$ μήκους $N = 2^J$ σε δύο επιμέρους σήματα $c(k)$ και $d(k)$ που έχουν ακριβώς το μισό μέγεθος του αρχικού. Το πρώτο σήμα προκύπτει ως η αποδεκατισμένη κατά παράγοντα 2 εκδοχή του φιλτραρισμένου σήματος $x(k)$ με ένα βαθυπερατό φίλτρο $\bar{h}_0(k)$, ενώ το δεύτερο σήμα προκύπτει ως η αποδεκατισμένη κατά παράγοντα 2 εκδοχή του φιλτραρισμένου σήματος $x(k)$ με ένα υψιπερατό φίλτρο $\bar{h}_1(k)$. Η διαδικασία αυτή επαναλαμβάνεται για J στάδια και κάθε φορά το σήμα που υπόκειται στο φιλτράρισμα με τη συστοιχία των φίλτρων είναι η έξοδος του βαθυπερατού φίλτρου του προηγούμενο σταδίου. Στην περίπτωση του Haar μετασχηματισμού τα φίλτρα που χρησιμοποιούμε είναι τα $h_0(k) = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$ και $h_1(k) = \left[\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right]$ και μας οδηγούν στο ακόλουθο σύνολο αναδρομικών εξισώσεων:

$$c_j(k) = \frac{c_{j-1}(2k) + c_{j-1}(2k+1)}{\sqrt{2}} \tag{2.24α'}$$

$$d_j(k) = \frac{c_{j-1}(2k) - c_{j-1}(2k+1)}{\sqrt{2}}\,, \tag{2.24β'}$$

για $j = 1, \dots J$, $k = 0, 1, \dots N/2^j - 1$, όπου $c_0(k) = x(k)$. Με αυτή τη διαδικασία το σήμα $x(k)$ μπορεί να αναπαρασταθεί ως ένα διάνυσμα που αποτελείται από τους συντελεστές κλιμάκωσης $c_J(k)$ της πιο αδρής κλίμακας και τους συντελεστές κυματιδίων $d_j(k)$ όλων των κλιμάκων ανάλυσης, $\mathbf{x} = \big[c_J(0) \dots c_J(N/2^J - 1), d_J(n), \dots d_1(0), \dots, d_1(N/2 - 1)\big]$.

Η εφαρμογή του Haar μετασχηματισμού σε 2-Δ σήματα (εικόνες) μπορεί να πραγματοποιηθεί ακολουθώντας τη διαχωρίσιμη στρατηγική της Ενοτ. 2.1.5. Πιο συγκεκριμένα, σε κάθε στάδιο/-κλίμακα εφαρμόζουμε τη μονοδιάστατη εκδοχή του μετασχηματισμού στις γραμμές (ή στις στήλες) της εικόνας παράγοντας με αυτό τον τρόπο δύο εικόνες που έχουν ακριβώς τον ίδιο αριθμό από γραμμές (στήλες) και το μισό αριθμό από στήλες (γραμμές). Στη συνέχεια, σε κάθε μια από τις δύο προκύπτουσες εικόνες εφαρμόζουμε το μετασχηματισμό αυτή τη φορά κατά μήκος των στηλών (γραμμών). Τελικά αυτό που λαμβάνουμε είναι 4 εικόνες που έχουν το μισό αριθμό από γραμμές και στήλες σε σχέση με την αρχική. Η διαδικασία επαναλαμβάνεται στις επόμενες κλίμακες όπου ως εικόνα εισόδου χρησιμοποιούμε την εικόνα που προέκυψε από το βαθυπερατό φιλτράρισμα κατά μήκος των γραμμών και των στηλών του προηγούμενου σταδίου. Με αυτή τη διαδικασία λαμβάνουμε μια εικόνα της οποίας οι τιμές αντιστοιχούν στους συντελεστές κλιμάκωσης και

τρεις εικόνες σε κάθε κλίμακα με τους συντελεστές λεπτομέρειας. Οι αναδρομικές εξισώσεις που περιγράφουν την παραπάνω διαδικασία είναι:

$$\begin{aligned}
c_j(k,\ell) =& \frac{c_{j-1}(2k,2\ell) + c_{j-1}(2k,2\ell+1) + c_{j-1}(2k,2\ell+1) + c_{j-1}(2k+1,2\ell+1)}{2} \\
d_j^h(k,\ell) =& \frac{[c_{j-1}(2k,2\ell) + c_{j-1}(2k,2\ell+1)] - [c_{j-1}(2k,2\ell+1) + c_{j-1}(2k+1,2\ell+1)]}{2} \\
d_j^v(k,\ell) =& \frac{[c_{j-1}(2k,2\ell) + c_{j-1}(2k+1,2\ell)] - [c_{j-1}(2k+1,2\ell) + c_{j-1}(2k+1,2\ell+1)]}{2} \\
d_j^d(k,\ell) =& \frac{[c_{j-1}(2k,2\ell) + c_{j-1}(2k+1,2\ell+1)] - [c_{j-1}(2k,2\ell+1) + c_{j-1}(2k+1,2\ell)]}{2} .
\end{aligned} \tag{2.25}$$

Επειδή στην 2-Δ περίπτωση οι συντελεστές λεπτομέρειας, d_j, προκύπτουν ως οι διαφορές των συντελεστών κλιμάκωσης της προηγούμενης κλίμακας για τρεις διαφορετικές κατευθύνσεις και μπορούν συνεπώς να εντοπίσουν ακμές στην οριζόντια, κατακόρυφη και διαγώνια διεύθυνση, χρησιμοποιούμε τον υπερδείκτη $o = \{h, v, d\}$, προκειμένου να μπορούμε να αναφερόμαστε στο κάθε είδος ξεχωριστά.

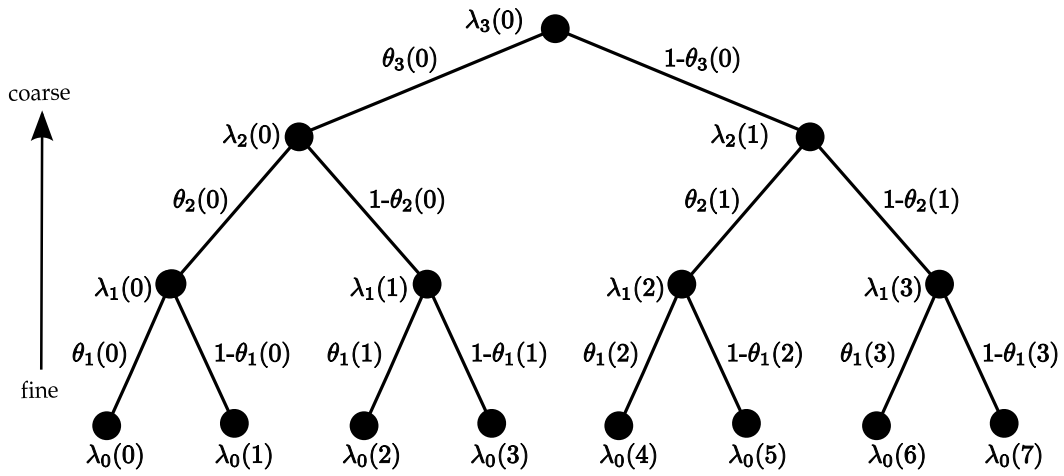
2.2.2 Πολλαπλασιαστική Πολυκλιμακωτή Ανάλυση 1-Δ Σημάτων

Όπως και ο Haar μετασχηματισμός, έτσι και η πολλαπλασιαστική ανάλυση [TN99, Kol99a, RRCB00] οδηγεί σε πολυκλιμακωτή αναπαράσταση 1-Δ σημάτων. Η πολυκλιμακωτή αναπαράσταση ενός σήματος $x(k)$ είναι εφικτή μέσω των παρακάτω αναδρομικών εξισώσεων:

$$\lambda_j(k) = \lambda_{j-1}(2k) + \lambda_{j-1}(2k+1) \tag{2.26α'}$$

$$\theta_j(k) = \frac{\lambda_{j-1}(2k)}{\lambda_{j-1}(2k) + \lambda_{j-1}(2k+1)} , \tag{2.26β'}$$

για $j = 1, \dots J$, $k = 0, 1, \dots N/2^j - 1$, όπου $\lambda_0(k) = x(k)$. Οι συντελεστές $\lambda_j(k)$ προκύπτουν μέσω της υποδιαίρεσης του σήματος σε δυαδικά σύνολα σε κάθε κλίμακα και, κατόπιν, του υπολογισμού της συνολικής τιμής κάθε συνόλου. Λόγω του τρόπου αυτού υπολογισμού τους, είναι βολικό να αναπαραστήσουμε τους συντελεστές λ_j ως τους κόμβους ενός δυαδικού δένδρου, όπως φαίνεται στο Σχ. 2.7. Αθροίζοντας τους συντελεστές λ_{j-1} που είναι τοποθετημένοι στους κόμβους "παιδιά", μπορούμε να υπολογίσουμε το συντελεστή λ_j του κόμβου "πατέρα". Τα κλαδιά του δένδρου, δηλαδή οι κόμβοι που βρίσκονται στην κλίμακα 0, αναπαριστούν το αρχικό μας σήμα. Οι συντελεστές $\theta_j(k)$ προκύπτουν ως το πηλίκο του αριστερού παιδιού με τον κόμβο πατέρα και μπορούν να συσχετιστούν με τις ακμές που συνδέουν τους κόμβους του δυαδικού δένδρου. Έτσι, προκειμένου να υπολογίσουμε την τιμή του κόμβου $\lambda_{j-1}(2k)$, που αποτελεί το αριστερό παιδί του κόμβου $\lambda_j(k)$, αρκεί να πολλαπλασιάσουμε την τιμή του κόμβου πατέρα με την τιμή $\theta_j(k)$ που βρίσκεται στην ακμή που συνδέει τους δύο κόμβους. Αντιστοίχως, για να υπολογίσουμε την τιμή του κόμβου $\lambda_{j-1}(2k)$ που αποτελεί το δεξιό παιδί του κόμβου $\lambda_j(k)$, θα πρέπει να πολλαπλασιάσουμε την τιμή του πατέρα με τον συμπληρωματικό συντελεστή $1-\theta_j(k)$ που βρίσκεται στην ακμή που συνδέει τους δύο κόμβους του δένδρου. Τέλος, για να υπολογίσουμε την τιμή ενός κλαδιού του δένδρου $\lambda_0(k)$, αρκεί να χρησιμοποιήσουμε την τιμή ενός κόμβου $\lambda_j(k)$ που αποτελεί πρόγονο του $\lambda_0(k)$ και να τον πολλαπλασιάσουμε με όλες τις τιμές $\theta_j(k)$ (ή/και $1-\theta_j(k)$) οι οποίες σχετίζονται με τις ακμές που βρίσκονται στο μονοπάτι το οποίο συνδέει τους δύο κόμβους. Διαπιστώνουμε, λοιπόν, ότι μέσω αυτής της πολυκλιμακωτής αναπαράστασης και της διαδικασίας που μόλις περιγράψαμε είναι δυνατή η επιστροφή στο αρχικό πεδίο στο οποίο ορίζεται το σήμα μας, δηλαδή, είναι εφικτή η τέλεια ανακατασκευή, αρκεί να έχουμε στη διάθεσή μας τους συντελεστές $\lambda_J(k)$ της πιο αδρής κλίμακας και τους συντελεστές λόγου $\theta_j(k)$ όλων των κλιμάκων ανάλυσης.



Σχήμα 2.7: Δυαδικό δένδρο παραγόμενο από πολυκλιμακωτή (ευκρινή-αδρή) ανάλυση του $\boldsymbol{\lambda}$ σε 3 κλίμακες. Γνωρίζοντας τις τιμές των παραγόντων διαχωρισμού $\theta_j(k)$ και την συνολική τιμή του σήματος, $\lambda_J(0)$, μπορούμε να ανακτήσουμε επακριβώς την αρχική αναπαράσταση του $\boldsymbol{\lambda}$.

Θα πρέπει να σημειώσουμε ότι και στην περίπτωση του Haar μετασχηματισμού, αλλά και γενικά των μετασχηματισμών κυματιδίων, είναι δυνατή η αναπαράσταση των συντελεστών κλιμάκωσης $c_j(k)$ ως κόμβων ενός δυαδικού δένδρου και η συσχέτιση των συντελεστών $d_j(k)$ με τις ακμές που συνδέουν τους κόμβους του δένδρου. Αυτό που διαφοροποιεί τα δύο είδη μετασχηματισμών είναι ότι στη δεύτερη περίπτωση, προκειμένου να "κατέβουμε" το δένδρο, προσθέτουμε στους κόμβους και δεν πολλαπλασιάζουμε με αυτούς τις τιμές που σχετίζονται με τις ακμές. Έτσι, ενώ στον Haar μετασχηματισμό, καθώς προωθούμαστε από την πιο ευκρινή κλίμακα ανάλυσης προς αδρότερες κλίμακες, τα λεπτομερή χαρακτηριστικά του σήματος κωδικοποιούνται με γραμμικό τρόπο στους συντελεστές λεπτομέρειας $\mathbf{d}_j$, στην περίπτωση της πολλαπλασιαστικής ανάλυσης τα χαρακτηριστικά αυτά κωδικοποιούνται με μη-γραμμικό τρόπο μέσω των συντελεστών λόγου $\boldsymbol{\theta}_j$. Οι συντελεστές αυτοί μπορούν να ερμηνευθούν ως παράγοντες διαχωρισμού των συντελεστών που εμπεριέχουν τα πιο αδρά χαρακτηριστικά του σήματος [Now99a].

Παρόλο που ο Haar μετασχηματισμός κυματιδίων και ο πολλαπλασιαστικός μετασχηματισμός περιγράφουν με διαφορετικό τρόπο την πληροφορία για τη λεπτομερή δομή του σήματος, υπάρχει μια ξεκάθαρη σχέση που συνδέει την αναπαράσταση του διακριτού σήματος $x(k)$, στο πολλαπλασιαστικό πεδίο, $[\lambda_J(0), \theta_J(0), \ldots, \theta_1(0), \ldots, \theta_1(N/2-1)]$, με την αναπαράσταση του στο πεδίο των κυματιδίων, $[c_J(0), d_J(0), \ldots, d_1(0), \ldots, d_1(N/2-1)]$. Συγκεκριμένα, είναι εύκολο να δείξει κανείς ότι οι συντελεστές $\lambda_j(k)$ συνιστούν ανακλιμακούμενες εκδοχές των Haar συντελεστών κλιμάκωσης $c_j(k)$, ενώ παράλληλα οι συντελεστές $\theta_j(k)$ αποτελούν διαιρετές κανονικοποιημένες εκδοχές των Haar συντελεστών λεπτομέρειας $d_j(k)$, επιπλέον μετατοπισμένες κατά 0.5,

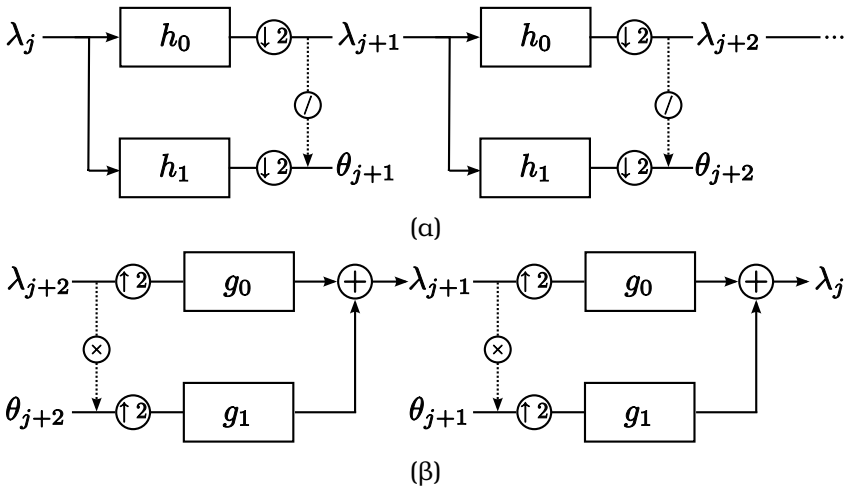
$$\lambda_j(k) = 2^{j/2} c_j(k) \tag{2.27α'}$$

$$\theta_j(k) = \frac{d_j(k)}{2c_j(k)} + \frac{1}{2}\,. \tag{2.27β'}$$

Η σχέση (2.27α') επαληθεύεται απευθείας μέσω της σύγκρισης των Εξ. (2.24α') και(2.26α'), όπου παρατηρούμε ότι η μόνη διαφορά τους είναι ο επιπλέον όρος $2^{-1/2}$ που εμφανίζεται στην πρώτη. Τώρα λόγω των (2.27α') και (2.24β'), η Εξ. (2.26β') μπορεί να γραφεί ως:

$$\theta_j(k) = \frac{c_{j-1}(2k)}{2^{1/2} c_j(k)} = \frac{2^{-1/2}\left(d_j(k) + c_j(k)\right)}{2^{1/2} c_j(k)} = \frac{d_j(k)}{2c_j(k)} + \frac{1}{2}\,, \tag{2.28}$$

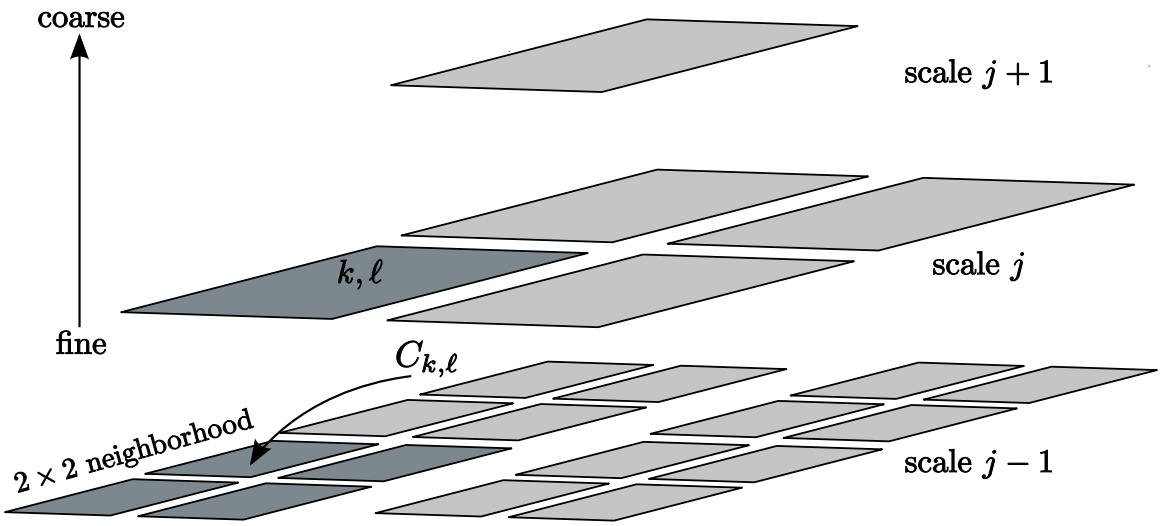
οπότε επαληθεύεται και η σχέση (2.27β').



Σχήμα 2.8: (α) Πολυκλιμακωτή αναπαράσταση ενός 1-Δ σήματος στο πεδίο του πολλαπλασιαστικού μετασχηματισμού μέσω επαναληπτικής ανάλυσης του σε υποζώνες με τη χρήση συστοιχίας φίλτρων. Σε κάθε στάδιο ανάλυσης αφού υποδειγματοληπτηθούν οι έξοδοι των φίλτρων h_0, h_1, η υποδειγματοληπτημένη έξοδος του φίλτρου h_1 κανονικοποιείται από την υποδειγματοληπτημένη έξοδο του φίλτρου h_0. (β) Τέλεια ανακατασκευή του σήματος ακολουθώντας αντίστροφη διαδικασία φιλτραρίσματος.

Το γεγονός της σύνδεσης των δύο αυτών μετασχηματισμών υποδηλώνει ότι βασικές ιδιότητες του μετασχηματισμού κυματιδίων, όπως οι αραιές αναπαραστάσεις, η κατά προσέγγιση αποσυσχέτιση των συντελεστών στο πεδίο του μετασχηματισμού, αλλά και η αποδοτική επεξεργασία των σημάτων μέσω αναδρομών (αδρή προς ευκρινή κλίμακα) κληρονομούνται και από τον πολλαπλασιαστικό μετασχηματισμό. Επίσης, είναι δυνατόν να συνδέσουμε την αναπαράσταση ενός σήματος σε πολλαπλές κλίμακες μέσω του πολλαπλασιαστικού μετασχηματισμού με την επαναληπτική ανάλυση του σήματος σε υποζώνες με τη χρήση συστοιχίας φίλτρων. Στο Σχ. 2.8(α) παρουσιάζεται η διαδικασία φιλτραρίσματος ενός σήματος μέσω συστοιχίας φίλτρων η οποία οδηγεί στην πολυκλιμακωτή αναπαράσταση του σήματος στο πεδίο του πολλαπλασιαστικού μετασχηματισμού. Τα φίλτρα που χρησιμοποιούνται κατά το στάδιο της ανάλυσης είναι τα $h_0(k) = [1\,1]$ και $h_1(k) = [0\,1]$. Προκειμένου να επιστρέψουμε στην αρχική μας αναπαράστασης ακολουθούμε την αντίστροφη διαδικασία όπως παρουσιάζεται στο Σχ. 2.8(β) με τα φίλτρα σύνθεσης να είναι ίσα με $g_0(k) = [0\,1]$ και $g_1(k) = [1\ -1]$.

Ο πολυκλιμακωτός πολλαπλασιαστικός μετασχηματισμός διαθέτει κάποιες ιδιότητες που τον καθιστούν καταλληλότερο από τον Haar μετασχηματισμό για την αναπαράσταση σημάτων σε συγκεκριμένες περιπτώσεις. Για παράδειγμα, όταν τα σήματα ενδιαφέροντος έχουν θετικές τιμές (π.χ. εικόνες), θα επιθυμούσε κανείς να επιβάλλει έναν περιορισμό στην επεξεργασία τους, ώστε το αποτέλεσμα που προκύπτει να διαθέτει μόνο θετικές τιμές. Ο συγκεκριμένος περιορισμός ικανοποιείται με φυσικό τρόπο στο πεδίο του πολλαπλασιαστικού μετασχηματισμού· εφόσον όλοι οι συντελεστές λόγου θ_j έχουν τιμές μεταξύ 0 και 1, το αποτέλεσμα θα είναι πάντα θετικό. Κάτι τέτοιο δεν είναι αληθές για τον Haar μετασχηματισμό ούτε γενικά για τους μετασχηματισμούς κυματιδίων, αφού η κατωφλιοποίηση των συντελεστών ενός θετικού σήματος μπορεί στην ανακατασκευή να οδηγήσει σε αρνητικές τιμές. Επιπλέον, η πολλαπλασιαστική ανάλυση μπορεί να χρησιμοποιηθεί για την μοντελοποίηση Poisson διαδικασιών. Κάτι τέτοιο είναι πολύ χρήσιμο, όταν αναλύουμε εικόνες οι οποίες έχουν συλληφθεί σε συνθήκες χαμηλού φωτισμού από συσκευές απεικόνισης που ανιχνέουν τον αριθμό των φωτονίων της φωτεινής πηγής. Σε αυτές τις περιπτώσεις οι εικόνες παρουσιάζουν αλλοιώσεις από κβαντικό θόρυβο ο οποίος διαθέτει Poisson στατιστικά χαρακτηρι-



Σχήμα 2.9: Απεικόνιση της γειτονιάς εικονοστοιχιων $C_{k,l}$ μεγέθους 2×2 σε μια κλίμακα $j-1$, η οποία συνδέεται με ένα εικονοστοιχείο της εικόνας με συντεταγμένες (k, ℓ) σε μια κλίμακα j.

στικά. Εφόσον ο πολλαπλασιαστικός μετασχηματισμός, όπως θα δούμε και αναλυτικά στο Κεφ. 3, έχει την δυνατότητα να "αποσυνδέει" το σήμα από το θόρυβο και, επιπλέον, τα σήματα μπορούν να μοντελοποιηθούν με αποδοτικό τρόπο στο συγκεκριμένο πεδίο, είναι δυνατόν να επιλύσουμε προβλήματα αποθορυβοποίησης/εκτίμησης κάνοντας χρήση εξαιρετικά δημοφιλών και ισχυρών Μπεϋζιανών (Bayesian) τεχνικών.

2.3 Μη-Γραμμική Πολυκλιμακωτή Ανάλυση Εικόνων

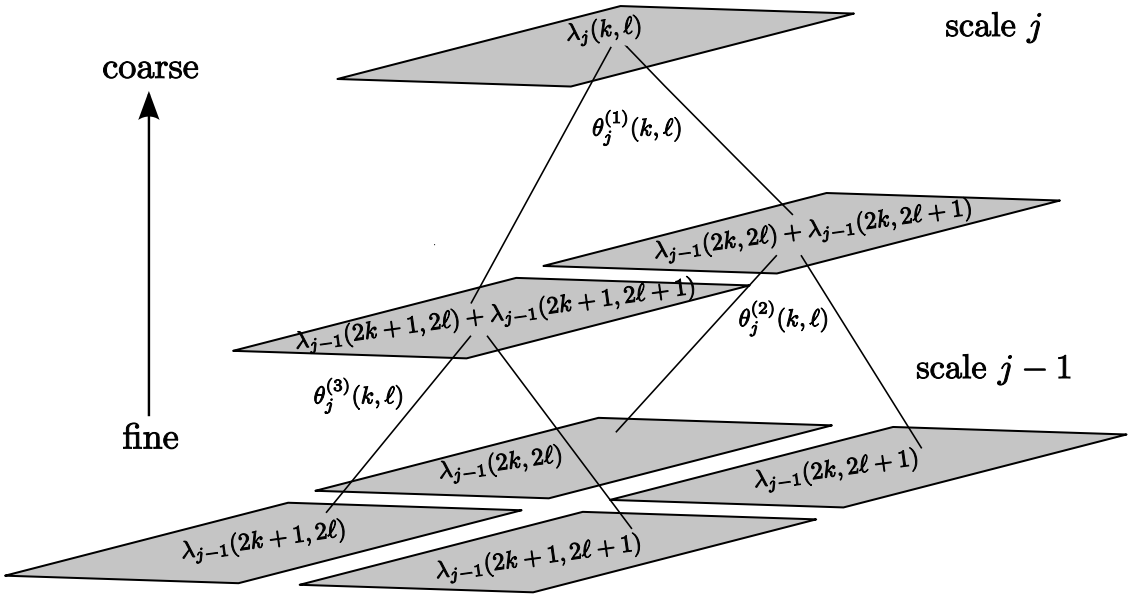
Η επέκταση του 1-Δ πολλαπλασιαστικού μετασχηματισμού για την αναπαράσταση σημάτων υψηλότερων διαστάσεων δεν είναι το ίδιο προφανής με την επέκταση του μετασχηματισμού κυματιδίων μέσω διαχωρίσιμων στρατηγικών σχεδίασης. Σε αυτή την ενότητα θα παρουσιάσουμε δύο στρατηγικές επέκτασης που έχουν προταθεί και αναφέρονται στην βιβλιογραφία. Επιπλέον, επειδή οι συγκεκριμένες επεκτάσεις εμφανίζουν το μειονέκτημα ότι δε διατηρούν την άμεση σχέση του 1-Δ μετασχηματισμού με τον Haar μετασχηματισμό κυματιδίων, θα παρουσιάσουμε μια νέα 2-Δ έκδοση που αντιμετωπίζει το συγκεκριμένο πρόβλημα.

2.3.1 Διαχωρίσιμος 2-Δ Πολλαπλασιαστικός Μετασχηματισμός

Έστω $\boldsymbol{\lambda}$ ένας 2-Δ πίνακας ο οποίος αποτελείται από τις τιμές έντασης φωτεινότητας μιας εικόνας για ένα διακριτό πλέγμα σημείων μεγέθους $N_1 \times N_2$. Ένας τρόπος για να επεκτείνουμε την πολυκλιμακωτή ανάλυση του πολλαπλασιαστικού μετασχηματισμού σε εικόνες είναι να εφαρμόσουμε τη μονοδιάστατη έκδοση με "διαχωρίσιμο" τρόπο όπως προτείνεται στα [TN99, Now99a]. Αν για κάθε εικονοστοιχείο της εικόνας που έχει συντεταγμένες (k, ℓ) σε μια κλίμακα j, ορίσουμε ως

$$C_{k,l} = \{(2k, 2\ell), (2k, 2\ell+1), (2k+1, 2\ell), (2k+1, 2\ell+1)\} \ , \tag{2.29}$$

το 2×2 σύνολο συντεταγμένων των εικονοστοιχείων τα οποία βρίσκονται στην ακριβώς επόμενη πιο ευκρινή κλίμακα, δες Σχ. 2.9, τότε οι αντίστοιχες σχέσεις ανάλυσης με αυτές της μοναδιάστατης



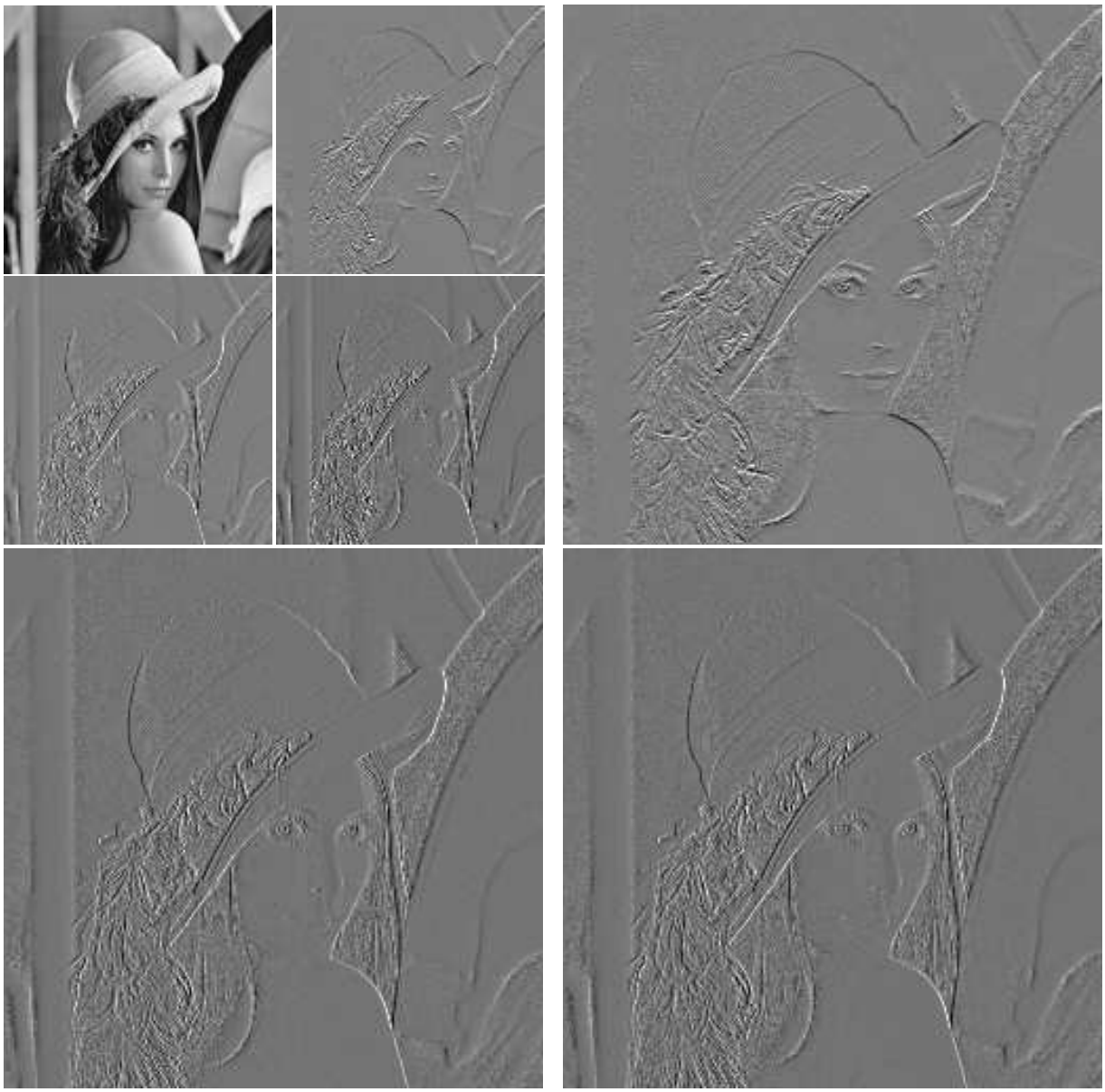
Σχήμα 2.10: Πολυκλιμακωτή ανάλυση 2-Δ εικόνας χρησιμοποιώντας το διαχωρίσιμο σχήμα παραγοντοποίησης των [TN99], [Now99a].

περίπτωσης (2.26) θα είναι:

$$\begin{aligned}
\lambda_j(k,\ell) &= \sum_{(k',\ell')\in C_{k,\ell}} \lambda_{j-1}(k',\ell') \\
\theta_j^{(1)}(k,\ell) &= \frac{\lambda_{j-1}(2k,2\ell)+\lambda_{j-1}(2k,2\ell+1)}{\lambda_j(k,\ell)} \\
\theta_j^{(2)}(k,\ell) &= \frac{\lambda_{j-1}(2k,2\ell)}{\lambda_{j-1}(2k,2\ell)+\lambda_{j-1}(2k,2\ell+1)} \\
\theta_j^{(3)}(k,\ell) &= \frac{\lambda_{j-1}(2k+1,2\ell)}{\lambda_{j-1}(2k+1,2\ell)+\lambda_{j-1}(2k+1,2\ell+1)}\,,
\end{aligned} \tag{2.30}$$

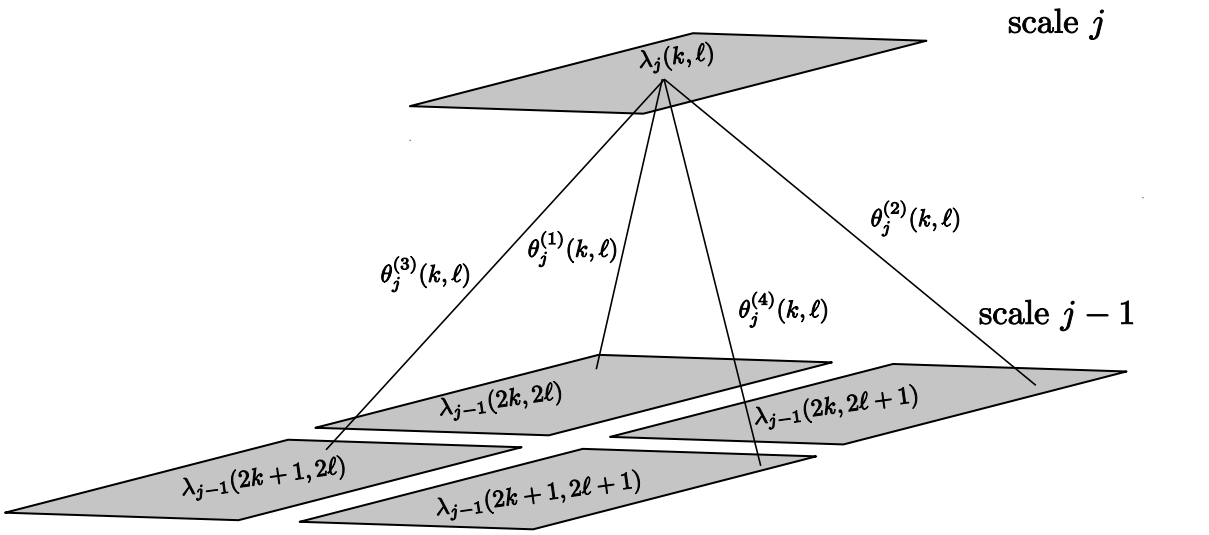
για $j=1,\dots J$, $k=0,\dots,N_1/2^j-1$, $\ell=0,\dots,N_2/2^j-1$ και $J=\min\{\log_2(N_1),\log_2(N_2)\}$. Κατά αναλογία με τη μονοδιάστατη περίπτωση της Ενοτ. 2.2.2, οι συντελεστές κλιμάκωσης $\lambda_j(k,\ell)$ προκύπτουν μέσω της υποδιαίρεσης της εικόνας σε τετράγωνα σε κάθε κλίμακα και, κατόπιν, του υπολογισμού της συνολικής τιμής κάθε τετραγώνου. Λόγω αυτού του τρόπου υπολογισμού μπορούμε να αναπαραστήσουμε τους συντελεστές $\lambda_j(k,\ell)$ ως τους κόμβους ενός δυαδικού δένδρου με ένα ενδιάμεσο επίπεδο, όπως φαίνεται στο Σχ. 2.10. Εδώ θα πρέπει να σημειώσουμε ότι δεν αναθέτουμε σε όλους τους κόμβους του δένδρου συντελεστές κλιμάκωσης. Συγκεκριμένα, στα ενδιάμεσα επίπεδα του δένδρου οι κόμβοι έχουν τιμές που αντιστοιχούν στο άθροισμα δυο γειτονικών συντελεστών κλιμάκωσης (δεξιός/αριστερός γείτονας). Τα κλαδιά του δένδρου αντιστοιχούν στην αρχική αναπαράσταση της εικόνας. Ο υπολογισμός των συντελεστών διαχωρισμού δεν είναι ακριβώς ανάλογος με αυτόν του μονοδιάστατου σχήματος. Για να υπολογίσουμε τα $\theta_j(k,\ell)$ χωρίζουμε τα τετράγωνα σε πάνω και κάτω μισά μέρη και στη συνέχεια το κάθε ένα από αυτά σε δεξιά και αριστερά μέρη, όπως φαίνεται στο Σχ. 2.10. Στη συνέχεια οι συντελεστές $\theta_j(k)$ προκύπτουν ως το πηλίκο του αριστερού παιδιού με τον κόμβο πατέρα και μπορούν να συσχετιστούν με τις ακμές που συνδέουν τους κόμβους του ειδικής μορφής δυαδικού δένδρου. Μέσω αυτής της αναπαράστασης για να υπολογίσουμε την τιμή ενός από τους κόμβους $\lambda_{j-1}(k',\ell')$ ή την τιμή ενός κλαδιού του δένδρου $\lambda_0(k,\ell)$μ αρκεί να χρησιμοποιήσουμε την τιμή ενός κόμβου $\lambda_j(k,\ell)$, ο οποίος αποτελεί

πρόγονο τους, και να τον πολλαπλασιάσουμε με όλες τις τιμές $\theta_j(k,\ell)$ (ή/και $1-\theta_j(k,\ell)$) οι οποίες σχετίζονται με τις ακμές που βρίσκονται στο μονοπάτι το οποίο συνδέει τους δύο κόμβους.



Σχήμα 2.11: Τιμή των συντελεστών που προκύπτουν από την ανάλυση της 512×512 εικόνας Lena μέσω του 2-Δ διαχωρίσιμου σχήματος για 2 επίπεδα ανάλυσης. Πάνω δεξιά αντιστοιχούν οι $\theta_j^{(1)}(k,\ell)$ συντελεστές, κάτω αριστερά οι $\theta_j^{(2)}(k,\ell)$ συντελεστές και κάτω δεξιά οι $\theta_j^{(3)}(k,\ell)$ συντελεστές.

Από εδώ και στο εξής θα αναφερόμαστε στη συγκεκριμένη 2-Δ πολυκλιμακωτή παραγοντοποίηση με τον όρο *διαχωρίσιμο* σχήμα. Σημειώνουμε, ότι όπως προκύπτει και από τις Εξ. (2.30) το διαχωρίσιμο σχήμα για μια 2×2 γειτονιά σε μια κλίμακα $j-1$, παράγει τους συντελεστές διαχωρισμού $\theta_j^{(i)}(k,\ell)$ για $i = 1,\ldots,3$, οι οποίοι ανήκουν σε τρεις υποζώνες. Οι συντελεστές θ αυτών των υποζωνών σε αντίθεση με την μονοδιάστατη περίπτωση δεν αντιστοιχίζονται 1-1 με τους συντελεστές λεπτομέρειας του 2-Δ Haar μετασχηματισμού. Στο Σχ. 2.11 απεικονίζονται οι 2-Δ συντελεστές $\lambda_j(k,\ell)$ και $\theta_j^{(i)}(k,\ell)$ που προκύπτουν από την ανάλυση μιας τυπικής εικόνας με το



Σχήμα 2.12: Πολυκλιμακωτή ανάλυση 2-Δ εικόνας χρησιμοποιώντας το τετραδικό σχήμα ανάλυσης των [Now99a, ECKVD04].

διαχωρίσιμο σχήμα.

2.3.2 Τετραδικός 2-Δ Πολλαπλασιαστικός Μετασχηματισμός

Εκτός από το διαχωρίσιμο σχήμα μελετούμε μια 2-Δ μη-διαχωρίσιμη πολυκλιμακωτή ανάλυση τετραδικής δενδρικής μορφής για την εικόνα $\boldsymbol{\lambda}$ διαστάσεων $N_1 \times N_2$ [Now99a, ECKVD04]. Σε αυτή την περίπτωση η πολυκλιμακωτή ανάλυση της εικόνας πραγματοποιείται μέσω των παρακάτω αναδρομικών εξισώσεων:

$$\begin{aligned}
\lambda_j(k,\ell) &= \sum_{(k',\ell')\in C_{k,\ell}} \lambda_{j-1}(k',\ell') \\
\theta_j^{(1)}(k,\ell) &= \frac{\lambda_{j-1}(2k,2\ell)}{\lambda_j(k,\ell)} \\
\theta_j^{(2)}(k,\ell) &= \frac{\lambda_{j-1}(2k,2\ell+1)}{\lambda_j(k,\ell)} \\
\theta_j^{(3)}(k,\ell) &= \frac{\lambda_{j-1}(2k+1,2\ell)}{\lambda_j(k,\ell)}\,,
\end{aligned} \tag{2.31}$$

για $j = 1, \ldots J$, $k = 0, \ldots, N_1/2^j - 1$, $\ell = 0, \ldots, N_2/2^j - 1$ και $J = \min\{\log_2(N_1), \log_2(N_2)\}$. Συγκρίνοντας τις Εξ. (2.30) και (2.31) διαπιστώνουμε ότι οι συντελεστές κλιμάκωσης υπολογίζονται με τον ίδιο ακριβώς τρόπο, ενώ η διαφορά των δύο μετασχηματισμών παρουσιάζεται στον τρόπο υπολογισμού των συντελεστών διαχωρισμού. Η διαδικασία υπολογισμού των συντελεστών κλιμάκωσης και των συντελεστών διαχωρισμού με βάση τις Εξ. (2.31) μας επιτρέπει να αναπαραστήσουμε τους συντελεστές $\lambda_j(k,\ell)$ ως τους κόμβους ενός τετραδικού δένδρου, όπως φαίνεται στο Σχ. 2.12. Αθροίζοντας τους συντελεστές $\lambda_{j-1}(k',\ell')$ που είναι τοποθετημένοι στους κόμβους "παιδιά" μπορούμε να υπολογίσουμε το συντελεστή λ_j του κόμβου "πατέρα". Οι συντελεστές $\theta_j(k,\ell)$ προκύπτουν όμοια με το 1-Δ μετασχηματισμό, δηλαδή ως το πηλίκο του κόμβου παιδιού με τον κόμβο πατέρα και μπορούν να συσχετιστούν με τις ακμές που συνδέουν τους κόμβους του

τετραδικού δένδρου. Επειδή ο συντελεστής διαχωρισμού, ο οποίος προκύπτει από τη διαίρεση του κόμβου πατέρα με το τέταρτο παιδί, συνδέεται με τους υπόλοιπους τρεις μέσω της σχέσης:

$$\theta_j^{(4)}(k,\ell) = 1 - \theta_j^{(1)}(k,\ell) - \theta_j^{(2)}(k,\ell) - \theta_j^{(3)}(k,\ell)\,,$$

είναι πλεονάζων και, επομένως, δεν απαιτείται ο υπολογισμός του. Για να "κατέβουμε" το δένδρο και να οδηγηθούμε στο αρχικό πεδίο αναπαράστασης, ακολουθούμε την ίδια ακριβώς διαδικασία, όπως και με το 1-Δ και το 2-Δ διαχωρίσιμο πολλαπλασιαστικό μετασχηματισμό. Από εδώ και στο εξής θα αναφερόμαστε στη συγκεκριμένη 2-Δ πολυκλιμακωτή παραγοντοποίηση με τον όρο *τετραδικό* σχήμα.

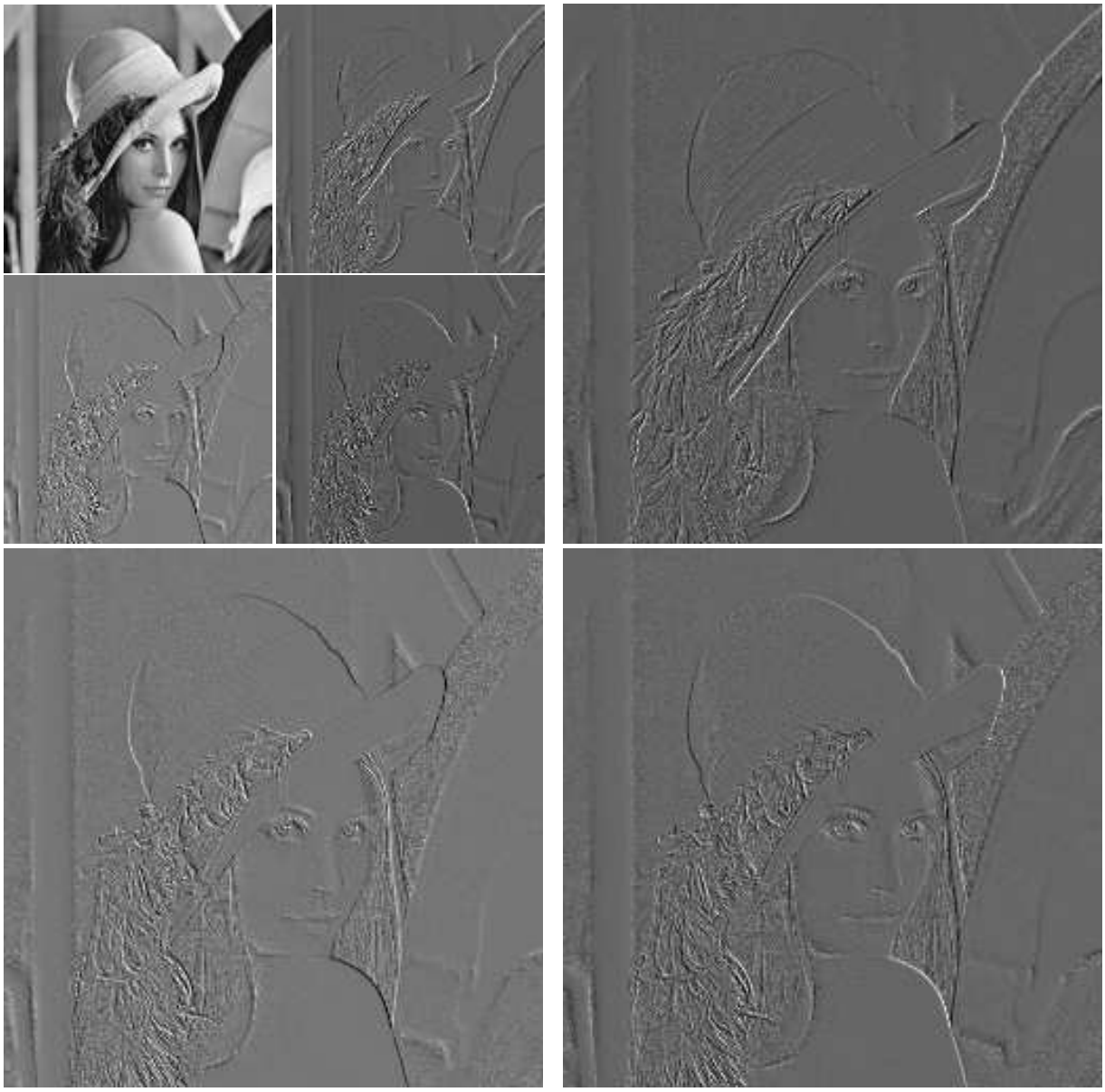
Παρόλη την άμεση αντιστοιχία υπολογισμού των συντελεστών διαχωρισμού με το 1-Δ μετασχηματισμό, ούτε και σε αυτή την περίπτωση οι συντελεστές θ δεν αντιστοιχίζονται 1-1 με τους συντελεστές λεπτομέρειας του 2-Δ Haar μετασχηματισμού. Προκειμένου να είναι δυνατή η σύγκριση του 2-Δ διαχωρίσιμου και του τετραδικού μετασχηματισμού πολυκλιμακωτής αναπαράστασης, στο Σχ. 2.11 απεικονίζονται οι 2-Δ συντελεστές $\lambda_j(k,\ell)$ και $\theta_j^{(i)}(k,\ell)$ οι οποίοι προκύπτουν από την ανάλυση της ίδιας τυπικής εικόνας (Lena) αυτή τη φορά με τη χρήση του τετραδικού πολλαπλασιαστικού μετασχηματισμού.

2.3.3 Poisson-Haar Μετασχηματισμός

Όπως είδαμε στις Ενοτ. 2.3.1 και 2.3.2 τόσο το διαχωρίσιμο όσο και το τετραδικό σχήμα είναι ικανά να παρέχουν κατάλληλες πολυκλιμακωτές αναπαραστάσεις για φυσικές εικόνες. Οι αναπαραστάσεις αυτές επιτυγχάνονται με την αναπαραμετροποίηση των τιμών έντασης φωτεινότητας της πηγαίας εικόνας σε ένα διάνυσμα $\boldsymbol{\lambda}_e$, το οποίο συνίσταται από ένα μοναδικό σημείο $\lambda_J(0,0)$ που εμπεριέχει τη συνολική ένταση της εικόνας και από το διάνυσμα του λόγου των τιμών έντασης $\boldsymbol{\theta}$. Τα στοιχεία του διανύσματος $\boldsymbol{\theta}$ ορίζονται σε κάθε περίπτωση ως οι λόγοι των τιμών των εντάσεων σε γειτονικές κλίμακες· δες την Εξ. (2.30) για το διαχωρίσιμο και την Εξ. (2.31) για το τετραδικό σχήμα. Αυτοί οι λόγοι εντάσεων μπορούν να ερμηνευθούν ως παράγοντες διαχωρισμού της έντασης και αντιστοιχούν στο ποσοστό της έντασης του κόμβου-πατέρα, μιας αδρής κλίμακας, το οποίο διαμοιράζεται σε κάθε κόμβο-παιδί της αμέσως επόμενης πιο ευκρινούς κλίμακας. Όπως ήδη έχουμε συζητήσει στην Ενότητα 2.2.2 οι παράγοντες διαχωρισμού είναι στενά συνδεδεμένοι με τους Haar συντελεστές λεπτομέρειας.

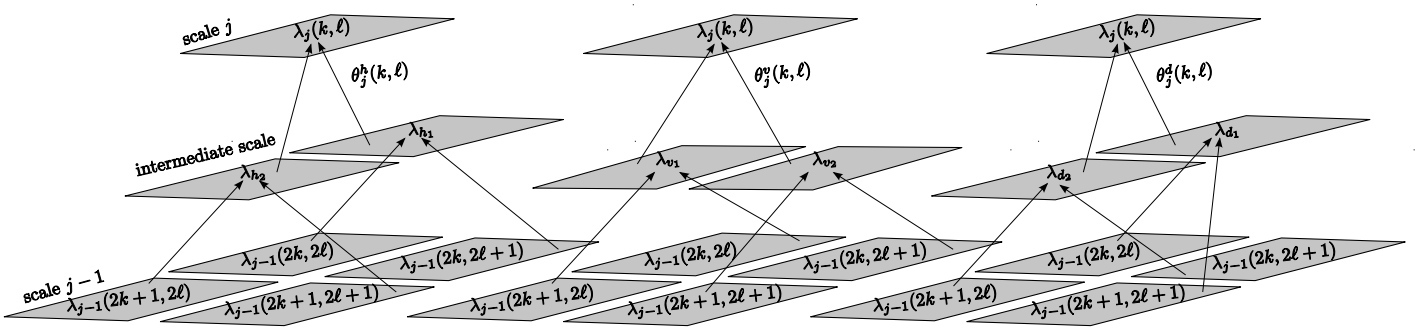
Αυτή η στενή σχέση υποδηλώνει ότι μπορούμε να ερμηνεύσουμε τα $\theta_j(k,\ell)$ ως τους συντελεστές που κωδικοποιούν την πληροφορία για τις ακμές της εικόνας στην κλίμακα j. Για το διαχωρίσιμο σχήμα του Σχ. 2.10, όπου οι μεταβλητές του λόγου εντάσεων ορίζονται στην Εξ. (2.30), η τιμή του συντελεστή $\theta_j^{(1)}(k,\ell)$ και οι τιμές των $\left\{\theta_j^{(2)}(k,\ell), \theta_j^{(3)}(k,\ell)\right\}$ συντελεστών χαρακτηρίζουν το περιεχόμενο των ακμών της εικόνας στην οριζόντια και κάθετη κατεύθυνση, αντιστοίχως, για την συγκεκριμένη κλίμακα και θέση, όπως άλλωστε προκύπτει και από την απευθείας εξέταση του Σχ. 2.11. Ειδικότερα, η ύπαρξη μιας βηματικής ακμής (step edge) σε κάθε κατεύθυνση υπονοείται εάν η τιμή του αντίστοιχου λόγου εντάσεων αποκλίνει σημαντικά από το 0.5. Υπό αυτό το πρίσμα, ένα μειονέκτημα του 2-Δ διαχωρίσιμου σχήματος είναι ότι εμφανίζει φτωχή κατευθυντική επιλεκτικότητα όσον αφορά στη σύλληψη διαγώνιων ακμών στην εικόνα. Δυστυχώς, ούτε το τετραδικό σχήμα είναι αποτελεσματικό στο συγκεκριμένο έργο. Πιο συγκεκριμένα, εάν όλοι οι λόγοι εντάσεων $\boldsymbol{\theta}_j(k,\ell)$ που ορίζονται στην Εξ. (2.31) είναι ίσοι με 0.25, τότε η εικόνα είναι ομαλή στη συγκεκριμένη περιοχή, αφού ο κόμβος-πατέρας ισοκατανέμει την τιμή έντασής του σε όλους τους κόμβους-παιδιά, τα οποία βρίσκονται στην αμέσως επόμενη ευκρινή κλίμακα. Σε διαφορετική περίπτωση δεν είναι τετριμμένο να αποφανθούμε για την ύπαρξη ακμής αλλά και για τον προσανατολισμό αυτής.

Για να επιτύχουμε βελτιωμένη κατευθυντική επιλεκτικότητα σε σύγκριση με τις δύο προηγούμενες προσεγγίσεις, προτείνουμε ένα πρωτότυπο πολυκλιμακωτό σχήμα ανάλυσης κατάλληλο



Σχήμα 2.13: Τιμή των συντελεστών που προκύπτουν από την ανάλυση της 512×512 εικόνας Lena μέσω του 2-Δ τετραδικού μετασχηματισμού για 2 επίπεδα ανάλυσης. Πάνω δεξιά αντιστοιχούν οι $\theta_j^{(1)}(k,\ell)$ συντελεστές, κάτω αριστερά οι $\theta_j^{(2)}(k,\ell)$ συντελεστές και κάτω δεξιά οι $\theta_j^{(3)}(k,\ell)$ συντελεστές. Σύγκρινε με το αντίστοιχο Σχ. 2.11 για το διαχωρίσιμο 2-Δ μετασχηματισμό.

για 2-Δ Poisson στοχαστικές διαδικασίες, το οποίο έχει σχεδιαστεί ώστε να μοντελοποιεί της διαγώνιες ακμές της εικόνας με σαφή τρόπο [LMP09, LPM09]. Το προτεινόμενο σχήμα καλείται *Poisson-Haar* λόγω του στενού δεσμού του με τον 2-Δ Haar μετασχηματισμό κυματιδίων. Όπως στο διαχωρίσμο, έτσι, και στο παρόν σχήμα υποθέτουμε ότι μεταξύ δύο κλιμάκων ανάλυσης, $j-1$ και j, υπάρχει ένα ενδιάμεσο επίπεδο. Η παρατήρηση του κόμβου-πατέρα, $\lambda_j(k,\ell)$, θεωρείται τότε ως το αποτέλεσμα δύο διαδοχικών βημάτων. Πρώτον, από τους τέσσερις κόμβους-παιδιά της κλίμακας $j-1$ παράγονται δύο ενδιάμεσοι όροι, οι οποίοι στο δεύτερο βήμα προστίθενται για να δώσουν την παρατήρηση του κόμβου-πατέρα, όπως απεικονίζεται στο Σχ. 2.14. Αυτή η διαδικασία προσομοιάζει στον τρόπο με τον οποίο το διαχωρίσμο σχήμα αναλύει την εικόνα. Ωστόσο, υπάρ-



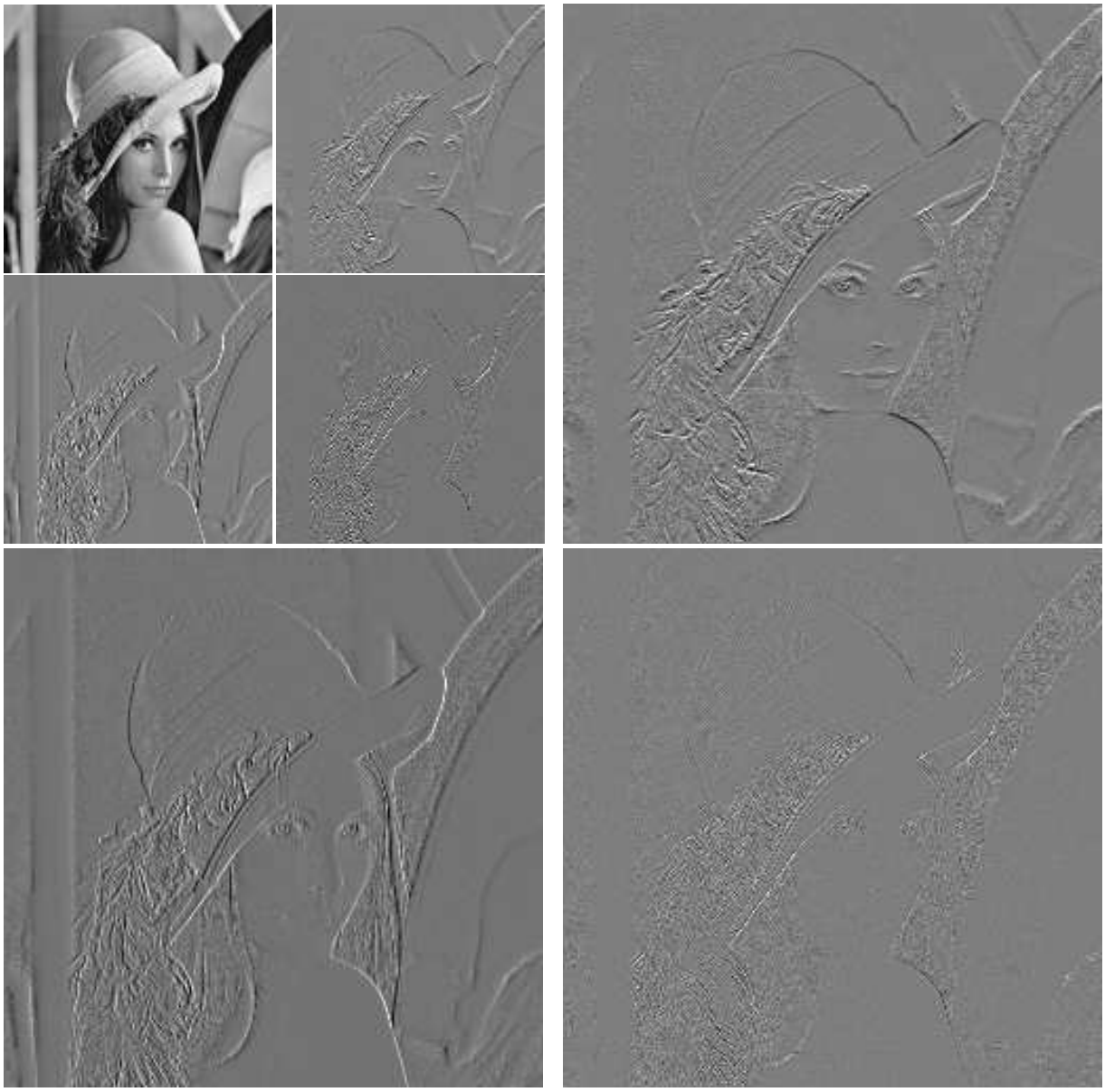
Σχήμα 2.14: Πολυκλιμακωτή αναπαράσταση εικόνας μέσω του Poisson-Haar μετασχηματισμού. Και τα τρία δένδρα καταλήγουν στην ίδια τιμή του κόμβου-πατέρα $\lambda_j(k,\ell)$ στην j κλίμακα. Η διαφορά ανάμεσα τους έγκειται στον τρόπο που δημιουργείται η ενδιάμεση κλίμακα, σχεδιασμένη έτσι ώστε οι 3 παράγοντες διαχωρισμού $\theta^h, \theta^v, \theta^d$ της Εξ. (2.32) να παρουσιάζουν ευαισθησία σε ακμές εικόνων διαφορετικού προσανατολισμού (οριζόντιος, κάθετος, ή διαγώνιος).

χει μια κρίσιμη διαφορά, ενώ στο διαχωρίσιμο σχήμα μόνο ένα ζεύγος από ενδιάμεσους όρους χρησιμοποιείται, δες το Σχ. 2.10, στο προτεινόμενο σχήμα εξετάζουμε και τους τρεις πιθανούς συνδυασμούς ζευγαριών στην κλίμακα $j-1$ που παράγουν τους ενδιάμεσους όρους. Κάθε συνδυασμός αντιστοιχεί σε ένα διαφορετικό δένδρο ανάλυσης και μπορεί να συλλάβει τις ακμές της εικόνας σε διαφορετικούς προσανατολισμούς (οριζόντιος, κάθετος ή διαγώνιος). Το πλεονέκτημα του σχήματος μας είναι ότι διαχειρίζεται συμμετρικά όλους τους τρεις προσανατολισμούς, ενώ το διαχωρίσιμο σχήμα λαμβάνει υπόψιν του μόνο τους οριζόντιους και κάθετους προσανατολισμούς.

Πιο συγκεκριμένα, το προτεινόμενο πολυκλιμακωτό σχήμα ορίζεται αναδρομικά ως εξής: Ας υποθέσουμε ότι παρατηρούμε τις μετρήσεις έντασης $\boldsymbol{\lambda}_j^c(k,\ell)$, οι οποίες αποτελούν απογόνους του κόμβου $\lambda_j(k,\ell)$ και βρίσκονται σε μια γειτονιά $C_{k,\ell}$. Τότε η παρατήρηση του κόμβου-πατέρα, $\lambda_j(k,\ell)$, μπορεί ισοδύναμα να ληφθεί ως το άθροισμα ενός οποιοδήποτε από τα τρία ξεχωριστά ζεύγη ενδιάμεσων όρων, $\{\lambda_{h_1},\lambda_{h_2}\}$, $\{\lambda_{v_1},\lambda_{v_2}\}$ και $\{\lambda_{d_1},\lambda_{d_2}\}$, όπως απεικονίζεται στο Σχ. 2.14. Στη συνέχεια, οι τρεις απαραίτητοι συντελεστές διαχωρισμού $\boldsymbol{\theta}$ υπολογίζονται ως το πηλίκο του άνω (ή δεξιού) κόμβου της ενδιάμεσης κλίμακας του κάθε δένδρου, με τον κόμβο-πατέρα. Η διαδικασία αυτή υπολογισμού των συντελεστών κλιμάκωσης και λόγου εντάσεων είναι δυνατόν να περιγραφεί μέσω των παρακάτω αναδρομικών εξισώσεων:

$$
\begin{aligned}
\lambda_j(k,\ell) &= \sum_{(k',\ell')\in C_{k,\ell}} \lambda_{j-1}(k',\ell') \\
\theta_j^h(k,\ell) &= \frac{\lambda_{j-1}(2k,2\ell)+\lambda_{j-1}(2k,2\ell+1)}{\lambda_j(k,\ell)} \\
\theta_j^v(k,\ell) &= \frac{\lambda_{j-1}(2k,2\ell)+\lambda_{j-1}(2k+1,2\ell)}{\lambda_j(k,\ell)} \\
\theta_j^d(k,\ell) &= \frac{\lambda_{j-1}(2k,2\ell)+\lambda_{j-1}(2k+1,2\ell+1)}{\lambda_j(k,\ell)}\,,
\end{aligned}
\tag{2.32}
$$

όπου ο υπερδείκτης $o=\{h,v,d\}$ δηλώνει τον προσανατολισμό (οριζόντια, κατακόρυφο και διαγώνιο). Συγκρίνοντας τους προκύπτοντες λόγους έντασης με εκείνους του 2-Δ διαχωρίσιμου σχήματος της Εξ. (2.30), παρατηρούμε ότι το Poisson-Haar σχήμα αναπαριστά με σαφή τρόπο όχι μόνο τις κάθετες και οριζόντιες, αλλά, και τις διαγώνιες ακμές (δες Σχ. 2.15), ενώ παράλληλα ο χειρισμός και των τριών προσανατολισμών γίνεται συμμετρικά. Επιπρόσθετα, σε αυτή την περίπτωση, κάτι

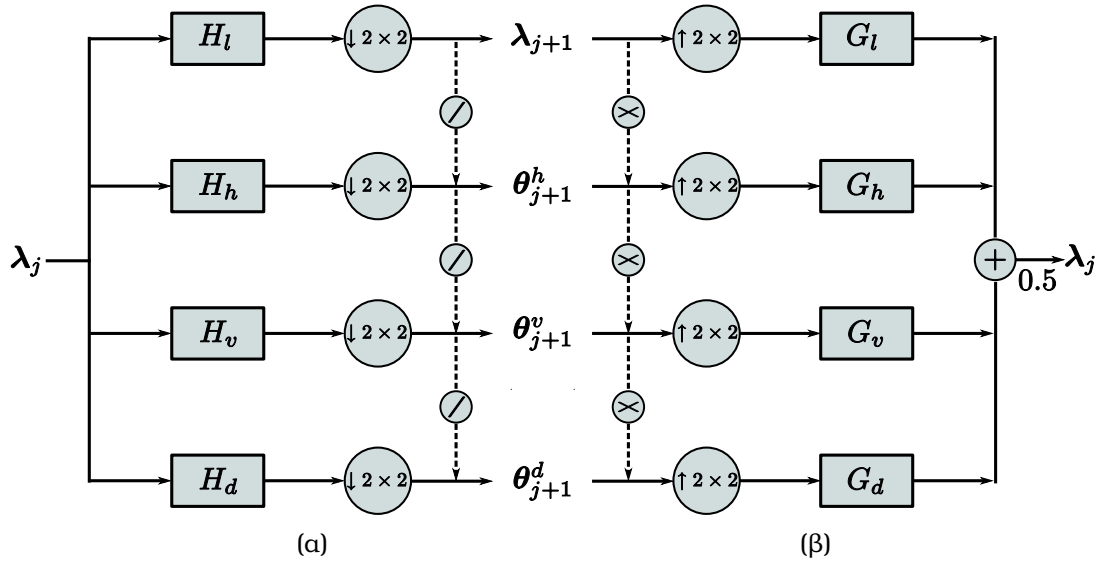


Σχήμα 2.15: Τιμή των συντελεστών που προκύπτουν από την ανάλυση της 512×512 εικόνας Lena μέσω του 2-Δ Poisson-Haar μετασχηματισμού για 2 επίπεδα ανάλυσης. Πάνω δεξιά αντιστοιχούν οι $\theta_j^{(h)}(k,\ell)$ συντελεστές, κάτω αριστερά οι $\theta_j^{(v)}(k,\ell)$ συντελεστές και κάτω δεξιά οι $\theta_j^{(d)}(k,\ell)$ συντελεστές. Σύγκρινε με τα αντίστοιχα Σχ. 2.11 και 2.13 για το διαχωρίσιμο και τετραδικό 2-Δ μετασχηματισμό.

που δεν ισχύει για τις περιπτώσεις του διαχωρίσιμου και τετραδικού σχήματος, αλλά, όμοια με την μονοδιάστατη περίπτωση που συζητήθηκε στην Ενότητα 2.2.2, υπάρχει μια άμεση σχέση μεταξύ των θ συντελεστών και των συντελεστών λεπτομέρειας d του 2-Δ Haar μετασχηματισμού κυματιδίων στην αντίστοιχη υποζώνη:

$$\theta_j^o(k,\ell) = \frac{d_j^o(k,\ell)}{2c_j(k,\ell)} + \frac{1}{2}\,.$$

Ακόμη, τα $\lambda_j(k,\ell)$ είναι ανακλιμακούμενες εκδοχές των 2-Δ Haar συντελεστών κλιμάκωσης, $\lambda_j(k,\ell) = 2^j c_j(k,\ell)$. Η εξίσωση (2.32) υπονοεί ότι μπορούμε να χρησιμοποιήσουμε μια ισοδύνα-



Σχήμα 2.16: Poisson-Haar πολυκλιμακωτή ανάλυση εικόνας μέσω της ανάλυσης της σε συχνοτικές υποζώνες με τη χρήση συστοιχιών φίλτρων. (α) Συστοιχία φίλτρων ανάλυσης, (β) Συστοιχία φίλτρων σύνθεσης.

μη διανυσματική αναπαράσταση για το διάνυσμα έντασης της πηγαίας εικόνας, $\boldsymbol{\lambda}$, της μορφής:

$$\boldsymbol{\lambda}_e = \left[\lambda_J(0,0), \boldsymbol{\theta}_J(0,0), \ldots, \boldsymbol{\theta}_1(0,0), \ldots, \boldsymbol{\theta}_1(N_1/2-1, N_2/2-1)\right] , \tag{2.33}$$

όπου τα διανύσματα $\boldsymbol{\theta}$ θα είναι της μορφής: $\boldsymbol{\theta}_j(k,\ell) = \left(\theta^h_j(k,\ell), \theta^v_j(k,\ell), \theta^d_j(k,\ell)\right)$. Η διαδικασία ανακατασκευής του $\boldsymbol{\lambda}$ από το $\boldsymbol{\lambda}_e$ (2.33) είναι ένα-προς-ένα και μπορεί να επιτευχθεί μέσω μιας αναδρομικής διαδικασίας, από την πιο αδρή στην πιο ευκρινή κλίμακα, για $j = J, J-1, \ldots, 1$:

$$\begin{aligned}
\lambda_{j-1}(2k, 2\ell) &= \lambda_j(k,\ell)\left(\theta^h_j(k,\ell) + \theta^v_j(k,\ell) + \theta^d_j(k,\ell) - 1\right)/2 \\
\lambda_{j-1}(2k, 2\ell+1) &= \lambda_j(k,\ell)\left(\theta^h_j(k,\ell) - \theta^v_j(k,\ell) - \theta^d_j(k,\ell) + 1\right)/2 \\
\lambda_{j-1}(2k+1, 2\ell) &= \lambda_j(k,\ell)\left(\theta^v_j(k,\ell) - \theta^h_j(k,\ell) - \theta^d_j(k,\ell) + 1\right)/2 \\
\lambda_{j-1}(2k+1, 2\ell+1) &= \lambda_j(k,\ell)\left(\theta^d_j(k,\ell) - \theta^h_j(k,\ell) - \theta^v_j(k,\ell) + 1\right)/2 .
\end{aligned} \tag{2.34}$$

2.3.4 Poisson-Haar Μετασχηματισμός και Συστοιχίες Φίλτρων

Λόγω του στενού δεσμού που εμφανίζει ο Poisson-Haar μετασχηματισμός με το 2-Δ Haar μετασχηματισμό κυματιδίων, είναι δυνατόν η εφαρμογή του πρώτου σε ένα 2-Δ σήμα να γίνει μέσω της επαναληπτικής ανάλυσης του σήματος σε υποζώνες από μια συστοιχία φίλτρων [LPM09]. Συγκεκριμένα, για να υπολογίσουμε ένα στάδιο του Poisson-Haar μετασχηματισμού σε μια εικόνα $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}$ μεγέθους $N_1 \times N_2$, αρχικά φιλτράρουμε την εικόνα με τα φίλτρα $H_l = \left(\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right)$, $H_h = \left(\begin{smallmatrix}0&0\\1&1\end{smallmatrix}\right)$, $H_v = \left(\begin{smallmatrix}0&1\\0&1\end{smallmatrix}\right)$, $H_d = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ και, στη συνέχεια, υποδειγματοληπτούμε την έξοδό τους κατά ένα παράγοντα 2 σε κάθε κατεύθυνση. Έπειτα, οι τρεις τελευταίες ζωνοπερατές εικόνες μεγέθους $N_1/2 \times N_2/2$ κανονικοποιούνται σημείο προς σημείο από τη βαθυπερατή εκδοχή της εικόνας, όπως απεικονίζεται στο Σχ. 2.16(α). Η διαδικασία φιλτραρίσματος και υποδειγματοληψίας μπορεί να συνεχιστεί μέχρι J φορές οδηγώντας σε μια πολυκλιμακωτή ανάλυση του $\boldsymbol{\lambda}$. Επειδή ο συγκεκριμένος μετασχηματισμός επιτρέπει την τέλεια ανακατασκευή, είναι δυνατόν από την πολυκλιμακωτή αναπαράσταση της εικόνας να οδηγηθούμε πίσω στην αρχική μας αναπαράσταση ακολουθώντας την αντίστροφη διαδικασία, όπως παρουσιάζεται στο Σχ. 2.16(β). Σε αυτή την

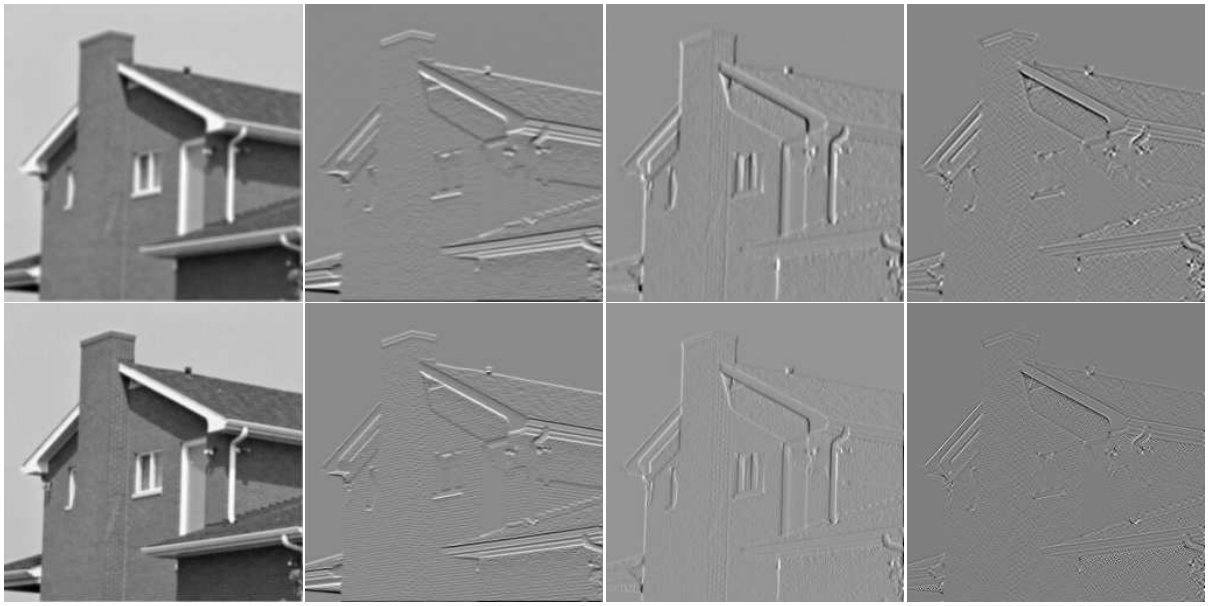
περίπτωση τα φίλτρα της συστοιχίας θα είναι τα $G_l = \left(\begin{smallmatrix} -1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, $G_h = \left(\begin{smallmatrix} 1 & 1 \\ -1 & -1 \end{smallmatrix}\right)$, $G_v = \left(\begin{smallmatrix} 1 & -1 \\ 1 & -1 \end{smallmatrix}\right)$, και $G_d = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$. Εδώ θα πρέπει να σημειώσουμε ότι σε αντίθεση με τον 2-Δ μετασχηματισμό Haar που μπορεί να υλοποιηθεί με διαχωρίσιμα φίλτρα στην περίπτωση μας κάποια από τα φίλτρα που χρησιμοποιούμε δεν επιδέχονται διαχωρισμού.

Υπερπλήρη Αναπαράσταση Η δυνατότητα υλοποίησης του Poisson-Haar μετασχηματισμού με τη χρήση συστοιχιών φίλτρων έχει το επιπλέον πλεονέκτημα ότι μας υποδεικνύει με ξεκάθαρο τρόπο πως μπορούμε να οδηγηθούμε σε μια υπερπλήρη αναπαράσταση. Ομοίως με το μη-κρίσιμα δειγματοληπτημένο μετασχηματισμό κυματιδίων (UDWT), είναι δυνατόν να καταλήξουμε σε μια υπερπλήρη αναπαράσταση της εικόνας εισόδου απαλείφοντας τα στάδια υποδειγματοληψίας κατά τη φάση της ανάλυσης. Με αυτό τον τρόπο μπορούμε να αντιμετωπίσουμε το φαινόμενο της φασματικής αναδίπλωσης το οποίο είναι υπεύθυνο για την εμφάνιση τεχνουργημάτων στην ανακατασκευασμένη εικόνα, όταν οι συντελεστές διαχωρισμού κατά την επεξεργασία τους έχουν τροποποιηθεί από την αρχική τους τιμή. Προκειμένου να οδηγηθούμε στην ολοκληρωμένη εκδοχή του υπερπλήρους Poisson-Haar μετασχηματισμού, θα πρέπει εκτός από την απαλοιφή της δειγματοληψίας τα φίλτρα, τα οποία χρησιμοποιούνται στο στάδιο ανάλυσης, να αποτελούν κατάλληλες υπερδειγματοληπτημένες εκδόσεις των αρχικών φίλτρων. Έτσι, στην πρώτη κλίμακα ανάλυσης κάνουμε χρήση των αρχικών φίλτρων, ενώ στην κλίμακα j χρησιμοποιούμε τα φίλτρα που προκύπτουν, αν στα αρχικά φίλτρα κάνουμε ένθεση $2^{j-1}-1$ μηδενικών μεταξύ γειτονικών γραμμών και στηλών. Έτσι, για παράδειγμα, στη δεύτερη κλίμακα ανάλυσης θα χρησιμοποιήσουμε αντί του φίλτρου H_l το φίλτρο $H_l^{(2)} = \left(\begin{smallmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{smallmatrix}\right)$. Την ίδια επεξεργασία θα υποστούν και τα υπόλοιπα φίλτρα της συστοιχίας ανάλυσης. Για την ανακατασκευή θα χρησιμοποιήσουμε πάλι τις εκτεταμένες εκδοχές των φίλτρων σύνθεσης χωρίς υπερδειγματοληψία των εισόδων των φίλτρων. Ένα παράδειγμα ανάλυσης εικόνας με τον μη-κρίσιμα δειγματοληπτημένο Poisson-Haar μετασχηματισμό για δύο επίπεδα ανάλυσης παρουσιάζεται στο Σχ. 2.17. Όπως και στον UDWT, έτσι και εδώ για L κλίμακες ανάλυσης ο βαθμός πλεονασμού (redundancy factor) θα είναι ίσος με $R = (L+1)$ για 1-Δ σήματα και $R_2 = 3(L+1)$ για εικόνες.

2.4 Ιδιότητες Αναπαράστασης Εικόνων στο Πεδίο Τοπικά Εντοπισμένων Μετασχηματισμών

Οι πολυκλιμακωτές αναπαραστάσεις που προκύπτουν μέσω των μετασχηματισμών, τους οποίους περιγράψαμε στις Ενοτ. 2.1 και 2.3, παρουσιάζονται ως ιδανικές για την στατιστική περιγραφή και μοντελοποίηση φυσικών εικόνων. Για το είδος των φυσικών εικόνων που μας ενδιαφέρει να μελετήσουμε, δηλαδή τις εικόνες περιορισμένης φωτεινότητας, θα χρησιμοποιήσουμε τις διάφορες εκδοχές του 2-Δ πολλαπλασιαστικού μετασχηματισμού. Για το συγκεκριμένο είδος εικόνων ο μετασχηματισμός αυτός είναι καταλληλότερος από το μετασχηματισμό κυματιδίων, αφού, όπως θα δείξουμε στο Κεφ. 3, συνδέεται στενά με τις Poisson στοχαστικές διαδικασίες και επιτρέπει τη στατιστική ανάλυση τους με αποδοτικό τρόπο.

Οι λόγοι που μπορεί να συνηγορούν στη χρήση κατάλληλων τοπικά εντοπισμένων μετασχηματισμών για τη μοντελοποίηση φυσικών εικόνων σχετίζονται με τις ιδιότητες που διαθέτουν οι συγκεκριμένοι μετασχηματισμοί. Για να θεωρείται ένας μετασχηματισμός ιδανικός υποψήφιος για τη μοντελοποίηση εικόνων, θα πρέπει η χρήση του να ευνοεί την ανάπτυξη αποδοτικών μεθόδων και αλγορίθμων για την επεξεργασία τους. Τέτοιες ελκυστικές ιδιότητες διαθέτει τόσο ο μετασχηματισμός κυματιδίων όσο και ο πολλαπλασιαστικός μετασχηματισμός, λόγω της στενής σχέσης που τους συνδέει. Στη συνέχεια, παραθέτουμε αυτές τις ιδιότητες αναφερόμενοι στο μετασχηματισμό κυματιδίων. Ωστόσο, οι ίδιες ή ανάλογες ιδιότητες ισχύουν και για τον πολλαπλασιαστικό μετασχηματισμό.



Σχήμα 2.17: Πολυκλιμακωτή ανάλυση της εικόνας house μεγέθους 256×256 με τον μη-κρίσιμα δειγματοληπτημένο Poisson-Haar μετασχηματισμό για δύο κλίμακες ανάλυσης. Από αριστερά προς τα δεξιά παρουσιάζονται οι συντελεστές κλιμάκωσης και οι συντελεστές διαχωρισμού για τρεις διαφορετικές κατευθύνσεις (οριζόντια, κατακόρυφη, διαγώνια). Τα επίπεδα ανάλυσης εμφανίζονται με σειρά από κάτω προς τα πάνω, δηλαδή το ευκρινέστερο επίπεδο βρίσκεται στην δεύτερη γραμμή και το αδρότερο στην πρώτη γραμμή.

- **Τοπικότητα**: Ο κάθε συντελεστής του μετασχηματισμού κυματιδίων αναπαριστά το χωρικό και συχνοτικό περιεχόμενο της εικόνας τοπικά. Η ιδιότητα του αυτή είναι ιδιαίτερα επιθυμητή, αφού τον καθιστά κατάλληλο να εφαρμοστεί σε μη-στάσιμες στοχαστικές διαδικασίες σε αντίθεση, για παράδειγμα, με το μετασχηματισμό Fourier.

- **Πολλαπλές Κλίμακες Ανάλυσης**: Τα κυματίδια (άτομα) τα οποία συνιστούν το πλαίσιο ή τη βάση του νέου χώρου αναπαράστασης των σημάτων, αποτελούν κλιμακούμενες και μετατοπισμένες εκδοχές ενός βασικού πατρικού και ενός μητρικού κυματιδίου. Η ιδιότητα αυτή επιτρέπει την ανάλυση των σημάτων σε ένα εμφωλευμένο σύνολο κλιμάκων.

- **Αποσυσχέτιση**: Ο μετασχηματισμός κυματιδίων είναι, κατά μία καλή προσέγγιση, μετασχηματισμός αποσυσχέτισης σε μια ευρεία κατηγορία στοχαστικών διαδικασιών. Επιπλέον, χάρη στην τοπικότητα του έχει την δυνατότητα να αποσυσχετίζει σε ικανοποιητικό βαθμό και μη-στάσιμες διαδικασίες. Η ιδιότητα της αποσυσχέτισης είναι από τις πιο σημαντικές που διαθέτει ο μετασχηματισμός κυματιδίων, αφού επιτρέπει την αποδοτική επεξεργασία εικόνων με την ανάπτυξη απλών αλγορίθμων και μεθόδων οι οποίες χειρίζονται κάθε συντελεστή ανεξάρτητα από τους υπόλοιπους.

- **Ανίχνευση Ακμών**: Οι συντελεστές κυματιδίων δρουν τοπικά ως ανιχνευτές ακμών. Οι ακμές στην εικόνα αναπαριστώνται με συντελεστές που έχουν μεγάλη τιμή στην αντίστοιχη θέση, ενώ οι ομαλές περιοχές της εικόνας αναπαριστώνται με συντελεστές που έχουν μικρές (περίπου μηδενικές) τιμές. Η ιδιότητα αυτή καταδεικνύει ότι ο μετασχηματισμός κυματιδίων έχει την ικανότητα να αναγνωρίζει δομικά στοιχεία της εικόνας τα οποία εμπίπτουν στο στάδιο της πρώιμης όρασης. Αντίστοιχη ιδιότητα διαθέτει και ο πολλαπλασιαστικός μετασχηματισμός με τη διαφορά ότι στην περίπτωση αυτή οι ομαλές περιοχές της εικόνας αναπαριστώνται

με συντελεστές που οι τιμές τους είναι κοντά στο 0.5 ή 0.25, ανάλογα με την επιλεγόμενη έκδοση του 2-Δ μετασχηματισμού, ενώ οι ακμές αναπαριστώνται με συντελεστές που οι τιμές τους αποκλίνουν σημαντικά από τις προαναφερθείσες τιμές.

- **Αραιές Αναπαραστάσεις**: Ο μετασχηματισμός κυματιδίων αλλά και οι υπερπλήρεις εκδοχές του, επιτρέπουν την αραιή αναπαράσταση εικόνων, δηλαδή μια αναπαράσταση όπου μόνο ένα μικρό ποσοστό των συντελεστών έχει μη-μηδενικές τιμές. Το γεγονός αυτό συνδέεται με το ότι οι ακμές σε μια εικόνα καταλαμβάνουν σημαντικά μικρότερο χώρο από τις ομαλές περιοχές και δίνει τη δυνατότητα για ικανοποιητική ανακατασκευή του αρχικού σήματος με σχετικά λίγους μη-μηδενικούς συντελεστές. Η ιδιότητα αυτή βρίσκει επιτυχή εφαρμογή στο πρόβλημα της κωδικοποίησης, αποκατάστασης και επεξεργασίας εικόνων [Fie94].

Η ιδιότητα της τοπικότητας σε συνδυασμό με την ιδιότητα των πολλαπλών κλιμάκων ανάλυσης επιτρέπει στο μετασχηματισμό κυματιδίων να ανιχνεύει ένα μεγάλο σύνολο χαρακτηριστικών των σημάτων, από υψίσυχνα μεταβατικά φαινόμενα και ακμές εώς αργά μεταβαλλόμενες αρμονικές. Η ικανότητα του μετασχηματισμού κυματιδίων να εφαρμόζεται με επιτυχία σε ένα ευρύ σύνολο σημάτων οδηγεί στην ιδιότητα των αραιών αναπαραστάσεων. Με βάση αυτή την ιδιότητα πολύπλοκα σήματα είναι δυνατόν να αναπαριστώνται μέσω ενός μικρού μόνο αριθμού ατόμων. Το αποτέλεσμα όλων αυτών των ιδιοτήτων είναι ότι οι μέθοδοι στατιστικής μοντελοποίησης και επεξεργασίας που εφαρμόζονται στο πεδίο των κυματιδίων είναι σε πολλές περιπτώσεις πολύ περισσότερο αποδοτικοί από αντίστοιχες τυπικές μεθόδους που εφαρμόζονται στο χωρικό ή στο συχνοτικό πεδίο.

2.5 Σύνοψη και Συμπεράσματα

Στο κεφάλαιο αυτό μελετήσαμε τοπικά εντοπισμένους μετασχηματισμούς για την πολυκλιμακωτή αναπαράσταση σημάτων και εικόνων σε πολλαπλές κλίμακες ανάλυσης. Με βάση τη θεωρία της πολυκλιμακωτής προσέγγισης (MRA) [Mal99] είναι δυνατή η μετάβαση από τον αρχικό χώρο αναπαράστασης των σημάτων σε πολυκλιμακωτές βάσεις μέσω του μετασχηματισμού κυματιδίων. Τέτοιου είδους αναπαραστάσεις, αλλά και οι υπερπλήρεις εκδοχές τους, είναι ιδανικές για τη μοντελοποίηση των υπό εξέταση σημάτων, αφού διαθέτουν κατάλληλες ιδιότητες που μας επιτρέπουν την ανάπτυξη απλών μεθόδων για την αποδοτική αντιμετώπιση πληθώρας προβλημάτων. Η χρήση αυτών των μετασχηματισμών γίνεται ακόμα πιο ελκυστική λόγω της σύνδεσής τους με την επαναληπτική ανάλυση των σημάτων σε συχνοτικές υποζώνες από συστοιχίες φίλτρων.

Για την ανάλυση και επεξεργασία εικόνων περιορισμένης φωτεινότητας, που αποτελεί ένα από τα βασικά προβλήματα που επιχειρούμε να αντιμετωπίσουμε στην παρούσα διατριβή, παρουσιάσαμε μια συγγενή κατηγορία μετασχηματισμών οι οποίοι αποτελούν καταλληλότερο υποψήφιο, λόγω της σύνδεσης τους με τις Poisson στοχαστικές διαδικασίες. Επιπλέον, η στενή σχέση που δείξαμε ότι υπάρχει μεταξύ του 1-Δ πολλαπλασιαστικού μετασχηματισμού και του Haar μετασχηματισμού κυματιδίων υποδηλώνει ότι και η συγκεκριμένη κατηγορία μετασχηματισμών κληρονομεί κάποιες από τις επιθυμητές ιδιότητες οι οποίες κατέστησαν το μετασχηματισμό κυματιδίων εξαιρετικό εργαλείο για την επεξεργασία σημάτων. Ωστόσο, ένα από τα προβλήματα που καλείται κανείς να αντιμετωπίσει κατά τη χρήση του 1-Δ πολλαπλασιαστικού μετασχηματισμού είναι ότι η επέκτασή του στην ανάλυση σημάτων μεγαλύτερων διαστάσεων, όπως οι εικόνες, δεν είναι προφανής. Για αυτό το λόγο εκτός από τις υπάρχουσες 2-Δ επεκτάσεις που έχουν προταθεί στη βιβλιογραφία αναπτύξαμε ένα πρωτότυπο 2-Δ μετασχηματισμό. Το βασικό πλεονέκτημα που παρουσιάζει ο συγκεκριμένος μετασχηματισμός σε σχέση με τις υπόλοιπες 2-Δ εκδοχές είναι ότι διατηρεί την απευθείας σύνδεση με το 2-Δ Haar μετασχηματισμό κυματιδίων. Το γεγονός αυτό δίνει τη δυνατότητα καλύτερης κωδικοποίησης βασικών δομών της εικόνας που τελικά συνεπάγεται, όπως θα δείξουμε και στο Κεφ. 3 την αποδοτικότερη επεξεργασία τους.

Κεφάλαιο 3

Βέλτιστη Αποθορυβοποίηση Poisson Εικόνων με χρήση Μπεϋζιανής Συμπερασματολογίας και Δενδρικών Στατιστικών Μοντέλων

Ο σχηματισμός εικόνων σε ένα πλήθος από σύγχρονα απεικονιστικά συστήματα επιτυγχάνεται μέσω της ανίχνευσης του αριθμού φωτονίων που προσπίπτουν σε διαφορετικά χωρικά σημεία ενός αισθητήρα μέσα σε ένα συγκεκριμένο χρονικό διάστημα παρατήρησης. Παράδειγμα ενός τέτοιου συστήματος αποτελεί η ψηφιακή κάμερα συζευγμένου φορτίου (Charged Couple Device - CCD). Συγκεκριμένα, μια CCD κάμερα περιλαμβάνει έναν ψηφιακό CCD ανιχνευτή που είναι ένας καταχωρητής ολίσθησης, δηλαδή, μια πολύ μικρή πλάκα πάνω στην οποία βρίσκονται διατεταγμένα έως και ένα εκατομμύριο στοιχεία ενός ημιαγώγιμου υλικού, ευαίσθητου στο φως (συνήθως πυρίτιο). Όταν ο ανιχνευτής εκτίθεται σε μια φωτεινή πηγή, σε καθένα απ' αυτά τα στοιχεία απελευθερώνονται ηλεκτρικά φορτία (ηλεκτρόνια) σε ευθεία αναλογία με τα φωτόνια που προσπίπτουν πάνω στο στοιχείο. Μετά την έκθεση στο φως ο αριθμός των συγκεντρωμένων ηλεκτρονίων στο κάθε στοιχείο καθορίζει την ένταση φωτεινότητας του αντίστοιχου σημείου πάνω στην οθόνη του υπολογιστή με τον οποίο είναι συνδεδεμένος ο ανιχνευτής CCD. Έτσι, η φωτογραφημένη εικόνα ανασυντίθεται σημείο προς σημείο στην οθόνη.

Σε τέτοια συστήματα απεικόνισης, όταν επικρατούν χαμηλές συνθήκες φωτισμού, όπως κατά κανόνα συμβαίνει κατά τη σύλληψη αστρονομικών [SH93, SM06] αλλά και βιοϊατρικών [SV82] εικόνων, η κύρια πηγή αλλοίωσης και παραμόρφωσης των λαμβανομένων εικόνων καλείται κβαντικός θόρυβος (quantum noise). Ο κβαντικός θόρυβος [Bar90] οφείλεται στις διακυμάνσεις του αριθμού των ανιχνευμένων φωτονίων, μια εγγενής αδυναμία της διακριτής φύσης της διαδικασίας ανίχνευσης, και αλλοιώνει τις εικόνες αυτές τόσο ποιοτικά όσο και ποσοτικά. Η προκύπτουσα αλλοίωση αποδεικνύεται σε πολλές περιπτώσεις ισχυρός παράγοντας που δυσκολεύει σημαντικά την περαιτέρω διαδικασία ανάλυσης των εικόνων και εξαγωγής χρήσιμων χαρακτηριστικών από αυτές. Συνεπώς, η ανάπτυξη μεθόδων και τεχνικών που θα υπερκεράσουν τις υπάρχουσες δυσκολίες είναι καθοριστικής σημασίας.

Σύμφωνα με το βασικό μοντέλο σχηματισμού εικόνων μέσω της ανίχνευσης των προσπιπτόντων φωτονίων, ο αριθμός των ανιχνευμένων φωτονίων σε κάθε εικονοστοιχείο ακολουθεί μια Poisson κατανομή. Συγκεκριμένα, με βάση αυτό το μοντέλο η λαμβανόμενη εικόνα μπορεί να θεωρηθεί ως το στιγμιότυπο μιας ετερογενούς Poisson στοχαστικής διαδικασίας. Αυτή η Poisson διαδικασία χαρακτηρίζεται από μια 2-Δ χωρικά μεταβαλλόμενη συνάρτηση έντασης φωτεινότητας η οποία ισούται με τη μέση τιμή της διαδικασίας [JKK05] και αντιστοιχεί στην πραγματική (αναλλοίωτη)

ένταση φωτεινότητας της εικόνας που θέλουμε να ανακτήσουμε (εικόνα πηγής). Η διακύμανση του αριθμού των φωτονίων ως προς τη μέση τιμή της διαδικασίας μπορεί λοιπόν να ερμηνευθεί ως θόρυβος. Αφού η διακύμανση της Poisson στοχαστικής διαδικασίας ισούται με τη συνάρτηση φωτεινότητας/εικόνα πηγής, ο θόρυβος που εμφανίζεται στη λαμβανόμενη εικόνα είναι χωρικά μεταβαλλόμενος και εξαρτάται από την πραγματική ένταση φωτεινότητας της εικόνας στο συγκεκριμένο σημείο. Αυτή η εξάρτηση του θορύβου από την πηγαία εικόνα μας αποτρέπει από το να χρησιμοποιήσουμε μια πληθώρα από πρόσφατες και καλά μελετημένες μεθόδους, ενδεικτικά αναφέρουμε τις [CB98, CNB98, ML99, MKRM99, CYV00, HKO01, SS02a, SS02b, SBK05], οι οποίες μοντελοποιούν αποδοτικά τα στατιστικά των εικόνων και οδηγούν σε αποτελεσματική αντιμετώπιση και απαλοιφή θορύβων που είναι προσθετικοί και ομογενείς, όπως, για παράδειγμα, ο γκαουσιανός θόρυβος. Από στατιστικής άποψης η αποθορυβοποίηση της λαμβανόμενης ψηφιακής εικόνας αντιστοιχεί στην εκτίμηση της διακριτοποιημένης γενεσιουργού έντασης μιας Poisson στοχαστικής διαδικασίας από ένα και μόνο στιγμιότυπό της. Για μια 2-Δ Poisson διαδικασία η διακριτοποιημένη γενεσιουργός ένταση, η οποία αντιστοιχεί στη φωτεινότητα της εικόνας πηγής, μπορεί να αναπαρασταθεί ως ένας 2-Δ πίνακας $\boldsymbol{\lambda}$, ενώ η θορυβώδης εικόνα παρατήρησης ως ένας 2-Δ πίνακας $\mathbf{x}$. Έτσι, σε κάθε χωρικό σημείο (i, j) ο αριθμός των ανιχνευμένων φωτονίων $x(i, j)$ μπορεί να ερμηνευθεί ως ένα στιγμιότυπο μιας τυχαίας μεταβλητής [1] $X(i, j)$ η οποία υπακούει την Poisson κατανομή με γενεσιουργό ένταση $\lambda(i, j)$.

3.1 Ανασκόπηση Βιβλιογραφίας και Κατηγοριοποίηση Μεθόδων

Στη διεθνή βιβλιογραφία έχει προταθεί πληθώρα τεχνικών για την επίλυση του προβλήματος εκτίμησης της γενεσιουργού έντασης μιας Poisson στοχαστικής διαδικασίας. Οι προτεινόμενες τεχνικές μπορούν να ταξινομηθούν σε τρεις μεγάλες κατηγορίες μεθόδων.

Προεπεξεργασία Poisson Μετρήσεων Μια κλασσική προσέγγιση, η οποία ακολουθείται αρκετά συχνά από πολλούς ερευνητές, περιλαμβάνει ένα πρώτο στάδιο προ-επεξεργασίας των Poisson μετρήσεων μέσω ενός σταθεροποιητικού μετασχηματισμού διακύμανσης (Variance Stabilizing Transform -VST) [Bar36, Ans48, FT50, Fis55, TV68, VT71]. Οι πιο δημοφιλείς VST μετασχηματισμοί είναι του Anscombe [Ans48] και του Fisz [Fis55] οι οποίοι εφαρμόζονται είτε στο χωρικό πεδίο [Don93] είτε στο πεδίο των Haar κυματιδίων [FN04, FDN07], αντιστοίχως. Ένας ανάλογος με τον Fisz μετασχηματισμό προτάθηκε πρόσφατα στο [Jan06] και μπορεί να εφαρμοσθεί με επιτυχία σε Poisson μετρήσεις που αναπαριστώνται σε οποιοδήποτε πεδίο κυματιδίων παράγοντας σε κάθε περίπτωση σταθεροποιημένους (stabilized) συντελεστές. Τέτοιου είδους μετασχηματισμοί αναμορφώνουν τα δεδομένα, ώστε ο θόρυβος που τα αλλοιώνει να προσεγγίζει γκαουσιανό θόρυβο σταθερής διακύμανσης. Έτσι, το πρόβλημα εκτίμησης ανάγεται στο κλασσικό πρόβλημα αποθορυβοποίησης εικόνων αλλοιωμένων από προσθετικό ομογενή γκαουσιανό θόρυβο. Μέσω αυτής της προ-επεξεργασίας είναι πλέον εφικτό να γίνει χρήση καθιερωμένων τεχνικών αποθορυβοποίησης που έχουν σχεδιασθεί για την αντιμετώπιση του γκαουσιανού θορύβου [Don94, Don95]. Μια πρόσφατη μέθοδος της ίδιας κατηγορίας προτάθηκε στο [ZFS08], όπου οι συγγραφείς εφαρμόζουν στα θορυβώδη δεδομένα ένα νέο πολυκλιμακωτό VST μετασχηματισμό (multiscale variance stabilizing transform - MS-VST). Στη συνέχεια, η αποθορυβοποίηση των εικόνων επιτυγχάνεται με την εφαρμογή συμβατικών μεθόδων αποθορυβοποίησης που λαμβάνουν χώρα είτε στο πεδίο των curvelets [SCD02] είτε στο πεδίο των ridgelets [CD99a].

Επεξεργασία στο πεδίο των κυματιδίων Μια εναλλακτική προσέγγιση περιλαμβάνει τεχνικές που εφαρμόζονται στο πεδίο των κυματιδίων οι οποίες όμως έχουν τροποποιηθεί κατάλληλα, ώστε

[1] Από εδώ και στο εξής οι τυχαίες μεταβλητές θα δηλώνονται με κεφαλαία γράμματα ενώ τα στιγμιότυπά τους με μικρά γράμματα.

να λαμβάνουν υπόψιν τους τα Poisson στατιστικά της εικόνας. Με αυτό τον τρόπο αποφεύγει κανείς την προ-επεξεργασία των Poisson δεδομένων. Ο Kolaczyk στα [Kol97, Kol99b, KD00] ανέπτυξε μια μέθοδο κατωφλιοποίησης των κυματιδιακών συντελεστών λεπτομέρειας που προκύπτουν από την εφαρμογή του Haar μετασχηματισμού. Σε αυτή τη μέθοδο εισήγαγε κατάλληλα τροποποιημένα κατώφλια λαμβάνοντας έτσι υπόψιν του τα συγκεκριμένα χαρακτηριστικά του Poisson θορύβου. Επίσης, οι Nowak και Baraniuk [NB99] πρότειναν μια μέθοδο φιλτραρίσματος στο πεδίο των κυματιδίων η οποία βασίζεται σε έναν εκτιμητή διεπικύρωσης, ενώ οι Αντωνιάδης και Σαπατίνας [AS01] πρότειναν μια προσαρμοστική μέθοδο Wiener φιλτραρίσματος.

Πολυκλιμακωτό Μπεϋζιανό πλαίσιο εργασίας Μια άλλη μεγάλη κατηγορία περιλαμβάνει μεθόδους που υιοθετούν ένα πολυκλιμακωτό Μπεϋζιανό πλαίσιο εργασίας [TN99, Now99a, Now99b, Kol99a, NK00, WN03, KN04, WN07a, WN07b] το οποίο βασίζεται αποκλειστικά στα στατιστικά των Poisson δεδομένων και προέκυψε από την ανεξάρτητη έρευνα των Timmerman και Nowak [TN99] και του Kolaczyk [Kol99a]. Ένα από τα κύρια πλεονεκτήματα των Μπεϋζιανών μεθόδων είναι ότι επιτρέπουν την ενσωμάτωση πρότερης γνώσης στη διαδικασία εκτίμησης. Επιπλέον, οι Μπεϋζιανές μέθοδοι συνδυαζόμενες με την πολυκλιμακωτή ανάλυση έχουν γίνει πολύ δημοφιλείς, λόγω των σημαντικών απλοποιήσεων που προσφέρουν στο πρόβλημα της εκτίμησης της γενεσιουργού έντασης. Συγκεκριμένα, αυτό το πλαίσιο εργασίας [TN99, Kol99a] περιλαμβάνει την ανάλυση της πιθανοφάνειας της Poisson στοχαστικής διαδικασίας σε πολλαπλές κλίμακες επιτρέποντας έτσι την εκτίμηση της γενεσιουργού έντασης με υπολογιστικά αποδοτικό τρόπο μέσω αναδρομικών σχημάτων. Επιπλέον, η ανάλυση της πιθανοφάνειας σε πολλαπλές κλίμακες συνοδεύεται με φυσικό τρόπο από ένα παραγοντοποιημένο στατιστικό μοντέλο της έντασης της πηγαίας εικόνας, όπου οι λόγοι των γενεσιουργών εντάσεων των Poisson τυχαίων μεταβλητών, οι οποίες βρίσκονται σε γειτονικές κλίμακες, μοντελοποιούνται ως συνδυασμοί από συζυγή παραμετρικά μοντέλα. Το συγκεκριμένο στατιστικό μοντέλο πρότερης πιθανότητας έχει αποδειχθεί πολύ αποτελεσματικό στην μοντελοποίηση εικόνων. Εξαιτίας αυτών των ιδιοτήτων οι πολυκλιμακωτές Μπεϋζιανές μέθοδοι παρουσιάζουν πολύ ικανοποιητική απόδοση, κυρίως στην περίπτωση των ετερογενών Poisson στοχαστικών διαδικασιών [BFS04] που είναι και η περίπτωση που συγκεντρώνει το μεγαλύτερο ενδιαφέρον.

Συνεισφορές Κεφαλαίου

Παρόλη τη δυναμική που παρουσιάζει το πολυκλιμακωτό Μπεϋζιανό πλαίσιο για την ανάλυση Poisson διαδικασιών, προηγούμενες εργασίες στη συγκεκριμένη ερευνητική περιοχή [TN99, Kol99a, NK00, LKA04] παρουσιάζουν ορισμένες αδυναμίες οι οποίες αποτρέπουν την ευρύτερη υιοθέτηση και εφαρμογή του. Στο παρόν κεφάλαιο προσπαθούμε να αντιμετωπίσουμε τα ακόλουθα προβλήματα:

1. *Αναπαράσταση των Εικόνων*: Μελέτη και ανάπτυξη πολυκλιμακωτών στρατηγικών ανάλυσης που προορίζονται αποκλειστικά για 2-Δ δεδομένα εικόνων και επιτρέπουν την καλύτερη κωδικοποίηση των δομικών τους στοιχείων.

2. *Εκτίμηση Παραμέτρων*: Εκτίμηση των παραμέτρων του μοντέλου πρότερης πιθανότητας απευθείας από τα παρατηρούμενα θορυβώδη Poisson δεδομένα, ώστε το μοντέλο να προσαρμόζεται με ακρίβεια στα στατιστικά της πηγαίας εικόνας.

3. *Δια-κλιμακωτές εξαρτήσεις*: Μοντελοποίηση των δια-κλιμακωτών εξαρτήσεων μεταξύ των συντελεστών λόγου εντάσεων που εμφανίζονται στις φυσικές εικόνες.

Οι κύριες συνεισφορές μας είναι: (1) Όσον αφορά στην αναπαράσταση εικόνων αντιμετωπίσαμε το συγκεκριμένο πρόβλημα μέσω της ανάπτυξης ενός νέου 2-Δ πολλαπλασιαστικού μετα-

σχηματισμού (Poisson-Haar) που παρουσιάσαμε στην Ενοτ. 2.3 του Κεφ. 2. Η στενή σχέση που συνδέει τον προτεινόμενο μετασχηματισμό με το 2-Δ Haar μετασχηματισμό κυματιδίων οδηγεί σε μια πολυκλιμακωτή αναπαράσταση εικόνων η οποία καταφέρνει να κωδικοποιεί αποδοτικότερα την πληροφορία των ακμών τους παρέχοντας έτσι καλύτερα αποτελέσματα στη μοντελοποίηση των εικόνων και κατά συνέπεια στο πρόβλημα εκτίμησης της έντασης φωτεινότητας. (2) Όσον αφορά στο πρόβλημα της εκτίμησης παραμέτρων προτείνουμε μια τεχνική Μεγιστοποίησης-Προσδοκίας (Expectation-Maximization - EM) για την εκτίμηση των παραμέτρων της πρότερης πιθανότητας μέσω μεγιστοποίησης της πιθανοφάνειας απευθείας από τα παρατηρούμενα Poisson δεδομένα, έτσι ώστε το μοντέλο να ειδικεύεται με ακρίβεια στα στατιστικά της πηγαίας εικόνας. Η ευρωστία της τεχνικής αυτής ενισχύεται με την ενσωμάτωση ενός κατάλληλου όρου κανονικοποίησης. Αυτός ο όρος προκύπτει από την εισαγωγή μιας συζυγούς πρότερης πιθανότητας στις παραμέτρους που θα εκτιμηθούν και μας οδηγεί στην εκτίμηση μέσω της μεγιστοποίησης της εκ των υστέρων πιθανότητας αντί για τη μεγιστοποίηση της πιθανοφάνειας. Η προτεινόμενη μέθοδος, η οποία βασίζεται στον EM αλγόριθμο, μπορεί να χρησιμοποιηθεί εξίσου αποδοτικά σε όλες τις αναπαραστάσεις εικόνων που προκύπτουν από την εφαρμογή των 2-Δ εκδοχών του πολλαπλασιαστικού μετασχηματισμού (δες Ενοτ. 2.3). (3) Το στατιστικό πλαίσιο εργασίας που χειρίζεται κάθε κλίμακα ανάλυσης ως ανεξάρτητη από όλες τις άλλες μπορεί να μην είναι καθόλου ρεαλιστικό για πολλές πραγματικές εφαρμογές. Οι κρυφές μαρκοβιανές δενδρικές δομές (HMT), όπως αυτές που έχουν προταθεί στα [CNB98, Now99a, LPH00, Wil02], μπορούν να μοντελοποιήσουν αποδοτικά τις δια-κλιμακωτές εξαρτήσεις μεταξύ των καταστάσεων των κατανομών των συντελεστών διαχωρισμού (συντελεστές λόγου έντασης) και είναι επομένως πιο κατάλληλες. Σε αυτή την περίπτωση το πρόβλημα της εκτίμησης των παραμέτρων είναι ακόμη πιο έκδηλο εξαιτίας των επιπλέον παραμέτρων που πρέπει να ειδικευτούν στα δεδομένα. Το πρόβλημα αυτό το αντιμετωπίζουμε επεκτείνοντας την EM τεχνική μας για την εκτίμηση των παραμέτρων, έτσι, ώστε να καλύπτει και αυτή την περίπτωση. Αυτή η μοντελοποίηση οδηγεί σε επιπλέον πλεονεκτήματα που επιτρέπουν την καλύτερη αντιμετώπιση του προβλήματος εκτίμησης της έντασης της πηγαίας εικόνας. Η αποτελεσματικότητα όλων των προτεινόμενων μεθόδων μας εξακριβώνεται πειραματικά συγκρίνοντας τα αποτελέσματά μας με αποτελέσματα από άλλες τεχνικές σε ένα σύνολο από εικόνες δοκιμής. Επιπλέον, εφαρμόζουμε τις προτεινόμενες μεθόδους για την αποθορυβοποίηση πραγματικών δορυφορικών αστρονομικών εικόνων, αλλά και βιοϊατρικών εικόνων.

Στη συνέχεια του κεφαλαίου και συγκεκριμένα στην Ενότητα 3.2 παρέχουμε μια γενική εισαγωγή στη στατιστική μοντελοποίηση εικόνων. Στην Ενότητα 3.3 συζητούμε το πολυκλιμακωτό Μπεϋζιανό πλαίσιο εργασίας για την εκτίμηση της γενεσιουργού έντασης μιας Poisson στοχαστικής διαδικασίας. Στην ενότητα 3.4 εξάγουμε Μπεϋζιανούς εκτιμητές για τα υπό συζήτηση μοντέλα, ενώ στις Ενότητες 3.5 και 3.6 περιγράφουμε τις EM τεχνικές μας για την εκτίμηση των παραμέτρων του ανεξάρτητου μεταξύ κλιμάκων μοντέλου αλλά και του HMT μοντέλου, αντιστοίχως. Πειραματικά αποτελέσματα και συγκρίσεις σε εικόνες αλλοιωμένες από Poisson θόρυβο παρουσιάζονται στην Ενότητα 3.7. Τέλος, αποτελέσματα σε αστρονομικές δορυφορικές εικόνες παρέχονται στην Ενότητα 3.8.

3.2 Στατιστική Μοντελοποίηση Εικόνων

Στην ερευνητική περιοχή της όρασης όλο και περισσότερη έμφαση δίνεται στη σπουδαιότητα της τεράστιας σε έκταση πρότερης πληροφορίας την οποία διαθέτει ο ανθρώπινος εγκέφαλος σχετικά με τη δομή του κόσμου που τον περιβάλλει. Προσφάτως έχει ξεκινήσει μια προσπάθεια φορμαλισμού αυτών των ιδεών υπό την ονομασία Μπεϋζιανή αντίληψη (Bayesian Perception) [KR96]. Βέβαια σύμφωνα με αρκετούς οι αρχές στις οποίες βασίζεται χρονολογούνται πίσω στο 19ο αιώνα, όπου ο Helmholtz έθεσε τα θεμέλια για τη σύνδεση της αντίληψης με την Μπεϋζιανή συμπερασματολογία (Bayesian Inference). Η Μπεϋζιανή συμπερασματολογία είναι η πιο κατάλληλη θεωρία

που μπορούμε να χρησιμοποιήσουμε, όταν ανακριβής ή ατελής πληροφορία συνδυάζεται με την ύπαρξη πρότερης πληροφορίας.

Προκειμένου να είμαστε σε θέση να κάνουμε χρήση της Μπεϋζιανής συμπερασματολογίας, θα πρέπει να υιοθετήσουμε μια στατιστική προσέγγιση ερμηνείας των εικόνων. Σύμφωνα με αυτή την προσέγγιση απαιτείται η ανάπτυξη στατιστικών μοντέλων για τις εικόνες του φυσικού κόσμου. Βέβαια είναι αυτονόητο ότι η ανάπτυξη στατιστικών μοντέλων ικανών να ερμηνεύουν το σύνολο των εικόνων είναι αδύνατη. Για αυτό το λόγο, ανάλογα με τις εικόνες που μας ενδιαφέρουν είναι απαραίτητη η ανάπτυξη στατιστικών μοντέλων εξειδικεύσιμων στη συγκεκριμένη κατηγορία. Τώρα το ερώτημα που τίθεται είναι με ποιο τρόπο μπορούμε να αξιοποιήσουμε την πρότερη γνώση που διαθέτουμε για τα στατιστικά των εικόνων. Εάν διαθέτουμε ένα αξιόπιστο πρότερο πιθανοτικό μοντέλο για τις εικόνες που μας ενδιαφέρουν, $P(\mathbf{y}|\boldsymbol{\theta})$, όπου $\boldsymbol{\theta}$ είναι ένα σύνολο παραμέτρων των οποίων η κατάλληλη επιλογή μας επιτρέπει την καλύτερη προσαρμογή του μοντέλου στα στατιστικά των εικόνων, και εάν επιπλέον διαθέτουμε ένα πιθανοτικό μοντέλο, $P(\mathbf{y}_o|\mathbf{y})$, το οποίο συνδέει την ατελή ή ανακριβή παρατηρούμενη εικόνα $\mathbf{y}_o$ με την πραγματική εικόνα $\mathbf{y}$, τότε εφαρμόζοντας το θεώρημα του Bayes μπορούμε να υπολογίσουμε την ύστερη συνάρτηση πιθανότητας

$$P(\mathbf{y}|\mathbf{y}_o, \boldsymbol{\theta}) = \frac{P(\mathbf{y}_o|\mathbf{y})\, P(\mathbf{y}|\boldsymbol{\theta})}{P(\mathbf{y}_o)}\,. \tag{3.1}$$

Η συνάρτηση ύστερης πιθανότητας μας παρέχει όλες τις δυνατές πληροφορίες που θα μπορούσαμε δυνητικά να συνάγουμε από τα παρατηρούμενα δεδομένα. Αντιλαμβάνεται λοιπόν κανείς ότι η επιλογή κατάλληλων πρότερων στατιστικών μοντέλων για τις εικόνες είναι καθοριστικής σημασίας όσον αφορά στη στατιστική αντιμετώπιση του προβλήματος.

Το πλαίσιο εργασίας που μόλις περιγράψαμε είναι αρκετά ισχυρό και ικανό να μας οδηγήσει σε αποτελεσματικές λύσεις, αλλά παράλληλα εμπεριέχει και πολλές προκλήσεις. Προκειμένου η συγκεκριμένη προσέγγιση να είναι πρακτικά εφαρμόσιμη, είναι απαραίτητη η υιοθέτηση πιθανοτικών μοντέλων που θα έχουν τέτοια δομή η οποία θα καθιστά υπολογιστικά εφικτές τις ακόλουθες λειτουργίες:

- *Συμπερασματολογία*: Θα πρέπει το υιοθετούμενο μοντέλο να επιτρέπει τον αποδοτικό υπολογισμό της ύστερης συνάρτησης πιθανότητας βάσει του κανόνα του Bayes (3.1) η οποία είναι απαραίτητη στη συναγωγή συμπερασμάτων για την πραγματική εικόνα $\mathbf{y}$.

- *Εκπαίδευση*: Με τη χρήση ενός συνόλου εικόνων εκπαίδευσης θα πρέπει να έχουμε τη δυνατότητα να προσαρμόσουμε τις παραμέτρους $\boldsymbol{\theta}$, βελτιστοποιώντας κάποιο στατιστικό κριτήριο, ώστε το μοντέλο να προσαρμόζεται με ακρίβεια στα στατιστικά των εικόνων της συγκεκριμένης κατηγορίας.

- *Δειγματοληψία/Σύνθεση*: Είναι άκρως επιθυμητό το μοντέλο μας να είναι αναγεννητικό, δηλαδή να έχουμε την δυνατότητα να λαμβάνουμε τυχαία δείγματα από την πρότερη, $P(\mathbf{y}|\boldsymbol{\theta})$, και την ύστερη, $P(\mathbf{y}|\mathbf{y}_o, \theta)$, συνάρτηση πιθανότητας των εικόνων. Μια τέτοια λειτουργία μας δίνει την δυνατότητα να αξιολογούμε τις δυνατότητες του μοντέλου που επιλέξαμε, αλλά και να επιβεβαιώνουμε ή όχι τις παραδοχές που κάναμε.

Με βάση τις παραπάνω απαιτήσεις γίνεται σαφές ότι η κατάλληλη επιλογή ενός πιθανοτικού μοντέλου περιγραφής των εικόνων είναι μια εξαιρετικά δύσκολη και σύνθετη διαδικασία. Σε αυτή την επιλογή θα πρέπει να συνυπολογίσουμε και τον καθορισμό των χαρακτηριστικών των εικόνων που προσδοκούμε το μοντέλο μας να συλλαμβάνει. Η λέξη χαρακτηριστικά συνήθως χρησιμοποιείται με έναν ασαφή τρόπο, προκειμένου να περιγράψει μια οποιαδήποτε συνάρτηση της εικόνας που αποτελεί αντικείμενο περαιτέρω επεξεργασίας. Στο παρόν κεφάλαιο θα μελετήσουμε πιθανοτικά μοντέλα που αφορούν στην *πρώιμη όραση*. Με τον όρο πρώιμη όραση [HHH09]

αναφερόμαστε στα αρχικά στάδια της επεξεργασίας του οπτικού σήματος, τα οποία συνήθως αναφέρονται στο στάδιο υπολογισμού χαρακτηριστικών χαμηλού επιπέδου από αυτό. Αυτό σημαίνει οτι τα συγκεκριμένα χαρακτηριστικά δεν έχουν τη δυνατότητα να περιγράψουν υψηλού επιπέδου ιδιότητες της εικόνας, όπως, για παράδειγμα, την ύπαρξη αντικειμένων. Ωστόσο, είναι ικανά να αναγνωρίσουν πρωτεύουσες δομές της εικόνας, όπως ακμές, καμπύλες, κηλίδες. Τέτοιου είδους δομές κωδικοποιούνται με αρκετή επιτυχία είτε από τους συντελεστές κυματιδίων είτε από τους συντελεστές διαχωρισμού που εξετάσαμε στο Κεφ. 2.

3.3 Πολυκλιμακωτή Μοντελοποίηση Poisson Στοχαστικών Διαδικασιών

3.3.1 Καθορισμός Προβλήματος

Αρχικά, προκειμένου να απλοποιήσουμε την ανάλυση, υποθέτουμε ότι η εικόνα ενδιαφέροντος είναι ένα μονοδιάστατο σήμα μήκους N, $\boldsymbol{\lambda} = \big[\lambda(0), \lambda(1), \ldots, \lambda(N-1)\big]$, όπου $\lambda(k)$, $k = 0, \ldots, N-1$ αναφέρεται στην πραγματική ένταση φωτεινότητας σε κάθε ξεχωριστό εικονοστοιχείο (pixel). Η συγκεκριμένη αναπαράσταση της εικόνας ως 1-Δ σήμα μπορεί να επιτευχθεί, για παράδειγμα, μέσω λεξικογραφικής διάταξης των τιμών των εικονοστοιχείων. Η εκτίμηση της έντασης της πηγαίας εικόνας $\boldsymbol{\lambda}$ βασίζεται στις θορυβώδεις μετρήσεις μας $\mathbf{x} = \big[x(0), x(1), \ldots, x(N-1)\big]$. Υιοθετώντας το βασικό μοντέλο σχηματισμού εικόνας, το διάνυσμα $\mathbf{x}$ αποτελείται από τα δείγματα παρατήρησης N τυχαίων μεταβλητών $X(k)$, οι οποίες δοθέντος του $\boldsymbol{\lambda}$, είναι ανεξάρτητες και κάθε μια ακολουθεί μια Poisson κατανομή με ένταση $\lambda(k)$, $X(k)|\lambda(k) \sim \text{Pois}\left(\lambda(k)\right)$. Ο ορισμός της Poisson κατανομής αλλά και οι ορισμοί όλων των κατανομών που χρησιμοποιούμε στο παρόν κεφάλαιο δίνονται στον Πίνακα 3.1. Στο Μπεϋζιανό πλαίσιο εργασίας το διάνυσμα $\boldsymbol{\lambda}$ δε θεωρείται πλέον ως ντετερμινιστικό, αλλά ως ένα διάνυσμα το οποίο περιέχει τις τιμές μια τυχαίας ακολουθίας $\boldsymbol{\Lambda}$, το συγκεκριμένο στιγμιότυπο της οποίας καλούμαστε να εκτιμήσουμε. Για να οδηγηθούμε σε ένα Μπεϋζιανό εκτιμητή για αυτό το πρόβλημα, θα πρέπει πρώτα να επιλέξουμε ένα μοντέλο πρότερης πιθανότητας, $p\left(\boldsymbol{\lambda}\right)$, για την τυχαία ακολουθία. Αφού καθορίσουμε το μοντέλο, μπορούμε να σχεδιάσουμε τεχνικές για την εύρεση ενός στατιστικά βέλτιστου εκτιμητή. Εάν επιλέξουμε ως κριτήριο ελαχιστοποίησης το μέσο τετραγωνικό σφάλμα (MSE), τότε θα καταλήξουμε στον εκτιμητή ελαχιστοποίησης μέσου τετραγωνικού σφάλματος (MMSE) [Kay93] ο οποίος δίνεται από τη σχέση:

$$\hat{\boldsymbol{\lambda}} = E\left\{\boldsymbol{\lambda}|\mathbf{x}\right\} = \int\limits_{\mathcal{R}^N} \boldsymbol{\lambda}\, p\left(\boldsymbol{\lambda}|\mathbf{x}\right)\, d\boldsymbol{\lambda}\,, \tag{3.2}$$

η οποία μπορεί να γραφεί εφαρμόζοντας τον κανόνα του Bayes ως[2]:

$$\hat{\boldsymbol{\lambda}} = \frac{\int \boldsymbol{\lambda}\, p\left(\mathbf{x}|\boldsymbol{\lambda}\right) p\left(\boldsymbol{\lambda}\right)\, d\boldsymbol{\lambda}}{\int p\left(\mathbf{x}|\boldsymbol{\lambda}\right) p\left(\boldsymbol{\lambda}\right)\, d\boldsymbol{\lambda}}\,. \tag{3.3}$$

Η περιθώρια κατανομή της θορυβώδους εικόνας $p\left(\mathbf{x}\right)$ που εμφανίζεται στον παρανομαστή της (3.3) ανήκει στην οικογένεια των πλεγμένων Poisson κατανομών [KX05]· η εκτίμηση των παραμέτρων για μια ενδιαφέρουσα υποκατηγορία αυτής της οικογενείας συζητείται στο [CLLT08]. Το Μπεϋζιανό πλαίσιο εργασίας είναι ελκυστικό για το πρόβλημα που καλούμαστε να αντιμετωπίσουμε, αλλά εμπεριέχει δύο σημαντικές προκλήσεις. Η πρώτη αφορά στον καθορισμό μιας κατάλληλης πολυδιάστατης πρότερης κατανομής για την ένταση της πηγαίας εικόνας, ενώ η δεύτερη αφορά στην

[2]Για λόγους απλότητας συμβολισμών δηλώνουμε το νόμο πιθανότητας που σχετίζεται είτε με μια συνεχή είτε με μία διακριτή τυχαία μεταβλητή X ως $p\left(x\right)$, και τον ερμηνεύουμε ανάλογα με την περίπτωση ως συνάρτηση πυκνότητας πιθανότητας ή ως συνάρτηση πιθανότητας.

Πίνακας 3.1: Ορισμοί των κατανομών των πιθανοτήτων που θα χρησιμοποιηθούν στο Κεφ. 3. Η συνάρτηση γάμα ορίζεται ως $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \mathrm{d}t$, D είναι η διάσταση των τυχαίων διανυσμάτων, και $n = \sum_{k=1}^{D} x_k$.

Μονοδιάστατες Κατανομές	
Poisson	$\mathrm{Pois}\,(x\|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$
Gamma	$\mathcal{G}\,(\lambda\|\alpha,\beta) = \lambda^{\alpha-1}\frac{e^{-\lambda/\beta}}{\beta^\alpha \Gamma(\alpha)}$
Binomial	$\mathrm{Bin}\,(x\|n,\theta) = \binom{n}{x}\theta^x\,(1-\theta)^{n-x}$
Beta	$\mathrm{Beta}\,(\theta\|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}\,(1-\theta)^{\beta-1}$

Πολυδιάστατες Κατανομές	
Multinomial	$\mathrm{Mult}\,(\mathbf{x}\|n,\boldsymbol{\theta}) = \frac{n!}{\prod_{k=1}^{D} x_k!} \cdot \prod_{k=1}^{D} \theta_k^{x_k}$
Dirichlet	$\mathrm{Dir}\,(\boldsymbol{\theta}\|\boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{k=1}^{D}\alpha_k\right)}{\prod_{k=1}^{D}\Gamma(\alpha_k)} \cdot \prod_{k=1}^{D}\theta_k^{\alpha_k-1}$
Polya	$\mathrm{Polya}\,(\mathbf{x}\|n,\boldsymbol{\alpha}) = \frac{n!}{\prod_{k=1}^{D} x_k!} \cdot \frac{\Gamma\left(\sum_{k=1}^{D}\alpha_k\right)}{\Gamma\left(n+\sum_{k=1}^{D}\alpha_k\right)} \cdot \prod_{k=1}^{D}\frac{\Gamma(x_k+\alpha_k)}{\Gamma(\alpha_k)}$

επίλυση του προβλήματος εκτίμησης σε ένα πολυ-διάστατο χώρο. Ειδικά το δεύτερο πρόβλημα είναι εξαιρετικά απαιτητικό και δύσκολο να αντιμετωπιστεί.

Προκειμένου να ξεπεραστούν τα παραπάνω προβλήματα, πρόσφατα προτάθηκε μια πολυκλιμακωτή ανάλυση Poisson στοχαστικών διαδικασιών [TN99, Kol99a] η οποία οδηγεί σε μια εφικτή λύση για το πρόβλημα της Μπεϋζιανής εκτίμησης. Στη συνέχεια της παρούσας ενότητας μελετούμε το πολυκλιμακωτό πλαίσιο εργασίας, το οποίο είναι κατάλληλο για μονοδιάστατα σήματα, και εξετάζουμε τους δυνατούς τρόπους επέκτασής του στην περίπτωση των εικόνων.

3.3.2 Παραγοντοποίηση Πιθανοφάνειας 1-Δ Σημάτων και Στατιστική Μοντελοποίηση

Αν δηλώσουμε με $\mathbf{x}_0 = \mathbf{x}$ και $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}$ τις αναπαραστάσεις των $\mathbf{x}$ και $\boldsymbol{\lambda}$, αντιστοίχως στην πιο λεπτομερή κλίμακα ανάλυσης, τότε μπορούμε να πετύχουμε μια πολυκλιμακωτή ανάλυσή τους μέσω των ακόλουθων επαναληπτικών εξισώσεων που προσομοιάζουν σε αυτές που οδηγούν στους μη-κανονικοποιημένους Haar συντελεστές κλιμάκωσης [TN99, Kol99a]:

$$x_j(k) = x_{j-1}(2k) + x_{j-1}(2k+1) \tag{3.4}$$

$$\lambda_j(k) = \lambda_{j-1}(2k) + \lambda_{j-1}(2k+1) \tag{3.5}$$

για $j = 1, \ldots, J$, $J = \log_2(N)$, και $k = 0, \ldots, N/2^j - 1$. Στις παραπάνω εξισώσεις το j δηλώνει την κλίμακα ανάλυσης (J είναι η πιο αδρή κλίμακα) και το k τη θέση στο αντίστοιχο διάνυσμα $\mathbf{x}_j$. Για λόγους απλοποίησης υποθέτουμε ότι το μήκος N του σήματος είναι δύναμη του δύο· σε αντίθετη περίπτωση η ίδια ανάλυση θα ισχύει και πάλι προσθέτοντας στο τέλος του σήματος μηδενικά (zero padding), ώστε το μήκος του να γίνει ίσο με δύναμη του δύο. Αυτή η διαδικασία παραγοντοποίησης του σήματος $\mathbf{x}$ υποκινείται από δύο βασικές ιδιότητες των Poisson διαδικασιών [BS00]: (1) Οι Poisson μετρήσεις σε μη-επικαλυπτόμενα χωρία είναι ανεξάρτητες δοθεισών των γενεσιουργών εντάσεων, και (2) το άθροισμα ανεξάρτητων Poisson τυχαίων μεταβλητών παραμένει Poisson κατανεμημένη. Έτσι λοιπόν, η τυχαία μεταβλητή $X_j(k)$, που προκύπτει ως το άθροισμα των Poisson

τυχαίων μεταβλητών $X_{j-1}(2k)$ και $X_{j-1}(2k+1)$, θα παραμείνει Poisson κατανεμημένη με ένταση $\lambda_j(k)$. Επιπλέον, ισχύει ότι για δύο ανεξάρτητες Poisson τυχαίες μεταβλητές, $X|\lambda_x \sim \text{Pois}(\lambda_x)$ και $Y|\lambda_y \sim \text{Pois}(\lambda_y)$, η δεσμευμένη κατανομή της X, δοθείσης της $X+Y$, είναι διωνυμική, δηλαδή $p(x|x+y,\lambda_x,\lambda_y) = \text{Bin}\left(x|x+y, \frac{\lambda_x}{\lambda_x+\lambda_y}\right)$ [BN03]. Η στατιστική αυτή σχέση μεταξύ γειτονικών κλιμάκων μας επιτρέπει την ακόλουθη παραγοντοποίηση της συνάρτησης πιθανοφάνειας μιας Poisson στοχαστικής διαδικασίας [Kol99a]:

$$p(\mathbf{x}|\boldsymbol{\lambda}) \;=\; \prod_{k=0}^{N-1} p\left(x_0(k)|\lambda_0(k)\right) \;=\; p\left(x_J(0)|\lambda_J(0)\right) \prod_{j=1}^{J} \prod_{k=0}^{\frac{N}{2^j}-1} \text{Bin}\left(x_{j-1}(2k)|x_j(k),\theta_j(k)\right) \;, \quad (3.6)$$

όπου $x_J(0)$ είναι ο συνολικός αριθμός των μετρήσεων έντασης φωτεινότητας της παρατηρούμενης εικόνας $\mathbf{x}$, $\lambda_J(0)$ είναι το συνολικό άθροισμα της έντασης της πηγαίας/καθαρής εικόνας $\boldsymbol{\lambda}$, και τα $\theta_j(k) \triangleq \frac{\lambda_{j-1}(2k)}{\lambda_j(k)}$ εκφράζουν το ποσοστό επιτυχίας της διωνυμικής κατανομής. Το σημαντικό ενδιαφέρον που παρουσιάζει η σχέση (3.6) είναι ότι κατά την παραγοντοποίηση της πιθανοφάνειας εμφανίζονται οι συντελεστές διαχωρισμού $\theta_j(k)$ και κλιμάκωσης $\lambda_J(0)$ που θα λάμβανε κανείς, αν εφάρμοζε τον πολυκλιμακωτό πολλαπλασιαστικό μετασχηματισμό (2.26) της Ενοτ. 2.2.2 στην πηγαία εικόνα $\boldsymbol{\lambda}$.

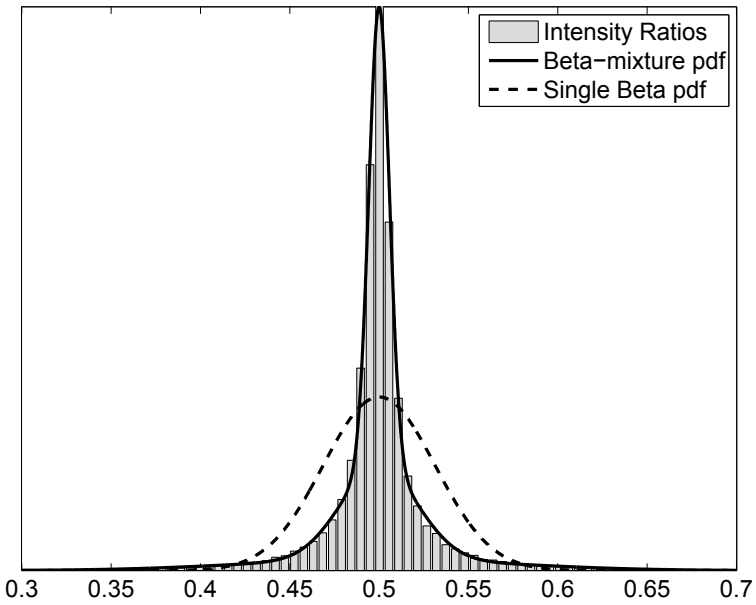
Για να συνεχίσουμε με το πολυκλιμακωτό Μπεϋζιανό πλαίσιο εργασίας, πρέπει επιπλέον να εκφράσουμε την πρότερη κατανομή $p(\boldsymbol{\lambda})$ σε μια παραγοντοποιημένη μορφή. Αναλύοντας το διάνυσμα έντασης $\boldsymbol{\lambda}$ με το 1-Δ πολλαπλασιαστικό μετασχηματισμό, οδηγούμαστε σε μια ισοδύναμη αναπαράσταση $\boldsymbol{\lambda}_e = \left[\lambda_J(0), \theta_J(0), \ldots, \theta_1(0)\ldots, \theta_1(N/2-1)\right]$. Θεωρώντας το $\lambda_J(0)$ και τα $\theta_j(k)$ ως τις παρατηρούμενες τιμές των τυχαίων μεταβλητών $\Lambda_J(0)$ και $\Theta_j(k)$, αντιστοίχως, και, επιπλέον, υποθέτοντας στατιστική ανεξαρτησία μεταξύ αυτών των μεταβλητών, τότε είναι δυνατόν να εκφράσουμε την πρότερη κατανομή για το $\boldsymbol{\lambda}_e$ στην παραγοντοποιημένη μορφή

$$p(\boldsymbol{\lambda}_e) = p\left(\lambda_J(0)\right) p(\boldsymbol{\theta}) = p\left(\lambda_J(0)\right) \prod_{j=1}^{J} \prod_{k=0}^{\frac{N}{2^j}-1} p\left(\theta_j(k)\right) \;. \quad (3.7)$$

Για να οδηγηθούμε σε πλήρη καθορισμό της πρότερης κατανομής (3.7), απαιτείται ο προσδιορισμός των όρων $p\left(\theta_j(k)\right)$. Λόγω του ότι η βήτα κατανομή είναι συζυγής με τη διωνυμική [Bis06], είναι πλεονεκτικό να συμπληρώσουμε τους διωνυμικούς όρους στη συνάρτηση πιθανοφάνειας (3.6) με μια πρότερη πιθανότητα μείγματος βήτα κατανομών

$$p\left(\theta_j(k)\right) = \sum_{m=1}^{M} \pi_{j,m} \,\text{Beta}\left(\theta_j(k)|\alpha_{j,m},\beta_{j,m}\right) \;, \quad (3.8)$$

όπου $\pi_{j,m}$ είναι το βάρος που ανατίθεται στην mοστή συνιστώσα του μείγματος για την jοστή κλίμακα ανάλυσης, τα $\alpha_{j,m}$ και $\beta_{j,m}$ είναι οι παράμετροι της συγκεκριμένης βήτα συνιστώσας και M είναι ο συνολικός αριθμός από τις συνιστώσες του μείγματος που χρησιμοποιούνται σε κάθε κλίμακα. Η καταλληλότητα της βήτα κατανομής εκτός από τη συζυγία που παρουσιάζει με τη διωνυμική κατανομή στηρίζεται και στο γεγονός ότι διαθέτει το ίδιο πεδίο ορισμού, διάστημα $[0\,1]$, με τους συντελεστές $\theta_j(k)$. Η χρήση στην Εξ. (3.8) ενός μείγματος από βήτα κατανομές αντί για μία μόνο συνιστώσα μας επιτρέπει να μοντελοποιήσουμε με μεγαλύτερη ακρίβεια τα στατιστικά της εικόνας. Η αυξημένη ακρίβεια αυτής της προσέγγισης απεικονίζεται στο Σχ. 3.1, όπου το ιστόγραμμα του λόγου των εντάσεων μιας καθαρής εικόνας μοντελοποιείται με μια συνιστώσα και ένα μείγμα βήτα κατανομών. Ένας διαφορετικός τρόπος για να δικαιολογήσουμε τη μοντελοποίηση των τυχαίων μεταβλητών $\Theta_j(k)$ με μείγμα βήτα κατανομών είναι να υποθέσουμε ότι σε κάθε



Σχήμα 3.1: Ιστόγραμμα των συντελεστών λόγου έντασης $\theta_1^{(2)}(k)$ της πηγαίας εικόνας Lena, δες Εξ. (2.30), μαζί με ταιριασμένες μια (διακεκομμένη γραμμή) και τρεις (συμπαγής γραμμή) συνιστώσες από ένα μείγμα συμμετρικών ($\alpha_{jm} = \beta_{jm}$) βήτα κατανομών. Οι παράμετροι του μοντέλου προκύπτουν μεγιστοποιώντας την πιθανοφάνεια· στην περίπτωση των πολλαπλών συνιστωσών γίνεται χρήση του ΕΜ αλγορίθμου.

κλίμακα ανάλυσης, κάθε τυχαία μεταβλητή $\Lambda_j(k)$ υπακούει μια γάμα κατανομή, όπου η επιλογή της γάμα οφείλεται στο ότι είναι συζυγής με την Poisson [BS00]. Συγκεκριμένα, εάν ισχύει $\Lambda_{j-1}(2k) \sim \mathcal{G}(\alpha_j, \delta)$ και $\Lambda_{j-1}(2k+1) \sim \mathcal{G}(\beta_j, \delta)$, τότε η τυχαία μεταβλητή $\Theta_j(k) = \frac{\Lambda_{j-1}(2k)}{\Lambda_j(k)}$ θα υπακούει την κατανομή Beta (α_j, β_j) [BN03]. Εκτός από τους παραπάνω ισχυρισμούς που χρησιμοποιήσαμε και οι οποίοι οδηγούν σε υπολογιστικά εφικτή λύση του προβλήματος, η υιοθέτηση μιας τέτοιας πολυκλιμακωτής πρότερης πιθανότητας στην εικόνα οδηγεί σε μια δομή συσχέτισης της έντασής της που αντιστοιχεί σε στοχαστικές διαδικασίες με περίπου $1/f$ φασματική συμπεριφορά [TN99]. Τέτοιες διαδικασίες συνδέονται με fractal περιγραφές και μπορούν να συλλάβουν συγκεκριμένες βασικές ιδιότητες των φυσικών εικόνων, όπως είναι οι μεγάλου εύρους εξαρτήσεις της έντασης.

3.3.3 Παραγοντοποίηση Πιθανοφάνειας Εικόνων και Στατιστική Μοντελοποίηση

Το στατιστικό πλαίσιο που περιγράψαμε στην Ενοτ. 3.3.2 και περιλαμβάνει την πολυκλιμακωτή παραγοντοποίηση της συνάρτησης πιθανοφάνειας και της πρότερης πιθανότητας είναι κατάλληλο για μονοδιάστατα σήματα. Για την περίπτωση που μας ενδιαφέρει, τα $\mathbf{x}$ και $\boldsymbol{\lambda}$ θα είναι 2-Δ εικόνες οι οποίες περιλαμβάνουν τις διακριτές παρατηρήσεις και τις πραγματικές τιμές της συνάρτησης έντασης φωτεινότητας, αντιστοίχως. Έτσι, προκειμένου να οδηγηθούμε σε αντίστοιχες σχέσεις με αυτές της Ενοτ. 3.3.2, αρχικά υιοθετούμε μια 2-Δ πολυκλιμακωτή ανάλυση των $\mathbf{x}_0 = \mathbf{x}$ και $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}$ που εκφράζεται μέσα από τις ακόλουθες αναδρομικές εξισώσεις:

$$x_j(k,\ell) = \sum_{(k',\ell')\in C_{k,\ell}} x_{j-1}(k',\ell') \tag{3.9}$$

$$\lambda_j(k,\ell) = \sum_{(k',\ell')\in C_{k,\ell}} \lambda_{j-1}(k',\ell')\,, \tag{3.10}$$

για $j = 1,\dots J$, $k = 0,\dots,N_1/2^j - 1$, $\ell = 0,\dots,N_2/2^j - 1$ και $J = \min\{\log_2(N_1), \log_2(N_2)\}$, με το $C_{k,\ell}$ να ορίζεται στην Εξ. (2.29). Ο τρόπος με τον οποίο πραγματοποιείται η πολυκλιμακωτή αυτή ανάλυση μας επιτρέπει να αντιστοιχίσουμε τις τιμές $x_j(k,\ell)$ στους κόμβους ενός τετραδικού δένδρου όπως του Σχ. 3.2. Ο συντελεστής της πιο αδρής κλίμακας $x_J(0,0)$, τοποθετείται στη ρίζα του δένδρου, ενώ οι συντελεστές της πιο λεπτομερούς κλίμακας στα κλαδιά του δένδρου. Η ρίζα του δένδρου αλλά και κάθε ενδιάμεσος κόμβος πλην των κλαδιών διαθέτει τέσσερα παιδιά. Με αυτή την αναπαράσταση μπορούμε να υπολογίσουμε τους συντελεστές μιας κλίμακας j ως το άθροισμα των τιμών των κόμβων-παιδιών τα οποία βρίσκονται στην κλίμακα $j-1$ του τετραδικού δένδρου. Όμοια με τη 1-Δ παραγοντοποίηση, λόγω του ότι οι τυχαίες μεταβλητές $X_{j-1}(k',l')$, με $(k',\ell') \in C_{k,\ell}$, είναι υπό συνθήκη ανεξάρτητες, οι τυχαίες μεταβλητές $X_j(k,\ell)$ παραμένουν κατά Poisson κατανεμημένες με γενεσιουργές εντάσεις $\lambda_j(k,\ell)$. Επιπλέον, είναι εύκολο να δείξει κανείς ότι και σε αυτή την περίπτωση εμφανίζονται στατιστικές εξαρτήσεις μεταξύ των συντελεστών κλιμάκωσης δύο γειτονικών επιπέδων του δένδρου. Συγκεκριμένα, δοθείσης της $X_j(k,\ell)$, η από κοινού υπό συνθήκη πιθανότητα του τυχαίου διανύσματος των τεσσάρων απογόνων (παιδιά) $\mathbf{X}_j^c(k,\ell) = \{X_{j-1}(k',\ell') : (k',\ell') \in C_{k,\ell}\}$ είναι πολυωνυμική [BS00]:

$$p\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell),\boldsymbol{\theta}_j(k,\ell)\right) = \text{Mult}\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell),\boldsymbol{\theta}_j(k,\ell)\right)\,, \tag{3.11}$$

με

$$\boldsymbol{\theta}_j(k,\ell) = \left\{\frac{\lambda_{j-1}(k',\ell')}{\lambda_j(k,\ell)} : \left(k',\ell'\right) \in C_{k,\ell}\right\}\,. \tag{3.12}$$

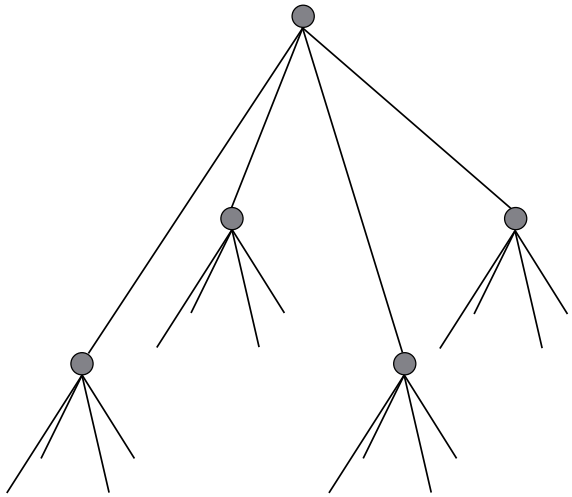
Το διάνυσμα $\boldsymbol{\theta}_j(k,\ell)$ εκφράζει το ποσοστό επιτυχίας της πολυωνυμικής κατανομής.

Οι στατιστικές σχέσεις που συνδέουν τις παρατηρήσεις $X_j(k,\ell)$ μεταξύ δύο γειτονικών κλιμάκων επιτρέπουν η συνάρτηση πιθανοφάνειας της εικόνας $\mathbf{x}$ να παραγοντοποιηθεί ως εξής (δες το Παράρτημα 3.Α' για την απόδειξη):

$$p\left(\mathbf{x}|\boldsymbol{\lambda}\right) = p\left(x_J(0,0)|\lambda_J(0,0)\right)\prod_{j=1}^{J}\prod_{k=0}^{\frac{N_1}{2^j}-1}\prod_{\ell=0}^{\frac{N_2}{2^j}-1} p\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell),\boldsymbol{\theta}_j(k,\ell)\right)\,. \tag{3.13}$$

Τετραδικός Πολλαπλασιαστικός Μετασχηματισμός Αν στην Εξ. (3.13) αντικαταστήσουμε τους όρους $p\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell),\boldsymbol{\theta}_j(k,\ell)\right)$ με αυτούς της Εξ. (3.11), τότε παρατηρούμε ότι κατά την παραγοντοποίηση της πιθανοφάνειας της εικόνας παρατήρησης $\mathbf{x}$, εμφανίζονται οι συντελεστές κλιμάκωσης $\boldsymbol{\lambda}_J(0,0)$ και διαχωρισμού $\boldsymbol{\theta}_j(k,\ell)$, τους οποίους θα λάμβανε κανείς, αν ανέλυε την πηγαία εικόνα $\boldsymbol{\lambda}$ σε πολλαπλές κλίμακες εφαρμόζοντας το 2-Δ τετραδικό πολλαπλασιαστικό μετασχηματισμό που παρουσιάσαμε στην Ενοτ. 2.3.2.

Όπως και στη 1-Δ περίπτωση, έτσι και εδώ, προκειμένου να μπορέσουμε να αξιοποιήσουμε το πολυκλιμακωτό Μπεϋζιανό πλαίσιο, θα πρέπει να χρησιμοποιήσουμε μια παραγοντοποιημένη πολυκλιμακωτή πρότερη πιθανότητα για το αναπαραμετροποιημένο διάνυσμα έντασης $\boldsymbol{\lambda}_e$. Αυτό είναι εφικτό θεωρώντας και πάλι ότι τα διανύσματα των συντελεστών $\boldsymbol{\theta}_j(k,\ell)$ της Εξ. (3.12) είναι



Σχήμα 3.2: Αντιστοίχιση των τιμών των συντελεστών $\lambda_j(k,\ell)$ και $x_j(k,\ell)$ στους κόμβους ενός τετραδικού δένδρου. Κάθε κόμβος του δένδρου, εκτός των κλαδιών, συνδέεται με τέσσερις κόμβους του αμέσως επόμενου επιπέδου.

αμοιβαία ανεξάρτητα, οπότε καταλήγουμε:

$$p\left(\boldsymbol{\lambda}_e\right) = p\left(\lambda_J(0,0)\right) \prod_{j=1}^{J} \prod_{k=0}^{\frac{N_1}{2^j}-1} \prod_{\ell=0}^{\frac{N_2}{2^j}-1} p\left(\boldsymbol{\theta}_j(k,\ell)\right) \ . \tag{3.14}$$

Για τη στατιστική μοντελοποίηση των τυχαίων διανυσμάτων $\boldsymbol{\Theta}_j(k,\ell)$, επειδή η κατανομή Dirichlet είναι συζυγής με την πολυωνυμική [Bis06], υιοθετούμε ένα μείγμα από Dirichlet συνιστώσες πρότερης πιθανότητας

$$p\left(\boldsymbol{\theta}_j(k,\ell)\right) = \sum_{m=1}^{M} \pi_{j,m}\, \mathrm{Dir}\left(\boldsymbol{\theta}_j(k,\ell)|\boldsymbol{\alpha}_{j,m}\right) \ , \tag{3.15}$$

όπου $\boldsymbol{\alpha}_{j,m}$ είναι το διάνυσμα παραμέτρων της mοστής συνιστώσας του μείγματος στην jοστή κλίμακα ανάλυσης. Ένας εναλλακτικός τρόπος δικαιολόγησης της επιλογής μας αυτής παρέχεται, αν λάβουμε υπόψιν μας ότι η Dirichlet, όπως και η βήτα κατανομή, επιδέχεται αναπαράστασης με όρους γάμα τυχαίων μεταβλητών. Μια παρόμοια στρατηγική μοντελοποίησης εξετάστηκε στο [ECKVD04] υπό το πρίσμα του προβλήματος της αποσυνέλιξης (deblurring)· παρόλα αυτά, οι συγγραφείς στο [ECKVD04] θεώρησαν την απλούστερη περίπτωση μιας πρότερης κατανομής η οποία μοντελοποιείται από μια μόνο Dirichlet συνιστώσα.

Το ερώτημα που τίθεται τώρα είναι κατά πόσο είναι δυνατή η πολυκλιμακωτή ανάλυση της εικόνας παρατήρησης $\mathbf{x}$ με διαφορετικούς τρόπους οι οποίοι θα εμφανίζουν στενή σχέση με την ανάλυση της πηγαίας εικόνας $\boldsymbol{\lambda}$ είτε με το 2-Δ διαχωρίσιμο είτε με τον Poisson-Haar πολλαπλασιαστικό μετασχηματισμό (δες Ενοτ. 2.3.1 και 2.3.3). Η αναζήτηση μιας τέτοιας δυνατότητας είναι καθοριστικής σημασίας, αφού θα επιτρέψει τη χρήση εναλλακτικών αναπαραστάσεων για την πηγαία εικόνα $\boldsymbol{\lambda}$, που με τη σειρά τους μπορούν να οδηγήσουν σε καλύτερη στατιστική μοντελοποίηση των δομικών χαρακτηριστικών της. Στην επόμενη Ενότητα μελετούμε πώς είναι δυνατό κάτι τέτοιο να επιτευχθεί για το διαχωρίσιμο και στη συνέχεια για τον Poisson-Haar μετασχηματισμό.

2-Δ Διαχωρίσιμος Πολλαπλασιαστικός Μετασχηματισμός Έστω ότι με $X_i,\ i = 1 \ldots 4$ συμβολίζουμε τέσσερις ανεξάρτητες Poisson τυχαίες μεταβλητές που χαρακτηρίζονται από τις αντίστοιχες

τιμές έντασης λ_i. Το άθροισμα αυτών των τυχαίων μεταβλητών, όπως αναφέραμε και προηγουμένως, θα είναι επίσης μια Poisson τυχαία μεταβλητή X με παράμετρο λ. Αν λάβουμε υπόψιν μας την παρακάτω ισοδυναμία (δες στο Παράρτημα 3.Β' για την απόδειξη):

$$\begin{aligned}&\text{Mult}\left(x_1, x_2, x_3, x_4 | x, \boldsymbol{\theta}\right) = \\ &\quad \text{Bin}\left(x_1 + x_2 | x, \frac{\lambda_1 + \lambda_2}{\lambda}\right) \text{Bin}\left(x_1 | x_1 + x_2, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right) \text{Bin}\left(x_3 | x_3 + x_4, \frac{\lambda_3}{\lambda_3 + \lambda_4}\right) ,\end{aligned} \tag{3.16}$$

διαπιστώνουμε ότι μπορούμε να ακολουθήσουμε μια εναλλακτική πολυκλιμακωτή ανάλυση της εικόνας $\mathbf{x}$ σε σχέση με αυτή που ακολουθήσαμε προηγουμένως. Με βάση τις Εξ. (3.11) και (3.16) μπορούμε να εκφράσουμε τους όρους $p\left(\mathbf{x}_j^c(k,\ell) | x_j(k,\ell), \boldsymbol{\theta}_j(k,\ell)\right)$ της συνάρτησης πιθανοφάνειας (3.13) για την εικόνα $\mathbf{x}$ ως εξής:

$$\begin{aligned}p\left(\mathbf{x}_j^c(k,\ell) | x_j(k,\ell), \boldsymbol{\theta}_j(k,\ell)\right) &= \text{Bin}\left(x_j^c(k,\ell) | x_j(k,\ell), \theta_j^{(1)}(k,\ell)\right) \\ &\times \text{Bin}\left(x_{j-1}(2k, 2\ell) | x_j^{c_1}(k,\ell), \theta_j^{(2)}(k,\ell)\right) \text{Bin}\left(x_{j-1}(2k+1, \ell) | x_j^{c_2}(k,\ell), \theta_j^{(3)}(k,\ell)\right)\end{aligned} \tag{3.17}$$

όπου οι παράμετροι $\left\{\theta_j^{(1)}, \theta_j^{(2)}, \theta_j^{(3)}\right\}$ ταυτίζονται με τους συντελεστές διαχωρισμού (2.30) του διαχωρίσιμου πολλαπλασιαστικού μετασχηματισμού και

$$\begin{aligned}x_j^{c_1}(k,\ell) &= x_{j-1}(2k, 2\ell) + x_{j-1}(2k, 2\ell+1) \\ x_j^{c_2}(k,\ell) &= x_{j-1}(2k+1, 2\ell) + x_{j-1}(2k+1, 2\ell+1) \\ x_j^{c}(k,\ell) &= x_j^{c_1}(k,\ell) + x_j^{c_2}(k,\ell) .\end{aligned} \tag{3.18}$$

Αν λοιπόν αναλύσουμε την εικόνα $\mathbf{x}$ σύμφωνα με τις Εξ. (3.18), τότε είναι δυνατόν να οδηγηθούμε σε μια παραγοντοποιημένη μορφή της συνάρτησης πιθανοφάνειας που περιλαμβάνει ως παραμέτρους τους συντελεστές κλιμάκωσης και διαχωρισμού οι οποίοι προκύπτουν κατά την πολυκλιμακωτή ανάλυση της $\boldsymbol{\lambda}$ με το διαχωρίσιμο σχήμα. Για τη στατιστική μοντελοποίηση των τυχαίων μεταβλητών $\Theta_j^{(i)}(k,\ell)$ θα χρησιμοποιήσουμε, για τους ίδιους λόγους με τη μονοδιάστατη περίπτωση, μείγμα από βήτα συνιστώσες.

Poisson-Haar Μετασχηματισμός Ένας τρίτος εναλλακτικός τρόπος ανάλυσης της εικόνας $\mathbf{x}$ προκύπτει μέσω της εφαρμογής των παρακάτω αναδρομικών εξισώσεων:

$$\begin{aligned}x_j(k,\ell) &= \sum_{(k',\ell') \in C_{k,\ell}} x_{j-1}(k',\ell') \\ x_j^h(k,\ell) &= x_{j-1}(2k, 2\ell) + x_{j-1}(2k, 2\ell+1) \\ x_j^v(k,\ell) &= x_{j-1}(2k, 2\ell) + x_{j-1}(2k+1, 2\ell) \\ x_j^d(k,\ell) &= x_{j-1}(2k, 2\ell) + x_{j-1}(2k+1, 2\ell+1) .\end{aligned} \tag{3.19}$$

Επειδή σε κάθε κλίμακα οι συντελεστές $x_j(k,\ell)$ και $x_j^o(k,\ell)$ για $o = \{h, v, d\}$ προκύπτουν ως αθροίσματα ανεξάρτητων Poisson τυχαίων μεταβλητών (Poisson μετρήσεις της αμέσως προηγούμενης κλίμακας ανάλυσης), θα διατηρούν και αυτές με τη σειρά τους τα Poisson στατιστικά. Επιπλέον, λόγω της σχέσης που συνδέει δυο Poisson τυχαίες μεταβλητές, οι υπό συνθήκη πιθανότητες $p\left(x_j^o(k,\ell) | x_j(k,\ell)\right)$ θα είναι δυωνυμικές

$$p\left(x_j^o(k,\ell) | x_j(k,\ell)\right) = \text{Bin}\left(x_j^o(k,\ell) | x_j(k,\ell), \theta_j^o(k,\ell)\right) , \tag{3.20}$$

όπου τα $\theta_j^o(k,\ell)$ αντιστοιχούν στους συντελεστές διαχωρισμού της Εξ. (2.32) που παράγονται κατά την πολυκλιμακωτή ανάλυση της πηγαίας εικόνας $\boldsymbol{\lambda}$ με τον Poisson-Haar μετασχηματισμό. Αυτό που απομένει τώρα είναι να επιλέξουμε μια στατιστική μοντελοποίηση της πηγαίας εικόνας $\boldsymbol{\lambda}$. Εφαρμόζοντας τον Poisson-Haar μετασχηματισμό, οδηγούμαστε σε μια ισοδύναμη διανυσματική αναπαράσταση για το διάνυσμα έντασης της πηγαίας εικόνας $\boldsymbol{\lambda}$, της μορφής:

$$\boldsymbol{\lambda}_e = \left[\lambda_J(0,0), \boldsymbol{\theta}_J(0,0), \ldots, \boldsymbol{\theta}_1(0,0), \ldots, \boldsymbol{\theta}_1(N_1/2-1, N_2/2-1)\right] , \tag{3.21}$$

όπου τα διανύσματα $\boldsymbol{\theta}$ είναι της μορφής: $\boldsymbol{\theta}_j(k,\ell) = \left(\theta_j^h(k,\ell), \theta_j^v(k,\ell), \theta_j^d(k,\ell)\right)$. Σε αυτή την περίπτωση κάθε στοιχείο του $\boldsymbol{\theta}_j(k,\ell)$ μπορεί να θεωρηθεί ως στιγμιότυπο μιας εκ των τριών ανεξάρτητων τυχαίων μεταβλητών $\Theta_j^h(k,\ell)$, $\Theta_j^v(k,\ell)$ και $\Theta_j^d(k,\ell)$. Αυτές οι μεταβλητές μπορούν να μοντελοποιηθούν ως μείγμα βήτα κατανομών (3.8) ακολουθώντας κανείς την ίδια λογική με αυτή που χρησιμοποιήσαμε για το διαχωρίσιμο και το 1-Δ σχήμα ανάλυσης. Η πρότερη πιθανότητα της πηγαίας εικόνας, όμοια με το τετραδικό σχήμα, επιδέχεται την πολυκλιμακωτή παραγοντοποίηση της Εξ. (3.14), με

$$p\left(\boldsymbol{\theta}_j(k,\ell)\right) = p(\theta_j^h(k,\ell)) p\left(\theta_j^v(k,\ell)\right) p(\theta_j^d(k,\ell)) . \tag{3.22}$$

Η επιλογή μιας ανεξάρτητης πρότερης πιθανότητας, $p\left(\boldsymbol{\theta}\right)$, για κάθε ένα ξεχωριστικό προσανατολισμό μπορεί να θεωρηθεί ανάλογη με το χειρισμό των κυματιδιακών συντελεστών λεπτομέρειας και συνιστά μια συνήθη υπόθεση που υιοθετείται στη βιβλιογραφία των κυματιδίων [RCB01, CB01].

3.4 Μπεϋζιανή Εκτίμηση Έντασης Φωτεινότητας

Στην παρούσα ενότητα οδηγούμαστε σε ένα στατιστικά βέλτιστο Μπεϋζιανό εκτιμητή για το αναπαραμετροποιημένο διάνυσμα $\boldsymbol{\lambda}_e$ της έντασης φωτεινότητας της πηγαίας εικόνας. Αυτός ο εκτιμητής ελαχιστοποιεί το Μπεϋζιανό κριτήριο του MSE [Kay93] στο μετασχηματισμένο πεδίο της έντασης. Ωστόσο, αν και ο μετασχηματισμός που εφαρμόζουμε είναι μη-γραμμικός, ο προκύπτων MMSE εκτιμητής είναι ισοδύναμος με τον εκτιμητή που θα λαμβάναμε, αν εργαζόμασταν στο αρχικό πεδίο ορισμού της εικόνας (δες το Παράρτημα 3.Γ' για την απόδειξη). Έχοντας εκτιμήσει το $\boldsymbol{\lambda}_e$, μπορούμε ευθέως να ανακτήσουμε το διάνυσμα έντασης $\boldsymbol{\lambda}$ της πηγαίας εικόνας μέσω μιας αναδρομικής διαδικασίας κατά μήκος των κλιμάκων ανάλυσης, όπως αυτή που ορίσαμε στην Εξ. (2.34) για το προτεινόμενο Poisson-Haar σχήμα ανάλυσης (ανάλογες αναδρομικές διαδικασίες ισχύουν για το 1-Δ πολλαπλασιαστικό αλλά και για το 2-Δ διαχωρίσιμο και τετραδικό σχήμα). Αξίζει να σημειώσουμε ότι θα μπορούσαμε να καταλήξουμε και σε άλλου τύπου εκτιμητές υπό το συγκεκριμένο πλαίσιο εργασίας υιοθετώντας εναλλακτικά κριτήρια βελτιστοποίησης, όπως το κριτήριο μεγιστοποίησης της ύστερης πιθανότητας (maximum a posteriori probability - MAP).

3.4.1 Πολυκλιμακωτή Παραγοντοποίηση Ύστερης Πιθανότητας

Έχοντας παραγοντοποιήσει τις πρότερες πιθανότητες (3.7), (3.14), (3.22) και τις συναρτήσεις πιθανοφάνειας (3.6), (3.13) σε πολλαπλές κλίμακες, είναι απλό να δείξουμε ότι και η ύστερη πιθανότητα $p\left(\boldsymbol{\lambda}_e|\mathbf{x}\right) = p\left(\lambda_J(0,0)|x_J(0,0)\right) p\left(\boldsymbol{\theta}|\mathbf{x}\right)$ επιδέχεται πολυκλιμακωτής παραγοντοποίησης. Η δυνατότητα παραγοντοποίησης της ύστερης πιθανότητας υποδηλώνει ότι το πρόβλημα εκτίμησης της έντασης της πηγαίας εικόνας μπορεί να αντιμετωπιστεί ξεχωριστά σε κάθε κλίμακα και θέση αντί να απαιτείται η επίλυση ενός ενιαίου πολύπλοκου και πολυδιάστατου προβλήματος. Σε αυτό το σημείο, προκειμένου να καταλήξουμε στην ύστερη κατανομή, θα εκμεταλλευτούμε το γεγονός ότι η πρότερη κατανομή που έχουμε υιοθετήσει για τη μοντελοποίηση του $\boldsymbol{\lambda}_e$ είναι συζυγής με τη συνάρτηση πιθανοφάνειας.

Για λόγους πληρότητας, στη συνέχεια, θα υπολογίσουμε εκτός από την ύστερη πιθανότητα της αναπαραμετροποιημένης εικόνας $\boldsymbol{\lambda}_e$ και την ύστερη πιθανότητα του 1-Δ σήματος $\boldsymbol{\lambda}_e$, προκειμένου να οδηγηθούμε στις εκφράσεις των αντίστοιχων MMSE εκτιμητών. Συνδυάζοντας τις συναρτήσεις πιθανοφάνειας (3.6) και (3.13) και τις πρότερες πιθανότητες (3.7), (3.14), καταλήγουμε στις ύστερες κατανομές για το 1-Δ δυαδικό και 2-Δ τετραδικό σχήμα ανάλυσης που έχουν τις ακόλουθες μορφές:

$$p\left(\boldsymbol{\lambda}_e|\mathbf{x}\right) = p\left(\lambda_J(0)|x_J(0)\right) \prod_{j=1}^{J} \prod_{k=0}^{\frac{N}{2^j}-1} p\left(\theta_j(k)|x_j(k), x_{j-1}(2k)\right) \tag{3.23}$$

$$p\left(\boldsymbol{\lambda}_e|\mathbf{x}\right) = p\left(\lambda_J(0,0)|x_J(0,0)\right) \prod_{j=1}^{J} \prod_{k=0}^{\frac{N_1}{2^j}-1} \prod_{\ell=0}^{\frac{N_2}{2^j}-1} p\left(\boldsymbol{\theta}_j(k,\ell)|\mathbf{x}_j^c(k,\ell)\right) \ . \tag{3.24}$$

Κάνοντας χρήση της ιδιότητας (μπορεί εύκολα να επιβεβαιωθεί με απευθείας αντικαταστάσεις)

$$\text{Mult}\left(\mathbf{x}|n,\boldsymbol{\theta}\right) \text{Dir}\left(\boldsymbol{\theta}|\boldsymbol{\alpha}\right) = \text{Polya}\left(\mathbf{x}|n,\boldsymbol{\alpha}\right) \text{Dir}\left(\boldsymbol{\theta}|\mathbf{x}+\boldsymbol{\alpha}\right) \ , \tag{3.25}$$

όπου η κατανομή Polya [JKN97] (που επίσης καλείται Dirichlet-πολυωνυμική) ορίζεται στον Πίνακα 3.1, μπορούμε να εκφράσουμε τους ύστερους παράγοντες ως μείγμα από Dirichlet συνιστώσες:

$$p\left(\boldsymbol{\theta}_j(k,\ell)|\mathbf{x}_j^c(k,\ell)\right) = \sum_{m=1}^{M} \gamma_m(z_j(k,\ell)) \text{Dir}\left(\boldsymbol{\theta}_j(k,\ell)|\mathbf{x}_j^c(k,\ell)+\boldsymbol{\alpha}_{j,m}\right) \ . \tag{3.26}$$

Εδώ η $z_j(k,\ell)$ υποδηλώνει ποια συνιστώσα του μείγματος είναι υπεύθυνη για τη δημιουργία του διανύσματος παρατήρησης $\mathbf{x}_j^c(k,\ell)$, ενώ το αντίστοιχο βάρος για την ανάθεση της ύστερης συνιστώσας του μείγματος ισούται με

$$\gamma_m(z_j(k,\ell)) = \frac{\pi_{j,m} \text{Polya}\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell),\boldsymbol{\alpha}_{j,m}\right)}{\sum_{n=1}^{M} \pi_{j,n} \text{Polya}\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell),\boldsymbol{\alpha}_{j,n}\right)} \ . \tag{3.27}$$

Σημειώνουμε ότι οι εκφράσεις (3.26) και (3.27) καλύπτουν και το 1-Δ πολλαπλασιαστικό σχήμα (3.23)· στην περίπτωση αυτή η κατανομή Dirichlet μετατρέπεται σε βήτα κατανομή που αποτελεί τη διδιάστατη εκδοχή της.

Σε όλα τα υπό θεώρηση πολυκλιμακωτά σχήματα η ύστερη κατανομή της συνολικής έντασης της πηγαίας εικόνας $\lambda_J(0,0)$ μπορεί να μοντελοποιηθεί ως γάμα κατανομή, $p\left(\lambda_J(0,0)|x_J(0,0)\right) = \mathcal{G}\left(x_J(0,0)+\gamma, \frac{\delta}{\delta+1}\right)$, της οποίας η μέση τιμή είναι $\hat{\lambda}_J(0,0) = \left(x_J(0,0)+\gamma\right)\delta/(\delta+1)$. Το αποτέλεσμα αυτό προκύπτει αναθέτοντας μια γάμα ύστερη πιθανότητα στη συνολική ένταση της εικόνας, δηλαδή $\Lambda_J(0,0) \sim \mathcal{G}\left(\gamma,\delta\right)$, λόγω του ότι η γάμα κατανομή είναι συζυγής με την Poisson. Μια τέτοια προσέγγιση ακολουθείται από τους συγγραφείς του [NK00]. Ωστόσο, αυτή η μοντελοποίηση δεν είναι κρίσιμη για την αποδοτικότητα του τελικού εκτιμητή. Για μια 2-Δ Poisson στοχαστική διαδικασία (Poisson μέτρηση της εικόνας) στην jοστή κλίμακα ανάλυσης ο τοπικός σηματοθορυβικός λόγος ορίζεται ως ο λόγος της μέσης τιμής του εικονοστοιχείου προς την τυπική απόκλισή του και ισούται με $SNR = \sqrt{\lambda_j(k,\ell)}$. Καθώς προσεγγίζουμε τις πιο αδρές κλίμακες ανάλυσης, η τιμή της έντασης $\lambda_j(k,\ell)$ αυξάνεται και ταυτόχρονα αυξάνεται και ο σηματοθορυβικός λόγος. Συνεπώς, αφού οι τιμές των μετρήσεων στις πιο αδρές κλίμακες είναι πρακτικά μεγάλες, έχουμε επιλέξει να χρησιμοποιήσουμε ως μια καλή και αξιόπιστη εκτίμηση της συνολικής γενεσιουργού έντασης $\lambda_J(0,0)$, την παρατηρούμενη συνολική ένταση, δηλαδή $\hat{\lambda}_J(0,0) \approx x_J(0,0)$.

3.4.2 Μπεϋζιανή Εκτίμηση Μέσου Τετραγωνικού Σφάλματος

Γνωρίζοντας την ύστερη κατανομή μπορούμε πλέον να υπολογίσουμε τον ύστερο εκτιμητή (MMSE εκτιμητής) του αναπαραμετροποιημένου διανύσματος έντασης της πηγαίας εικόνας $\boldsymbol{\lambda}_e$. Αρχικά, θα υπολογίσουμε τον ύστερο εκτιμητή για το τετραδικό σχήμα και το 1-Δ πολλαπλασιαστικό σχήμα και, στη συνέχεια, θα παρουσιάσουμε τους αντίστοιχους εκτιμητές για το διαχωρίσιμο και το Poisson-Haar πολυκλιμακωτό σχήμα ανάλυσης, επειδή τα δύο τελευταία προκύπτουν ως ειδικές περιπτώσεις του πρώτου. Λόγω της παραγοντοποίησης της ύστερης κατανομής (3.24) και της υπόθεσης ανεξαρτησίας των συντελεστών του λόγου εντάσεων $\boldsymbol{\theta}$ στο μοντέλο πρότερης πιθανότητας της Εξ. (3.14), μπορούμε να βρούμε το βέλτιστο εκτιμητή για κάθε στοιχείο του $\boldsymbol{\theta}_j(k,\ell)$ ξεχωριστά ως

$$\hat{\theta}_j^i(k,\ell) = E\left\{\theta_j^i(k,\ell)|\mathbf{x}_j^c(k,\ell)\right\} = \int_{\mathcal{R}^D} \theta_j^i(k,\ell) p\left(\boldsymbol{\theta}_j(k,\ell)|\mathbf{x}_j^c(k,\ell)\right)\, d\boldsymbol{\theta}_j(k,\ell)$$

$$= \sum_{m=1}^{M} \gamma_m(z_j(k,\ell)) \left(\frac{x_j^{c,i}(k,\ell) + \alpha_{j,m}^i}{x_j(k,\ell) + \alpha_{j,m}^0}\right), \quad (3.28)$$

όπου το $\theta_j^i(k,\ell)$ είναι το iοστό στοιχείο του διανύσματος $\boldsymbol{\theta}_j(k,\ell)$ και $\alpha_{j,m}^0 = \sum_{t=1}^{D} \alpha_{j,m}^t$.

Για τη μονοδιάστατη ανάλυση ο MMSE εκτιμητής αποτελεί ειδική περίπτωση του (3.28)

$$\hat{\theta}_j(k) = \sum_{m=1}^{M} \gamma_m(z_j(k)) \left(\frac{x_{j-1}(2k) + \alpha_{j,m}}{x_j(k) + \alpha_{j,m} + \beta_{j,m}}\right). \quad (3.29)$$

όπου το βάρος $\gamma_m(z_j(k))$ θα ισούται με:

$$\gamma_m(z_j(k)) = \frac{\pi_{j,m}\,\text{Polya}\left(x_{j-1}(2k)|x_j(k),\alpha_{j,m},\beta_{j,m}\right)}{\sum_{n=1}^{M} \pi_{j,n}\,\text{Polya}\left(x_{j-1}(2k)|x_j(k),\alpha_{j,n},\beta_{j,n}\right)}. \quad (3.30)$$

Ο εκτιμητής του διαχωρίσιμου σχήματος για τη 2-Δ περίπτωση είναι της ίδιας μορφής με αυτόν της μονοδιάστατης περίπτωσης. Για περισσότερες πληροφορίες ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο [TN99]. Τέλος, οι βέλτιστες εκτιμήσεις για τις τυχαίες μεταβλητές $\left(\Theta_j^h(k,\ell), \Theta_j^v(k,\ell), \Theta_j^d(k,\ell)\right)$, που εμφανίζονται στο προτεινόμενο πολυκλιμακωτό Poisson-Haar σχήμα, είναι οι:

$$\hat{\theta}_j^o(k,\ell) = \sum_{m=1}^{M} \gamma_m(z_j^o(k,\ell)) \left(\frac{x_j^o(k,\ell) + \alpha_{j,m}^o}{x_j(k,\ell) + \alpha_{j,m}^o + \beta_{j,m}^o}\right) \quad (3.31)$$

με

$$\gamma_m(z_j^o(k,\ell)) = \frac{\pi_{j,m}\,\text{Polya}\left(x_j^o(k,\ell)|x_j(k,\ell),\alpha_{j,m}^o,\beta_{j,m}^o\right)}{\sum_{n=1}^{M} \pi_{j,n}\,\text{Polya}\left(x_j^o(k,\ell)|x_j(k,\ell),\alpha_{j,n}^o,\beta_{j,n}^o\right)}, \quad (3.32)$$

όπου το $x_j^o(k,\ell)$ ορίζεται στην Εξ. (3.19) και το o υποδηλώνει τον επιλεγμένο προσανατολισμό.

3.4.3 Επαναληπτική Εκτίμηση και Κυκλικές Μετατοπίσεις

Η επιλογή ενός εκ των τριών 2-Δ σχημάτων ανάλυσης μας επιτρέπει να καταλήξουμε σε ένα βέλτιστο εκτιμητή για το διάνυσμα $\boldsymbol{\lambda}_e$. Επειδή το $\boldsymbol{\lambda}_e$ αποτελεί ένα μετασχηματισμό του διανύσματος

έντασης της πηγαίας εικόνας $\boldsymbol{\lambda}$, καταλήγουμε στην τελική μας εκτίμηση για την εικόνα ενδιαφέροντος εφαρμόζοντας έναν αναδρομικό κατά μήκος των κλιμάκων αλγόριθμο, ώστε να οδηγηθούμε στο αρχικό πεδίο ορισμού της έντασης της εικόνας. Ο αλγόριθμος ξεκινά στην πιο αδρή κλίμακα ανάλυσης, όπου τα $\hat{\lambda}_J(0,0)$ και $\hat{\boldsymbol{\theta}}_J(0,0)$ χρησιμοποιούνται για τον υπολογισμό των εκτιμήσεων $\hat{\lambda}_j(k,\ell)$ της αμέσως πιο ευκρινούς κλίμακας. Όσον αφορά στο προτεινόμενο Poisson-Haar σχήμα ανάλυσης οι εξισώσεις που μας οδηγούν στην επόμενη κλίμακα είναι αυτές στην (2.34) με τη μόνη διαφορά ότι στη θέση των πραγματικών τιμών $\boldsymbol{\theta}$ χρησιμοποιούμε τις εκτιμήσεις τους. Ανάλογες εξισώσεις ισχύουν και για τα άλλα δύο σχήματα. Αυτή η διαδικασία επαναλαμβάνεται, έως ότου φτάσουμε στην ευκρινέστερη κλίμακα ανάλυσης $j = 0$, που είναι και το αρχικό πεδίο ορισμού της εικόνας.

Ένα μειονέκτημα των τριών υπό συζήτηση σχημάτων ανάλυσης είναι ότι η προκύπτουσα δενδρική δομή οδηγεί σε μη-στασιμότητα στο χώρο. Αυτό οφείλεται στο γεγονός ότι γειτονικά εικονοστοιχεία της εικόνας στην πιο ευκρινή κλίμακα είναι πιθανόν να έχουν κοινούς απογόνους σε πολύ αδρές κλίμακες του δένδρου [TN99, LPH00, Wil02]. Η ύπαρξη αυτής της μη-στασιμότητας τυπικά οδηγεί στην εμφάνιση τεχνουργημάτων στις τελικές εκτιμήσεις της έντασης της εικόνας. Μια εξήγηση από διαφορετική οπτική γωνία για την εμφάνιση τεχνουργημάτων στην αποκατεστημένη εικόνα συζητήσαμε στο Κεφ. 2. Η εξήγηση αυτή σχετίζεται με τη φασματική αναδίπλωση που λαμβάνει χώρα, λόγω της υποδειγματοληψίας κατά τη φάση της ανάλυσης. Τέτοιου είδους επιπτώσεις απαντώνται πολύ συχνά και στον κρίσιμα δειγματοληπτημένο μετασχηματισμό κυματιδίων.

Προκειμένου να επιτευχθεί η εν μέρει απαλοιφή αυτών των τεχνουργημάτων, είναι δυνατόν να γίνει χρήση μιας τεχνικής κυκλικής μετατόπισης [CD95] η οποία μπορεί να οδηγήσει σε σημαντικές βελτιώσεις στην ποιότητα των εκτιμήσεων έντασης. Αυτή η τεχνική συνίσταται στην κυκλική μετατόπιση των παρατηρούμενων δεδομένων, στην εκτίμηση της γενεσιουργού έντασης με τον τρόπο που περιγράψαμε στην Ενοτ. 3.4.2 και στην αντίστροφη μετατόπιση των αποτελεσμάτων στην αρχική τους θέση. Η διαδικασία αυτή επαναλαμβάνεται για όλες τις δυνατές μετατοπίσεις. Ως τελική εκτίμηση λαμβάνεται η μέση τιμή των αποτελεσμάτων που προήλθαν από όλες τις μετατοπίσεις. Μια άλλη σχετικά ισοδύναμη λύση είναι η χρήση υπερπλήρων αναπαραστάσεων. Στις υπερπλήρεις αναπαραστάσεις, όπως έχουμε ήδη αναφέρει, ο αριθμός των συντελεστών στο νέο χώρο αναπαράστασης είναι μεγαλύτερος από τον αρχικό αριθμό των εικονοστοιχείων της εικόνας. Μια τέτοια προσέγγιση μπορεί να ακολουθηθεί άμεσα στην περίπτωση του προτεινόμενου Poisson-Haar μετασχηματισμού, όπως παρουσιάσαμε στην Ενοτ. 2.3.4. Αν κανείς υιοθετήσει τη συγκεκριμένη προσέγγιση, τότε η διαδικασία εκτίμησης των συντελεστών $\theta_j^o(k,\ell)$ δεν παρουσιάζει καμία επιπλέον δυσκολία, αφού παραμένει ακριβώς η ίδια με τη μέθοδο που παρουσιάσαμε στην Ενοτ. 3.4.2.

3.5 Εκτίμηση Παραμέτρων με τη χρήση του EM αλγορίθμου

Οι εκτιμήσεις, οι οποίες προέκυψαν με βάση τον MMSE εκτιμητή στην Ενότητα 3.4.2, απαιτούν τη γνώση των παραμέτρων του μοντέλου $\boldsymbol{\mu} = \{\boldsymbol{\pi}_j, \boldsymbol{\alpha}_j\}$, δηλαδή τα βάρη και τις παραμέτρους του μείγματος βήτα/Dirichlet κατανομών που ελέγχουν τους παράγοντες διαχωρισμού $\Theta_j(k,\ell)$. Θα μπορούσε κάποιας να θεωρήσει αυτές τις παραμέτρους τυχαίες, να τους επιβάλει μια υπερ-πρότερη πιθανότητα $p(\boldsymbol{\mu})$, και να εργαστεί με το προκύπτον ιεραρχικό Μπεϋζιανό μοντέλο [GCSR03]. Ωστόσο, μια τέτοια προσέγγιση μπορεί να είναι εξαιρετικά απαιτητική υπολογιστικά. Για αυτό το λόγο επιλέγουμε να αντιμετωπίσουμε το πρόβλημα μέσω ενός εμπειρικού Μπεϋζιανού πλαισίου το οποίο αντιστοιχεί στο να εκτιμήσουμε τις παραμέτρους $\boldsymbol{\mu}$ για κάθε θορυβώδη εικόνα και να τις θεωρήσουμε σταθερές για το έργο εκτίμησης της έντασης της πηγαίας εικόνας.

Προηγούμενες ερευνητικές εργασίες, οι οποίες ακολούθησαν το εμπειρικό Μπεϋζιανό πλαίσιο, δεν κατάφεραν να αντιμετωπίσουν ικανοποιητικά το πρόβλημα εκτίμησης των παραμέτρων.

Οι συγγραφείς στο [TN99] μοντελοποιούν τις τυχαίες μεταβλητές $\Theta_j(k,\ell)$ με μείγμα 3 συμμετρικών βήτα κατανομών, όπου οι παράμετροι των βήτα επιλέγονται ευριστικά για κάθε κλίμακα ανάλυσης. Επιπλέον, υποθέτουν ότι το ένα από τα βάρη του μείγματος είναι γνωστό και εκτιμούν τα υπόλοιπα δύο χρησιμοποιώντας τη μέθοδο εξίσωσης ροπών (moment matching) [Day69]. Στην πραγματικότητα, επειδή το $\pi_{j,1}$ θεωρείται γνωστό και το $\pi_{j,3}$ περιορίζεται να είναι ίσο με $\pi_{j,3} = 1 - \pi_{j,1} - \pi_{j,2}$, μόνο το $\pi_{j,2}$ χρειάζεται να εκτιμηθεί. Η συγκεκριμένη μέθοδος είναι αρκετά περιοριστική και είναι κατάλληλη μόνο στην περίπτωση που επιλέγεται μείγμα έως 3 συνιστώσες για τη μοντελοποίηση της πρότερης πιθανότητας των τυχαίων μεταβλητών $\Theta_j(k,\ell)$. Έχει, επιπλέον, το μειονέκτημα ότι συχνά οδηγεί σε μη συμβατά αποτελέσματα, αφού ένα από τα βάρη του μείγματος μπορεί να λάβει και αρνητική τιμή, όπως έχει σημειωθεί και στο [LKA04]. Μια διαφορετική προσέγγιση ακολουθήθηκε στο [LKA04], όπου οι συγγραφείς αρχικά χρησιμοποιούν μια βοηθητική μέθοδο αποθορυβοποίησης [Don94, Don95] η οποία βασίζεται στην κατωφλιοποίηση των κυματιδιακών συντελεστών λεπτομέρειας, για να οδηγηθούν σε μια εκτίμηση των καθαρών παραγόντων διαχωρισμού, $\Theta_j(k,\ell)$. Με βάση τους συντελεστές αυτούς, στη συνέχεια, οι παράμετροι $\boldsymbol{\mu}$ ειδικεύονται με τον EM αλγόριθμο. Ένα μειονέκτημα αυτής της προσέγγισης είναι ότι, επειδή βασίζεται σε μια μέθοδο αποθορυβοποίησης που δε λαμβάνει υπόψιν της τα Poisson στατιστικά της εικόνας, είναι επιρρεπής σε πιθανή αποτυχία της βοηθητικής μεθόδου. Επιπλέον, η απόδοση του έργου εκτίμησης χρησιμοποιώντας τις παραμέτρους που ελήφθησαν με τη συγκεκριμένη διαδικασία είναι περιορισμένη, υπό την έννοια ότι οι εκτιμήσεις των παραγόντων διαχωρισμού, οι οποίες προέκυψαν μέσω της βοηθητικής μεθόδου, δε βελτιστοποιούν με ευθύ τρόπο κάποιο κριτήριο που να συνδέεται με τα στατιστικά του θορύβου και, συνεπώς, διαφέρουν από τις βέλτιστες δυνατές εκτιμήσεις.

Για να αντιμετωπίσουμε αυτούς τους περιορισμούς, αναζητούμε έναν εκτιμητή μέγιστης πιθανοφάνειας (maximum likelihood - ML), προκειμένου να συμπεράνουμε τις παραμέτρους του μοντέλου $\boldsymbol{\mu}$ *απευθείας* από τα παρατηρούμενα Poisson δεδομένα. Επειδή η πολυκλιμακωτή προσέγγιση μας επιτρέπει τον ανεξάρτητο χειρισμό κάθε κλίμακας, δε θα χρησιμοποιήσουμε πλέον τον δείκτη j για λόγους ευκολίας. Επίσης, αντί για το (k,ℓ), θα χρησιμοποιήσουμε το δείκτη n, για να δηλώσουμε τη θέση, υποθέτοντας ότι έχουμε μετατρέψει την παρατηρούμενη 2-Δ εικόνα σε ένα διάνυσμα $\mathbf{x}$ μήκους N μέσω λεξικογραφικής αναδιάταξης. Η βασική ιδέα είναι να περιθωριοποιήσουμε (integrate out) τους παράγοντες διαχωρισμού $\boldsymbol{\theta}_n$ που δεν έχουμε στη διάθεσή μας, και, επομένως, να συνδέσουμε τις παραμέτρους του μοντέλου $\boldsymbol{\mu}$ απευθείας με τις παρατηρήσεις $\mathbf{x}$. Κάνοντας χρήση της ιδιότητας (3.25), αυτή η διαδικασία μπορεί να υπολογιστεί σε κλειστή μορφή οδηγώντας στην παραγοντοποιημένη συνάρτηση πιθανοφάνειας:

$$p(\mathbf{x}|\boldsymbol{\mu}) = \int p(\mathbf{x},\boldsymbol{\theta}|\boldsymbol{\mu})\, d\boldsymbol{\theta} = \int p(\mathbf{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\boldsymbol{\mu})\, d\boldsymbol{\theta} = \prod_{n=1}^{N} \int p(\mathbf{x}_n^c|x_n,\boldsymbol{\theta}_n)\, p(\boldsymbol{\theta}_n|\boldsymbol{\mu})\, d\boldsymbol{\theta}_n$$
$$= \prod_{n=1}^{N} \left(\sum_{m=1}^{M} \pi_m \int p(\mathbf{x}_n^c|x_n,\boldsymbol{\theta}_n)\, p(\boldsymbol{\theta}_n|\boldsymbol{\alpha}_m)\, d\boldsymbol{\theta}_n \right) = \prod_{n=1}^{N} \sum_{m=1}^{M} \pi_m \,\text{Polya}\,(\mathbf{x}_n^c|x_n,\boldsymbol{\alpha}_m)\ , \quad (3.33)$$

όπου το $\mathbf{x}_n^c$ είναι οι Poisson μετρήσεις των παιδιών του κόμβου x_n. Για το διαχωρίσιμο και το προτεινόμενο πολυκλιμακωτό Poisson-Haar σχήμα ανάλυσης η $p(\mathbf{x}_n^c|x_n,\boldsymbol{\theta}_n)$ και η $p(\boldsymbol{\theta}_n|\boldsymbol{\alpha}_m)$ αντιστοιχούν στη διωνυμική και στο μείγμα βήτα κατανομών με $\boldsymbol{\alpha}_m = [\alpha_m, \beta_m]$, ενώ για το τετραδικό σχήμα στην πολυωνυμική και στο μείγμα Dirichlet κατανομών, αντιστοίχως.

Η εύρεση των ML παραμέτρων αντιστοιχεί στην εύρεση των παραμέτρων $\boldsymbol{\mu}$ που μεγιστοποιούν τη συνάρτηση πιθανοφάνειας στο λογαριθμικό πεδίο $L(\boldsymbol{\mu}) = \log p(\mathbf{x}|\boldsymbol{\mu})$. Η απευθείας βελτιστοποίηση της $L(\boldsymbol{\mu})$ είναι εξαιρετικά δύσκολη εξαιτίας των πολλαπλών συνιστωσών του μείγματος. Για αυτό το λόγο θέτουμε σε χρήση τον αλγόριθμο EM [Bis06] και εργαζόμαστε με την πλήρη συνάρτηση πιθανοφάνειας. Σύμφωνα με τον αλγόριθμο EM, σε κάθε δείγμα παρατήρησης $\mathbf{x}_n^c$ ανατίθεται μια κρυφή δυαδική τυχαία διανυσματική μεταβλητή $\mathbf{z}_n$, M διαστάσεων. Σε κάθε στιγ-

μιότυπο της $\mathbf{z}_n$ μόνο ένα συγκεκριμένο στοιχείο z_{nm} μπορεί να ισούται με μονάδα, ενώ όλα τα υπόλοιπα θα πρέπει να είναι μηδέν. Το στοιχείο z_{nm} είναι ίσο με μονάδα μόνο, εάν η mοστή συνιστώσα του μείγματος των Polya κατανομών είναι υπεύθυνη για την παραγωγή των παρατηρήσεων $\mathbf{x}_n^c$, δοθείσης της παρατήρησης του κόμβου-πατέρα x_n. Με $\mathbf{z}$ συμβολίζουμε το σύνολο των αναθέσεων των διακριτών συνιστωσών μείγματος για όλα τα δείγματα παρατήρησης, δηλαδή $\mathbf{z} = \{\mathbf{z}_1 \dots \mathbf{z}_n\}$. Η περιθώρια κατανομή για κάθε $\mathbf{z}_n$ καθορίζεται με όρους των βαρών μείγματος, έτσι ώστε $p(z_{nm} = 1) = \pi_m$. Βασιζόμενοι σε αυτή τη σημειολογία η συνάρτηση $L(\boldsymbol{\mu})$ μπορεί να θεωρηθεί ως η ατελής λογαριθμική πιθανοφάνεια και η $L_c(\boldsymbol{\mu}) = \log p(\mathbf{x}, \mathbf{z}|\boldsymbol{\mu})$ ως η πλήρης. Αν τώρα υποθέσουμε ότι εκτός από τα δείγματα παρατήρησης μας δίνονται επιπλέον και οι τιμές των αντίστοιχων κρυφών μεταβλητών στο $\mathbf{z}$, μπορούμε να εκφράσουμε την πλήρη λογαριθμική πιθανοφάνεια με τη μορφή:

$$L_c(\boldsymbol{\mu}) = \log \prod_{n=1}^{N} \prod_{m=1}^{M} \left[\pi_m \,\text{Polya}\left(\mathbf{x}_n^c | x_n, \boldsymbol{\alpha}_m\right)\right]^{z_{nm}}$$
$$= \sum_{n=1}^{N} \sum_{m=1}^{M} z_{nm} \log(\pi_m) + \sum_{n=1}^{N} \sum_{m=1}^{M} z_{nm} \log \text{Polya}\left(\mathbf{x}_n^c | x_n, \boldsymbol{\alpha}_m\right). \quad (3.34)$$

Είναι πολύ πιο εύκολο να εργαστούμε με αυτή τη μορφή, επειδή ο λογάριθμος δεν επενεργεί πλέον στα αθροίσματα των συνιστωσών του μείγματος, αλλά απευθείας στην Polya κατανομή. Ωστόσο, το πρόβλημα τώρα είναι ότι οι πραγματικές τιμές των κρυφών μεταβλητών μας είναι άγνωστες. Για αυτό το λόγο αντί να θεωρήσουμε την πλήρη λογαριθμική πιθανοφάνεια υπολογίζουμε τη μέση τιμή της ως προς την ύστερη κατανομή των κρυφών μεταβλητών. Αυτός ο υπολογισμός λαμβάνει χώρα στο βήμα εύρεσης της μέσης τιμής του αλγορίθμου (Ε-βήμα), όπου η υπό συνθήκη μέση τιμή της πλήρους πιθανοφάνειας, δοθέντων των παρατηρούμενων δεδομένων και των πρόσφατων εκτιμήσεων των παραμέτρων, είναι της μορφής:

$$E\left[L_c(\boldsymbol{\mu}) | \mathbf{x}, \boldsymbol{\mu}^{(l)}\right] = \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_m(z_n) \log(\pi_m) + \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_m(z_n) \log \text{Polya}\left(\mathbf{x}_n^c | x_n, \boldsymbol{\alpha}_m\right), \quad (3.35)$$

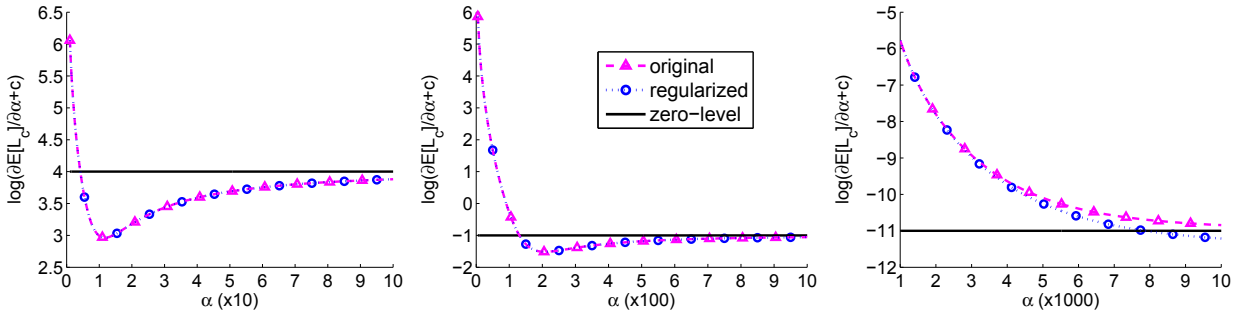
όπου η $\gamma_m(z_n) = E\{z_{nm} = 1 | \mathbf{x}\}$ ορίζεται όπως στις εξισώσεις (3.27) και (3.32), αναλόγως με το επιλεγμένο σχήμα ανάλυσης. Παρατηρούμε ότι η μέση τιμή της πλήρους λογαριθμικής πιθανοφάνειας (3.35) συνίσταται σε δύο ξεχωριστούς όρους. Επομένως, για να εκτιμήσουμε τις ανανεωμένες παραμέτρους του μοντέλου $\boldsymbol{\mu}^{(l+1)} = \left\{\boldsymbol{\pi}^{(l+1)}, \boldsymbol{\alpha}^{(l+1)}\right\}$ μπορούμε να μεγιστοποιήσουμε τον πρώτο όρο ως προς το $\boldsymbol{\pi}$, και το δεύτερο ως προς το $\boldsymbol{\alpha}$. Αυτό αποτελεί το βήμα μεγιστοποίησης του αλγορίθμου (Μ-βήμα). Η μεγιστοποίηση της Εξ. (3.35) ως προς το $\boldsymbol{\pi}$, υπό τον περιορισμό $\sum_{m=1}^{M} \pi_m = 1$, οδηγεί στα ανανεωμένα βάρη του μείγματος

$$\pi_m^{(l+1)} = \frac{1}{N} \sum_{n=1}^{N} \gamma_m(z_n). \quad (3.36)$$

Προκειμένου να λάβουμε τις ανανεωμένες παραμέτρους $\boldsymbol{\alpha}^{(l+1)}$, πρέπει πρώτα να υπολογίσουμε τις μερικές παραγώγους της λογαριθμικής πιθανοφάνειας (3.35) ως προς τα α_m^t

$$\frac{\partial E\{L_c\}}{\partial \alpha_m^t} = \sum_{n=1}^{N} \gamma_m(z_n) \left[\psi\left(x_n^{c,t} + \alpha_m^t\right) - \psi\left(a_m^t\right) + \psi\left(\sum_{t=1}^{D} \alpha_m^t\right) - \psi\left(x_n + \sum_{t=1}^{D} \alpha_m^t\right)\right], \quad (3.37)$$

όπου $\psi(\cdot)$ είναι η συνάρτηση δίγαμα [AS72] η οποία ορίζεται ως $\psi(x) = \partial \log \Gamma(x) / \partial x$ και ο υπερδείκτης $t = 1, \dots, D$, με $D = 4$ για το τετραδικό σχήμα και $D = 2$ για τα άλλα δύο σχήματα



Σχήμα 3.3: Η μερική παράγωγος της λογαριθμικής συνάρτησης πιθανοφάνειας ως προς τη συμμετρική παράμετρο της βήτα κατανομής α_m και η ομαλοποιημένη εκδοχή της. Οι τρεις γραφικές αναπαραστάσεις αντιστοιχούν στη συνιστώσα με τη μικρή (αριστερά), μεσαία (κέντρο), και μεγάλη (δεξιά) τιμή της α_m παραμέτρου. Η επιβαλλόμενη ομαλοποίηση φράζει αποτελεσματικά τη ρίζα της συνιστώσας με την μεγαλύτερη παράμετρο, ενώ ουσιαστικά αφήνει ακέραιες τις δύο άλλες ρίζες.

ανάλυσης, δηλώνει το t-οστό στοιχείο του αντίστοιχου διανύσματος. Οι ανανεωμένες παράμετροι μπορούν να βρεθούν θέτοντας την (3.37) ίση με μηδέν και λύνοντας ως προς α_m^t. Δυστυχώς αυτή η εξίσωση είναι μη γραμμική· έτσι είναι αδύνατη η εύρεση μιας λύσης για τις παραμέτρους σε κλειστή μορφή. Επιπλέον, η κατανομή Polya δεν ανήκει στην εκθετική οικογένεια κατανομών και, επομένως, δε διαθέτει επαρκή στατιστικά (sufficient statistics) [GG00]. Συνεπώς, για να οδηγηθούμε σε μια λύση, πρέπει να καταφύγουμε στη χρήση μιας επαναληπτικής μεθόδου. Σε αυτό το σημείο πρέπει να σημειώσουμε ότι το συγκεκριμένο πρόβλημα επίλυσης της εξίσωσης (3.37) απαντάται και σε άλλες ερευνητικές περιοχές, όπως είναι η μοντελοποίηση φυσικής γλώσσας και κειμένου [MP94, SMY07], αλλά και στη μοντελοποίηση ακολουθιών DNA [SKB+96]. Σε αυτές τις περιπτώσεις οι συγγραφείς είτε εφαρμόζουν τη μέθοδο μέγιστης καθόδου (gradient descent) [DHS01] είτε μια επαναληπτική μέθοδο σταθερού σημείου, όπως προτείνεται στο [Min03].

Για τη βελτιστοποίηση της (3.35) προτείνουμε μια πρωτότυπη τεχνική που εφαρμόζει την επαναληπτική μέθοδο *Newton-Raphson* [PTVF07] για την επίλυση της Εξ. (3.37). Ένα πρόβλημα που αντιμετωπίζουν οι συμβατικές τεχνικές είναι ότι σε εικόνες με ομαλό cartoon-like περιεχόμενο το ιστόγραμμα των παραγόντων διαχωρισμού $\boldsymbol{\theta}$ παρουσιάζει ισχυρές κορυφές (στο 0.25 για το τετραδικό και στο 0.5 για το διαχωρίσιμο και το Poisson-Haar σχήμα ανάλυσης), με αποτέλεσμα οι προκύπτουσες παράμετροι α_m^t να είναι πολύ μεγάλες. Το κριτήριο ML μπορεί τότε να οδηγήσει σε υπερ-ειδίκευση (over-fitting), με τη μεγαλύτερη από τις α_m^t παραμέτρους να αυξάνεται διαρκώς χωρίς κανέναν περιορισμό σε κάθε επανάληψη του ΕΜ αλγορίθμου (η αντίστοιχη συνιστώσα του μείγματος βήτα κατανομών διαρκώς κορυφώνεται υψηλότερα στο σημείο 0.5). Αριθμητικά, η διαδικασία εύρεσης της μεγάλης τιμής της ρίζας της παραγώγου της λογαριθμικής πιθανοφάνειας (3.37) καθίσταται επίσης ασταθής και εξαιρετικά χρονοβόρα. Για να αντιμετωπίσουμε αυτό το θέμα, προσθέτουμε έναν όρο ομαλοποίησης $-\epsilon \sum_{m=1}^{M} \sum_{t=1}^{D} \alpha_m^t$ στην (3.35), όπου το ϵ είναι μια μικρή θετική σταθερά. Μέσω αυτής της προσέγγισης, ακόμα και για μεγάλες τιμές της α_m^t, καταφέρνουμε να βρούμε μια ρίζα στην (3.37) και συνακόλουθα ένα μέγιστο για την Εξ. (3.35), ενώ για μικρές και μεσαίες τιμές η πραγματική τιμή της ρίζας επηρεάζεται ελάχιστα, όπως απεικονίζεται και στο Σχ. 3.3. Ο συγκεκριμένος όρος ομαλοποίησης μπορεί να ερμηνευθεί ως ο καθορισμός μια συζυγούς πρότερης πιθανότητας για το $\boldsymbol{\alpha}_m$ που οδηγεί στη MAP εκτίμηση αντί για την τυπική ML εκτίμηση στο M-βήμα του αλγορίθμου [Bis06] (για περισσότερες λεπτομέρειες δες στο Παράρτημα 3.Δ'). Ο προκύπτων ποινικοποιημένος ΕΜ αλγόριθμος είναι εξαιρετικά εύρωστος στην πράξη.

Ο αλγόριθμος ΕΜ απαιτεί κατάλληλη αρχικοποίηση. Προκειμένου να καταστήσουμε τον αλ-

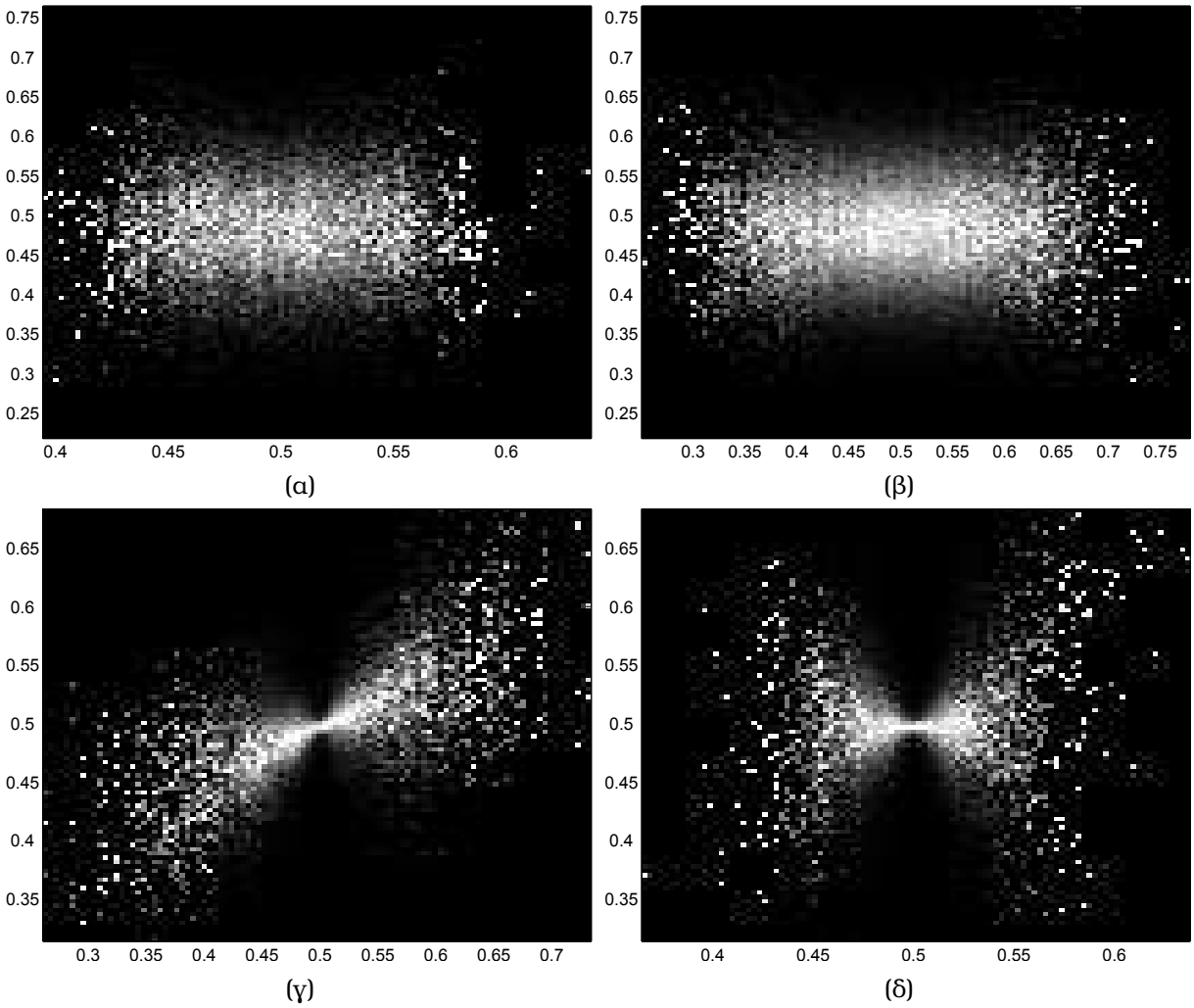
γόριθμό μας για την εκτίμηση των παραμέτρων πλήρως αυτοματοποιημένο, έχουμε αναπτύξει μια τεχνική αυτοδύναμης εκκίνησης (bootstrapping). Ειδικότερα, ξεκινάμε από μια μοναδική συνιστώσα μείγματος της οποίας οι παράμετροι ειδικεύονται μέσω της μεγιστοποίησης της πιθανοφάνειας. Στην συνέχεια, ο αριθμός των συνιστωσών αυξάνεται κατά έναν, έως ότου καταλήξουμε στον επιθυμητό αριθμό. Αυτό επιτυγχάνεται με τον κατέπανάληψη διαχωρισμό εκείνης της συνιστώσας του μείγματος που διαθέτει το μεγαλύτερο συνολικό βάρος ανάθεσης $\gamma_m = \sum_{n=1}^{N} \gamma_m(z_n)$. Κάθε διαχωρισμός οδηγεί σε δύο συνιστώσες των οποίων οι αρχικές παράμετροι είναι μετατοπισμένες εκδοχές της γεννήτριας συνιστώσας. Στην περίπτωση που διαθέτουμε εύλογες αρχικοποιήσεις από την εκπαίδευση σε παρόμοιες εικόνες, μπορούμε να χρησιμοποιήσουμε αυτές για να επιταχύνουμε τη διαδικασία εκπαίδευσης.

3.6 Μοντελοποίηση Εξαρτήσεων Μεταξύ Κλιμάκων με Κρυφά Μαρκοβιανά Δένδρα

Στην Ενότητα 3.3, προκειμένου να οδηγηθούμε σε μια παραγοντοποιημένη μορφή της συνάρτησης πρότερης πιθανότητας $p(\boldsymbol{\lambda}_e)$, θεωρήσαμε τους συντελεστές $\boldsymbol{\theta}$ σε κάθε κλίμακα ανάλυσης ως τα στιγμιότυπα ανεξάρτητων τυχαίων μεταβλητών οι οποίες παράγονται από ένα μείγμα βήτα/Dirichlet κατανομών. Η επιλογή της ανεξαρτησίας στηρίζεται στην ύπαρξη της στενής σχέσης που συνδέει τον πολλαπλασιαστικό μετασχηματισμό και το μετασχηματισμό κυματιδίων. Όπως ήδη έχει αναφερθεί στο Κεφ. 2, ο μετασχηματισμός κυματιδίων δρα, σε καλή προσέγγιση, ως μετασχηματισμός αποσυσχέτισης, δηλαδή οι συντελεστές λεπτομέρειας που προκύπτουν μέσα από την εφαρμογή του σε μια εικόνα τείνουν να είναι ασυσχέτιστοι μεταξύ τους. Είναι λοιπόν λογικό κανείς να υποθέτει ότι και οι συντελεστές διαχωρισμού $\boldsymbol{\theta}$ διαθέτουν τη συγκεκριμένη ιδιότητα.

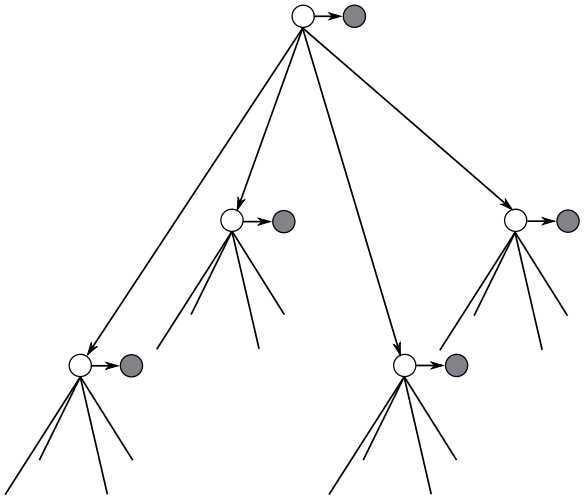
Στην πράξη τόσο ο μετασχηματισμός κυματιδίων όσο και ο πολλαπλασιαστικός μετασχηματισμός δεν μπορούν να εξαλείψουν εντελώς τις εξαρτήσεις μεταξύ των συντελεστών. Η διαπίστωση αυτή ενισχύεται περαιτέρω και από την ιδιότητα που έχει βρεθεί οτι διαθέτει ο μετασχηματισμός κυματιδίων [MZ92, MH92, OR94, WSW01], δηλαδή ότι οι συντελεστές του εμφανίζονται κυρίως σε συστάδες στο χώρο ή στην κλίμακα. Αυτό σημαίνει ότι, εάν ένας συγκεκριμένος συντελεστής έχει μεγάλη/μικρή τιμή, τότε είναι πολύ πιθανόν οι γειτονικοί του συντελεστές στο χώρο ή στην κλίμακα να έχουν και αυτοί με τη σειρά τους μεγάλη/μικρή τιμή. Η συγκεκριμένη ιδιότητα κληρονομείται και από τους συντελεστές διαχωρισμού, όπως απεικονίζεται και στο Σχ. 3.4. Σε αυτό το σχήμα παρουσιάζονται τέσσερα από κοινού ιστογράμματα και οι υπό συνθήκη κατανομές για δύο διαφορετικά σύνολα συντελεστών. Το πρώτο σύνολο που καλούμε σύνολο "παιδιού" περιλαμβάνει τους συντελεστές θ που έχουν υπολογιστεί σε μια λεπτομερή κλίμακα ανάλυσης και για ένα συγκεκριμένο προσανατολισμό. Το δεύτερο σύνολο καλείται σύνολο "πατέρα" και περιλαμβάνει τους συντελεστές που είτε υπολογίστηκαν σε μια αδρότερη κλίμακα ανάλυσης είτε στην ίδια κλίμακα αλλά για ένα διαφορετικό προσανατολισμό. Κάθε στήλη των 2-Δ αναπαραστάσεων αντιστοιχεί στο 1-Δ από κοινού ιστόγραμμα $p(\text{child}|\text{parent})$ για μια σταθερή τιμή του συντελεστή "πατέρα". Η ένταση φωτεινότητας αντιστοιχεί στην τιμή της πιθανότητας. Στα Σχ. 3.4(α) και 3.4(β) απεικονίζονται τα από κοινού ιστογράμματα των συντελεστών που προέκυψαν από μια συνθετική εικόνα γκαουσιανού θορύβου. Όπως αναμενόταν, τα δύο σύνολα συντελεστών είναι ανεξάρτητα μεταξύ τους, επειδή το σχήμα που προκύπτει σε κάθε ιστόγραμμα είναι ανεξάρτητο από την τιμή του πατέρα. Στα Σχ. 3.4(γ) και 3.4(δ) εμφανίζονται οι συντελεστές που προέκυψαν κατά την ανάλυση μιας τυπικής φυσικής εικόνας. Στο Σχ. 3.4(γ) οι συντελεστές του συνόλου "πατέρα" υπολογίστηκαν σε μια αδρότερη κλίμακα από αυτή του συνόλου "παιδιού" αλλά για τον ίδιο προσανατολισμό, ενώ στο Σχ. 3.4(δ) οι συντελεστές του συνόλου "πατέρα" υπολογίστηκαν στην ίδια κλίμακα με αυτή του "παιδιού" αλλά για διαφορετικό προσανατολισμό. Και στα δύο αυτά σχήματα βλέπουμε ότι, παρόλο που οι συντελεστές είναι κατά προσέγγιση ασυσχέτιστοι, εμφανίζουν στατιστικές εξαρτήσεις υψηλότερων τάξεων. Πιο συγκεκριμένα, η υπό συνθήκη κατανομή των τιμών για τους συντελεστές

του συνόλου “παιδιού” εμφανίζει μια τυπική απόκλιση που μεταβάλλεται με βάση την τιμή του συντελεστή “πατέρα” δημιουργώντας στο ιστόγραμμα ένα σχήμα πεταλούδας (bow-tie). Το γεγονός αυτό επιβεβαιώνει ότι εκτός από τους συντελεστές κυματιδίων και οι συντελεστές διαχωρισμού διαθέτουν την ιδιότητα της εμφάνισης κατά συστάδες στο χώρο και στην κλίμακα.



Σχήμα 3.4: Από κοινού ιστογράμματα και οι υπό συνθήκη κατανομές για δύο διαφορετικά σύνολα συντελεστών διαχωρισμού θ (“πατέρα”-“παιδιού”). (α)-(γ) Από κοινού ιστόγραμμα των συντελεστών του συνόλου “παιδιού” (λεπτομερής κλίμακα ανάλυσης) και του συνόλου “πατέρα” (αδρότερη κλίμακα ανάλυσης) για μια συνθετική εικόνα γκαουσιανού θορύβου και μια τυπική φυσική εικόνα, αντιστοίχως. (β)-(δ) Από κοινού ιστόγραμμα των συντελεστών του συνόλου “παιδιού” και του συνόλου “πατέρα” που προέκυψαν στην ίδια κλίμακα ανάλυσης αλλά για διαφορετικούς προσανατολισμούς για μια συνθετική εικόνα γκαουσιανού θορύβου και μια τυπική φυσική εικόνα, αντιστοίχως.

Έχοντας υπόψιν μας την παραπάνω συζήτηση καταλήγουμε στο ότι η υπόθεση της ανεξαρτησίας που υιοθετήσαμε στις προηγούμενες ενότητες περιορίζει την απόδοση του μοντέλου που την υιοθετεί, αφού δε λαμβάνει υπόψιν του τις εξαρτήσεις υψηλότερων τάξεων. Για να αντιμετωπίσουμε αυτό το θέμα, μπορούμε να ενσωματώσουμε στο πολυκλιμακωτό πλαίσιο ανάλυσης των Poisson διαδικασιών το κρυφό μαρκοβιανό δενδρικό (hidden Markov tree - HMT) μοντέλο το οποίο αρχικά χρησιμοποιήθηκε για το πρόβλημα αποθορυβοποίησης μονοδιάστατων σημάτων αλλοιωμένων



Σχήμα 3.5: Η δομή του κρυφού μαρκοβιανού δένδρου (βαθμού 4) που χρησιμοποιείται για τη μοντελοποίηση δια-κλιμακωτών εξαρτήσεων των συντελεστών διαχωρισμού εικόνων. Οι λευκοί κόμβοι συμβολίζουν τις διακριτές κρυφές μεταβλητές $\mathbf{z}_n$, ενώ οι σκιασμένοι κόμβοι συμβολίζουν τις παρατηρήσεις. Τα βέλη δηλώνουν στατιστικές εξαρτήσεις με τη φορά του βέλους να καθορίζει ποια είναι η εξαρτημένη τυχαία μεταβλητή. Στην περίπτωση που εξαλειφθούν οι εξαρτήσεις μεταξύ των κρυφών καταστάσεων (λευκοί κόμβοι), τότε οδηγούμαστε στο ανεξάρτητο δένδρο (hidden independent tree -HIT) της Ενοτ. 3.5

από γκαουσιανό θόρυβο [CNB98], ενώ στη συνέχεια επεκτάθηκε και σε 2-Δ σήματα στο [RCB01]. Το HMT μοντελοποιεί με αποδοτικό τρόπο τις δια-κλιμακωτές εξαρτήσεις μεταξύ των συνιστωσών του μείγματος και, επομένως, μπορεί να προσφέρει πρόσθετα πλεονεκτήματα στη διαδικασία εκτίμησης της γενεσιουργού έντασης. Η υιοθέτηση του HMT μοντέλου μας επιτρέπει να συλλάβουμε την εγγενή ιεραρχική δομή των δεδομένων και παράλληλα να εκμεταλλευτούμε την ύπαρξη αποδοτικών αναδρομικών μεταξύ κλιμάκων αλγορίθμων που έχουν σχεδιαστεί να επενεργούν σε δενδρικούς γράφους. Παρότι η χρήση της HMT δομής στο πρόβλημα της μοντελοποίησης εικόνων περιορισμένης φωτεινότητας είχε ήδη προταθεί στο [Now99a], δεν έχει τύχει μεγάλης αποδοχής σε αυτό το πλαίσιο, λόγω του ότι μέχρι τώρα δεν υπήρχε μια ικανοποιητική λύση για το πρόβλημα της εκτίμησης των παραμέτρων. Αυτό το πρόβλημα το αντιμετωπίζουμε επεκτείνοντας την EM τεχνική που παρουσιάσαμε στην Ενοτ. 3.5, προκειμένου να καλύπτει και την περίπτωση του HMT.

Το μοντέλο HMT είναι παρόμοιο με το κρυφό μαρκοβιανό μοντέλο (hidden Markov model - HMM) το οποίο χρησιμοποιείται πάρα πολύ συχνά στις εφαρμογές ανάλυσης φωνής [Rab89]. Το HMT αποτελεί ένα πολύ σημαντικό στατιστικό εργαλείο, αφού έχει τη δυνατότητα να μοντελοποιεί τις δια-κλιμακωτές εξαρτήσεις μεταξύ των συντελεστών διαχωρισμού με έναν έμμεσο και ταυτόχρονα πολύ αποδοτικό τρόπο. Σε κάθε συντελεστή παρατήρησης που αναπαρίσταται ως κόμβος ενός δένδρου ανατίθεται μια διακριτή κρυφή μεταβλητή $\mathbf{z}_n$ η οποία, όπως ορίσαμε στην Ενότητα 3.5, είναι υπεύθυνη για τις αναθέσεις των συνιστωσών του μείγματος. Στη συνέχεια, επιβάλλεται μια μαρκοβιανή δενδρική δομή εξάρτησης μεταξύ αυτών των κρυφών καταστάσεων, όπως φαίνεται και στο Σχ. 3.5. Πιο συγκεκριμένα, έστω ότι το $\boldsymbol{\mu} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\alpha}\}$ είναι το σύνολο των παραμέτρων του HMT μοντέλου· αυτές αντιστοιχούν σε (1) π_i: πρότερη πιθανότητα κατάληψης της i κατάστασης στη ρίζα του δένδρου, (2) A_{ij}: πιθανότητα μετάβασης από πατέρα που βρίσκεται στην j κατάσταση σε παιδί που βρίσκεται στην i κατάσταση και (3) $\boldsymbol{\alpha}_i$: οι παράμετροι της Polya κατανομής που αντιστοιχούν στην i κατάσταση. Έστω επίσης ότι το n διατρέχει όλους τους $N+1$ κόμβους του δένδρου, όπου $n = 0$ αντιστοιχεί στη ρίζα του δένδρου, $n = N$ αντιστοιχεί στον τελευταίο κόμβο

στην πιο ευκρινή κλίμακα, το $a(n) = \lfloor (n+1)/D \rfloor$ συμβολίζει τον πατέρα του κόμβου n[3] και το $\mathbf{x}_n^c$ είναι το διάνυσμα με τις τιμές των παιδιών του x_n. Το μοντέλο που υιοθετήσαμε ικανοποιεί επιπλέον και την ιδιότητα της υπό συνθήκη ανεξαρτησίας, δηλαδή ότι το $\mathbf{z}_n$ είναι ανεξάρτητο του $\mathbf{z}_{a(a(n))}$, δοθέντος του $\mathbf{z}_{a(n)}$. Επιπλέον, θεωρούμε ότι, δοθεισών των κρυφών μεταβλητών, όλα τα δείγματα παρατήρησης είναι αμοιβαία ανεξάρτητα. Βασιζόμενοι στις παραπάνω υποθέσεις η πρότερη πιθανότητα της κρυφής διακριτής ακολουθίας καταστάσεων $p(\mathbf{z}) = p(\mathbf{z}_0, \ldots, \mathbf{z}_N)$ μπορεί να εκφρασθεί ως

$$p(\mathbf{z}) = \prod_{m=1}^{M} \pi_m^{z_{0m}} \prod_{n=1}^{N} \prod_{m'=1}^{M} \prod_{m=1}^{M} A_{m,m',j(n)}^{z_{nm} z_{a(n)m'}}, \tag{3.38}$$

ενώ η δεσμευμένη πιθανοφάνεια $p(\mathbf{x}|\mathbf{z})$ ως:

$$p(\mathbf{x}|\mathbf{z}) = \prod_{n=0}^{N} \prod_{m=1}^{M} \text{Polya}\left(\mathbf{x}_n^c | x_n, \boldsymbol{\alpha}_{j(n),m}\right)^{z_{nm}}, \tag{3.39}$$

όπου $P(z_{0m} = 1) = \pi_m$, $P\left(z_{nm} = 1 | z_{a(n)m'} = 1\right) = A_{m,m',j(n)}$ και το $j(n)$ δηλώνει το επίπεδο του δένδρου στο οποίο βρίσκεται ο συγκεκριμένος κόμβος. Η εξίσωση (3.39) συνδέει τις παραμέτρους του μοντέλου με τις θορυβώδεις παρατηρήσεις του διανύσματος $\mathbf{x}$ και οδηγεί σε ένα μείγμα από Polya-κατανομές, ακριβώς όπως και στην Ενότητα 3.5.

Διαθέτοντας πλέον τις εκφράσεις για την πρότερη πιθανότητα των κρυφών καταστάσεων, αλλά και των δεσμευμένων πιθανοτήτων των παρατηρήσιμων δειγμάτων, δοθεισών των κρυφών μεταβλητών, έχουμε τη δυνατότητα να βρούμε μια έκφραση της πλήρους λογαριθμικής συνάρτησης πιθανοφάνειας $L_c(\boldsymbol{\mu})$ η οποία είναι απαραίτητη για την εκπαίδευση στον ΕΜ αλγόριθμο:

$$\begin{aligned} L_c(\boldsymbol{\mu}) &= \log p(\mathbf{x}, \mathbf{z}|\boldsymbol{\mu}) = \log p(\mathbf{z}|\boldsymbol{\mu}) + \log p(\mathbf{x}|\mathbf{z}, \boldsymbol{\mu}) \\ &= \sum_{m=1}^{M} z_{0m} \log \pi_m + \sum_{n=1}^{N} \sum_{m'=1}^{M} \sum_{m=1}^{M} z_{nm} z_{a(n)m'} \log A_{m,m',j(n)} \\ &\quad + \sum_{n=0}^{N} \sum_{m=1}^{M} z_{nm} \log \text{Polya}\left(\mathbf{x}_n^c | x_n, \boldsymbol{\alpha}_{j(n),m}\right). \end{aligned} \tag{3.40}$$

Όπως μπορούμε να παρατηρήσουμε, η πλήρης λογαριθμική συνάρτηση πιθανοφάνειας δίνεται σε διαχωρίσιμη μορφή ως προς τις παραμέτρους του μοντέλου. Συνεπώς, μπορούμε να βρούμε την τιμή της κάθε παραμέτρου απλά μεγιστοποιώντας τον αντίστοιχο όρο. Ωστόσο, όπως και στην προηγούμενη ενότητα, έτσι και εδώ δε γνωρίζουμε τις τιμές των κρυφών μεταβλητών. Για ακόμη μία φορά η τυπική προσέγγιση είναι να μεγιστοποιήσουμε την αναμενόμενη τιμή της $L_c(\boldsymbol{\mu})$ ως προς την ύστερη κατανομή των κρυφών μεταβλητών. Έτσι, στο Ε-βήμα χρησιμοποιούμε τις τρέχουσες εκτιμήσεις των παραμέτρων του μοντέλου $\boldsymbol{\mu}^{(l)}$, ώστε να βρούμε την ύστερη κατανομή και στη συνέχεια χρησιμοποιούμε αυτή την τιμή για να υπολογίσουμε τη μέση τιμή της πλήρους πιθανοφάνειας ως συνάρτηση των παραμέτρων $\boldsymbol{\mu}$. Η δεσμευμένη μέση τιμή της πλήρους συνάρτησης πιθανοφάνειας που εκτιμάται στο Ε-βήμα θα είναι της μορφής:

$$\begin{aligned} E\left\{L_c(\boldsymbol{\mu}) | \mathbf{x}, \boldsymbol{\mu}^{(l)}\right\} &= \sum_{m=1}^{M} \gamma_m(z_0) \log \pi_m + \sum_{n=1}^{N} \sum_{m'=1}^{M} \sum_{m=1}^{M} \xi_{m,m'}\left(z_n, z_{a(n)}\right) \log A_{m,m',j(n)} \\ &\quad + \sum_{n=0}^{N} \sum_{m=1}^{M} \gamma_m(z_n) \log \text{Polya}\left(\mathbf{x}_n^c | x_n, \boldsymbol{\alpha}_{j(n),m}\right). \end{aligned} \tag{3.41}$$

[3] Η τιμή του D αντιστοιχεί στον αριθμό των παιδιών ενός κόμβου του δένδρου. Έτσι, για το τετραδικό σχήμα $D = 4$, ενώ για το διαχωρίσιμο και το Poisson-Haar σχήμα $D = 2$.

Θα πρέπει να σημειώσουμε ότι στην περίπτωση του HMT μοντέλου όλες οι παρατηρήσεις από όλες τις κλίμακες τυγχάνουν επεξεργασίας όλες μαζί ως σύνολο σε αντίθεση με την ανεξάρτητη περίπτωση κατά την οποία μόνο οι παρατηρήσεις από μια κλίμακα χρησιμοποιούνται κάθε φορά. Εφαρμόζοντας τον αλγόριθμο *ανοδικής-καθοδικής διάδοσης* (upward-downward) [CNB98], μπορούμε με αποδοτικό τρόπο να υπολογίσουμε τη δεσμευμένη πιθανότητα $\gamma_m(z_n) = P\left(z_{nm} = 1|\mathbf{x}\right)$ που απαιτείται για τις εκτιμήσεις των (3.28) και (3.31), όπως και την από κοινού πιθανότητα των καταστάσεων $\xi_{m,m'}(z_n, z_{a(n)}) = P\left(z_{nm} = 1, z_{a(n)m'} = 1|\mathbf{x}\right)$. Έπειτα, στο M-βήμα λαμβάνουμε νέες εκτιμήσεις για τις παραμέτρους του μοντέλου. Ο υπολογισμός των παραμέτρων $\boldsymbol{\pi}$ και $\mathbf{A}$ πραγματοποιείται όπως περιγράφεται στο [CNB98], ενώ η μεγιστοποίηση της Εξ. (3.41) ως προς το $\boldsymbol{\alpha}_{j(n),m}$ γίνεται με ακριβώς τον ίδιο τρόπο, όπως στην ανεξάρτητη περίπτωση που συζητήσαμε στην Ενότητα 3.5. Η ομαλοποίηση της λύσης είναι και εδώ εξίσου σημαντική για να επιτύχουμε ευρωστία των αποτελεσμάτων.

Σε αυτό το σημείο πρέπει να τονίσουμε ότι στην περίπτωση του πολυκλιμακωτού Poisson-Haar σχήματος ανάλυσης εικόνων, το οποίο προτείναμε και παρουσιάσαμε στην Ενότητα 2.3.3, εφαρμόζουμε τρία ανεξάρτητα HMT μοντέλα, προκειμένου να μοντελοποιήσουμε τις δια-κλιμακωτές εξαρτήσεις των οριζόντιων, κάθετων και διαγώνιων συντελεστών λόγου έντασης. Συγκεκριμένα, κάθε ένα από τα HMT μοντέλα είναι υπεύθυνο για τη μοντελοποίηση των εξαρτήσεων μεταξύ των κλιμάκων μιας εκ των τριών τυχαίων μεταβλητών $\boldsymbol{\Theta}^h$, $\boldsymbol{\Theta}^v$ και $\boldsymbol{\Theta}^d$. Επιπροσθέτως, το διάνυσμα με τις παρατηρήσεις των κόμβων-παιδιών, το οποίο εμφανίζεται στην Polya κατανομή, αποτελείται από τους όρους της ενδιάμεσης κλίμακας, δηλαδή $\mathbf{x}_n^{c,o} = [x_n^o, x_n - x_n^o]$ με το o να υποδηλώνει τον συγκεκριμένο προσανατολισμό και να δέχεται μια από τις τιμές (h, v, d). Αυτοί οι όροι λαμβάνονται με τον τρόπο που περιγράφηκε στην Εξ. (3.19) και το άθροισμά τους οδηγεί στο δείγμα παρατήρησης του κόμβου-πατέρα x_n. Η εφαρμογή τριών ανεξάρτητων HMT μοντέλων είναι επίσης συνηθισμένη στην περίπτωση του 2-Δ μετασχηματισμού κυματιδίων κατά την οποία ένα HMT μοντέλο χρησιμοποιείται για τη μοντελοποίηση των συντελεστών λεπτομέρειας σε κάθε υποζώνη ανάλυσης [RCB01]. Κατά συνέπεια, το προκύπτον Poisson-Haar HMT μοντέλο θα παρουσιάζει βελτιωμένη κατευθυντική επιλεκτικότητα εν συγκρίσει με τις HMT παραλλαγές που χρησιμοποιούν το διαχωρίσιμο και το τετραδικό σχήμα ανάλυσης.

3.7 Πειράματα και Συγκρίσεις

Προκειμένου να επιβεβαιώσουμε την αποδοτικότητα των προτεινόμενων μεθόδων εκτίμησης έντασης της πηγαίας εικόνας, παραθέτουμε πειραματικές συγκρίσεις με άλλες ανταγωνιστικές μεθόδους αποθορυβοποίησης. Οι μέθοδοι που θα χρησιμοποιηθούν στις συγκρίσεις μας ανήκουν σε δύο διαφορετικές κατηγορίες. Στην Ενότητα 3.7.1 συγκρίνουμε μεθόδους οι οποίες έχουν σχεδιασθεί αποκλειστικά για να εφαρμόζονται σε Poisson δεδομένα, όπως συμβαίνει και με τις δικές μας προτεινόμενες μεθόδους. Στην Ενότητα 3.7.2 παραθέτουμε συγκρίσεις με μεθόδους οι οποίες σε ένα πρώτο βήμα προεπεξεργάζονται τα Poisson δεδομένα μέσω ενός σταθεροποιητικού μετασχηματισμού διακύμανσης (VST), προκειμένου να μετασχηματίσουν τα στατιστικά του θορύβου και, στη συνέχεια, εφαρμόζουν τεχνικές αποθορυβοποίησης που έχουν σχεδιασθεί για να χειρίζονται ομογενή προσθετικό γκαουσιανό θόρυβο.

Όλες οι συγκρίσεις μας πραγματοποιήθηκαν σε ένα σύνολο από πέντε τυπικές εικόνες ελέγχου, κλίμακας του γκρι, των 8-bit και απεικονίζονται στο Σχ. 3.6. Το μέγεθός τους είναι 512×512 εικονοστοιχεία εκτός από την εικόνα cameraman της οποίας το μέγεθος είναι 256×256 εικονοστοιχεία. Η απόδοση των υπό εξέταση τεχνικών ελέγχθηκε για επτά διαφορετικά επίπεδα έντασης τα οποία αντιστοιχούν σε διαφορετικά επίπεδα θορύβου. Συγκεκριμένα, κάθε εικόνα κανονικοποιήθηκε, ώστε τα εικονοστοιχεία της να έχουν μέγιστες τιμές $(1\,, 2\,, 3\,, 4\,, 5\,, 10\,, 20)$. Στη συνέχεια, οι υλοποιήσεις των Poisson μετρήσεων πραγματοποιήθηκαν χρησιμοποιώντας μια γεννήτρια τυχαίων Poisson αριθμών. Επειδή ο θόρυβος Poisson

Σχήμα 3.6: Το σύνολο όλων των εικόνων που χρησιμοποιούνται για τις συγκρίσεις μεταξύ των αλγορίθμων. Οι εικόνες είναι, από αριστερά προς τα δεξιά, οι Lena, Barbara, boat, fingerprint και cameraman.

εξαρτάται από τις τιμές έντασης της πραγματικής εικόνας με τοπικό $SNR = \sqrt{\lambda_j(k,\ell)}$, το επίπεδο του θορύβου αυξάνεται όσο η ένταση της εικόνας μειώνεται. Η μέση τιμή της έντασης για κάθε επιλεγμένη μέγιστη τιμή σε όλες τις εικόνες κυμαίνεται στα παρακάτω εύρη $(0.47 - 0.56,\ 0.94 - 1.11,\ 1.40 - 1.67,\ 1.88 - 2.22,\ 2.35 - 2.78,\ 4.70 - 5.55,\ 9.40 - 11.10)$, καλύπτοντας έτσι μια ευρεία κλίμακα από συνθήκες θορύβου.

3.7.1 Συγκρίσεις με Μπεϋζιανές Μεθόδους Σχεδιασμένες Αποκλειστικά για Poisson Δεδομένα

Σε αυτή την ενότητα συγκρίνουμε την απόδοση των προτεινόμενων αλγορίθμων εκτίμησης της γενεσιουργού έντασης (πηγαία εικόνα) με τις Poisson μεθόδους αποθορυβοποίησης των Timmerman και Nowak (TN) [TN99] και των Lu, Kim και Anderson (LKA) [LKA04]. Και οι δύο αυτές μέθοδοι υιοθετούν ένα πολυκλιμακωτό Μπεϋζιανό πλαίσιο εργασίας για την ανάλυση εικόνων περιορισμένης φωτεινότητας και, ουσιαστικά, διαφέρουν στη μέθοδο εκτίμησης των παραμέτρων που χρησιμοποιούν, όπως περιγράφηκε συνοπτικά στην Ενότητα 3.5.

Όλα τα αποτελέσματα που παρουσιάζονται αφορούν στις εκδοχές των μεθόδων οι οποίες αναλύουν τις εικόνες σε 5 κλίμακες και κάνουν χρήση της τεχνικής της κυκλικής μετατόπισης. Οι εκδοχές αυτές προκύπτουν λαμβάνοντας τη μέση τιμή των εκτιμήσεων που παράγονται από κάθε μέθοδο για διάφορες κυκλικές μετατοπίσεις της θορυβώδους εικόνας [CD95]. Στα πειράματα που διενεργήσαμε χρησιμοποιήσαμε 16 τυχαίες κυκλικές μετατοπίσεις στην περίπτωση της 256×256 εικόνας του cameraman και 32 για όλες τις άλλες 512×512 εικόνες με μόνη διαφορά την LKA μέθοδο στην οποία χρησιμοποιήσαμε 64 κυκλικές μετατοπίσεις, όπως ακριβώς περιγράφεται στο [LKA04]. Ο υπολογισμός της μέσης τιμής των παραγόμενων εκτιμήσεων οδηγεί σε εικόνες υψηλότερης ποιότητας και κατορθώνει να μειώνει σημαντικά τα τεχνουργήματα που προκύπτουν λόγω του φαινομένου της μη-στασιμότητας στο χώρο (δες την Ενότητα 3.4.3). Σε αυτό το σημείο θα πρέπει να αναφέρουμε ότι από την πειραματική μας εμπειρία η χρήση περισσότερων μετατοπίσεων οδηγεί σε αμελητέες βελτιώσεις συγκριτικά με τον επιπλέον υπολογιστικό χρόνο που απαιτείται. Ακόμη, επειδή η μέθοδος TN δεν μπορεί να χειριστεί την εκτίμηση παραμέτρων για περισσότερες από τρεις συμμετρικές συνιστώσες μείγματος, χρησιμοποιούμε και εμείς στα ακόλουθα πειράματα αυτόν τον αριθμό συνιστωσών, προκειμένου να επιτύχουμε ακριβείς συγκρίσεις μεταξύ των μεθόδων. Ωστόσο, οι τεχνικές μας μπορούν να εφαρμοστούν εξίσου αποδοτικά για την εκτίμηση των παραμέτρων και σε μη-συμμετρικές κατανομές για οποιοδήποτε αριθμό συνιστωσών μείγματος περιοριζόμενες μόνο από το υπολογιστικό κόστος και την ποσότητα των δεδομένων εκπαίδευσης.

Η ποιότητα των παραγόμενων εικόνων από όλες τις μεθόδους αποτιμάται με όρους κορυφοτιμής σηματοθορυβικού λόγου (Peak Signal to noise ratio - PSNR) και ορίζεται ως $10\log_{10}\left(\text{peak}^2/\text{MSE}\right)$. Με "peak" δηλώνουμε τη μέγιστη ένταση της πηγαίας/"καθαρής" εικόνας

Πίνακας 3.2: Εκτίμηση της έντασης φωτεινότητας εικόνων αλλοιωμένων από συνθετικό Poisson θόρυβο χρησιμοποιώντας μείγμα κατανομών με 3 συμμετρικές συνιστώσες. Τα αποτελέσματα παρουσιάζονται σε όρους PSNR (dB) για διάφορα επίπεδα Poisson θορύβου.

	PSNR (dB) / Methods								
Image/ Peak Int.	noisy	TN	LKA	Sep IND	Quad IND	PH. IND	Sep HMT	Quad HMT	PH HMT
Cam./1	3.29	19.04	19.23	19.48	19.84	19.68	19.88	20.11	20.03
Cam./2	6.29	20.04	20.39	20.79	21.00	20.82	21.22	21.48	21.41
Cam./3	8.06	20.66	21.05	21.38	21.62	21.52	22.10	22.24	22.31
Cam./4	9.30	21.16	21.52	21.77	22.39	22.19	22.74	22.74	22.90
Cam./5	10.26	21.59	21.91	22.22	22.84	22.65	23.22	23.18	23.37
Cam./10	13.28	22.86	23.18	23.85	24.08	23.93	24.65	24.64	24.97
Cam./20	16.29	24.35	24.67	24.99	25.90	25.77	26.39	26.23	26.61
Lena/1	2.97	21.32	21.84	22.16	22.32	22.30	22.47	22.58	22.66
Lena/2	5.96	22.48	23.05	23.25	23.56	23.48	23.77	23.86	23.91
Lena/3	7.73	23.23	23.76	23.99	24.30	24.17	24.55	24.60	24.69
Lena/4	8.97	23.79	24.29	24.55	24.81	24.68	25.08	25.18	25.29
Lena/5	9.95	24.19	24.72	25.01	25.24	25.11	25.59	25.68	25.78
Lena/10	12.96	25.53	25.98	26.23	26.60	26.48	27.03	27.09	27.21
Lena/20	15.97	26.97	27.31	27.69	28.15	27.90	28.45	28.41	28.66
Boat/1	2.94	21.06	21.17	21.37	21.43	21.44	21.62	21.66	21.76
Boat/2	5.95	21.90	22.14	22.21	22.36	22.37	22.58	22.60	22.77
Boat/3	7.71	22.43	22.71	22.79	22.97	23.02	23.22	23.20	23.45
Boat/4	8.96	22.82	23.11	23.22	23.36	23.41	23.63	23.65	23.90
Boat/5	9.93	23.17	23.46	23.59	23.70	23.77	24.08	24.04	24.31
Boat/10	12.94	24.23	24.51	24.65	25.00	25.05	25.32	25.21	25.57
Boat/20	15.95	25.43	25.73	26.10	26.23	26.34	26.59	26.48	26.96
Barb./1	3.22	19.66	20.00	20.22	20.29	20.29	20.37	20.40	20.48
Barb./2	6.22	20.51	20.85	20.97	21.07	21.05	21.20	21.18	21.27
Barb./3	7.99	21.01	21.31	21.43	21.52	21.53	21.65	21.62	21.72
Barb./4	9.24	21.35	21.64	21.78	21.84	21.87	21.98	21.96	22.07
Barb./5	10.20	21.62	21.87	22.01	22.06	22.08	22.22	22.19	22.33
Barb./10	13.21	22.47	22.58	22.79	22.84	22.94	23.14	23.05	23.45
Barb./20	16.22	23.34	23.30	23.74	23.81	24.05	24.65	24.36	24.92
Fgrpt./1	2.56	16.42	16.70	17.24	17.19	17.28	17.34	17.23	17.39
Fgrpt./2	5.56	17.09	17.24	18.28	18.23	18.33	18.46	18.33	18.55
Fgrpt./3	7.33	17.70	17.68	19.02	18.97	19.08	19.27	19.09	19.36
Fgrpt./4	8.57	18.16	18.03	19.60	19.55	19.66	19.86	19.66	19.94
Fgrpt./5	9.55	18.58	18.36	20.05	20.01	20.12	20.32	20.12	20.42
Fgrpt./10	12.56	20.19	19.80	21.56	21.54	21.66	21.79	21.62	21.91
Fgrpt./20	15.58	22.02	21.79	23.13	23.15	23.26	23.32	23.19	23.46

και το MSE είναι το μέσο τετραγωνικό σφάλμα μεταξύ της αποκατεστημένης και πηγαίας εικόνας. Στον Πίνακα 3.2 παρουσιάζονται τα αποτελέσματα όλων των εικόνων ελέγχου σε όλες τις συνθήκες θορύβου που θεωρήσαμε. Το PSNR υπολογίζεται χρησιμοποιώντας 10 ανεξάρτητα πειράματα σε κάθε περίπτωση, ώστε τα αποτελέσματα να μην είναι πολωμένα. Όπως μπορούμε ξεκάθαρα να διαπιστώσουμε, χρησιμοποιώντας τις παραμέτρους του μοντέλου που προέκυψαν πλήρως αυτο-

ματοποιημένα από τις EM τεχνικές μας (μέσω της διαδικασίας αυτοδύναμης εκκίνησης) και τους προτεινόμενους αλγορίθμους, τα αποτελέσματα που λάβαμε είναι ποσοτικά καλύτερα συγκριτικά με αυτά των υπό εξέταση ανταγωνιστικών μεθόδων. Συγκεκριμένα, τα σχήματα ανάλυσης Sep-IND (διαχωρίσιμο), Quad-IND (τετραδικό), και PH-IND (Poisson-Haar), υπολογίζοντας τις παραμέτρους τους ανεξάρτητα σε κάθε κλίμακα μέσω της EM τεχνικής μας, παρουσιάζουν κατά μέσο όρο για όλες τις εικόνες και τις συνθήκες θορύβου μια βελτίωση περίπου της τάξης των $0.5 - 1$ dB σε σύγκριση με τα καλύτερα αποτελέσματα των μεθόδων TN [TN99] και LKA [LKA04]. Αν επιπλέον μοντελοποιήσουμε τις δια-κλιμακωτές εξαρτήσεις με το HMT μοντέλο, τότε οι μέθοδοί μας Sep-HMT, Quad-HMT, και PH-HMT επιτυγχάνουν μια περαιτέρω βελτίωση 0.5 dB. Όσον αφορά στις συγκρίσεις μεταξύ των δικών μας μεθόδων τότε στην περίπτωση που εκτιμούμε τις παραμέτρους ανεξάρτητα σε κάθε κλίμακα, το τετραδικό και το Poisson-Haar σχήμα συμπεριφέρονται περίπου το ίδιο και ελαφρώς καλύτερα από το διαχωρίσιμο. Ωστόσο, όταν εξετάζουμε τα αντίστοιχα σχήματα στην περίπτωση κατά την οποία η εκτίμηση γίνεται με τα HMT μοντέλα, τότε το PH-HMT, είναι καλύτερο δίνοντας κατά μέσο όρο μια βελτίωση περίπου της τάξης των 0.25 dB έναντι των άλλων.

Η αποτελεσματικότητα των μεθόδων μας σε σύγκριση με τις άλλες μεθόδους μπορεί και οπτικά να αποτιμηθεί από τα αντιπροσωπευτικά παραδείγματα αποθορυβοποίησης των εικόνων cameraman και boat που παρουσιάζονται στα Σχήματα. 3.7 και 3.8. Στο πρώτο σχήμα παρουσιάζουμε την αποθορυβοποιημένη εικόνα του cameraman εφαρμόζοντας τις ανταγωνιστικές μεθόδους TN και LKA και την προτεινόμενη Quad-HMT, ενώ στο δεύτερο σχήμα παρουσιάζουμε τα αποτελέσματα των μεθόδων TN, LKA, και της προτεινόμενης PH-HMT μεθόδου για την εικόνα boat. Από αυτά τα σχήματα μπορούμε να διαπιστώσουμε ότι οι προτεινόμενες μέθοδοι αφαιρούν αποδοτικότερα το θόρυβο και την ίδια στιγμή ελαχιστοποιούν τα τεχνουργήματα σε σχέση με τις άλλες μεθόδους. Επιπλέον, από το Σχ. 3.8 μπορούμε να επιβεβαιώσουμε τη δυνατότητα του προτεινόμενου Poisson-Haar σχήματος ανάλυσης να διατηρεί καλύτερα τη δομή των ακμών της εικόνας λόγω της βελτιωμένης κατευθυντικής επιλεκτικότητάς του. Όσον αφορά στον αριθμό των συνιστωσών μείγματος, επαναλαμβάνοντας τα ίδια πειράματα χρησιμοποιώντας μόνο μια συνιστώσα οδηγούμαστε σε μείωση της απόδοσης κατά περίπου $1.5 - 2$ dB, ενώ χρησιμοποιώντας 2 συνιστώσες παρατηρούμε μια μείωση μόνο $0.15 - 0.2$ dB σε σχέση με την περίπτωση των 3 συνιστωσών. Επιπροσθέτως, η χρήση 4 συνιστωσών οδηγεί σε αμελητέα αύξηση της απόδοσης.

Σχετικά με το υπολογιστικό κόστος μετρήσαμε το χρόνο εκτέλεσης όλων των εξεταζόμενων μεθόδων για την εικόνα του cameraman με μέγιστη ένταση 20. Ο πιο γρήγορος χρόνος μετρήθηκε για την TN μέθοδο με $8s$, ενώ η LKA μέθοδος απαιτεί περίπου $60s$. Ο υπολογιστικός χρόνος που χρειάστηκαν οι προτεινόμενες μέθοδοι για πλήρως αυτοματοποιημένη εκπαίδευση χρησιμοποιώντας μη-βελτιστοποιημένο κώδικα MATLAB, είναι $58s$ και $45s$ αντιστοίχως για τις Quad-IND και Quad-HMT μεθόδους, $89s$ και $80s$ για τις Sep-IND και Sep-HMT και $115s$ και $92s$ για τις PH-IND και PH-HMT μεθόδους. Σημειώνουμε ότι στις HMT εκδοχές δε χρησιμοποιήθηκε η τεχνική αυτοδύναμης εκκίνησης, αντίθετα οι εκτιμώμενες παράμετροι των αντίστοιχων ανεξάρτητων εκδοχών αποτέλεσαν τις αρχικές παραμέτρους για τον EM. Σημαντική μείωση του υπολογιστικού κόστους με μικρή μόνο απώλεια απόδοσης μπορεί να επιτευχθεί χρησιμοποιώντας 2 συνιστώσες μείγματος. Μια επιπλέον δυνατότητα για περαιτέρω μείωση είναι η απευθείας χρήση αρχικών παραμέτρων στον EM αλγόριθμο, βασιζόμενες σε πρότερη γνώση από προηγούμενη εκπαίδευση σε παρόμοιες εικόνες.

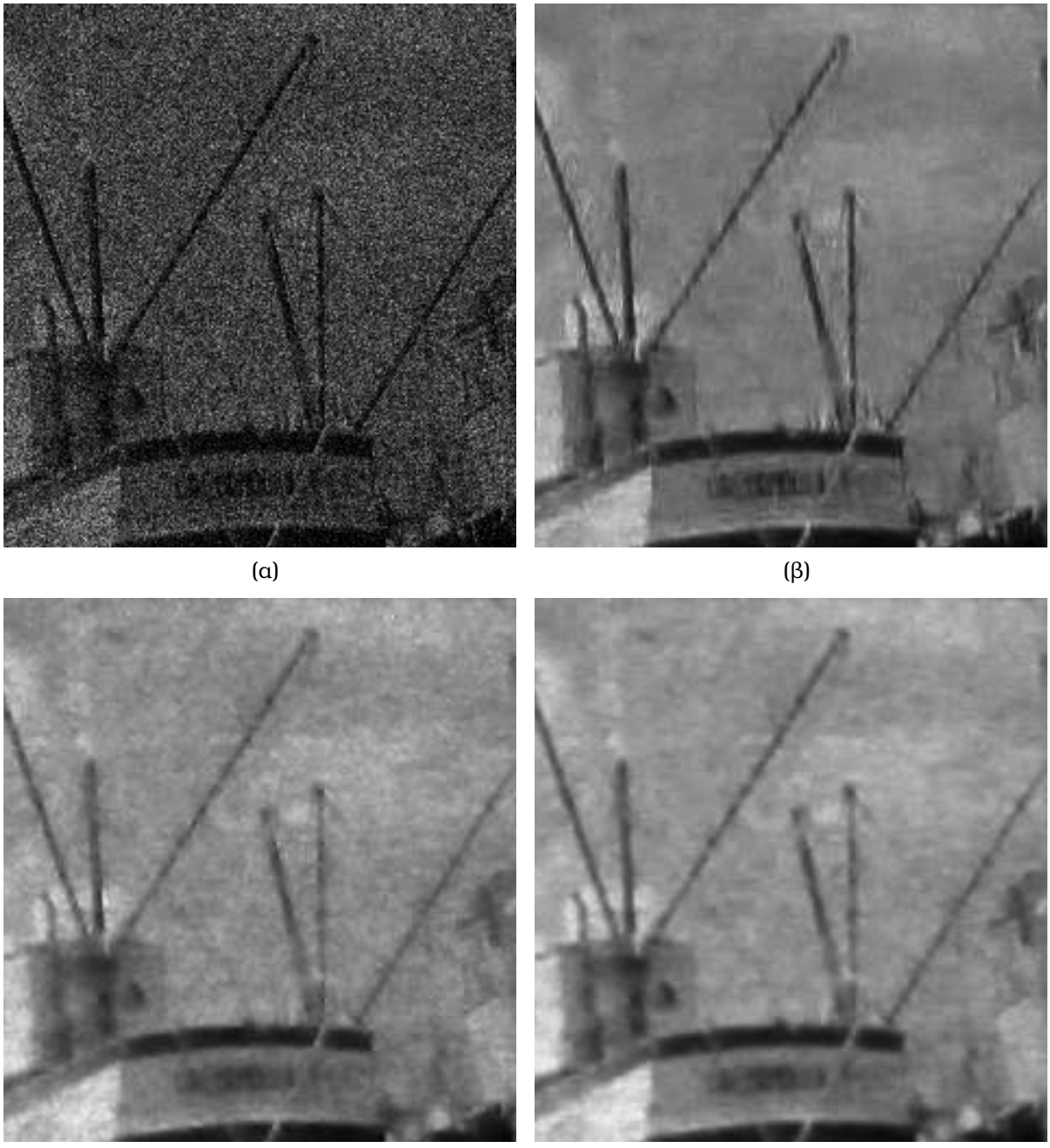
3.7.2 Συγκρίσεις με VST Μεθόδους

Σε αυτή την ενότητα παρέχουμε συγκριτικά αποτελέσματα της καλύτερης μεθόδου μας, PH-HMT, με VST μεθόδους οι οποίες εφαρμόζουν τις τρέχουσες state-of-the-art τεχνικές αποθορυβοποίησης. Ακολουθώντας τη VST προσέγγιση, αρχικά εφαρμόζουμε στα θορυβώδη δεδομένα, $\mathbf{x}$, τον Anscombe μετασχηματισμό [Ans48], προκειμένου να αναμορφώσουμε τα στατιστικά του θορύβου, ώστε στα μετασχηματισμένα δεδομένα, $\mathbf{y} = 2\sqrt{\mathbf{x} + 3/8}$, ο θόρυβος προσεγγιστικά να

(α) (β)

(γ) (δ)

Σχήμα 3.7: Αποτελέσματα αποθορυβοποίησης της εικόνας cameraman με μέγιστη τιμή έντασης 20 και συνθετικό Poisson θόρυβο. Ο μέσος αριθμός φωτονίων ανά εικονοστοιχείο στη θορυβώδη εικόνα είναι 9.4. (α) Θορυβώδης εικόνα (PSNR=16.29), (β) Quad-HMT (PSNR=26.23), (γ) TN [TN99] (PSNR=24.35), (δ) LKA [LKA04] (PSNR=24.67).



(α) (β)

(γ) (δ)

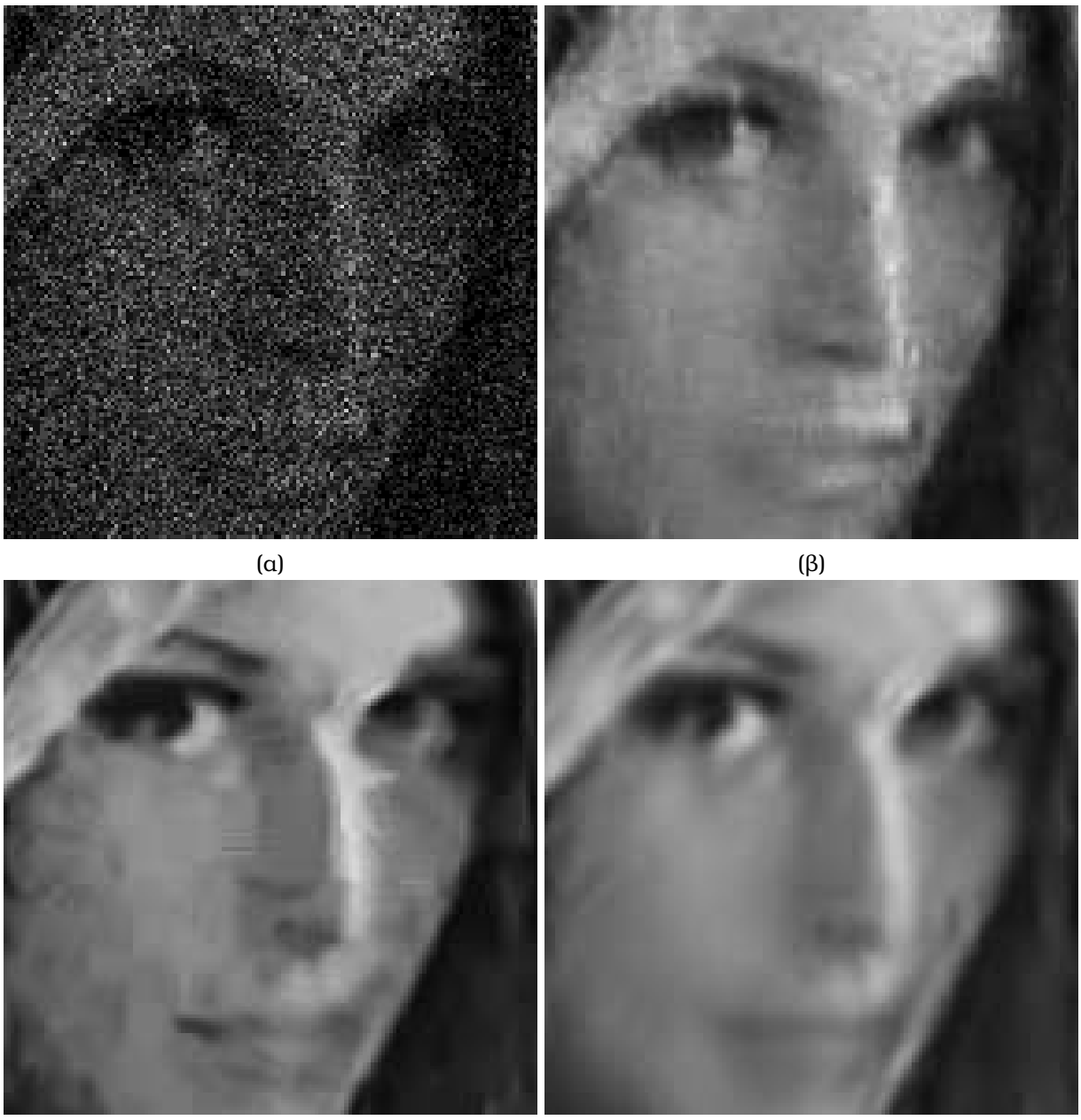
Σχήμα 3.8: Αποτελέσματα αποθορυβοποίησης της εικόνας boat με μέγιστη τιμή έντασης 10 και συνθετικό Poisson θόρυβο. Ο μέσος αριθμός φωτονίων ανά εικονοστοιχείο στη θορυβώδη εικόνα είναι 5. (α) Θορυβώδης εικόνα (PSNR=12.94), (β) PH-HMT (PSNR=25.57),(γ) TN [TN99] (PSNR=24.23), (δ) LKA [LKA04] (PSNR=24.51).

ισούται με γκαουσιανό θόρυβο μοναδιαίας διακύμανσης. Στη συνέχεια, η αποθορυβοποίηση της μετασχηματισμένης εικόνα $\mathbf{y}$ πραγματοποιείται κάνοντας χρήση μεθόδων που έχουν σχεδιαστεί για να χειρίζονται αποκλειστικά γκαουσιανά δεδομένα. Το τελικό αποτέλεσμα της αποθορυβοποίησης λαμβάνεται μέσω του αντίστροφου μετασχηματισμού Anscombe, $\hat{\mathbf{x}} = (\hat{\mathbf{y}}/2)^2 - 1/8$. Θα πρέπει να επισημάνουμε ότι ο εφαρμοζόμενος αντίστροφος Anscombe μετασχηματισμός, ο οποίος δεν ταυτίζεται με τον αλγεβρικό τύπο αντιστροφής, $\hat{\mathbf{x}}_{alg} = (\hat{\mathbf{y}}/2)^2 - 3/8$, αντισταθμίζει την πόλωση της μέσης τιμής που εισάγεται από τον ευθύ μετασχηματισμό. Ως γκαουσιανές τεχνικές αποθορυβοποίησης μέσα στο VST πλαίσιο εργασίας εφαρμόζουμε δύο εναλλακτικές state-of-the-art μεθόδους. Η πρώτη μέθοδος είναι η BLS-GSM (Bayes least squares Gaussian scale mixture) [PSWS03] που είναι μια παραμετρική τεχνική και η δεύτερη είναι η BM3D (block-matching and 3D filtering) [DFKE07] η οποία ανήκει στην κατηγορία των πρόσφατα προτεινόμενων μη-τοπικών, μη-παραμετρικών τεχνικών. Επιπλέον, παρέχουμε συγκρίσεις μεταξύ της PH-HMT μεθόδου μας και της πολυκλιμακωτής MS-VST μεθόδου του [ZFS08]. Όπως και στην προηγούμενη ενότητα, έτσι και εδώ όλα τα αποτελέσματα ελήφθησαν ως ο μέσος όρος 10 ανεξάρτητων πειραμάτων.

Από τα αποτελέσματα που παρουσιάζονται στον Πίνακα 3.3, παρατηρούμε ότι η μέθοδός μας για την περίπτωση μικρού αριθμού Poisson μετρήσεων (χαμηλή φωτεινότητα εικόνων), δηλαδή για μέγιστη ένταση των εικονοστοιχείων μικρότερη του 5, δίνει σχεδόν πάντα καλύτερα αποτελέσματα από τις BLS-GSM και BM3D μεθόδους. Αυτό είναι αναμενόμενο, επειδή ο Anscombe VST μετασχηματισμός στις περιοχές χαμηλού σηματοθορυβικού λόγου δεν καταφέρνει να προσεγγίσει ικανοποιητικά τον Poisson θόρυβο με γκαουσιανό. Για τις περιπτώσεις μεσαίου επιπέδου φωτεινότητας η μέθοδός μας εξακολουθεί να είναι συγκρίσιμη με τις άλλες δύο μεθόδους με εξαίρεση τις εικόνες Barbara και fingerprint. Αυτή η διαφορά απόδοσης για τις δύο τελευταίες εικόνες οφείλεται στην πλούσια πληροφορία υφής του περιεχομένου τους την οποία ο αλγόριθμός μας δεν αξιοποιεί πλήρως· από την άλλη μεριά, οι μη-τοπικοί αλγόριθμοι αποθορυβοποίησης, όπως ο BM3D, είναι πολύ αποτελεσματικοί στο να αξιοποιούν αποδοτικά τα περιοδικά μοτίβα που κυριαρχούν στις εικόνες υφής. Τέλος, για υψηλά επίπεδα φωτεινότητας (υψηλό SNR) ο σταθεροποιητικός μετασχηματισμός διακύμανσης οδηγεί σε δεδομένα που ακολουθούν με μεγάλη ακρίβεια γκαουσιανά στατιστικά και, συνεπώς, οι αλγόριθμοι BLS-GSM και BM3D παρουσιάζουν καλύτερη απόδοση από την προτεινόμενη τεχνική.

Πέρα από τη σύγκριση με όρους PSNR, προκειμένου να έχουμε και μια ποιοτική αίσθηση της τυπικής συμπεριφοράς της προτεινόμενης μεθόδου σε σχέση με τις άλλες δύο VST τεχνικές, παρουσιάζουμε στο Σχ. 3.9 συγκριτικά αποτελέσματα αποθορυβοποίησης για την εικόνα Lena, εστιάζοντας στο πρόσωπο του ατόμου, η οποία έχει αλλοιωθεί από συνθετικό Poisson θόρυβο. Από αυτό το παράδειγμα μπορεί κανείς να δει ότι, παρά το γεγονός ότι η προτεινόμενη μέθοδος εμφανίζει χαμηλότερο PSNR, εισάγει σαφώς λιγότερα οπτικά τεχνουργήματα συγκριτικά με τις άλλες δύο μεθόδους. Η ιδιότητα αυτή της μεθόδου PH-HMT είναι σημαντική σε εφαρμογές, όπως η επεξεργασία αστρονομικών και βιοϊατρικών εικόνων, στις οποίες είναι κρίσιμο η διαδικασία αποθορυβοποίησης να οδηγεί σε αποτελέσματα που να είναι όσο το δυνατόν πιο πιστά στο πραγματικό οπτικό περιεχόμενο της εικόνας. Όσον αφορά το υπολογιστικό κόστος της VST τεχνικής χρησιμοποιώντας μια εκ των BM3D και BLS-GSM μεθόδων ως ενδιάμεσες ρουτίνες γκαουσιανής αποθορυβοποίησης μετρήσαμε, όπως και στην Ενότητα 3.7.1, τους χρόνους εκτέλεσης για την εικόνα cameraman με μέγιστη ένταση φωτεινότητας 20. Η γρηγορότερη εκτέλεση πραγματοποιήθηκε από την BM3D, $3s$ με χρήση βελτιστοποιημένου MATLAB/MEX κώδικα, ενώ ο χρόνος εκτέλεσης της BLS-GSM ήταν $18s$.

Τέλος, παρέχουμε συγκρίσεις της PH-HMT μεθόδου με τα καλύτερα αποτελέσματα της MS-VST μεθόδου σε τέσσερις διαφορετικές εικόνες που αναφέρονται στο [ZFS08]. Επειδή η υλοποίηση της συγκεκριμένης μεθόδου δεν είναι ακόμα διαθέσιμη, εφαρμόσαμε τη μέθοδό μας μόνο στα επίπεδα έντασης φωτεινότητας που αναφέρονται για κάθε εικόνα στο [ZFS08]. Επι-



(α) (β)

(γ) (δ)

Σχήμα 3.9: Αποτελέσματα αποθορυβοποίησης της εικόνας Lena με μέγιστη τιμή έντασης 40 και συνθετικό Poisson θόρυβο. Ο μέσος αριθμός φωτονίων ανά εικονοστοιχείο στη θορυβώδη εικόνα είναι 2.5. (α) Αλλοιωμένη εικόνα (PSNR=9.95), (β) PH-HMT(PSNR=25.78), (γ) BM3D [DFKE07] (PSNR=26.52), (δ) BLS-GSM [PSWS03] (PSNR=26.16). Σημειώνουμε ότι, παρόλο που η προτεινόμενη μέθοδος δίνει χαμηλότερο PSNR, εισάγει σαφώς λιγότερα οπτικά τεχνουργήματα συγκριτικά με τις άλλες δύο VST μεθόδους.

Πίνακας 3.3: Συγκριτικά αποτελέσματα της PH-HMT μεθόδου με δύο VST μεθόδους που χρησιμοποιούν τις BLS-GSM [PSWS03] και BM3D [DFKE07] τεχνικές ως ενδιάμεσες ρουτίνες για γκαουσιανή αποθορυβοποίηση. Τα αποτελέσματα παρουσιάζονται σε όρους PSNR (dB) για διάφορες τιμές εντάσεων που αντιστοιχούν σε διαφορετικά επίπεδα Poisson θορύβου.

Methods	Peak Intensities						
	1	2	3	4	5	10	20
Cameraman							
noisy	3.29	6.29	8.06	9.30	10.26	13.28	16.29
BM3D	14.90	20.13	22.18	23.41	24.16	26.13	27.69
BLS-GSM	14.37	18.99	20.72	21.64	22.43	24.53	26.54
PH-HMT	20.03	21.41	22.31	22.90	23.37	24.97	26.61
Lena							
noisy	2.97	5.96	7.73	8.97	9.95	12.96	15.97
BM3D	16.14	22.72	24.93	25.91	26.52	28.30	29.94
BLS-GSM	15.84	21.99	24.24	25.36	26.16	28.02	29.56
PH-HMT	22.66	23.91	24.69	25.29	25.78	27.21	28.66
Boat							
noisy	2.94	5.95	7.71	8.96	9.93	12.94	15.95
BM3D	15.83	21.62	23.33	24.15	24.70	26.27	27.84
BLS-GSM	15.57	20.86	22.68	23.65	24.32	25.94	27.40
PH-HMT	21.76	22.77	23.45	23.90	24.31	25.57	26.96
Barbara							
noisy	3.22	6.22	7.99	9.24	10.20	13.21	16.22
BM3D	15.31	20.79	22.76	23.74	24.41	26.34	28.20
BLS-GSM	14.96	19.87	21.33	21.92	22.57	24.66	26.50
PH-HMT	20.48	21.27	21.72	22.07	22.33	23.45	24.92
Fingerprint							
noisy	2.56	5.56	7.33	8.57	9.55	12.56	15.58
BM3D	14.56	19.39	20.93	21.67	22.20	23.81	25.39
BLS-GSM	13.92	18.87	20.53	21.34	21.88	23.39	24.96
PH-HMT	17.39	18.55	19.36	19.94	20.42	21.91	23.46

πλέον, για να επιτρέψουμε άμεσες συγκρίσεις, εφαρμόσαμε το *κανονικοποιημένο μέσο ολοκληρώσιμο τετραγωνικό σφάλμα* (normalized mean integrated square error - NMISE) που χρησιμοποιείται από τους συγγραφείς ως μετρικό ποιότητας. Το συγκεκριμένο μετρικό ορίζεται ως $\text{NMISE} = E\left\{\left(\sum_i^N \left(\hat{\lambda}_i - \lambda_i\right)^2 / \lambda_i\right) / N\right\}$, όπου $\hat{\lambda}_i$ είναι η εκτιμώμενη ένταση φωτεινότητας. Στα αποτελέσματα που παρουσιάζονται στον Πίνακα 3.4, η μέθοδός μας PH-HMT έδωσε καλύτερα αποτελέσματα από την MS-VST σε 3 από τις 4 εικόνες, γεγονός που υποδηλώνει ότι είναι ανταγωνιστική με τις state-of-the-art VST μεθόδους.

Πίνακας 3.4: Συγκριτικά αποτελέσματα της PH-HMT μεθόδου με τα καλύτερα αποτελέσματα της MSVST μεθόδου, τα οποία αναφέρονται στο [ZFS08], για τέσσερις εικόνες (Spots, Galaxy, Barbara, Cells). Τα αποτελέσματα παρουσιάζονται σε όρους NMISE (η μικρότερη τιμή υποδηλώνει καλύτερη απόδοση).

Methods	Images/Intensity range			
	Spots [0.03 , 5.02]	Galaxy [0 , 5]	Barbara [0.93 , 15.73]	Cells [0.53 , 16.93]
noisy	1.002	1.002	0.999	1.002
MS-VST	0.069	0.035	0.170	0.078
PH-HMT	0.048	0.030	0.159	0.082

3.7.3 Αποθορυβοποίηση Εικόνων με Πραγματικό Poisson Θόρυβο

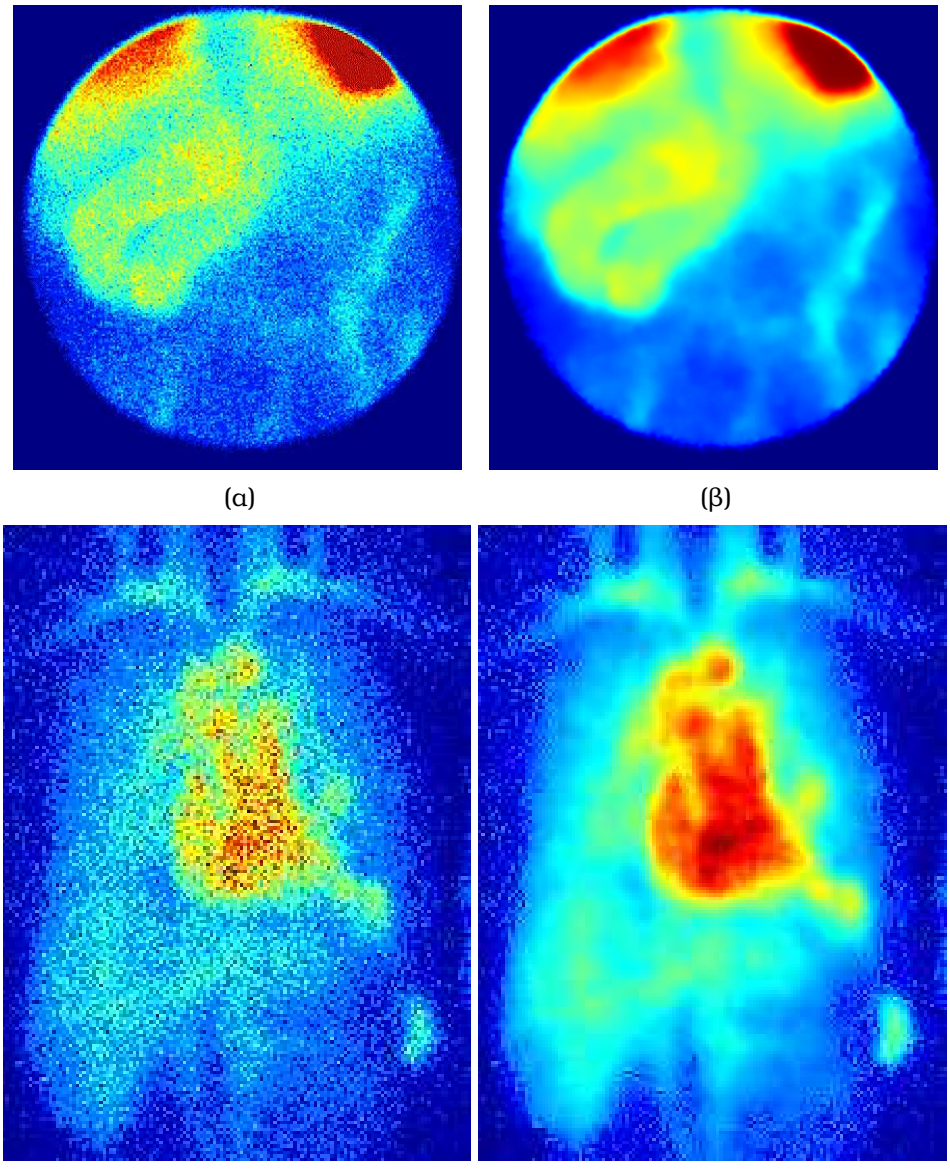
Εκτός από τις συγκρίσεις που παρείχαμε με εναλλακτικές μεθόδους αποθορυβοποίησης σε εικόνες ελέγχου, παραθέτουμε και τα αποτελέσματα που προέκυψαν εφαρμόζοντας δύο από τις προτεινόμενες μεθόδους, Quad-HMT και PH-HMT, σε βιοϊατρικές εικόνες με πραγματικό θόρυβο. Ο στόχος μας είναι να εξετάσουμε κατά πόσο οι προτεινόμενες μέθοδοι μπορούν να εφαρμοστούν με επιτυχία και σε πραγματικές συνθήκες.

Οι βιοϊατρικές εικόνες που χρησιμοποιήσαμε έχουν ληφθεί με απεικονιστικές τεχνικές της πυρηνικής ιατρικής. Σε αυτή την περίπτωση για την καταγραφή των εικόνων απαιτείται η εισαγωγή ραδιενεργών φαρμάκων στον οργανισμό. Καθώς το ραδιενεργό φάρμακο διασπάται, εκπέμπονται ακτίνες γ οι οποίες μπορούν να διαπεράσουν τους ιστούς του ανθρώπινου σώματος και να ανιχνευθούν από μια ειδική κάμερα, καθιστώντας δυνατή τη μελέτη φυσιολογικών και βιοχημικών διαδικασιών του ασθενούς. Η συσκευή καταγραφής μετατρέπει την ανιχνευόμενη ακτινοβολία σε φως με αποτέλεσμα η τιμή της έντασης κάθε εικονοστοιχείου της εικόνας να σχετίζεται με έναν αριθμό ανιχνευόμενων φωτονίων. Οι τιμές αυτές ακολουθούν την κατανομή Poisson, λόγω της τυχαίας φύσης της διάσπασης του χορηγούμενου ραδιενεργού υλικού.

Στο Σχ. 3.10 παρουσιάζονται δύο ακατέργαστες βιοϊατρικές εικόνες που λάβαμε από το [WUM] και οι αντίστοιχες αποθορυβοποιημένες εκδοχές τους με τη χρήση των προτεινόμενων μεθόδων. Το Σχ. 3.10(α) απεικονίζει μια βιοϊατρική εικόνα κοιλίας, ενώ το Σχ. 3.10(β) παρουσιάζει την εκτίμηση της πραγματικής φωτεινότητας που προέκυψε εφαρμόζοντας τη μέθοδο Quad-HMT. Ο μέσος αριθμός φωτονίων στην αρχική εικόνα είναι 3.5. Αυτό υποδηλώνει ότι η αλλοίωση της εικόνας είναι σημαντική, όπως μπορούμε να παρατηρήσουμε και οπτικά. Στο Σχ. 3.10(γ) αποτυπώνεται μια βιοϊατρική εικόνα θώρακα. Σε αυτή την περίπτωση ο μέγιστος αριθμός φωτονίων είναι 5.4. Το Σχ. 3.10(δ) απεικονίζει την εκτίμηση της πραγματικής φωτεινότητας της εικόνας με την εφαρμογή της PH-HMT μεθόδου. Από τις παραπάνω εικόνες βλέπει κανείς ότι οι προτεινόμενες μέθοδοι επιτυγχάνουν να βελτιώσουν την ποιότητα των αρχικών εικόνων και ταυτόχρονα διατηρούν αναλλοίωτη τη δομή τους.

3.8 Εφαρμογές σε Αστρονομικά Δορυφορικά Δεδομένα

Το ενδιαφέρον μας για την εκτίμηση της έντασης ετερογενών Poisson στοχαστικών διαδικασιών υποκινείται από την άμεση σχέση που εμφανίζει με το πρόβλημα της επεξεργασίας δορυφορικών εικόνων. Οι δορυφορικές εικόνες, όπως για παράδειγμα οι αστρονομικές, συλλαμβάνονται συνήθως υπό συνθήκες χαμηλού φωτισμού και εντάσσονται επομένως στην κατηγορία εικόνων περιορισμένης φωτεινότητας (photon-limited). Όλες αυτές οι εικόνες συνήθως υφίστανται το φαι-

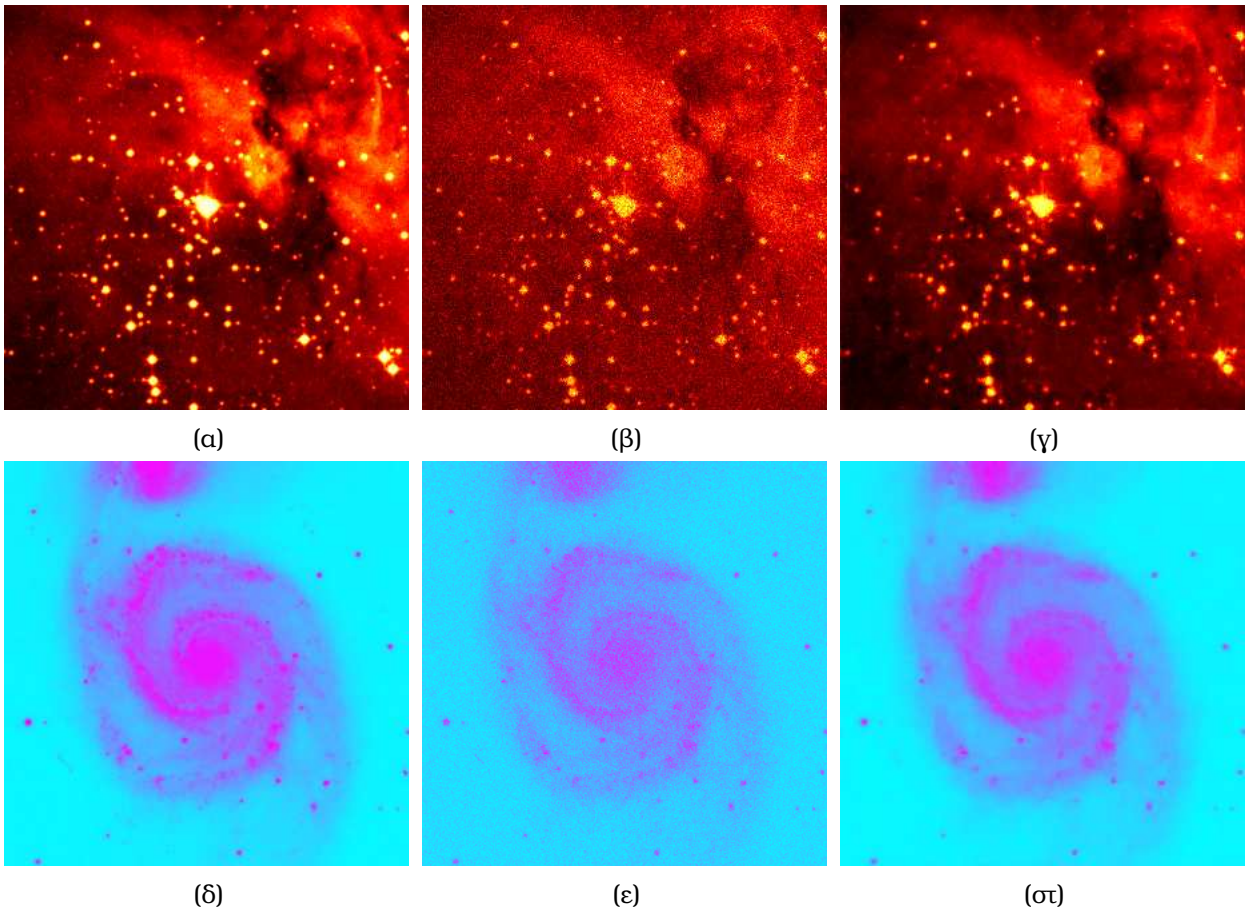


Σχήμα 3.10: Αποθορυβοποίηση βιοϊατρικών εικόνων. (α) Αρχική εικόνα κοιλίας, (β) Αποτέλεσμα εκτίμησης με την προτεινόμενη Quad-HMT μέθοδο, (γ) Αρχική εικόνα θώρακα, (δ) Αποτέλεσμα εκτίμησης με την προτεινόμενη PH-HMT μέθοδο.

νόμενο του κβαντικού θορύβου εξαιτίας της διακύμανσης των ανιχνευόμενων φωτονίων κατά τη διάρκεια της σύλληψής τους. Σε τέτοιες περιπτώσεις ο αριθμός των ανιχνευθέντων φωτονίων μπορεί να μοντελοποιηθεί με επιτυχία ως να προέκυπτε από την υλοποίηση μιας χρονικά ετερογενούς και χωρικά μεταβαλλόμενης Poisson στοχαστικής διαδικασίας. Στην παρούσα ενότητα, προκειμένου να αποτιμήσουμε τις δυνατότητες των προτεινόμενων μεθόδων για τέτοιου είδους εφαρμογές, παρουσιάζουμε αποτελέσματα εκτίμησης της έντασης φωτεινότητας σε δορυφορικές αστρονομικές εικόνες.

3.8.1 Δορυφορικές Εικόνες με Συνθετικό Poisson Θόρυβο

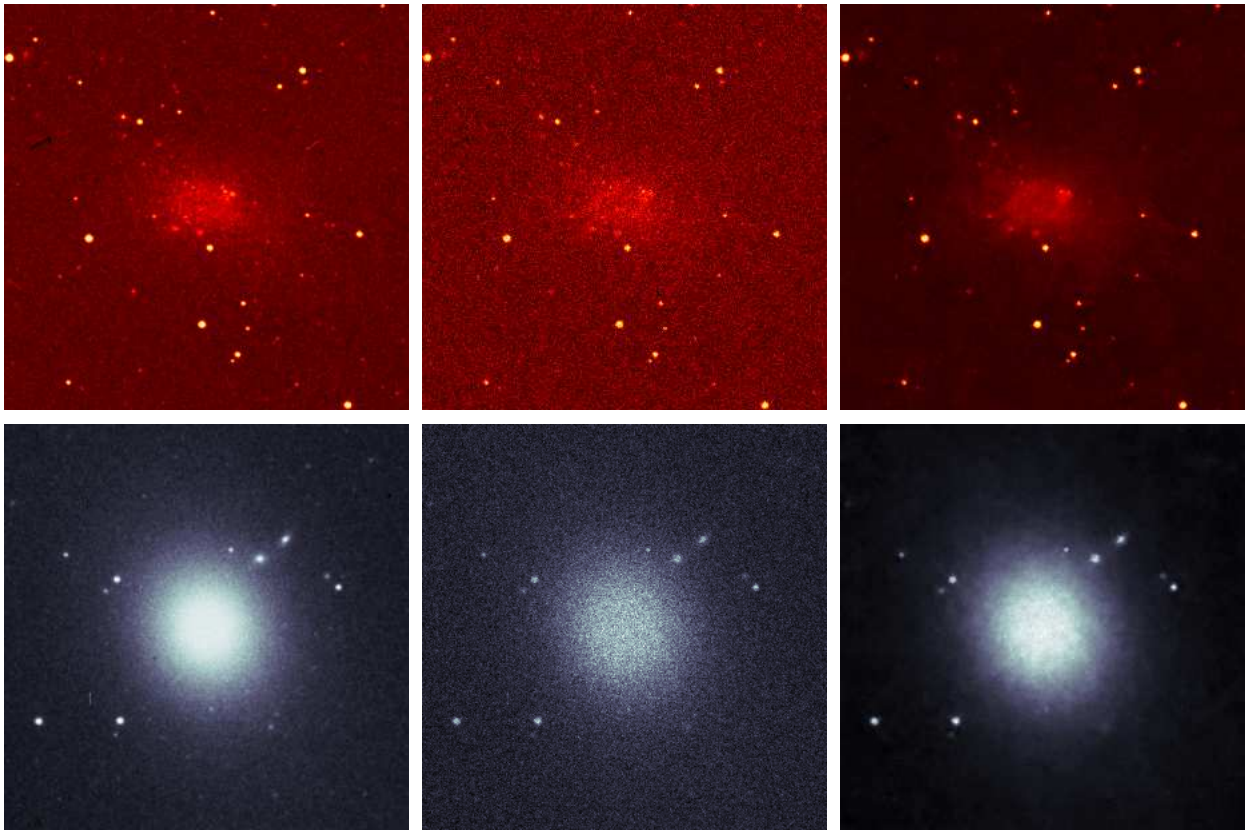
Στην ενότητα αυτή παρουσιάζουμε πειραματικά αποτελέσματα των προτεινόμενων μεθόδων σε μια σειρά από αστρονομικές δορυφορικές εικόνες. Στις εικόνες αυτές έχουμε προσθέσει συνθετικό Poisson θόρυβο ακολουθώντας την ίδια διαδικασία με αυτή που περιγράψαμε στην Ενοτ. 3.7. Για κάθε εικόνα παρουσιάζουμε την εκτίμηση που προέκυψε από τη χρήση μίας εκ των προτεινόμενων μεθόδων, από εκείνες που υϊοθετούν το HMT στατιστικό μοντέλο. Παράλληλα, επειδή στη διάθεση μας έχουμε την πραγματική ένταση φωτεινότητας της εικόνας, παρέχουμε το SNR της θορυβώδους εικόνας αλλά και της αποκατεστημένης εκδοχής της, προκειμένου εκτός από το οπτικό αποτέλεσμα να έχουμε και μια ποσοτική ένδειξη της βελτίωσης που επιτυγχάνεται.



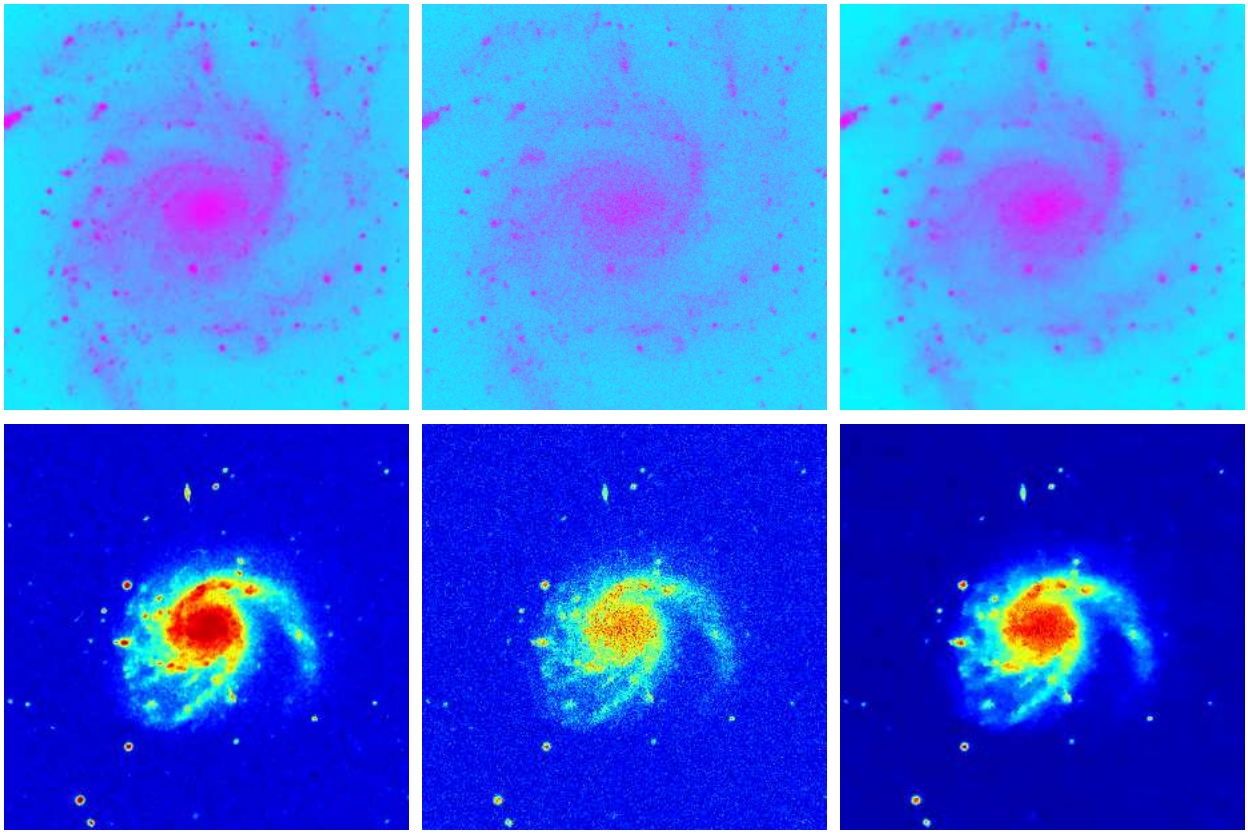
(α) (β) (γ)

(δ) (ε) (στ)

Σχήμα 3.11: Αποθορυβοποίηση αστρονομικών δορυφορικών εικόνων με την PH-HMT μέθοδο. (α) Πηγαία εικόνα του αστρικού συστήματος Eta Carinae, (β) Θορυβώδης εικόνα (PSNR=22.10 dB), (γ) Αποτέλεσμα εκτίμησης (PSNR=30.62 dB). (δ) Πηγαία εικόνα του γαλαξία Whirlpool, (ε) Θορυβώδης εικόνα (PSNR=21.80 dB), (στ) Αποτέλεσμα εκτίμησης (PSNR=32.86 dB).



Σχήμα 3.12: Αποθορυβοποίηση αστρονομικών δορυφορικών εικόνων με την Quad-HMT μέθοδο. (α) Πηγαία εικόνα του γαλαξία Sextans B, (β) Θορυβώδης εικόνα (PSNR=24.92 dB), (γ) Αποτέλεσμα εκτίμησης (PSNR=34.33 dB). (δ) Πηγαία εικόνα του γαλαξία Messier 87, (ε) Θορυβώδης εικόνα (PSNR=22.75 dB), (στ) Αποτέλεσμα εκτίμησης (PSNR=34.98 dB).



Σχήμα 3.13: Αποθορυβοποίηση αστρονομικών δορυφορικών εικόνων με την Sep-HMT μέθοδο. (α) Πηγαία εικόνα του γαλαξία Pinwheel, (β) Θορυβώδης εικόνα (PSNR=21.54 dB), (γ) Αποτέλεσμα εκτίμησης (PSNR=30.39 dB). (δ) Πηγαία εικόνα του γαλαξία Messier 99, (ε) Θορυβώδης εικόνα (PSNR=24.54 dB), (στ) Αποτέλεσμα εκτίμησης (PSNR=34.06 dB).

3.8.2 Δορυφορικές Εικόνες με Πραγματικό Θόρυβο

Στην προηγούμενη ενότητα παρουσιάσαμε αποτελέσματα αποθορυβοποίησης σε δορυφορικά δεδομένα με συνθετικό Poisson θόρυβο. Ωστόσο, για να θεωρείται μια μέθοδος αποδοτική δεν αρκεί μόνο να παρουσιάζει καλή απόδοση σε ελεγχόμενες συνθήκες. Για αυτό το λόγο και προκειμένου να αποτιμήσουμε τις δυνατότητες των προτεινόμενων μεθόδων και σε πραγματικές συνθήκες θορύβου, παρουσιάζουμε κάποια ενδεικτικά αποτελέσματα σε αστρονομικές εικόνες χαμηλής φωτεινότητας, οι οποίες ελήφθησαν από το δορυφόρο XMM-Newton (X-ray Multi-Mirror).

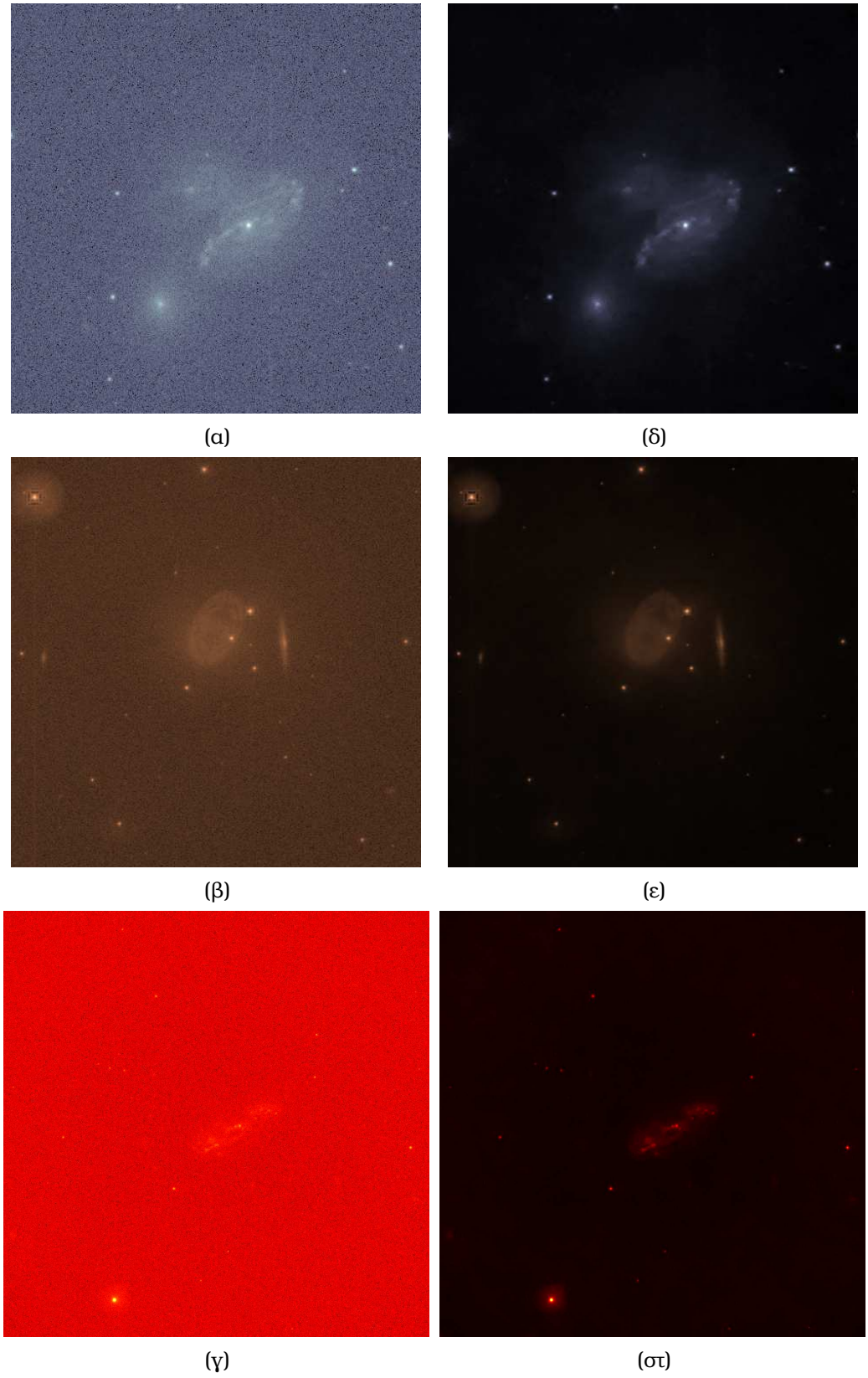
Ο δορυφόρος XMM-Newton ανήκει στην ευρωπαϊκή υπηρεσία διαστήματος (European Space Agency - ESA) και εκτοξεύτηκε το Δεκέμβριο του 1999 στα πλαίσια μιας διαστημική αποστολής η οποία αποτελεί μέρος του ευρύτερου ερευνητικού διαστημικού προγράμματος Horizon 2000. Ο συγκεκριμένος δορυφόρος διαθέτει τρία X-ray τηλεσκόπια υψηλής ευκρίνειας που παρουσιάζουν δυνατότητα κάλυψης πολύ μεγάλων περιοχών του ουρανού. Ο συνδυασμός της μεγάλης περιοχής κάλυψης και της δυνατότητας των τηλεσκοπίων να πραγματοποιούν καταγραφές δεδομένων χωρίς διακοπή για πολύ μεγάλα χρονικά διαστήματα, οδηγεί σε παρατηρήσεις πολύ υψηλής ευαισθησίας. Επειδή η γήινη ατμόσφαιρα εμποδίζει την είσοδο των ακτίνων X ώστε αυτές να καταγραφούν από επίγεια τηλεσκόπια, μόνο ένα τηλεσκόπιο στο διάστημα μπορεί να ανιχνεύσει και να μελετήσει επουράνιες πηγές ακτίνων X. Η αποστολή του δορυφόρου XMM-Newton βοηθάει τους επιστήμονες να μελετήσουν και να προσπαθήσουν να λύσουν ένα σημαντικό αριθμό από κοσμικά μυστήρια, όπως είναι η ύπαρξη μαύρων οπών αλλά και ο χρόνος γένεσης του σύμπαντος.

Στο Σχ. 3.14(α)-(γ) παρουσιάζονται τρεις δορυφορικές αστρονομικές εικόνες χαμηλής φωτεινότητας που εμφανίζουν σημαντικές αλλοιώσεις από θόρυβο. Οι εικόνες αυτές απεικονίζουν κατά σειρά τους γαλαξίες με κωδικές ονομασίες NGC 3227, BPM 16274 και NGC 1808, και τις προμηθευτήκαμε από το Ινστιτούτο Διαστημικής MSSL (Mullard Space Science Laboratory). Στα Σχ. 3.14(δ)-(στ) παρουσιάζονται οι εκτιμήσεις για την πραγματική έντασή τους που προέκυψαν με τη χρήση των μεθόδων Sep-HMT, Quad-HMT και PH-HMT, αντιστοίχως. Συμπερασματικά, μπορούμε να πούμε ότι τόσο από τα παραπάνω αποτελέσματα σε πραγματικά δορυφορικά αστρονομικά δεδομένα όσο και από τα αποτελέσματα που λάβαμε στην Ενοτ. 3.7.3 για τις πραγματικές εικόνες πυρηνικής ιατρικής, οι μέθοδοι που προτείναμε εμφανίζουν ικανοποιητική απόδοση και βελτιώνουν αισθητά την οπτική ποιότητα των εικόνων σε σχέση με τις αρχικές, χωρίς ταυτόχρονα να επηρεάζουν σε σημαντικό βαθμό τα δομικά τους χαρακτηριστικά.

3.8.3 Ανίχνευση σε Δορυφορικές Εικόνες

Στον κλάδο της Αστρονομίας και της Αστροφυσικής, εκτός από το πρόβλημα εκτίμησης της πραγματικής έντασης φωτεινότητας των λαμβανόμενων δορυφορικών εικόνων, ένα εξίσου σημαντικό πρόβλημα αποτελεί και η ανίχνευση των πηγών σε μια εικόνα. Πιο συγκεκριμένα, οι ερευνητές ενδιαφέρονται να ανιχνεύσουν και να καταγράψουν από τα δορυφορικά δεδομένα, το σύνολο των αστρονομικών πηγών που βρίσκονται σε μια συγκεκριμένη περιοχή έρευνας του διαστήματος.

Η ανάγκη για την αναγνώριση και ακριβή ποσοτικοποίηση αντικειμένων ενδιαφέροντος σε αστρονομικές παρατηρήσεις, με τη βοήθεια αυτόματων στατιστικών μεθόδων, γίνεται ολοένα και πιο επιτακτική. Το γεγονός αυτό, οφείλεται, κατά πρώτο λόγο, στο ότι οι σύγχρονες διαστημικές αποστολές καταγράφουν τεράστιες σε έκταση περιοχές του διαστήματος. Συνεπώς, η αναγνώριση όλων των αστρονομικών πηγών είναι αδύνατον να πραγματοποιηθεί αποκλειστικά από ανθρώπους. Σε μια τέτοια διαδικασία είναι εξαιρετικά κρίσιμο να μπορεί κανείς να ρωτά ακριβείς στατιστικές ερωτήσεις σχετικά με τα δεδομένα των καταγραφών. Οι ερωτήσεις αυτού του είδους σχετίζονται με θέματα όπως κατά πόσο υπάρχει αστρονομική πηγή σε μια συγκεκριμένη περιοχή του ουρανού, αν αυτή η πηγή είναι σημειακή ή εκτεταμένη, τι είδους παράμετροι μπορεί να τη χαρακτηρίσουν. Προκειμένου να είναι δυνατή η εξέλιξη της επιστήμης και η απάντηση αυτών των ερωτημάτων απαιτείται η ανάπτυξη μεθόδων που θα έχουν τη δυνατότητα ανίχνευσης των πηγών με ακρίβεια.



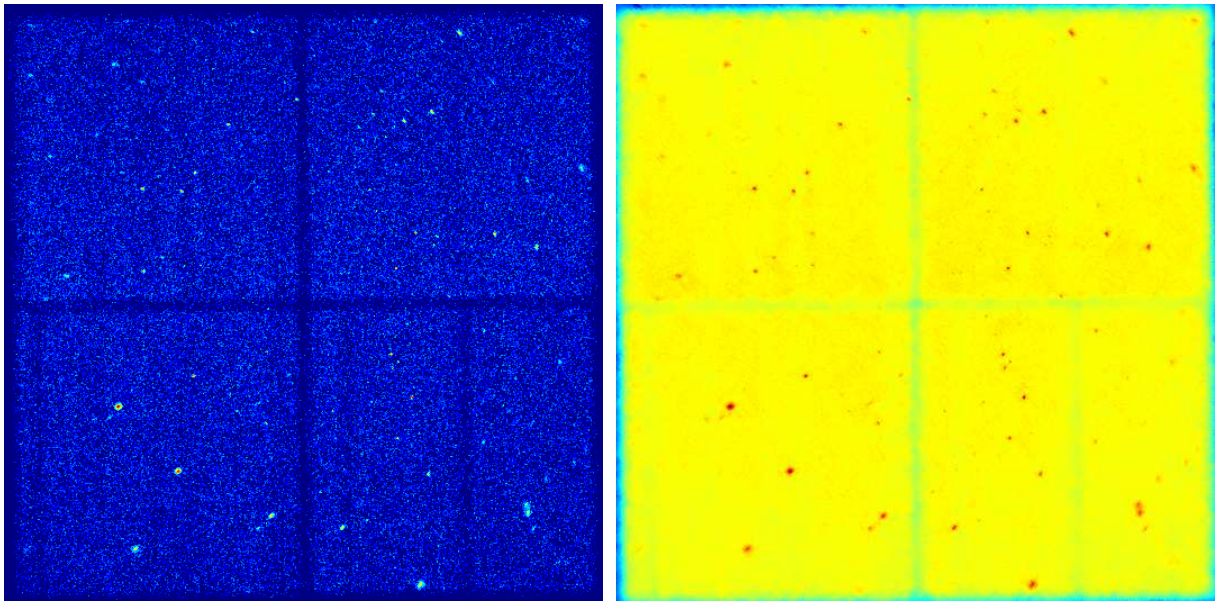
(α) (δ)

(β) (ε)

(γ) (στ)

Σχήμα 3.14: Αποθορυβοποίηση αστρονομικών δορυφορικών εικόνων που καταγράφηκαν από το δορυφόρο XMM-Newton. (α)-(γ) Περιορισμένης φωτεινότητας αστρονομικές εικόνες. (δ)-(στ) Αποτέλεσμα εκτίμησης με τις προτεινόμενες Sep-HMT, Quad-HMT και PH-HMT μεθόδους, αντιστοίχως.

Προκειμένου, να ανιχνεύσει κανείς πηγές από αστρονομικά δεδομένα θα πρέπει πρώτα να αντιμετωπίσει έναν αριθμό από προκλήσεις. Μια από τις σημαντικότερες προκλήσεις είναι και η ύπαρξη του κβαντικού θορύβου στις καταγραφές, ο οποίος μπορεί να μας οδηγήσει αν δεν τον λάβουμε επαρκώς υπόψιν μας, σε λανθασμένα συμπεράσματα για την ύπαρξη ή όχι πηγών αλλά και για το μέγεθος τους. Στο Σχ. 3.15 παρουσιάζουμε ένα παράδειγμα δορυφορικής καταγραφής μιας περιοχής του διαστήματος, με πολλαπλές αστρονομικές πηγές, η οποία ελήφθη από τη διαστημική αποστολή του δορυφόρου Chandra. Στο Σχ. 3.15(α) εμφανίζεται η αρχική καταγραφή ενώ στο Σχ. 3.15(β) η ίδια καταγραφή μετά την επεξεργασία της με την PH-HMT μέθοδο.



(α) (β)

Σχήμα 3.15: (α) Εικόνα με πολλαπλές αστρονομικές πηγές που καταγράφηκε από το δορυφόρο Chandra. (α) Αρχική καταγραφή, (β) Αποτέλεσμα επεξεργασίας των δεδομένων με τη μέθοδο PH-HMT.

Έχοντας εφαρμόσει τις προτεινόμενες μεθόδους εκτίμησης τόσο σε συνθετικά όσο και σε πραγματικά δεδομένα, διαθέτουμε ισχυρές ενδείξεις ότι το στατιστικό μοντέλο που υιοθετήσαμε μπορεί να μοντελοποιήσει με αρκετή επιτυχία τα στατιστικά χαρακτηριστικά αστρονομικών εικόνων, πάνω στις οποίες επιδρά ο κβαντικός θόρυβος. Το γεγονός αυτό, μας δίνει το κίνητρο να ερευνήσουμε κατά πόσο είναι δυνατή η ανάπτυξη μιας στατιστικής μεθόδου για την αυτόματη ανίχνευση αστρονομικών πηγών. Μια ιδέα που εξετάζουμε αλλά δεν έχει ακόμα ολοκληρωθεί, στηρίζεται στην πολυκλιμακωτή αναπαράσταση των εικόνων και στη χρήση ενός HMT μοντέλου δύο κρυφών καταστάσεων. Η πρώτη κατάσταση μοντελοποιεί την ύπαρξη πηγής σε μια συγκεκριμένη κλίμακα ανάλυσης και μια περιοχή της εικόνας, ενώ η δεύτερη κατάσταση μοντελοποιεί την απουσία πηγής. Στη συνέχεια είναι δυνατή η εκτίμηση, με τη χρήση της προτεινόμενης τεχνικής της Ενοτ. 3.6, των αντίστοιχων παραμέτρων του μοντέλου για κάθε κατάσταση. Προκειμένου, να συνάγουμε αν σε μια συγκεκριμένη κλίμακα ανάλυσης j και σε μια περιοχή C υπάρχει ή όχι αστρονομική πηγή, μπορούμε να χρησιμοποιήσουμε κάποιο στατιστικό κριτήριο, όπως για παράδειγμα τον παράγοντα Bayes των δύο καταστάσεων:

$$\mathrm{BF}(\mathbf{x}) = \frac{p\left(z_{jC}=1|\mathbf{x}\right)p\left(z_{jC}=0\right)}{p\left(z_{jC}=0|\mathbf{x}\right)p\left(z_{jC}=1\right)} > 1\,, \tag{3.42}$$

όπου $p(z_{jC}=1)$ και $p(z_{jC}=0)$ είναι οι αρχικές πιθανότητες ύπαρξης ή απουσίας πηγής. Όταν ο λόγος των δύο ποσοτήτων είναι μεγαλύτερος της μονάδας τότε το πιθανότερο είναι ότι στο συγκεκριμένο σημείο υπάρχει αστρονομική πηγή, ενώ σε διαφορετική περίπτωση καταλήγουμε στην απουσία πηγής. Η διαδικασία αυτή μπορεί να γίνει αναδρομικά από την αδρότερη κλίμακα ανάλυσης στην λεπτομερέστερη, οπότε και οδηγούμαστε στο τελικό αποτέλεσμα της ανίχνευσης. Η ιδέα της πολυκλιμακωτής ανάλυσης είναι αρκετά ελπιδοφόρα για την αντιμετώπιση του συγκεκριμένου προβλήματος, δεδομένου ότι στις αδρότερες κλίμακες, όπως ήδη έχει αναφερθεί, η επίδραση του κβαντικού θορύβου είναι σαφώς περιορισμένη. Παράλληλα, η δυνατότητα του HMT μοντέλου να διαδίδει την πληροφορία μεταξύ των κλιμάκων μπορεί να δώσει πολύ καλύτερα αποτελέσματα ανίχνευσης σε σχέση με εναλλακτικές μεθόδους.

3.9 Σύνοψη και Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάσαμε ένα στατιστικό μοντέλο για την ανάλυση Poisson στοχαστικών διαδικασιών σε πολλαπλές κλίμακες και για την εκτίμηση της γενεσιουργού έντασης φωτεινότητάς τους. Το μοντέλο αυτό βρίσκει εφαρμογή στην επεξεργασία εικόνων περιορισμένης φωτεινότητας που παρουσιάζουν αλλοιώσεις από σηματο-εξαρτώμενο Poisson θόρυβο. Για τη βελτίωση αυτού του στατιστικού μοντέλου προτείναμε έναν αποδοτικό και εύρωστο αλγόριθμο εκτίμησης των παραμέτρων του απευθείας από τις θορυβώδεις παρατηρήσεις. Η εύρεση των παραμέτρων αυτών είναι καθοριστικής σημασίας, όπως δείξαμε και με πειραματικά αποτελέσματα. Ακόμη, μελετήσαμε μια ακριβέστερη εκδοχή του στατιστικού μοντέλου η οποία κάνει χρήση αναδρομικών ως προς την κλίμακα μαρκοβιανών δομών, με σκοπό τη μοντελοποίηση των στατιστικών εξαρτήσεων που εμφανίζουν οι συντελεστές διαχωρισμού μεταξύ γειτονικών κλιμάκων ανάλυσης. Επεκτείναμε την προτεινόμενη μέθοδο εκτίμησης των παραμέτρων, ώστε να καλύπτει και τη συγκεκριμένη περίπτωση.

Η απόδοση και η πρακτική σημασία των προτεινόμενων μεθόδων και του πολυκλιμακωτού Poisson-Haar σχήματος αναπαράστασης αποτιμήθηκαν μέσα από συγκρίσεις με εναλλακτικές μεθόδους που ανήκουν σε διαφορετικές κατηγορίες. Οι προτεινόμενες τεχνικές αποδείχθηκαν εξαιρετικά ανταγωνιστικές, ειδικότερα για τις περιπτώσεις εκείνες που η παρουσία του θορύβου ήταν ιδιαίτερα έντονη. Τέλος, προκειμένου να εξετάσουμε κατά πόσο οι τεχνικές που προτείναμε μπορούν να εφαρμοστούν με επιτυχία και σε δεδομένα με πραγματικό θόρυβο, τις χρησιμοποιήσαμε για την αποθορυβοποίηση πραγματικών δορυφορικών αστρονομικών εικόνων και βιοϊατρικών εικόνων. Τα αποτελέσματα που προέκυψαν είναι αρκετά ενθαρρυντικά και αποτελούν μια σαφή ένδειξη ότι το στατιστικό μοντέλο που υιοθετήσαμε καταφέρνει να μοντελοποιεί με επιτυχία τα στατιστικά των φυσικών εικόνων.

3.Α΄ Παραγοντοποίηση της Poisson Συνάρτησης Πιθανοφάνειας σε Τετραδικό Δένδρο

Αρχικά θεωρούμε ότι η εικόνα παρατήρησης $\mathbf{x}$ στην πιο λεπτομερή κλίμακα ανάλυσης έχει μετατραπεί μέσω κατάλληλης αναδιάταξης των εικονοστοιχείων της, σε ένα διάνυσμα μεγέθους $N = N_1 \times N_2$, δηλαδή $\mathbf{x}_0 = \big[x_0(0,0), \ldots, x_0(0, N_2 - 1), \ldots x_0(N_1 - 1, N_2 - 1)\big]$. Επειδή δεδομένων των γενεσιουργών εντάσεων $\boldsymbol{\Lambda}_0(k,\ell)$ όλες οι τυχαίες μεταβλητές $X_0(k,\ell)$ είναι υπό συνθήκη ανεξάρτητες, η συνάρτηση πιθανοφάνειας $p\left(\mathbf{x}|\boldsymbol{\lambda}\right)$ μπορεί να γραφεί ως:

$$p\left(\mathbf{x}|\boldsymbol{\lambda}\right) = \prod_{k=0}^{N_1-1} \prod_{\ell=0}^{N_2-1} p\left(x_0(k,\ell)|\lambda_0(k,\ell)\right) . \tag{3.43}$$

Επιπλέον, επειδή σε κάθε κλίμακα ανάλυσης j οι παραγόμενες παρατηρήσεις $x_j(k,\ell)$ αντιστοιχούν στα δείγματα παρατήρησης υπό συνθήκη ανεξάρτητων Poisson τυχαίων μεταβλητών $X_j(k,\ell)$, θα ισχύει:

$$p\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell), \boldsymbol{\theta}_j(k,\ell)\right) = \frac{\prod\limits_{(k',\ell')\in C_{k,\ell}} p\left(x_{j-1}(k',\ell')|\lambda_{j-1}(k',\ell')\right)}{p\left(x_j(k,\ell)|\lambda_j(k,\ell)\right)} , \tag{3.44}$$

που οδηγεί στην

$$p\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell), \boldsymbol{\theta}_j(k,\ell)\right) p\left(x_j(k,\ell)|\lambda_j(k,\ell)\right) = \prod_{(k',\ell')\in C_{k,\ell}} p\left(x_{j-1}(k',\ell')|\lambda_{j-1}(k',\ell')\right) . \tag{3.45}$$

Η Poisson πιθανοφάνεια, με την βοήθεια των (3.44) και (3.45) μπορεί να ξαναγραφεί ως:

$$\begin{aligned} p\left(\mathbf{x}|\boldsymbol{\lambda}\right) &= \prod_{k=0}^{(N_1/2)-1} \prod_{\ell=0}^{(N_2/2)-1} \prod_{(k',\ell')\in C_{k,\ell}} p\left(x_0(k',\ell')|\lambda_0(k',\ell')\right) \\ &\overset{(3.45)}{=} \prod_{k=0}^{(N_1/2)-1} \prod_{\ell=0}^{(N_2/2)-1} p\left(\mathbf{x}_1^c(k,\ell)|x_1(k,\ell)\right) \cdot \prod_{k=0}^{(N_1/2)-1} \prod_{\ell=0}^{(N_2/2)-1} p\left(x_1(k,\ell)|\lambda_1(k,\ell)\right) . \end{aligned} \tag{3.46}$$

Ακολουθώντας την ίδια διαδικασία επαναληπτικά στο δεύτερο όρο του γινομένου στο δεξί μέλος της εξίσωσης, τελικά λαμβάνουμε:

$$p\left(\mathbf{x}|\boldsymbol{\lambda}\right) = p\left(x_J(0,0)|\lambda_J(0,0)\right) \prod_{j=1}^{J} \prod_{k=0}^{\frac{N_1}{2^j}-1} \prod_{\ell=0}^{\frac{N_2}{2^j}-1} p\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell), \boldsymbol{\theta}_j(k,\ell)\right) , \tag{3.47}$$

με $p\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell), \boldsymbol{\theta}_j(k,\ell)\right) = \text{Mult}\left(\mathbf{x}_j^c(k,\ell)|x_j(k,\ell), \boldsymbol{\theta}_j(k,\ell)\right)$. Αποδείξαμε λοιπόν την παραγοντοποίηση της πιθανοφάνειας σε πολλαπλές κλίμακες.

3.Β΄ Ανάπτυξη της Πολυωνυμικής Κατανομής σε Γινόμενο Δυωνυμικών Κατανομών

Παρατήρηση 1. *Έστω ότι διαθέτουμε ένα σύνολο τεσσάρων Poisson τυχαίων μεταβλητών $\{X_i\}_{i=1\ldots4}$ με παραμέτρους λ_i και $X = \sum_{i=1}^{4} X_i$. Τότε η συνάρτηση της υπό συνθήκης από κοινού πιθανότητας των X_i δοθέντων των λ_i και X θα εκφράζεται ως πολυωνυμική κατανομή. Σε αυτή την περίπτωση μπορούμε να δείξουμε ότι ισχύει:*

$$\begin{aligned} &\textit{Mult}\left(x_1, x_2, x_3, x_4|x, \boldsymbol{\lambda}\right) = \\ &\qquad \textit{Bin}\left(x_1|x_1 + x_2, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right) \textit{Bin}\left(x_1 + x_2|x, \frac{\lambda_1 + \lambda_2}{\lambda}\right) \textit{Bin}\left(x_3|x_3 + x_4, \frac{\lambda_3}{\lambda_3 + \lambda_4}\right) . \end{aligned}$$

Απόδειξη. Από τον ορισμό της πολυωνυμικής κατανομής έχουμε:

$$\text{Mult}\left(x_1, x_2, x_3, x_4 | x, \boldsymbol{\lambda}\right) = \binom{x}{x_1\, x_2\, x_3\, x_4} \prod_{i=1}^{4} \left(\frac{\lambda_i}{\lambda}\right)^{x_i} .$$

Τους δύο όρους γινομένου στο δεξί μέλος της παραπάνω εξίσωσης μπορούμε να τους γράψουμε ισοδύναμα ως:

$$\begin{aligned}\binom{x}{x_1\, x_2\, x_3\, x_4} =& \frac{x!}{x_1! x_2! x_3! x_4!} = \frac{(x_1+x_2)!}{x_1! x_2!} \frac{x!}{(x_1+x_2)!\,(x_3+x_4)!} \frac{(x_3+x_4)!}{x_3! x_4!} \\ =& \binom{x_1+x_2}{x_1} \cdot \binom{x}{x_1+x_2} \cdot \binom{x_3+x_4}{x_3}\end{aligned}$$

και

$$\begin{aligned}\prod_{i=1}^{4} \left(\frac{\lambda_i}{\lambda}\right)^{x_i} =& \left(\frac{\lambda_1}{\lambda_1+\lambda_2} \frac{\lambda_1+\lambda 2}{\lambda}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1+\lambda_2} \frac{\lambda_1+\lambda 2}{\lambda}\right)^{x_2} \\ & \times \left(\frac{\lambda_3}{\lambda_3+\lambda_4} \frac{\lambda_3+\lambda 4}{\lambda}\right)^{x_3} \left(\frac{\lambda_4}{\lambda_3+\lambda_4} \frac{\lambda_3+\lambda 4}{\lambda}\right)^{x_4} \\ =& \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{x_2} \left(\frac{\lambda_1+\lambda_2}{\lambda}\right)^{(x_1+x_2)} \\ & \times \left(\frac{\lambda_3+\lambda_4}{\lambda}\right)^{(x_3+x_4)} \left(\frac{\lambda_3}{\lambda_3+\lambda_4}\right)^{x_3} \left(\frac{\lambda_4}{\lambda_3+\lambda_4}\right)^{x_4}\end{aligned}$$

Αν τώρα χρησιμοποιήσουμε τον ορισμό της κατανομής βήτα από τον Πίνακα 3.1 οδηγούμαστε στο ζητούμενο. □

3.Γ΄ Ισοδυναμία της MMSE Εκτίμησης στο Πεδίο Ορισμού της Εικόνας και στο Μετασχηματισμένο Πεδίο

Σε αυτή την ενότητα θα αποδείξουμε ότι η MMSE εκτίμηση που κανείς λαμβάνει στο πεδίο του πολλαπλασιαστικού μετασχηματισμού για την αναπαραμετροποιημένη εικόνα $\boldsymbol{\lambda}_e$, είναι ισοδύναμη με την MMSE εκτίμηση για την εικόνα $\boldsymbol{\lambda}$ στο αρχικό πεδίο ορισμού της. Η απόδειξη που παραθέτουμε αφορά στον εκτιμητή της μονοδιάστατης περίπτωσης. Ωστόσο, η 2-Δ περίπτωση της απόδειξης μπορεί να ληφθεί με ανάλογο τρόπο.

Αρχικά συμβολίζουμε με $\boldsymbol{\lambda} = \lambda_J(0) \cdot \boldsymbol{\Phi}(\boldsymbol{\theta})$ την ένα-ένα αντιστοίχιση του διανύσματος έντασης της πηγαίας εικόνας $\boldsymbol{\lambda}$ μεγέθους N στο αρχικό πεδίο ορισμού της, με το αναπαραμετροποιημένο διάνυσμα $\boldsymbol{\lambda}_e = [\lambda_J(0), \boldsymbol{\theta}]$ στο πεδίο του πολλαπλασιαστικού μετασχηματισμού. Το $\boldsymbol{\Phi}(\boldsymbol{\theta})$ σε αυτή την περίπτωση ορίζεται ως $\boldsymbol{\Phi}(\boldsymbol{\theta}) = [\Phi_1(\boldsymbol{\theta}) \ \ldots \ \Phi_N(\boldsymbol{\theta})]^T$, όπου

$$\Phi_i(\boldsymbol{\theta}) = \prod_{j=1}^{J} g_{i,j}(\theta_{i,j}) \,. \tag{3.48}$$

Η συνάρτηση $g_{i,j}(\theta_{i,j})$ παρέχει ως έξοδο είτε το συντελεστή $\theta_{i,j}$ αφήνοντας τον έτσι αναλλοίωτο είτε το συμπληρωματικό του $1 - \theta_{i,j}$. Ο συντελεστής $\theta_{i,j}$ αποτελεί πρόγονο του λ_i εικονοστοιχείου της εικόνας $\boldsymbol{\lambda}$ και βρίσκεται στην κλίμακα j. Ουσιαστικά, η εξίσωση $\lambda_J(0) \cdot \Phi_i(\boldsymbol{\theta})$ περιγράφει με μαθηματικό τρόπο τη διαδικασία που πρέπει κανείς να ακολουθήσει προκειμένου να καταλήξει από τον πιο αδρό συντελεστή κλιμάκωσης $\lambda_J(0)$ στο εικονοστοιχείο λ_i (δες την Ενοτ. 2.2.2).

Βασιζόμενοι στην παραπάνω σημειογραφία και την ιδιότητα παραγοντοποίησης της ύστερης πιθανότητας $p(\boldsymbol{\theta}|\mathbf{x})$ στην Εξ. (3.23) μπορούμε να εκφράσουμε τον MMSE εκτιμητή για κάθε συνιστώσα του $\boldsymbol{\lambda}$, λ_i ως

$$\begin{aligned}\hat{\lambda}_i =& \int_{\boldsymbol{\Lambda}} \lambda_i \, p_{\boldsymbol{\Lambda}|\mathbf{X}}(\boldsymbol{\lambda}|\mathbf{x}) \; d\boldsymbol{\lambda} \\ =& \hat{\lambda}_J(0) \int_{\boldsymbol{\Theta}} \Phi_i(\boldsymbol{\theta}) \, p_{\boldsymbol{\Theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x}) \, d\boldsymbol{\theta} \\ =& \hat{\lambda}_J(0) \prod_{j=1}^{J} \int_0^1 g_{i,j}(\theta_{i,j}) p(\theta_{i,j}|\mathbf{x}) \; d\theta_{i,j} \\ =& \hat{\lambda}_J(0) \prod_{j=1}^{J} g_{i,j}(\hat{\theta}_{i,j}) = \hat{\lambda}_J(0) \Phi_i(\hat{\boldsymbol{\theta}}) \, .\end{aligned} \tag{3.49}$$

Αποδείξαμε επομένως την ισοδυναμία του εκτιμητή MMSE για ένα συγκεκριμένο στοιχείο λ_i του $\boldsymbol{\lambda}$, στο πεδίο της εικόνας και στο πεδίο του μετασχηματισμού. Αυτό υπονοεί την γενική ισοδυναμία $\hat{\boldsymbol{\lambda}} = \hat{\lambda}_j(0) \cdot \boldsymbol{\Phi}(\hat{\boldsymbol{\theta}})$.

3.Δ΄ Ποινικοποιημένος Αλγόριθμος Μεγιστοποίησης Προσδοκίας

Για να δείξουμε πως μπορούμε να οδηγηθούμε στην ποινικοποιημένη έκδοση του αλγορίθμου μεγιστοποίησης προσδοκίας (EM) θεωρούμε ότι οι παράγοντες διαχωρισμού $\boldsymbol{\Theta}$ παράγονται από ένα μείγμα κατανομής με Dirichlet συνιστώσες. Η θεώρηση αυτή είναι γενική και καλύπτει και το μείγμα βήτα κατανομών ως ειδική περίπτωση [4].

Η κατανομή Dirichlet ανήκει στην οικογένεια των εκθετικών κατανομών και συνεπώς μπορεί να γραφεί στην μορφή [BS00]:

$$\mathrm{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha}) = f(\boldsymbol{\theta}) \, g(\boldsymbol{\alpha}) \, e^{\phi(\boldsymbol{\alpha})^T u(\boldsymbol{\theta})} \, , \tag{3.50}$$

με $f(\boldsymbol{\theta}) = 1$, $g(\boldsymbol{\alpha}) = 1/B(\boldsymbol{\alpha})$ όπου $B(\boldsymbol{\alpha}) = \dfrac{\prod_{t=1}^{D} \Gamma(\alpha_t)}{\Gamma\left(\sum_{t=1}^{D} \alpha_t\right)}$, $\phi(\boldsymbol{\alpha}) = \begin{pmatrix} \alpha_1 - 1 \\ \vdots \\ \alpha_D - 1 \end{pmatrix}$ και $u(\boldsymbol{\theta}) = \begin{pmatrix} \ln \theta_1 \\ \vdots \\ \ln \theta_D \end{pmatrix}$.

Επειδή για κάθε μέλος της εκθετικής οικογένειας κατανομών υπάρχει μια συζυγής πρότερη πιθανότητα που μπορεί να γραφεί στην μορφή:

$$p(\boldsymbol{\alpha}|\mathbf{v}, \eta) \propto g(\boldsymbol{\alpha})^{\eta} \, e^{\phi(\boldsymbol{\alpha})^T \mathbf{v}} \, , \tag{3.51}$$

μια κατάλληλη πρότερη πιθανότητα για τις παραμέτρους $\boldsymbol{\alpha}$ της Dirichlet κατανομής θα είναι:

$$p(\boldsymbol{\alpha}|\mathbf{v}, \eta) \propto \frac{1}{B(\boldsymbol{\alpha})^{\eta}} e^{-\sum_{t=1}^{D} (\alpha_t v_t)} \, . \tag{3.52}$$

Προσθέτοντας τον προκύπτοντα λογαριθμικό όρο πρότερης πιθανότητας $\log p(\boldsymbol{\alpha}|\mathbf{v}, \eta)$ στην πλήρη συνάρτηση λογαριθμικής πιθανοφάνειας της Εξ. (3.34) οδηγούμαστε στην MAP (ποινικοποιημένος ML) αντί για την τυπική ML λύση [Bis06]. Αυτό αντιστοιχεί στην πρόσθεση στο δεξί μέλος της εξίσωσης Εξ. (3.35) του όρου ποινής

$$-\eta \log B(\boldsymbol{\alpha}) - \sum_{m=1}^{M} \sum_{t=1}^{D} \alpha_m^t v_m^t + (\mathrm{const}) \, . \tag{3.53}$$

[4] Η βήτα κατανομή αποτελεί ειδική περίπτωση της Dirichlet όταν ο αριθμός των τυχαίων μεταβλητών είναι δύο.

Επιλέγοντας $\eta = 0$ και $v_m^t = \epsilon$, καταλήγουμε στον απλοποιημένο όρο ομαλοποίησης $-\epsilon \sum_{m=1}^{M} \sum_{t=1}^{D} \alpha_m^t$ που μας προσφέρει μια εύρωστη λύση στο πρόβλημα βελτιστοποίησης στο Μ-βήμα του ΕΜ αλγορίθμου.

Μέρος ΙΙ

Πολυαισθητηριακή Επεξεργασία Ακουστικών Σημάτων

Κεφάλαιο 4

Βέλτιστη Αποθορυβοποίηση Ακουστικών Σημάτων με τη Χρήση Συστοιχίας Ψηφιακών Αισθητήρων

Τα τελευταία χρόνια, τα συστήματα που βασίζονται σε συστοιχίες αισθητήρων έχουν τύχει μεγάλου ενδιαφέροντος από τη διεθνή επιστημονική κοινότητα. Πρόσφατες ερευνητικές προσπάθειες υποδεικνύουν ότι οι πολυαισθητηριακές τεχνικές είναι πολλά υποσχόμενες στην περιοχή της επεξεργασίας σημάτων. Ιδιαίτερο ενδιαφέρον παρουσιάζει η χρήση αυτών των συστημάτων σε εφαρμογές ασύρματων επικοινωνιών. Το κύριο πλεονέκτημα των πολυαισθητηριακών τεχνικών έναντι των κλασσικών μονοκαναλικών λύσεων (χρήση ενός αισθητήρα) είναι ότι εκμεταλλεύονται πέραν της φασματικής πληροφορίας και τη χωρική πληροφορία (συνήθως ο θόρυβος και τα ακουστικά σήματα προέρχονται από πηγές που βρίσκονται σε διαφορετικά σημεία του χώρου). Ο χωρικός διαχωρισμός επιτυγχάνεται με τη χρήση τεχνικών διαμόρφωσης δέσμης (beamforming) [VB88]. Είναι, λοιπόν, λογικό να υπάρχει η προσδοκία ότι οι συστοιχίες αισθητήρων μπορούν να οδηγήσουν σε καλύτερα αποτελέσματα ιδιαίτερα σε περιβάλλοντα με έντονη παρουσία θορύβου και αντήχησης. Ωστόσο, σε πολλές περιπτώσεις, η απόδοση αυτών των αλγορίθμων δεν είναι επαρκής και απαιτείται η χρήση τεχνικών ύστερου-φιλτραρίσματος στην έξοδο του διαμορφωτή δέσμης για την περαιτέρω ενίσχυση της ποιότητας του σήματος.

Ένα από τα πιο κοινά κριτήρια που χρησιμοποιούνται για την αποκατάσταση αλλοιωμένων από θόρυβο ακουστικών σημάτων είναι το μέσο τετραγωνικό σφάλμα (mean squared error - MSE). Η ελαχιστοποίησή του στην περίπτωση των πολυαισθητηριακών συστημάτων οδηγεί στο πολυκαναλικό Wiener φίλτρο. Αυτό το βέλτιστο, ως προς το MSE, πολυκαναλικό φίλτρο έχει αποδειχθεί ότι μπορεί να αναλυθεί σε έναν διαμορφωτή δέσμης (beamformer) ελάχιστης διακύμανσης και απόκρισης χωρίς παραμόρφωση (minimum variance distortionless response - MVDR) ακολουθούμενο από ένα μονοκαναλικό Wiener ύστερο-φίλτρο [SBM01, Tre02]. Ωστόσο, το MSE δεν αποτελεί το μοναδικό κατάλληλο κριτήριο ελαχιστοποίησης όσον αφορά στην αποθορυβοποίηση ακουστικών σημάτων [EM84, EM85]. Εξίσου κατάλληλα κριτήρια είναι και αυτά που βασίζονται στην ελαχιστοποίηση του MSE του φασματικού πλάτους ή του MSE του λογαρίθμου του φασματικού πλάτους. Η ελαχιστοποίηση αυτών των κριτηρίων οδηγεί στον εκτιμητή του φασματικού πλάτους βραχέος χρόνου (STSA - Short Time Spectral Amplitude) [EM84] και στον εκτιμητή του λογαρίθμου του φασματικού πλάτους βραχέος χρόνου (log-STSA) [EM85], αντιστοίχως. Αυτοί οι εκτιμητές έχουν επίσης αποδειχθεί ότι μπορούν να αναλυθούν σε ένα MVDR διαμορφωτή δέσμης ακολουθούμενο από ένα ύστερο-φίλτρο [BR02, BRB$^+$04]. Γενικώς, η χρήση των ύστερων-φίλτρων επιτυγχάνει μεγαλύτερη απομάκρυνση θορύβου από ό,τι ο MVDR διαμορφωτής μόνος του [BSK98]. Συνεπώς, η ενσωμάτωσή τους δύναται να οδηγήσει σε σημαντική ενίσχυση του σηματοθορυβικού λόγου (SNR).

Παρά τα βέλτιστα θεωρητικά αποτελέσματα τα Wiener, STSA και log-STSA ύστερα-φίλτρα είναι δύσκολο να εφαρμοσθούν στην πράξη. Αυτό οφείλεται στο ότι απαιτείται η εκ των προτέρων γνώση των στατιστικών 2ης τάξης τόσο του σήματος πηγής όσο και του θορύβου που αλλοιώνει το σήμα. Για την αντιμετώπιση αυτού του προβλήματος έχει προταθεί πλήθος λύσεων στη διεθνή βιβλιογραφία [Zel88, FS96, MS97, BSK01, CB02, MB03, Coh04, LA05, LAS06]. Μια αρκετά συνηθισμένη μέθοδος για την εκτίμηση των ποσοτήτων, οι οποίες απαιτούνται για το σχηματισμό των ύστερων-φίλτρων, βασίζεται στη χρήση των αυτο- και ετερο-φασμάτων ισχύος των πολυαισθητηριακών σημάτων εισόδου [SBM01, Zel88, MB03]. Μια από τις πρώτες μεθόδους για την εκτίμηση του ύστερου-φίλτρου αποδίδεται στον Zelinski [Zel88]. Η μέθοδος αυτή μελετήθηκε περαιτέρω από τους Marro et al. [MMS88]. Η γενικευμένη έκδοση του αλγορίθμου του Zelinski βασίζεται στην υπόθεση ότι το πεδίο θορύβου είναι χωρικά ασυσχέτιστο. Μια τέτοια θεώρηση, όμως, δεν είναι καθόλου ρεαλιστική για τις περισσότερες των περιπτώσεων, αφού η συσχέτιση του θορύβου μεταξύ των αισθητήρων μπορεί να είναι σημαντική ιδιαίτερα στις κλίμακες βραχειών συχνοτήτων. Μια πιο ακριβής θεώρηση για το μοντέλο του πεδίου θορύβου προτάθηκε από τους McCowan και Bourlard [MB03] οι οποίοι αντικατέστησαν την προηγούμενη υπόθεση με μια πιο γενική, την υπόθεση της γνώσης της συνάρτησης συνοχής (coherence function) του πεδίου θορύβου. Με αυτόν τον τρόπο επέκτειναν την προηγούμενη μέθοδο και ανέπτυξαν ένα πιο γενικό σχήμα ύστερου φιλτραρίσματος. Ωστόσο, ένα βασικό μειονέκτημα που χαρακτηρίζει και τις δύο μεθόδους είναι ότι το φάσμα ισχύος του θορύβου υπερεκτιμάται [MB03, FK97], με αποτέλεσμα τα προκύπτοντα φίλτρα να είναι υποβέλτιστα. Επιπλέον, οι δύο αυτές μέθοδοι εκτίμησης μπορούν να εφαρμοσθούν μόνο στον υπολογισμό της μορφής του φίλτρου που προκύπτει κατά την ελαχιστοποίηση του κριτηρίου MSE το οποίο, όπως προαναφέρθηκε, οδηγεί στο πολυκαναλικό Wiener φίλτρο.

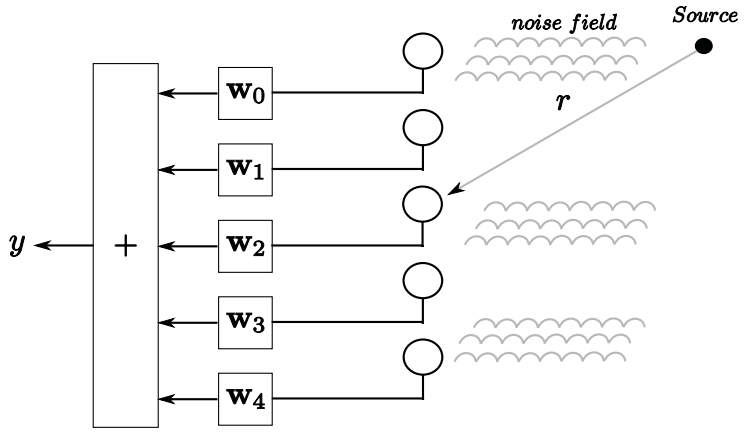
Η έρευνά μας στην περιοχή αυτή εστιάστηκε στη μελέτη και ανάπτυξη νέων αλγορίθμων για την εκτίμηση της μορφής του ύστερου-φίλτρου με απώτερο στόχο τη βέλτιστη αποθορυβοποίηση ακουστικών σημάτων. Η μέθοδος που αναπτύξαμε βασίζεται σε στατιστικές μεθόδους επεξεργασίας και μοντελοποίησης σημάτων. Είναι εύρωστη απέναντι στο θόρυβο, γενική, καθώς μπορεί να χρησιμοποιηθεί σε διαφορετικά περιβάλλοντα θορύβου και, ακόμη, μπορεί να εφαρμοσθεί στην εκτίμηση της μορφής γραμμικών και μη-γραμμικών ύστερων-φίλτρων αναλόγως με το επιλεγόμενο κριτήριο ελαχιστοποίησης.

Στη συνέχεια του κεφαλαίου και συγκεκριμένα στην Ενοτ. 4.1 ορίζουμε το πρόβλημα της πολυαισθητηριακής αποθορυβοποίησης ακουστικών σημάτων και παρέχουμε κάποιες βασικές εισαγωγικές έννοιες της πολυαισθητηριακής επεξεργασίας. Στην Ενοτ. 4.2 παρουσιάζουμε τις υπάρχουσες τεχνικές χωρικού φιλτραρίσματος, ενώ στην Ενοτ. 4.3 εξετάζουμε τα στατιστικά κριτήρια που οδηγούν σε βέλτιστες τεχνικές αποθορυβοποίησης. Επίσης, στην Ενοτ. 4.4 παρουσιάζουμε το προτεινόμενο σχήμα εκτίμησης των παραμέτρων που απαιτούνται για την υλοποίηση των βέλτιστων φίλτρων. Τέλος, στην Ενοτ. 4.5 παραθέτουμε συγκριτικά πειραματικά αποτελέσματα της προτεινόμενης μεθόδου με άλλες εναλλακτικές μεθόδους αποθορυβοποίησης, ενώ στην Ενοτ. 4.6 παρέχεται μια σύνοψη του κεφαλαίου αλλά και τα συμπεράσματα που προέκυψαν κατά την παρούσα μελέτη.

4.1 Καθορισμός Προβλήματος

Θεωρούμε μια συστοιχία με N αισθητήρες διατεταγμένους σε γραμμική διάταξη η οποία είναι τοποθετημένη σε ένα περιβάλλον με έντονη παρουσία θορύβου. Έστω ότι το επιθυμητό σήμα πηγής βρίσκεται σε απόσταση r και υπό κλίση γωνίας θ από το κέντρο της συστοιχίας, όπως φαίνεται στο Σχ. 4.1. Το παρατηρούμενο σήμα, $x_i(n)$, $i = 0, \ldots, N-1$, στον iοστό αισθητήρα αντιστοιχεί σε μια γραμμικά φιλτραρισμένη έκδοση του σήματος πηγής $s(n)$ με την επιπλέον προσθήκη μιας συνιστώσας θορύβου $v_i(n)$:

$$x_i(n) = d_i(n; \theta, r) * s(n) + v_i(n)\,, \tag{4.1}$$



Σχήμα 4.1: Συστοιχία 5 αισθητήρων για τη σύλληψη ακουστικών σημάτων.

όπου $d_i(n;\theta,r)$ είναι η κρουστική απόκριση του ακουστικού μονοπατιού από την επιθυμητή πηγή προς τον iοστό αισθητήρα, ενώ το σύμβολο $*$ δηλώνει την πράξη της συνέλιξης. Η κρουστική απόκριση κάθε ακουστικού μονοπατιού σε περιβάλλον χωρίς αντήχηση, που είναι και η περίπτωση που εξετάζουμε, μπορεί να μοντελοποιηθεί ως μια εξασθενημένη και καθυστερημένη Kronecker δέλτα συνάρτηση $d_i(n;\theta,r) = \alpha_i(\theta,r)\delta(n-\tau_i(\theta,r))$, όπου α_i είναι ο παράγοντας εξασθένισης και τ_i είναι η χρονική καθυστέρηση εκφρασμένη σε χρονικά δείγματα. Αυτή η καθυστέρηση αναπαριστά τον επιπρόσθετο χρόνο που χρειάζεται το σήμα πηγής για να ταξιδέψει προς τον iοστό αισθητήρα, αφού προηγουμένως έχει φτάσει στο κέντρο της συστοιχίας. Σε αυτή την περίπτωση η Εξ. (4.1) παίρνει τη μορφή

$$x_i(n) = a_i(\theta,r)s(n-\tau_i(\theta,r)) + v_i(n)\,, \tag{4.2}$$

η οποία σε αναπαράσταση πινάκων μπορεί να γραφεί ως

$$\begin{pmatrix} x_0(n) \\ x_1(n) \\ \vdots \\ x_{N-1}(n) \end{pmatrix} = \begin{pmatrix} \alpha_0(\theta,r)s(n-\tau_0(\theta,r)) \\ \alpha_1(\theta,r)s(n-\tau_1(\theta,r)) \\ \vdots \\ \alpha_{N-1}(\theta,r)s(n-\tau_{N-1}(\theta,r)) \end{pmatrix} + \begin{pmatrix} v_0(n) \\ v_1(n) \\ \vdots \\ v_{N-1}(n) \end{pmatrix}$$

$$\mathbf{x}(n) = \boldsymbol{\alpha}\mathbf{s}(n-\boldsymbol{\tau}(\theta,r)) + \mathbf{v}(n)\,. \tag{4.3}$$

Η ανάλυση που πρόκειται να ακολουθήσουμε μπορεί να γίνει αποδοτικότερα, αν μεταβούμε στο πεδίο των συχνοτήτων. Ωστόσο, επειδή τόσο τα ακουστικά σήματα όσο και οι συνιστώσες θορύβου είναι μη-στάσιμες διαδικασίες [RS78], διαθέτουν δηλαδή χρονικά μεταβαλλόμενα χαρακτηριστικά, είναι απαραίτητο να ακολουθηθεί μια βραχέος χρόνου ανάλυση. Σύμφωνα με αυτή την ανάλυση τα παρατηρούμενα σήματα διαιρούνται στο χρόνο σε επικαλυπτόμενα πλαίσια και σε κάθε πλαίσιο εφαρμόζεται μια συνάρτηση παραθύρου. Στη συνέχεια, κάθε πλαίσιο αναλύεται με τη βοήθεια του βραχύχρονου μετασχηματισμού Fourier (STFT - Short Time Fourier Transform) [AR77, All77]. Υποθέτοντας ότι η απόκριση συχνότητας είναι χρονικά αμετάβλητη, μπορούμε να εκφράσουμε τις παρατηρηθείσες πληροφορίες στο κοινό χρονο-συχνοτικό πεδίο ως:

$$\begin{pmatrix} X_0(k,\ell) \\ X_1(k,\ell) \\ \vdots \\ X_{N-1}(k,\ell) \end{pmatrix} = \begin{pmatrix} \alpha_0(\theta,r)e^{-j\omega_k\tau_0(\theta,r)} \\ \alpha_1(\theta,r)e^{-j\omega_k\tau_1(\theta,r)} \\ \vdots \\ \alpha_{N-1}(\theta,r)e^{-j\omega_k\tau_{N-1}(\theta,r)} \end{pmatrix} S(k,\ell) + \begin{pmatrix} V_0(k,\ell) \\ V_1(k,\ell) \\ \vdots \\ V_{N-1}(k,\ell) \end{pmatrix}$$

$$\mathbf{X}(k,\ell) = \mathbf{D}(k;\theta,r)S(k,\ell) + \mathbf{V}(k,\ell)\,, \tag{4.4}$$

όπου τα k και ℓ είναι δείκτες που αναφέρονται στα συχνοτικά και χρονικά πλαίσια, αντιστοίχως, ενώ το $\mathbf{D}(k;\theta,r) = [D_0(k;\theta,r)\ D_1(k;\theta,r)\ \ldots D_{N-1}(k;\theta,r)]^T$ καλείται διάνυσμα οδήγησης της συστοιχίας [Tre02] και είναι ένα μιγαδικό διάνυσμα που εμπεριέχει όλα τα χωρικά χαρακτηριστικά αυτής. Για την περίπτωση περιβάλλοντος χωρίς αντήχηση το iοστό στοιχείο του διανύσματος οδήγησης μπορεί να γραφεί ως $D_i(k;\theta,r) = \alpha_i(\theta,r)e^{-j\omega_k\tau_i(\theta,r)}$ [DM03], με ω_k τη διακριτού χρόνου γωνιακή συχνότητα που αντιστοιχεί στο kοστό συχνοτικό πλαίσιο.

Χρησιμοποιώντας το παραπάνω μοντέλο στόχος μας είναι να εκτιμήσουμε το πηγαίο σήμα $s(n)$ με ένα βέλτιστο τρόπο έχοντας στη διάθεσή μας μόνο τις θορυβώδεις παρατηρήσεις στις εξόδους των αισθητήρων. Προτού όμως οδηγηθούμε στην περιγραφή βέλτιστων τεχνικών για την εκτίμηση των ακουστικών σημάτων ενδιαφέροντος, θα πρέπει πρώτα, προκειμένου να έχουμε μια καλύτερη οπτική του προβλήματος και των εργαλείων που θα χρησιμοποιήσουμε, να αναφερθούμε σε κάποια βασικά θέματα της πολυαισθητηριακής επεξεργασίας αλλά και των συστοιχιών αισθητήρων.

4.1.1 Χωρική Δειγματοληψία

Στο πρόβλημα που καλούμαστε να αντιμετωπίσουμε για τη σύλληψη των ακουστικών σημάτων κάνουμε χρήση μιας διακριτής συστοιχίας από αισθητήρες. Αυτή η συστοιχία αισθητήρων μπορεί να θεωρηθεί ότι προκύπτει δειγματοληπτώντας μια συνεχή συστοιχία (aperture) στο χώρο. Όπως θα περίμενε κανείς με βάση τη θεωρία ψηφιακής επεξεργασίας σημάτων, η δειγματοληψία αυτή μπορεί να προκαλέσει κάποια ανεπιθύμητα φαινόμενα, όπως είναι η αναδίπλωση συχνοτήτων (aliasing). Στη χρονική δειγματοληψία είναι ευρέως γνωστό το θεώρημα του Nyquist το οποίο ορίζει την ελάχιστη απαιτούμενη συχνότητα δειγματοληψίας, προκειμένου να αποφύγουμε την αναδίπλωση συχνοτήτων στο δειγματοληπτημένο σήμα. Μια ανάλογη απαίτηση υπάρχει και για τη χωρική δειγματοληψία, αφού η συστοιχία αποτελείται από μεμονωμένους αισθητήρες οι οποίοι δειγματοληπτούν χωρικά το περιβάλλον στο οποίο τοποθετούνται. Ένα από τα προβλήματα που προκύπτουν από την ανεπαρκή χωρική δειγματοληψία είναι η εμφάνιση ψευδών κύριων λοβών στη συνάρτηση μεταφοράς της συστοιχίας [Zio95]. Το αποτέλεσμα είναι η συστοιχία των αισθητήρων να παρουσιάζει αδυναμία στο να ξεχωρίσει σήματα τα οποία διαδίδονται από κατευθύνσεις, που αντιστοιχούν σε χωρικές συχνότητες στις οποίες εμφανίζονται οι ψευδείς λοβοί, από σήματα που διαδίδονται από τις χωρικές συχνότητες που βρίσκονται στη ζώνη μετάβασης της συστοιχίας. Σύμφωνα με το θεώρημα του Nyquist, το οποίο αναφέρεται στη χρονική δειγματοληψία, ένα σήμα πρέπει να δειγματοληπτείται με συχνότητα δειγματοληψίας f_s (περίοδος δειγματοληψίας T_s) τέτοια, ώστε να ισχύει η παρακάτω σχέση:

$$f_s = \frac{1}{T_s} \geq 2f_{max}\,, \tag{4.5}$$

όπου f_{max} αντιστοιχεί στη μέγιστη συχνοτική συνιστώσα που περιέχεται στο φάσμα συχνοτήτων του σήματος. Ένα ανάλογο θεώρημα ισχύει και για τη χωρική δειγματοληψία [Zio95]. Σύμφωνα με το θεώρημα αυτό, προκειμένου η αναδίπλωση των χωρικών συχνοτήτων να μην έχει καμία επίδραση στα λαμβανόμενα σήματα, θα πρέπει η απόσταση, d, μεταξύ γειτονικών αισθητήρων της συστοιχίας να ικανοποιεί τη σχέση $d \leq \frac{\lambda_{min}}{2}$, όπου λ_{min} είναι το ελάχιστο μήκος κύματος του σήματος ενδιαφέροντος.

4.1.2 Πηγές Κοντινού και Μακρινού Πεδίου

Οι πηγές των ακουστικών σημάτων και οι πηγές θορύβου μπορούν να βρίσκονται είτε κοντά στη συστοιχία των αισθητήρων είτε μακριά από αυτήν. Η πρώτη περίπτωση ονομάζεται κατάσταση *κοντινού-πεδίου* (near-field), ενώ η δεύτερη κατάσταση *μακρινού-πεδίου* (far-field). Στην περίπτωση, όπου έχουμε πηγές μακρινού-πεδίου, μπορούμε να υποθέσουμε ότι έχουμε μετάδοση επίπεδων κυμάτων και ότι τα σήματα που λαμβάνονται από όλους τους αισθητήρες της συστοιχίας παρουσιάζουν την ίδια εξασθένιση, $\alpha_i(\theta, r) = 1$. Αντίθετα, στην περίπτωση κοντινού-πεδίου πηγών θα πρέπει να λάβουμε υπόψιν μας τόσο τη σφαιρική μετάδοση των κυμάτων όσο και τη διαφορετική εξασθένιση. Ένας τυπικός κανόνας είναι ότι η υπόθεση μακρινού-πεδίου παύει να είναι πλέον έγκυρη [Mai94], όταν

$$r < \frac{d_{tot}^2 f_s}{c}\,, \tag{4.6}$$

όπου d_{tot} είναι το συνολικό μήκος της συστοιχίας των αισθητήρων, f_s η συχνότητα δειγματοληψίας και c είναι η ταχύτητα διάδοσης του ήχου ($c \approx 340\frac{m}{s}$).

Στην περίπτωση πηγών μακρινού πεδίου είναι δυνατόν να υπολογίσουμε τη χρονική καθυστέρηση σε κάθε αισθητήρα της συστοιχίας κάνοντας χρήση απλών γεωμετρικών ιδιοτήτων, ως

$$\tau_i(\theta, r) = \frac{d_i \cos\theta}{c} f_s\,. \tag{4.7}$$

Στην αντίθετη περίπτωση, κατά την οποία η πηγή του ακουστικού σήματος βρίσκεται σε απόσταση r που την τοποθετεί στο κοντινό πεδίο (4.6), αποδεικνύεται ότι η απόσταση του πηγαίου σήματος από τον iοστό αισθητήρα της συστοιχίας είναι ίση με:

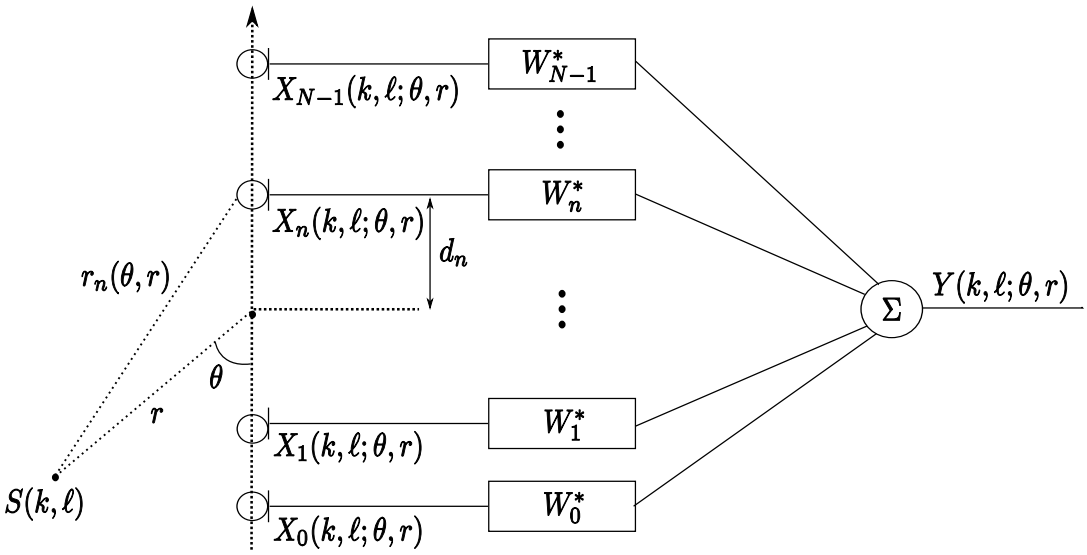
$$r_i(\theta, r) = \left(r^2 + d_i^2 + 2 r d_i \cos\theta\right)^{\frac{1}{2}}\,, \tag{4.8}$$

όπου d_i είναι η απόσταση του iοστού αισθητήρα από το κέντρο της συστοιχίας (δες Σχ.4.2). Τότε, η χρονική καθυστέρηση $\tau_n(\theta, r)$ και ο παράγοντας εξασθένισης του σήματος $a_n(\theta, r)$ δίνονται από τις παρακάτω σχέσεις:

$$\begin{aligned} \tau_i(\theta, r) =& \frac{r_i(\theta, r) - r}{c} f_s \\ a_i(\theta, r) =& \frac{r}{r_i(\theta, r)}\,. \end{aligned} \tag{4.9}$$

4.2 Βασικές Τεχνικές Χωρικού Φιλτραρίσματος

Μια από τις πιο απλές και ταυτόχρονα εύρωστες μεθόδους χωρικού φιλτραρίσματος είναι το φιλτράρισμα μέσω *διαμόρφωσης δέσμης* (*beamforming*). Ο όρος beamforming αναφέρεται σε μια μεγάλη κατηγορία αλγορίθμων πολυαισθητηριακής επεξεργασίας στην οποία καθορίζοντας τα βάρη των αισθητήρων, $W(k, \ell)$, (βλέπε Σχ. 4.2) έχουμε τη δυνατότητα να διαμορφώσουμε και να κατευθύνουμε τη δέσμη της συστοιχίας με τέτοιο τρόπο, ώστε να εστιάσουμε τις δυνατότητες σύλληψής της σε μια συγκεκριμένη κατεύθυνση και να απορρίψουμε σήματα που μεταδίδονται από διαφορετικές κατευθύνσεις. Συνεπώς, η τεχνική του φιλτραρίσματος διαμόρφωσης δέσμης είναι ανάλογη με το ζωνοπερατό φιλτράρισμα. Μια τέτοια προσέγγιση εμφανίζει αρκετά πλεονεκτήματα στην επεξεργασία και αποθορυβοποίηση ακουστικών σημάτων. Αυτό συμβαίνει, γιατί σε ένα τυπικό περιβάλλον η πηγή του ακουστικού σήματος συνήθως βρίσκεται σε διαφορετική χωρική τοποθεσία από τις παρεμβαλλόμενες πηγές θορύβου. Επομένως, εκμεταλλευόμενες τη χωρική διάσταση του προβλήματος, οι συστοιχίες αισθητήρων προσπαθούν να εξασφαλίσουν ένα υψηλής



Σχήμα 4.2: Φιλτράρισμα διαμόρφωσης δέσμης στο κοινό χρονο-συχνοτικό πεδίο.

ποιότητας ακουστικό σήμα χωρίς να είναι απαραίτητο η πηγή του σήματος να βρίσκεται κοντά στον αισθητήρα.

Η γενική μορφή του σήματος εξόδου από ένα τέτοιο σύστημα διαμόρφωσης δέσμης δίνεται από τη σχέση [JD93]

$$Y(k,\ell) = \sum_{i=0}^{N-1} X_i(k,\ell) W_i^*(k,\ell) = \mathbf{W}^H(k,\ell)\mathbf{X}(k,\ell)\,, \tag{4.10}$$

όπου $\mathbf{W}^H(k,\ell) = \left[W_0^*(k,\ell) W_1^*(k,\ell) \ldots W_{N-1}^*(k,\ell)\right]$ είναι τα βάρη των αισθητήρων και τα $(\cdot)^*$, $(\cdot)^H$ συμβολίζουν, αντιστοίχως, το μιγαδικό συζυγή και την ερμιτιανή συμμετρία. Συνδυάζοντας τις Εξ. (4.4) και (4.10) μπορούμε να εκφράσουμε το σήμα εξόδου του συστήματος διαμόρφωσης δέσμης ως

$$Y(k,\ell) = \underbrace{\mathbf{W}^H(k,\ell)\mathbf{D}(k;\theta,r)S(k,\ell)}_{Y_s(k,\ell)} + \underbrace{\mathbf{W}^H(k,\ell)\mathbf{V}(k,\ell)}_{Y_v(k,\ell)}\,, \tag{4.11}$$

όπου $Y_s(k,\ell)$ είναι η συνιστώσα του ακουστικού σήματος πηγής και $Y_v(k,\ell)$ είναι η συνιστώσα του θορύβου.

4.2.1 Κέρδος Συστοιχίας

Μια συνηθισμένη ποσότητα που χρησιμοποείται για την αξιολόγηση της απόδοσης των συστημάτων διαμόρφωσης δέσμης είναι το *κέρδος συστοιχίας* το οποίο δηλώνει τη βελτίωση που επιτυγχάνεται σε όρους σηματοθορυβικού λόγου (SNR), αν χρησιμοποιήσουμε μια συστοιχία αισθητήρων [CZO87, CZK86]. Μαθηματικά, μπορούμε να εκφράσουμε το κέρδος συστοιχίας ως

$$G = \frac{\text{SNR}_{out}}{\text{SNR}_{in}}\,, \tag{4.12}$$

όπου SNR_{out} είναι ο σηματοθορυβικός λόγος στην έξοδο της συστοιχίας και SNR_{in} είναι η μέση τιμή του σηματοθορυβικού λόγου των σημάτων εισόδου [1]. Ο σηματοθορυβικός λόγος στην έξοδο

[1] Για λόγους απλότητας και χωρίς απώλεια της γενικότητας θα παραλείπονται από εδώ και στο εξής οι δείκτες k και ℓ.

της συστοιχίας θα ισούται με $\text{SNR}_{out} = \frac{\Phi_{ys}}{\Phi_{yv}}$, όπου τα Φ_{y_s} και Φ_{y_v} αντιστοιχούν στο φάσμα ισχύος των σημάτων Y_s και Y_v. Το φάσμα ισχύος του σήματος Y_s μπορεί να υπολογιστεί ως

$$\Phi_{y_s} = E\left\{Y_s Y_s^*\right\} = \Phi_s \left|\mathbf{W}^H \mathbf{D}\right|^2 , \tag{4.13}$$

όπου Φ_s είναι το φάσμα ισχύος του πηγαίου σήματος. Ομοίως το φάσμα ισχύος του Y_v δίνεται από τη σχέση

$$\Phi_{y_v} = E\left\{Y_v Y_v^*\right\} = \mathbf{W}^H \mathbf{\Phi}_v \mathbf{W} , \tag{4.14}$$

όπου

$$\mathbf{\Phi}_v = \begin{pmatrix} \Phi_{v_0} & \Phi_{v_0 v_1} & \cdots & \Phi_{v_0 v_{N-1}} \\ \Phi_{v_1 v_0} & \Phi_{v_1} & & \\ \vdots & & \ddots & \\ \Phi_{v_{N-1} v_0} & \cdots & & \Phi_{v_{N-1}} \end{pmatrix} \tag{4.15}$$

είναι ο πίνακας ετεροφάσματος ισχύος των συνιστωσών του θορύβου. Αν ορίσουμε το παράγοντα κανονικοποίησης f [2] ως

$$f = \frac{trace(\mathbf{\Phi}_v)}{N} = \frac{\sum_{i=0}^{N-1} \Phi_{v_i}}{N} , \tag{4.16}$$

τότε ο πίνακας $\mathbf{\Phi}_v$ μπορεί να γραφεί ως

$$\mathbf{\Phi}_v = f \cdot \bar{\mathbf{\Phi}}_v \tag{4.17}$$

όπου $\bar{\mathbf{\Phi}}_v$ είναι ο κανονικοποιημένος πίνακας ετεροφάσματος ισχύος του θορύβου. Τελικά, χρησιμοποιώντας τις Εξ.(4.14) και (4.17) το φάσμα ισχύος του σήματος Y_v ισούται με

$$\Phi_{y_v} = f \cdot \mathbf{W}^H \bar{\mathbf{\Phi}}_v \mathbf{W} \tag{4.18}$$

Από τις Εξ. (4.13) και (4.18) το SNR στην έξοδο της συστοιχίας ισούται με

$$SNR_{out} = \frac{\Phi_s \left|\mathbf{W}^H \mathbf{D}\right|^2}{f \cdot \mathbf{W}^H \bar{\mathbf{\Phi}}_v \mathbf{W}} . \tag{4.19}$$

Αντιστοίχως, η μέση τιμή για το SNR εισόδου μπορεί να υπολογιστεί ως:

$$SNR_{in} = \frac{E\left\{(\mathbf{D}S)^H S\mathbf{D}\right\}}{E\left\{\mathbf{V}^H \mathbf{V}\right\}} = \frac{\Phi_s \mathbf{D}^H \mathbf{D}}{\sum_{i=1}^{N-1} \Phi_{v_i}} = \frac{\Phi_s \sum_{i=0}^{N-1} |\alpha_i(\theta, r)|^2}{N \cdot f} . \tag{4.20}$$

Τελικά το κέρδος της συστοιχίας θα έχει την παρακάτω μορφή:

$$G = \underbrace{\frac{N}{\sum_{i=0}^{N-1} |\alpha_i(\theta, r)|^2}}_{const.} \cdot \frac{\left|\mathbf{W}^H \mathbf{D}\right|^2}{\mathbf{W}^H \mathbf{\Phi}_v \mathbf{W}} , \tag{4.21}$$

όπου στα βάρη έχουμε ενσωματώσει τον όρο κανονικοποίησης. Για την περίπτωση του μακρινού-πεδίου πηγής θα ισχύει $\alpha_i(\theta, r) = 1 \, \forall i$, επομένως η σταθερά της Εξ. (4.21) θα ισούται με μονάδα.

[2] Το f επιβάλλει το ίχνος του πίνακα να ισούται με N.

4.2.2 Υπερκατευθυντικό Φιλτράρισμα Διαμόρφωσης Δέσμης

Μια από τις πιο γνωστές και συχνά χρησιμοποιούμενες τεχνικές χωρικού φιλτραρίσματος είναι το υπερκατευθυντικό φιλτράρισμα διαμόρφωσης δέσμης. Ο όρος 'υπέρ' σε έναν υπερκατευθυντικό διαμορφωτή δέσμης (beamformer) δηλώνει ότι ο συγκεκριμένος διαμορφωτής μεγιστοποιεί το κέρδος της συστοιχίας $G(k, \ell; \theta, r)$ προς την κατεύθυνση της πηγής του ακουστικού σήματος για ένα γνωστό πεδίο θορύβου. Αυτός ο διαμορφωτής έχει τη δυνατότητα να καταστέλλει το θόρυβο που έρχεται από όλες τις άλλες κατευθύνσεις και, παράλληλα, να αφήνει ανεπηρέαστο το επιθυμητό σήμα που διαδίδεται από μια κύρια κατεύθυνση.

Προκειμένου να βρούμε τα βέλτιστα βάρη του υπερκατευθυντικού διαμορφωτή, θα πρέπει να βρούμε εκείνα τα $\mathbf{W}$ για τα οποία το κέρδος της συστοιχίας (4.21) μεγιστοποιείται. Αυτό είναι ισοδύναμο με το να ελασχιστοποιήσουμε το φάσμα ισχύος της συνιστώσας θορύβου, Y_v, στην έξοδο του συστήματος διαμόρφωσης (4.14). Η ελαχιστοποίηση αυτή, προκειμένου να έχει νόημα, θα πρέπει να συμπληρωθεί από το κριτήριο μη-παραμόρφωσης, $\mathbf{W}^H\mathbf{D} = 1$, το οποίο απαιτεί στην περίπτωση απουσίας θορύβου το σήμα εισόδου να παραμένει αναλλοίωτο στην έξοδο του διαμορφωτή. Έτσι, τα βέλτιστα βάρη μπορούν να βρεθούν μέσω του παρακάτω προβλήματος ελαχιστοποίησης

$$\min_{\mathbf{W}} \mathbf{W}^H\mathbf{\Phi}_v\mathbf{W} \quad \text{s.t.} \quad \mathbf{W}^H\mathbf{D}(\theta, r) = 1\,. \tag{4.22}$$

Κάνοντας χρήση της μεθόδου βελτιστοποίησης Lagrange [DHS01], καταλήγουμε στα βέλτιστα βάρη, $\mathbf{W}$, που δίνονται από τη σχέση

$$\mathbf{W}^H = \frac{\mathbf{D}^H\mathbf{\Phi}_v^{-1}}{\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}}\,, \tag{4.23}$$

και μας οδηγούν στον υπερκατευθυντικό διαμορφωτή δέσμης [CZO87, BS01] που είναι γνωστός και ως διαμορφωτής ελάχιστης διακύμανσης και απόκρισης χωρίς παραμόρφωση (Minimum Variance Distortionless Response - MVDR). Όπως θα διαπιστώσουμε στη συνέχεια, ο MVDR διαμορφωτής παίζει πολύ σημαντικό ρόλο στην πολυαισθητηριακή επεξεργασία σημάτων και εμφανίζεται συχνά ως ένα πρώτο υποσύστημα φιλτραρίσματος.

4.3 Βέλτιστη Αποθορυβοποίηση Ακουστικών Σημάτων

Παρά την καλή θεωρητικά συμπεριφορά του υπερκατευθυντικού διαμορφωτή δέσμης, στην πράξη σε πολλές εφαρμογές η απόδοσή του δεν είναι επαρκής. Για αυτό το λόγο, συχνά υιοθετούνται εναλλακτικά κριτήρια τα οποία μπορούν να μας οδηγήσουν σε συστήματα που παρουσιάζουν μια στατιστικά βέλτιστη απόδοση. Στην παρούσα ενότητα, εστιάζουμε στη χρήση τριών διαφορετικών κριτηρίων βελτιστοποίησης για την αποθορυβοποίηση ακουστικών σημάτων. Τα κριτήρια αυτά έχει αποδειχθεί ότι οδηγούν σε εκτιμητές που μπορούν να αναλυθούν σε ένα MVDR διαμορφωτή δέσμης ακολουθούμενο από ένα μονοκαναλικό ύστερο-φίλτρο. Οι υπό εξέταση εκτιμητές είναι οι εξής: ο εκτιμητής μέσου τετραγωνικού σφάλματος MMSE, ο εκτιμητής μέσου τετραγωνικού σφάλματος του φασματικού πλάτους βραχέος χρόνου (MMSE STSA) και ο εκτιμητής μέσου τετραγωνικού σφάλματος του λογαρίθμου του φασματικού πλάτους βραχέος χρόνου (MMSE log-STSA).

Για την υλοποίηση των παραπάνω εκτιμητών καθίσταται απαραίτητο να γίνεται γνωστή η εκ των προτέρων συνάρτηση πυκνότητας πιθανότητας των συντελεστών Fourier τόσο του πηγαίου σήματος όσο και του θορύβου. Επειδή στην πράξη κάτι τέτοιο δε συμβαίνει, και επιπλέον, η μέτρηση της συνάρτησης πυκνότητας πιθανότητας είναι μια πολύπλοκη και επίπονη διαδικασία, υιοθετούνται οι παρακάτω υποθέσεις [EM84] υποκινούμενες από το κεντρικό οριακό θεώρημα:

1. Το πηγαίο σήμα είναι μια γκαουσιανή στοχαστική διαδικασία με μηδενική μέση τιμή και φάσμα ισχύος Φ_s.

2. Τα λαμβανόμενα από τους αισθητήρες σήματα, τα οποία αποτιμώνται ως θόρυβος, είναι γκαουσιανές στοχαστικές διαδικασίες με μηδενική μέση τιμή και πίνακα ετερο-φάσματος ισχύος $\mathbf{\Phi}_v$.

3. Το πηγαίο σήμα είναι ασυσχέτιστο με τα σήματα θορύβου και οι συντελεστές Fourier κάθε στοχαστικής διαδικασίας είναι ανεξάρτητοι σε διαφορετικές συχνότητες.

Μετά τον καθορισμό του στατιστικού μοντέλου είναι πλέον δυνατόν να προχωρήσουμε στην εύρεση της μαθηματικής μορφής των προαναφερθέντων εκτιμητών.

4.3.1 Πολυκαναλικός MMSE Εκτιμητής

Εφόσον έχουμε υποθέσει ότι το πηγαίο σήμα και τα σήματα θορύβου είναι διανυσματικές γκαουσιανές στοχαστικές διαδικασίες, ο MMSE εκτιμητής ανάγεται σε γραμμικό εκτιμητή. Στη συνέχεια, οδηγούμαστε στο συγκεκριμένο εκτιμητή μέσα από μια διανυσματική προσέγγιση [Kay93].

Το βέλτιστο διάνυσμα $\mathbf{W}_{opt}$ είναι αυτό που μετασχηματίζει το διάνυσμα $\mathbf{X}$ με τις θορυβώδεις παρατηρήσεις στην καλύτερη δυνατή προσέγγιση του πηγαίου σήματος S, σύμφωνα με το κριτήριο MSE. Για την εύρεση αυτού του βέλτιστου διανύσματος, το οποίο συνιστά το πολυκαναλικό Wiener φίλτρο, πρέπει να ελαχιστοποιήσουμε το MSE στην έξοδο του διαμορφωτή δέσμης. Στο κοινό χωροσυχνοτικό πεδίο το σφάλμα στην έξοδο του διαμορφωτή δέσμης ορίζεται ως: $\mathcal{E} = S - \mathbf{W}^H\mathbf{X}$ και η βέλτιστη λύση, υποθέτοντας ότι ο πίνακας $\mathbf{\Phi}_x$ είναι αντιστρέψιμος, δίνεται από τη σχέση [SBM01]:

$$\mathbf{W}_{opt} = \mathbf{\Phi}_x^{-1}\mathbf{\Phi}_{xs}\,, \tag{4.24}$$

όπου $\mathbf{\Phi}_{xs}$ είναι το διάνυσμα του ετερο-φάσματος ισχύος μεταξύ του πηγαίου σήματος και των σημάτων θορύβου. Αν κάνουμε χρήση της υπόθεσης ότι το πηγαίο σήμα και ο θόρυβος είναι ασυσχέτιστα μεταξύ τους, τότε θα ισχύουν τα εξής:

$$\begin{aligned} \mathbf{\Phi}_x =& \Phi_s\mathbf{D}\mathbf{D}^H + \mathbf{\Phi}_v \\ \mathbf{\Phi}_{xs} =& \Phi_s\mathbf{D}\,. \end{aligned} \tag{4.25}$$

Συνεπώς, το διάνυσμα βέλτιστων βαρών μπορεί να γραφεί ως

$$\mathbf{W}_{opt} = \mathbf{\Phi}_x^{-1}\Phi_s\mathbf{D} = \left[\Phi_s\mathbf{D}\mathbf{D}^H + \mathbf{\Phi}_v\right]^{-1}\Phi_s\mathbf{D}\,. \tag{4.26}$$

Από το Λήμμα Αντιστροφής Πίνακα [Mey00] γνωρίζουμε ότι ο αντίστροφος ενός πίνακα της μορφής $\left[\mathbf{A}^{-1} + \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H\right]^{-1}$ μπορεί να γραφεί στη μορφή:

$$\left[\mathbf{A}^{-1} + \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H\right]^{-1} = \mathbf{A} - \mathbf{A}\mathbf{B}\left(\mathbf{C} + \mathbf{B}^H\mathbf{A}\mathbf{B}\right)^{-1}\mathbf{B}^H\mathbf{A} \tag{4.27}$$

Αν στην Εξ. (4.27) προβούμε στις αντικαταστάσεις

$$\mathbf{A} = \mathbf{\Phi}_v^{-1},\ \mathbf{B} = \sqrt{\Phi_s}\mathbf{D} \quad \text{και} \quad \mathbf{C} = 1,$$

και λάβουμε επιπλέον υπόψιν μας ότι το $k = \mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}$ είναι μια πραγματική βαθμωτή μεταβλητή, τότε ο πίνακας $\left[\Phi_s\mathbf{D}\mathbf{D}^H + \mathbf{\Phi}_v\right]^{-1}$, μπορεί να γραφεί ως

$$\begin{aligned} \left[\Phi_s\mathbf{D}\mathbf{D}^H + \mathbf{\Phi}_v\right]^{-1} =& \mathbf{\Phi}_v^{-1} - \Phi_s\mathbf{\Phi}_v^{-1}\mathbf{D}\left(1 + \Phi_s\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}\right)^{-1}\mathbf{D}^H\mathbf{\Phi}_v^{-1} \\ =& \mathbf{\Phi}_v^{-1} - \frac{\Phi_s\mathbf{\Phi}_v^{-1}\mathbf{D}\mathbf{D}^H\mathbf{\Phi}_v^{-1}}{\left(1 + \Phi_s\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}\right)} \end{aligned} \tag{4.28}$$

Συνεπώς, η (4.26) με τη βοήθεια της (4.28) μπορεί να εκφρασθεί ως

$$\begin{aligned}\mathbf{W}_{opt} &= \left[\mathbf{\Phi}_v^{-1}\mathbf{D} - \frac{\Phi_s\mathbf{\Phi}_v^{-1}\mathbf{D}\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}}{\left(1+\Phi_s\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}\right)}\right]\Phi_s \\ &= \left[1 - \frac{\Phi_s\cdot k}{(1+\Phi_s\cdot k)}\right]\Phi_s\mathbf{\Phi}_v^{-1}\mathbf{D} \\ &= \left(\frac{\Phi_s}{\left(\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}\right)^{-1}+\Phi_s}\right)\frac{\mathbf{\Phi}_v^{-1}\mathbf{D}}{\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}}\end{aligned} \tag{4.29}$$

Η εξίσωση αυτή υποδηλώνει ότι η (4.24) μπορεί να αναλυθεί περαιτέρω σε έναν MVDR διαμορφωτή δέσμης ακολουθούμενο από ένα μονοκαναλικό Wiener φίλτρο το οποίο επιδρά στην έξοδό του [SBM01, Tre02]:

$$\mathbf{W}_{opt}^H = \underbrace{\frac{\mathbf{D}^H\mathbf{\Phi}_v^{-1}}{\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}}}_{\mathbf{W}_{mvdr}^H}\cdot\underbrace{\left(\frac{\Phi_s}{\Phi_s+\Phi_n}\right)}_{\text{Wiener post-filter}}, \tag{4.30}$$

όπου Φ_n είναι το φάσμα ισχύος του θορύβου στην έξοδο του MVDR. Η μεταβλητή Φ_n μπορεί να γραφεί ως:

$$\Phi_n = \mathbf{W}_{mvdr}^H\mathbf{\Phi}_v\mathbf{W}_{mvdr} = \left(\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}\right)^{-1}. \tag{4.31}$$

Από την (4.30) μπορούμε εύκολα να βρούμε τον MMSE εκτιμητή ως: $\hat{S} = \mathbf{W}_{opt}^H\mathbf{X}$.

4.3.2 Βέλτιστοι Μη-Γραμμικοί Εκτιμητές

Από αντιληπτικής άποψης η πληροφορία που λαμβάνουμε από τη φάση του φάσματος ενός ακουστικού σήματος δεν εκτιμάται ως το ίδιο σημαντική με την πληροφορία που λαμβάνουμε από το φασματικό πλάτος του [Var85]. Επομένως, η προσέγγιση της εκτίμησης του φασματικού πλάτους των ακουστικών σημάτων παρουσιάζεται προτιμότερη έναντι της εκτίμησης του μιγαδικού φάσματός τους. Αν γράψουμε $S(k,\ell) = A(k,\ell)e^{j\psi(k,\ell)}$, όπου $A(k,\ell)$ είναι το φασματικό πλάτος βραχέος χρόνου και $\psi(k,\ell)$ η φασματική φάση, τότε ο MMSE-STSA εκτιμητής για την kοστή φασματική συνιστώσα προκύπτει ως η λύση της υπό συνθήκη μέσης τιμής [EM84]:

$$\hat{A} = E\left\{A \,|\, x_0(\cdot),\ldots,x_{N-1}(\cdot)\right\}, \tag{4.32}$$

όπου το $E\{\cdot\}$ δηλώνει τη στατιστική μέση τιμή. Αφού η $\{x_0(\cdot),\ldots,x_{N-1}(\cdot)\}$ και η $\{X_0(\cdot),\ldots,X_{N-1}(\cdot)\}$ είναι ισοδύναμες αναπαραστάσεις και, επιπλέον, οι συντελεστές Fourier κάθε διαδικασίας είναι ασυσχέτιστοι σε διαφορετικές συχνότητες, δηλαδή το $X_i(k_1)$ είναι ανεξάρτητο του $X_j(k_2)$ για $k_1 \neq k_2$, η (4.32) μπορεί να εκφραστεί ως:

$$\begin{aligned}\hat{A} =& E\left\{A \,|\, \{X_1,\ldots,X_{N-1}\} = \mathbf{X}\right\} \\ =& \int_0^\infty A\left(\int_0^{2\pi} p\left(A,\psi \,|\, \mathbf{X}\right)\mathrm{d}\psi\right)\mathrm{d}A,\end{aligned} \tag{4.33}$$

όπου $p\left(A,\psi\right)$ είναι η από κοινού πιθανότητα του πλάτους και της φάσης των ακουστικών σημάτων.

Ομοίως με τον MMSE-STSA εκτιμητή, ο MMSE log-STSA εκτιμητής ελαχιστοποιεί το MSE του λογαρίθμου του φασματικού πλάτους. Στην πραγματικότητα, αυτή η μετρική αλλοίωσης σύμφωνα με το [EM85], είναι καταλληλότερη. Σε αυτή την περίπτωση, ο εκτιμητής προκύπτει ως αποτέλεσμα της ακόλουθης δεσμευμένης μέσης τιμής:

$$\hat{A}_{log} = \exp\left(E\left\{\ln\left(A\right) \,|\, \mathbf{X}\right\}\right). \tag{4.34}$$

Το υπό θεώρηση γκαουσιανό στατιστικό μοντέλο οδηγεί στη Rayleigh από κοινού πιθανότητα:

$$p(A,\psi) = \frac{A}{\pi\Phi_s}\exp\left(-\frac{A^2}{\Phi_s}\right). \tag{4.35}$$

Επιπροσθέτως, η υπό συνθήκη πιθανότητα $p(\mathbf{X}\,|\,A,\psi)$ ισούται με:

$$p(\mathbf{X}\,|\,A,\psi) = \frac{1}{\pi^N \det\left(\mathbf{\Phi}_v\right)}\exp\left(-\left(\mathbf{X}^H - S^*\mathbf{D}^H\right)\mathbf{\Phi}_v^{-1}\left(\mathbf{X}-\mathbf{D}S\right)\right) \tag{4.36}$$

Αυτή η υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας μπορεί να εκφρασθεί ως το γινόμενο δύο συναρτήσεων:

$$p(\mathbf{X}\,|\,A,\psi) = g(A,T(\mathbf{X}))h(\mathbf{X}), \tag{4.37}$$

όπου η g εξαρτάται μόνο από το A και το $T(\mathbf{X})$ και η h εξαρτώνται μόνο από τον πίνακα $\mathbf{X}$ των θορυβωδών παρατηρήσεων. Το $T(\mathbf{X})$ είναι η έξοδος του MVDR διαμορφωτή δέσμης:

$$T(\mathbf{X}) = \frac{\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{X}}{\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}} = \mathbf{W}_{mvdr}^H\mathbf{X}. \tag{4.38}$$

Σύμφωνα με το Θεώρημα Παραγοντοποίησης [Poo98], το $T(\mathbf{X})$ αποδεικνύεται ότι συνιστά επαρκές στατιστικό για το A. Επιπλέον, σύμφωνα με το [BR02], το $T(\mathbf{X})$ λειτουργεί ως επαρκές στατιστικό για το S, αλλά και για κάθε συνάρτηση του S ή του $\rho(S)$. Οι παραπάνω προτάσεις οδηγούν στη διαπίστωση ότι για κάθε ύστερη πιθανότητα του S, η υπό συνθήκη πιθανότητα του S ή μιας συνάρτησης $\rho(S)$ ως προς τις θορυβώδεις παρατηρήσεις $\mathbf{X}$ είναι ισοδύναμη με την υπό συνθήκη πιθανότητα ως προς το $T(\mathbf{X})$:

$$p(\rho(S)\,|\,\mathbf{X}) = p(\rho(S)\,|\,T(\mathbf{X})). \tag{4.39}$$

Λαμβάνοντας υπόψιν αυτή την ισότητα, μπορεί πλέον εύκολα να αποδειχθεί ότι η δεσμευμένη μέση τιμή του $\rho(S)$ ως προς το $\mathbf{X}$ είναι ισοδύναμη με [BR02]:

$$E\left\{\rho(S)\,|\,\mathbf{X}\right\} = E\left\{\rho(S)\,|\,T(\mathbf{X})\right\}. \tag{4.40}$$

Το αποτέλεσμα αυτό είναι εξαιρετικής σημασίας και θα χρησιμοποιηθεί για την εύρεση της τελικής μορφής των MMSE-STSA και MMSE log-STSA εκτιμητών.

4.3.2.1 Πολυκαναλικός MMSE-STSA Εκτιμητής

Για την εύρεση του MMSE-STSA εκτιμητή χρησιμοποιούμε τη (4.40) στην περίπτωση που ισχύει $\rho(S) = A$ λαμβάνοντας:

$$\hat{A} = E\left\{A\,|\,Y = T(\mathbf{X})\right\}. \tag{4.41}$$

Θα πρέπει, δηλαδή, να εκτιμήσουμε τη δεσμευμένη μέση τιμή του φασματικού πλάτους ως προς την έξοδο του MVDR διαμορφωτή δέσμης. Ανακαλώντας ότι ο MVDR διαμορφωτής ικανοποιεί το κριτήριο της μη-παραμόρφωσης, στην έξοδό του θα έχουμε:

$$Y = S + \frac{\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{V}}{\mathbf{D}^H\mathbf{\Phi}_v^{-1}\mathbf{D}}. \tag{4.42}$$

Η κλειστή μορφή της (4.41) προκύπτει ως [EM84]:

$$\hat{A} = G\left(u\right)R, \tag{4.43}$$

$$G(u) = \Gamma(1.5)\frac{\sqrt{u}}{\gamma}\exp\left(-\frac{u}{2}\right)\left[\left(1+u\right)I_0\left(\frac{u}{2}\right) + uI_1\left(\frac{u}{2}\right)\right], \tag{4.44}$$

όπου R είναι το φασματικό πλάτος του Y, $Y(k,\ell) = R(k,\ell)e^{j\vartheta(k,\ell)}$, Γ είναι η συνάρτηση Γάμα και I_0, I_1 είναι οι τροποποιημένες Bessel συναρτήσεις μηδενικής και πρώτης τάξης [AS72], αντιστοίχως. Η μεταβλητή u ορίζεται ως:

$$u = \frac{\xi}{1+\xi} \cdot \gamma, \tag{4.45}$$

όπου τα ξ και γ είναι γνωστά ως *εκ των προτέρων* και *εκ των υστέρων* SNR, αντιστοίχως, και ορίζονται ως:

$$\xi = \frac{\Phi_s}{\Phi_n}, \quad \gamma = \frac{R^2}{\Phi_n}. \tag{4.46}$$

Έχοντας εκτιμήσει το φασματικό πλάτος $\hat{A}$, μπορούμε πλέον να χρησιμοποιήσουμε τη φάση της θορυβώδους εξόδου του MVDR για να υπολογίσουμε το βελτιωμένης ποιότητας ακουστικό σήμα ως $\hat{S} = \hat{A}e^{j\vartheta}$. Η όλη διαδικασία είναι ισοδύναμη με το να επεξεργαστούμε, αρχικά, όλες τις θορυβώδεις παρατηρήσεις μέσω του MVDR διαμορφωτή δέσμης και με το να εφαρμόσουμε, στη συνέχεια, στο μονοκαναλικό πλέον σήμα εξόδου Y ένα ύστερο-φιλτράρισμα με απόκριση συχνότητας $G(u)$ η οποία δίνεται από τη (4.44).

4.3.2.2 Πολυκαναλικός MMSE log-STSA Εκτιμητής

Για την εύρεση του MMSE log-STSA εκτιμητή χρησιμοποιούμε ακόμα μια φορά τη (4.40) για την περίπτωση στην οποία ισχύει $\rho(S) = \ln(A)$ λαμβάνοντας:

$$\hat{A}_{log} = E\left\{\ln(A) \,|\, Y = T(\mathbf{X})\right\}. \tag{4.47}$$

Θα πρέπει, δηλαδή, να εκτιμήσουμε τη δεσμευμένη μέση τιμή του λογαρίθμου του φασματικού πλάτους ως προς την έξοδο του MVDR διαμορφωτή δέσμης. Σε αυτή την περίπτωση, η κλειστή μορφή της έκφρασης (4.47) ισούται με [EM85]:

$$\hat{A}_{log} = G_{log}(u)R, \tag{4.48}$$

$$G_{log}(u) = \frac{\xi}{1+\xi} \exp\left(\frac{1}{2}\int_u^\infty \frac{e^{-t}}{t}\,\mathrm{d}t\right), \tag{4.49}$$

όπου R είναι το φασματικό πλάτος του Y στη (4.42) και τα ξ και γ ορίζονται στη (4.46). Όπως και προηγουμένως, μπορούμε να θεωρήσουμε ότι το αποκατεστημένο ακουστικό σήμα $\hat{S}$ μπορεί να ληφθεί, αν αρχικά επεξεργαστούμε μέσω του MVDR διαμορφωτή δέσμης τις θορυβώδεις παρατηρήσεις $\mathbf{X}$, και, στη συνέχεια, στην έξοδό του εφαρμόσουμε ένα ύστερο-φίλτρο με απόκριση συχνότητας αυτήν που δίνεται στην Εξ. (4.49).

4.4 Εκτίμηση Ύστερου-Φίλτρου

Στην περίπτωση του MVDR διαμορφωτή δέσμης, το διάνυσμα $\mathbf{W}_{mvdr}^H$ στη σχέση (4.23) μπορεί να υπολογιστεί λόγω του ότι η μορφή του είναι ανεξάρτητη από τα δεδομένα. Στην πραγματικότητα, ακόμα και στην περίπτωση που δε μας είναι εκ των προτέρων γνωστός ο πίνακας ετερο-φάσματος ισχύος του θορύβου $\mathbf{\Phi}_v$, μπορούμε να αποδείξουμε ότι υπάρχει λύση η οποία εξαρτάται μόνο από τον πίνακα ετερο-φάσματος ισχύος των θορυβωδών παρατηρήσεων $\mathbf{\Phi}_x$. Παρατηρώντας ότι ο $\mathbf{\Phi}_x$ μπορεί να ξαναγραφεί ως $\mathbf{\Phi}_x = \Phi_s\mathbf{D}\mathbf{D}^H + \mathbf{\Phi}_v$, υπό τη θεώρηση ότι το ακουστικό πηγαίο σήμα και ο θόρυβος είναι ασυσχέτιστα μεταξύ τους και χρησιμοποιώντας την έκφραση για το $\mathbf{\Phi}_x^{-1}$, που δείξαμε στην Εξ. (4.28), μπορούμε να εκφράσουμε το $\mathbf{D}^H\mathbf{\Phi}_x^{-1}$ ως:

$$\mathbf{D}^H\mathbf{\Phi}_x^{-1} = \mathbf{D}^H\mathbf{\Phi}_v^{-1} - \frac{(\Phi_s/\Phi_n)\,\mathbf{D}^H\mathbf{\Phi}_v^{-1}}{(1+\Phi_s/\Phi_n)} = \frac{\mathbf{D}^H\mathbf{\Phi}_v^{-1}}{(1+\Phi_s/\Phi_n)} \tag{4.50}$$

Πλέον, είναι εύκολο να αποδειχθεί ότι ισχύει η ακόλουθη ισότητα:

$$\mathbf{W}_{mvdr}^H = \frac{\mathbf{D}^H\mathbf{\Phi}_x^{-1}}{\mathbf{D}^H\mathbf{\Phi}_x^{-1}\mathbf{D}}\,. \tag{4.51}$$

Στη γενικότερη περίπτωση, στην οποία δεν ισχύει η υπόθεση της μη-συσχέτισης του θορύβου με το σήμα πηγής, ο διαμορφωτής δέσμης που προκύπτει στην Εξ. (4.51) ονομάζεται διαμορφωτής δέσμης ελάχιστης ισχύος και απόκρισης χωρίς παραμόρφωση (minimum power distortionless response - MPDR) [Tre02]. Ο διαμορφωτής αυτός προκύπτει ελαχιστοποιώντας το φάσμα ισχύος της εξόδου υπό τον περιορισμό ότι απουσία θορύβου το σήμα εισόδου θα παραμένει αναλλοίωτο στην έξοδο του διαμορφωτή.

Αντίθετα με τα χωρικά αυτά φίλτρα κάνοντας μια επισκόπηση των εξισώσεων (4.30), (4.44) και (4.49), διαπιστώνουμε ότι, προκειμένου να υπολογίσουμε την τελική μορφή των υπό μελέτη ύστερων-φίλτρων, απαιτείται προηγουμένως να εκτιμήσουμε τις ποσότητες Φ_s και Φ_n. Για την εκτίμηση των παραπάνω ποσοτήτων προτείνουμε μια πρωτότυπη μέθοδο κάνοντας χρήση της μιγαδικής συνάρτησης συνοχής [Elk01].

4.4.1 Ανάλυση του Πεδίου Θορύβου

Σε εφαρμογές στις οποίες χρησιμοποιούνται συστοιχίες αισθητήρων, τα πεδία θορύβου μπορούν να κατηγοριοποιηθούν σύμφωνα με το βαθμό συσχέτισης των λαμβανόμενων σημάτων θορύβου από διαφορετικούς μέσα στο χώρο αισθητήρες. Μια συνήθης μετρική, η οποία χρησιμοποιείται για το χαρακτηρισμό ενός πεδίου θορύβου, είναι η μιγαδική συνάρτηση συνοχής. Η συνάρτηση συνοχής μεταξύ δύο σημάτων x_i και x_j τα οποία εντοπίζονται σε διαφορετικές χωρικές συντεταγμένες ισούται με [Elk01]:

$$C_{x_ix_j}(\omega) = \frac{\Phi_{x_ix_j}(\omega)}{\sqrt{\Phi_{x_i}(\omega)\Phi_{x_j}(\omega)}}\,, \tag{4.52}$$

όπου $\Phi_{x_ix_j}$ είναι το ετερο-φάσμα ισχύος μεταξύ των δύο στοχαστικών διαδικασιών και Φ_{x_i}, Φ_{x_j} τα φάσματα ισχύος τους. Η συνάρτηση συνοχής είναι μια κανονικοποιημένη συνάρτηση πυκνότητας ετερο-φάσματος. Συγκεκριμένα, η κανονικοποίηση περιορίζει την (4.52), ώστε το τετραγωνικό πλάτος της συνάρτησης συνοχής να κείται στο διάστημα $0 \leq |C_{x_ix_j}|^2 \leq 1$.

Σε ένα πεδίο διαχυτικού ή σφαιρικά ισοτροπικού θορύβου, θόρυβος ίσης ενέργειας μεταδίδεται προς όλες τις κατευθύνσεις ταυτοχρόνως. Σε ένα τέτοιο πεδίο θορύβου οι αισθητήρες μιας συστοιχίας λαμβάνουν σήματα θορύβου τα οποία είναι συσχετισμένα κυρίως στις κλίμακες βραχειών συχνοτήτων και έχουν περίπου την ίδια ενέργεια. Το μοντέλο του σφαιρικά ισοτροπικού θορύβου μπορεί να θεωρηθεί ως κατάλληλο μοντέλο για διάφορα περιβάλλοντα θορύβου (π.χ. αυτοκίνητα και γραφεία [MS97, MB03]). Μια ικανοποιητική προσέγγιση της μιγαδικής συνάρτησης συνοχής [Elk01] για ένα τέτοιο πεδίο θορύβου δίνεται στην παρακάτω σχέση:

$$C_{v_iv_j}(\omega) = \frac{\sin(\omega f_s r/c)}{\omega f_s r/c},\ \forall\omega\,, \tag{4.53}$$

όπου το $v_{i,j}$ συμβολίζει το θόρυβο που λαμβάνεται από τους αισθητήρες i και j, r είναι η απόσταση μεταξύ των αισθητήρων, c είναι η ταχύτητα του ήχου και ω είναι η γωνιακή συχνότητα διακριτού χρόνου. Για τα πειράματα που πραγματοποιήσαμε σε αυτή την εργασία θεωρήσαμε την ύπαρξη ενός τέτοιου πεδίου θορύβου.

4.4.2 Γενικευμένη Εκτίμηση Ύστερου-Φίλτρου

Στην ενότητα αυτή προτείνουμε μια πρωτότυπη μέθοδο εκτίμησης για τον υπολογισμό των υπό μελέτη ύστερων-φίλτρων. Η προτεινόμενη μέθοδος είναι κατάλληλη για ένα πλήθος διαφορετικών πεδίων θορύβου και είναι βέλτιστη ως προς όλα τα κριτήρια ελαχιστοποίησης που έχουμε ήδη μελετήσει (MSE, MSE-STSA, MSE log-STSA). Μια γενική εποπτική εικόνα του συνολικού συστήματος για τη βέλτιστη αποθορυβοποίηση ακουστικών σημάτων με χρήση συστοιχίας αισθητήρων παρέχεται στο Σχήμα 4.3. Θα πρέπει να σημειώσουμε ότι για διαφορετικές περιπτώσεις (διαφορετικά κριτήρια ελαχιστοποίησης), το σύστημα αποθορυβοποίησης διαφοροποιείται ως προς το είδος του ύστερου-φίλτρου που χρησιμοποιείται στην έξοδο του MVDR διαμορφωτή δέσμης. Συγκεκριμένα, το συνολικό σύστημα περιλαμβάνει τα ακόλουθα στάδια:

1. Τα πολλαπλά λαμβανόμενα σήματα εισόδου διοχετεύονται σε ένα υποσύστημα χρονικής ευθυγράμμισης. Οι έξοδοι αυτού του υποσυστήματος αντιστοιχούν στις κανονικοποιημένες και χρονικά ευθυγραμμισμένες εισόδους των αισθητήρων. Η διαδικασία αυτή πραγματοποιείται για την αντιμετώπιση των χρονικών καθυστερήσεων και παραμορφώσεων των σημάτων κατά τη φάση της διάδοσής τους από την πηγή προς τους αισθητήρες της συστοιχίας. Τα σήματα στην έξοδο του υποσυστήματος μπορούν να αποδοθούν με μορφή πίνακα ως $\mathbf{X}' = \mathbf{I} \cdot S + \mathbf{V}'$, με $\mathbf{I} = [1, \ldots, 1]^T$ ένα $N \times 1$ διάνυσμα στήλης [3].

2. Οι πολλαπλές (πλέον χρονικά ευθυγραμμισμένες) θορυβώδεις παρατηρήσεις προβάλλονται μέσω του MVDR διαμορφωτή δέσμης στη μονοκαναλική έξοδο Y της σχέσης (4.42). Στην έξοδο αυτή διαπιστώνεται θόρυβος ο οποίος παρουσιάζει την ελάχιστη δυνατή διακύμανση.

3. Ένα από τα υπό μελέτη ύστερα-φίλτρα, αναλόγως με το ποιο κριτήριο ελαχιστοποίησης έχουμε επιλέξει, εφαρμόζεται στην έξοδο Y για την περαιτέρω βελτίωσή της.

4.4.2.1 Εκτίμηση Φάσματος Πηγαίου Σήματος

Με βάση τις μέχρι τώρα θεωρήσεις και την επιπρόσθετη υπόθεση ενός ομογενούς πεδίου θορύβου, στο οποίο όλοι οι αισθητήρες λαμβάνουν σήματα θορύβου που έχουν την ίδια ισχύ ($\Phi_{v_i} = \Phi_v \, \forall i$), ο υπολογισμός των αυτο- και ετερο-φασμάτων ισχύος των χρονικά ευθυγραμμισμένων σημάτων εισόδου οδηγεί στις παρακάτω εξισώσεις:

$$\Phi_{x_i x_j} = \Phi_s + \Phi_{v_i v_j} \tag{4.54}$$

$$\Phi_{x_i} = \Phi_s + \Phi_v. \tag{4.55}$$

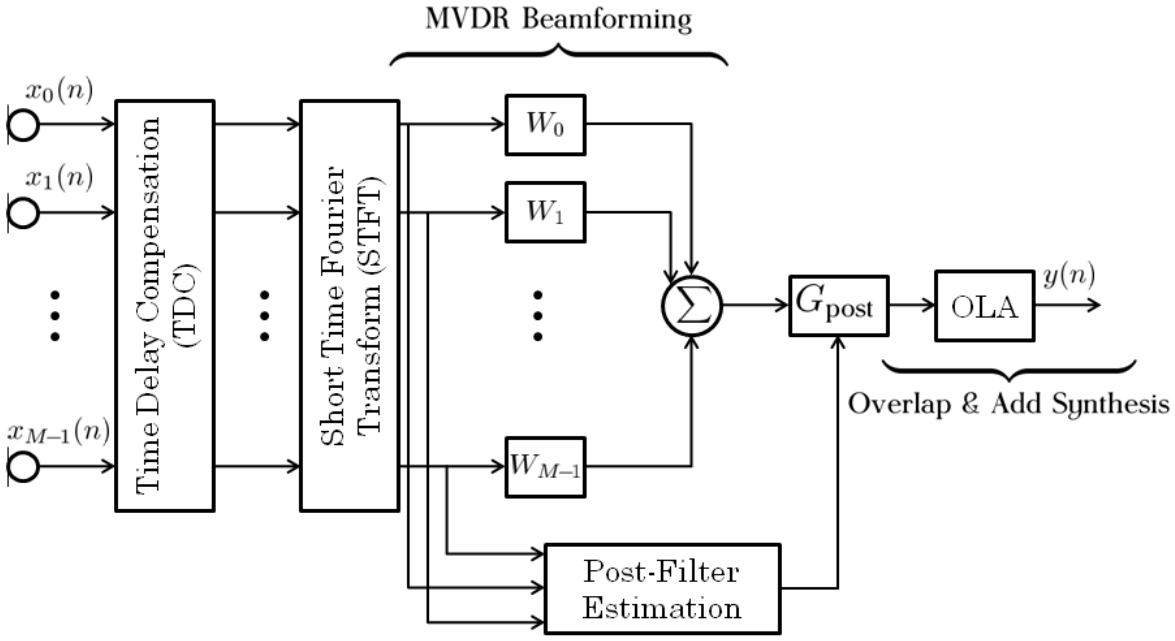
Αν, επιπλέον, υπάρχει διαθέσιμη μια εκτίμηση της συνάρτησης συνοχής, τότε προκύπτει άμεσα - αντικαθιστώντας στην εξίσωση (4.52) τα x_i και x_j με v_i και v_j, αντιστοίχως - ότι το ετερο-φάσμα ισχύος του θορύβου $\Phi_{v_i v_j}$ δίνεται από:

$$\Phi_{v_i v_j} = \Phi_v C_{v_i v_j}. \tag{4.56}$$

Οι εξισώσεις (4.54)–(4.56) σχηματίζουν ένα 3×3 γραμμικό σύστημα εξισώσεων. Παρατηρώντας ότι ισχύει $\Phi_{x_i} = \Phi_{x_j}$ και επιλύοντας ως προς Φ_s, παίρνουμε:

$$\hat{\Phi}_s^{ij} = \frac{\mathrm{Re}\left\{\hat{\Phi}_{x_i x_j}\right\} - \frac{1}{2}\left(\hat{\Phi}_{x_i} + \hat{\Phi}_{x_j}\right) \mathrm{Re}\left\{\hat{C}_{v_i v_j}\right\}}{1 - \mathrm{Re}\left\{\hat{C}_{v_i v_j}\right\}}. \tag{4.57}$$

[3] Από εδώ και στο εξής, όταν χρησιμοποιείται το $\mathbf{X}$, θα αναφέρεται στις χρονικά ευθυγραμμισμένες εκδόσεις των σημάτων $\mathbf{X}'$.



Σχήμα 4.3: Πολυαισθητηριακό σύστημα αποθορυβοποίησης ακουστικών σημάτων με χρήση ύστερου-φιλτραρίσματος.

Η προκύπτουσα λύση αντιστοιχεί στην εκτίμηση του Φ_s κάνοντας χρήση των αυτο- και ετερο-φασμάτων ισχύος μεταξύ των αισθητήρων i και j. Ο συμβολισμός $\hat{(\cdot)}$ χρησιμοποιείται για τις εκτιμώμενες ποσότητες. Αν λάβουμε τη μέση τιμή των φασμάτων ισχύος των εισόδων i και j της συστοιχίας, μπορούμε να βελτιώσουμε την ακρίβεια της εκτίμησης. Η χρήση του τελεστή $\mathrm{Re}\{\cdot\}$ είναι επιβεβλημένη λόγω του ότι εξ ορισμού το φάσμα ισχύος πρέπει να είναι πραγματικός αριθμός. Η ακρίβεια της εκτίμησης μπορεί να βελτιωθεί περαιτέρω λαμβάνοντας τη μέση τιμή όλων των $\binom{N}{2}$ δυνατών συνδυασμών των εισόδων i και j. Έτσι, τελικά προκύπτει:

$$\hat{\Phi}_s = \frac{2}{N(N-1)} \sum_{i=0}^{N-2} \sum_{j=i+1}^{N-1} \hat{\Phi}_s^{ij} \,. \tag{4.58}$$

Η εξαγωγή του συγκεκριμένου αποτελέσματος για την εκτίμηση του αριθμητή του Wiener ύστερου-φίλτρου (4.30) έγινε αρχικά από τους McCowan και Bourlard [MB03]. Ωστόσο, το αποτέλεσμα αυτό αποτελεί μέρος και της δικής μας εκτεταμένης μεθόδου η οποία μπορεί να εφαρμοσθεί σε όλα τα κριτήρια ελαχιστοποίησης. Οι συγγραφείς στο [MB03], προκειμένου να υπολογίσουν τη συνολική απόκριση συχνότητας, εκτιμούν τον παρανομαστή $\Phi_s + \Phi_n$ της (4.30) ως τη μέση τιμή του αθροίσματος των N φασμάτων ισχύος Φ_{x_i}:

$$\Phi_s + \Phi_n = \sum_{i=0}^{N-1} \Phi_{x_i} \,. \tag{4.59}$$

Αυτή η προσέγγιση, όμως, οδηγεί σε μια υποβέλτιστη λύση [MB03, FK97], επειδή υπερεκτιμά το φάσμα ισχύος του θορύβου στην έξοδο του MVDR διαμορφωτή δέσμης. Αυτό αποδίδεται στο γεγονός ότι δε λαμβάνεται υπόψιν η μερική απαλοιφή του θορύβου που έχει ήδη πραγματοποιηθεί από το διαμορφωτή δέσμης.

4.4.2.2 Εκτίμηση Φάσματος Θορύβου

Σε αυτή την ενότητα προτείνουμε μια ακριβέστερη μέθοδο για την εκτίμηση της ποσότητας Φ_n η οποία οδηγεί στη βέλτιστη λύση. Επιπροσθέτως, πρέπει να αναφέρουμε ότι η προτεινόμενη μέθοδος, σε αντίθεση με αυτές των McCowan και Bourlard [MB03] και του Zelinski [Zel88], παρέχει μια ξεχωριστή εκτίμηση του φάσματος ισχύος του θορύβου, Φ_n, στην έξοδο του διαμορφωτή δέσμης. Η εκτίμηση αυτή μπορεί να χρησιμοποιηθεί και για τον υπολογισμό των αποκρίσεων συχνότητας των μη-γραμμικών ύστερων-φίλτρων που δίνονται στις σχέσεις (4.44) και (4.49).

Με βάση την υπόθεση ενός ομογενούς πεδίου θορύβου και κάνοντας χρήση της (4.31), το Φ_n μπορεί να γραφεί ως:

$$\Phi_n = \Phi_v \mathbf{W}_{mvdr}^H \mathbf{C}_v \mathbf{W}_{mvdr} = \frac{\Phi_v}{\mathbf{D}^H \mathbf{C}_v^{-1} \mathbf{D}}, \tag{4.60}$$

όπου $\mathbf{C}_v$ είναι ο πίνακας συνοχής του πεδίου θορύβου και ορίζεται ως:

$$\mathbf{C}_v = \begin{pmatrix} 1 & C_{v_0 v_1} & \dots & C_{v_0 v_{N-1}} \\ C_{v_1 v_0} & 1 & & \\ \vdots & & \ddots & \\ C_{v_{N-1} v_0} & \dots & & 1 \end{pmatrix}. \tag{4.61}$$

Συνεπώς, προκειμένου να εκτιμήσουμε το Φ_n, απαιτείται μόνο η εκτίμηση του Φ_v. Επιλύοντας το σύστημα εξισώσεων (4.54)-(4.56) ως προς Φ_v, λαμβάνουμε:

$$\hat{\Phi}_v^{ij} = \frac{\frac{1}{2}\left(\hat{\Phi}_{x_i} + \hat{\Phi}_{x_j}\right) - \mathrm{Re}\left\{\hat{\Phi}_{x_i x_j}\right\}}{1 - \mathrm{Re}\left\{\hat{C}_{v_i v_j}\right\}}, \tag{4.62}$$

που είναι η εκτίμηση του Φ_v κάνοντας χρήση των αυτο- και ετερο-φασμάτων ισχύος μεταξύ των αισθητήρων i και j. Ακολουθώντας παρόμοια λογική με αυτή που χρησιμοποιήσαμε για τον υπολογισμό του Φ_s, μπορούμε να επιτύχουμε βελτιωμένη ακρίβεια λαμβάνοντας το μέσο όρο των φασμάτων ισχύος μεταξύ των εισόδων i και j και τη μέση τιμή μεταξύ όλων των δυνατών συνδυασμών των αισθητήρων:

$$\hat{\Phi}_v = \frac{2}{N(N-1)} \sum_{i=0}^{N-2} \sum_{j=i+1}^{N-1} \hat{\Phi}_v^{ij}. \tag{4.63}$$

Θα πρέπει να τονιστεί ότι στην περίπτωση που $\hat{C}_{v_i v_j} = 1$, για κάθε $i \neq j$, η εκτίμηση των $\hat{\Phi}_s^{ij}$ (4.57) και $\hat{\Phi}_v^{ij}$ (4.62) οδηγεί σε μια όχι καλά ορισμένη λύση. Ένας απλός τρόπος για την αποφυγή του συγκεκριμένου προβλήματος είναι να περιορίσουμε το μοντέλο της συνάρτησης συνοχής, ώστε $\hat{C}_{v_i v_j} < 1$, για κάθε $i \neq j$.

Μια εναλλακτική προσέγγιση για την εκτίμηση του παρανομαστή της απόκρισης συχνότητας του Wiener ύστερου-φίλτρου $\Phi_s + \Phi_n$ (4.30) είναι να εκτιμήσουμε το φάσμα ισχύος Φ_y απευθείας από την έξοδο του διαμορφωτή δέσμης. Ωστόσο, σε μια τέτοια περίπτωση η εκτίμηση υπολείπεται ακρίβειας, αφού για τον υπολογισμό της έχουμε ένα μόνο σήμα εξόδου αντί των N σημάτων που χρησιμοποιούμε στην προτεινόμενη μέθοδο.

Πολλές φορές, ο MVDR αδυνατεί για πρακτικούς λόγους να απομακρύνει ικανοποιητικά το θόρυβο στις κλίμακες βραχειών συχνοτήτων (δες Σχ. 4.4). Για να αντιμετωπίσουμε αυτή τη δυσλειτουργία, μπορούμε αντί του Φ_n να χρησιμοποιήσουμε μια τροποποιημένη έκδοσή του εκφρασμένη ως:

$$\Phi_n = \begin{cases} \Phi_v, \text{ για } \omega \leq \omega_1 \\ \Phi_n, \text{ για } \omega > \omega_1 \end{cases},$$

όπου το ω_1 ορίζει ένα κάτω φράγμα για την περιοχή των βραχειών κλιμάκων συχνοτήτων. Αφού, πλέον, έχουμε εκτιμήσει τις ποσότητες Φ_s και Φ_n, η εύρεση των υπό μελέτη ύστερων-φίλτρων των σχέσεων (4.30), (4.44) και (4.49) μπορεί να επιτευχθεί άμεσα.

4.5 Πειραματικές Μετρήσεις και Αποτελέσματα

Για να επιβεβαιώσουμε την αποτελεσματικότητα της προτεινόμενης μεθόδου για την εκτίμηση των ύστερων-φίλτρων, συγκρίνουμε την απόδοσή της με άλλες πολυαισθητηριακές τεχνικές αποθορυβοποίησης υπό τη θεώρηση ενός πεδίου σφαιρικά ισοτροπικού θορύβου. Σε αυτές συμπεριλαμβάνονται ο MVDR διαμορφωτής δέσμης [BS01], το γενικευμένο Zelinski ύστερο-φίλτρο [Zel88] και το ύστερο-φίλτρο του McCowan [MB03]. Επιπροσθέτως, παρέχουμε συγκρίσεις με τα αποτελέσματα αποθορυβοποίησης που προκύπτουν, αν στην έξοδο του MVDR διαμορφωτή δέσμης χρησιμοποιήσουμε την "decision directed" μέθοδο εκτίμησης [EM84]. Αυτή είναι μια μονοκαναλική μέθοδος που χρησιμοποιείται για την εκτίμηση της απόκρισης συχνότητας του ύστερου-φίλτρου χωρίς να εκμεταλλεύεται την πολλαπλή πληροφορία που λαμβάνεται από τους αισθητήρες της συστοιχίας.

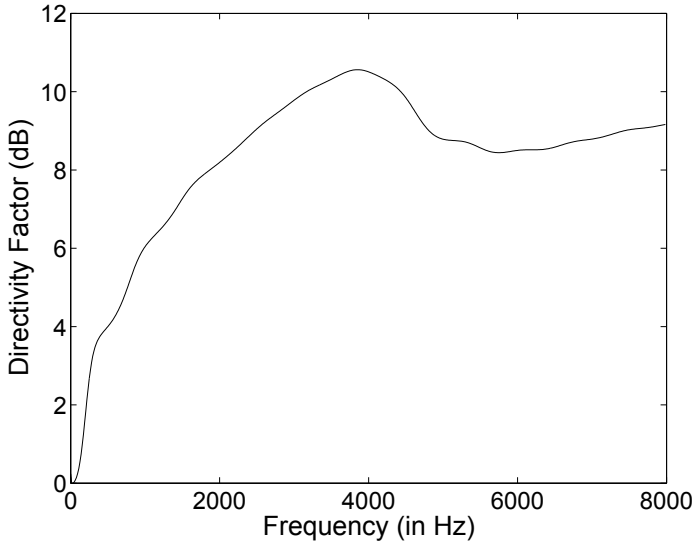
4.5.1 Συλλογή Ακουστικών Δεδομένων και Πρακτική Υλοποίηση του Συστήματος Αποθορυβοποίησης

Τα πολυαισθητηριακά δεδομένα ακουστικών σημάτων, τα οποία χρησιμοποιήθηκαν στις πειραματικές μετρήσεις, προέρχονται από τη CMU βάση δεδομένων [Sul96]. Οι ηχογραφήσεις των σημάτων πραγματοποιήθηκαν με ρυθμό δειγματοληψίας 16 kHz σε περιβάλλον εργαστηρίου με τη χρήση γραμμικής συστοιχίας αισθητήρων. Η συστοιχία αποτελείται από 8 αισθητήρες σε απόσταση 7 εκατοστών μεταξύ τους. Τοποθετήθηκε πάνω σε μια βάση και οι ομιλητές βρίσκονταν σε απόσταση 1 μέτρου από το κέντρο της. Για κάθε ηχογραφημένο σήμα από τους αισθητήρες υπάρχει διαθέσιμη και η αντίστοιχη ηχογράφηση ενός καθαρού σήματος ελέγχου. Το περιβάλλον της ηχογράφησης περιελάμβανε πολλαπλές πηγές θορύβου, όπως ανεμιστήρες, κλιματιστικά και μονάδες ηλεκτρονικών υπολογιστών. Οι υπάρχουσες συνθήκες θορύβου προσεγγίζουν ικανοποιητικά ένα πεδίο σφαιρικά ισοτροπικού θορύβου. Ο χρόνος αντήχησης του δωματίου μετρήθηκε στα 240 ms και το μέσο SNR των ηχογραφήσεων στα 6.5 dB. Το συνολικό πλήθος των δεδομένων απαρτίζεται από 130 εκφωνήσεις, 10 ομιλητές με 13 εκφωνήσεις ο καθένας.

Οι χρονικά ευθυγραμμισμένες θορυβώδεις παρατηρήσεις στις εισόδους των αισθητήρων διαιρούνται σε χρονικά πλαίσια των 400 (25 ms) δειγμάτων με επικάλυψη 300 δειγμάτων ($\approx$ 19 ms) μεταξύ γειτονικών πλαισίων. Κάθε πλαίσιο παραθυροποιείται με ένα παράθυρο Hamming και, στη συνέχεια, εφαρμόζεται μια βραχέος χρόνου STFT ανάλυση. Κατόπιν, οι μετασχηματισμένες είσοδοι προωθούνται στον MVDR διαμορφωτή δέσμης. Προκειμένου να ξεπεραστούν τα σφάλματα κέρδους και φάσης των αισθητήρων, καθώς και το πρόβλημα εισαγωγής επιπλέον θορύβου από τα ηλεκτρονικά της συστοιχίας, το διάνυσμα βαρών του MVDR διαμορφωτή δέσμης υπολογίζεται υπό έναν *white noise gain* περιορισμό [CZK86]. Οι παραγόμενες αποκρίσεις συχνότητας των ύστερων-φίλτρων, εφαρμόζοντας κάθε μία από τις υπό εξέταση μεθόδους, προκύπτουν χρησιμοποιώντας τις θορυβώδεις παρατηρήσεις ως εισόδους στο σύστημα αποθορυβοποίησης (βλέπε Σχήμα 4.3). Τα αυτο- και ετερο-φάσματα ισχύος Φ_{x_i} και $\Phi_{x_i x_j}$ υπολογίζονται χρησιμοποιώντας τη βραχέος χρόνου μέθοδο εκτίμησης που προτείνεται στο [ABB77]:

$$\hat{\Phi}_{x_i x_j}(k,\ell) = \alpha \hat{\Phi}_{x_i x_j}(k,\ell-1) + (1-\alpha) x_i(k,\ell) x_j^*(k,\ell)\,. \tag{4.64}$$

Η μέθοδος αυτή δίνει βελτιωμένες εκτιμήσεις και ομαλότερα φάσματα και μπορεί να θεωρηθεί ως ένα επαναληπτικό Welch περιοδόγραμμα. Ο όρος α στη (4.64) είναι ένας αριθμός κοντά στη μονάδα και το σύμβολο * υποδηλώνει το συζυγή μιγαδικό. Τελικά, το βελτιωμένης ποιότητας



Σχήμα 4.4: Απεικόνιση του παράγοντα κατευθυντικότητας του MVDR διαμορφωτή δέσμης ο οποίος περιγράφει την ικανότητα του διαμορφωτή δέσμης να περιορίσει το θόρυβο. Η ικανότητά του αυτή παρουσιάζεται μικρή στην περιοχή των κλιμάκων βραχειών συχνοτήτων.

σήμα εξόδου μετασχηματίζεται ξανά στο πεδίο του χρόνου χρησιμοποιώντας τη μέθοδο σύνθεσης Επικάλυψης και Αθροίσματος (OLA - Overlap and Add) [RS78].

4.5.2 Πειράματα Αποθορυβοποίησης Ακουστικών Σημάτων

Προκειμένου να είμαστε σε θέση να συγκρίνουμε την προτεινόμενη μέθοδο ύστερου-φιλτραρίσματος με τις υπόλοιπες πολυαισθητηριακές μεθόδους αποθορυβοποίησης και τη μονοκαναλική “decision directed" μέθοδο εκτίμησης, χρησιμοποιούμε πέντε διαφορετικές αντικειμενικές μετρικές για την ποιότητα του εξαγόμενου ακουστικού σήματος. Για να αποτιμήσουμε το μέγεθος εξάλειψης του θορύβου, κάνουμε χρήση της αύξησης του τμηματικού σηματοθορυβικού λόγου (SSNRE - Segmental Signal-to-Noise Ratio Enhancement). Αυτή η μετρική ισούται με τη διαφορά του τμηματικού SNR της ενισχυμένης εξόδου από τη μέση τιμή του τμηματικού SNR των θορυβωδών παρατηρήσεων στις εισόδους των αισθητήρων. Το τμηματικό SNR ορίζεται στο [HP98] και αποτελεί καταλληλότερο κριτήριο απόδοσης από το κλασσικό SNR. Χρονικά πλαίσια του σήματος με SNR πάνω από $35dB$ δε συνεισφέρουν σημαντικά στη συνολική ποιότητα του σήματος. Επιπλέον, χρονικά πλαίσια τα οποία αποτελούνται μόνο από σιγή μπορούν να έχουν SNR με ακραία αρνητικές τιμές οι οποίες δεν αντανακλούν την πραγματική συνεισφορά των συγκεκριμένων πλαισίων στο συνολικό σήμα. Για αυτούς τους λόγους, το SNR σε κάθε χρονικό πλαίσιο περιορίζεται να κείται σε ένα εύρος τιμών $[-10, -35]$ dB. Για να αποτιμήσουμε την ακουστική ποιότητα των ενισχυμένων σημάτων εξόδου, χρησιμοποιούμε τις μετρικές Log-Area-Ratio distance (LAR), Log-Likelihood Ratio (LLR), Itakura-Saito Distortion (IS) [HP98] και τη Log-Spectral Distance (LSD) [Coh04]. Οι συγκεκριμένες μετρικές έχει βρεθεί ότι παρουσιάζουν υψηλή συσχέτιση με τον τρόπο που αντιλαμβάνεται ο ανθρώπινος εγκέφαλος την ακουστική ποιότητα [FS96]. Οι χαμηλές τιμές στις παραπάνω τέσσερις μετρικές ποιότητας υποδηλώνουν υψηλή ακουστική ποιότητα.

Τα αποτελέσματα από όλες τις προαναφερθείσες μετρικές παρουσιάζονται στον Πίνακα 4.1 και αφορούν σε όλες τις υπό μελέτη μεθόδους. Τα αποτελέσματα αυτά έχουν ληφθεί ως οι μέσοι όροι από όλη τη βάση δεδομένων. Με το πρόθεμα “dd” δηλώνονται τα αποτελέσματα που ελήφθησαν

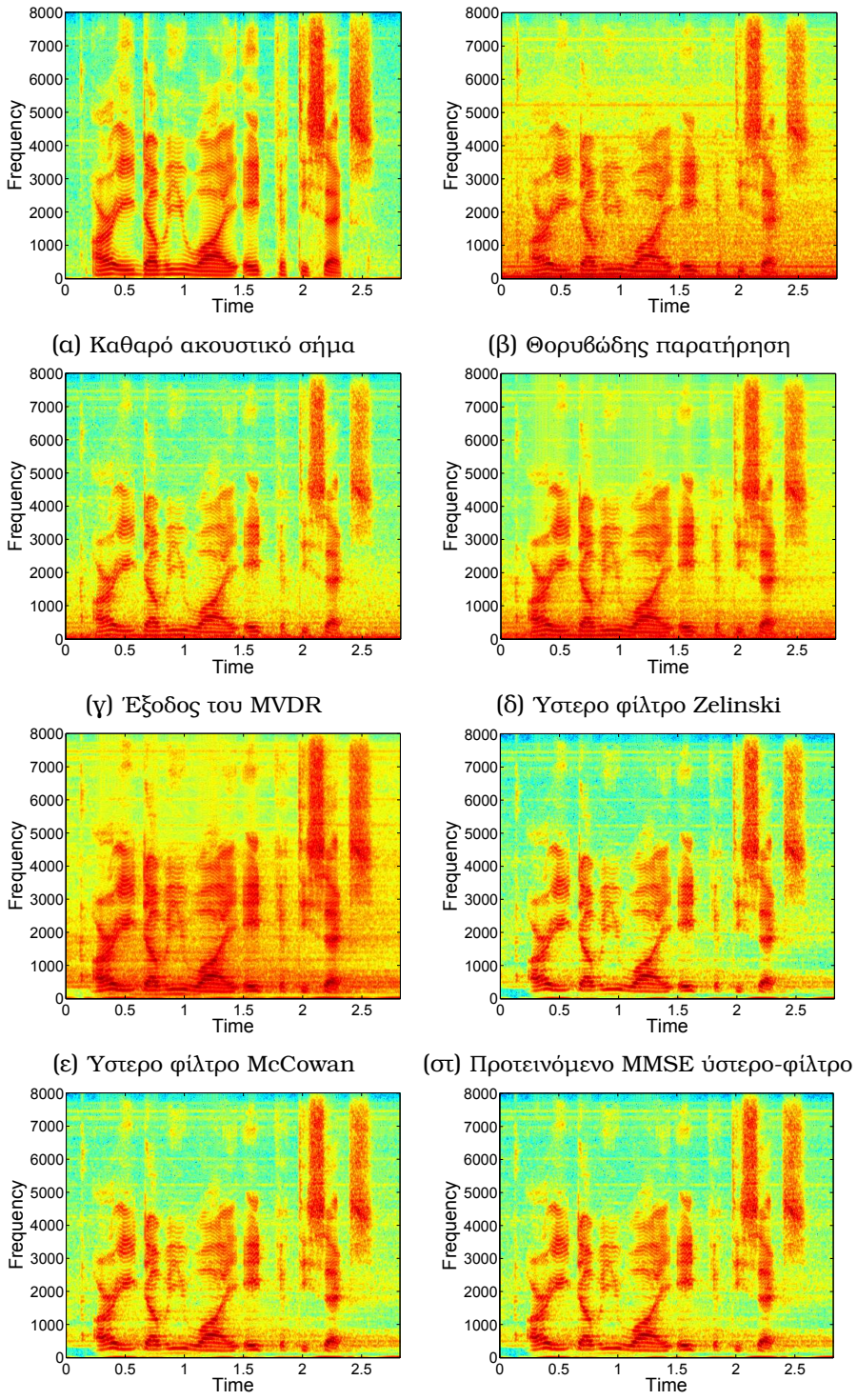
χρησιμοποιώντας την "decision directed" μέθοδο. Στις τελευταίες τρεις σειρές του Πίνακα 4.1 παρουσιάζονται τα αποτελέσματα που προέκυψαν κάνοντας χρήση της προτεινόμενης μεθόδου για τα τρία διαφορετικά κριτήρια ελαχιστοποίησης. Επιπλέον, στο Σχήμα 4.5 απεικονίζονται τυπικά φασματογραφήματα για κάθε μια από τις πολυαισθητηριακές μεθόδους, οι οποίες έχουν μέχρι στιγμής αναφερθεί, καθώς και φασματογραφήματα του καθαρού σήματος πηγής και της θορυβώδους εισόδου στον κεντρικό αισθητήρα της συστοιχίας.

Πίνακας 4.1: Ποιοτικές μετρήσεις στα δεδομένα που προέκυψαν από τα πειράματα στη βάση δεδομένων CMU .

	SSNRE (dB)	IS	LAR	LLR (dB)	LSD (dB)
Noisy Input	––	1.973	8.314	6.920	8.341
MVDR	0.024	1.260	10.969	6.409	6.794
Zelinski	0.097	5.950	13.774	9.912	6.166
McCowan	5.707	3.279	6.764	7.156	3.775
MMSEdd *	3.071	13.518	8.575	8.666	4.921
STSAdd	2.137	6.984	9.439	6.936	5.519
log-STSAdd	2.621	11.559	9.392	7.417	5.135
MMSE	6.361	0.988	4.425	4.742	3.511
STSA	6.221	0.992	4.431	4.727	3.525
log-STSA	6.320	0.989	4.425	4.733	3.512

* Το πρόθεμα "dd" δηλώνει την "decision directed" μέθοδο.
* Οι χαμηλές τιμές σε όλες τις μετρικές ποιότητας πλην του SSNRE υποδηλώνουν υψηλή ακουστική ποιότητα.

Παρατηρώντας τόσο τα αποτελέσματα που παρουσιάζονται στον Πίνακα 4.1 όσο και τα φασματογραφήματα του Σχήματος (4.5), μπορούμε να διαπιστώσουμε ότι ούτε η απομονωμένη χρήση του MVDR διαμορφωτή δέσμης, αλλά ούτε και το ύστερο-φίλτρο του Zelinski απαλείφουν ικανοποιητικά το θόρυβο σε σχέση με τις άλλες τέσσερις πολυαισθητηριακές μεθόδους και την "decision directed" προσέγγιση. Συγκεκριμένα, από τα Σχήματα 4.5(γ) και 4.5(δ) διαπιστώνουμε ότι οι δύο αυτές μέθοδοι παρουσιάζονται αδύναμες να απαλείψουν το θόρυβο από τις βραχείες κλίμακες συχνοτήτων. Η αδυναμία αυτή του MVDR διαμορφωτή δέσμης αποδίδεται στο γεγονός ότι το μεγαλύτερο μέρος της ενέργειας του θορύβου είναι συγκεντρωμένο στις βραχείες κλίμακες συχνοτήτων, όπου ο διαμορφωτής δέσμης παρουσιάζει μειωμένο παράγοντα κατευθυντικότητας (βλέπε Σχ. 4.4). Η "φτωχή" απόδοση του ύστερου-φίλτρου του Zelinski είναι αναμενόμενη, καθώς η συγκεκριμένη μέθοδος είναι βασισμένη στην υπόθεση ότι το πεδίο θορύβου είναι χωρικά ασυσχέτιστο. Αυτή η υπόθεση οδηγεί σε ένα ακατάλληλο μοντέλο για τις επικρατούσες συνθήκες θορύβου. Υιοθετώντας τη συνολική υπόθεση ότι ο θόρυβος σε όλες τις συχνότητες είναι ασυσχέτιστος μεταξύ των καναλιών, το ύστερο-φίλτρο του Zelinski απαλείφει επιτυχώς το θόρυβο στις μεσαίες και υψηλές κλίμακες συχνοτήτων, αλλά δεν έχει καμία απολύτως επίδραση στη χαμηλή κλίμακα συχνοτήτων, όπου η συσχέτιση είναι σημαντική. Μια επιπλέον εξήγηση δίδεται και στο [FS96] στο οποίο αποδεικνύεται ότι η μέθοδος του Zelinski μπορεί να έχει ανεκτή απόδοση μόνο για χρόνους αντήχησης πάνω από 300 ms. Για μικρούς χρόνους αντήχησης η ποιότητα του σήματος εισόδου μπορεί να είναι χειρότερη από αυτή του σήματος εξόδου. Από την άλλη πλευρά, το ύστερο-φίλτρο του McCowan παρουσιάζει καλύτερη απόδοση από τις προηγούμενες δύο μεθόδους. Αυτό οφείλεται στο ότι η εκτίμηση του φάσματος του πηγαίου σήματος πραγματοποιείται χρησιμοποιώντας τη συσχέτιση του θορύβου μεταξύ των αισθητήρων της συστοιχίας. Ωστόσο, η απόδοσή του εξακολουθεί να είναι κατώτερη από αυτή των προτεινόμενων ύστερων-φίλτρων για τους λόγους που έχουν ήδη αναφερθεί. Τέλος, όσον αφορά στην "decision directed" μέθοδο, η απαλοιφή του θορύβου που επιτυγχάνει είναι μεγαλύτερη από αυτή των δύο πρώτων μεθόδων, αλλά με κόστος τη χαμηλότερη



(α) Καθαρό ακουστικό σήμα (β) Θορυβώδης παρατήρηση

(γ) Έξοδος του MVDR (δ) Ύστερο φίλτρο Zelinski

(ε) Ύστερο φίλτρο McCowan (στ) Προτεινόμενο MMSE ύστερο-φίλτρο

(ζ) Προτεινόμενο STSA ύστερο-φίλτρο (η) Προτεινόμενο log-STSA ύστερο-φίλτρο

Σχήμα 4.5: Φασματογραφήματα της εκφώνησης "r-e-w-y-8-56". (α) Καθαρό ακουστικό σήμα. (β) Θορυβώδης παρατήρηση στον κεντρικό αισθητήρα της συστοιχίας (IS=1.44). (γ) Παρατηρήσιμο ακουστικό σήμα στην έξοδο του MVDR διαμορφωτή δέσμης (SSNRE=0.02 dB, IS=0.90). (δ) Γενικευμένο ύστερο-φίλτρο Zelinski (SSNRE=0.17 dB, IS=2.89). (ε) Ύστερο φίλτρο McCowan (SSNRE=3.95 dB, IS=2.08). (στ) MMSE (SSNRE=4.54 dB, IS=0.81). (ζ) STSA (SSNRE=4.46 dB, IS=0.82). (η) log-STSA (SSNRE=4.52 dB, IS=0.81).

ποιότητα του σήματος εξαιτίας των έντονων παραμορφώσεων της εξόδου.

Από τα παρεχόμενα αποτελέσματα μπορούμε να συμπεράνουμε ότι η απόδοση των μεθόδων που χρησιμοποιούν το προτεινόμενο σχήμα εκτίμησης των ύστερων φίλτρων υπερτερεί συνολικά των αποδόσεων των υπολοίπων υπό εξέταση μεθόδων στη συγκεκριμένη βάση δεδομένων σε όλες τις αντικειμενικές μετρικές ποιότητας. Επιπροσθέτως, παρατηρώντας τα Σχήματα 4.5(α)–4.5(η) μπορούμε να διαπιστώσουμε ότι τα φασματογραφήματα που πλησιάζουν περισσότερο σε αυτό του καθαρού σήματος εισόδου είναι εκείνα που προκύπτουν, αν εφαρμοστεί η προτεινόμενη μέθοδος στην εκτίμηση της απόκρισης συχνότητας των ύστερων-φίλτρων. Αυτό αιτιολογείται με βάση το γεγονός ότι τα παραγόμενα ύστερα-φίλτρα, λόγω της ακριβέστερης εκτίμησης του φάσματος ισχύος του θορύβου, επιτυγχάνουν ικανοποιητική απαλοιφή του θορύβου σε όλες τις συχνοτικές κλίμακες (βραχείες - μεσαίες - υψηλές), ενώ παράλληλα, παρέχουν υψηλότερη ακουστική ποιότητα χωρίς να παραμορφώνουν περαιτέρω το σήμα. Επιπλέον, τα παρόμοια βελτιωμένα αποτελέσματα που λαμβάνονται με τα διαφορετικά κριτήρια ελαχιστοποίησης (MSE, MSE-STSA, MSE log-STSA), υποδεικνύουν ότι η προτεινόμενη μέθοδος εκτίμησης μπορεί να χρησιμοποιηθεί επιτυχώς και στις τρεις προκύπτουσες μεθόδους αποθορυβοποίησης.

4.6 Σύνοψη και Συμπεράσματα

Στο κεφάλαιο αυτό μελετήσαμε μεθόδους αποθορυβοποίησης ακουστικών σημάτων με τη χρήση συστοιχίας αισθητήρων. Η χρήση μιας διάταξης από αισθητήρες αντί ενός μοναδικού αισθητήρα δίνει τη δυνατότητα να εκμεταλλευτεί κανείς την επιπλέον πληροφορία που διατίθεται για το ακουστικό περιβάλλον. Το γεγονός αυτό οδηγεί στην ανάπτυξη χωρικών φίλτρων που είναι ικανά να αντιμετωπίσουν με μεγαλύτερη επιτυχία την επίδραση του θορύβου σε σχέση με κλασσικές μονοκαναλικές μεθόδους. Για ακόμη μεγαλύτερη ενίσχυση των αλλοιωμένων από θόρυβο σημάτων είναι δυνατή η χρήση στατιστικών κριτηρίων που οδηγούν σε βέλτιστα αποτελέσματα.

Η δυσκολία που υπάρχει στην υλοποίηση των στατιστικά βέλτιστων φίλτρων είναι ότι σε αντίθεση με τα χωρικά φίλτρα απαιτείται η γνώση παραμέτρων που εξαρτώνται άμεσα από τα σήματα ενδιαφέροντος, δηλαδή τα σήματα πηγής που θέλουμε να εκτιμήσουμε. Για την αντιμετώπιση αυτού του προβλήματος προτείναμε μια μέθοδο εκτίμησης των απαιτούμενων παραμέτρων. Το πλεονέκτημα που παρουσιάζει η συγκεκριμένη μέθοδος έναντι εναλλακτικών είναι ότι μπορεί να εφαρμοστεί για τον υπολογισμό της απόκρισης συχνότητας μιας ευρείας κατηγορίας φίλτρων που προκύπτουν ως αποτέλεσμα βελτιστοποίησης διαφόρων στατιστικών κριτηρίων, ενώ, παράλληλα, παρουσιάζει μεγαλύτερη ακρίβεια στις παρεχόμενες εκτιμήσεις.

Κεφάλαιο 5

Πολυαισθητηριακή Επεξεργασία για την Αυτόματη Ανίχνευση και Αναγνώριση Ακουστικών Σημάτων

Το πρόβλημα της αυτόματης αναγνώρισης ακουστικών σημάτων, αλλά και ειδικότερα της αυτόματης αναγνώρισης φωνής (Automatic Speech Recognition - ASR) έχει απασχολήσει και απασχολεί εδώ και αρκετές δεκαετίες ένα μεγάλο τμήμα της επιστημονικής κοινότητας. Στόχος της μηχανικής αναγνώρισης είναι η προσπάθεια αντιστοίχισης των ακουστικών σημάτων με ακρίβεια και αποδοτικότητα με διακριτά σύμβολα και η αποτύπωσή τους με τέτοιο τρόπο, ώστε να είναι κατανοητός από έναν υπολογιστή [RS07]. Τα κίνητρα για τη μελέτη και επίλυση του συγκεκριμένου προβλήματος ποικίλλουν και σχετίζονται τόσο με την ανάπτυξη εφαρμογών επικοινωνίας μεταξύ ανθρώπου-υπολογιστή (αυτόματη αναγνώριση φωνής) όσο και με την ανάπτυξη εφαρμογών αναγνώρισης και κατηγοριοποίησης μεγάλου όγκου δεδομένων και πληροφορίας (αυτόματη αναγνώριση προτύπων).

Η ραγδαία τεχνολογική ανάπτυξη ειδικά των τελευταίων ετών επέτρεψε στους επιστήμονες να σημειώσουν σημαντική πρόοδο στην αντιμετώπιση του συγκεκριμένου προβλήματος [RJ93]. Με τη χρήση εξελιγμένων μεθόδων επεξεργασίας σημάτων και αναγνώρισης προτύπων είναι πλέον δυνατή η με μεγάλη ακρίβεια αυτόματη αναγνώριση φωνής ή άλλων ακουστικών σημάτων σε ελεγχόμενο περιβάλλον. Παρόλα αυτά, η ανάπτυξη συστημάτων αυτόματης αναγνώρισης που θα έχουν σταθερή απόδοση ανεξάρτητα από τον περιβάλλοντα χώρο και τις συνθήκες θορύβου παραμένει ένα σημαντικό πρόβλημα προς επίλυση. Πράγματι, ενώ, για παράδειγμα, αρκετά σύγχρονα συστήματα αναγνώρισης φωνής επιτυγχάνουν ακρίβεια αναγνώρισης απομονωμένων ψηφίων (isolated digits) της τάξης του 99%, τα ίδια συστήματα παρουσιάζουν δραματική μείωση στην απόδοσή τους όταν, τα υπό αναγνώριση ψηφία εμφανίζουν έστω και μικρές αλλοιώσεις λόγω ύπαρξης θορύβου [Gon95, OSM98, OMS01]. Είναι λοιπόν επιτακτική η ανάγκη ανάπτυξης νέων μεθόδων και τεχνικών που θα συμβάλλουν στη δημιουργία εύρωστων στο θόρυβο συστημάτων αναγνώρισης.

Το πρόβλημα της εύρωστης αναγνώρισης είναι δυνατόν να αντιμετωπιστεί σε διάφορα επίπεδα επεξεργασίας, όπως κατά την προεπεξεργασία του ακουστικού σήματος (αποθορυβοποίηση), κατά την εξαγωγή χαρακτηριστικών (feature extraction) ή τέλος, στο επίπεδο των στατιστικών μοντέλων. Οι τεχνικές ενίσχυσης των ακουστικών σημάτων (acoustic signal enhancement), όπως αυτές που μελετήσαμε και στο Κεφ. 4, χρησιμοποιούνται για τη βελτίωση της ποιότητας των σημάτων που διοχετεύονται στο σύστημα αναγνώρισης [Lim83, Eph92]. Η επίδρασή τους, ωστόσο, δεν είναι προφανής, αφού δεν υπάρχει απευθείας σύνδεση της ενίσχυσης της αντιληπτικής ποιότητας του σήματος με την απόδοση της αναγνώρισης, ακόμα και αν το στάδιο της προεπεξεργασίας είναι κοινό κατά τη φάση της εκπαίδευσης και ελέγχου του στατιστικού συστήματος. Η δεύτερη

κατηγορία τεχνικών περιλαμβάνει μεθόδους των οποίων στόχος είναι η εξαγωγή συμπαγών και εύρωστων συνόλων χαρακτηριστικών που θα μπορούν να αντιμετωπίσουν με επιτυχία το φαινόμενο της ασυμφωνίας του ακουστικού χώρου μεταξύ των καθαρών και θορυβωδών σημάτων. Τέλος, η τρίτη κατηγορία αποτελείται από τεχνικές για την προσαρμοστική μοντελοποίηση των παραμέτρων των στατιστικών μοντέλων. Στο παρόν κεφάλαιο, διερευνάται η δυνατότητα βελτίωσης της διαδικασίας αυτόματης αναγνώρισης σε συνθήκες θορύβου με την ανάπτυξη μιας μεθόδου που ανήκει στην δεύτερη κατηγορία, δηλαδή στη διαδικασία εξαγωγής εύρωστων στο θόρυβο χαρακτηριστικών. Η προτεινόμενη μέθοδος βασίζεται σε ένα σχήμα αποδιαμόρφωσης των ακουστικών σημάτων σε πλήθος φασματικών συνιστωσών, ενώ, παράλληλα, εκμεταλλεύεται τις πολλαπλές εκδοχές του ακουστικού σήματος που λαμβάνονται από μια συστοιχία αισθητήρων.

Στη συνέχεια του κεφαλαίου και συγκεκριμένα στην Ενοτ. 5.1 παρέχουμε μια σύντομη εισαγωγή στις τεχνικές ακουστικής επεξεργασίας που απαιτούνται για την εξαγωγή διανυσμάτων χαρακτηριστικών. Στην Ενοτ. 5.2 μελετούμε τις ιδιότητες δύο βασικών ενεργειακών τελεστών που χρησιμοποιούνται για τη λήψη ενεργειακών μετρήσεων και παρέχουμε ένα πλαίσιο ενοποίησής τους. Στην Ενοτ. 5.3 με βάση την προηγούμενη ανάλυση αναπτύσσουμε ένα πολυαισθητηριακό σχήμα αποδιαμόρφωσης το οποίο αποτελεί και τη βάση του πολυαισθητηριακού σχήματος που προτείνουμε για την εξαγωγή εύρωστων στο θόρυβο χαρακτηριστικών. Τέλος, στην Ενοτ. 5.4 παρουσιάζουμε πειραματικά αποτελέσματα, προκειμένου να επιβεβαιώσουμε την αποδοτικότητα του προτεινόμενου σχήματος, ενώ στην Ενοτ. 5.5 παρέχεται μια σύνοψη των θεμάτων που μελετήθηκαν.

5.1 Επεξεργασία Ακουστικών Σημάτων

Το πρώτο κύριο μέρος ενός συστήματος αυτόματης αναγνώρισης αποτελεί η επεξεργασία των ακουστικών σημάτων (frontend). Η ακουστική πληροφορία, η οποία εμπεριέχεται στα σήματα εισόδου, είναι δυνατόν να κωδικοποιηθεί αποτελεσματικά οδηγώντας σε σημαντική μείωση του αρχικού μεγέθους τους. Η κωδικοποίηση και συμπίεση των δεδομένων είναι εξαιρετικής σημασίας, αφού μόνο υπό αυτές τις συνθήκες μπορεί να καταστεί υπολογιστικά εφικτή η διαδικασία αναγνώρισης. Με την επεξεργασία των ακουστικών σημάτων η ακουστική πληροφορία κωδικοποιείται σε μια ακολουθία από διανύσματα χαρακτηριστικών τα οποία είναι αυτά που τελικά θα χρησιμοποιηθούν από το δεύτερο κύριο τμήμα του συστήματος αυτόματης αναγνώρισης, δηλαδή αυτό που αποτελείται από τα υποσυστήματα εκπαίδευσης και αποκωδικοποίησης των στατιστικών μοντέλων (backend).

5.1.1 Συστοιχίες Φίλτρων

Η επεξεργασία των ακουστικών σημάτων από συστοιχίες φίλτρων (filterbanks) απαντάται πολύ συχνά ως βασικό συστατικό μεθόδων ανάλυσης και εξαγωγής χαρακτηριστικών που βασίζονται στον υπολογισμό της ενέργειας. Στην επεξεργασία σημάτων μια συστοιχία φίλτρων είναι μια διάταξη από ζωνοπερατά φίλτρα που μπορούν να αναλύσουν το σήμα εισόδου σε πολλαπλές συνιστώσες με την κάθε μια από αυτές να διατηρεί την φασματική πληροφορία από μια βασική συχνοτική υποζώνη του αρχικού σήματος. Αυτού του είδους η ανάλυση είναι εμπνευσμένη από ενδείξεις ότι το φυσιολογικό ακουστικό σύστημα επεξεργάζεται το ακουστικό σήμα φιλτράροντας το συχνοτικό του περιεχόμενο σε διαφορετικές επικαλυπτόμενες συχνοτικές περιοχές [All94]. Σε μια τέτοια διαδικασία καθοριστικής σημασίας είναι η επιλογή των θέσεων των φίλτρων σε συνδυασμό με το εύρος ζώνης τους αλλά και το είδος τους.

5.1.1.1 Μη-Γραμμική Mel και Bark Κλίμακα Συχνοτήτων

Ένας αρκετά διαδεδομένος τρόπος κατανομής των φίλτρων στη συχνότητα είναι η τοποθέτησή τους σύμφωνα με την Mel κλίμακα που είναι εμπνευσμένη από τη φυσιολογία του ανθρώπινου αυτιού. Η κλίμακα Mel προκύπτει από τη γραμμική κλίμακα συχνοτήτων μέσω του μετασχηματισμού

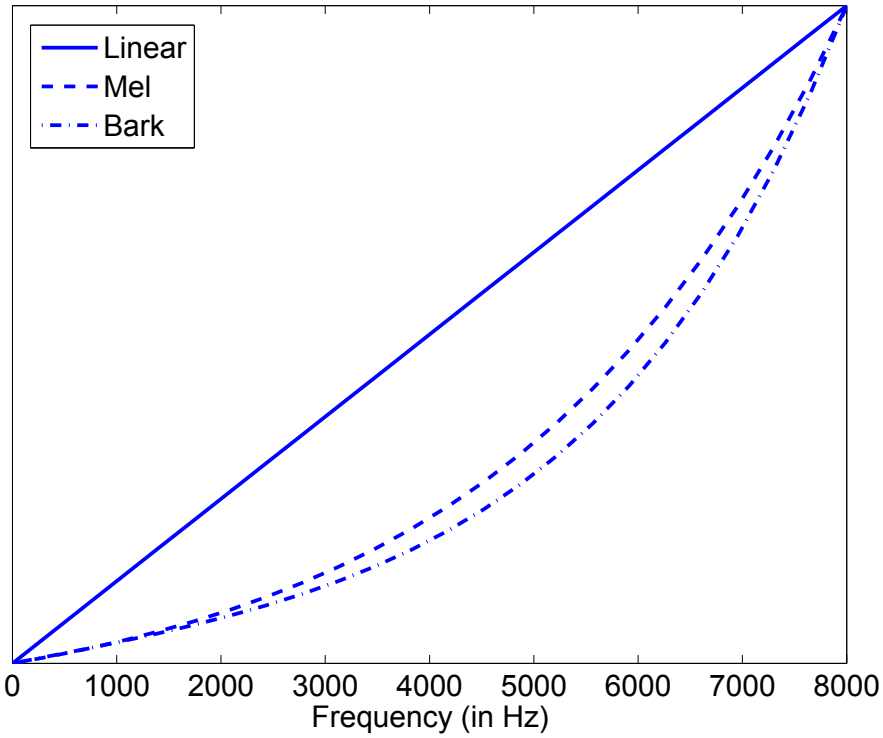
$$mel(f) = 1127 \log\left(f/700 + 1\right), \tag{5.1}$$

όπου $0 \leq f \leq F_s/2$ και F_s η συχνότητα δειγματοληψίας. Τα φίλτρα τοποθετούνται με ομοιόμορφο τρόπο στην κλίμακα των Mel συχνοτήτων ο οποίος αντιστοιχεί σε μη-ομοιόμορφη τοποθέτησή τους στη γραμμική κλίμακα. Τα πολύ δημοφιλή MFCC - (Mel Frequency Cepstrum Coefficients) χαρακτηριστικά [DM80] προκύπτουν μέσα από μια διαδικασία που περιλαμβάνει την επεξεργασία των ακουστικών σημάτων στη Mel κλίμακα συχνοτήτων.

Μια άλλη ευρύτατα διαδεδομένη κλίμακα συχνοτήτων είναι η επίσης μη-γραμμική Bark κλίμακα που προτάθηκε ύστερα από μελέτες στην ακουστική αντίληψη. Η Bark κλίμακα χρησιμοποιείται για την εξαγωγή των εξίσου αποδοτικών PLP - (Perceptual Linear Prediction) χαρακτηριστικών [Her90] και προκύπτει μέσω του παρακάτω μετασχηματισμού της γραμμικής κλίμακας συχνοτήτων

$$bark(f) = \frac{26.81 f}{f + 1960} - 0.53\,, \tag{5.2}$$

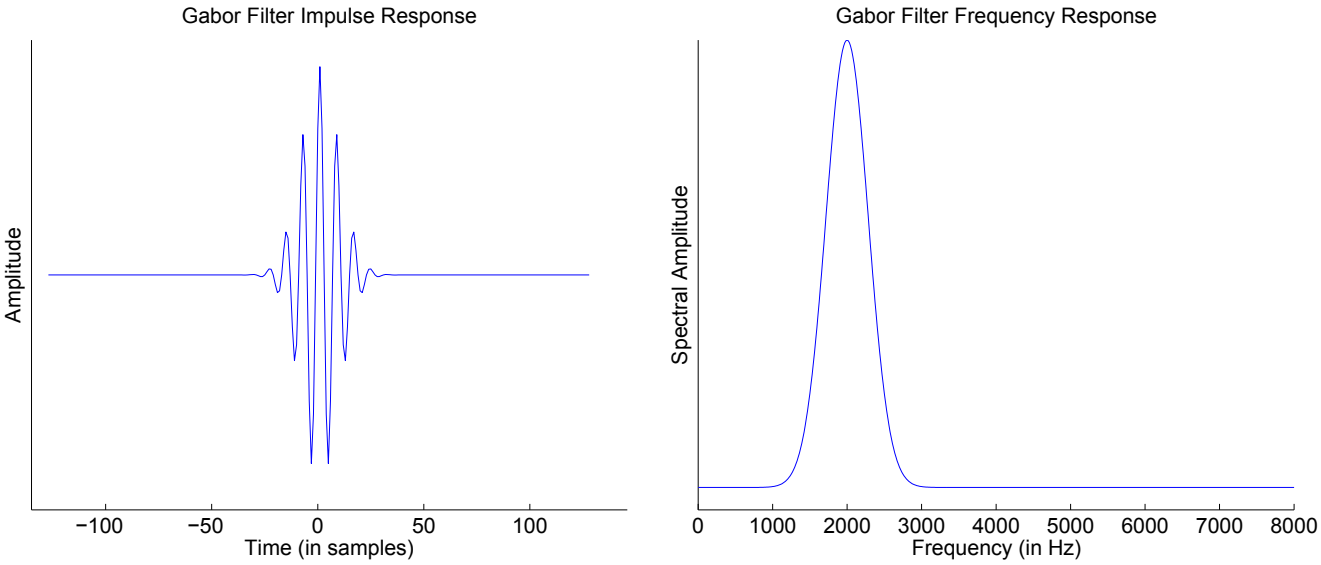
για $0 \leq f \leq F_s/2$. Στο Σχ. 5.1 απεικονίζεται ο τρόπος τοποθέτησης των φίλτρων στη γραμμική κλίμακα σε ένα εύρος ζώνης 8 kHz, αν αυτή γίνει με ομοιόμορφο τρόπο στη Mel και τη Bark κλίμακα συχνοτήτων.



Σχήμα 5.1: Ομοιόμορφη τοποθέτηση των φίλτρων στη γραμμική, Mel και Bark κλίμακα συχνοτήτων.

5.1.2 Φίλτρα για Ακουστική Επεξεργασία

Για την ακουστική επεξεργασία σημάτων έχουν προταθεί στη διεθνή βιβλιογραφία αρκετά φίλτρα με τα πιο γνωστά και ευρέως διαδεδομένα να είναι τα τριγωνικά [DM80], τα Gabor [Gab46] και τα Gammatone [IP97]. Τα τριγωνικά φίλτρα που ορίζονται στο συχνοτικό πεδίο χρησιμοποιούνται για την εξαγωγή των MFCCs διανυσμάτων χαρακτηριστικών [DM80], λόγω της απλότητας υλοποίησής τους και της υπολογιστικής ευκολίας που συνεπάγονται, ενώ, παράλληλα, οδηγούν και σε πολύ υψηλά ποσοστά αναγνώρισης.



(α) (β)

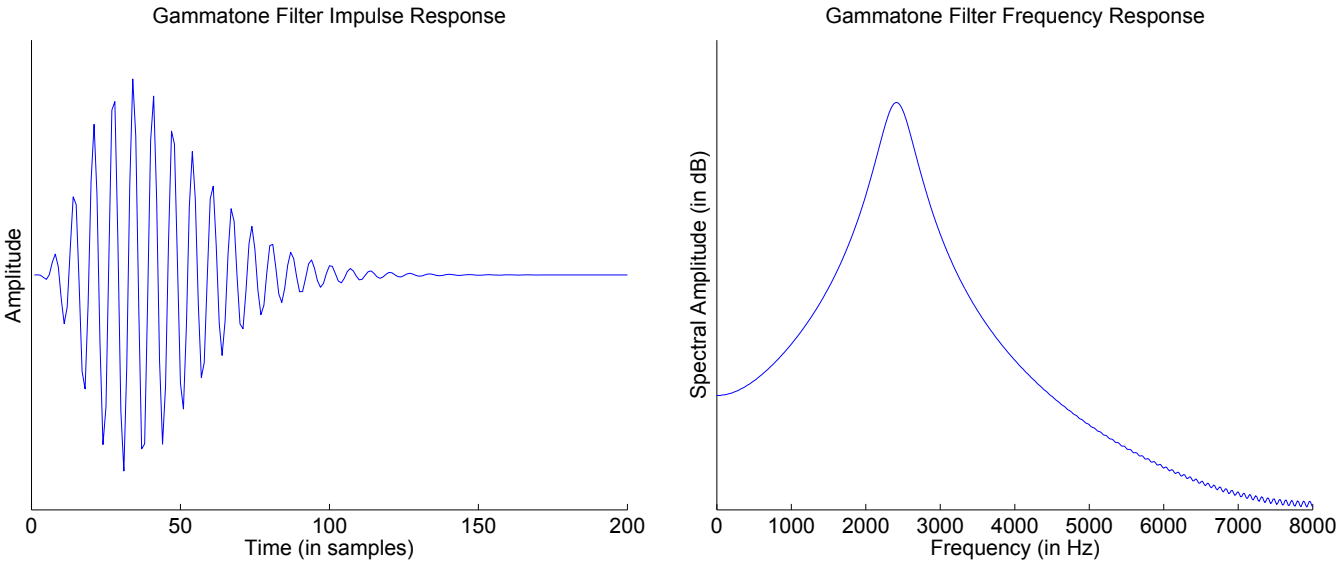
Σχήμα 5.2: (α) Κρουστική απόκριση και (β) απόκριση συχνότητας του Gabor φίλτρου με κεντρική συχνότητα $f_c = 2000$ Hz και ενεργό εύρος ζώνης 200 Hz.

Μια άλλη οικογένεια, αυτή των Gabor φίλτρων, εμφανίζεται επίσης ως ιδανική υποψήφιος για την ακουστική επεξεργασία και ανάλυση, λόγω των ευνοϊκών ιδιοτήτων που διαθέτει. Τα Gabor φίλτρα [Gab46] είναι διαμορφωμένες συναρτήσεις Gauss από ημίτονα (φέροντα) και η κρουστική απόκριση και η απόκριση συχνότητάς τους δίνονται από:

$$h(t) = e^{-\alpha^2 t^2} \cos \omega_c t \tag{5.3}$$

$$H(\omega) = \frac{\sqrt{\pi}}{2a} \left[e^{-\frac{(\omega-\omega_c)^2}{4\alpha^2}} + e^{-\frac{(\omega+\omega_c)^2}{4\alpha^2}} \right], \tag{5.4}$$

όπου ω_c είναι η κεντρική συχνότητα του φίλτρου και α είναι η παράμετρος που καθορίζει το ενεργό εύρος ζώνης του φίλτρου (rms εύρος ζώνης). Στο Σχ. 5.2 απεικονίζεται η κρουστική απόκριση και η απόκριση συχνότητας ενός Gabor φίλτρου με κεντρική συχνότητα $f_c = 2000$ Hz και ενεργό εύρος ζώνης 200 Hz. Η χρήση των Gabor φίλτρων σε πολλές περιπτώσεις προτιμάται για την ακουστική επεξεργασία, λόγω του ότι είναι συμπαγή, ομαλά και στη μιγαδική τους εκδοχή, όταν το φέρον είναι μιγαδικό, επιτυγχάνουν την ελάχιστη δυνατή αβεβαιότητα. Αυτό σημαίνει ότι το γινόμενο του rms εύρους ζώνης συχνοτήτων τους με τη χρονική rms διάρκειά τους δίνει τη μικρότερη δυνατή τιμή σε σχέση με όλα τα άλλα φίλτρα. Η τιμή αυτή, μάλιστα, συμπίπτει με το κάτω όριο που προκύπτει από την ανισότητα της αρχής της αβεβαιότητας [Gab46]. Το γεγονός αυτό πρακτικά δηλώνει ότι τα μιγαδικά Gabor φίλτρα επιτρέπουν ίδιας ευκρίνειας ανάλυση τόσο στο χρονικό όσο και στο συχνοτικό πεδίο.



Σχήμα 5.3: (α) Κρουστική απόκριση και (β) απόκριση συχνότητας του Gammatone φίλτρου με κεντρική συχνότητα $f_c = 2000$ Hz και εύρος ζώνης $b = 200$ Hz.

Τέλος, τα gammatone είναι γραμμικά φίλτρα που περιγράφονται από μια κρουστική απόκριση η οποία προκύπτει από το γινόμενο μιας συνάρτησης Γάμα και ενός ημιτόνου

$$h(t) = \alpha t^{n-1} e^{-2\pi bt} \cos \omega_c t \,, \tag{5.5}$$

όπου α το πλάτος, n η τάξη, b το εύρος και ω_c η κεντρική συχνότητα του φίλτρου. Παράδειγμα ενός τέτοιου φίλτρου δίνεται στο Σχ. 5.3 το οποίο έχει κεντρική συχνότητα $f_c = 2000$ Hz και εύρος ζώνης $b = 200$ Hz. Τα gammatone φίλτρα έχει δειχθεί ότι περιγράφουν δεδομένα ενεργοποίησης πρωτευόντων νευρώνων τα οποία έχουν συλλεγεί από το ακουστικό σύστημα της γάτας.

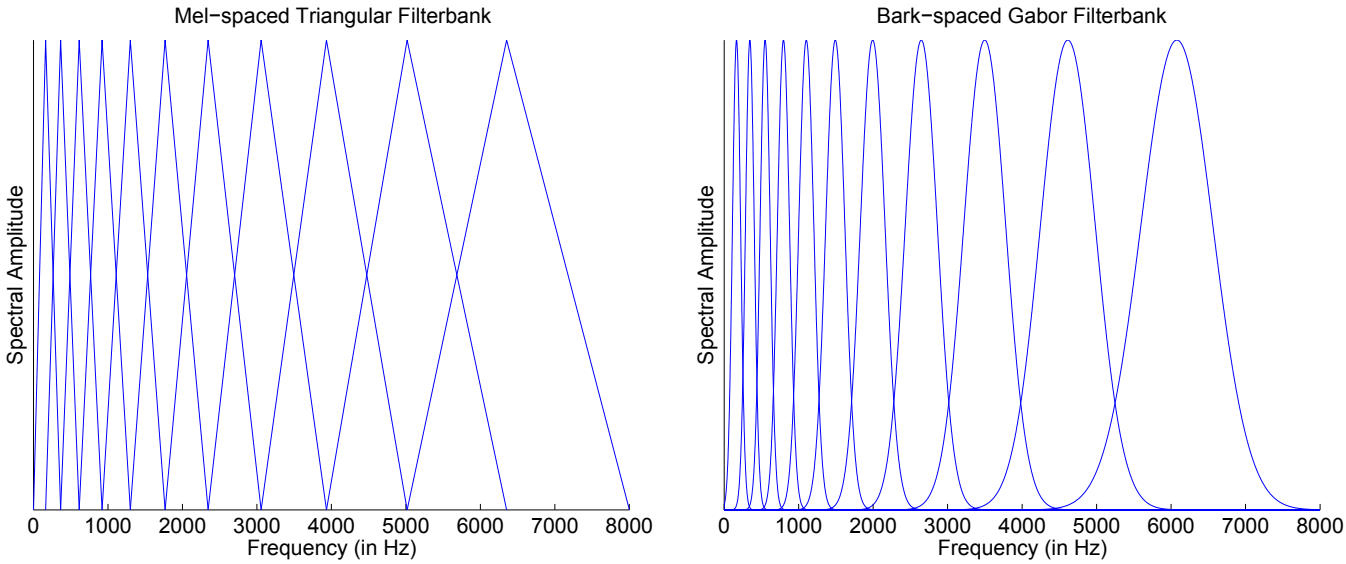
5.1.3 Ενεργειακές Μετρήσεις

Μετά την ανάλυση των ακουστικών σημάτων σε φασματικές συνιστώσες μέσω μιας επιλεγμένης συστοιχίας φίλτρων μια τυπική διαδικασία για την εξαγωγή χαρακτηριστικών διανυσμάτων περιλαμβάνει τον υπολογισμό της ζωνοπερατής ενέργειας σε κάθε μια από τις προκύπτουσες ζώνες συχνοτήτων. Στη συνέχεια, ανάλογα με το είδος των ενεργειακών χαρακτηριστικών που επιθυμεί κανείς να υπολογίσει, ακολουθεί το τελικό στάδιο κατά το οποίο οι τιμές της ενέργειας μετασχηματίζονται κατάλληλα για την τελική εξαγωγή των διανυσμάτων. Για τον υπολογισμό των τιμών της ενέργειας έχει προταθεί στη διεθνή βιβλιογραφία η χρήση δύο εναλλακτικών ενεργειακών τελεστών. Ο πρώτος είναι ο τελεστής τετραγωνικού πλάτους που, όπως δηλώνει και το όνομα του, υπολογίζει την ενέργεια ενός σήματος ως το τετράγωνο του πλάτους του, ενώ ο δεύτερος είναι ο ενεργειακός τελεστής Teager-Kaiser τον οποίο και θα περιγράψουμε στη συνέχεια.

5.1.3.1 Ενεργειακός Τελεστής Teager-Kaiser

Για έναν αρμονικό ταλαντωτή με μάζα m προσαρτημένη στην άκρη ενός ελατηρίου με σταθερά k, η διαφορική εξίσωση κίνησης δίνεται από τη σχέση

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0 \,, \tag{5.6}$$



Σχήμα 5.4: Συστοιχία από (α) 11 τριγωνικά φίλτρα ομοιόμορφα διατεταγμένα στην Mel κλίμακα και (β) 11 Gabor φίλτρα ομοιόμορφα διατεταγμένα στη Bark κλίμακα συχνοτήτων.

με λύση τη $x(t) = a\cos[\omega_0(t) + \theta]$, όπου a το πλάτος και ω_0 η συχνότητα της ταλάντωσης. Η στιγμιαία ενέργεια E_0 αυτού του ταλαντωτή είναι σταθερή και ισούται με το άθροισμα της κινητικής και δυναμικής του ενέργειας

$$E_0 = \frac{1}{2}kx^2 + \frac{1}{2}m\dot{x}^2 \Rightarrow E_0 = \frac{1}{2}ma^2\omega_0^2\,. \tag{5.7}$$

Συνεπώς, η ενέργεια ενός γραμμικού ταλαντωτή εξαρτάται τόσο από το πλάτος όσο και από τη συχνότητα της ταλάντωσης. Λαμβάνοντας υπόψιν του αυτό το μηχανικό ανάλογο ο Teager [TT90] αντιλήφθηκε την ανάγκη ανάλυσης των σημάτων με βάση την ενέργεια που απαιτείται για την παραγωγή τους και ανέπτυξε ένα μη γραμμικό ενεργειακό τελεστή τον οποίο συστηματικά παρουσίασε στην επιστημονική κοινότητα ο Kaiser [Kai90a, Kai90b]. Ο *Teager-Kaiser* ενεργειακός τελεστής (ΤΕΟ) Ψ μπορεί να ανιχνεύσει τη στιγμιαία ενέργεια μιας πηγής που εκτελεί μια γραμμική αρμονική ταλάντωση. Όταν ο τελεστής Ψ εφαρμόζεται σε ένα σήμα συνεχούς χρόνου, τότε η έξοδός του δίνεται από:

$$\Psi\left[x(t)\right] = \dot{x}(t)^2 - x(t)\ddot{x}(t)\,, \tag{5.8}$$

όπου τα $\dot{x}(t)$ και $\ddot{x}(t)$ συμβολίζουν την πρώτη και δεύτερη παράγωγο του $x(t)$, αντιστοίχως. Εφαρμόζοντας τον τελεστή αυτόν σε ένα AM-FM σήμα της μορφής $x(t) = a(t)\cos\left(\phi(t)\right)$, στην έξοδο του τελεστή λαμβάνουμε το ενεργειακό γινόμενο

$$\Psi\left[a(t)\cos\left(\phi(t)\right)\right] \approx a(t)^2\omega(t)^2\,, \tag{5.9}$$

όπου $\omega(t) = \dot{\phi}(t)$ είναι η στιγμιαία γωνιακή συχνότητα. Το σφάλμα της προσέγγισης είναι αμελητέο [MKQ93a] για την περίπτωση που το στιγμιαίο πλάτος και η συχνότητα δε μεταβάλλονται πολύ γρήγορα ή έντονα ως προς τη μέση τιμή του $\omega(t)$.

Για το πρόβλημα της αποδιαμόρφωσης AM-FM σημάτων οι Maragos, Kaiser και Quatieri χρησιμοποίησαν την ιδιότητα (5.9) του Teager ενεργειακού τελεστή και ανέπτυξαν τον Αλγόριθμο

Αποδιαμόρφωσης Ενέργειας Συνεχούς Χρόνου [MKQ93b] (Continuous-Time Energy Separation Algorithm - CESA) ο οποίος επεκτάθηκε και σε εικόνες στο [Eva07]. Ο αλγόριθμος αυτός έχει τη δυνατότητα για γενικά σήματα AM-FM να υπολογίζει το στιγμιαίο πλάτος και τη στιγμιαία φάση του σήματος ως:

$$\frac{\Psi\left[x(t)\right]}{\sqrt{\Psi\left[\dot{x}(t)\right]}} = |a(t)| \tag{5.10}$$

$$\sqrt{\frac{\Psi\left[\dot{x}(t)\right]}{\Psi\left[x(t)\right]}} = \omega(t)\,. \tag{5.11}$$

Επιπλέον, η χρήση του επεκτάθηκε και στην αποδιαμόρφωση πιο σύνθετων σημάτων, όπως τα σήματα φωνής τα οποία μοντελοποιούνται ως άθροισμα περισσοτέρων από ένα AM-FM σημάτων. Για να αντιμετωπιστεί η εγγενής αδυναμία του ενεργειακού τελεστή να εφαρμόζεται σε σήματα με πολλαπλές φασματικές συνιστώσες, προηγείται της ενεργειακής μέτρησης ζωνοπερατό φιλτράρισμα με στόχο την απομόνωση των επιμέρους φασματικών συνιστωσών [MKQ93b, BQ93]. Κάτι τέτοιο επιτυγχάνεται με τη χρήση συστοιχιών φίλτρων, όπως είναι τα Gabor φίλτρα [Gab46, PM99b].

Ο Teager-Kaiser ενεργειακός τελεστής παρουσιάζει το πλεονέκτημα έναντι του ενεργειακού τελεστή τετραγωνικού πλάτους ότι στον υπολογισμό της ενέργειας λαμβάνει υπόψιν του πέρα από το πλάτος του σήματος και τη συχνότητά του. Συνεπώς, παρέχει πλουσιότερη "ενεργειακή πληροφορία". Το γεγονός αυτό ώθησε πολλούς ερευνητές στην εκτεταμένη χρήση του αντί του κλασσικού τελεστή ενέργειας κατά την ανάπτυξη μεθόδων και αλγορίθμων σε πληθώρα εφαρμογών, όπως αναγνώριση φωνής [Pot96, JC99, PM01, DMP05a, DMP05b], ανάλυση και σύνθεση φωνής [PM99a], ανίχνευση ακουστικής δραστηριότητας [PM99b, EM06, KVE+08].

5.2 Ενοποίηση Ενεργειακών Μετρήσεων

Στην ενότητα αυτή μελετούμε τους δύο ενεργειακούς τελεστές, ΤΕΟ και τελεστή τετραγωνικού πλάτους. Βασικός λόγος εξέτασης των τελεστών αυτών είναι η πολύ διαδεδομένη χρήση τους στη λήψη ζωνοπερατών ενεργειακών μετρήσεων οι οποίες χρησιμοποιούνται κατά τη διαδικασία εξαγωγής διανυσμάτων χαρακτηριστικών σε συστήματα αυτόματης αναγνώρισης. Κύριος στόχος της μελέτης μας είναι να εξετάσουμε τις ιδιότητες των τελεστών και να παρουσιάσουμε ένα πλαίσιο ενοποίησης των μετρήσεων που λαμβάνονται από τη χρήση τους. Από εδώ και στο εξής τη μέτρηση που θα λαμβάνουμε στην έξοδο του ΤΕΟ θα τη δηλώνουμε ως Teager ενέργεια, ενώ την αντίστοιχη μέτρηση στην έξοδο του ενεργειακού τελεστή τετραγωνικού πλάτους ως τυπική ενέργεια.

5.2.1 Ισοδυναμία Ενεργειακών Μετρήσεων

Θεωρούμε ένα στάσιμο υπό την ευρεία έννοια (Wide Sense Stationary-WSS) πραγματικό σήμα $x(t)$ και ένα γραμμικό χρονικά αμετάβλητο (ΓΧΑ) φίλτρο $h(t)$. Στην ανάλυση που ακολουθεί αρχικά υποθέτουμε ότι το φίλτρο $h(t)$ είναι επίσης πραγματικό. Στην Ενοτ. 5.2.3 αίρουμε το συγκεκριμένο περιορισμό και δείχνουμε ότι τα αποτελέσματα και τα συμπεράσματα, τα οποία προκύπτουν, ισχύουν και στη γενικότερη περίπτωση που το φίλτρο $h(t)$ είναι μιγαδικό.

Η φιλτραρισμένη εκδοχή του σήματος $x(t)$ από το φίλτρο $h(t)$ δίνεται από τη σχέση

$$y(t) = x(t) * h(t)\,,$$

όπου το σύμβολο $*$ δηλώνει την πράξη της συνέλιξης, ενώ η μέση τυπική ενέργεια του παραγόμενου σήματος $y(t)$ ισούται με

$$E\left\{|y(t)|^2\right\} = E\left\{y(t)y^*(t)\right\} = R_y\left(0\right) . \tag{5.12}$$

Αντιστοίχως, η μέση Teager ενέργεια του $y(t)$, υποθέτοντας ότι αυτό είναι δεύτερης τάξης παραγωγίσιμο, θα ισούται με

$$E\left\{\Psi\left[y(t)\right]\right\} = E\left\{\dot{y}(t)^2 - y(t)\ddot{y}(t)\right\} = R_{\dot{y}}\left(0\right) - R_{y\ddot{y}}\left(0\right) . \tag{5.13}$$

Προκειμένου να καταλήξουμε σε μια σχέση ισότητας μεταξύ των δύο παραπάνω ενεργειακών μετρήσεων, θα προχωρήσουμε στην αναλυτική έκφραση των ποσοτήτων $R_y\left(\tau\right)$, $R_{\dot{y}}\left(\tau\right)$ και $R_{y\ddot{y}}\left(\tau\right)$.

Μπορούμε να συνδέσουμε τη συνάρτηση αυτοσυσχέτισης του σήματος $y(t)$, $R_y\left(\tau\right)$, με τη συνάρτηση αυτοσυσχέτισης του αρχικού μας σήματος $x(t)$, $R_x\left(\tau\right)$, ως εξής:

$$\begin{aligned} R_y\left(\tau\right) =& E\left\{y(t+\tau)y(t)\right\} = E\left\{y(t+\tau)\int_{-\infty}^{\infty} h(u)x(t-u)\,du\right\} \\ =& \int_{-\infty}^{\infty} h(u)E\left\{y(t+\tau)x(t-u)\right\}\,du = \int_{-\infty}^{\infty} h(u)R_{yx}\left(\tau+u\right)\,du \\ =& h(-\tau) * R_{yx}\left(\tau\right) \end{aligned} \tag{5.14}$$

$$\begin{aligned} R_{yx}\left(\tau\right) =& E\left\{y(t+\tau)x(t)\right\} = E\left\{x(t)\int_{-\infty}^{\infty} h(u)x(t+\tau-u)\,du\right\} \\ =& \int_{-\infty}^{\infty} h(u)E\left\{x(t+\tau-u)x(t)\right\}\,du = \int_{-\infty}^{\infty} h(u)R_x\left(\tau-u\right)\,du \\ =& h(\tau) * R_x\left(\tau\right) \end{aligned} \tag{5.15}$$

Συνδυάζοντας τις Εξ. (5.14) και (5.15) καταλήγουμε στην Εξ. (5.16) που συνδέει τη συνάρτηση αυτοσυσχέτισης του αρχικού μας σήματος, $x(t)$, με την αντίστοιχη συνάρτηση αυτοσυσχέτισης της φιλτραρισμένης του εκδοχής, $y(t)$:

$$R_y\left(\tau\right) = h(-\tau) * h(\tau) * R_x\left(\tau\right) . \tag{5.16}$$

Για να υπολογίσουμε τις υπόλοιπες ποσότητες $R_{\dot{y}}\left(\tau\right)$ και $R_{y\ddot{y}}\left(\tau\right)$, θα βασιστούμε στην παραπάνω ανάλυση και, επιπλέον, θα κάνουμε χρήση της ιδιότητας της αντιμετάθεσης της συνέλιξης με την παραγώγιση:

$$\frac{d}{dt}\left(x(t) * h(t)\right) = x(t) * \frac{d}{dt}h(t) = \frac{d}{dt}x(t) * h(t) . \tag{5.17}$$

Λόγω της συγκεκριμένης ιδιότητας, το σήμα $\dot{y}(t)$ μπορεί να γραφεί ως $\dot{y}(t) = x(t) * \dot{h}(t)$ και, επομένως, προκύπτει άμεσα από την Εξ. (5.16) ότι

$$R_{\dot{y}}\left(\tau\right) = \dot{h}(-\tau) * \dot{h}(\tau) * R_x\left(\tau\right) . \tag{5.18}$$

Επειδή επίσης ισχύει ότι $\ddot{y}(t) = x(t) * \ddot{h}(t)$, η ετεροσυσχέτιση $R_{y\ddot{y}}\left(\tau\right)$ μπορεί να γραφεί ως

$$\begin{aligned} R_{y\ddot{y}}(\tau) =& E\left\{y(t+\tau)\ddot{y}(t)\right\} = E\left\{y(t+\tau)\int_{-\infty}^{\infty}\ddot{h}(u)x(t-u)\,du\right\} \\ =& \int_{-\infty}^{\infty}\ddot{h}(u)E\left\{y(t+\tau)x(t-u)\right\}\,du = \int_{-\infty}^{\infty}\ddot{h}(u)R_{yx}(\tau+u)\,du \\ =& \ddot{h}(-\tau) * R_{yx}(\tau)\ . \end{aligned} \tag{5.19}$$

Συνδυάζοντας τις Εξ. (5.15) και (5.19) λαμβάνουμε τελικά

$$R_{y\ddot{y}}(\tau) = \ddot{h}(-\tau) * h(\tau) * R_x(\tau)\ . \tag{5.20}$$

Κάνοντας πάλι χρήση της αντιμετάθεσης της συνέλιξης με την παραγώγιση μπορούμε να εκφράσουμε την ετεροσυχέτιση $R_{y\ddot{y}}(\tau)$ ως:

$$\begin{aligned} R_{y\ddot{y}}(\tau) =& -\frac{d}{d\tau}\left(\dot{h}(-\tau)\right) * h(\tau) * R_x(\tau) \\ =& -\dot{h}(-\tau) * \dot{h}(\tau) * R_x(\tau) = -R_{\dot{y}}(\tau)\ . \end{aligned} \tag{5.21}$$

Από τις (5.13) και (5.21) προκύπτει τελικά ότι

$$E\left\{\Psi\left[y(t)\right]\right\} = 2R_{\dot{y}}(0)\ . \tag{5.22}$$

Εξετάζοντας τις Εξ. (5.12) και (5.22) είναι πλέον εύκολο να διαπιστώσουμε ότι η μέση Teager ενέργεια του σήματος $y(t)$ ισούται με δύο φορές τη μέση τυπική ενέργεια του σήματος $\dot{y}(t)$,

$$E\left\{\Psi\left[y(t)\right]\right\} = 2E\left\{\left|\dot{y}(t)\right|^2\right\}\ . \tag{5.23}$$

Μια αντίστοιχη σχέση για ντετερμινιστικά σήματα έχει δειχθεί ότι ισχύει στο [MKQ93b]. Η παραπάνω σχέση είναι πολύ σημαντική, αφού καταρχάς μας δίνει τη δυνατότητα να υπολογίζουμε με πολύ αποδοτικό τρόπο τη μέση Teager ενέργεια ενός σήματος. Ουσιαστικά, προκειμένου να υπολογίσουμε τη μέση Teager ενέργεια του σήματος $y(t)$, αρκεί να υπολογίσουμε τη μέση τυπική ενέργεια της παραγώγου του. Επειδή όμως το σήμα $y(t)$ προκύπτει από τη συνέλιξη του αρχικού σήματος $x(t)$ με το φίλτρο $h(t)$, αλλά και λόγω της δυνατότητας αντιμετάθεσης της συνέλιξης με την παραγώγιση, μπορούμε να ερμηνεύσουμε τη μέση Teager ενέργεια είτε ως τη μέση τυπική ενέργεια του φιλτραρισμένου σήματος $x(t)$ με το φίλτρο $\dot{h}(t)$ ή ως τη μέση τυπική ενέργεια του φιλτραρισμένου σήματος $\dot{x}(t)$ με το φίλτρο $h(t)$,

$$E\left\{\Psi\left[x(t) * h(t)\right]\right\} = 2E\left\{\left|x(t) * \dot{h}(t)\right|^2\right\} = 2E\left\{\left|\dot{x}(t) * h(t)\right|^2\right\}\ . \tag{5.24}$$

Με αυτό τον τρόπο υπολογισμού μειώνεται σημαντικά η μεγάλη υπολογιστική πολυπλοκότητα, αφού αντί για τρεις συνελίξεις που θα χρειαζόμασταν για ένα σήμα, $y(t) = x(t) * h(t)$, με βάση τον ορισμό της Εξ. (5.8) απαιτείται πλέον ο υπολογισμός μόνο μιας. Μάλιστα, για την αντιμετώπιση του συγκεκριμένου προβλήματος έχει προταθεί στο [EKM05] μια μέθοδος υπολογισμού της Teager ενέργειας στο συχνοτικό πεδίο στο οποίο οι συνελίξεις μετατρέπονται σε πολλαπλασιασμούς. Ένα επιπλέον πλεονέκτημα που παρουσιάζει η σύνδεση της μέσης Teager με τη μέση τυπική ενέργεια είναι η δημιουργία ενός κριτηρίου για την επιλογή της κατάλληλης ενεργειακής μέτρησης, Teager ή τυπικής ενέργεια, ανάλογα με το πρόβλημα που αντιμετωπίζουμε, όπως θα δείξουμε και στην Ενοτ. 5.2.4. Σε αυτό το σημείο θα πρέπει να αναφέρουμε ότι ανάλογες σχέσεις ισοδυναμίας ισχύουν και στην περίπτωση που κανείς θεωρήσει σήματα μεγαλύτερων διαστάσεων, όπως, για παράδειγμα, εικόνες ή βίντεο.

5.2.2 Ερμηνεία στο Συχνοτικό Πεδίο

Ιδιαίτερο ενδιαφέρον παρουσιάζει και η σύγκριση των δύο ενεργειακών μετρήσεων στο συχνοτικό πεδίο, καθώς θα μας φανεί αρκετά χρήσιμη στην Ενοτ. 5.2.4, όπου θα εξετάσουμε τη συμπεριφορά τους παρουσία θορύβου.

Η Εξ. (5.12) μπορεί να εκφρασθεί στο πεδίο της συχνότητας χρησιμοποιώντας τη σχέση που συνδέει την αυτοσυσχέτιση του σήματος στο πεδίο του χρόνου με το φάσμα ισχύος του στο πεδίο Fourier [Pap91]. Έτσι, θα έχουμε:

$$E\left\{|y(t)|^2\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_y(\omega)\,d\omega\,, \tag{5.25}$$

όπου $\Phi_y(\omega)$ το φάσμα ισχύος του σήματος $y(t)$. Από τη σχέση (5.16) η Εξ. (5.25) γράφεται ως

$$E\left\{|y(t)|^2\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty}|H(\omega)|^2\,\Phi_x(\omega)\,d\omega\,, \tag{5.26}$$

όπου $|H(\omega)|$ η απόκριση πλάτους του φίλτρου $h(t)$ και $\Phi_x(\omega)$ το φάσμα ισχύος του αρχικού σήματος $x(t)$. Αντιστοίχως, με βάση την Εξ. (5.22) η μέση Teager ενέργεια θα ισούται με

$$E\left\{\Psi\left[y(t)\right]\right\} = \frac{1}{\pi}\int_{-\infty}^{\infty}\Phi_{\dot{y}}(\omega)\,d\omega\,, \tag{5.27}$$

όπου $\Phi_{\dot{y}}(\omega)$ το φάσμα ισχύος της παραγώγου του σήματος $y(t)$. Με βάση την Εξ. (5.18) και τη σχέση

$$\frac{d^n}{dt^n}f(t) \overset{\mathcal{F}}{\longleftrightarrow} (j\omega)^n F(\omega)\,, \tag{5.28}$$

η οποία συνδέει το μετασχηματισμό Fourier ενός σήματος με τις παραγώγους του, η μέση Teager ενέργεια αναλύεται περαιτέρω σε

$$E\left\{\Psi\left[y(t)\right]\right\} = \frac{1}{\pi}\int_{-\infty}^{\infty}\omega^2\,|H(\omega)|^2\,\Phi_x(\omega)\,d\omega\,. \tag{5.29}$$

Παρατηρούμε, λοιπόν, ότι ο όρος ω^2 στο ολοκλήρωμα της (5.27) είναι και ο μόνος που διαφοροποιεί τις δύο ενεργειακές μετρήσεις. Τον επιπλέον αυτό όρο μπορούμε να τον αποδώσουμε είτε στο φίλτρο $h(t)$ είτε στο σήμα $x(t)$, προκειμένου να καταλήξουμε στην ίδια ερμηνεία της μέσης Teager ενέργειας με αυτήν που κάναμε στην Ενοτ. 5.2.1. Δηλαδή, η μέση Teager ενέργεια μπορεί να ιδωθεί ως η μέση τυπική ενέργεια του φιλτραρισμένου σήματος $x(t)$ με το φίλτρο $\dot{h}(t)$ ή ως η μέση τυπική ενέργεια του φιλτραρισμένου σήματος $\dot{x}(t)$ με το φίλτρο $h(t)$.

5.2.3 Επέκταση Ενεργειακού Τελεστή για Μιγαδικά Σήματα

Η έννοια του Teager ενεργειακού τελεστή επεκτάθηκε και στην περίπτωση μιγαδικών σημάτων με βάση ένα γενικευμένο τελεστή για διανυσματικά σήματα [MB95]. Ο μιγαδικός Teager τελεστής C για ένα μιγαδικό σήμα $f(t) = f_r(t) + jf_i(t)$ ορίζεται ως

$$C\left[f(t)\right] = \left|\dot{f}(t)\right|^2 - \mathrm{Re}\left[f^*(t)\ddot{f}(t)\right]\,, \tag{5.30}$$

όπου το $()^*$ δηλώνει το μιγαδικό συζυγή του ορίσματος και $f_r = \mathrm{Re}\,(f)$ και $f_i = \mathrm{Im}\,(f)$. Εύκολα κανείς διαπιστώνει ότι

$$C\,(f) = \Psi\,(f_r) + \Psi\,(f_i) = \Psi\left[\mathrm{Re}\,(f)\right] + \Psi\left[\mathrm{Im}\,(f)\right]\,. \tag{5.31}$$

Δηλαδή, η εφαρμογή του ενεργειακού τελεστή σε μιγαδικά σήματα καταλήγει στην ανεξάρτητη ανάλυση του πραγματικού και του φανταστικού μέρους. Η ιδιότητα αυτή έρχεται σε πλήρη συμφωνία με τον υπολογισμό της τυπικής ενέργειας για μιγαδικά σήματα, όπου

$$|f|^2 = |f_r|^2 + |f_i|^2 \ . \tag{5.32}$$

Με βάση τα παραπάνω μπορούμε άμεσα να επεκτείνουμε τα αποτελέσματα της ανάλυσης των Ενοτ. 5.2.1 και 5.2.2 στη γενικότερη περίπτωση κατά την οποία το φίλτρο $h(t)$ είναι μιγαδικό, $h(t) = h_r(t) + jh_i(t)$. Σε μια τέτοια περίπτωση το παραγόμενο σήμα $y(t)$ θα είναι επίσης μιγαδικής μορφής και θα έχουμε:

$$\begin{aligned} E\left\{C\left[y(t)\right]\right\} =& E\left\{\Psi\left[y_r(t)\right]\right\} + E\left\{\Psi\left[y_i(t)\right]\right\} \\ =& 2\left(E\left\{|y_r(t)|^2\right\} + E\left\{|y_i(t)|^2\right\}\right) = 2E\left\{|y(t)|^2\right\} \ , \end{aligned} \tag{5.33}$$

όπου $y_r(t) = y(t) * h_r(t)$ και $y_i(t) = y(t) * h_i(t)$.

5.2.4 Ανάλυση Ενεργειακών Μετρήσεων σε Θόρυβο

Στην ενότητα αυτή παρουσιάζονται κάποια θεωρητικά αποτελέσματα σχετικά με τη συμπεριφορά των δύο υπό εξέταση ενεργειακών τελεστών παρουσία θορύβου. Η συμπεριφορά του Teager τελεστή σε θόρυβο για μια ειδική κατηγορία σημάτων, σήματα που χαρακτηρίζονται από διαμορφώσεις στο πλάτος και τη συχνότητα (AM-FM), μελετάται στο [BQ93]. Μια τέτοια ανάλυση είναι αρκετά σημαντική για την κατανόηση των ιδιοτήτων των δύο ενεργειακών τελεστών και θα μας επιτρέψει την κατάλληλη επιλογή ενεργειακής μέτρησης ανάλογα με τις επικρατούσες συνθήκες θορύβου.

Έστω ότι διαθέτουμε ένα σήμα εισόδου $x(t)$ το οποίο αποτελείται από το σήμα ενδιαφέροντος $s(t)$ και μια προσθετική συνιστώσα θορύβου $n(t)$. Το σήμα εισόδου φιλτράρεται από ένα ΓΧΑ φίλτρο $g(t)$ και στην έξοδό του λαμβάνουμε το σήμα $y(t)$:

$$y(t) = x(t) * g(t) = (s(t) + n(t)) * g(t) = s_g(t) + n_g(t) \, , \tag{5.34}$$

όπου $s_g(t)$ και $n_g(t)$ είναι οι φιλτραρισμένες εκδοχές των σημάτων $s(t)$ και $n(t)$, αντιστοίχως. Στη συνέχεια θα υπολογίσουμε και θα συγκρίνουμε τη μέση τυπική και τη μέση Teager ενέργεια του σήματος $y(t)$.

$$\begin{aligned} E\left\{|y(t)|^2\right\} =& E\left\{\left(s_g(t) + n_g(t)\right)\left(s_g(t) + n_g(t)\right)^*\right\} \\ =& E\left\{s_g(t)s_g^*(t) + s_g(t)n_g^*(t) + n_g(t)s_g^*(t) + n_g(t)n_g^*(t)\right\} \\ =& R_{s_g}(0) + R_{(s,n)_g}(0) + R_{(n,s)_g}(0) + R_{n_g}(0) \ . \end{aligned} \tag{5.35}$$

Ακολουθώντας την ίδια διαδικασία με αυτή της Ενοτ. 5.2.1 μπορούμε να δείξουμε ότι:

$$\begin{aligned} R_{s_g}(\tau) + R_{(s,n)_g}(\tau) + R_{(n,s)_g}(\tau) + R_{n_g}(\tau) = \\ g^*(-\tau) * g(\tau) * \left(R_s(\tau) + R_{(s,n)}(\tau) + R_{(n,s)}(\tau) + R_n(\tau)\right) \end{aligned} \tag{5.36}$$

και, επομένως, η μέση τυπική ενέργεια κάνοντας χρήση του θεωρήματος του Parseval θα ισούται με:

$$\begin{aligned} E\left\{|y(t)|^2\right\} =& \frac{1}{2\pi}\int_{-\infty}^{\infty} |G(\omega)|^2 \left(\Phi_s(\omega) + \Phi_n(\omega) + \Phi_{(s,n)}(\omega) + \Phi_{(n,s)}(\omega)\right) \, d\omega \\ =& \frac{1}{2\pi}\int_{-\infty}^{\infty} |G(\omega)|^2 \, \Phi_s(\omega) \, d\omega + \underbrace{\frac{1}{2\pi}\int_{-\infty}^{\infty} |G(\omega)|^2 \, \Phi_{\text{noise}}(\omega) \, d\omega}_{\text{όρος λάθους}} . \end{aligned} \tag{5.37}$$

Η ποσότητα Φ_{noise} αντιστοιχεί στο άθροισμα του φάσματος ισχύος των συνιστωσών που εμφανίζονται λόγω της παρουσίας θορύβου. Ιδανικά κανείς θα επιθυμούσε την πλήρη απαλοιφή αυτού του όρου, ώστε η τελική μέτρηση ενέργειας που λαμβάνουμε να οφείλεται μόνο στην παρουσία του σήματος ενδιαφέροντος, $s(t)$. Προκειμένου να βρούμε μια αντίστοιχη έκφραση της μέσης Teager ενέργειας, θα αξιοποιήσουμε το γεγονός ότι η μέση Teager ενέργεια ισούται με δύο φορές τη μέση τυπική ενέργεια της παραγώγου του σήματος $y(t)$. Επομένως, θα έχουμε:

$$E\left\{\Psi\left[y(t)\right]\right\} = \frac{1}{\pi}\int_{-\infty}^{\infty}\omega^2\left|G(\omega)\right|^2\Phi_s(\omega)\,d\omega + \underbrace{\frac{1}{\pi}\int_{-\infty}^{\infty}\omega^2\left|G(\omega)\right|^2\Phi_{\text{noise}}(\omega)\,d\omega}_{\text{όρος λάθους}}. \tag{5.38}$$

Όπως και στην περίπτωση της τυπικής ενέργειας, έτσι και για την Teager ενέργεια στη μέτρηση που λαμβάνει κανείς εμφανίζονται επιπλέον όροι που οφείλονται στην παρουσία θορύβου. Το ενδιαφέρον μας, ωστόσο, επικεντρώνεται στο ποια από τις δύο μετρήσεις είναι πιο εύρωστη στην επίδραση των συνιστωσών του θορύβου. Για να οδηγηθούμε σε κάποιο συμπέρασμα, θα πρέπει να λάβουμε υπόψιν μας τόσο τα φασματικά χαρακτηριστικά του σήματος ενδιαφέροντος όσο και του θορύβου. Αν υποθέσουμε ότι οι συχνοτικές συνιστώσες του σήματος $s(t)$ καλύπτουν όλο το εύρος ζώνης, τότε από την Εξ. (5.38) παρατηρούμε ότι, στην περίπτωση που το φασματικό περιεχόμενο του θορύβου είναι ισχυρότερο στη χαμηλή ζώνη συχνοτήτων ($1/f^n$ θόρυβος), η μέση Teager ενέργεια είναι η καταλληλότερη μέτρηση. Αυτό συμβαίνει, γιατί η συγκεκριμένη μέτρηση θα επηρεάζεται λιγότερο από το θόρυβο εξαιτίας του επιπλέον όρου ω^2 ο οποίος ενισχύει το φάσμα του σήματος, ενώ, παράλληλα, δεν ενισχύει στον ίδιο βαθμό τις φασματικές συνιστώσες του θορύβου. Αντίθετα, όταν το φασματικό περιεχόμενο του θορύβου είναι ενισχυμένο στην υψηλή ζώνη συχνοτήτων (f^n θόρυβος), η ενεργειακή μέτρηση θα επηρεάζεται σε πολύ σημαντικό βαθμό από το θόρυβο· επομένως, σε μια τέτοια περίπτωση καταλληλότερη είναι η μέτρηση της μέσης τυπικής ενέργειας. Τέλος, στην περίπτωση λευκού θορύβου, στην οποία οι φασματικές συνιστώσες έχουν όλες το ίδιο πλάτος και καλύπτουν ολόκληρο το εύρος ζώνης, θα πρέπει να εξετάσουμε το φασματικό περιεχόμενο του σήματος ενδιαφέροντος για να επιλέξουμε την πιο αξιόπιστη μέτρηση. Θα πρέπει να σημειωθεί ότι η παραπάνω θεωρητική ανάλυση και τα συμπεράσματα που προκύπτουν έρχονται σε συμφωνία και ερμηνεύουν τα πειραματικά αποτελέσματα που παρουσιάζονται στο [DPM09].

5.2.5 Επέκταση Αποτελεσμάτων σε Σήματα Διακριτού Χρόνου

Στις Ενοτ. 5.2.1-5.2.4 η θεωρητική ανάλυση που παρουσιάσαμε για τις ζωνοπερατές ενεργειακές μετρήσεις αφορούσε σε σήματα συνεχούς χρόνου. Ωστόσο, επειδή το ενδιαφέρον μας επικεντρώνεται ως επί το πλείστον σε σήματα διακριτού χρόνου, στην παρούσα ενότητα θα συζητήσουμε τη δυνατότητα επέκτασης των παραπάνω αποτελεσμάτων στην περίπτωση εκείνη που τα υπό μελέτη σήματα είναι διακριτού χρόνου.

Το διακριτό αντίστοιχο του Teager ενεργειακού τελεστή είναι ο τελεστής Ψ_d [MQK91, MKQ93b, MP95] ο οποίος, όταν εφαρμοστεί σε ένα σήμα $x(n)$, μας δίνει

$$\Psi_d\left[x(n)\right] = x(n)^2 - x(n-1)x(n+1)\,. \tag{5.39}$$

Στο [MKQ93b] οι συγγραφείς έδειξαν ότι χρησιμοποιώντας συγκεκριμένους συνδυασμούς προσεγγίσεων της χρονικής παραγώγισης από χρονικές διαφορές, τότε κανείς μπορεί από το συνεχή τελεστή Ψ να οδηγηθεί σε μια έκφραση στενά συνδεδεμένη με το διακριτό τελεστή Ψ_d.

Η επέκταση των αποτελεσμάτων, τα οποία συζητήθηκαν στις προηγούμενες ενότητες, για την περίπτωση σημάτων διακριτού χρόνου δεν είναι άμεση όταν κανείς χρησιμοποιήσει το διακριτό τελεστή. Ωστόσο, το φιλτράρισμα που προηγείται της μέτρησης δίνει την δυνατότητα αντί για το

διακριτό ενεργειακό τελεστή Ψ_d κανείς να χρησιμοποιήσει το συνεχή τελεστή Ψ [DM06, KEM09] και, στη συνέχεια, να δειγματοληπτήσει το σήμα που εμφανίζεται στην έξοδό του κατά τις χρονικές στιγμές $t = nT_s$

$$\Psi\left[x(t)\right] = \dot{x}(t)^2 - x(t) * \ddot{x}(t)|_{t=nTs}\,, \tag{5.40}$$

όπου T_s η περίοδος δειγματοληψίας. Η διαδικασία αυτή είναι δυνατή λόγω της αντιμετάθεσης της παραγώγισης με την συνέλιξη. Έτσι, αν $x(t) = s(t) * h(t)$, όπου $s(t)$ αποτελεί την επέκταση του σήματος $s(n)$ στο συνεχή χρόνο και $h(t)$ είναι ένα συνεχές φίλτρο, τότε η Teager ενέργεια του $x(t)$ θα δίνεται από την ακόλουθη σχέση:

$$\begin{aligned}\Psi\left[s(t) * h(t)\right] &= \left[\frac{d}{dt}\left(s(t) * h(t)\right)\right]^2 - \left(s(t) * h(t)\right)\left[\frac{d^2}{dt^2}\left(s(t) * h(t)\right)\right] \\ &= \left[s(t) * \frac{dh(t)}{dt}\right]^2 - \left(s(t) * h(t)\right)\left[s(t) * \frac{d^2h(t)}{dt^2}\right]\,.\end{aligned} \tag{5.41}$$

Από τη σχέση (5.41) προκύπτει ότι η δειγματοληψία της εξόδου του τελεστή συνεχούς χρόνου, Ψ, συνεπάγεται την συνέλιξη του διακριτού σήματος $s(n)$ με τις δειγματοληπτημένες παραγώγους του φίλτρου $h(t)$

$$\Psi\left[s(t) * h(t)\right]|_{t=nT_s} = \left[s(n) * \dot{h}(t)|_{t=nT_s}\right]^2 - s(n) * h(t)|_{t=nT_s}\left[s(n) * \ddot{h}(t)|_{t=nT_s}\right]\,. \tag{5.42}$$

Ο συγκεκριμένος υπολογισμός της ενέργειας εμφανίζει σημαντικά πλεονεκτήματα σε σχέση με το διακριτό υπολογισμό της δίνοντας πιο ακριβή αποτελέσματα [DM06, KEM09], αφού αποφεύγει την προσέγγιση των παραγώγων με χρονικές διαφορές κατά την οποία εισάγονται σφάλματα μοντελοποίησης. Με βάση τα παραπάνω γίνεται σαφές ότι, εφόσον κανείς χρησιμοποιεί το συγκεκριμένο τρόπο υπολογισμού της Teager ενέργειας σε διακριτά σήματα, τα θεωρητικά αποτελέσματα που λάβαμε στις προηγούμενες ενότητες για το συνεχή χρόνο επεκτείνονται και σε αυτή την περίπτωση. Αυτό συμβαίνει λόγω του ότι ο υπολογισμός της μέσης τυπικής ενέργειας δεν παρουσιάζει ουσιαστικές διαφορές μεταξύ συνεχούς και διακριτού χρόνου.

5.3 Πολυαισθητηριακή-Πολυφασματική Αποδιαμόρφωση Ακουστικών Σημάτων

Όπως ήδη αναφέραμε στην Ενοτ. 5.1, το πρώτο στάδιο για την αυτόματη αναγνώριση στηρίζεται στην εξαγωγή διανυσμάτων χαρακτηριστικών από τα παρατηρούμενα ακουστικά σήματα. Τα χαρακτηριστικά αυτά επιλέγονται με τέτοιο τρόπο, ώστε να κωδικοποιούν αποδοτικά την ακουστική πληροφορία των σημάτων εισόδου σε όσο το δυνατόν λιγότερους συντελεστές. Ωστόσο, ένα από τα σημαντικά προβλήματα που προκύπτουν είναι η επιλογή κατάλληλων χαρακτηριστικών τα οποία θα παρουσιάζουν "αντοχή" στην επίδραση του θορύβου, προκειμένου η αυτόματη αναγνώριση να είναι εφικτή ακόμα και σε δυσμενείς συνθήκες.

Στην ενότητα αυτή προτείνουμε ένα νέο σχήμα εξαγωγής χαρακτηριστικών το οποίο βασίζεται σε μια πολυαισθητηριακή-πολυφασματική αποδιαμόρφωση των σημάτων εισόδου. Το προτεινόμενο σχήμα εξαγωγής εκτός από την πληροφορία για το φασματικό περιεχόμενο των σημάτων εισόδου εκμεταλλεύεται και τη χωρική πληροφορία που παρέχεται από τη συστοιχία των αισθητήρων. Το γεγονός αυτό έχει ως αποτέλεσμα το προτεινόμενο σχήμα να εμφανίζει τη δυνατότητα, όπως θα δείξουμε και πειραματικά, σημαντικής βελτίωσης των ποσοστών αναγνώρισης σε συνθήκες θορύβου.

Στις ενότητες που ακολουθούν παρουσιάζουμε το σχήμα πολυαισθητηριακής-πολυφασματικής αποδιαμόρφωσης, μελετούμε την επίδραση του θορύβου και καταλήγουμε στο προτεινόμενο πολυαισθητηριακό σχήμα εξαγωγής εύρωστων χαρακτηριστικών.

5.3.1 Πολυαισθητηριακή - Πολυφασματική Αποδιαμόρφωση και Ενεργειακές Μετρήσεις

Θεωρούμε μια συστοιχία με M αισθητήρες διατεταγμένους σε τυχαία διάταξη η οποία είναι τοποθετημένη σε περιβάλλον με έντονη παρουσία θορύβου. Η συστοιχία αυτή χρησιμοποιείται για τη σύλληψη της κυματομορφής ενός σήματος πηγής $s(t)$. Το παρατηρούμενο σήμα $y'_m(t)$, $m = 0, \ldots, M-1$ στον mοστό αισθητήρα αντιστοιχεί σε μια γραμμικά φιλτραρισμένη έκδοση του σήματος πηγής $s(t)$ με την επιπλέον προσθήκη μιας συνιστώσας θορύβου. Το φίλτρο με το οποίο συνελίσσεται το σήμα πηγής μοντελοποιεί την κρουστική απόκριση του μονοπατιού που αυτό ακολουθεί κατά τη μετάδοσή του προς τους αισθητήρες της συστοιχίας. Εξαιτίας των διαφορετικών μονοπατιών που ακολουθεί το μεταδιδόμενο σήμα, προκειμένου να φτάσει στους αισθητήρες, τα σήματα που τελικά κανείς παρατηρεί παρουσιάζουν διαφορετικές χρονικές καθυστερήσεις και παραμορφώσεις. Για την αντιμετώπιση αυτού του φαινομένου τα λαμβανόμενα από τους αισθητήρες σήματα διοχετεύονται σε ένα υποσύστημα χρονικής ευθυγράμμισης. Οι έξοδοι αυτού του υποσυστήματος, όταν δεν υπάρχει αντήχηση στο περιβάλλον, αντιστοιχούν στις κανονικοποιημένες και χρονικά ευθυγραμμισμένες εισόδους των αισθητήρων και μπορούν να αναπαρασταθούν ως:

$$y_m(t) = s(t) + v_m(t),\ m = 0, \ldots, M-1\,. \tag{5.43}$$

Η προσθετική συνιστώσα θορύβου, $v_m(t)$, θεωρούμε ότι είναι μια WSS στοχαστική διαδικασία με συνάρτηση αυτοσυσχέτισης $R_{v_m}(\tau)$ και πυκνότητα φάσματος ισχύος $\Phi_{v_m}(\omega)$. Επιπλέον, θεωρούμε ότι οι συνιστώσες θορύβου που επενεργούν σε δύο διαφορετικούς αισθητήρες της συστοιχίας $m,\ k$ είναι από κοινού WSS στοχαστικές διαδικασίες με συνάρτηση ετεροσυσχέτισης $R_{v_{(mk)}}(\tau)$ και πυκνότητα ετεροφάσματος ισχύος $\Phi_{v_{(mk)}}(\omega)$.

Όπως και στα μονοκαναλικά (ένας αισθητήρας) σχήματα αποδιαμόρφωσης, έτσι και στο υπό μελέτη πολυαισθητηριακό σχήμα κάθε κανονικοποιημένο σήμα αναλύεται σε K σήματα στενού εύρους ζώνης μέσω μιας συστοιχίας φίλτρων, όπως αυτές που αναφέραμε στην Ενοτ.5.1.1. Στη συνέχεια της ανάλυσης θα συμβολίζουμε με y_{mj} το σήμα το οποίο ελήφθη από το mοστό αισθητήρα και φιλτραρίστηκε από το jοστό φίλτρο της συστοιχίας. Η συγκεκριμένη διαδικασία αποδιαμόρφωσης μπορεί να εκφρασθεί με μαθηματικό τρόπο ως:

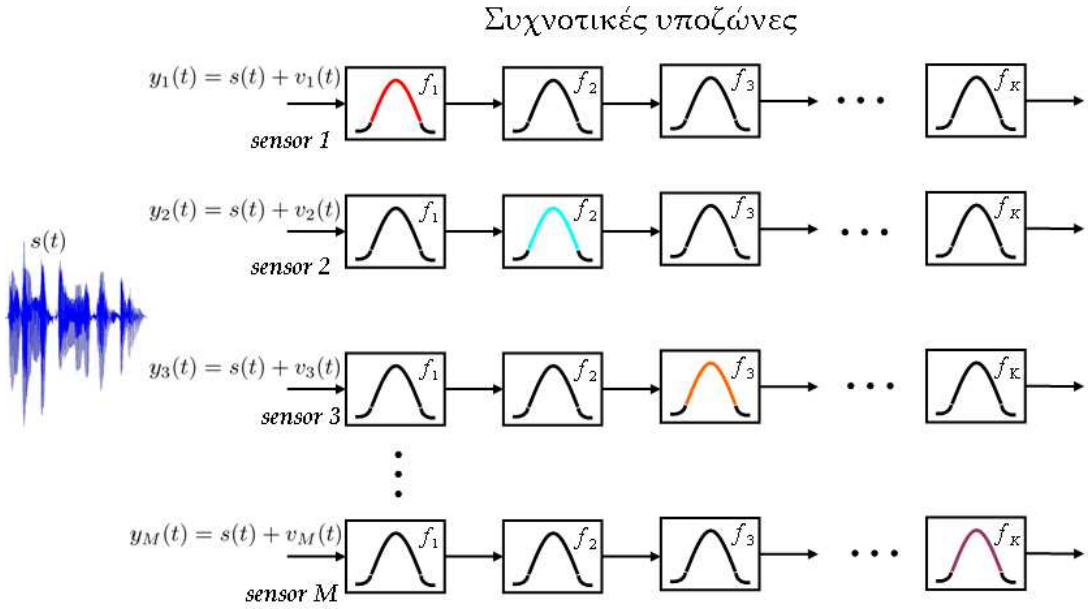
$$y_{mj}(t) = y_m(t) * g_j(t),\ j = 0, \ldots, K-1, \tag{5.44}$$

ενώ αποτυπώνεται και σχηματικά στο Σχ. 5.5.

Στόχος μας τώρα είναι να μελετήσουμε πως επιδρά ο θόρυβος στις πολυαισθητηριακές ενεργειακές μετρήσεις οι οποίες χρησιμοποιούνται σε επόμενο στάδιο για την εξαγωγή διανυσματικών χαρακτηριστικών. Στην ακόλουθη ανάλυση θα χρησιμοποιήσουμε ως ενεργειακή μέτρηση τη μέση Teager ενέργεια. Ωστόσο, λόγω της σχέσης (5.23), η οποία συνδέει τη μέση τυπική και τη μέση Teager ενέργεια, τα αποτελέσματα που προκύπτουν γενικεύονται άμεσα, όπως θα δείξουμε και στην περίπτωση στην οποία κανείς επιλέξει ως μέτρηση την τυπική ενέργεια.

Η cross-Teager ενέργεια [Kai93] δύο ζωνοπερατών σημάτων που συνέλαβε ένα συγκεκριμένο ζεύγος αισθητήρων $(m\,, k)$ και τα οποία έχουν φιλτραριστεί από το jοστό φίλτρο της συστοιχίας είναι ίση με:

$$\Psi_c\left[y_{mj}(t), y_{kj}(t)\right] = \dot{y}_{mj}(t)\dot{y}_{kj}(t) - y_{mj}(t)\ddot{y}_{kj}(t)\,. \tag{5.45}$$



Σχήμα 5.5: Ανάλυση των πολλαπλών σημάτων εισόδου στη διάταξη των αισθητήρων σε φασματικές συνιστώσες. Οι φασματικές συνιστώσες προκύπτουν από το φιλτράρισμα των εισόδων με μια συστοιχία φίλτρων.

Μπορούμε να αναλύσουμε περαιτέρω αυτή την έκφραση χρησιμοποιώντας τις εξισώσεις (5.43) και (5.44) και να λάβουμε τελικά:

$$\Psi_c\left[y_{mj}(t), y_{kj}(t)\right] = \Psi\left[s_j(t)\right] + \Psi_c\left[v_{mj}(t), v_{kj}(t)\right] + \Psi_c\left[s_j(t), v_{kj}(t)\right] + \Psi_c\left[v_{mj}(t), s_j(t)\right] . \quad (5.46)$$

Όπως παρατηρούμε, η cross-Teager ενέργεια αποτελείται από την Teager ενέργεια του ζωνοπερατού σήματος ενδιαφέροντος $s_j(t)$ και από επιπλέον τρεις όρους, στο δεξιό μέλος της Εξ. (5.46), οι οποίοι οφείλονται στην παρουσία θορύβου και, επομένως, είναι ανεπιθύμητοι, αφού εισάγουν σφάλμα στις μετρήσεις μας. Για να υπολογίσουμε τη μέση τιμή της Εξ. (5.46), θα εκφράσουμε αρχικά τη μέση τιμή της cross-Teager ενέργειας σε μια μορφή ανάλογη με αυτή της Εξ. (5.22). Ακολουθώντας την ίδια διαδικασία με την Ενοτ. 5.2.1 μπορεί κανείς να δείξει ότι η μέση τιμή της cross-Teager ενέργειας δύο φιλτραρισμένων σημάτων $x(t)$, $z(t)$ είναι ίση με:

$$E\left\{\Psi_c\left[x(t), z(t)\right]\right\} = 2R_{\dot{x}\dot{z}}\left(0\right) , \quad (5.47)$$

όπου $R_{\dot{x}\dot{z}}\left(\tau\right)$ είναι η ετεροσυσχέτιση των παραγώγων των σημάτων $x(t)$, $z(t)$. Έτσι, λοιπόν, από τις Εξ. (5.22), (5.46) και (5.47) προκύπτει ότι

$$\begin{aligned} E\left\{\Psi_c\left[y_{mj}(t), y_{kj}(t)\right]\right\} =& 2R_{\dot{y}_{(m,k)j}}\left(0\right) \\ =& 2R_{\dot{s}_j}\left(0\right) + 2\left(R_{\dot{v}_{(m,k)j}}\left(0\right) + R_{\dot{s}_j\dot{v}_{kj}}\left(0\right) + R_{\dot{v}_{mj}\dot{s}_j}\left(0\right)\right) \end{aligned} \quad (5.48)$$

η οποία είναι ισοδύναμη στο συχνοτικό πεδίο με

$$\begin{aligned} E\left\{\Psi_c\left[y_{mj}(t), y_{kj}(t)\right]\right\} &= \frac{1}{\pi}\int_{-\infty}^{\infty} \Phi_{\dot{y}_{(m,k)j}}(\omega)\, d\omega \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \Phi_{\dot{s}_j}(\omega)\, d\omega + \frac{1}{\pi}\int_{-\infty}^{\infty} \left(\Phi_{\dot{v}_{(m,k)j}}(\omega) + \Phi_{\dot{s}_j\dot{v}_{mj}}(\omega) + \Phi_{\dot{v}_{kj}\dot{s}_j}(\omega)\right) d\omega , \end{aligned} \quad (5.49)$$

με το $\Phi_{\dot{y}_{(m,k)j}}(\omega)$ να συμβολίζει το ετερο- φάσμα ισχύος των σημάτων $\dot{y}_{mj}(t)$ και $\dot{y}_{kj}(t)$. Ο δεύτερος όρος στο δεξί μέλος της Εξ. (5.49) αναπαριστά τη συγκέντρωση ισχύος του θορύβου στη συχνοτική ζώνη διάβασης του φίλτρου $g_j(t)$ και τον συμβολίζουμε με $\Gamma_{(m,k)j}$,

$$\begin{aligned}\Gamma_{(m,k)j} =& \frac{1}{\pi}\int_{-\infty}^{\infty}\left(\Phi_{\dot{v}_{(m,k)j}}(\omega)+\Phi_{\dot{s}_j\dot{v}_{mj}}(\omega)+\Phi_{\dot{v}_{kj}\dot{s}_j}(\omega)\right)d\omega \\ =& \frac{1}{\pi}\int_{-\infty}^{\infty}\omega^2\left|G(\omega)\right|^2\left(\Phi_{v_{(m,k)j}}(\omega)+\Phi_{s_jv_{mj}}(\omega)+\Phi_{v_{kj}s_j}(\omega)\right)d\omega\,.\end{aligned} \tag{5.50}$$

Λαμβάνοντας υπόψιν μας τις Εξ. (5.27), (5.49) και (5.50) μπορούμε να γράψουμε τελικά την cross-Teager ενέργεια των δύο σημάτων εισόδου στην απλούστερη μορφή:

$$E\left\{\Psi_c\left[y_{mj}(t),y_{kj}(t)\right]\right\}=E\left\{\Psi\left[s_j(t)\right]\right\}+\Gamma_{(m,k)j}\,. \tag{5.51}$$

Σε αυτό το σημείο θα πρέπει να αναφέρουμε ότι στην ειδική περίπτωση κατά την οποία υπολογίζουμε την cross-Teager ενέργεια στον ίδιο αισθητήρα, δηλαδή $m=k$, η προκύπτουσα τιμή θα ισούται με την απλή Teager ενέργεια

$$E\left\{\Psi\left[y_{mj}(t)\right]\right\}=E\left\{\Psi\left[s_j(t)\right]\right\}+\Gamma_{mj}\,. \tag{5.52}$$

Αν λάβουμε υπόψιν μας την ανισότητα [Pap91],

$$R_{xy}^2(\tau)\le R_x(0)\,R_y(0) \tag{5.53}$$

και θέσουμε $x=\dot{y}_{mj}$, $y=\dot{y}_{kj}$, τότε αποδεικνύεται ότι ισχύει (δες [Pap91] σελ. 331-332 για την απόδειξη)

$$\left|\int_{-\infty}^{\infty}\Phi_{\dot{y}_{(m,k)j}}(\omega)\,d\omega\right|^2\le\int_{-\infty}^{\infty}\Phi_{\dot{y}_{mj}}(\omega)\,d\omega\int_{-\infty}^{\infty}\Phi_{\dot{y}_{kj}}(\omega)\,d\omega\,. \tag{5.54}$$

Από την ανισότητα (5.54) προκύπτει άμεσα ότι:

$$\left|E\left\{\Psi_c\left[y_{mj}(t),y_{kj}(t)\right]\right\}\right|^2\le E\left\{\Psi\left[y_{mj}(t)\right]\right\}E\left\{\Psi\left[y_{kj}(t)\right]\right\}\,, \tag{5.55}$$

δηλαδή, το τετράγωνο της απόλυτης μέσης τιμής της cross-Teager ενέργειας μεταξύ δύο αισθητήρων (m,k) είναι μικρότερο ή ίσο με το γινόμενο της μέσης Teager ενέργειας που κανείς μετράει στον κάθε έναν αισθητήρα ξεχωριστά. Αν κανείς αντί για τη μέση cross-Teager ενέργεια υπολόγιζε τη μέση cross-τυπική ενέργεια, θα λάμβανε τα ίδια αποτελέσματα με τη διαφορά ότι στη σχέση (5.49) και κατ' επέκταση και στα επόμενα αποτελέσματα αντί για το φάσμα της παραγώγου των υπό μελέτη σημάτων θα είχε το φάσμα των σημάτων αυτών πολλαπλασιαζόμενο με μια σταθερά. Έτσι, πάλι θα καταλήγαμε στην αντίστοιχη ανισότητα

$$\left|E\left\{y_{mj}(t)y_{kj}^*(t)\right\}\right|^2\le E\left\{\left|y_{mj}(t)\right|^2\right\}E\left\{\left|y_{kj}(t)\right|^2\right\}\,. \tag{5.56}$$

Οι ανισότητες που προκύπτουν στις Εξ. (5.55) και (5.56) έχουν πολύ μεγάλη σημασία και, όπως θα δείξουμε στην επόμενη ενότητα, θα αποτελέσουν το βασικό μας εργαλείο για την ανάπτυξη ενός γενικού και εύρωστου στο θόρυβο αλγορίθμου πολυαισθητηριακής εξαγωγής χαρακτηριστικών.

5.3.2 Πολυαισθητηριακή Εξαγωγή Ενεργειακών Χαρακτηριστικών

Ορμώμενοι από την ανάλυση της Ενοτ. 5.3.1 αλλά και από το γεγονός ότι τα πλέον διαδεδομένα και αποδοτικά διανύσματα χαρακτηριστικών, όπως είναι τα MFCC [DM80], PLP [Her87, Her90] και TECC [DMP05a], προκύπτουν μέσα από διαδικασίες στις οποίες βασικό στάδιο αποτελεί ο υπολογισμός της μέσης τυπικής ή Teager ενέργειας, προτείνουμε ένα πολυαισθητηριακό σχήμα εξαγωγής ενεργειακών χαρακτηριστικών. Το προτεινόμενο σχήμα συνδυάζει τα οφέλη της "πλούσιας" ακουστικής πληροφορίας, η οποία παρέχεται από τις ενεργειακές μετρήσεις, αλλά και τα οφέλη της χωρικής πληροφορίας για το ακουστικό περιβάλλον που μας παρέχει η συστοιχία των αισθητήρων.

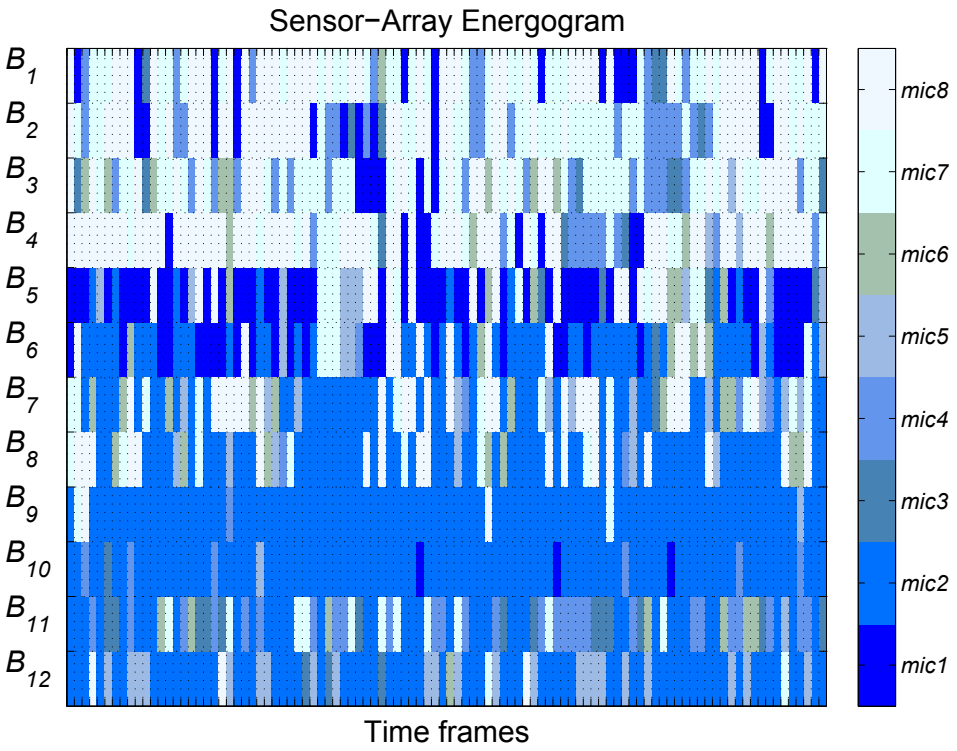
Όπως αναφέραμε και στην εισαγωγή, συστήματα αναγνώρισης τα οποία εκπαιδεύονται σε ελεγχόμενο περιβάλλον χωρίς θόρυβο επιτυγχάνουν πολύ ικανοποιητικά ποσοστά αναγνώρισης. Το γεγονός αυτό υποδεικνύει ότι τα διανυσματικά χαρακτηριστικά, τα οποία βασίζονται σε ενεργειακές μετρήσεις και χρησιμοποιούνται για την κωδικοποίηση της ακουστικής πληροφορίας, επιτυγχάνουν αυξημένη διαχωρισιμότητα. Η μειωμένη απόδοση αυτών των συστημάτων παρουσιάζεται, όταν τα σήματα εισόδου εμφανίζουν αλλοιώσεις λόγω της παρουσίας θορύβου και, συνεπώς, και οι μετρήσεις που κανείς λαμβάνει από αυτά είναι εσφαλμένες.

Στο προτεινόμενο σχήμα ο στόχος μας είναι να εξάγουμε ενεργειακές μετρήσεις που θα προσεγγίζουν τη μέτρηση την οποία θα λάμβανε κανείς, αν αντί για το σήμα $y_{mj}(t)$ είχε στη διάθεσή του το καθαρό ζωνοπερατό σήμα $s_j(t)$. Κάτι τέτοιο είναι εφικτό ελαχιστοποιώντας τον όρο Γ_{mj} (βλέπε την Εξ. (5.52)) που εκφράζει τη συγκέντρωση ισχύος του θορύβου μέσα στη ζώνη διάβασης του φίλτρου $g_j(t)$. Δεδομένου ότι, λόγω των πολλαπλών αισθητήρων που περιλαμβάνει η συστοιχία μας, έχουμε στη διάθεσή μας περισσότερα από ένα στιγμιότυπα του σήματος εισόδου $y(t)$, μια απλή ιδέα μείωσης της επίδρασης του θορύβου είναι να υπολογίσει κανείς για την jοστή υποζώνη φιλτραρίσματος τη μέση τιμή της ενέργειας (Teager ή τυπικής) σε όλα τα διαθέσιμα στιγμιότυπα. Στη συνέχεια, να επιλέξει εκείνη τη μέτρηση που έχει τη μικρότερη τιμή. Η συγκεκριμένη ενεργειακή μέτρηση είναι σαφές ότι αντιστοιχεί σε εκείνη που προσεγγίζει καλύτερα από όλες τις υπόλοιπες μετρήσεις τη μέση ενέργεια την οποία θα λαμβάναμε, αν είχαμε στη διάθεσή μας το αναλλοίωτο ζωνοπερατό σήμα $s_j(t)$. Αυτό οφείλεται στο γεγονός ότι η μέση ενέργεια του ζωνοπερατού πηγαίου σήματος $s_j(t)$ παραμένει ίδια σε όλες μας τις μετρήσεις, $E\left\{\Psi\left[y_{mj}\right]\right\}$ για $m = 0, 1, \ldots, M-1$ και $j = 0, 1, \ldots, K-1$, ενώ ο μόνος όρος που παρουσιάζει διακύμανση είναι ο όρος σφάλματος Γ_{mj} εξαιτίας της διαφορετικής ισχύος του θορύβου που επηρεάζει το ακουστικό σήμα εισόδου σε κάθε αισθητήρα της συστοιχίας. Το γεγονός αυτό επιβεβαιώνεται και στην πράξη, όπως παρουσιάζεται και στο Σχ. 5.6, όπου απεικονίζεται το 'ενεργόγραμμα' για μια διάταξη 8 αισθητήρων στην οποία τα παρατηρούμενα σήματα έχουν αλλοιωθεί από πραγματικό θόρυβο. Το ενεργόγραμμα απεικονίζει σε ποιον αισθητήρα της συστοιχίας μετράται η μικρότερη ενέργεια για ένα συγκεκριμένο χρονικό πλαίσιο του σήματος (frame) και σε μια συγκεκριμένη συχνοτική υποζώνη. Όπως παρατηρούμε, η επίδραση του θορύβου δεν είναι ομοιόμορφη σε όλες τις μετρήσεις, επομένως, μπορούμε να εκμεταλλευτούμε αυτό το γεγονός και να επιλέγουμε κάθε φορά τις πιο ακριβείς ενεργειακές μετρήσεις.

Ένας τρόπος για την περαιτέρω μείωση της επίδρασης του θορύβου στις ενεργειακές μετρήσεις μπορεί να προκύψει λαμβάνοντας υπόψιν μας τις ανισότητες (5.55) και (5.56). Για την περίπτωση υπολογισμού της μέσης Teager ενέργειας θα έχουμε από τις Εξ. (5.51), (5.52), (5.55) και μέσω απλών αντικαταστάσεων ότι

$$\left|\Gamma_{(mk)j}\right|^2 + 2E\left\{\Psi\left[s_j(t)\right]\right\}\mathrm{Re}\left[\Gamma_{(mk)j}\right] \leq \Gamma_{mj}\Gamma_{kj} + E\left\{\Psi\left[s_j(t)\right]\right\}\left(\Gamma_{mj} + \Gamma_{kj}\right) , \qquad (5.57)$$

όπου στο πρώτο μέλος της ανισότητας εμφανίζονται οι cross-όροι θορύβου από τον υπολογισμό της cross-Teager ενέργειας, ενώ στο δεύτερο μέλος οι αντίστοιχοι όροι από το γινόμενο των απλών Teager ενεργειών σε δύο αισθητήρες ξεχωριστά. Ουσιαστικά, αυτό που δηλώνει η συγκεκριμένη

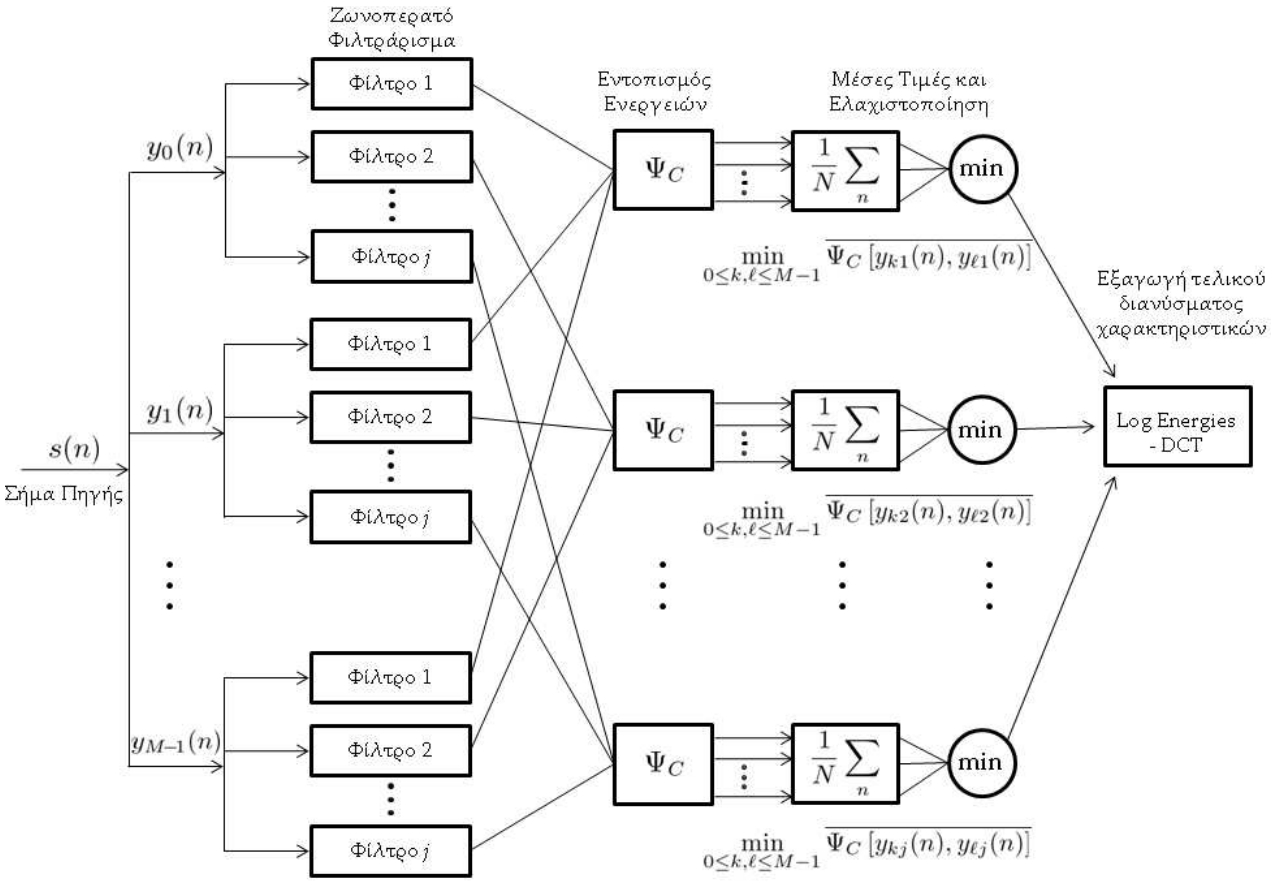


Σχήμα 5.6: Ενεργόγραμμα διάταξης 8 αισθητήρων. Το γράφημα αυτό απεικονίζει σε ποιον από τους αισθητήρες της διάταξης υπολογίζεται η πιο αξιόπιστη ενεργειακή μέτρηση (η πλησιέστερη στην πραγματική μέτρηση) ανά συχνοτική υποζώνη και χρονικό πλαίσιο.

ανισότητα είναι ότι η επίδραση του θορύβου στον υπολογισμό του τετραγώνου της μέσης cross-Teager ενέργειας είναι μικρότερη ή ίση με αυτήν που εμφανίζεται στο γινόμενο των μέσων ενεργειών τις οποίες υπολογίζουμε σε δύο αισθητήρες ξεχωριστά. Συνεπώς, υπολογίζοντας την τετραγωνική ρίζα της απόλυτης μέσης τιμής της cross-Teager ενέργειας μεταξύ δύο αισθητήρων, θα λάβουμε μια τιμή η οποία θα προσεγγίζει καλύτερα την τιμή μέτρησης, την οποία θα λαμβάναμε απευθείας από το καθαρό ζωνοπερατό σήμα $s_j(t)$, από ό,τι θα την προσέγγιζε μία τουλάχιστον τιμή των δύο άλλων μετρήσεων. Μάλιστα, στην περίπτωση που οι δύο αισθητήρες είναι αρκετά απομακρυσμένοι μεταξύ τους, δηλαδή μπορούμε να θεωρήσουμε ότι ο θόρυβος στις δύο χωρικές θέσεις είναι ασυσχέτιστος, και επιπλέον το σήμα πηγής είναι ασυσχέτιστο με το θόρυβο, τότε η μέτρηση που θα λάβουμε θα συμπίπτει με την "καθαρή" μέτρηση.

Επειδή σε μια συστοιχία M αισθητήρων έχουμε στη διάθεσή μας $\binom{M+1}{2}$ πιθανά ζεύγη [1], είναι δυνατόν για κάθε υποζώνη φιλτραρίσματος να υπολογίσουμε όλες τις cross-Teager ενέργειες και, στη συνέχεια, να διαλέξουμε αυτή με τη μικρότερη τιμή. Η εν λόγω μέτρηση, βάσει των παραπάνω παρατηρήσεων, θα προσεγγίζει καλύτερα την ενεργειακή μέτρηση του καθαρού σήματος το οποίο έχει φιλτραριστεί από το αντίστοιχο φίλτρο. Το πλεονέκτημα αυτής της προσέγγισης σε σχέση με άλλες μεθόδους είναι οτι οι μετρήσεις λαμβάνονται απευθείας από τα θορυβώδη δεδομένα χωρίς αυτά να υπόκεινται σε καμία άλλη επεξεργασία που πιθανώς θα αλλοίωνε το φασματικό τους περιεχόμενο. Το γεγονός αυτό είναι εξαιρετικής σημασίας, γιατί αντίθετα με το ανθρώπινο ακουστικό σύστημα, το οποίο δεν επηρεάζεται σε μεγάλο βαθμό από αλλοιώσεις του φάσματος, οι μηχανές αναγνώρισης είναι αρκετά ευαίσθητες και μικρές μεταβολές οδηγούν σε σημαντικές

[1] Η διάταξη των αισθητήρων σε κάθε ζεύγος δε μας ενδιαφέρει, γιατί γενικά ισχύει $|E\{\Psi_c[x(t), y(t)]\}| = |E\{\Psi_c[y(t), x(t)]\}|$ και $|E\{x(t)y^*(t)\}| = |E\{y(t)x^*(t)\}|$.



Σχήμα 5.7: Πολυαισθητηριακή εξαγωγή ενεργειακών χαρακτηριστικών. Πολυκαναλικό φιλτράρισμα σε κάθε αισθητήρα της διάταξης, εντοπισμός των cross-ενεργειών (Teager ή τυπικής ενέργειας) και επιλογή της ελάχιστης μέσης τιμής ανά συχνοτική υποζώνη. Οι επιλεγόμενες ενεργειακές μετρήσεις χρησιμοποιούνται για την εξαγωγή του τελικού διανύσματος χαρακτηριστικών που θα αξιοποιηθεί στην αυτόματη αναγνώριση. Τα παράθυρα ανάλυσης είναι μήκους N δειγμάτων.

επιδράσεις στα ποσοστά αναγνώρισης [DA08].

Με βάση τις παραπάνω παρατηρήσεις προτείνουμε ένα πολυαισθητηριακό σχήμα εξαγωγής χαρακτηριστικών το οποίο περιγράφεται στον Αλγόριθμο 1 και απεικονίζεται στο Σχ. 5.7 ως μπλοκ διάγραμμα. Το προτεινόμενο σχήμα εξαγωγής αποτελεί γενίκευση μιας προγενέστερης εργασίας μας [LMK08] και μπορεί να χρησιμοποιηθεί για την εξαγωγή χαρακτηριστικών διανυσμάτων τα οποία βασίζονται είτε στην μέση τυπική είτε στη μέση Teager ενέργεια. Παράλληλα, επειδή τελικά για όλα τα σήματα εισόδου διατηρούμε μια μόνο μέτρηση σε κάθε συχνοτική υποζώνη, το παραγόμενο διάνυσμα χαρακτηριστικών θα είναι της ίδιας μορφής με αυτό που κανείς θα λάμβανε, αν είχε στη διάθεσή του ένα μόνο αισθητήρα για τη σύλληψη των ακουστικών σημάτων. Αυτό δίνει τη δυνατότητα χρήσης μεθόδων οι οποίες έχουν προταθεί στη βιβλιογραφία για την περαιτέρω ενίσχυση των εξαγόμενων χαρακτηριστικών, όπως είναι οι τεχνικές κανονικοποίησης της μέσης τιμής (CMN - cepstral mean normalization) [Ata74], της διακύμανσης (CMV - cepstral variance normalization) και του ιστογράμματος (CMN - cepstral histogram normalization) [dlTPS+05] του cepstrum, αλλά και παραλλαγών τους, όπως η RASTA (relative spectral filtering) τεχνική [HM94]. Στόχος αυτών των τεχνικών κανονικοποίησης είναι η εφαρμογή ενός μετασχηματισμού στα παρατηρούμενα διανυσματικά χαρακτηριστικά. Ο μετασχηματισμός αυτός θα πρέπει να εξαλείφει τυχόν διακυμάνσεις που δε συνδέονται με τις ακουστικές μετεγγραφές, ενώ, παράλληλα, να μειώνει

Αλγόριθμος 1 Αλγόριθμος για την πολυαισθητηριακή εξαγωγή χαρακτηριστικών.

1. Χρησιμοποιούμε μια συστοιχία φίλτρων όμοια με αυτές που περιγράψαμε στην Ενοτ. 5.1.1, για να αναλύσουμε κάθε ένα από τα M σήματα εισόδου της συστοιχίας σε K ζωνοπερατά σήματα.
2. Εκτιμούμε τη βραχέως χρόνου μέση τιμή είτε της cross-τυπικής είτε της cross-Teager ενέργειας όλων των δυνατών (m, k) ζευγών από αισθητήρες για κάθε ένα από τα K ζωνοπερατά σήματα. Η επιλογή της κατάλληλης μέτρησης εξαρτάται από το είδος χαρακτηριστικών που θέλουμε να υπολογίσουμε. Το παράθυρο που χρησιμοποιούμε για τη βραχέως χρόνου επεξεργασία καλύπτει σήματα διάρκειας $25ms$ και μετατοπίζεται σε νέα δείγματα του σήματος $10ms$.
3. Για κάθε ζωνοπερατό πλαίσιο της συστοιχίας φίλτρων επιλέγουμε εκείνη τη μέση τιμή της ενεργειακής μέτρησης, από όλα τα $\binom{M+1}{2}$ αποτελέσματα, η οποία εμφανίζει τη μικρότερη τιμή.
4. Αναλόγως με το είδος των ενεργειακών χαρακτηριστικών που επιθυμούμε να υπολογίσουμε (MFCC, PLP, TECC), ακολουθούμε τα επόμενα βήματα της αντίστοιχης διαδικασίας εξαγωγής.

τις όποιες διαφορές υπάρχουν μεταξύ των διανυσμάτων χαρακτηριστικών τα οποία έχουν εξαχθεί υπό διαφορετικές συνθήκες. Τέτοιες διαφορές εμφανίζονται μεταξύ διανυσμάτων χαρακτηριστικών που χρησιμοποιούνται κατά την εκπαίδευση και αυτών που χρησιμοποιούνται κατά τον έλεγχο του στατιστικού συστήματος.

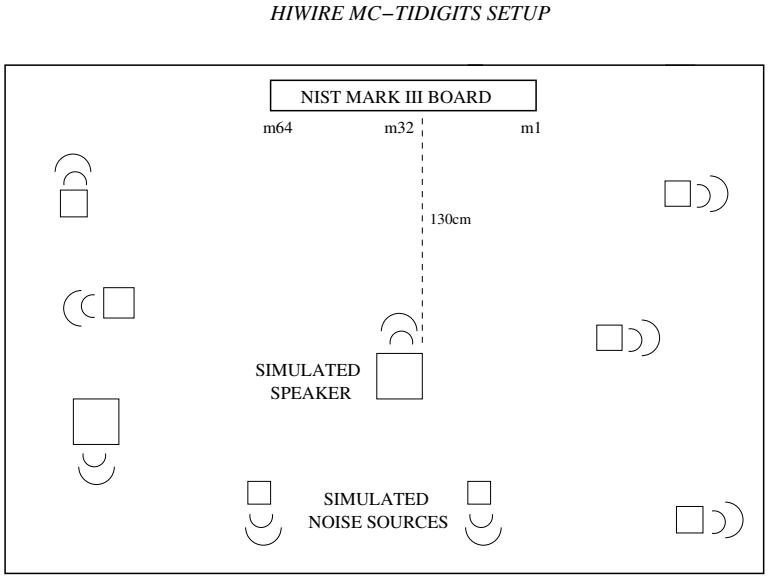
5.4 Πειραματικά Αποτελέσματα

Για να αποτιμήσουμε την αποδοτικότητα του προτεινόμενου πολυαισθητηριακού - πολυφασματικού σχήματος αποδιαμόρφωσης και της συνολικής μεθόδου εξαγωγής χαρακτηριστικών, μελετούμε την απόδοσή τους στο πρόβλημα της αυτόματης αναγνώρισης φωνής (ASR) παρουσία θορύβου. Η προτεινόμενη μέθοδος πολυαισθητηριακής εξαγωγής ενεργειακών διανυσματικών χαρακτηριστικών, την οποία από εδώ και στο εξής θα καλούμε MSCE (MultiSensor Cross Energy-based), συγκρίνεται με τα αποτελέσματα που προκύπτουν από τη χρήση εναλλακτικών μεθόδων. Τα πειράματα που θα παρουσιάσουμε διακρίνονται σε δύο είδη. Το πρώτο είδος αφορά σε αποτελέσματα αναγνώρισης στα οποία τα διανύσματα χαρακτηριστικών βασίζονται στη μέτρηση της μέσης τυπικής ενέργειας, ενώ το δεύτερο είδος πειραμάτων σε αποτελέσματα που προκύπτουν, όταν τα διανυσματικά χαρακτηριστικά βασίζονται στην μέση Teager ενέργεια.

Μια από τις εναλλακτικές μεθόδους που θα χρησιμοποιηθεί στις συγκρίσεις του πρώτου είδους πειραμάτων είναι η εξαγωγή των MFCC χαρακτηριστικών [DM80] από τον κεντρικό αισθητήρα της συστοιχίας. Σε αυτή την περίπτωση δεν εφαρμόζεται καμία τεχνική για την εξάλειψη της επίδρασης του θορύβου. Οι υπόλοιπες μέθοδοι, τις οποίες χρησιμοποιούμε στις συγκρίσεις μας, βασίζονται στην εξαγωγή των MFCC χαρακτηριστικών από την αποθορυβοποιημένη έκδοση των σημάτων εισόδου που προκύπτει στην έξοδο ενός συστήματος διαμόρφωσης δέσμης (βλέπε Κεφ. 4). Σε αυτή την περίπτωση χρησιμοποιούμε τρεις διαφορετικές τεχνικές διαμόρφωσης δέσμης, την Delay and Sum (DS) [VB88], την MPDR (Minimum Power Distortionless Response) [Tre02] και την MVDR (Minimum Variance Distortionless Response) [CZO87] διαμόρφωση. Η συγκεκριμένη προσέγγιση του συνδυασμού των σημάτων εισόδου στη συστοιχία και της εξαγωγής διανυσμάτων χαρακτηριστικών από το προκύπτον αποθορυβοποιημένο σήμα είναι αρκετά συνηθισμένη [OMSG97, HKDS99] και μπορεί να θεωρηθεί ως η πιο διαδεδομένη πρακτική στην περίπτωση αναγνώρισης φωνής παρουσία θορύβου σε πολυαισθητηριακά περιβάλλοντα. Για το δεύτερο είδος πειραμάτων χρησιμοποιούμε πάλι τις ίδιες μεθόδους με τη διαφορά ότι τα διανυσματικά χαρακτηριστικά, τα οποία εξάγονται σε κάθε περίπτωση, αποτελούν μια παραλλαγή των MFCCs στην οποία αντί για την τυπική υπολογίζεται η Teager ενέργεια. Τέλος, να σημειώσουμε ότι για την ακουστική επεξεργασία των σημάτων και στα δύο είδη πειραμάτων, τα οποία παρουσιάζουμε, χρησιμοποιήσαμε συστοιχία από 26 τριγωνικά φίλτρα κατανεμημένα ομοιόμορφα στη Mel κλίμακα συχνοτήτων.

5.4.1 Περιγραφή Βάσης Δεδομένων και Στατιστικού Συστήματος

Το σύνολο των ακουστικών δεδομένων που χρησιμοποιήσαμε για τα ASR πειράματα προέρχεται από την HIWIRE-MCTIDIGITS βάση η οποία αποτελεί ένα υποσύνολο της TIDIGITS βάσης δεδομένων. Η συγκεκριμένη βάση απαρτίζεται από εκφωνήσεις συνδεδεμένων ψηφίων και περιέχει περίπου 3000 εγγραφές από 52 άνδρες και 52 γυναίκες ομιλητές. Η συλλογή των δεδομένων έγινε σε ένα ειδικά σχεδιασμένο δωμάτιο του ερευνητικού ινστιτούτου ITC-IRST (The Center for Scientific and Technological Research), το οποίο χαρακτηρίζεται από πολύ μικρούς χρόνους αντήχησης - πρακτικά είναι σχεδόν ανηχοϊκό, με την NIST Mark III κονσόλα που αποτελείται από μια γραμμική συστοιχία 64 αισθητήρων (δες Σχ. 5.8). Από τους 64 αισθητήρες στα πειράματα που παρουσιάζουμε χρησιμοποιήσαμε μόνο τους 8 κεντρικούς. Η απόσταση μεταξύ γειτονικών αισθητήρων είναι 2 εκ., ενώ η δειγματοληψία των σημάτων έγινε σε συχνότητα 16 KHz. Ο κάθε ομιλητής (πηγή του ακουστικού σήματος) κατά τις εκφωνήσεις βρισκόταν ακριβώς μπροστά από τη συστοιχία και σε απόσταση 1.3 μέτρων από το κέντρο της. Οι πληροφορίες για την απόσταση του ομιλητή από το κέντρο της συστοιχίας, αλλά και για την απόσταση μεταξύ των αισθητήρων, κατέστησαν δυνατή τη χρονική ευθυγράμμιση των σημάτων εισόδου η οποία αποτελεί απαραίτητη προϋπόθεση για την περαιτέρω επεξεργασία. Όλες οι καταγραφές έχουν αλλοιωθεί από θόρυβο σε διάφορα επίπεδα SNR. Το είδος του θορύβου που χρησιμοποιήθηκε προέρχεται από πραγματικό θόρυβο που καταγράφηκε στο θάλαμο διακυβέρνησης ενός αεροσκάφους και το φασματικό του περιεχόμενο είναι χρωματισμένο.



Σχήμα 5.8: Σχεδιάγραμμα του δωματίου στο οποίο πραγματοποιήθηκαν οι καταγραφές των εκφωνήσεων.

Τα πειράματα αναγνώρισης πραγματοποιήθηκαν με τη χρήση του στατιστικού συστήματος HMM Toolkit (HTK) [YEK+02]. Ειδικότερα για τη φάση της εκπαίδευσης και του ελέγχου χρησιμοποιήσαμε τα κρυφά μαρκοβιανά μοντέλα (hidden Markov models - HMMs) πλήρους ψηφίου και τύπου δεξιά-αριστερά [RJ93, GY07] με το κάθε ένα να έχει 14 καταστάσεις. Κάθε κατάσταση μοντελοποιείται με μείγμα τριών γκαουσιανών κατανομών. Τα HMMs εκπαιδεύθηκαν σε ένα σύνολο από "καθαρά" δεδομένα (2100 εκφωνήσεις) και ελέγχθηκαν σε 7 σύνολα δεδομένων (7 PSNRs $\times$ 900 εκφωνήσεις) εκ των οποίων τα 6 παρουσιάζουν αλλοιώσεις από θόρυβο. Ως διανύσματα χαρακτηριστικών, για κάθε χρονικό πλαίσιο του ακουστικού σήματος, χρησιμοποιήσαμε τους 12 μεγαλύτερους από τους 26 συνολικά συντελεστές οι οποίοι προκύπτουν από την εφαρμογή του

DCT μετασχηματισμού στις 26 λογαριθμικές ενέργειες που υπολογίζονται στην έξοδο της συστοιχίας των τριγωνικών φίλτρων. Τα χαρακτηριστικά αυτά στη συνέχεια επεκτάθηκαν χρησιμοποιώντας τις χρονικές παραγώγους τους ($\Delta, \Delta\Delta$) και είχαν τελικό μέγεθος 36 συντελεστών.

5.4.2 Αποτελέσματα Αναγνώρισης Φωνής

Στην παρούσα ενότητα παρουσιάζονται τα αποτελέσματα αναγνώρισης που προέκυψαν για όλες τις συνθήκες ελέγχου και από όλες τις υπό εξέταση μεθόδους εξαγωγής χαρακτηριστικών. Η πρώτη συνθήκη ελέγχου είναι ίδια με τη συνθήκη εκπαίδευσης, δηλαδή, τα δεδομένα δεν παρουσιάζουν καμία αλλοίωση από το θόρυβο. Ωστόσο, τα δεδομένα στα οποία εκπαιδεύτηκε το σύστημα αναγνώρισης είναι διαφορετικά από αυτά που χρησιμοποιούνται κατά τον έλεγχο. Στις επόμενες 6 συνθήκες τα δεδομένα ελέγχου παρουσιάζουν αλλοιώσεις για διαφορετικά επίπεδα θορύβου με το SNR να κυμαίνεται από -5–20 dB με βήμα 5 dB. Στον Πίνακα 5.1 παρέχονται οι μέσες τιμές των αποτελεσμάτων αναγνώρισης για όλες τις συνθήκες ελέγχου με τα εξαγόμενα διανύσματα χαρακτηριστικών να βασίζονται στη μέση τυπική ενέργεια. Στον Πίνακα 5.2 παρουσιάζονται τα αντίστοιχα αποτελέσματα, όταν τα διανύσματα χαρακτηριστικών βασίζονται στη μέση Teager ενέργεια. Στους Πίνακες 5.1 και 5.2 αναφέρονται τα καλύτερα δυνατά αποτελέσματα που προέκυψαν για κάθε μέθοδο μοντελοποιώντας τις καταστάσεις του στατιστικού συστήματος με 1 έως 3 γκαουσιανές κατανομές.

Πίνακας 5.1: Αποτελέσματα αναγνώρισης (μέση ακρίβεια αναγνώρισης λέξεων (%) για διαφορετικές συνθήκες εκπαίδευσης/ελέγχου) με τη μέτρηση της τυπικής ενέργειας.

Correct Word Accuracies (%)							
SNR/ Methods	clean	20dB	15dB	10dB	5dB	0dB	-5dB
MFCC	96.23	94.07	86.25	71.95	52.69	34.50	16.67
DS+MFCC	96.16	93.25	85.69	72.66	55.81	37.77	22.74
MVDR+MFCC	**97.19**	93.61	88.18	77.40	61.52	44.47	25.45
MPDR+MFCC	96.58	94.78	90.51	79.95	65.26	44.34	26.09
MSCE	96.80	**96.80**	**95.57**	**92.32**	**84.66**	**69.05**	**45.36**

Πίνακας 5.2: Αποτελέσματα αναγνώρισης (μέση ακρίβεια αναγνώρισης λέξεων (%) για διαφορετικές συνθήκες εκπαίδευσης/ελέγχου) με τη μέτρηση της Teager ενέργειας.

Correct Word Accuracies (%)							
SNR/ Methods	clean	20dB	15dB	10dB	5dB	0dB	-5dB
TECC	96.42	94.40	88.80	79.45	64.53	44.52	24.11
DS+TECC	**96.94**	95.98	91.40	83.27	69.78	49.75	32.30
MVDR+TECC	96.72	95.39	90.65	82.45	66.02	49.61	31.05
MPDR+TECC	96.60	94.82	90.93	84.14	70.44	52.22	34.61
MSCE	96.14	**96.68**	**95.42**	**91.98**	**83.40**	**65.90**	**39.28**

Μελετώντας τα πειραματικά αποτελέσματα που προέκυψαν διαπιστώνει κανείς ότι στην περί-

πτωση απουσίας θορύβου, δηλαδή, όταν η εκπαίδευση και ο έλεγχος του συστήματος αναγνώρισης πραγματοποιείται στις ίδιες συνθήκες, η προτεινόμενη μέθοδος MSCE παρουσιάζει μια ελαφρώς χειρότερη επίδοση σε σχέση με τη μέθοδο που δίνει τα καλύτερα αποτελέσματα αναγνώρισης, MVDR και DS διαμόρφωση δέσμης για χαρακτηριστικά που βασίζονται στην τυπική και την Teager ενέργεια, αντιστοίχως. Ωστόσο, για διαφορετικές συνθήκες εκπαίδευσης και ελέγχου η προτεινόμενη μέθοδος είναι σταθερά καλύτερη από όλες τις υπόλοιπες ανεξάρτητα από το είδος της ενεργειακής μέτρησης που χρησιμοποιούμε. Συγκεκριμένα, παρατηρούμε ότι, καθώς ο σηματοθορυβικός λόγος μειώνεται και η επίδραση του θορύβου γίνεται όλο και μεγαλύτερη, η βελτίωση που επιτυγχάνεται με τη μέθοδο MSCE σε σχέση με τις υπόλοιπες μεθόδους αυξάνεται. Μάλιστα, διαπιστώνουμε ότι, ενώ οι υπόλοιπες μέθοδοι παρουσιάζουν μια σημαντική πτώση των ποσοστών αναγνώρισης για τις περιπτώσεις στις οποίες ο σηματοθορυβικός λόγος είναι μικρότερος των 15 dB, η προτεινόμενη μέθοδος εμφανίζει σαφώς καλύτερη ευρωστία στην επίδραση του θορύβου με υψηλά ποσοστά αναγνώρισης ακόμη και για τα 5 dB. Επιπλέον, και για τις πιο δύσκολες συνθήκες θορύβου, 0 dB και -5 dB, τα αποτελέσματα που λαμβάνουμε είναι αρκετά υψηλά συγκρινόμενα με αυτά των υπολοίπων μεθόδων.

Εκτός από τη σύγκριση μεταξύ του προτεινόμενου σχήματος εξαγωγής χαρακτηριστικών και των εναλλακτικών σχημάτων είναι επίσης αρκετά ενδιαφέρον και χρήσιμο να εξετάσουμε τη συμπεριφορά των δύο κατηγοριών ενεργειακών χαρακτηριστικών. Παρατηρούμε ότι, στην περίπτωση που τα εξαγόμενα χαρακτηριστικά βασίζονται στην Teager ενέργεια, τα αποτελέσματα των εναλλακτικών μεθόδων παρουσιάζουν σημαντική βελτίωση για όλες τις συνθήκες θορύβου. Το γεγονός αυτό οφείλεται εν μέρει στο ότι τα χωρικά φίλτρα διαμόρφωσης δέσμης DS, MVDR και MPDR, όπως ήδη έχουμε αναφέρει και στο Κεφ. 4, παρουσιάζουν περιορισμένη δυνατότητα αντιμετώπισης του θορύβου στην περιοχή των χαμηλών συχνοτήτων. Στην περίπτωση, λοιπόν, που χρησιμοποιούμε αντί για τη μέση τυπική τη μέση Teager ενέργεια, λόγω της εμφάνισης του επιπλέον όρου ω^2 (βλέπε Εξ. (5.27)), επιτυγχάνουμε με τεχνητό τρόπο αύξηση του SNR αντιμετωπίζοντας σε κάποιο βαθμό τις αδυναμίες των χωρικών φίλτρων. Βέβαια αυτή η τεχνητή αύξηση του SNR και κατ' επέκταση και η βελτίωση των ποσοστών αναγνώρισης είναι δυνατή, επειδή ο συγκεκριμένος θόρυβος που εξετάσαμε είναι χρωματισμένος και μορφής $1/f$. Στη γενική περίπτωση, όπως αναφέραμε και στην Ενοτ. (5.2.4), θα πρέπει κάθε φορά να εξετάζουμε τόσο το φασματικό περιεχόμενο του θορύβου όσο και του σήματος ενδιαφέροντος, προκειμένου να επιλέγουμε την καταλληλότερη ενεργειακή μέτρηση.

5.5 Σύνοψη και Συμπεράσματα

Στο κεφάλαιο αυτό αρχικά μελετήσαμε δύο ενεργειακούς τελεστές που χρησιμοποιούνται ευρέως στην εξαγωγή διανυσματικών χαρακτηριστικών με σκοπό την αυτόματη αναγνώριση. Η ύπαρξη του ενοποιητικού πλαισίου, το οποίο αποδείξαμε για τις μετρήσεις που λαμβάνονται από τους τελεστές αυτούς, αναδεικνύει τη δυνατότητα θέσπισης ενός κριτηρίου επιλογής του καταλληλότερου τελεστή ανάλογα με τις επικρατούσες συνθήκες θορύβου. Το γεγονός αυτό είναι εξαιρετικής σημασίας, αφού, όπως τελικά προκύπτει, η χρήση κάθε μιας από τις δύο ενεργειακές μετρήσεις παρουσιάζεται ευνοϊκή για διαφορετικά είδη θορύβου. Συνεπώς, είναι δυνατή η λήψη πιο εύρωστων μετρήσεων και κατ' επέκταση και η βελτίωση των ποσοστών αναγνώρισης απλώς επιλέγοντας κανείς κάθε φορά τον κατάλληλο ενεργειακό τελεστή.

Μέσω της πολυαισθητηριακής ανάλυσης που ακολουθήθηκε είναι δυνατόν πέραν της φασματικής πληροφορίας να χρησιμοποιηθεί και η χωρική πληροφορία που παρέχεται από τη συστοιχία των αισθητήρων. Το γεγονός αυτό καθιστά εφικτή την ανάπτυξη ενός σχήματος εξαγωγής χαρακτηριστικών τα οποία παρουσιάζονται ανθεκτικότερα στο θόρυβο. Η μεγαλύτερη ευρωστία των εξαγόμενων χαρακτηριστικών οφείλεται στη δυνατότητα λήψης πολλαπλών μετρήσεων, αλλά και στη μη-ομοιόμορφη επίδραση του θορύβου σε διαφορετικές χωρικές συντεταγμένες. Ο συνδυασμός

αυτών των δύο γεγονότων έχει ως αποτέλεσμα τη σημαντική βελτίωση των ποσοστών αναγνώρισης ακόμα και σε εξαιρετικά δύσκολες συνθήκες.

Κεφάλαιο 6

Σύνοψη Διατριβής και Μελλοντική Έρευνα

6.1 Ερευνητική Συνεισφορά

Η εργασία μας κινείται στην περιοχή της επεξεργασίας και ανάλυσης (πολυαισθητηριακών, πολυκαναλικών ή πολυφασματικών) σημάτων και εικόνων και έχει ως βασικά πεδία εφαρμογής την ψηφιακή αποκατάσταση μιας ειδικής κατηγορίας φυσικών εικόνων που εμφανίζονται σε ερευνητικές περιοχές, όπως η αστρονομία και η βιοϊατρική, την αποθορυβοποίηση ακουστικών σημάτων που λαμβάνονται από συστοιχίες αισθητήρων, καθώς και την εξαγωγή χαρακτηριστικών από αυτά. Οι συνεισφορές της διατριβής μπορούν να συνοψισθούν στα ακόλουθα σημεία:

- Μελέτη και ανάπτυξη τοπικά εντοπισμένων μη-γραμμικών μετασχηματισμών που είναι κατάλληλοι για την ανάλυση φυσικών εικόνων σε πολλαπλές κλίμακες. Η συγκεκριμένη κατηγορία μετασχηματισμών παρουσιάζεται ως καταλληλότερη σε σχέση με εναλλακτικούς στην ανάλυση εικόνων που εμφανίζουν αλλοιώσεις από σηματο-εξαρτώμενο θόρυβο (Poisson), αφού εμφανίζει τη δυνατότητα αποσύνδεσης αυτού του είδους θορύβου από την πραγματική πληροφορία της εικόνας. Επιπλέον, στο συγκεκριμένο ερευνητικό πλαίσιο αποδείξαμε την ύπαρξη μιας στενής σχέσης που συνδέει την κατηγορία των μετασχηματισμών που μελετήσαμε με το μετασχηματισμό κυματιδίων. Η σύνδεση αυτή αποδεικνύεται εξαιρετικά σημαντική αφού αφενός υποδηλώνει οτι βασικές ιδιότητες του μετασχηματισμού κυματιδίων κληρονομούνται από τους μη-γραμμικούς μετασχηματισμούς αφετέρου επιτρέπει την αποδοτική υλοποίησή τους μέσα από μια επαναληπτική εφαρμογή συστοιχιών FIR φίλτρων.

- Μοντέλα στατιστικής περιγραφής των εικόνων στο πεδίο των μη-γραμμικών τοπικά εντοπισμένων μετασχηματισμών. Σε αυτό το πλαίσιο προτείναμε ένα νέο βελτιωμένο στατιστικό μοντέλο για την ανάλυση Poisson στοχαστικών διαδικασιών (εικόνες με Poisson θόρυβο) και την εκτίμηση της γενεσιουργού έντασής τους. Το μοντέλο αυτό είναι πιθανοτικό και χρησιμοποιεί αναδρομικές ως προς την κλίμακα μαρκοβιανές δομές που επιτρέπουν τη στατιστική συμπερασματολογία για την πραγματική ένταση της εικόνας με αποδοτικό τρόπο. Επιπλέον, επειδή το μοντέλο αυτό είναι παραμετρικό, αναπτύξαμε μια αυτοματοποιημένη, αποδοτική και εύρωστη μέθοδο εκτίμησης των παραμέτρων του απευθείας από τα θορυβώδη δεδομένα αντιμετωπίζοντας κατ' αυτό τον τρόπο το πρόβλημα επιλογής των κατάλληλων παραμέτρων που βελτιστοποιούν την απόδοση του μοντέλου.

- Εκτός από την ανάλυση και τη στατιστική επεξεργασία εικόνων το ενδιαφέρον μας επικεντρώθηκε και στο πρόβλημα της αποθορυβοποίησης ακουστικών σημάτων. Για την αντιμετώπιση αυτού του προβλήματος μελετήσαμε στατιστικά βέλτιστες μεθόδους επεξεργασίας σημάτων

τα οποία λαμβάνονται από πολλαπλούς αισθητήρες (πολυαισθητηριακή επεξεργασία σημάτων). Για να εφαρμόσει κανείς αυτές τις μεθόδους θα πρέπει προηγουμένως να εκτιμήσει τις παραμέτρους από τις οποίες εξαρτάται ένα φίλτρο που καλείται ύστερο φίλτρο. Για την εκτίμηση αυτών των παραμέτρων αναπτύξαμε μια νέα τεχνική η οποία είναι γενική και μπορεί να εκτιμήσει διαφορετικές εκφράσεις του ύστερου φίλτρου (γραμμικές και μη-γραμμικές) που προκύπτουν από τη βελτιστοποίηση εναλλακτικών στατιστικών κριτηρίων.

- Εύρωστη εξαγωγή διανυσμάτων χαρακτηριστικών από ακουστικά σήματα με σκοπό την αυτόματη κατηγοριοποίηση και αναγνώριση. Σε αυτό το πλαίσιο ασχοληθήκαμε με τη μελέτη και ανάπτυξη ενός πολυαισθητηριακού σχήματος αποδιαμόρφωσης ακουστικών σημάτων σε φασματικές συνιστώσες. Με βάση αυτό το σχήμα και την απόδειξη ύπαρξης ενός ενοποιητικού πλαισίου για δύο ενεργειακούς τελεστές που χρησιμοποιούνται στη λήψη ενεργειακών μετρήσεων, αναπτύξαμε ένα πολυαισθητηριακό σχήμα εξαγωγής εύρωστων στο θόρυβο χαρακτηριστικών. Τα χαρακτηριστικά αυτά μπορούν να χρησιμοποιηθούν στο πρόβλημα της αυτόματης αναγνώρισης σημάτων.

Κατάλογος με τις σχετικές μας δημοσιεύσεις παρατίθεται στο Παράρτημα Α΄.

6.2 Κατευθύνσεις Μελλοντικής Έρευνας

Μια σειρά ζητημάτων που σχετίζονται με την ερευνητική μας προσπάθεια στα πλαίσια της διδακτορικής διατριβής παραμένουν ανοικτά και θα απασχολήσουν το ερευνητικό μας ενδιαφέρον στο άμεσο μέλλον. Τα ζητήματα αυτά αφορούν τόσο σε βελτιώσεις των προτεινόμενων τεχνικών για τα προβλήματα που μελετήσαμε όσο και σε πιθανές επεκτάσεις τους με σκοπό την εφαρμογή τους σε συγγενή προβλήματα. Πιο συγκεκριμένα, για τις μεθόδους που αναπτύξαμε στο πρώτο μέρος της διατριβής μας αξίζει να αναφερθούμε στα ακόλουθα ζητήματα:

- Σχετικά με τις αναπαραστάσεις εικόνων, όπως αναφέραμε και στο Κεφ. 2, η επεξεργασία των παραγόμενων συντελεστών στο πεδίο του πολλαπλασιαστικού μετασχηματισμού οδηγεί στην εμφάνιση τεχνουργημάτων κατά την ανακατασκευή. Το πρόβλημα αυτό είναι δυνατόν να αντιμετωπιστεί με τη χρήση υπερπλήρων αναπαραστάσεων. Ωστόσο, ένα ζήτημα που παραμένει ανοικτό είναι ότι στην περίπτωση της υπερπλήρους αναπαράστασης που αναπτύξαμε (μη-δειγματοληπτημένος Poisson-Haar μετασχηματισμός) δεν είναι ξεκάθαρο πως μπορεί να εφαρμοσθεί μια στατιστική δομή που θα μοντελοποιεί τις δια-κλιμακωτές εξαρτήσεις των συντελεστών. Προς αυτή την κατεύθυνση αξίζει να διερευνηθεί η δυνατότητα ανάπτυξης ενός νέου μη-κρίσιμα δειγματοληπτημένου μετασχηματισμού που θα εμφανίζει μια δενδρική δομή και συνεπώς θα επιτρέπει τη χρήση του εξαιρετικά ισχυρού HMT μοντέλου.

- Όπως συζητήθηκε στο Κεφ. 3 εκτός από τις στατιστικές εξαρτήσεις που εμφανίζουν οι συντελεστές μεταξύ κλιμάκων και τις οποίες μοντελοποιούμε με το HMT μοντέλο, υπάρχουν και εξαρτήσεις εντός της ίδιας κλίμακας ανάλυσης. Είναι λοιπόν προφανές ότι η αναζήτηση μιας στατιστικής δομής που θα μπορεί να μοντελοποιεί αποτελεσματικά ταυτοχρόνως και τα δύο είδη στατιστικών εξαρτήσεων θα επιτρέψει την ανάπτυξη εξαιρετικά αποδοτικών μεθόδων επεξεργασίας.

- Οι στατιστικές δομές που μελετήσαμε για τη μοντελοποίηση των εικόνων εκτός από το πρόβλημα της αποθορυβοποίησης έχουν εφαρμοσθεί επιτυχώς και σε επιπρόσθετα προβλήματα, στην περίπτωση γκαουσιανού θορύβου, όπως η συμπίεση και κατάτμηση εικόνων. Έχει ενδιαφέρον λοιπόν να εξετάσουμε κατά πόσο είναι εφικτή η χρήση των πολυκλιμακωτών μοντέλων που αναπτύξαμε στην αυτόματη ανίχνευση χαρακτηριστικών σε εικόνες με Poisson θόρυβο. Η δυνατότητα αυτή μπορεί να φανεί ιδιαίτερα χρήσιμη κατά την επεξεργασία

αστρονομικών εικόνων. Στην περίπτωση αυτή οι επιστήμονες επιθυμούν την ύπαρξη στατιστικών εργαλείων που μπορούν να τους καθοδηγήσουν σχετικά με την παρουσία ή όχι αστρονομικών πηγών (όπως γαλαξίες κ.α.), ιδιαίτερα όταν οι εικόνες παρουσιάζουν σοβαρές αλλοιώσεις λόγω θορύβου.

Όσον αφορά στις μεθόδους που αναπτύξαμε στο δεύτερο μέρος της διατριβής μπορούμε να αναφερθούμε στα εξής:

- Το πολυαισθητηριακό σχήμα εξαγωγής που αναπτύξαμε παρουσιάζει αρκετά καλή συμπεριφορά σε σχετικά αντίξοες συνθήκες θορύβου. Ωστόσο, τα αποτελέσματα αναγνώρισης που παρουσιάσαμε στο Κεφ. 5 είναι αρκετά πρώιμα και σε καμία περίπτωση δεν μπορούμε να θεωρήσουμε ότι δεν επιδέχονται βελτιώσεων. Προς αυτή την κατεύθυνση έχει ενδιαφέρον να ερευνήσουμε τόσο τη δυνατότητα ενσωμάτωσης και διαφορετικών, πιθανόν πιο εύρωστων μετρήσεων αλλά και τη δυνατότητα για περαιτέρω επεξεργασία των εξαγόμενων χαρακτηριστικών. Με τον τρόπο αυτό ενδεχομένως να οδηγηθούμε σε επιπλέον βελτιώσεις των αποτελεσμάτων αναγνώρισης, ιδιαίτερα στις περιπτώσεις έντονης παρουσίας θορύβου.
- Ιδιαίτερα χρήσιμη είναι η εξέταση της αποδοτικότητας του προτεινόμενου σχήματος εξαγωγής χαρακτηριστικών μέσα από συγκρίσεις με ένα ευρύτερο σύνολο εναλλακτικών μεθόδων. Εξίσου χρήσιμη είναι και η πειραματική μελέτη της προτεινόμενης μεθόδου σε μεγάλες βάσεις όπου τα δεδομένα εμφανίζουν αλλοιώσεις από πραγματικό θόρυβο.

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

Παράρτημα Α΄

Δημοσιεύσεις Διατριβής

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

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

- Μοντελοποίηση φυσικών εικόνων περιορισμένης φωτεινότητας στο πεδίο τοπικά εντοπισμένων μετασχηματισμών και ανάπτυξη πολυκλιμακωτών μεθόδων για την αντιμετώπιση του Poisson θορύβου [J2, C5, C6]. Σε συνεργασία με τους Γ. Παπανδρέου και Π. Μαραγκό.
- Μέθοδοι πολυαισθητηριακής επεξεργασίας για αποθορυβοποίηση ακουστικών σημάτων [J1, C1]. Σε συνεργασία με τον Π. Μαραγκό.
- Πολυαισθητηριακή - Πολυφασματική επεξεργασία ακουστικών σημάτων για εξαγωγή χαρακτηριστικών και αναγνώριση [C2, C3, C4]. Σε συνεργασία με τους Π. Μαραγκό, Α. Κατσαμάνη και Δ. Δημητριάδη.

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

Δημοσιεύσεις σε Περιοδικά

J1. S. Lefkimmiatis και P. Maragos: A Generalized Estimation Approach for Linear and Nonlinear Microphone Array Post-Filters. Speech Communication, 49:657–666, Ιούλιος 2007.

J2. S. Lefkimmiatis, 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. S. Lefkimmiatis και P. Maragos: Optimum Post-Filter Estimation for Noise Reduction in Multichannel Speech Processing. Στο Proc. 14th European Signal Processing Conf. (EUSIPCO-06), Florence, Italy, Sept. 2006.

C2. S. Lefkimmiatis, D. Dimitriadis, και P. Maragos: An Optimum Microphone Array Post-Filter for Speech Applications. Στο Proc. Int. Conf. on Spoken Language Processing (InterSpeech 2006 - ICSLP), Pittsburgh PA, USA, Sep. 2006., σελίδες 2142–2145, 2006.

C3. D. Dimitriadis, P. Maragos, και S. Lefkimmiatis: Multiband, Multisensor Features for Robust Speech Recognition. Στο Proc. Int. on Speech Technology and Communication (InterSpeech 2007 - EuroSpeech), Antwerp, Belgium, Aug. 2007.

C4. S. Lefkimmiatis, P. Maragos, και A. Katsamanis: Multisensor Multiband Cross-Energy Tracking for Feature Extraction and Recognition. Στο Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-2008), Las Vegas, Nevada, U.S.A., Mar.-Apr. 2008.

C5. S. Lefkimmiatis, G. Papandreou, και P. Maragos: Photon-limited image denoising by inference on multiscale models. Στο Proc. IEEE Int. Conf. on Image Processing (ICIP-08), San Diego, CA, October, 2008.

C6. S. Lefkimmiatis, 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-09), Cairo, Egypt, November, 2009.

Βιβλιογραφία

[ABB77] Allen, J. B., D. A. Berkley, και J. Blauert: *Multimicrophone signal-processing technique to remove room reverberation from speech signals*. J. Acoust. Soc. Amer., 62(4):912–915, 1977.

[All77] Allen, J. B.: *Short-term spectral analysis and synthesis and modification by discrete Fourier transform*. IEEE Trans. Acoust. Speech Signal Processing, 25(3):235–238, 1977.

[All94] Allen, J. B.: *How do humans process and recognize speech?* IEEE Trans. Speech and Audio Processing, 4(2):567–577, 1994.

[Ans48] Anscombe, F. J.: *The transformation of Poisson, binomial and negative-binomial data*. Biometrika, 35(3):246–254, 1948.

[AR77] Allen, J. B. και L.R. Rabiner: *A unified theory of short-time spectrum analysis and synthesis*. Proc. IEEE, 65(11):1558–1564, 1977.

[AS72] Abramowitz, M. και I. A. Stegun: *Handbook of Mathematical Functions*. Dover, 1972.

[AS01] Antoniadis, A. και T. Sapatinas: *Wavelet shrinkage for natural exponential families with quadratic variance functions*. Biometrica, 88:805–820, 2001.

[Ata74] Atal, B. S.: *Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification*. J. Acoust. Soc. Amer., 55:1304–1312, 1974.

[BA83] Burt, P. J. και E. H. Adelson: *The Laplacian pyramid as a compact image code*. IEEE Trans. Commun., 31(4):532–540, 1983.

[Bar36] Bartlett, M. S.: *The square root transformation in analysis of variance*. J. Roy. Statist. Soc. (Suppl.), 3(1):68–78, 1936.

[Bar90] Barett, H.: *Objective assessment of image quality: effects of quantum noise and object variability*. J. Opt. Soc. Am. A, 7(7):1266–1278, 1990.

[BFS04] Besbeas, P., I. D. Feis, και T. Sapatinas: *A comparative simulation study of wavelet shrinkage estimators for Poisson counts*. Internat. Statist. Rev., 72(2):209–237, 2004.

[Bis06] Bishop, C.: *Pattern Recognition and Machine Learning*. Springer, 2006.

[BN03] Balahrishnan, N. και V. B. Nevzorov: *A Primer on Statistical Distributions*. Wiley, 2003.

[BQ93] Bovik, A.C. και P. Maragos T.F. Quatieri: *AM-FM energy detection and separation in noise using multiband energy operators*. IEEE Trans. Signal Processing, 41:3245–3265, 1993.

[BR02] Balan, R. και J. Rosca: *Microphone array speech enhancement by Bayesian estimation of spectral amplitude and phase*. Στο *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, σελίδες 209–213, 2002.

[BRB+04] Balan, R., J. Rosca, C. Beaugeant, V. Gilg, και T. Fingscheidt: *Generalized stochastic principle for microphone array speech enhancement and applications to car environments*. Στο *Proc. 14th European Signal Processing Conf. (EUSIPCO-04)*, 2004.

[BS00] Bernardo, J. και A. Smith: *Bayesian Theory*. Wiley, 2000.

[BS01] Bitzer, J. και K. U. Simmer: *Superdirective microphone arrays*. Στο Brandstein, M. και D. Ward (επιμελητές): *Microphone Arrays: Signal Processing Techniques and Applications*, κεφάλαιο 2, σελίδες 19–38. Springer Verlag, 2001.

[BSK98] Bitzer, J., K. U. Simmer, και K. Kammeyer: *Multichannel noise reduction - algorithms and theoritical limits*. Στο *Proc. 6th European Signal Processing Conf. (EUSIPCO-98)*, 1998.

[BSK01] Bitzer, J., K. U. Simmer, και K. D. Kammeyer: *Multi-microphone noise reduction techniques as front-end devices for speech recognition*. Speech Communication, 34:3–12, 2001.

[CB98] Choi, H. και R. Baraniuk: *Analysis of wavelet-domain wiener filters*. Στο *IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, 1998.

[CB01] Choi, H. και R. G. Baraniuk: *Multiscale image segmentation using wavelet-domain hidden Markov models*. IEEE Trans. Image Process., 10(9):1309–1321, 2001.

[CB02] Cohen, I. και B. Berdugo: *Microphone array post-filtering for nonstationary noise suppresion*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, τόμος 1, σελίδες 901–904, 2002.

[CD95] Coifman, R. και D. Donoho: *Translation invariant de-noising*. Στο *Lecture Notes in Statistics*, σελίδες 125–150. Springer Verlag, 1995.

[CD99a] Candès, E. και D. Donoho: *Ridgelets: the key to high dimensional intermittency?* Philosophical Transactions of the Royal Society of London A, 357:2495–2509, 1999.

[CD99b] Candès, E. και D. L. Donoho: *Curvelets: a surprisingly effective nonadaptive representation for objects with edges*. Στο Cohen, A., C. Rabut, και L.L. Schumaker (επιμελητές): *Curve and Surface Fitting*. Vanderbilt University Press, 1999.

[CD99c] Candès, E. και D. L. Donoho: *Ridgelets and their derivatives: Representation of images with edges*. Στο Cohen, A., C. Rabut, και L.L. Schumaker (επιμελητές): *Curve and Surface Fitting*. Vanderbilt University Press, 1999.

[CDDY05] Candès, E.J., L. Demanet, D.L. Donoho, και L. Ying: *Fast discrete curvelet transforms*. Multiscale Model Simulations, 5:861–899, 2005.

[CLLT08] Chatelain, F., S. Lambert-Lacroix, και J. Y. Tourneret: *Pairwise likelihood estimation for multivariate mixed Poisson models generated by gamma intensities*. Statistics and Computing, 2008.

[CNB98] Crouse, M., R. Nowak, και G. Baraniuk: *Wavelet-based statistical signal processing using hidden Markov models*. IEEE Trans. Signal Process., 46(4):886–902, 1998.

[Coh04] Cohen, I.: *Multichannel post-filtering in nonstationary noise environments*. IEEE Trans. Signal Processing, 52(5):1149–1160, 2004.

[CYV00] Chang, S. G., B. Yu, και M. Vetterli: *Spatially adaptive wavelet thresholding with context modeling for image denoising*. IEEE Trans. Image Process., 9(9):1522–1531, 2000.

[CZK86] Cox, H., R. M. Zeskind, και T. Kooij: *Practical supergain*. IEEE Trans. Speech and Audio Processing, 34(3):393–398, 1986.

[CZO87] Cox, H., R. M. Zeskind, και M. W. Owen: *Robust adaptive beamforming*. IEEE Trans. Speech and Audio Processing, 35(10):1365–1376, 1987.

[DA08] Droppo, J. και A. Acero: *Environmental robustness*. Στο Benesty, J., M. M. Sondhi, και Y. Huang (επιμελητές): *Handbook of Speech Processing*, κεφάλαιο 33, σελίδες 653–679. Springer, 2008.

[Dau92] Daubechies, I.: *Ten Lectures on Wavelets*. SIAM, 1992.

[Day69] Day, N. E.: *Estimating the components of a mixture of normal distributions*. Biometrika, 56(3):463–474, 1969.

[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, 2007.

[DHS01] Duda, R. O., P. E. Hart, και D.G. Stork: *Pattern Classification*. Willey, δεύτερη έκδοση, 2001.

[dlTPS+05] Torre, A. de la, A. M. Peinado, J.C. Segura, J. L. Perez-Cordoba, M. C. Benitez, και A. J. Rubio: *Histogram equalization of speech representation for robust speech recognition*. IEEE Trans. Speech and Audio Processing, 13(3):355–366, 2005.

[DM80] Davis, S. B. και P. Mermelstein: *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*. IEEE Trans. Acoust. Speech Signal Processing, 28(4):357–366, 1980.

[DM03] Doclo, S. και M. Moonen: *Design of far-field and near-field broadband beamformers using eigenfilters*. Speech Communication, 83:2641–2672, 2003.

[DM06] Dimitriadis, D. και P. Maragos: *Continuous energy demodulation methods and application to speech analysis*. Speech Communication, 48:819–837, 2006.

[DML07] Dimitriadis, D., P. Maragos, και S. Lefkimmiatis: *Multiband, multisensor features for robust speech recognition*. Στο *Proc. Int. on Speech Technology and Communication (InterSpeech)*, σελίδες 246–249, 2007.

[DMP05a] Dimitriadis, D., P. Maragos, και A. Potamianos: *Auditory Teager energy cepstrum coefficients for robust speech recognition*. Στο *Proc. Int. on Speech Technology and Communication (InterSpeech)*, σελίδες 3013–3016, 2005.

[DMP05b] Dimitriadis, D., P. Maragos, και A. Potamianos: *Robust AM-FM features for speech recognition*. IEEE Signal Processing Lett., 12(9):621–624, 2005.

[Don93] Donoho, D. L.: *Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data*. Στο *Proc. Symposia in Applied Mathematics*, τόμος 47, σελίδες 173–205, 1993.

[Don94] Donoho, D. L.: *Ideal spatial adaptation by wavelet shrinkage*. Στο *Biometrika*, τόμος 81, σελίδες 425–455, 1994.

[Don95] Donoho, D. L.: *De-noising by soft-thresholding*. Στο *IEEE Trans. Inf. Theory*, τόμος 41, σελίδες 613–627, 1995.

[DPM09] Dimitriadis, D., A. Potamianos, και P. Maragos: *A comparison of the squared energy and Teager-Kaiser operators for short-term energy estimation in additive noise*. IEEE Trans. Signal Process., 57:2569–2581, 2009.

[DV05] Do, M. N. και M. Vetterli: *The contourlet transform: an efficient directional multiresolution image representation*. IEEE Trans. Image Process., 5(12):2091–2106, 2005.

[ECKVD04] Esch, D. N., A. Connors, M. Karovska, και D. A. Van Dyk: *An image restoration technique with error estimates*. The Astrophysical Journal, 610:1213–1227, 2004.

[EKM05] Evangelopoulos, G., I. Kokkinos, και P. Maragos: *Advances in variational image segmentation using AM-FM models: Regularized demodulation and probabilistic cue integration*. Στο *Proc. of Third International Workshop, VLSM 2005*, σελίδες 121–136, 2005.

[Elk01] Elko, G. W.: *Spatial coherence function for differential microphones in isotropic noise fields*. Στο Brandstein, M. και D. Ward (επιμελητές): *Microphone Arrays: Signal Processing Techniques and Applications*, κεφάλαιο 4, σελίδες 61–85. Springer Verlag, 2001.

[EM84] Ephraim, Y. και D. Mallah: *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*. IEEE Trans. Acoust. Speech Signal Processing, 32(6):1109–1121, 1984.

[EM85] Ephraim, Y. και D. Mallah: *Speech enhancement using a minimum mean-square error log-spectral amplitude estimator*. IEEE Trans. Acoust. Speech Signal Processing, 33(2):443–445, 1985.

[EM06] Evangelopoulos, G. και P. Maragos: *Multiband modulation energy tracking for noisy speech detection*. IEEE Trans. on Audio, Speech and Language Process., 14(6):2024–2038, 2006.

[Eph92] Ephraim, Y.: *Gain-adapted hidden Markov models for recognition of clean and noisy speech*. IEEE Trans. Signal Processing, 40:1303–1316, 1992.

[Eva07] Evangelopoulos, G.: *Texture and Image Microstructure Analysis with Modulation Models, Energy and Variational Techniques: Detection & Separation*. Διδακτορική διατριβή, N.T.U.A., School of E.C.E., 2007.

[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, 1991.

[FDN07] Fryźlewicz, P., V. Delouille, και G. P. Nason: *Goes-8 x-ray sensor variance stabilization using the multiscale data-driven Haar-Fisz transform*. J. Roy. Statist. Soc. C, 56:99–116, 2007.

[Fie94] Field, D. J.: *What is the goal of sensory coding.* Neural Computation, (6):559–601, 1994.

[Fis55] Fisz, M.: *The limiting distribution of a function of two independent random variables and its statistical application.* Colloquium Mathematicum, 3:138–146, 1955.

[FK97] Fischer, S. και K. D. Kammeyer: *Broadband beamforming with adaptive postfiltering for speech acquisition in noisy environments.* Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, τόμος 1, σελίδες 359–362, 1997.

[FN04] Fryźlewicz, P. και G. P. Nason: *A Haar-Fisz algorithm for Poisson intensity estimation.* J. Comp. Graph. Stat., 13:621–638, 2004.

[FS96] Fischer, S. και K. U. Simmer: *Beamforming microphone arrays for speech acquisition in noisy environments.* Speech Communication, 20:215–227, 1996.

[FT50] Freeman, M. F. και J. W. Tukey: *Transformations related to the angular and the square root.* Ann. Math. Statist., 21(4):607–611, 1950.

[FVU05] Feilner, M., D. Van De Ville, και M. Unser: *An orthogonal family of quincunx wavelets with continuously adjustable order.* IEEE Trans. Image Process., 14(4):499–510, 2005.

[Gab46] Gabor, D.: *Theory of communication.* Journal Inst. of Elec. Eng. London, 93:429–457, 1946.

[GCSR03] Gelman, A., J.B. Carlin, H.S. Stern, και D.B. Rubin: *Bayesian Data Analysis.* Chapman & Hall, δεύτερη έκδοση, 2003.

[GG00] Gelfand, A. και M. Ghosh: *Generalized linear models: A Bayesian view.* Στο Dey, D.K., S. K. Ghosh, και B. K. Mallick (επιμελητές): *Generalized Linear Models: A Bayesian Perspective*, κεφάλαιο 1, σελίδες 3–22. Marcel Dekker Press, 2000.

[Gon95] Gong, Y.: *Speech recognition in noisy environments: A survey.* Speech Communication, 16:261–291, 1995.

[GY07] Gales, M. και S. Young: *The application of Hidden Markov Models in Speech Recognition.* Now Publishers, 2007.

[HaJMT89] Holschneider, M., R. Kronland Martinet adn J. Morlet, και P. Tchamitchian: *A real-time algorithm for signal analysis wuith the help of the wavelet transform.* Στο *Wavelets: Time-Frequency Methods and Phase Space*, σελίδες 289–297. Springer Verlag, 1989.

[Her87] Hermansky, H.: *An efficient speaker independent automatic speech recognition by simulation of some properties of human auditory perception.* Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, σελίδες 1156–1162, 1987.

[Her90] Hermansky, H.: *Perceptual linear predictive (plp) analysis of speech.* J. Acoust. Soc. Amer., 87(4):1738–1752, 1990.

[HHH09] Hyvárinen, A., J. Hurri, και P. O. Hoyer: *Natural Image Statistics.* Springer, 2009.

[HKDS99] Hughes, T., H. Kim, J. DiBiase, και H. Silverman: *Performance of an hmm speech recognizer using a real-time tracking microphone array as input.* IEEE Trans. Speech and Audio Processing, 7:346–349, 1999.

[HKO01] Hyvárinen, A., K. Karhunen, και E. Oja: *Independent Component Analysis*. Wiley, 2001.

[HM94] Hermansky, H. και N. Morgan: *Rasta processing of speech*. IEEE Trans. Speech and Audio Processing, 2(4):578–589, 1994.

[HP98] Hansen, J. H. L. και B. L. Pellom: *An effective quality evaluation protocol for speech enhancement algorithms*. Στο *Int. Conf. Spoken Language Processing (ICSLP)*, σελίδες 2819–2822, 1998.

[HW06] Heil, C. και D. F. Walnut (επιμελητές): *Fundamental papers in wavelet theory*. Princeton University Press, 2006.

[IP97] Irino, T. και R. D. Patterson: *A time-domain level-dependent auditory filter: The gammarchip*. J. Acoust. Soc. Amer., 101:412–419, 1997.

[Jan06] Jansen, M.: *Multiscale poisson data smoothing*. J. Roy. Statist. Soc. B, 68:27–48, 2006.

[JC99] Jabloun, F. και A. Enis Cetin: *The Teager energy based feature parameters for robust speech recognition in car noise*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, σελίδες 273–276, 1999.

[JD93] Johnson, D. H. και D. E. Dudgeon: *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993.

[JKK05] Johnson, N.L, A.W. Kemp, και S. Kotz: *Univariate Discrete Distributions*. Wiley, τρίτη έκδοση, 2005.

[JKN97] Johnson, N.L, S. Kotz, και N.Balakrishnan: *Discrete Multivariate Distributions*. Wiley, 1997.

[Kai90a] Kaiser, F.: *On a simple algorithm to calculate the energy of a signal*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, τόμος 1, σελίδες 381–384, 1990.

[Kai90b] Kaiser, J. F.: *On Teager's energy algorithm and it's generalization to continuous signals*. Στο *4th IEEE Digital Signal Processing Workshop*, 1990.

[Kai93] Kaiser, J. F.: *Some useful properties of Teager's energy operators*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, τόμος 3, σελίδες 149–152, 1993.

[Kay93] Kay, S. M.: *Fundamentals of Statistical Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.

[KC07a] Kovacevic, J. και A. Chebira: *Life beyond bases: The advent of frames (Part i)*. IEEE Signal Process. Mag., 24(4):86–104, 2007.

[KC07b] Kovacevic, J. και A. Chebira: *Life beyond bases: The advent of frames (Part ii)*. IEEE Signal Process. Mag., 24(5):115–125, 2007.

[KD00] Kolaczyk, E. D. και D.D. Dixon: *Nonparametric estimation of intensity maps using haar wavelets and poisson noise characteristics*. The Astrophysical Journal, 534:490–505, 2000.

[KEM09] Kokkinos, I., G. Evangelopoulos, και P. Maragos: *Texture analysis and segmentation using modulation features, generative models and weighted curve evolution*. IEEE Trans. Pattern Analyis and Machine Intelligence, 31(1):142–157, 2009.

[Kin98] Kingsbury, N. G.: *The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters*. Στο *8th IEEE DSP Workshop*, 1998.

[Kin00] Kingsbury, N.G.: *A dual-tree complex wavelet transform with improved orthogonality and symmetry properties*. Στο *Int. Conf. Image Process. (ICIP)*, σελίδες 375–378, 2000.

[Kin01] Kingsbury, N.G.: *Complex wavelets for shift invariant analysis and filtering of signals*. Appl. Comput. Harmon. Anal., 10(3):234–253, 2001.

[KN04] Kolaczyk, E. D. και R.D. Nowak: *Multiscale likelihood analysis and complexity penalized estimation*. Annals of Statistics, 32:500–527, 2004.

[Kol97] Kolaczyk, E.D.: *Non-parametric estimation of gamma-ray burst intensities using haar wavelets*. The Astrophysical Journal, 483:340–349, 1997.

[Kol99a] Kolaczyk, E.D.: *Bayesian multiscale models for Poisson processes*. J. Amer. Stat. Assoc., 94(447):920–933, 1999.

[Kol99b] Kolaczyk, Eric D.: *Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds*. Statistica Sinica, 9:119–135, 1999.

[KR96] Knill, D. C. και W. Richards: *Perception as Bayesian Inference*. Cambridge University Press, 1996.

[KS00] Kovacevic, J. και W. Sweldens: *Wavelet families of increasing order in arbitrary dimensions*. IEEE Trans. Image Process., 9(3):480–496, 2000.

[KV92] Kovacevic, J. και M. Vetterli: *Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for* $\mathbb{R}^n$. IEEE Trans. Inf. Theory, 38(2):533–555, 1992.

[KVE+08] Kotti, M., D. Ververidis, G. Evangelopoulos, I. Panagakis, C.Kotropoulos, P. Maragos, και I. Pitas: *Audio-assisted movie dialogue detection*. IEEE on Circuits and Systems for Video Technology, 18(11):1618–1627, 2008.

[KX05] Karlis, D. και E. Xekalaki: *Mixed Poisson distributions*. Int. Stat. Rev., 73(1):35–58, 2005.

[LA05] Li, J. και M. Akagi: *A hybrid microphone array post-filter in a diffuse noise field*. Στο *Proc. Int. on Speech Technology and Communication (InterSpeech)*, σελίδες 2313–2316, 2005.

[LAS06] Li, J., M. Akagi, και Y. Suzuki: *Improved hybrib microphone array post-filter by integrating a robust speech absence probability estimator for speech enhancement*. Στο *Int. Conf. Spoken Language Processing (ICSLP)*, σελίδες 2130–2133, 2006.

[LDM06] Lefkimmiatis, S., D. Dimitriadis, και P. Maragos: *An optimum microphone array post-filter for speech applications*. Στο *Int. Conf. Spoken Language Processing (ICSLP)*, σελίδες 2142–2145, 2006.

[Lim83] Lim, J. S.: *Speech Enhancement*. Prentice Hall, 1983.

[LKA04] Lu, H., Y. Kim, και J. Anderson: *Improved Poisson intensity estimation: Denoising application using Poisson data*. IEEE Trans. Image Process., 13(8):1128–1135, 2004.

[LM06] Lefkimmiatis, S. και P. Maragos: *Optimum post-filter estimation for noise reduction in multichannel speech processing*. Στο *Proc. 14th European Signal Processing Conf. (EUSIPCO-06)*, 2006.

[LM07] Lefkimmiatis, S. και P. Maragos: *A generalized estimation approach for linear and nonlinear microphone array post-filters*. Speech Communication, 49:657–666, 2007.

[LMK08] Lefkimmiatis, S., P. Maragos, και A. Katsamanis: *Multisensor multiband cross-energy tracking for feature extraction and recognition*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, σελίδες 4741–4744, 2008.

[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. Image Process., 18(8), 2009.

[LPH00] Lafertè, J. M., P. Pèrez, και F. Heitz: *Discrete Markov image modeling and inference on the quadtree*. IEEE Trans. Image Process., 9(3):390–404, 2000.

[LPM08] Lefkimmiatis, S., G. Papandreou, και P. Maragos: *Photon-limited image denoising by inference on multiscale models*. Στο *Int. Conf. Image Process. (ICIP)*, σελίδες 2332–2335, 2008.

[LPM09] Lefkimmiatis, S., G. Papandreou, και P. Maragos: *Poisson-Haar transform: A nonlinear multiscale representation for photon-limited image denoising*. Στο *Int. Conf. Image Process. (ICIP)*, 2009.

[Mai94] Mailloux, R. J.: *Phased Array Antenna Handbook*. Artech House, 1994.

[Mal89a] Mallat, S.: *Multiresolution approximations and wavelet orthonormal bases of* $\mathbf{L}^2$. Trans. Amer. Math. Soc., 315(1):69–87, 1989.

[Mal89b] Mallat, S.: *A theory of multiresolution signal decomposition: The wavelet transform*. IEEE Trans. Pattern Anal. Mach. Intell., 11(7):674–693, 1989.

[Mal99] Mallat, S.: *A Wavelet Tour of Signal Processing*. Academic Press, δεύτερη έκδοση, 1999.

[MB95] Maragos, P. και A.C. Bovik: *Image demodulation using multidimensional energy separation*. Journal of the Optical Society of America A, 12(9):1867–1875, 1995.

[MB03] McCowan, I. A. και H. Bourlard: *Microphone array post-filter based on noise field coherence*. IEEE Trans. Speech and Audio Processing, 11(6):709–716, 2003.

[Mey92] Meyer, Y.: *Wavelets and Operators: Advanced Mathematics*. Cambridge University Press, 1992.

[Mey93] Meyer, Y.: *Wavelets: Algorithms and Applications*. SIAM, 1993.

[Mey00] Meyer, C. D.: *Matrix Analysis and Applied Linear Algebra*. Siam, 2000.

[MH92] Mallat, S. και W. Hwang: *Singularity detection and processing with wavelets*. IEEE Trans. Inf. Theory, 38(2):617–643, 1992.

[Min03] Minka, T.: *Estimating a Dirichlet distribution*. Τεχνική αναφορά, Microsoft Research, 2003. http://research.microsoft.com/~minka/papers/dirichlet.

[MKQ93a] Maragos, P., J. F. Kaiser, και T. F. Quatieri: *Energy separation in signal modulations with application to speech analysis*. IEEE Trans. Signal Processing, 41(10):3024–3051, 1993.

[MKQ93b] Maragos, P., J. F. Kaiser, και T. F. Quatieri: *On amplitude and frequency demodulation using energy operators*. IEEE Trans. Signal Processing, 41(4):1532–1550, 1993.

[MKRM99] Mihçak, M.K., I. Kozintsev, K. Ramchandran, και P. Moulin: *Low complexity image denoising based on statistical modeling of wavelet coefficients*. IEEE Signal Process. Lett., 12(6):300–303, 1999.

[ML99] Moulin, P. και J. Liu: *Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors*. IEEE Trans. Inf. Theory, 45(3):909–919, 1999.

[MMS88] Marro, C., Y. Mahieux, και K. U. Simmer: *Analysis of noise reduction techniques based on microphone arrays with postfiltering*. IEEE Trans. Speech and Audio Processing, 6(3):240–259, 1988.

[MP94] MacKay, D. J. C. και L. Peto: *An hierarchical Dirichlet language model*. Natural Language Engineering, 1(3):1–19, 1994.

[MP95] Maragos, P. και A. Potamianos: *Higher order differential energy operators*. IEEE Signal Processing Lett., 2(8):152–154, 1995.

[MQK91] Maragos, P., T. F. Quatieri, και J. F. Kaiser: *Speech nonlinearitites, modulations and energy operators*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, σελίδες 421–424, 1991.

[MS97] Meyer, J. και K. U. Simmer: *Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, τόμος 2, σελίδες 1167–1170, 1997.

[MZ92] Mallat, S. και S. Zhong: *Characterization of signals from multiscale edges*. IEEE Trans. Pattern Anal. Mach. Intell., 14(7):710–732, 1992.

[NB99] Nowak, R. D. και R. G. Baraniuk: *Wavelet-domain filtering for photon imaging systems*. IEEE Trans. Image Process., 8(5):666–678, 1999.

[NK00] Nowak, R. D. και E. D. Kolaczyk: *A statistical multiscale framework for Poisson inverse problems*. IEEE Trans. Inf. Theory, 46(5):1811–1825, 2000.

[Now99a] Nowak, R.: *Multiscale hidden Markov models for Bayesian image analysis*. Στο Vidacovic, B. και P. Muller (επιμελητές): *Bayesian Inference in Wavelet Based Models*. Springer Verlag, 1999.

[Now99b] Nowak, R.: *Multiscale hidden Markov models for photon-limited imaging*. Στο *Proc. SPIE Conf. 3816*, σελίδες 321–332, 1999.

[OMS01] Omologo, M., M. Matassoni, και P. Svaizer: *Speech recognition with microphone arrays*. Στο Brandstein, M. και D. Ward (επιμελητές): *Microphone Arrays: Signal Processing Techniques and Applications*, κεφάλαιο 15, σελίδες 331–353. Springer Verlag, 2001.

[OMSG97] Omologo, M., M. Matassoni, P. Svaizer, και D. Giuliani: *Microphone array based speech recognition with different talker-array positions*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, σελίδες 227–230, 1997.

[OR94] Orchard, M. T. και K. Ramchandran: *An investigation of wavelet-based image coding using an entropy-constrained quantization framework*. Στο *Proc. Data Compression Conf.*, σελίδες 341–350, 1994.

[OSM98] Omologo, M., P. Svaizer, και M. Matassoni: *Environmental conditions and acoustic transduction in hands-free speech recognition*. Speech Communication, 25:75–95, 1998.

[Pap91] Papoulis, A.: *Probability, Random Variables, and Stochastic Processes*. International Edition, McGraw-Hill, New York, 1991.

[PM99a] Potamianos, A. και P. Maragos: *Speech analysis and synthesis using an AM-FM modulation model*. Speech Communication, 28:195–209, 1999.

[PM99b] Potamianos, A. και P. Maragos: *Speech formant frequency and bandwidth tracking using multiband energy demodulation*. J. Acoust. Soc. Amer., 9:3795–3806, 1999.

[PM01] Potamianos, A. και P. Maragos: *Time-frequency distributions for automatic speech recognition*. IEEE Trans. Speech and Audio Processing, 9:196–200, 2001.

[PM08] Papandreou, G. και P. Maragos: *Image inpainting with a wavelet domain hidden Markov tree model*. Στο *Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, σελίδες 773–776, 2008.

[Poo98] Poor, H. V.: *An Introduction to Signal Detection and Estimation*. Springer Verlag, 1998.

[Pot96] Potamianos, A.: *Speech Processing Applications Using an AM-FM Modulation Model*. Διδακτορική διατριβή, Harvard University, 1996.

[PSWS03] Portilla, J., V. Strela, M. 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.

[PTVF07] Press, W. H., S. A. Teukolsky, W. T. Vetterling, και B. P. Flannery: *Numerical Recipes*. Cambridge, τρίτη έκδοση, 2007.

[Rab89] Rabiner, L. R.: *A tutorial on hidden Markov models and selected applications in speech recognition*. Proc. IEEE, 77:257–286, 1989.

[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, 2001.

[RJ93] Rabiner, L. R και B. W. Juang: *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

[RRCB00] Romberg, J., R. Riedi, H. Choi, και R. G. Baraniuk: *Multiplicative multiscale image decompositions: Analysis and modeling*. Στο *SPIE's 45th Annual Meeting, Internat. Symp. on Optical Science and Technology*, 2000.

[RS78] Rabiner, L. R και R. W. Schafer: *Digital Signal Processing of Speech Signals*. Prentice Hall, 1978.

[RS07] Rabiner, L. R και R. W. Schafer: *Introduction to Digital Speech Processing*. Now Publishers, 2007.

[SA90] Simoncelli, E. P. και E. H. Adelson: *Nonseparable extensions of quadrature mirror filters to multiple dimensions*. Proc. IEEE, 78(4):652–664, 1990.

[SBK05] Selesnick, I.W., R.G. Baraniuk, και N.G. Kingsbury: *The dual-tree complex wavelet transform*. IEEE Signal Process. Mag., 22:123–151, 2005.

[SBM01] Simmer, K. U., J. Bitzer, και C. Marro: *Post-filtering techniques*. Στο Brandstein, M. και D. Ward (επιμελητές): *Microphone Arrays: Signal Processing Techniques and Applications*, κεφάλαιο 3, σελίδες 39–60. Springer Verlag, 2001.

[SCD02] Starck, J. L., E. Candès, και D. Donoho: *The curvelet transform for image denoising*. IEEE Trans. Image Process., 11(6):131–141, 2002.

[SF95] Simoncelli, E. P. και W. T. Freeman: *The steerable pyramid: A flexible architecture for multi-scale derivative computation*. Στο *Int. Conf. Image Process. (ICIP)*, σελίδες 444–447, 1995.

[SF07] Starck, J. L. και J. Fadilli: *The undecimated wavelet decomposition and its reconstruction*. IEEE Trans. Image Process., 16(2):297–309, 2007.

[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, 1992.

[SH93] Snyder, D. L και A. M. Hammoud: *Image recovery from data acquired with a charge-coupled-device camera*. J. Opt. Soc. Am. A, 10(5):1014–1023, 1993.

[Sha93] Shapiro, J. M.: *Embedded image coding using zerotrees of wavelet coefficients*. IEEE Trans. Signal Process., 41(12):3445–3462, 1993.

[She92] Shensa, M. J.: *Discrete wavelet transforms: Wedding the à trous and Mallat algorithms*. IEEE Trans. Signal Process., 40(9):2464–2482, 1992.

[SKB+96] Sjölander, K., K. Karplus, M. Brown, R. Hughey, A. Krogh, I. S. Mian, και D. Haussler: *Dirichlet mixtures: A method for improved detection of weak but significant protein sequence homology*. Comput. Appl. Biosci., 12(4):327–345, 1996.

[SM06] Starck, J. L. και F. Murtagh: *Astronomical Image and Data Analysis*. Springer, δεύτερη έκδοση, 2006.

[SMY07] Sadamitsu, Kugatsu, Takuya Mishina, και Mikio Yamamoto: *Topic-based language models using Dirichlet mixtures*. Systems and Computers in Japan, 38(12):1771–1779, 2007.

[SN96] Strang, G. και T. Nguyen: *Wavelets and FilterBanks*. Wellesley-Cambridge Press, 1996.

[SS02a] Sendur, L. και I.W. Selesnick: *Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency*. IEEE Trans. Signal Process., 50(11):2744–2756, 2002.

[SS02b] Sendur, L. και I.W. Selesnick: *Bivariate shrinkage with local variance estimation*. IEEE Signal Process. Lett., 9(12):438–441, 2002.

[Sul96] Sullivan, Tom: *CMU microphone array database*, 1996. `http://www.speech.cs.cmu.edu/databases/micarray`.

[SV82] Shepp, L. A. και Y. Vardi: *Maximum likelihood reconstruction for emission tomography*. IEEE Trans. Med. Imag., 1:113–122, 1982.

[TN99] Timmerman, K. και R. Nowak: *Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging*. IEEE Trans. Inf. Theory, 45(3):846–862, 1999.

[Tre02] Trees, H. L. Van: *Optimum Array Processing*. Wiley, 2002.

[TT90] Teager, H. M. και S. M. Teager: *Evidence of nonlinear sound production mechanisms in the vocal tract*. Στο *Speech Production and Speech Modelling*, σελίδες 241–261. Kluwer Academic, 1990.

[TV68] Tweedie, M. C. K. και A. Veevers: *The inversion of cumulant operators for power-series distributions, and the approximate stabilization of variance by transformations*. J. Amer. Stat. Assoc., 63(321):321–328, 1968.

[Vai93] Vaidyanathan, P. P.: *Multirate Systems and FilterBanks*. Prentice Hall, 1993.

[Var85] Vary, P.: *Noise suppresion by spectral magnitude estimation - mechanism and theoritical limits*. Signal Processing, 8(4):387–400, 1985.

[VB88] Veen, B. D. Van και K. M. Buckley: *Beamforming: A versatile approach to spatial filtering*. IEEE ASSP Magazine, 5:4–24, 1988.

[VK95] Vetterli, M. και J. Kovacevic: *Wavelets and Subband Coding*. Prentice Hall, 1995.

[VT71] Veevers, A. και M. C. K. Tweedie: *Variance-stabilizing transformation of a poisson variate by a beta function*. Appl. Statist., 20(3):304–308, 1971.

[Wil02] Willsky, A.: *Multiresolution Markov models for signal and image processing*. Proc. IEEE, 90(8):1396–1458, 2002.

[WN03] Willet, R. και R.D Nowak: *A multiscale approach for recovering edges and surfaces in photon-limited medical imaging*. IEEE Trans. Inf. Theory, 22, 2003.

[WN07a] Willet, R. και R.D Nowak: *Multiscale analysis of photon-limited astronomical images*. Statistical Challenges in Modern Astronomy (SCMA) IV, 2007.

[WN07b] Willet, R. και R.D Nowak: *Multiscale poisson intensity and density estimation*. IEEE Trans. Inf. Theory, 53:3171–3187, 2007.

[WSW01] Wainwright, M. J., E. P. Simoncelli, και A. Willsky: *Random cascades on wavelet trees and their use in analyzing and modeling natural images*. Appl. and Comput. Harm. Anal., 11:89–123, 2001.

[WUM] Division of Nuclear Medicine, Washington University School of Medicine, St. Louis, MO, USA. `http://gamma.wustl.edu/`.

[YEK+02] Young, S., G. Evermann, D. Kersgaw, D. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, και P. Woodland: *The HTK Book (for HTK Version 3.2)*. Cambridge Research Lab: Entropics, 2002.

[Zel88] Zelinski, R.: *A microphone array with adaptive post-filtering for noise reduction in reverberant rooms*. Στο *Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, τόμος 5, σελίδες 2578–2581, 1988.

[ZFS08] Zhang, B., M.J. Fadili, και J. L. Starck: *Wavelets, ridgelets and curvelets for Poisson noise removal*. IEEE Trans. Image Process., 17(7):1093–1108, 2008.

[Zio95] Ziomek, L. J.: *Fundamental of Acoustic Field Theory and Space-time Signal Processing*. CRC Press, 1995.

Βιογραφικό Σημείωμα

Ο Λευκιμμιάτης Σταμάτης γεννήθηκε στην Αθήνα το 1980 όπου και ολοκλήρωσε τις εγκύκλιες σπουδές του. Το 1999 εισήχθη στην Πολυτεχνική Σχολή Μηχανικών Ηλεκτρονικών Υπολογιστών & Πληροφορικής του Πανεπιστημίου Πατρών. Τον Ιούλιο του 2004 έλαβε το δίπλωμα του Μηχανικού Ηλεκτρονικών Υπολογιστών & Πληροφορικής με βαθμό άριστα (8,62/10) και κατετάγη στο άνω 1% της τάξης. Η διπλωματική του εργασία είχε ως αντικείμενο την Επεξεργασία Σημάτων και Ψηφιακών Επικοινωνιών με κύριο θέμα την ισοστάθμιση σε συστήματα διακριτής πολυτονικής διαμόρφωσης. Κατά τη διάρκεια των ετών 2004-2009 συνέχισε τις σπουδές του ως Υποψήφιος Διδάκτωρ της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσοβίου Πολυτεχνείου στον τομέα Σημάτων, Ελέγχου και Ρομποτικής αποτελώντας μέλος της ερευνητικής ομάδας Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων. Η διατριβή του με τίτλο "Μέθοδοι Πολυμεσικών Σημάτων για Βέλτιστη Αποθορυβοποίηση και Ανίχνευση με Εφαρμογές σε Δορυφορικά Δεδομένα" επικεντρώθηκε στη στατιστική ανάλυση και μοντελοποίηση σημάτων και εικόνων με στόχο την ανάπτυξη νέων τεχνικών ψηφιακής επεξεργασίας. Η διατριβή του εντάχθηκε στο ερευνητικό πρόγραμμα της ΓΓΕΤ, ΠΕΝΕΔ-2003 στο πεδίο "Επεξεργασία Σημάτων και Εικόνων με Εφαρμογές σε Διαστημικές και Δορυφορικές Επικοινωνίες", ενώ, παράλληλα, ο ίδιος συμμετείχε ως ερευνητής του ΕΠΙΣΕΥ σε ερευνητικά προγράμματα της Ευρωπαϊκής Ένωσης με αντικείμενο την πολυαισθητηριακή επεξεργασία ακουστικών σημάτων και την αναγνώριση φωνής. Την άνοιξη του 2009 επισκέφθηκε το Ερευνητικό Ινστιτούτο Διαστημικής MSSL όπου συνεργάσθηκε με τον καθηγητή Peter Muller σε θέματα ανάλυσης και επεξεργασίας διαστημικών δορυφορικών εικόνων. Τα ερευνητικά του ενδιαφέροντα εντοπίζονται στη γενικότερη περιοχή της ανάλυσης εικόνων και επεξεργασίας σημάτων και ειδικότερα στα πεδία της μη-γραμμικής επεξεργασίας σημάτων, ανάλυσης ήχου και εικόνας, στατιστικής μοντελοποίησης εικόνων και αναγνώρισης προτύπων με εφαρμογές στην Επεξεργασία Δορυφορικών και Βιοϊατρικών Εικόνων, την Επεξεργασία Πολυαισθητηριακών Σημάτων και την Αναγνώριση Φωνής. Από το 2005 είναι μέλος του Τεχνικού Επιμελητηρίου Ελλάδος (ΤΕΕ) και από το 2008 μέλος του διεθνούς οργανισμού IEEE. Έχει δημοσιεύσει 8 άρθρα σε διεθνή περιοδικά και συνέδρια με κριτές, ενώ είναι τακτικός κριτής σε περιοδικά των οργανισμών IEEE και Eurasip.