



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Μη-γραμμική Διάχυση στην Όραση Υπολογιστών
και Στατιστικά Μοντέλα Σχήματος με Εφαρμογές
στην Ανάλυση Εικόνων Αρθρωτών Φωνητικού και
Νοηματικού Λόγου**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΑΝΑΣΤΑΣΙΟΣ Ν. ΡΟΥΣΣΟΣ

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

Αθήνα, Οκτώβριος 2010



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Μη-γραμμική Διάχυση στην Όραση Υπολογιστών
και Στατιστικά Μοντέλα Σχήματος με Εφαρμογές
στην Ανάλυση Εικόνων Αρθρωτών Φωνητικού και
Νοηματικού Λόγου**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΑΝΑΣΤΑΣΙΟΣ Ν. ΡΟΥΣΣΟΣ

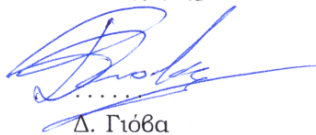
Διπλωματούχος Ηλεκτρολόγος Μηχανικός · Μηχανικός Υπολογιστών Ε.Μ.Π.

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

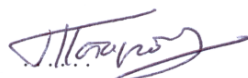
Εγκρίθηκε από την επταμελή εξεταστική επιτροπή στις ...06/10/ 2010:



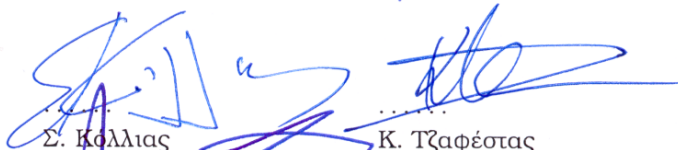
.....
Π. Μαραγκός
Καθηγητής Ε.Μ.Π.



.....
Δ. Γιόβας
Καθηγήτρια Ε.Μ.Π.



Γ. Ποταμιάνος
Ερευνητής Α',
ΕΚΕΦΕ ΔΗΜΟΚΡΙΤΟΣ



.....
Σ. Κόλλιας
Καθηγητής Ε.Μ.Π.



.....
Ν. Σταυρακάκης
Καθηγητής Ε.Μ.Π.



.....
Ι. Κόκκινος
Επίκουρος Καθηγητής Ecole Centrale Paris

Αθήνα, Οκτώβριος 2010

...

ΑΝΑΣΤΑΣΙΟΣ Ν. ΡΟΥΣΣΟΣ

Υποψήφιος Διδάκτορας Ηλεκτρολόγος Μηχανικός · Μηχανικός Ηλεκτρονικών Υπολογιστών Ε.Μ.Π.

Copyright © Αναστασιος Ν. Ρουσσος, 2010.

Με επιφύλαξη παντός δικαιώματος. All rights reserved

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

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

Περιεχόμενα

1	Εισαγωγή	1
1.1	Χρήση Μερικών Διαφορικών Εξισώσεων και Στατιστικών Μοντέλων Σχήματος	1
1.2	Κύριες Εφαρμογές	3
1.2.1	Έρευνα του Ανθρώπινου Συστήματος Παραγωγής Φωνής	3
1.2.2	Αυτόματη Αναγνώριση Νοηματικής Γλώσσας	9
1.3	Ερευνητικές Συνεισφορές	11
2	Επισκόπηση Ερευνητικών Περιοχών	17
2.1	Μερικές Διαφορικές Εξισώσεις	17
2.1.1	Χώροι Κλίμακας	18
2.1.2	Μη-γραμμική Διάχυση	20
2.1.3	Κατάτμηση και Ενεργά Περιγράμματα	22
2.2	Στατιστικά Μοντέλα Σχήματος	23
2.3	Παρακολούθηση της Γλώσσας Ομιλητή σε Βίντεο Υπερήχων	24
2.4	Οπτική Επεξεργασία Βίντεο Νοηματικής Γλώσσας	24
3	Μοντέλο Μη-γραμμικής Διάχυσης για την Αντιστρέψιμη Παρεμβολή Εικόνων	27
3.1	Εισαγωγή	27
3.2	Προκαταρκτικά Θέματα και Υπόβαθρο	28
3.2.1	Συνθήκη Αντιστρεψιμότητας για την Παρεμβολή	28
3.2.2	Συνθήκη Αντιστρεψιμότητας στο Πεδίο των Συχνοτήτων	30
3.2.3	Παρεμβολή Μηδενικού Γεμίσιματος Συχνοτήτων	30
3.2.4	Παρεμβολή Βασισμένη στην Ολική Μεταβολή (ΟΜ)	31
3.2.5	Μέθοδος Belahmidi-Guichard (BG)	32
3.2.6	Μοντέλο ΜΔΕ των Tschumperlé-Deriche (TD)	32
3.3	Προτεινόμενη ΜΔΕ Ανισοτροπικής Διάχυσης-Προβολής	34
3.3.1	Συνθήκη Αντιστρεψιμότητας για το Προτεινόμενο Μοντέλο	35
3.3.2	ΜΔΕ του Προτεινόμενου Μοντέλου	36
3.3.3	Έκφραση του Τελεστή Προβολής	36
3.3.4	Ιδιότητες του Μοντέλου	37
3.3.5	Αριθμητική Υλοποίηση	38
3.4	Πειραματικά Αποτελέσματα και Συγκρίσεις	39
3.4.1	Πειραματικό Πλαίσιο	39
3.4.2	Παραδείγματα από τα Πειραματικά Αποτελέσματα	41
3.4.3	Συγκεντρωτικές Μετρήσεις Απόδοσης	44

3.4.4	Παρεμβολή Βιοϊατρικών Εικόνων	45
3.5	Θεωρητικές Λεπτομέρειες	46
3.6	Συμπεράσματα	50
4	Μέθοδοι Διάχυσης με Τανυστές Εξαγμένες από Μεταβολικά Πλαίσια	51
4.1	Εισαγωγή	51
4.2	Γενίκευση του Συναρτησιακού Beltrami Βασισμένη στα Τεμάχια Εικόνας . .	52
4.3	Γενικευμένο Συναρτησιακό Βασισμένο στον Δομικό Τανυστή	53
4.4	Νέες Μέθοδοι σαν Ειδικές Περιπτώσεις του Προτεινόμενου Συναρτησιακού .	54
4.4.1	Τανυστική Ολική Μεταβολή	54
4.4.2	Γενικευμένη Ροή Beltrami	54
4.4.3	Άλλες Ενδιαφέρουσες Περιπτώσεις και Σχέσεις με Προηγούμενες Ερ- γασίες	55
4.5	Πειραματικά Αποτελέσματα και Συγκρίσεις	55
4.6	Θεωρητικές Λεπτομέρειες	58
4.6.1	Εξαγωγή του Γενικευμένου Συναρτησιακού Beltrami	58
4.6.2	Περίγραμμα της Απόδειξης του Θεωρήματος 4.3.1	60
4.7	Συμπεράσματα	61
5	Παρακολούθηση της Γλώσσας σε Εικόνες Υπερήχων με Ενεργά Μοντέλα Εμφάνισης	63
5.1	Εισαγωγή	63
5.2	Προκαταρκτικά	64
5.2.1	Σύστημα Καταγραφής Δεδομένων	64
5.2.2	Χρήση των Ακτίνων Χ για την Μοντελοποίηση Σχήματος	64
5.2.3	Εκτίμηση της Πόζας του Πλέγματος Φωνητικής Οδού	65
5.3	Προεπεξεργασία των Πλαισίων Υπερήχων	65
5.4	Ενεργό Μοντέλο Παρακολούθησης της Γλώσσας	68
5.4.1	Αναπαράσταση της Εμφάνισης της Γλώσσας στα Πλαίσια Υπερήχων .	68
5.4.2	Μοντελοποίηση της Μεταβολής του Σχήματος	69
5.4.3	Μοντελοποίηση της Μεταβολής της Υψής	70
5.4.4	Παρακολούθηση μέσω του Ταίριασματος του Μοντέλου	71
5.5	Πειραματικά Αποτελέσματα	71
6	Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας	75
6.1	Εισαγωγή	75
6.2	Αρχική Παρακολούθηση Χεριών και Κεφαλιού Βασισμένη σε Οπτική Προεπεξεργασία	76
6.2.1	Κατάτμηση και Εντοπισμός Δέρματος	77
6.2.2	Παρακολούθηση και Διαχείριση Αλληλεπικαλύψεων	80
6.2.3	Στατιστικός Προσδιορισμός των Παραμέτρων.	80
6.3	Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος-Εμφάνισης	81
6.3.1	Αναπαράσταση Χεριών με Εικόνες Σχήματος-Εμφάνισης	81
6.3.2	Μοντελοποίηση των Εικόνων Σχήματος-Εμφάνισης	82
6.3.3	Εκπαίδευση του Γραμμικού Συνδυασμού του ΜΣΕ	84
6.3.4	Άμεσο Ταίριασμα του Μοντέλου	89

6.3.5 Ομαλοποιημένο Ταίριασμα του ΜΣΕ με Στατική και Δυναμική Πρότερη Πληροφορία	90
6.4 Πειράματα Ταξινόμησης Χειρομορφών	98
6.4.1 Δεδομένα, Επισημείωση και Κλάσεις Χειρομορφών	98
6.4.2 Οργάνωση Πειραμάτων	99
6.4.3 Συγκρινόμενες Μέθοδοι Εξαγωγής Χαρακτηριστικών Χειρομορφών . .	100
6.4.4 Αποτελέσματα και Συγκρίσεις	101
7 Σύνοψη Προόδου και Κατευθύνσεις Μελλοντικής Έρευνας	105
7.1 Ερευνητική Συνεισφορά	105
7.2 Κατευθύνσεις Μελλοντικής Έρευνας	106
Βιβλιογραφία	108
Α' Κατάλογος Δημοσιεύσεων του Συγγραφέα	121

Κατάλογος Σχημάτων

1.1	Μεσο-οβελιαία όψη της φωνητικής οδού.	4
1.2	Πλαίσια από στερεοσκοπική βιντεοσκόπηση του προσώπου ομιλητή κατά το πρωτόκολλο του ASPI.	6
1.3	Απεικόνιση υπερήχων της γλώσσας ομιλητή.	7
1.4	Ακολουθίες ακτίνων X για την απεικόνιση της φωνητικής οδού ομιλητή. . . .	7
1.5	Απεικόνιση MRI της φωνητικής οδού.	8
1.6	Απεικόνιση δυναμικών MRI.	8
1.7	Αλληλεπικαλύψεις και αλλαγές στην διαμόρφωση των χεριών κατά τον νοηματισμό.	10
1.8	Μεταβλητότητα στη διαμόρφωση και την πόζα του κυρίαρχου χεριού.	10
1.9	Παρεμβολή (4×4) μιας MRI εικόνας της φωνητικής οδού με την μέθοδο που εισάγαμε.	12
1.10	Αποθορυβοποίηση μιας ακολουθίας εικόνων ακτίνων X της φωνητικής οδού ομιλητή χρησιμοποιώντας την προτεινόμενη Τανυστική ΟΜ.	13
1.11	Προσαρμοστικό φιλτράρισμα εικόνων υπερήχων της γλώσσας ομιλητή, με την μέθοδο που προτείναμε στο [7].	13
1.12	Παρακολούθηση της καμπύλης της γλώσσας σε βίντεο υπερήχων με το μοντέλο που εισάγαμε στο [100].	14
1.13	Ταίριασμα του προτεινόμενου μοντέλου χειρομορφών σε ένα τμήμα βίντεο συνεχούς νοηματισμού.	14
1.14	Πειράματα ταξινόμησης χειρομορφών.	15
2.1	Χώροι κλίμακας μιας εικόνας ψαριού.	19
3.1	Συνθήκη αντιστρεψιμότητας για την άγνωστη λύση $u(x)$ της παρεμβολής. . .	29
3.2	Εικόνες αναφοράς από το χρησιμοποιούμενο σύνολο εικόνων.	40
3.3	Λεπτομέρειες των εικόνων αναφοράς.	40
3.4	Αποτελέσματα 4×4 παρεμβολής για την εικόνα αναφοράς #7.	42
3.5	Σύγκριση των δύο εκδοχών της προτεινόμενης μεθόδου παρεμβολής.	43
3.6	Αποτελέσματα 4×4 παρεμβολής για την εικόνα αναφοράς #5.	43
3.7	Εξέλιξη της προτεινόμενης ΜΔΕ για την 4×4 παρεμβολή μίας έγχρωμης εικόνας.	45
3.8	Μετρήσεις κατά τη διάρκεια της εξέλιξης της προτεινόμενης ΜΔΕ παρεμβολής. . .	46
3.9	Συγκεντρωτικές μετρήσεις απόδοσης των μεθόδων παρεμβολής.	47
3.10	Παρεμβολή (4×4) μιας βιοϊατρικής εικόνας της φωνητικής οδού.	48
3.11	Εξαγωγή του $\hat{p}_{m_1, m_2}^{(n_1, n_2)}$	49

4.1	Εφαρμογή της Τανυστικής ΟΜ σε μια ακολουθία εικόνων ακτίνων X της φωνητικής οδού ομιλητή.	56
4.2	Αποθορυβοποίηση χρησιμοποιώντας την ΟΜ και την Τανυστική ΟΜ.	57
4.3	Αποθορυβοποίηση μιας έγχρωμης εικόνας με την κλασική και γενικευμένη Ροή Beltrami.	58
4.4	Συγκεντρωτικές μετρήσεις PSNR για διαφορετικές μεθόδους διάχυσης.	58
5.1	Εικόνες υπερήχων και ακτίνων X ομιλητή, με υπέρθεση του ευθυγραμμισμένου πλέγματος της φωνητικής οδού.	64
5.2	Κατασκευή του πυρήνα $k(r, \phi; r', \phi')$, ο οποίος συσχετίζεται με την εικόνα υπερήχων.	65
5.3	Προσαρμοστικό φιλτράρισμα ακολουθίας εικόνων υπερήχων με την προτεινόμενη μέθοδο.	67
5.4	Προτεινόμενη αναπαράσταση της εμφάνισης της γλωσσικής καμπύλης.	68
5.5	Εφαρμογή της προτεινόμενης μεθόδου για την ανίχνευση της γλώσσας.	72
5.6	Σύγκριση των αποτελεσμάτων των μεθόδων ανίχνευσης γλώσσας.	72
5.7	Ποσοτικές συγκρίσεις μεθόδων ανίχνευσης της γλώσσας ομιλητή.	73
6.1	Επισκόπηση της εξόδου του αρχικού υποσυστήματος παρακολούθησης των χεριών και του κεφαλιού.	77
6.2	Μοντελοποίηση χρώματος δέρματος.	78
6.3	Ενδεικτικά αποτελέσματα του συστήματος εξαγωγής μάσκας δέρματος και μορφολογικής κατάταμής της.	79
6.4	Αρχική παρακολούθηση των χεριών και του κεφαλιού νοηματιστή.	81
6.5	Δείγματα από το σύνολο εκπαίδευσης του Μοντέλου Σχήματος-Εμφάνισης (ΜΣΕ) του χεριού νοηματιστή.	83
6.6	1-1 αφινική ευθυγράμμιση δύο εικόνων Σχήματος-Εμφάνισης (ΣΕ).	86
6.7	Ημιαυτόματη διαδικασία αφινικής ευθυγράμμισης ενός συνόλου εκπαίδευσης εικόνων Σχήματος-Εμφάνισης.	88
6.8	Εκμάθηση των εικόνων του γραμμικού συνδυασμού του ΜΣΕ μέσω της εφαρμογής PCA.	89
6.9	Άμεσο ταίριασμα του Μοντέλου Σχήματος-Εμφάνισης.	91
6.10	Μήτρες στάθμισης που χρησιμοποιούνται για την πρόβλεψη των παραμέτρων $\lambda(n)$	94
6.11	Πλαίσια δοκιμής πρόβλεψης για την εκπαίδευση των πρότερων δυναμικών πληροφοριών σχετικά με το λ	95
6.12	Ομαλοποιημένο ταίριασμα του ΜΣΕ.	97
6.13	Τάξεις χειρομορφών για το Πείραμα-A ταξινόμησης χειρομορφών.	99
6.14	Τάξεις χειρομορφών για το Πείραμα-B ταξινόμησης χειρομορφών.	100
6.15	Γεωμετρικά χαρακτηριστικά περιοχής (ΓΧΠ) του χεριού [1].	102
6.16	Αποτελέσματα ταξινόμησης χειρομορφών.	103

ΠΕΡΙΛΗΨΗ

Στην παρούσα έρευνα, αναπτύσσονται μοντέλα Μερικών Διαφορικών Εξισώσεων τύπου Μη-γραμμικής Διάχυσης για την απλοποίηση και την βελτίωση της ποιότητας εικόνων. Χρησιμοποιούνται επίσης μεθοδολογίες Στατιστικών Μοντέλων Σχήματος για τον οπτικό εντοπισμό και αναγνώριση αντικειμένων. Οι κύριες εφαρμογές που μελετώνται αφορούν την ανάλυση οπτικών δεδομένων των αρθρωτών λόγου, σε δύο διαφορετικές μορφές του : τον φωνητικό λόγο, του οποίου οι αρθρωτές βρίσκονται στην φωνητική οδό, και τον νοηματικό λόγο, του οποίου οι αρθρωτές είναι κυρίως τα χέρια, τα χείλη και τα μάτια. Η ανάλυση τέτοιων οπτικών δεδομένων συνεισφέρει στην έρευνα και την τεχνολογία που σχετίζονται με την ομιλία και την νοηματική γλώσσα.

Πιο συγκεκριμένα, σχεδιάζεται μία μέθοδος μη-γραμμικής διάχυσης για την βελτίωση της ευκρίνειας διανυσματικών εικόνων, η οποία αποδίδει αποτελέσματα υψηλής ακρίβειας, με περιορισμένα ανεπιθύμητα ψεύδεργα (artifacts). Επίσης, εισάγεται ένα θεωρητικό πλαίσιο το οποίο βασίζεται στον δομικό τανυστή της εικόνας και γενικεύει διάφορες μεταβολικές μεθόδους μη-γραμμικής διάχυσης για την αποκατάσταση εικόνων. Με βάση το πλαίσιο αυτό, προτείνονται νέες μέθοδοι διάχυσης, οι οποίες συνδυάζουν τα πλεονεκτήματα διαφόρων προηγούμενων προσεγγίσεων.

Ακόμα, σχεδιάζεται ένα στατιστικό μοντέλο για την καμπύλη της γλώσσας ομιλητή, το οποίο ενσωματώνει πρότερη γνώση σχήματος και μπορεί να εκτιμήσει ολόκληρη την καμπύλη της γλώσσας σε ακολουθίες εικόνων υπερήχων. Τέλος, παρουσιάζεται μία νέα μοντελοποίηση χειρομορφών νοηματικής γλώσσας, η οποία προσφέρει μία περιγραφική και συμπαγή αναπαράσταση των διαμορφώσεων του χεριού του νοηματιστή και έχει εφαρμοστεί με επιτυχία στην ταξινόμηση χειρομορφών. Αναπτύσσεται μία διαδικασία εύρωστου ταιριάσματος του μοντέλου αυτού σε βίντεο συνεχούς νοηματισμού, ενσωματώνοντας πρότερη στατική και δυναμική πληροφορία σχετικά με τις χειρομορφές.

ABSTRACT

In this work, Nonlinear Diffusion models for image simplification and image restoration are developed. In addition, the framework of Statistical Shape Models is used for object detection and recognition. The main applications that are studied are related to the image analysis of speech articulators, for two different speech types: voiced speech, whose articulators are located in the vocal tract, and signed speech, whose main articulators are the hands, lips and eyes. This kind of image analysis contributes to the voiced and signed speech research and technology.

More precisely, we design a nonlinear diffusion method for vector-valued image interpolation, which yields accurate results with reduced artifacts. In addition, we introduce a theoretical framework that is based on the image structure tensor and generalizes several variational methods of nonlinear diffusion for image restoration. Based on this framework, we propose some new diffusion methods that combine the advantages of various existing approaches.

Furthermore, we design a statistical model for the tongue contour of a speaker, model which incorporates prior shape knowledge and is able to robustly estimate the tongue contour in ultrasound image sequences. Finally, we present a new modeling of sign language handshapes, which offers a descriptive and compact representation of hand configurations and has been successfully applied on the handshape classification. We develop a robust model fitting process in continuous signing videos, by incorporating static and dynamic prior handshape information.

ΠΡΟΛΟΓΟΣ

Η διατριβή που κρατάτε στα χέρια σας ή βλέπετε στην οθόνη του υπολογιστή σας είναι το αποτέλεσμα της διδακτορικής μου έρευνας στο εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων του Εθνικού Μετσόβιου Πολυτεχνείου, από τον Οκτώβριο του 2005 έως τον Οκτώβριο του 2010. Βλέποντας αυτή την προσπάθεια ολοκληρωμένη αισθάνομαι ικανοποίηση και ανακούφιση. Με τον τρόπο αυτό κλείνει ένας κύκλος κοπιαστικής δουλειάς και δημιουργικής περιπλάνησης στον κόσμο της έρευνας. Το σίγουρο είναι ότι θα θυμάμαι αυτή την πορεία με νοσταλγία. Παρόλο που για την διδακτορική μου έρευνα αφιέρωσα ένα σημαντικό κομμάτι του εαυτού μου, η διατριβή αυτή δεν μπορεί παρά να αποτελεί μόνο μία προβολή αυτής της ερευνητικής διαδικασίας, η οποία με την σειρά της ήταν μία από τις πτυχές του ταξιδιού μου κατά τα πέντε αυτά χρόνια, τα τόσο πλούσια σε προσωπικά και όχι μόνο γεγονότα. Θεωρώ όμως πιο σημαντικό να μιλήσω εδώ για τους ανθρώπους οι οποίοι συνεισέφεραν στο να φτάσω σε αυτό το σημείο.

Πρώτα από όλα, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντά μου, Καθ. *Πέτρο Μαραγκό*, για όλα όσα μου προσέφερε κατά τη διάρκεια της πολυετούς συνεργασίας μας. Η συνεργασία και η αλληλεπίδραση μαζί του υπήρξε για μένα μία πολύ σημαντική εμπειρία. Η επιστημονική του διαδρομή καθώς και η ευρύτητα των επιστημονικών του γνώσεων και ενδιαφερόντων λειτούργησαν σαν έμπνευση για μένα. Τον ευχαριστώ επίσης για την αμέριστη στήριξή του και τις πολύτιμες συμβουλές του σχετικά με τα επόμενα βήματά μου.

Θέλω ακόμα να ευχαριστήσω τα υπόλοιπα μέλη της τριμελούς συμβουλευτικής επιτροπής μου, Καθ. *Στέφανο Κόλλη* και Καθ. *Κώστα Τζαφέστα*, για τις χρήσιμες συμβουλές και υποδείξεις τους καθ' όλη τη διάρκεια του διδακτορικού μου. Ευχαριστώ επίσης τους Καθ. *Διδώ Γιώβα*, Καθ. *Νίκο Σταυρακάκη*, Δρ. *Γεράσιμο Ποταμίανο* και Καθ. *Ιάσωνα Κόκκινο* για την τιμή που μου έκαναν να συμμετάσχουν στην επταμελή εξεταστική επιτροπή του διδακτορικού μου καθώς και για τα εποικοδομητικά τους σχόλια και παρατηρήσεις.

Θα ήθελα επίσης να ευχαριστήσω τους φίλους, ερευνητικούς συνεργάτες και συνταξιδιώτες *Βασίλη Πιτσικάλη*, *Νάσο Κατσαμάνη*, *Γιώργο Ευαγγελόπουλο*, *Σταύρο Θεοδωράκη*, *Γιώργο Παπανδρέου*, *Σταμάτη Λευκιμμάτη*, *Νάνσυ Ζηλαντίτση*, *Ισίδωρο Ροδομαγουλιάκη* και *Michael Aron*, για την πολύτιμη αλληλεπίδραση που είχα μαζί τους. Για τις συμβουλές, τις δημιουργικές συζητήσεις, την συμπαράσταση, την αλληλεγγύη καθώς και για το εξαιρετικό κλίμα που δημιουργούσαν εντός και εκτός εργαστηρίου. Βρίσκονταν δίπλα μου, πάντα πρόθυμοι να μοιραστούν μαζί μου την χαρά, την απογοήτευση, την ελπίδα, την αγανάκτηση, την ματαίωση και την αισιοδοξία.

Ευχαριστώ ακόμα τους φίλους, συμπότες και συνοδοιπόρους που είχα την τύχη να γνωρίσω στα προπτυχιακά φοιτητικά μου χρόνια, *Χρήστο Ν.*, *Γιάννη Κ.*, *Τάσο Κ.*, *Τάσο*

Ν., Λήδα Ρ., Γιάννη Μ., Πέτρο Ξ., Κατερίνα Α., Κώστα Τ., Ιωάννα Κ., Ειρήνη Γ. και Πόπη Μ., γιατί μου συμπαραστάθηκαν πιστεύοντας σε μένα περισσότερο από ό,τι εγώ καθώς και γιατί υπήρξαν σημαντικό αντίβαρο που συνέβαλε στην επαφή μου με την πραγματικότητα εκτός διδακτορικού. Τους ευχαριστώ θερμά για τα ξενύχτια, τα ποτά, τα γέλια, τις πολύωρες συζητήσεις, τα οράματα και τις αγωνίες που μοιράστηκαν μαζί μου.

Θέλω επίσης να ευχαριστήσω τους αδερφικούς μου φίλους, *Αλέξανδρο Λ., Γιάννη Κ., Δημήτρη Γ., Γιώργο Σ., Νίκο Μ., Μιχάλη Γ. και Κατερίνα Ρ.,* με τους οποίους μας ενώνουν αναμνήσεις και δεσμοί δεκαετιών, γιατί ήταν και είναι πάντα δίπλα μου για να συμμετάσχουν στις χαρούμενες στιγμές μου καθώς και να με υποστηρίξουν στις δύσκολες.

Αν και ο αγώνας για αυτή τη διατριβή άρχισε το 2005, τα εφόδια και οι βάσεις που με έκαναν να αντεπεξέλθω είχαν δημιουργηθεί πολλά χρόνια πριν. Γι' αυτό, χρωστώ αμέτρητη ευγνωμοσύνη στους γονείς μου *Νίκο και Βασιλική,* τον αδερφό μου *Γιάννη,* τους παππούδες και γιαγιάδες μου, *Γιάννη, Τζένη, Τάσο και Ελένη* και τους θείους μου *Δημήτρη και Πόπη.* Ιδιαίτερα για τους γονείς μου, αισθάνομαι ευγνώμων μεταξύ πολλών άλλων γιατί προσπάθησαν να μου εμφυσήσουν τον ενθουσιασμό για την επιστήμη, τον μεθοδικό τρόπο σκέψης, την επιμονή στους στόχους και την αισιόδοξη στάση απέναντι στις δυσκολίες. Επίσης, δεν έχω λόγια να ευχαριστήσω τον αδερφό μου για την πολύτιμη συμπαραστάσή του, την υποστήριξή του και τις συμβουλές του.

Τέλος, ευχαριστώ από τα βάθη της καρδιάς μου την σύντροφό μου *Μαριάνθη* για όλα αυτά που απλόχερα μου χάρισε κατά τη διάρκεια και αυτών των πέντε χρόνων του διδακτορικού μου. Με αστείρευτη κατανόηση, υπομονή και αυταπάτηση με στήριξε ακόμα και στις πιο δύσκολες στιγμές. Ήταν σταθερά πλάι μου, να με βοηθήσει να αποδρώ από τον κόσμο του θετικισμού και της βελτιστοποίησης ενεργειών και να μπαίνω σε άλλα μαγικά σύμπαντα, χωρίς νόρμες και μετρικά.

Τάσος Ρούσσο

Κεφάλαιο 1

Εισαγωγή

Η Όραση Υπολογιστών (Computer Vision) ασχολείται με την εξαγωγή της συμβολικής πληροφορίας (περιγραφή αντικειμένων) που ενυπάρχει σε μία εικόνα ή σε ένα σύνολο εικόνων. Απώτερος στόχος της είναι η κατασκευή συστημάτων, τα οποία θα αντιλαμβάνονται τον ορατό κόσμο με ανθρώπινη ευφυΐα. Βέβαια, η σύγχρονη Επιστήμη και Τεχνολογία απέχει ακόμα πολύ από τον στόχο αυτό. Η Όραση Υπολογιστών εμφανίστηκε για πρώτη φορά σαν επιστημονικό πεδίο την δεκαετία του 1960 και έκτοτε έχει γνωρίσει μία ραγδαία ανάπτυξη, συνδυάζοντας πολλά επιστημονικά πεδία.

Η Όραση Υπολογιστών παρουσιάζει μία μεγάλη ποικιλία εφαρμογών, μερικές εκ των οποίων είναι η Επεξεργασία πληροφορίας σε Εικόνες, η Βϊοιατρική Έρευνα και Τεχνολογία, το Περιβάλλον και οι Γεωεπιστήμες, η Τηλεπισκόπηση, η Ρομποτική, η Επικοινωνία Ανθρώπου - Μηχανής, οι Ψηφιακές Εικαστικές Τέχνες καθώς και τα Ευφυή Συστήματα Ψηφιοποίησης και Ανάλυσης Βιβλίων.

Στην διδακτορική μας έρευνα, αναπτύσσουμε μοντέλα Μερικών Διαφορικών Εξισώσεων (ΜΔΕ) τύπου μη-γραμμικής διάχυσης για απλοποίηση και βελτίωση της ποιότητας γενικών διανυσματικών εικόνων. Χρησιμοποιούμε επίσης μεθοδολογίες Στατιστικών Μοντέλων Σχήματος για τον οπτικό εντοπισμό και αναγνώριση αντικειμένων κάνοντας χρήση πρότερης πληροφορίας σχήματος. Όσον αφορά τις εφαρμογές, ασχολούμαστε κυρίως με προβλήματα που σχετίζονται με τη μελέτη του λόγου, για δύο διαφορετικές μορφές του: τον *φωνητικό λόγο* και τον *νοηματικό λόγο*. Πιο συγκεκριμένα, ασχολούμαστε με την ανάλυση οπτικών δεδομένων των *αρθρωτών λόγου*, ανάλυση η οποία συνεισφέρει στην έρευνα και την τεχνολογία που σχετίζονται με τις δύο προαναφερθείσες μορφές λόγου. Αρθρωτές λόγου ονομάζονται τα μέρη του ανθρώπινου σώματος που εκφέρουν τον λόγο: Στον **φωνητικό λόγο** είναι κυρίως τα χείλη, η μύτη, τα δόντια, η γλώσσα, ο ουρανίσκος, ο φάρυγγας, η γλωττίδα και ο λάρυγγας ενώ στον **νοηματικό λόγο** είναι κυρίως τα χέρια, τα χείλη και τα μάτια.

1.1 Χρήση Μερικών Διαφορικών Εξισώσεων και Στατιστικών Μοντέλων Σχήματος

Μία από τις πιο σύγχρονες και ελπιδοφόρες μεθοδολογίες για πολλά προβλήματα της Όρασης Υπολογιστών είναι η χρήση μοντέλων Μερικών Διαφορικών Εξισώσεων (ΜΔΕ). Η μεθοδολογία αυτή άρχισε να χρησιμοποιείται συστηματικά στην Όραση Υπολογιστών

προς τα τέλη της δεκαετίας του 1980 και έχει γίνει πλέον δημοφιλής λόγω διαφόρων πλεονεκτημάτων που παρουσιάζει (βλ. Ενότητα 2.1 για λεπτομέρειες). Μία από τις σημαντικές συνεισφορές της έγκειται στην ανάπτυξη της θεωρίας των *χώρων κλίμακας* (scale spaces), οι οποίοι εισάγουν μία ιεραρχία στις δομές μιας εικόνας μέσω της κατασκευής μιας ακολουθίας από ολοένα και περισσότερο απλοποιημένες αναπαραστάσεις της εικόνας. Κατά τα τελευταία χρόνια, έχουν προταθεί ποικίλοι τύποι χώρων κλίμακας, εκ των οποίων ιδιαίτερα διαδεδομένοι είναι αυτοί που δημιουργούνται από *μη-γραμμική διάχυση* της εικόνας εισόδου.

Πολλά από τέτοιου είδους μοντέλα μη-γραμμικής διάχυσης εφαρμόζονται επίσης και σε άλλες εφαρμογές της Επεξεργασίας Εικόνων και της Όρασης Υπολογιστών, όπως αποθορυβοποίηση, παρεμβολή ή γενικότερα αποκατάσταση εικόνων. Στην κατηγορία των μοντέλων μη-γραμμικής διάχυσης για την παρεμβολή εικόνων ανήκει και η μέθοδος που εισάγαμε στα [101, 102] και παρουσιάζουμε στο Κεφάλαιο 3. Στο μοντέλο αυτό, επιβάλλουμε έναν περιορισμό σχετικά με την αντιστρεψιμότητα της παρεμβολής, ο οποίος εκμεταλλεύεται αποτελεσματικά την διαθέσιμη πληροφορία της εικόνας εισόδου. Η προτεινόμενη διάχυση, η οποία είναι μη-γραμμική αλλά και ανισοτροπική, επιτελεί προσαρμοστική εξομάλυνση ενώ παράλληλα ικανοποιεί τον περιορισμό αντιστρεψιμότητας, χάρη σε έναν κατάλληλο τελεστή προβολής. Σε σύγκριση με τις κλασικές και άλλες υπάρχουσες μεθόδους παρεμβολής με ΜΔΕ, η μέθοδός μας φαίνεται να αυξάνει την ακρίβεια του αποτελέσματος και να ελαττώνει τα ανεπιθύμητα ψεύδεργα (artifacts), όπως το θόλωμα και την διατάραξη των ακμών.

Διάφορες μέθοδοι μη-γραμμικής διάχυσης μπορούν επίσης να προκύψουν από μεταβολικά προβλήματα, μέσω της ελαχιστοποίησης ενός κατάλληλου συναρτησιακού. Ένα από τα πιο διαδεδομένα συναρτησιακά τέτοιου είδους είναι η *Ολική Μεταβολή* (Total Variation) [106], η οποία δεν επιβάλλει «ποινή» στις ασυνέχειες της εικόνας αλλά μόνο στις ισχυρές ταλαντώσεις και έτσι μπορεί να αφαιρεθεί ο θόρυβος χωρίς να θολώνονται οι ακμές. Όπως θα παρουσιαστεί στο Κεφάλαιο 4, εισάγαμε πρόσφατα ένα θεωρητικό πλαίσιο το οποίο γενικεύει την Ολική Μεταβολή καθώς και αρκετές άλλες μεταβολικές μεθόδους μη-γραμμικής διάχυσης [103]. Το συναρτησιακό που προτείνουμε εφαρμόζεται σε γενικές διανυσματικές εικόνες και βασίζεται στον *δομικό τανυστή*, ο οποίος περιγράφει την γεωμετρία των δομών της εικόνας στην γειτονιά του κάθε σημείου. Όπως αποδείξαμε, η ελαχιστοποίηση του συναρτησιακού αυτού οδηγεί σε μία γενική μη-γραμμική και ανισοτροπική διάχυση η οποία είναι κανονικοποιημένη (regularized), υπό την έννοια ότι ο τανυστής διάχυσης περιλαμβάνει συνελιξείς με έναν πυρήνα. Βασισμένοι στο αποτέλεσμα αυτό, προτείναμε δύο νέες μεθόδους διάχυσης, οι οποίες συνδυάζουν τα πλεονεκτήματα των μεταβολικών μεθόδων και των μεθόδων που βασίζονται σε τανυστές.

Μία άλλη συνεισφορά των ΜΔΕ στην Όραση Υπολογιστών σχετίζεται με τα Μοντέλα *Ενεργών Περιγραμμάτων* (Active Contours) για την κατάτμηση εικόνων [25, 64, 66]. Τα μοντέλα αυτά συνήθως ξεκινούν από μία αρχική κλειστή καμπύλη απλού σχήματος και την παραμορφώνουν σταδιακά έως ότου αυτή περιγράψει το σύνορο του αντικειμένου που θέλουν να κατατμήσουν. Αυτό γίνεται εφικτό υιοθετώντας πάλι μία μεταβολική προσέγγιση, μέσω της ελαχιστοποίησης κάποιου κατάλληλα σχεδιασμένου συναρτησιακού ενέργειας, το οποίο όμως δρα πάνω σε καμπύλες αντί σε εικόνες. Στα πλαίσια της έρευνάς μας, διερευνήσαμε την καταλληλότητα τέτοιων μοντέλων για τις εφαρμογές ανάλυσης εικόνων αρθρωτών λόγου. Γρήγορα όμως έγινε αντιληπτό ότι, λόγω της δυσκολίας των συγκεκριμένων προβλημάτων, ήταν απαραίτητη η ενσωμάτωση πρότερης πληροφορίας σχετικά με

τα σχήματα προς εντοπισμό. Η λεπτομερής μελέτη μοντέλων ενεργών περιγραμμάτων που ενσωματώνουν πρότερη πληροφορία μας οδήγησε στο να σχηματίσουμε την άποψη ότι εμφανίζουν κάποια μειονεκτήματα σημαντικά για τις συγκεκριμένες εφαρμογές: πρώτον, περιέχουν έναν εγγενή πλεονασμό ως προς την μοντελοποίηση του σχήματος και δεύτερον δεν μπορούν να χειριστούν αξιόπιστα τον εντοπισμό *ανοικτών* καμπυλών (για λεπτομέρειες, βλ. Ενότητα 2.1.3.1).

Έτσι, καταλήξαμε ότι είναι πιο αποτελεσματική η υιοθέτηση του πλαισίου των *Στατιστικών Μοντέλων Σχήματος* [35, 37, 83], τα οποία εφαρμόζουν πρότερη γνώση σχετικά με το σχήμα των αντικειμένων καθώς επίσης σχετικά με πρόσθετα χαρακτηριστικά τους, όπως την εμφάνισή τους στις εικόνες.

Βασισμένοι στη μεθοδολογία αυτή, προτείναμε στο [100] μια νέα μέθοδο για την αυτόματη ανίχνευση και παρακολούθηση της καμπύλης της γλώσσας σε ακολουθίες εικόνων υπερήχων (βλ. Κεφάλαιο 5). Η μέθοδος αυτή, η οποία αποτελεί μια παραλλαγή των *Ενεργών Μοντέλων Εμφάνισης* (Active Appearance Models - AAM) [35] ειδικά προσαρμοσμένη για την συγκεκριμένη εφαρμογή, ενσωματώνει πρότερη γνώση για το σχήμα της γλώσσας και μπορεί να εκτιμήσει εύρωστα και με ακρίβεια ολόκληρη την καμπύλη της γλώσσας σε μια ακολουθία εικόνων.

Στην κατηγορία των στατιστικών μοντέλων σχήματος ανήκει επίσης η μοντελοποίηση των χειρομορφών νοηματικής γλώσσας που προτείναμε στο [104, 105] (βλ. Κεφάλαιο 6). Η μοντελοποίηση αυτή είναι αναλλοίωτη σε αφινικούς μετασχηματισμούς και βασίζεται σε μια υβριδική αναπαράσταση σε επίπεδο εικόνας τόσο του σχήματος όσο και της εμφάνισης του χεριού. Με τη διαδικασία του ταιριάσματος του μοντέλου, καταλήγουμε σε χαρακτηριστικά (features) χειρομορφής τα οποία προσφέρουν μία συμπαγή και περιγραφική αναπαράσταση των σχηματισμών του χεριού. Όπως καταδεικνύεται από εκτενή πειράματα, η χρήση των χαρακτηριστικών αυτών στην κατηγοριοποίηση χειρομορφών είναι ιδιαίτερα αποτελεσματική.

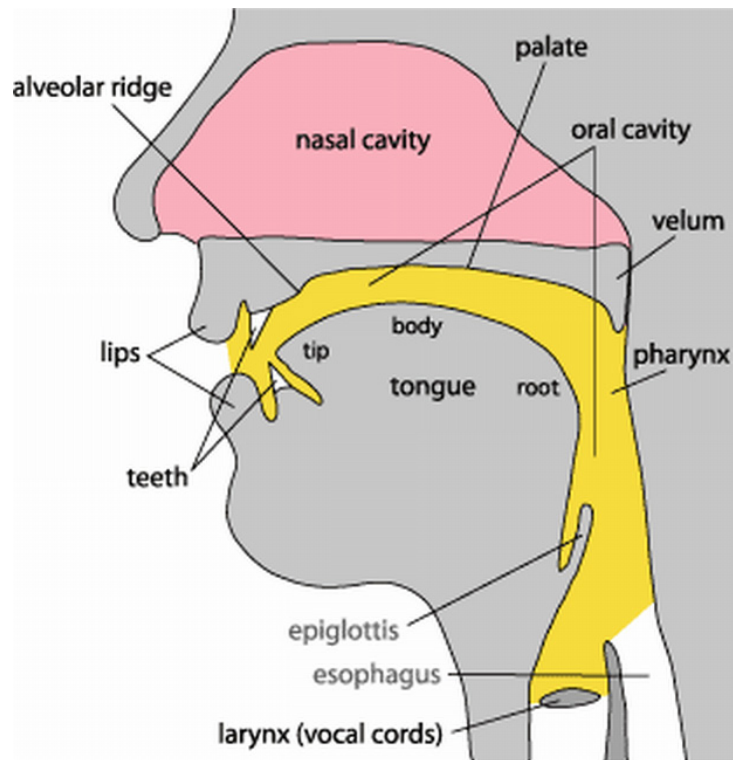
Μία πιο λεπτομερής περιγραφή των ερευνητικών περιοχών με τις οποίες ασχολούμαστε στην διδακτορική μας έρευνα παρέχεται στο Κεφάλαιο 2.

1.2 Κύριες Εφαρμογές

1.2.1 Έρευνα του Ανθρώπινου Συστήματος Παραγωγής Φωνής

Στον φωνητικό λόγο, η πλειοψηφία των αρθρωτών σχηματίζει την *φωνητική οδό*, η οποία ευθύνεται για την τελική μορφή του ήχου της φωνής, βλ. Σχ. 1.1. Η ανάλυση και μοντελοποίηση των αρθρωτών αυτών και γενικότερα του ανθρώπινου συστήματος παραγωγής φωνής είναι μία αναπτυσσόμενη ερευνητική περιοχή με πληθώρα εφαρμογών. Δύο σημαντικά υποπροβλήματα είναι η αρθρωτή σύνθεση φωνής και η οπτικοακουστική αντιστροφή φωνής. Στην ενότητα αυτή, κάνουμε μια σύντομη αναφορά στα σχετικά ζητήματα και για περισσότερες λεπτομέρειες παραπέμπουμε τον ενδιαφερόμενο αναγνώστη στις αναφορές [6, 50, 53, 65, 69, 81, 117].

Η *Αρθρωτή Σύνθεση Φωνής* συνίσταται στην παραγωγή τεχνητής φωνής χρησιμοποιώντας κάποιο μοντέλο που μιμείται το ανθρώπινο σύστημα παραγωγής φωνής, εξαρτάται δηλαδή από παραμέτρους που σχετίζονται με την γεωμετρία και την δυναμική της φωνητικής οδού. Η σύνθεση φωνής αυτού του τύπου, σε αντίθεση με την διαδεδομένη στις μέρες μας



Σχήμα 1.1: Μεσο-οβελιαία όψη της φωνητικής οδού [54].

σύνθεση που βασίζεται στην συρραφή κομματιών φυσικού λόγου από κάποια βάση δεδομένων, προσφέρει μεγάλη ευελιξία και διαισθητικό έλεγχο πολλών όψεων του λόγου, όπως συναισθηματική κατάσταση και χροιά του ομιλητή. Παράλληλα, η μοντελοποίηση αυτή περιέχει εγγενώς την απαραίτητη πληροφορία για την σύνθεση και του οπτικού μέρους της φωνητικής επικοινωνίας, δηλαδή την εικόνα των χειλιών και του υπόλοιπου προσώπου ενός εικονικού ομιλητή. Τα παραπάνω καθιστούν την αρθρωτή σύνθεση φωνής ελπιδοφόρα σε σχέση με την μελλοντική δυνατότητά της να αποδίδει ένα αποτέλεσμα χωρίς την ενοχλητική για τον άνθρωπο αίσθηση της τεχνητής φωνής. Προς το παρόν όμως, η αρθρωτή σύνθεση φωνής βρίσκεται ακόμα σε σχετικά πρώιμο στάδιο, λόγω της μεγάλης πολυπλοκότητας της μοντελοποίησης και της σχετικά περιορισμένης επιστημονικής έρευνας που έχει συντελεστεί παγκοσμίως στο συγκεκριμένο πρόβλημα.

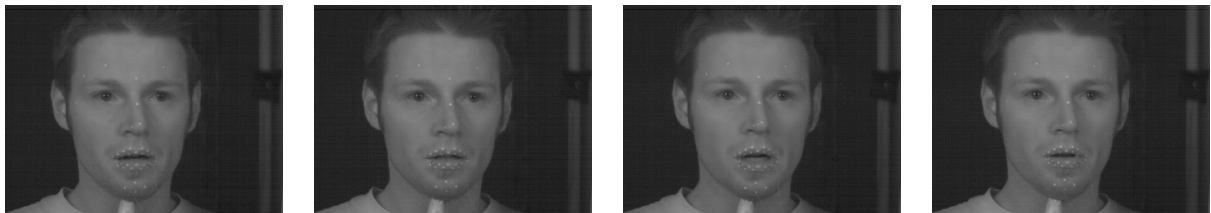
Ένα στενά συνδεδεμένο πρόβλημα είναι το αντίστροφο από το παραπάνω, δηλαδή η *Οπτικοακουστική Αντιστροφή Φωνής*. Συνίσταται στην ανάκτηση του σχήματος και της δυναμικής της φωνητικής οδού λαμβάνοντας σαν είσοδο το ακουστικό σήμα φωνής, συμπληρωμένο από την ακολουθία εικόνων του προσώπου του ομιλητή. Η ολοκληρωμένη λύση αυτού του προβλήματος, το οποίο είναι ακόμα ανοικτό ερευνητικά, θα μπορούσε να αποτελέσει μία τομή στην έρευνα και την τεχνολογία της φωνής, διότι μία αναπαράσταση του σήματος φωνής σε επίπεδο φωνητικής οδού θα ήταν ωφέλιμη τόσο θεωρητικά όσο και πρακτικά σε πολλές εφαρμογές επεξεργασίας φωνής (π. χ. εκμάθηση γλώσσας, λογοθεραπεία, αυτόματη αναγνώριση φωνής, κωδικοποίηση φωνής). Ιδιαίτερο ενδιαφέρον παρουσιάζει η χρήση της μεθοδολογίας της ανάλυσης μέσω σύνθεσης, στην οποία αναζητούνται οι βέλτιστοι παράμετροι της φωνητικής οδού με τις οποίες ένας κατάλληλος αρθρωτός συνθέτης παράγει ένα σήμα φωνής όσο το δυνατόν πιο παρόμοιο με το

πραγματικό.

Ένα κρίσιμο και ιδιαίτερα απαιτητικό κομμάτι της σύγχρονης έρευνας του ανθρώπινου συστήματος παραγωγής φωνής αποτελείται από την καταγραφή και την επεξεργασία οπτικών δεδομένων των αρθρωτών κατά τη διάρκεια της ομιλίας. Η ποιότητα τέτοιων δεδομένων έχει τα τελευταία χρόνια βελτιωθεί, επιτρέποντας ήδη σημαντικές προόδους στην ακουστική και αρθρωτή μοντελοποίηση της ομιλίας. Από την άλλη όμως, απέχει ακόμα αρκετά από το να φτάσει σε ένα αξιόπιστο επίπεδο. Επιπρόσθετα, δεν υπάρχει, μέχρι στιγμής τουλάχιστον, κάποια τεχνική καταγραφής η οποία από μόνη της να προσφέρει *ικανοποιητική* πληροφορία σχετικά με την θέση, το σχήμα και την δυναμική *όλων* των αρθρωτών της φωνητικής οδού. Για αυτούς τους λόγους, συνήθως χρησιμοποιούνται και συνδυάζονται διάφορες απεικονιστικές τεχνικές, στην πλειοψηφία τους *βιοϊατρικής* προέλευσης (π.χ. στην έρευνά μας, χρησιμοποιούμε κυρίως δεδομένα από το πρωτόκολλο καταγραφής του ευρωπαϊκού προγράμματος *ASPI* [7, 69] στο οποίο συμμετείχαμε). Οι κυριότερες από αυτές τις τεχνικές απεικόνισης, κάθε μια από τις οποίες προσφέρει ένα μέρος της συνολικής επιθυμητής πληροφορίας αλλά παρουσιάζει παράλληλα κάποιους περιορισμούς, είναι οι ακόλουθες:

- **Κάμερα ή ζευγάρι στερεοσκοπικών καμερών**, οι οποίες βιντεοσκοπούν το πρόσωπο του ομιλητή, έτσι ώστε να καταγράφονται τα χείλη, το σαγόκι και γενικότερα οι παραμορφώσεις του προσώπου που είναι ορατές εξωτερικά, βλ. Σχ. 1.2. Η απεικόνιση αυτή καταγράφει ένα μικρό μέρος της φωνητικής οδού, αλλά μπορεί να χρησιμοποιηθεί όπως είπαμε σαν συμπληρωματική πληροφορία φωνής στην είσοδο της οπτικοακουστικής αντιστροφής φωνής.
- **Απεικόνιση υπερήχων της φωνητικής οδού**, στην οποία είναι ορατή η κίνηση ενός μέρους της επιφάνειας της γλώσσας, βλ. Σχ. 1.3. Η απεικόνιση αυτή μπορεί να χρησιμοποιηθεί σχετικά εύκολα, δεν εκθέτει το υποκείμενο σε ακτινοβολία και επιτυγχάνει ρυθμούς δειγματοληψίας που μπορούν να καταγράψουν την γρήγορη κίνηση της γλώσσας. Από την άλλη όμως, οι εικόνες υπερήχων περιέχουν υψηλές ποσότητες θορύβου τύπου *speckle*, κάποια μέρη της καμπύλης της γλώσσας δεν είναι καθόλου ορατά ενώ τα υπόλοιπα είναι μόνο ασθενώς ορατά.
- **Ακολουθίες εικόνων ακτίνων X**, στις οποίες απεικονίζεται ολόκληρη η φωνητική οδός μέσω του μηχανισμού «σκίασης» των ακτίνων X: Κάθε εικονοστοιχείο των 2D εικόνων αντιστοιχεί όχι σε ένα σημείο αλλά σε μια ευθεία που τέμνει εγκάρσια το πρόσωπο, βλ. Σχ. 1.4. Παλαιότερα, αυτή ήταν η απεικονιστική τεχνική που κυριαρχούσε στην έρευνα της άρθρωσης φωνής, αλλά λόγω των σαφών πλέον κινδύνων για την υγεία, τις τελευταίες δεκαετίες έχει ελαχιστοποιηθεί η χρήση της, όσον αφορά τόσο τη συχνότητα συλλογής τέτοιων δεδομένων όσο και τον χρόνο και την ένταση της έκθεσης στις ακτίνες. Έτσι, είναι διαθέσιμα πολυάριθμα δεδομένα τέτοιου τύπου από καταγραφές παλαιότερων δεκαετιών, τα οποία όμως συχνά είναι διαταραγμένα από σημαντικό θόρυβο και θόλωμα, λόγω των περιορισμών της τεχνολογίας της αντίστοιχης εποχής.
- **Απεικόνιση MRI**, η οποία παρέχει μία 3D εικόνα ολόκληρης της φωνητικής οδού σε ικανοποιητική ποιότητα, βλ. Σχ. 1.5. Όμως η καταγραφή αυτή δεν περιλαμβάνει καθόλου δυναμική πληροφορία. Επίσης, ο ομιλητής πρέπει να κάθεται ξαπλωμένος

ανάσκελα σε μία στενή σήραγγα, κάτι το οποίο εμποδίζει την φυσικότητα του λόγου, ενώ το μηχάνημα καταγραφής παράγει θόρυβο υψηλής έντασης. Τέλος, ένα άλλο σημαντικό μειονέκτημα είναι ότι η διάρκεια της καταγραφής αυτής είναι σχετικά μεγάλη και το υποκείμενο πρέπει να εκφέρει παρατεταμένα κάποιο φώνημα παραμένοντας τελείως ακίνητο ώστε οι αποκτημένες εικόνες να μην είναι διαταραγμένες. Με στόχο την υπέρβαση των μειονεκτημάτων αυτών, τα τελευταία χρόνια έχει αρχίσει να χρησιμοποιείται μια νέα τεχνική, αυτή των *δυναμικών MRI* (dynamic / real time MRI). Η τεχνική αυτή επιτρέπει την καταγραφή δυναμικής πληροφορίας, με το κόστος όμως της σημαντικά χαμηλότερης ποιότητας (ισχυρότερος θόρυβος και μικρότερη χωρική ευκρίνεια) και της καταγραφής μόνο της μεσο-οβελιαίας τομής του προσώπου, βλ. Σχ. 1.6.

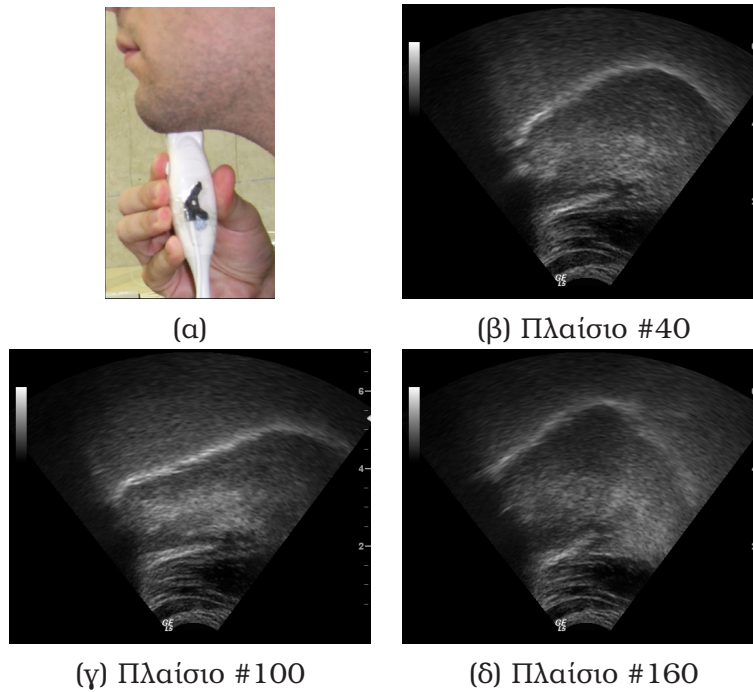


Κάμερα 1



Κάμερα 2

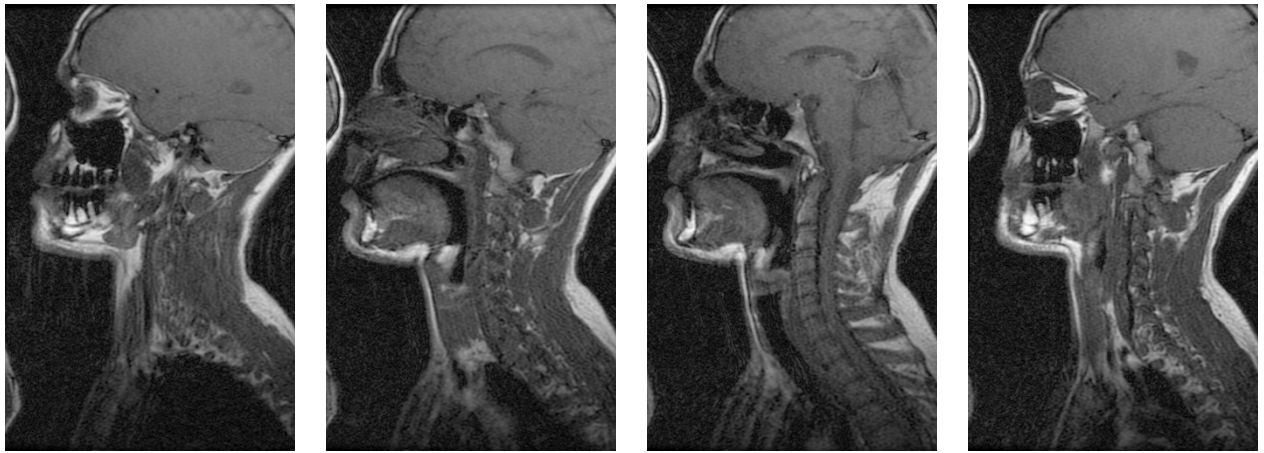
Σχήμα 1.2: Πλαίσια από βιντεοσκόπηση του προσώπου ομιλητή με ζευγάρι στερεοσκοπικών καμερών κατά το πρωτόκολλο του ASPI [7, 69] (κάθε κάμερα καταγράφει 640×480 εικονοστοιχεία/πλαίσιο, 120 Hz).



Σχήμα 1.3: Απεικόνιση υπερήχων της γλώσσας ομιλητή. (α) Τοποθέτηση του probe υπερήχων κάτω από το πηγούνι. (β)-(δ) Πλαίσια από καταγεγραμμένη ακολουθία εικόνων υπερήχων κατά το πρωτόκολλο του ASPI [7, 69] (532×434 εικονοστοιχεία ανά πλαίσιο, περίπου 0.17 mm/εικονοστοιχείο, 66 Hz).



Σχήμα 1.4: Ακολουθίες ακτίνων X για την απεικόνιση της φωνητικής οδού ομιλητή: πλαίσια μίας καταγραφής του 1974 από το *Université Laval* [97] (φιλμ 35 mm, 50 Hz - ψηφιοποιημένο από τους Munhall et al. [89]).



(α) Τομή #4

(β) Τομή #7

(γ) Τομή #10

(δ) Τομή #13

Σχήμα 1.5: Απεικόνιση MRI της φωνητικής οδού: 4 από τις 16 τομές μιας 3Δ στατικής καταγραφής κατά το πρωτόκολλο του ASPI [6, 69] (0.63 mm/εικονοστοιχείο κάθε τομής, πάχος τομών 3 mm, συνολικός χρόνος καταγραφής 65 δευτερόλεπτα).



(α) Πλαίσιο #1

(β) Πλαίσιο #17

(γ) Πλαίσιο #33

Σχήμα 1.6: Απεικόνιση δυναμικών MRI: 3 πλαίσια από μία ακολουθία μεσο-οβελιαίων τομών της φωνητικής οδού ομιλητή καταγεγραμμένη σύμφωνα με το πρωτόκολλο του [49] (2.34 mm/εικονοστοιχείο, 9 Hz).

Όπως διαφαίνεται από την προηγηθείσα απαρίθμηση, η ποιότητα των βιοϊατρικών εικόνων της φωνητικής οδού είναι συχνά μη ικανοποιητική και χρειάζεται βελτίωση, ώστε η περαιτέρω ανάλυση των δεδομένων να είναι αξιόπιστη. Αυτό μπορεί να επιτευχθεί πρώτα από όλα μέσω της βελτίωσης των τεχνικών βιοϊατρικής απεικόνισης, η οποία όμως δεν λύνει το πρόβλημα στα ήδη καταγεγραμμένα δεδομένα, συνεπάγεται αύξηση του κόστους καταγραφής και φτάνει σε έναν κορεσμό λόγω φυσικών περιορισμών της διαδικασίας καταγραφής. Ένας δεύτερος, αρκετά αποτελεσματικός, τρόπος είναι η εκ των υστέρων ψηφιακή επεξεργασία των καταγεγραμμένων εικόνων. Αυτό είναι ένα από τα θέματα που απασχολούν την έρευνά μας και για το οποίο, όπως θα δούμε σε επόμενα κεφάλαια, έχουμε αναπτύξει και εφαρμόσει μοντέρνες τεχνικές της Όρασης Υπολογιστών.

Το επόμενο σημαντικό βήμα επεξεργασίας είναι η καθεαυτή αξιοποίηση των παραπάνω οπτικών δεδομένων μέσω της εξαγωγής της γεωμετρίας και της δυναμικής διαφόρων μερών της φωνητικής οδού. Αυτόματες μέθοδοι για τέτοιου είδους ανακατασκευές αποτελούν επίσης μέρη της διδακτορικής μας έρευνας. Για παράδειγμα, έχουμε προτείνει μια νέα μέθοδο για την αυτόματη ανίχνευση και παρακολούθηση της καμπύλης της γλώσσας σε ακολουθίες εικόνων υπερήχων (Κεφάλαιο 5).

1.2.2 Αυτόματη Αναγνώριση Νοηματικής Γλώσσας

Οι νοηματικές γλώσσες χρησιμεύουν κυρίως σαν ένας εναλλακτικός ή συμπληρωματικός τρόπος επικοινωνίας μεταξύ ανθρώπων. Επίσης αποτελούν τον τρόπο επικοινωνίας που οι περισσότεροι κωφοί προτιμούν, μιας και συχνά παρουσιάζουν περιορισμένη ικανότητα χειρισμού της συμβατικής γλώσσας, ακόμα και για ανάγνωση ή γραφή. Αξίζει να σημειωθεί ότι εκτιμάται ότι οι προ-γλωσσικοί κωφοί αποτελούν περισσότερο από το 1% του συνολικού πληθυσμού της Ευρώπης.

Τα οπτικά πρότυπα των νοηματικών γλωσσών, σε αντίθεση με τα ακουστικά πρότυπα των προφορικών γλωσσών, σχηματίζονται μέσω των χειρομορφών και των κινήσεων των χεριών, καθώς επίσης και μέσω άλλων χαρακτηριστικών, όπως οι εκφράσεις του προσώπου και οι κινήσεις των χειλιών. Η γραμματική της νοηματικής γλώσσας είναι θεμελιακά διαφορετική από αυτήν της προφορικής γλώσσας. Η δομή μιας πρότασης στην προφορική γλώσσα είναι γραμμική, υπό την έννοια ότι η μία λέξη διαδέχεται την άλλη, ενώ στην νοηματική γλώσσα υπάρχει μία ταυτόχρονη δομή με παράλληλη χωρική και χρονική διαμόρφωση, η οποία μεταφέρει πλούσια πληροφορία σχετικά με τον χρόνο, τον τόπο, το πρόσωπο ή το κατηγορημα.

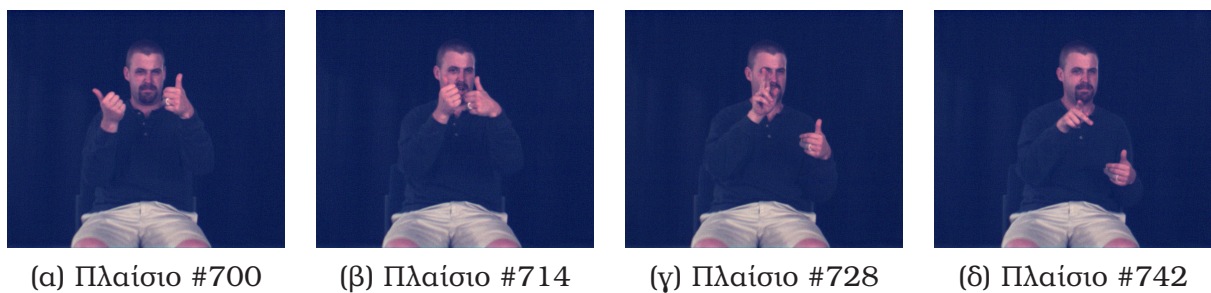
Η εκφραστικότητα των νοηματικών γλωσσών διευκολύνει την αλληλεπίδραση και την ανταλλαγή πληροφορίας όχι μόνο μεταξύ κωφών αλλά επίσης σε καταστάσεις όπου η ομιλία είναι δύσκολη, π.χ. σε θορυβώδεις χώρους εργασίας ή σαν χειρονομίες που συνοδεύουν την ομιλία. Στοιχεία της νοηματικής γλώσσας μπορούν επίσης να χρησιμοποιηθούν στην Επικοινωνία Ανθρώπου - Μηχανής.

Δυστυχώς, πολύ λίγα ακούοντα άτομα γνωρίζουν να επικοινωνούν στην νοηματική γλώσσα. Επίσης η παρουσία μεταφραστών που μεταφράζουν από την νοηματική στην προφορική γλώσσα και αντίστροφα είναι σε πολλές περιπτώσεις αδύνατη ή πολύ δύσκολη. Εξάλλου, τα κωφά άτομα παρουσιάζουν συνήθως μεγάλες δυσκολίες στην ανάγνωση και την γραφή. Ο λόγος για αυτό είναι ότι, ενώ τα ακούοντα άτομα μαθαίνουν και αντιλαμβάνονται την γραπτή γλώσσα σαν μία οπτική αναπαράσταση της προφορικής τους γλώσσας, για τα κωφά άτομα αυτή η αντιστοιχία δεν υπάρχει και τα γράμματα (τα οποία

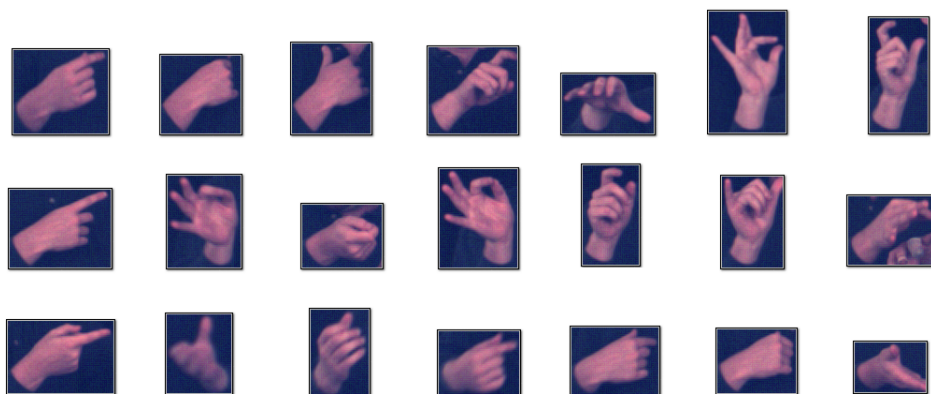
κωδικοποιούν φωνήματα) είναι απλά σύμβολα χωρίς νόημα. Αυτό για παράδειγμα έχει την αρνητική συνέπεια πολλοί κωφοί να μην μπορούν να χρησιμοποιήσουν τον Παγκόσμιο Ιστό ή να επικοινωνήσουν μέσω ηλεκτρονικού ταχυδρομείου με την ίδια ευκολία με τους ακούοντες. Τα παραπάνω οδηγούν σε προβλήματα κοινωνικής ένταξης των κωφών και βάζουν εμπόδια στην προσπάθειά τους για έναν ανεξάρτητο και αυτοσυντηρούμενο τρόπο ζωής.

Για όλους τους προαναφερθέντες λόγους, η αυτόματη αναγνώριση καθώς και η αυτόματη σύνθεση νοηματικής γλώσσας μπορούν να βοηθήσουν σημαντικά κάποιες όψεις της κοινωνικής ζωής των κωφών καθώς και των υπολοίπων ανθρώπων του περιβάλλοντός τους.

Στα πλαίσια της διδακτορικής μας έρευνας, ασχολούμαστε με διάφορα υποπροβλήματα του γενικότερου προβλήματος της Αυτόματης Αναγνώρισης Νοηματικής Γλώσσας (AANΓ), όπως είναι ο εντοπισμός και η παρακολούθηση των χεριών του νοηματιστή καθώς και η εξαγωγή χαρακτηριστικών (features) που να περιγράφουν αξιόπιστα την πόζα και την διαμόρφωσή τους. Η αποτελεσματική αντιμετώπιση των υποπροβλημάτων αυτών είναι ιδιαίτερα κρίσιμη για την τελική επιτυχία ενός συστήματος AANΓ. Πρόκειται όμως για έναν αρκετά δύσκολο στόχο, λόγω κυρίως της γρήγορης κίνησης των χεριών, των αλληλεπικαλύψεων καθώς και της μεγάλης μεταβλητότητας στην πόζα και το σχήμα του χεριού, βλ. Σχ. 1.7, 1.8.



Σχήμα 1.7: Αλληλεπικαλύψεις και αλλαγές στην διαμόρφωση των χεριών κατά τον νοηματισμό: 4 πλαίσια από ένα βίντεο της βάσης συνεχούς Αμερικανικής Νοηματικής Γλώσσας BU400 [45] (648×484 εικονοστοιχεία/πλαίσιο, 60 Hz).



Σχήμα 1.8: Μεταβλητότητα στη διαμόρφωση και την πόζα του κυρίαρχου χεριού: παραδείγματα εξαγωγής των σχετικών λεπτομερειών από βίντεο της βάσης BU400 [45].

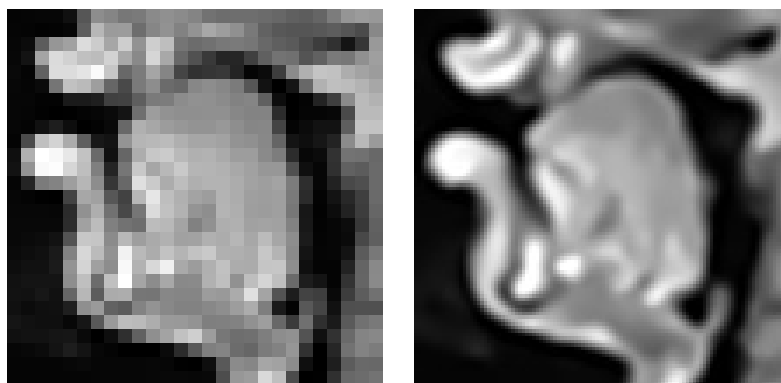
Όπως θα παρουσιαστεί στο Κεφάλαιο 6, έχουμε αναπτύξει μία μοντελοποίηση των χειρομορφών νοηματικής γλώσσας η οποία προσφέρει μία συμπαγή αναπαράσταση, ιδιαίτερα αποτελεσματική για την κατηγοριοποίηση χειρομορφών.

1.3 Ερευνητικές Συνεισφορές

Οι κύριες συνεισφορές της μέχρι τώρα διδακτορικής μας έρευνας μπορούν να συνοψισθούν ως εξής:

- Έχει σχεδιαστεί ένα μοντέλο μη-γραμμικής διάχυσης για την παρεμβολή διανυσματικών εικόνων (βλ. Κεφάλαιο 3 και [101, 102]). Στο μοντέλο αυτό, επιβάλλουμε έναν περιορισμό σχετικά με την αντιστρεψιμότητα της παρεμβολής, ο οποίος εκμεταλλεύεται αποτελεσματικά την διαθέσιμη πληροφορία της εικόνας εισόδου. Η προτεινόμενη διάχυση, η οποία είναι μη-γραμμική αλλά και ανισοτροπική, επιτελεί προσαρμοστική εξομάλυνση ενώ παράλληλα ικανοποιεί τον περιορισμό αντιστρεψιμότητας, χάρη σε έναν κατάλληλο τελεστή προβολής. Η μέθοδος αυτή μπορεί μεταξύ άλλων να συνεισφέρει στην βελτίωση της ευκρίνειας εικόνων της φωνητικής οδού ομιλητή. Μεγεθύνει με ακρίβεια τέτοιου είδους βιοϊατρικές εικόνες, ακόμα και όταν αυτές έχουν ιδιαίτερα χαμηλή ευκρίνεια, περιορίζοντας τα ανεπιθύμητα ψεύδεργα (artifacts) που συνήθως εμφανίζονται στην παρεμβολή εικόνων, βλ. Σχ. 1.9.
- Έχει προταθεί και μελετηθεί ένα συναρτησιακό για διανυσματικές εικόνες που γενικεύει διάφορες υπάρχουσες μεταβολικές μεθόδους ενώ οδηγεί και σε νέες μεθόδους μη-γραμμικής διάχυσης που βασίζονται σε τανυστές (βλ. Κεφάλαιο 4 και [103]). Το συναρτησιακό αυτό βασίζεται στον *δομικό τανυστή*, ο οποίος περιγράφει την γεωμετρία των δομών της εικόνας στην γειτονιά του κάθε σημείου. Όπως αποδείξαμε, η ελαχιστοποίηση του συναρτησιακού οδηγεί σε μία μη-γραμμική και ανισοτροπική διάχυση η οποία είναι κανονικοποιημένη (regularized), υπό την έννοια ότι ο τανυστής διάχυσης περιλαμβάνει συνελίξεις με έναν πυρήνα. Βασισμένοι στο αποτέλεσμα αυτό, προτείναμε δύο νέες μεθόδους διάχυσης, οι οποίες συνδυάζουν τα πλεονεκτήματα των μεταβολικών μεθόδων και των μεθόδων που βασίζονται σε τανυστές. Οι μέθοδοι αυτές μπορούν να εφαρμοστούν με επιτυχία και για την απομάκρυνση του θορύβου από βιοϊατρικές εικόνες της φωνητικής οδού, βλ. Σχ. 1.10.
- Έχει αναπτυχθεί μία τεχνική προσαρμοστικού φιλτραρίσματος ειδικά σχεδιασμένη για εικόνες υπερήχων της γλώσσας ομιλητή (βλ. Ενότητα 5.3 και [7]). Η τεχνική αυτή βασίζεται επίσης στον δομικό τανυστή της εικόνας, αλλά αντί να εφαρμόζει διάχυση, φιλτράρει την εικόνα με έναν χωρικά μεταβαλλόμενο πυρήνα ο οποίος ενισχύει τις 1D φωτεινές ζώνες της εικόνας. Με αυτόν τον τρόπο απομακρύνεται ο έντονος speckle θόρυβος και βελτιώνεται σημαντικά η ορατότητα της γλώσσας, βλ. Σχ. 1.11.
- Έχει σχεδιασθεί ένα στατιστικό μοντέλο σχήματος για τον αυτόματο εντοπισμό και παρακολούθηση της καμπύλης της γλώσσας σε βίντεο υπερήχων (βλ. Κεφάλαιο 5 και [100]). Το μοντέλο αυτό ενσωματώνει πρότερη γνώση για το σχήμα της γλώσσας και μπορεί να εκτιμήσει εύρωστα και με ακρίβεια ολόκληρη την καμπύλη της γλώσσας σε μια ακολουθία εικόνων, βλ. Σχ. 1.12.

- Τέλος, έχει προταθεί μία νέα μοντελοποίηση των χειρομορφών νοηματικής γλώσσας η οποία εμπεριέχει αφινικούς μετασχηματισμούς και βασίζεται σε μια υβριδική αναπαράσταση σε επίπεδο εικόνας τόσο του σχήματος όσο και της εμφάνισης του χεριού, χωρίς να χρειάζεται ενδεικτικά σημεία (βλ. Κεφάλαιο 6 και [104]). Για το εύρωστο ταίριασμα του αντίστοιχου μοντέλου σε βίντεο συνεχούς νοηματισμού, έχει αναπτυχθεί μία διαδικασία βελτιστοποίησης που ενσωματώνει πρότερη πληροφορία σχετικά με το σχήμα και τη δυναμική των χειρομορφών. Με το ταίριασμα του μοντέλου (βλ. Σχ. 1.13), καταλήγουμε σε χαρακτηριστικά χειρομορφής τα οποία προσφέρουν μία συμπαγή και περιγραφική αναπαράσταση των σχηματισμών του χεριού. Όπως καταδεικνύεται από εκτενή πειράματα, η χρήση των χαρακτηριστικών αυτών στην κατηγοριοποίηση χειρομορφών είναι ιδιαίτερα αποτελεσματική, Σχ. 1.14.

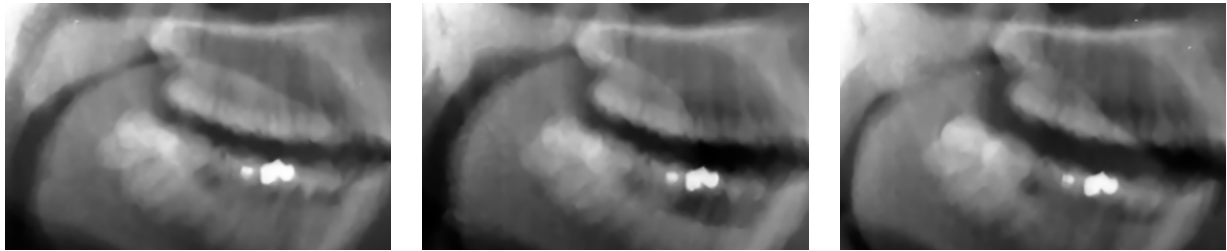


(α) Είσοδος (μεγεθυμένη με επανάληψη των pixels) (β) Προτεινόμενη Μέθοδος

Σχήμα 1.9: Παρεμβολή (4×4) μιας MRI εικόνας της φωνητικής οδού με χαμηλή ευκρίνεια, χρησιμοποιώντας την μέθοδο που εισάγαμε στα [101, 102].

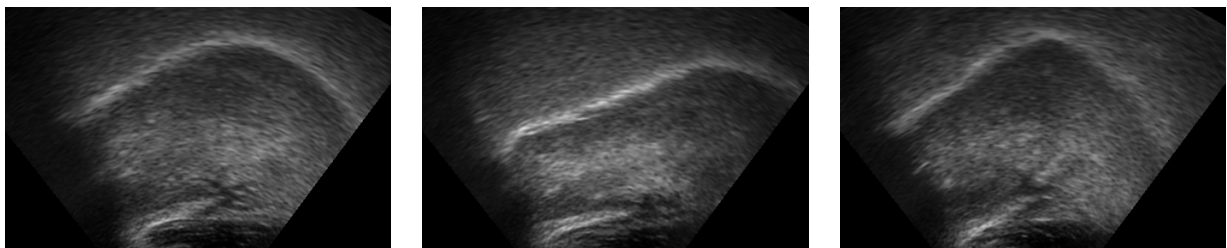


Ακολουθία εισόδου

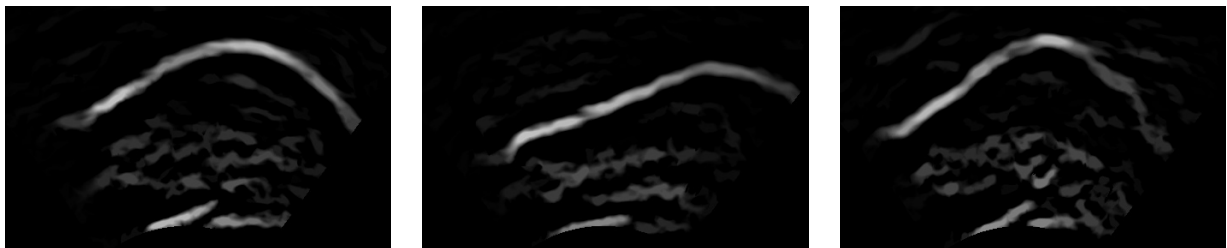


Ακολουθία εξόδου

Σχήμα 1.10: Αποθρομβοποίηση μιας ακολουθίας εικόνων ακτίνων X της φωνητικής οδού ομιλητή χρησιμοποιώντας την Τανυστική ΟΜ που προτείναμε στο [103].

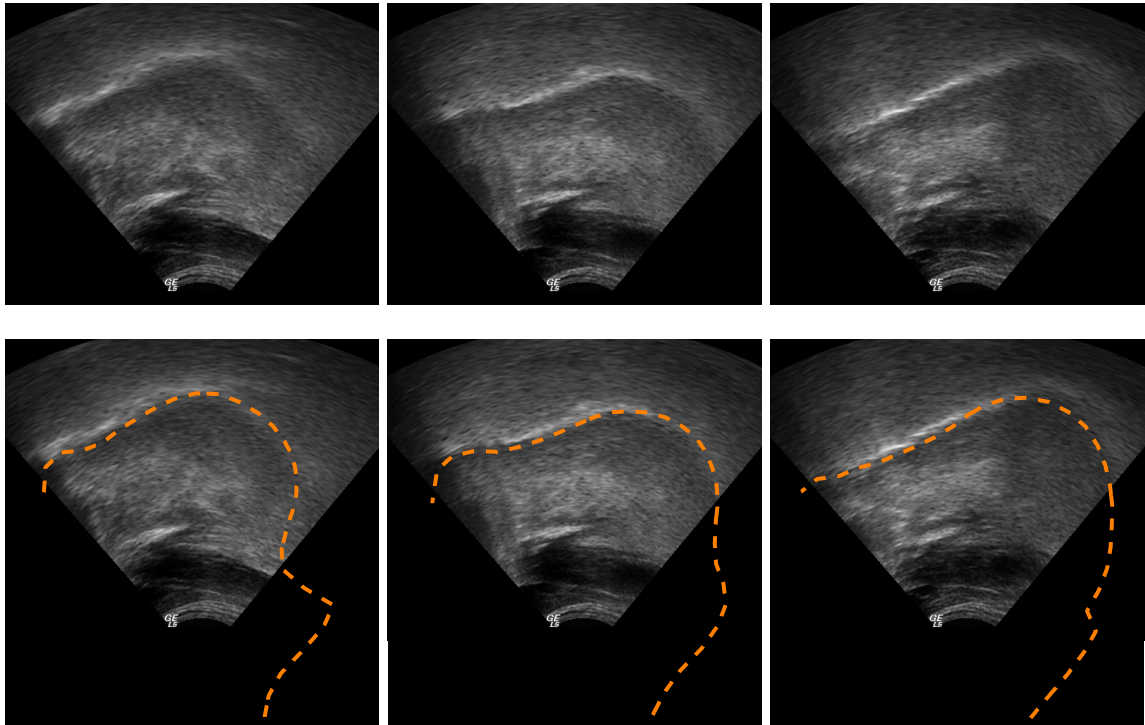


Ακολουθία εισόδου

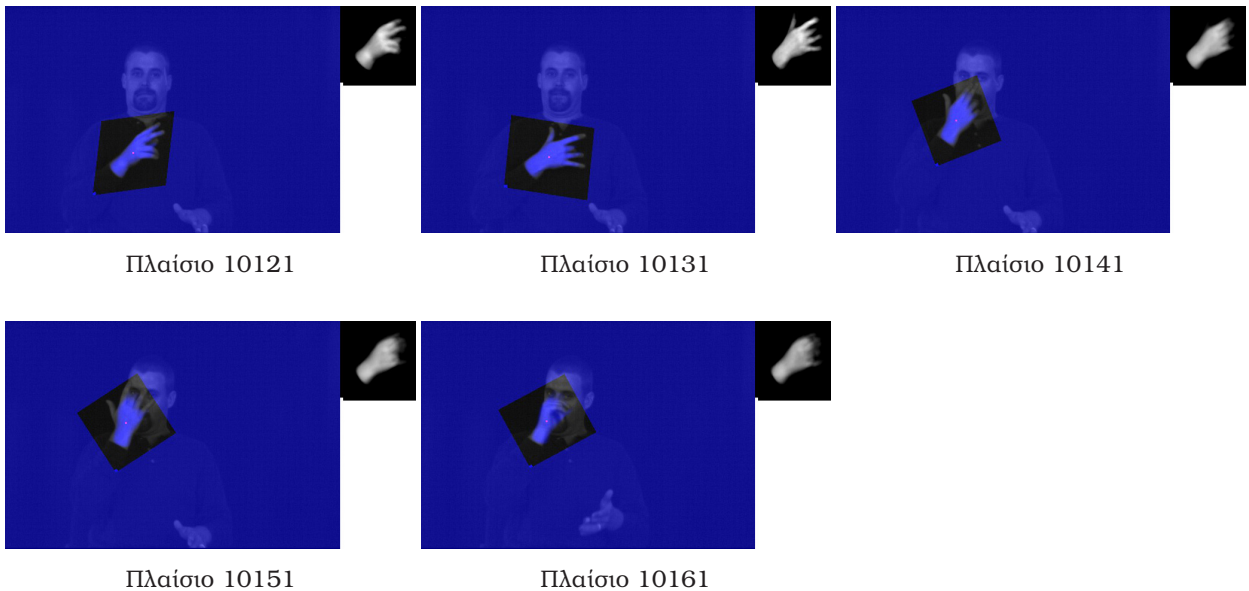


Ακολουθία εξόδου

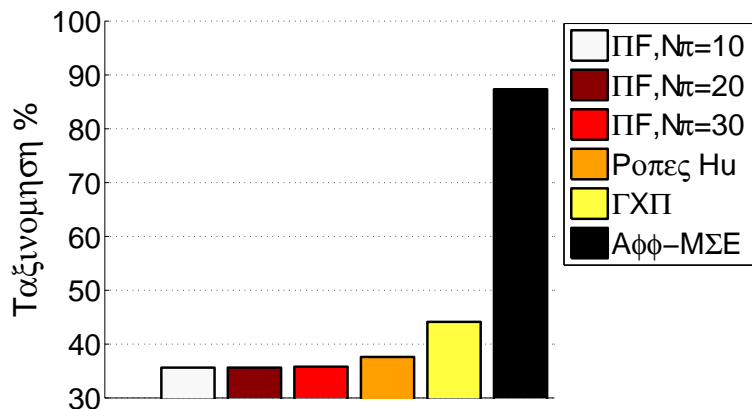
Σχήμα 1.11: Προσαρμοστικό φιλτράρισμα εικόνων υπερήχων της γλώσσας ομιλητή, με την μέθοδο που προτείναμε στο [7].



Σχήμα 1.12: Παρακολούθηση της καμπύλης της γλώσσας σε βίντεο υπερήχων με το μοντέλο που εισάγαμε στο [100]. *Πρώτη σειρά:* Τρία πλαίσια της ακολουθίας εισόδου. *Δεύτερη σειρά:* Εκτίμηση ολόκληρης της καμπύλης της γλώσσας, ενσωματώνοντας πρότερη πληροφορία σχήματος.



Σχήμα 1.13: Ταίριασμα του προτεινόμενου μοντέλου χειρομορφών σε ένα τμήμα βίντεο συνεχούς νοηματισμού. Σε κάθε αρχική εικόνα πλαισίου, υπερθέτουμε την ανακατασκευη βασισμένη στο μοντέλο, ενώ στην πάνω δεξιά γωνία δείχνουμε κάθε φορά την ίδια ανακατασκευή, αλλά στον χώρο του μοντέλου, όπου η ανακατασκευή είναι «κανονικοποιημένη» ως προς τους αφινικούς μετασχηματισμούς.



Σχήμα 1.14: Πειράματα ταξινόμησης χειρομορφών. *Επάνω*: τάξεις χειρομορφών που χρησιμοποιήθηκαν στο πείραμα. *Κάτω*: αποτελέσματα ταξινόμησης χρησιμοποιώντας εξαγωγή χαρακτηριστικών με την προτεινόμενη (Αφφ-ΜΣΕ) [104] καθώς και με άλλες βασικές μεθόδους της βιβλιογραφίας (για λεπτομέρειες βλ. Ενότητα 6.4).

Κεφάλαιο 2

Επισκόπηση Ερευνητικών Περιοχών

Στο κεφάλαιο αυτό, παρουσιάζονται οι κύριες μεθοδολογίες της Όρασης Υπολογιστών που έχουμε υιοθετήσει στην έρευνά μας, δηλαδή οι Μερικές Διαφορικές Εξισώσεις καθώς και τα Στατιστικά Μοντέλα Σχήματος. Επιπρόσθετα, γίνεται μία επισκόπηση μερικών από τις βασικότερες μεθόδους της βιβλιογραφίας που σχετίζονται με τις εφαρμογές που μελετάμε, δηλαδή την αυτόματη ανάλυση βίντεο της γλώσσας ομιλητή καθώς και ανάλυση βίντεο νοηματικής γλώσσας.

2.1 Μερικές Διαφορικές Εξισώσεις

Μία από τις πιο σύγχρονες και ελπιδοφόρες μεθοδολογίες για πολλά προβλήματα της Όρασης Υπολογιστών είναι η χρήση μοντέλων Μερικών Διαφορικών Εξισώσεων (ΜΔΕ). Οι ΜΔΕ άρχισαν να χρησιμοποιούνται συστηματικά στην Όραση Υπολογιστών σχετικά πρόσφατα, προς τα τέλη της δεκαετίας του 1980 και έκτοτε έχουν γνωρίσει μία αρκετά ευρεία ανάπτυξη στον τομέα αυτό. Η μεθοδολογία αυτή έχει γίνει πλέον δημοφιλής στην Επεξεργασία Εικόνας και την Όραση Υπολογιστών μιας και παρουσιάζει σημαντικά πλεονεκτήματα, μερικά εκ των οποίων είναι τα ακόλουθα:

- Επειδή οι ΜΔΕ εμφανίζονται κατεξοχήν στη μαθηματική μοντελοποίηση φυσικών φαινομένων (π.χ. διάχυση θερμότητας, διάδοση κύματος, μεταφορά ύλης, ηλεκτρομαγνητισμός), μπορεί να αποκτηθεί φυσική διαίσθηση για τη συμπεριφορά της λύσης τους και να σχεδιαστούν μέθοδοι επεξεργασίας με επιθυμητές ιδιότητες μέσω της μίμησης και παραλλαγής φυσικών διεργασιών.
- Αρκετές μέθοδοι ΜΔΕ προκύπτουν από μεταβολικά προβλήματα όπου ένα κατάλληλα σχεδιασμένο συναρτησιακό ενέργειας ελαχιστοποιείται, επομένως χρησιμοποιούν έναν εννοιολογικά καθαρό φορμαλισμό στον οποίο οι υποθέσεις του μοντέλου γίνονται μόνο στο συναρτησιακό και η λύση προκύπτει ως βελτιστοποίηση των επιλεγμένων κριτηρίων.
- Χάρη στα γερά θεμέλια των ΜΔΕ από τα Εφαρμοσμένα Μαθηματικά και την Μαθηματική Φυσική, μπορούν να αξιοποιηθούν πληθώρα έτοιμων αποτελεσμάτων σχετικά με την ύπαρξη, μοναδικότητα και συνέχεια των λύσεων καθώς και με τον σχεδιασμό ευσταθών αλγορίθμων για την αριθμητική τους προσέγγιση.

- Η μεθοδολογία των ΜΔΕ προσφέρει νέες ερμηνείες διάφορων κλασικών μεθόδων (π.χ. γκαουσιανή εξομάλυνση, median φιλτράρισμα, dilation και erosion) μέσα σε ενοποιητικά πλαίσια καθώς και γενικεύσεις των μεθόδων αυτών.
- Τα μοντέλα ΜΔΕ είναι εγγενώς συνεχή, σε αντίθεση με αρκετές κλασικές μεθόδους όπου οι εικόνες θεωρούνται εξαρχής *διακριτές*. Έτσι, ο σχεδιασμός και η θεωρητική αξιολόγηση των διαφόρων μεθόδων επεξεργασίας γίνονται στον συνεχή χώρο, χωρίς να εξαρτώνται από κάποιο πλέγμα ορισμού της διακριτής εικόνας. Το γεγονός αυτό προσφέρει περισσότερη διαίσθηση καθώς επίσης και ευκολία στην σχεδίαση και υλοποίηση μεθόδων που είναι αναλλοίωτες ως προς περιστροφές ή άλλους γεωμετρικούς μετασχηματισμούς της εικόνας ¹.

2.1.1 Χώροι Κλίμακας

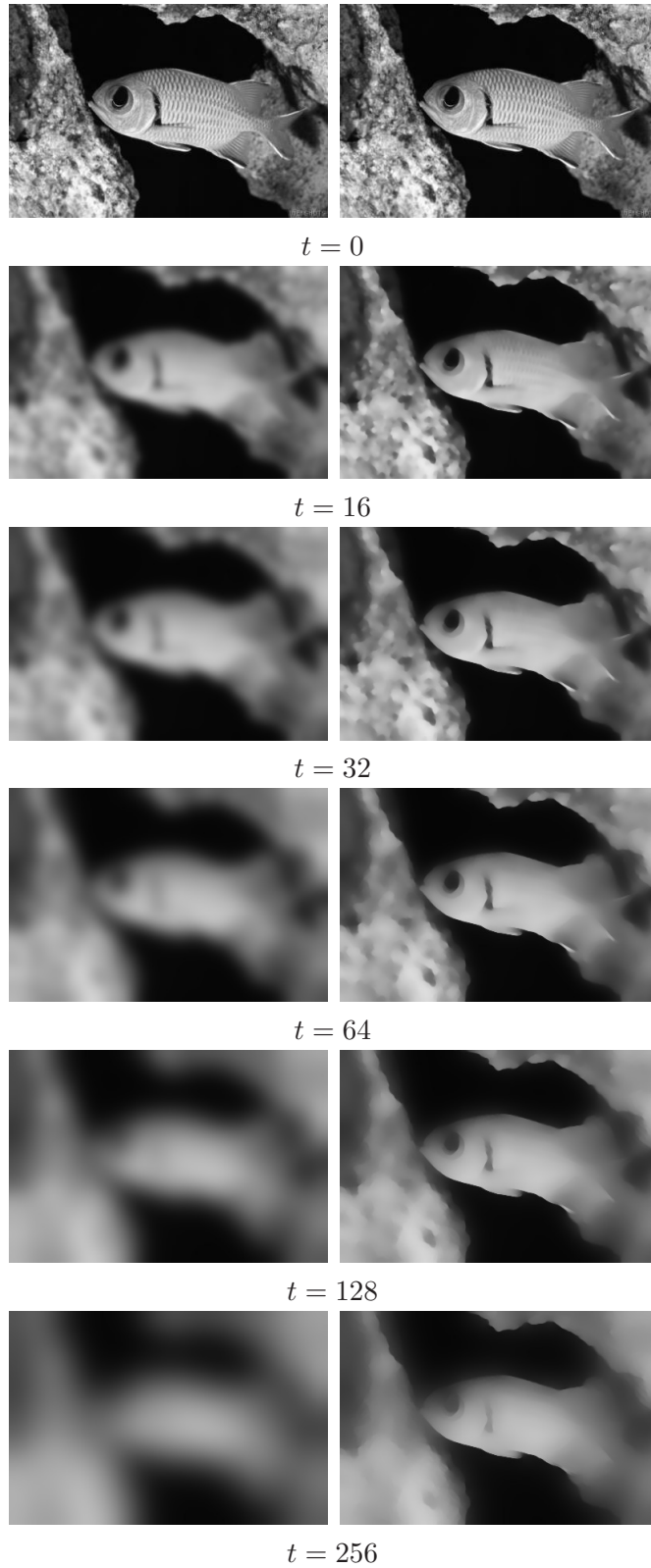
Η εφαρμογή των ΜΔΕ στην Όραση Υπολογιστών έχει μεταξύ άλλων συνεισφέρει σημαντικά στην ανάπτυξη της θεωρίας των *χώρων κλίμακας* (scale spaces) [139]. Ένας χώρος κλίμακας μιας εικόνας είναι μία μονοπαραμετρική οικογένεια εικόνων που περιλαμβάνει ολοένα και περισσότερο απλοποιημένες και αφαιρετικές αναπαραστάσεις της αρχικής εικόνας, βλ. Σχ. 2.1. Με τον τρόπο αυτό, εισάγεται μία *ιεραρχία* στις δομές της, αφού για μεγάλες τιμές της παραμέτρου κλίμακας, μόνο τα κυρίαρχα αντικείμενα της εικόνας παραμένουν εμφανή. Τέτοιου είδους ιεραρχίες βοηθούν στην μετάβαση από την απλή αναπαράσταση της εικόνας ως διδιάστατο σήμα φωτεινοτήτων σε μία υψηλού επιπέδου περιγραφή των απεικονιζόμενων αντικειμένων.

Αξίζει να σημειωθεί ότι οι ΜΔΕ αποτελούν ουσιαστικά την «φυσική γλώσσα» περιγραφής των χώρων κλίμακας: Όπως έχει αποδειχθεί, όλοι οι συνεχείς χώροι κλίμακας που ικανοποιούν έναν μικρό αριθμό αξιωμάτων, τα οποία σχετίζονται με θεμελιώδεις επιθυμητές ιδιότητες, μπορούν να μοντελοποιηθούν από (γενικά μη-γραμμικές) ΜΔΕ δεύτερης τάξης [3].

Οι χώροι κλίμακας που προτάθηκαν αρχικά δημιουργούνται μέσω του βαθυπερατού φιλτραρίσματος της εικόνας με ισοτροπικούς γκαουσιανούς πυρήνες αυξανόμενου χωρικού εύρους [136, 139], το οποίο μπορεί να περιγραφεί από την ΜΔΕ γραμμικής διάχυσης θερμότητας [67]. Αυτή η απλή διαδικασία όμως, δεν απομακρύνει μόνο τα χαρακτηριστικά μικρής κλίμακας της εικόνας, αλλά επίσης θολώνει και εξαπλώνει χωρικά σημαντικές δομές της τρέχουσας κλίμακας (βλ. Σχ. 2.1, αριστερά). Με στόχο την αντιμετώπιση τέτοιου είδους περιορισμών, έχουν προταθεί διάφοροι τύποι μη-γραμμικών χώρων κλίμακας, εκ των οποίων δύο σημαντικές κατηγορίες είναι οι **μορφολογικοί** χώροι κλίμακας [3, 77, 80, 82, 86, 87, 109, 127] και οι χώροι κλίμακας που δημιουργούνται από τη **μη-γραμμική διάχυση**, η οποία αποτελεί το θέμα της ακόλουθης Ενότητας.

Πριν προχωρήσουμε όμως, αξίζει να σημειώσουμε ότι, εκτός από την δημιουργία χώρων κλίμακας, τα μοντέλα μη-γραμμικής διάχυσης χρησιμοποιούνται και σε διάφορες άλλες εφαρμογές της Όρασης Υπολογιστών, με κυριότερη την αποθορυβοποίηση (denoising) εικόνων. Στην περίπτωση της αποθορυβοποίησης, η εικόνα μπορεί να εξελιχθεί σύμφωνα με τις ίδιες ΜΔΕ μη-γραμμικής διάχυσης, αλλά το ζητούμενο εδώ είναι η επιλογή μόνο μιας συγκεκριμένης κλίμακας έτσι ώστε η εικόνα να έχει όσο το δυνατόν καλύτερη

¹Σε αυτές τις περιπτώσεις, το «αναλλοίωτο» νοείται υπό την έννοια ότι η δράση του μετασχηματισμού και της μεθόδου πάνω στην εικόνα μπορούν να αντιστραφούν, οδηγώντας στο ίδιο αποτέλεσμα.



Σχήμα 2.1: Χώροι κλίμακας μιας εικόνας ψαριού. *Αριστερά:* γραμμική διάχυση θερμότητας (συνελίξεις με γκαουσιανούς πυρήνες αυξανόμενου εύρους). *Δεξιά:* μη-γραμμική διάχυση των Catté et al. [27]. Το σχήμα αναπαράγεται από το [78].

ποιότητα.

2.1.2 Μη-γραμμική Διάχυση

Όπως έχει προαναφερθεί, η μη-γραμμική διάχυση εισήχθη στην Όραση Υπολογιστών με στόχο να ξεπεράσει τις αδυναμίες των γραμμικών χώρων κλίμακας. Ένα από τα πρώτα μοντέλα τέτοιου είδους προτάθηκε από τους Perona και Malik [95] και περιγράφεται από την ακόλουθη ΜΔΕ :

$$\begin{aligned} \frac{\partial u(x, y, t)}{\partial t} &= \operatorname{div} (g(\|\nabla u\|^2) \nabla u) , \\ u(x, y, 0) &= u_0(x, y) \end{aligned} \quad (2.1)$$

όπου t είναι μία παράμετρος εικονικού χρόνου που μοντελοποιεί την αύξηση της κλίμακας, $u_0(x, y)$ είναι η εικόνα εισόδου και $u(x, y, t)$ είναι η εξελισσόμενη εικόνα ή αλλιώς η οικογένεια εικόνων του χώρου κλίμακας. Επίσης, η συνάρτηση g του συντελεστή διάχυσης είναι φθίνουσα, έτσι ώστε να ευνοείται η εξομάλυνση της εικόνας u εντός ομοιογενών περιοχών και να αποτρέπεται η ανάμειξη μεταξύ γειτονικών ετερογενών περιοχών και το θόλωμα των ακμών (στις οποίες το $\|\nabla u\|^2$ είναι μεγάλο).

Έκτοτε λοιπόν, έχουν αναπτυχθεί διάφορες τροποποιήσεις και βελτιώσεις της παραπάνω μη-γραμμικής διάχυσης (για κάποιες πιο αναλυτικές επισκοπήσεις της σχετικής βιβλιογραφίας βλ. [10, 99, 122, 135]). Στην κατηγορία αυτή ανήκει η σημαντική συνεισφορά από τους Catté et al. [27], οι οποίοι πρότειναν μία κανονικοποίηση της (2.1) χρησιμοποιώντας μία εξομαλυμένη εκδοχή της εικόνας στον υπολογισμό του συντελεστή διάχυσης, βλ. Σχ. 2.1, δεξιά. Με τον τρόπο αυτό, ο συντελεστής διάχυσης επηρεαζόταν λιγότερο από τον θόρυβο καθώς επίσης έγινε εφικτό να αποδειχθεί η ύπαρξη, μοναδικότητα και ομαλότητα της λύσης της αντίστοιχης ΜΔΕ.

Λίγο αργότερα, ο Weickert [132, 135] πρότεινε την εφαρμογή στις εικόνες μιας γενικευμένου τύπου διάχυσης, η οποία δεν είναι μόνο μη-γραμμική αλλά επίσης *ανισοτροπική*, ελέγχεται δηλαδή από έναν ανισοτροπικό *τανυστή* που προσαρμόζεται πιο ευέλικτα στις τοπικές δομές της εικόνας. Ο ίδιος συγγραφέας [133] επέκτεινε την μέθοδο αυτή στην περίπτωση έγχρωμων ή γενικών διανυσματικών εικόνων: η διάχυση εξελίσσεται σε κάθε κανάλι της εικόνας χωριστά, αλλά η σύζευξη των καναλιών επιτυγχάνεται αποτελεσματικά, μέσω της χρήσης κοινού τανυστή διάχυσης σε όλα τα κανάλια, για κάθε σημείο και κάθε χρονική στιγμή.

Μία άλλη μέθοδος μη-γραμμικής διάχυσης για διανυσματικές εικόνες προτάθηκε από τους Sapiro και Ringach [108]. Στην μέθοδο αυτή, η διάχυση είναι πλήρως ανισοτροπική, μιας και συντελείται μόνο στην κατεύθυνση ελάχιστης μεταβολής της εικόνας σε κάθε σημείο. Λόγω όμως του γεγονότος ότι η διάχυση συντελείται μόνο σε μία κατεύθυνση, η προσαρμογή στις τοπικές δομές της εικόνας δεν είναι πάντα αποτελεσματική.

Οι Tschumperlé και Deriche [122, 124] πρότειναν μία από τις πιο αποτελεσματικές μεθόδους ΜΔΕ για την βελτίωση της ποιότητας διανυσματικών εικόνων. Και στην μέθοδο αυτή, εφαρμόζεται σε κάθε εικόνα εισόδου ανισοτροπική διάχυση που χρησιμοποιεί τανυστές για να προσαρμόζεται στις δομές της (βλ. Ενότητα 3.2.6 για λεπτομέρειες). Χάρη στον προσεκτικό σχεδιασμό των εξισώσεων που προσαρμόζουν τον ανισοτροπικό τανυστή διάχυσης στις τοπικές δομές της εικόνας, η μέθοδος αυτή προσφέρει έναν ευέλικτο και αποτελεσματικό έλεγχο στην διαδικασία της διάχυσης.

Οι κυριότερες εφαρμογές της Όρασης Υπολογιστών στις οποίες χρησιμοποιούνται τα μοντέλα μη-γραμμικής διάχυσης είναι η δημιουργία χώρων κλίμακας, η αποθορυβοποίηση (denoising) [10, 135], η παρεμβολή εικόνων (interpolation) [14, 56, 76, 124], η ανακατασκευή χαμένων κομματιών της εικόνας (inpainting) [16, 26, 30], η ταυτόχρονη απομάκρυνση θορύβου και θολώματος από εικόνες [28, 31, 68]. Στην κατηγορία των μοντέλων μη-γραμμικής διάχυσης για την παρεμβολή εικόνων ανήκει και η μέθοδος που εισάγαμε στα [101, 102] και παρουσιάζουμε στο Κεφάλαιο 3.

2.1.2.1 Μεταβολικές Μέθοδοι

Αρκετές μέθοδοι μη-γραμμικής διάχυσης μπορούν επίσης να προκύψουν από μεταβολικά προβλήματα, μέσω της ελαχιστοποίησης ενός κατάλληλου συναρτησιακού. Ένα ιδιαίτερα δημοφιλές συναρτησιακό τέτοιου είδους είναι η *Ολική Μεταβολή* (Total Variation) [106]:

$$E[u] = \int_{\Omega} \|\nabla u\| dx \quad (2.2)$$

Η ελαχιστοποίηση αυτού του συναρτησιακού χρησιμοποιώντας τις Euler-Lagrange εξισώσεις και την μέθοδο ταχύτερης καθόδου οδηγεί σε μία ΜΔΕ που είναι ειδική περίπτωση της (2.1) με $g(s^2) = 1/s$. Η Ολική Μεταβολή (ΟΜ) δεν επιβάλλει «ποινή» στις ασυνέχειες της εικόνας αλλά μόνο στις ισχυρές ταλαντώσεις, επομένως μπορεί να αφαιρεθεί ο θόρυβος χωρίς να θολώνονται οι ακμές. Στις εργασίες [19, 29, 107] προτείνονται γενικεύσεις της Ολικής Μεταβολής για την περίπτωση διανυσματικών εικόνων.

Μια ενδιαφέρουσα μεταβολική ερμηνεία της γενικής εξίσωσης μη-γραμμικής (ισοτροπικής) διάχυσης (2.1) προκύπτει χρησιμοποιώντας το ακόλουθο συναρτησιακό που γενικεύει την Ολική Μεταβολή [122, 134]:

$$E[u] = \int_{\Omega} \varphi(\|\nabla u\|^2) dx \quad (2.3)$$

όπου φ είναι μια αύξουσα συνάρτηση. Χρησιμοποιώντας ξανά την μέθοδο ταχύτερης καθόδου για την ελαχιστοποίηση του συναρτησιακού, καταλήγουμε σε μία ΜΔΕ της μορφής (2.1) με $g(s^2) = 2\varphi'(s^2)$.

Αν τώρα στο παραπάνω συναρτησιακό (2.3) διαλέξουμε $\phi(s^2) = \sqrt{1 + s^2}$, τότε μπορεί να δειχθεί ότι αντιστοιχεί στο εμβαδόν της 2Δ επιφάνειας που αναπαριστά την εικόνα $u(x, y)$ σαν γραφική παράσταση στον 3Δ χώρο (βλ. π.χ. [32]). Επομένως, η ροή προς την ελαχιστοποίηση του συναρτησιακού αυτού οδηγεί στην μέθοδο των *Ελάχιστων Επιφανειών* (Minimal Surfaces). Οι Sochen et al. [112] γενικεύουν την προσέγγιση αυτή στην περίπτωση διανυσματικών εικόνων, ερμηνεύοντας μία διανυσματική εικόνα u με N κανάλια σαν μια 2Δ επιφάνεια $(x, u(x))$ εμβυθισμένη (embedded) σε έναν χώρο $N+2$ διαστάσεων. Προτείνουν λοιπόν την εξέλιξη της εικόνας σύμφωνα με την *Ροή Beltrami* (Beltrami flow), η οποία είναι μία ροή ανισοτροπικής διάχυσης προς την ελαχιστοποίηση του εμβαδού της επιφάνειας αυτής.

Όπως θα παρουσιαστεί στο Κεφάλαιο 4, εισάγαμε πρόσφατα ένα θεωρητικό πλαίσιο το οποίο γενικεύει περαιτέρω διάφορες υπάρχουσες μεθόδους μη-γραμμικής διάχυσης, όπως είναι η Ολική Μεταβολή και το Συναρτησιακό Beltrami [103].

2.1.3 Κατάτμηση και Ενεργά Περιγράμματα

Μία άλλη σημαντική συνεισφορά των ΜΔΕ στην Όραση Υπολογιστών σχετίζεται με τα Μοντέλα *Ενεργών Περιγραμμάτων* (Active Contours) ή αλλιώς *Μοντέλα Φιδιού* (Snake Models) για την κατάτμηση εικόνων. Τα μοντέλα αυτά ξεκινούν από μία αρχική κλειστή καμπύλη συνήθως απλού σχήματος και την παραμορφώνουν σταδιακά έως ότου αυτή περιγράψει το σύνορο του αντικειμένου που θέλουν να κατατμήσουν. Σε πολλές περιπτώσεις επίσης, μπορούν παράλληλα να εξελίσσονται πολλαπλές καμπύλες οι οποίες να εντοπίζουν περισσότερα του ενός αντικείμενα. Τα παραπάνω επιτυγχάνονται, όπως και σε άλλα μοντέλα ΜΔΕ που συναντήσαμε στις προηγούμενες ενότητες, μέσω της ελαχιστοποίησης κάποιου κατάλληλα σχεδιασμένου συναρτησιακού ενέργειας, αλλά σε αυτή την περίπτωση πρόκειται για συναρτησιακό μιας καμπύλης αντί μιας εικόνας. Για την ελαχιστοποίηση αυτή χρησιμοποιείται η μέθοδος της ταχύτερης καθόδου και η καμπύλη εξελίσσεται σύμφωνα με μία ΜΔΕ, η οποία, στις περιπτώσεις αποτελεσματικής λειτουργίας, ισορροπεί όταν η καμπύλη περιγράψει το αντικείμενο.

Τα ενεργά περιγράμματα προτάθηκαν για πρώτη φορά από τους Kass et al. [64], αλλά η σχετική μέθοδος παρουσιάζει τα μειονεκτήματα ότι **1)** ακόμα και όταν χρειάζεται, το μοντέλο δεν μπορεί απευθείας να διαιρέσει ή να συγχωνεύσει τις εξελισσόμενες καμπύλες και ότι **2)** η τιμή του συναρτησιακού της καμπύλης δεν εξαρτάται μόνο από την γεωμετρία της αλλά και από την παραμετροποίησή της. Μία καθοριστική συνεισφορά που υπερβαίνει τα προαναφερθέντα μειονεκτήματα είναι το μοντέλο των *Γεωδαιτικών Ενεργών Περιγραμμάτων* (Geodesic Active Contours) [25, 66]. Στο μοντέλο αυτό, ελαχιστοποιείται το *γεωδαιτικό* μήκος της καμπύλης, το οποίο υπολογίζεται σε μία πολλαπλότητα (manifold) Riemann εφοδιασμένη με μία μετρική εξαρτώμενη από χαρακτηριστικά της εικόνας. Τούτη η ελαχιστοποίηση οδηγεί σε μία ΜΔΕ για την εξέλιξη της καμπύλης η οποία, όπως και οι περισσότερες ΜΔΕ εξέλιξης καμπυλών, υλοποιείται αξιόπιστα χρησιμοποιώντας τη μέθοδο των *Επιπεδοσυνόλων* (Level sets) [110]. Στην υλοποίηση των επιπεδοσυνόλων, το ενεργό περίγραμμα καθορίζεται έμμεσα ως το μηδενικό επιπεδοσύνολο μιας βαθμωτής συνάρτησης εμβύθισης (embedding) με πεδίο ορισμού το επίπεδο της εικόνας, πράγμα το οποίο προσφέρει εκτός των άλλων το πλεονέκτημα ότι οι αλλαγές στην τοπολογία της καμπύλης γίνονται με φυσικό τρόπο. Έτσι, η εξέλιξη της καμπύλης υλοποιείται μέσω της εξέλιξης της συνάρτησης εμβύθισης, σύμφωνα με μία ΜΔΕ που προκύπτει από την ΜΔΕ εξέλιξης της καμπύλης.

2.1.3.1 Ενσωμάτωση Πρότερης Πληροφορίας Σχήματος

Στα πλαίσια της διδακτορικής μας έρευνας, διερευνήσαμε την καταλληλότητα εφαρμογής μοντέλων ενεργών περιγραμμάτων στα προβλήματα του εντοπισμού και παρακολούθησης της γλώσσας ομιλητή απεικονιζόμενη με υπερήχους αλλά και των χεριών από βίντεο νοηματικής γλώσσας. Γρήγορα έγινε αντιληπτό ότι, λόγω της δυσκολίας των συγκεκριμένων προβλημάτων, ήταν απαραίτητη η ενσωμάτωση στη μεθοδολογία των ενεργών περιγραμμάτων πρότερης πληροφορίας σχετικά με τα σχήματα προς εντοπισμό. Στις μεθόδους που πραγματοποιούν έναν τέτοιο συνδυασμό (βλ. π.χ. [15, 38, 70, 98]), προστίθενται κατάλληλοι *όροι πρότερου σχήματος* είτε κατευθείαν στην ΜΔΕ της εξέλιξης της καμπύλης είτε στο ενεργειακό συναρτησιακό της καμπύλης, υιοθετώντας έναν πιθανοτικό φορμαλισμό. Με τον τρόπο αυτό, η κατάτμηση γίνεται πιο σθεναρή απέναντι στον θόρυβο ή σε επικαλύψεις μέρους των αντικειμένων.

Όμως, κατά την άποψή μας, οι μέθοδοι αυτές περιέχουν έναν εγγενή πλεονασμό ως προς την μοντελοποίηση του σχήματος. Πιο συγκεκριμένα, περιορισμός σχετικά με το σχήμα προς εντοπισμό δεν υπεισέρχεται μόνο μέσω των όρων πρότερου σχήματος αλλά επίσης έμμεσα, μέσω των κλασικών όρων των ενεργών περιγραμμάτων: Βασικό συστατικό των ενεργών περιγραμμάτων είναι η κανονικοποίηση (regularization) της καμπύλης ώστε να είναι ομαλή, πράγμα το οποίο ουσιαστικά αντιστοιχεί και αυτό σε μία συνθήκη για το σχήμα, συνθήκη μάλιστα που είναι γενική και δεν αντικατοπτρίζει το συγκεκριμένο σχήμα προς μοντελοποίηση. Επιπρόσθετα, οι μέθοδοι αυτές, λόγω του ότι βασίζονται στην θεωρία των επιπεδοσυνόλων, δεν μπορούν να χειριστούν αξιόπιστα τον εντοπισμό ανοικτών καμπυλών. Το γεγονός αυτό αποτελεί σημαντικό μειονέκτημα για τις εφαρμογές που μελετάμε, ιδιαίτερα για τον εντοπισμό της γλώσσας στις εικόνες υπερήχων, μιας και το κομμάτι του περιγράμματος της γλώσσας που μας ενδιαφέρει είναι μία ανοικτή καμπύλη. Αντίθετα, πιο κατάλληλα για τις εφαρμογές που μελετάμε φαίνονται να είναι τα Στατιστικά Μοντέλα Σχήματος, τα οποία παρουσιάζονται στην επόμενη ενότητα. Στα μοντέλα αυτά, μπορούν να μοντελοποιηθούν ευρύτερες κατηγορίες σχημάτων και οι περιορισμοί σχήματος εισέρχονται με πιο σαφή τρόπο.

2.2 Στατιστικά Μοντέλα Σχήματος

Τα Στατιστικά Μοντέλα Σχήματος [35,37,83] εφαρμόζουν πρότερη γνώση σχετικά με το σχήμα των αντικειμένων, καθώς επίσης σχετικά με πρόσθετα χαρακτηριστικά τους, όπως την εμφάνισή τους (διακυμάνσεις φωτεινότητας στις εικόνες) ή τις χωρικές τους ιδιότητες. Τα μοντέλα αυτά είναι *αναγεννητικά* (generative), που σημαίνει ότι είναι τόσο πλήρη ώστε να έχουν την δυνατότητα να αναγεννήσουν σχετικά ρεαλιστικές εικόνες των αντικειμένων που μοντελοποιούν. Με αυτόν τον τρόπο, ο εντοπισμός αλλά και η κατηγοριοποίηση ενός αντικειμένου σε μια εικόνα μπορεί να διατυπωθεί σαν ένα πρόβλημα ταιριάσματος: Δεδομένης μιας εικόνας που περιέχει ένα τέτοιο αντικείμενο, αυτό μπορεί να εντοπισθεί και να κατηγοριοποιηθεί προσαρμόζοντας τις παραμέτρους του μοντέλου έτσι ώστε το μοντέλο να αναγεννήσει μία συνθετική εικόνα που να είναι όσο το δυνατόν πιο παρόμοια με την πραγματική.

Ένα άλλο βασικό χαρακτηριστικό των μοντέλων αυτών έγκειται στο ότι είναι *παραμορφώσιμα*, δηλαδή μπορούν να παραμορφωθούν ώστε να λάβουν υπόψη την μεταβλητότητα της κατηγορίας αντικειμένων που μοντελοποιούν. Τα πρότυπα της μεταβλητότητας αυτής μαθαίνονται στατιστικά, μέσω της εκτίμησης των απαραίτητων στατιστικών ποσοτήτων από ένα σύνολο παραδειγμάτων τέτοιων αντικειμένων. Έτσι, μπορεί να επιτευχθεί μία κατάλληλη παραμετροποίηση αυτής της μεταβλητότητας, η οποία να προσφέρει μία συμπαγή και αξιόπιστη αναπαράσταση των αντικειμένων. Κάποιες από τις πιο βασικές δουλειές της κατηγορίας αυτής είναι τα *Ενεργά Μοντέλα Σχήματος* (Active Shape Models - ASM) [37] και τα *Ενεργά Μοντέλα Εμφάνισης* (Active Appearance Models - AAM) [35]. Στην κατηγορία των στατιστικών μοντέλων σχήματος ανήκουν επίσης η μέθοδος εντοπισμού της καμπύλης της γλώσσας σε βίντεο υπερήχων που προτείναμε στο [100] (βλ. Κεφάλαιο 5) καθώς και η μοντελοποίηση των χειρομορφών νοηματικής γλώσσας που εισάγαμε στα [104, 105] (βλ. Κεφάλαιο 6).

2.3 Παρακολούθηση της Γλώσσας Ομιλητή σε Βίντεο Υπερήχων

Ο αυτόματος εντοπισμός και παρακολούθηση της γλώσσας ομιλητή σε ακολουθίες εικόνων υπερήχων είναι ένα αρκετά δύσκολο πρόβλημα, λόγω του ότι οι εικόνες υπερήχων περιέχουν υψηλές ποσότητες θορύβου τύπου speckle, κάποια μέρη της καμπύλης της γλώσσας δεν είναι ορατά ενώ τα υπόλοιπα είναι μόνο ασθενώς ορατά. Στην βιβλιογραφία αναφέρονται σχετικά λίγες μέθοδοι για το πρόβλημα αυτό.

Μία από τις πρώτες προσπάθειες αντιμετώπισης του προβλήματος αυτού έγινε από τους Unser και Stone [125], οι οποίοι χρησιμοποίησαν κάποια στοιχειώδη βήματα προεπεξεργασίας για την απομάκρυνση του θορύβου, απλή παραγωγή για την ενίσχυση των ακμών της γλώσσας και τέλος έναν αλγόριθμο δυναμικού προγραμματισμού για την εκτίμηση των σημείων που ανήκουν στη γλώσσα. Οι Li et al. [71] ανέπτυξαν το *EdgeTrak*, ένα δημόσια διαθέσιμο ημιαυτόματο σύστημα για την παρακολούθηση της γλώσσας σε βίντεο υπερήχων. Το σύστημα αυτό βασίζεται σε ένα *μοντέλο ενεργών περιγραμμάτων* (ή αλλιώς *μοντέλο Φιδιού* (Snake model), λόγω της εμφάνισης της καμπύλης καθώς εξελίσσεται) το οποίο έχει σχεδιαστεί για την εφαρμογή αυτή και συνδυάζει πληροφορίες σχετικά με την κλίση των ακμών της εικόνας, τη φωτεινότητα και τον προσανατολισμό της καμπύλης. Το σύστημα αυτό αποδίδει αρκετά καλά σε υπακολουθίες πλαισίων όπου το ίδιο μέρος της γλώσσας είναι παντού ορατό, αλλά όταν ένα μέρος εξαφανίζεται, εισάγονται σημαντικά σφάλματα στην ανακατασκευασμένη καμπύλη της γλώσσας και στη συνέχεια το σύστημα δεν έχει τη δυνατότητα να ανακάμψει αυτόματα. Για αυτό το λόγο, είναι αναγκαίες συχνές διορθώσεις από τον χρήστη. Οι ίδιοι συγγραφείς πρότειναν λίγο αργότερα [72] διάφορες βελτιώσεις, χρησιμοποιώντας την μεθοδολογία των επιπεδοσυνόλων.

Βελτιώσεις στην μέθοδο του *EdgeTrak* εισήχθησαν επίσης πιο πρόσφατα από τους Aron et al. [7]. Η σχετική μέθοδος βασίζεται επίσης σε μοντέλα Φιδιού, αλλά εφαρμόζει προεπεξεργασία στα πλαίσια υπερήχων για την βελτίωση της ορατότητας της γλώσσας και επιβάλλει περιορισμούς στα σύνορα της εξελισσόμενης καμπύλης για να εμποδίσει την συρρίκνωσή της. Επίσης, η καμπύλη αρχικοποιείται σε κάθε πλαίσιο χρησιμοποιώντας την πληροφορία από την οπτική ροή μεταξύ δύο διαδοχικών πλαισίων και από δύο ηλεκτρομαγνητικούς (HM) αισθητήρες που είναι κολλημένοι στην γλώσσα. Αυτή η μέθοδος, στην οποία θα αναφερόμαστε ως *Περιορισμένο Μοντέλο Φιδιού*, είναι πιο ακριβής από το σύστημα *EdgeTrak*. Από την άλλη, χρειάζεται και αυτή διορθώσεις από τον χρήστη, αν και λιγότερο συχνά από ό,τι το *EdgeTrak*.

2.4 Οπτική Επεξεργασία Βίντεο Νοηματικής Γλώσσας

Το αρχικό στάδιο ενός συστήματος επεξεργασίας βίντεο νοηματικής γλώσσας είναι συνήθως ο εντοπισμός των χεριών, ο οποίος υλοποιείται χρησιμοποιώντας διάφορους τύπους οπτικών χαρακτηριστικών καθώς και, σε πολλές περιπτώσεις, συνδυασμούς τους. Τέτοιου είδους χαρακτηριστικά είναι το χρώμα του δέρματος, η πληροφορία ακμών της εικόνας, το σχήμα, η κίνηση και τα ανατομικά μοντέλα των χεριών. Η κατάτμηση του χρώματος του δέρματος έχει χρησιμοποιηθεί για τον εντοπισμό των χεριών από πολλές μεθόδους, π.χ. [5, 111, 114, 141, 143]. Η πληροφορία χρώματος είναι κατάλληλη λόγω του χαρακτηριστικού χρώματος του δέρματος των ανθρώπων. Κάποιος βαθμός ευρωστίας απέναντι

σε αλλαγές του φωτισμού μπορεί να επιτευχθεί διαλέγοντας χώρους χρώματος, όπως οι *HSV*, *CIE-Lab*, *YCbCr*, που διαχωρίζουν αποτελεσματικά τις συνιστώσες χρωματικότητας (chromaticity), βλ. [63, 121] για βιβλιογραφικές επισκοπήσεις και ποσοτικές αποτιμήσεις διαφορετικών μοντέλων χρωματικότητας του δέρματος. Στα [39, 61] χρησιμοποιούνται κριτήρια (cues) κίνησης, κάνοντας την υπόθεση ότι τα χέρια είναι τα μόνα κινούμενα αντικείμενα, με τον κορμό και το κεφάλι του νοηματοιστή να παραμένουν σχετικά σταθερά και το φόντο του βίντεο να παραμένει στατικό. Στην μέθοδο του [91], τα χέρια εντοπίζονται με βάση μόνο την πληροφορία σχήματος, χρησιμοποιώντας ένα *ενισχυμένο δέντρο ταξινόμησης* (boosted classifier tree).

Το δεύτερο στάδιο προς την αυτόματη κατανόηση των κινήσεων των χεριών του νοηματοιστή είναι συνήθως η παρακολούθηση των χεριών, δηλαδή η αντιστοίχιση από πλαίσιο σε πλαίσιο των κατατηθεισών περιοχών δέρματος ή των εξαγμένων χαρακτηριστικών. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας μοντέλα βασισμένα σε «σταγόνες» (blobs) [114, 119], στην εμφάνιση [61] ή στο περίγραμμα [33, 39] των χεριών. Οι αλληλεπικαλύψεις χεριών και κεφαλιού που συμβαίνουν συχνά κατά τη διάρκεια του αυθόρμητου νοηματοισμού κάνουν το πρόβλημα αυτό ιδιαίτερα δύσκολο. Με στόχο την ευρωστία απέναντι σε τέτοιες αλληλεπικαλύψεις και σε γρήγορες κινήσεις, οι μέθοδοι των [111, 143] εφαρμόζουν πιθανοτική συλλογιστική για την ταυτόχρονη ανάθεση ετικετών κεφαλιού και χεριών στις υποψήφιας περιοχές. Μία άλλη δυσκολία της παρακολούθησης των χεριών έγκειται στην πολυπλοκότητα του φόντου. Οι περισσότερες εργασίες χρησιμοποιούν δεδομένα με ομοιόμορφο φόντο, π.χ. [13, 141, 143], αλλά υπάρχουν διάφορες άλλες μέθοδοι που εφαρμόζουν αφαίρεση του φόντου, π.χ. [1, 33].

Το επόμενο κρίσιμο στάδιο που πρέπει να αντιμετωπιστεί, ταυτόχρονα ή ύστερα από τον εντοπισμό των χεριών, είναι η εξαγωγή χαρακτηριστικών της διαμόρφωσης του κάθε χεριού. Ένα απλό και σύνθετο χαρακτηριστικό θέσης είναι οι 2Δ ή 3Δ συντεταγμένες του κεντροειδούς του χεριού [13, 39, 114, 119, 130]. Επίσης, για τον χαρακτηρισμό της κίνησης του χεριού έχουν χρησιμοποιηθεί χαρακτηριστικά κίνησης [21, 33, 141], όπως τροχιές των εικονοστοιχείων που αντιστοιχούν στο χέρι ή οπτική ροή. Πιο περίπλοκοι περιγραφητές χεριού σχετίζονται με το σχήμα ή την εμφάνιση του χεριού [24, 73]. Οι κατατημημένες εικόνες του χεριού κανονικοποιούνται ως προς το μέγεθος, τον επίπεδο προσανατολισμό ή/και την φωτεινότητα, και συχνά στη συνέχεια εφαρμόζεται Ανάλυση Πρωτεύουσών Συνιστωσών (Principal Component Analysis) για την ελάττωση των διαστάσεων (dimensionality reduction) και την περιγραφική αναπαράσταση του σχήματος του χεριού, π.χ. [17, 20, 41, 44, 46, 118, 140]. Επίσης, στο πρόβλημα παρακολούθησης και αναγνώρισης χειρομορφών έχουν εφαρμοστεί Ενεργά Μοντέλα Σχήματος και Εμφάνισης (Active Shape and Appearance Models) [2, 20, 61]. Αρκετές εργασίες χρησιμοποιούν γεωμετρικές μετρήσεις που σχετίζονται με το σχήμα του χεριού, όπως ροπές σχήματος [60, 114] ή μεγέθη και αποστάσεις μεταξύ των δακτύλων, της παλάμης και του πίσω μέρους του χεριού [12, 13]. Κάποιες άλλες μέθοδοι χρησιμοποιούν το περίγραμμα του χεριού για να εξαγάουν διάφορα αναλλοίωτα χαρακτηριστικά, όπως είναι οι Περιγραφητές Fourier [33, 34]. Μία άλλη δημοφιλής μέθοδος εξαγωγής χαρακτηριστικών σχήματος προκύπτει από την χρήση του λεγόμενου *Ιστογράμματος Κατευθυνόμενων Κλίσεων* (Histogram of Oriented Gradients - HOG) [40], το οποίο είναι ένα από κοινού ιστόγραμμα κβαντισμένων κατευθύνσεων της κλίσης της εικόνας και μετατοπισμένης θέσης, στην γειτονιά κάθε εικονοστοιχείου. Στα άρθρα [24, 73], τέτοιου είδους περιγραφητές χρησιμοποιούνται για τα χέρια των νοηματοιστών και συνδυάζονται τόσο η εμφάνιση όσο

και το σχήμα των χεριών.

Εκτός από τις μεθόδους που επεξεργάζονται ή μοντελοποιούν κατευθείαν τις 2Δ εικόνες των χεριών, υπάρχουν μέθοδοι που βασίζονται σε 3Δ μοντέλα, έτσι ώστε να εκτιμήσουν τις σχετικές γωνίες των δακτύλων και την 3Δ πόζα του χεριού, π.χ. [1, 9, 52, 59, 115, 116]. Αυτές οι μέθοδοι έχουν το πλεονέκτημα ότι έχουν την δυνατότητα οι διαδικασίες της παρακολούθησης και εξαγωγής χαρακτηριστικών του χεριού να είναι ανεξάρτητες από την οπτική γωνία της κάμερας, από την άλλη όμως η διαδικασία ταιριάσματος του μοντέλου κινδυνεύει να είναι ιδιαίτερα αργή και μη-εύρωστη. Υπάρχουν επίσης κάποια 3Δ μοντέλα που χρησιμοποιούν πολλαπλές κάμερες για να διασφαλίσουν 3Δ μετρήσεις και να ξεπεράσουν το πρόβλημα των αλληλεπικαλύψεων [84, 130].

Διάφορες εργασίες δίνουν ιδιαίτερη έμφαση στην εξαγωγή πληροφορίας σχετικά με την πόζα ολόκληρου του σώματος καθώς και την θέση και τις εκφράσεις του προσώπου. Τέτοιου είδους πληροφορία βοηθάει στην εύρωστη παρακολούθηση των χεριών του νοηματιστή και παράλληλα είναι σημαντική για την αναγνώριση της νοηματικής γλώσσας, μιας και σχηματίζει γραμματικά και προσωδιακά στοιχεία της γλώσσας. Στα άρθρα [42, 88, 129] για παράδειγμα, με βάση την ανάλυση και μοντελοποίηση του σχήματος και των εκφράσεων του προσώπου, αναπτύσσονται μέθοδοι για την παρακολούθηση και την εξαγωγή χαρακτηριστικών των εκφράσεων του προσώπου. Άλλες εργασίες, π.χ. [12, 13, 58], συνδυάζουν κριτήρια χρώματος με ανατομικούς περιορισμούς έτσι ώστε να εντοπίσουν, παράλληλα με τα χέρια, το κεφάλι, τον κορμό και τους ώμους. Αυτό τους επέτρεψε να εκφράσουν τη θέση και κίνηση των χεριών σε σχέση με το σώμα του νοηματιστή. Οι Downton και Drouet [43] κατασκεύασαν ένα 3Δ ιεραρχικό κυλινδρικό μοντέλο του πάνω μέρους του ανθρώπινου σώματος και εξήγαγαν τις κινηματικές παραμέτρους υλοποιώντας μία διαδικασία προβολής και ταιριάσματος με τις ακμές της εικόνας που είχαν εντοπιστεί σε κάθε πλαίσιο. Πιο πρόσφατα, ο Buehler και οι συνεργάτες του [23] πρότειναν μία μέθοδο για τον εντοπισμό και την εύρωστη παρακολούθηση των χεριών αλλά και των βραχιόνων του νοηματιστή, κάτι το οποίο διευκολύνει την παρακολούθηση των χεριών σε περιπτώσεις αλληλεπικαλύψεων. Αυτού του είδους η κατεύθυνση στην οπτική επεξεργασία βίντεο νοηματικής γλώσσας φαίνεται ελπιδοφόρα αλλά δεν έχει ακόμα διερευνηθεί σε βάθος. Αντίθετα, υπάρχουν αρκετές αξιόλογες εργασίες που μελετούν το γενικότερο πρόβλημα του εντοπισμού και εκτίμησης των κινήσεων του ανθρώπινου σώματος κατά τη διάρκεια διαφόρων δραστηριοτήτων, π.χ. [22, 51, 62, 120, 126].

Κεφάλαιο 3

Μοντέλο Μη-γραμμικής Διάχυσης για την Αντιστρέψιμη Παρεμβολή Εικόνων

3.1 Εισαγωγή

Η παρεμβολή εικόνας ανήκει στα θεμελιώδη προβλήματα της επεξεργασίας εικόνων και συχνά απαιτείται σε διάφορες λειτουργίες ανάλυσης εικόνων. Είναι λοιπόν ιδιαίτερου ενδιαφέροντος για ποικίλες εφαρμογές όπως επεξεργασία βιοϊατρικών, εναέριων και δορυφορικών εικόνων, αναγνώριση κειμένου, βίντεο υψηλής ευκρίνειας, και υψηλής ποιότητας απεικόνιση και εκτύπωση. Στη παρούσα εργασία, χρησιμοποιούμε τον όρο *παρεμβολή εικόνας* (image interpolation) με την έννοια της διαδικασίας που παίρνει σαν είσοδο μία διακριτή εικόνα και ανακτά μία συνεχή εικόνα ή μία διακριτή εικόνα με υψηλότερη ευκρίνεια. Η δεύτερη περίπτωση, όπου η εικόνα εξόδου είναι διακριτή, εμφανίζεται στην βιβλιογραφία και με διάφορα άλλα ονόματα: *ψηφιακό ζουμ* (digital zooming), *μεγέθυνση εικόνας* (image magnification), *δειγματοληψία προς τα πάνω* (upsampling), *βελτίωση της ευκρίνειας* (resolution enhancement).

Υπάρχει μία μεγάλη ποικιλία μεθόδων παρεμβολής εικόνας, οι οποίες μπορούν να ταξινομηθούν σε δύο κύριες κατηγορίες, τις γραμμικές και τις μη-γραμμικές μεθόδους (βλ. [85] για μία λεπτομερή ανασκόπηση). Οι *γραμμικές* μέθοδοι, π.χ. δικυβική (bicubic) τετραγωνική (quadratic) και spline παρεμβολή, εφαρμόζουν συνέλιξη των δειγμάτων της εικόνας με έναν συγκεκριμένο πυρήνα, το οποίο είναι ισοδύναμο με ένα βαθυπερατό φιλτράρισμα. Οι μέθοδοι αυτές οδηγούν σε γρήγορους αλγορίθμους, αλλά δεν μπορούν να ανακατασκευάσουν αποτελεσματικά το υψίσυχο μέρος των εικόνων και αναπόφευκτα εισάγουν ψεύδεργα (artifacts). Οι μη-γραμμικές μέθοδοι εφαρμόζουν μία επεξεργασία που προσαρμόζεται στις τοπικές γεωμετρικές δομές της εικόνας, με κύριο στόχο να ανακατασκευάσουν με ακρίβεια τις ακμές της εικόνας. Αυτή η κατηγορία περιέχει μεταξύ άλλων μεταβολικές μεθόδους (π.χ. [4, 56, 76]) και μεθόδους βασισμένες σε ΜΔΕ (π.χ. [14, 124]), μερικές από τις οποίες θα παρουσιαστούν σε ακόλουθες ενότητες. Τέτοιου είδους μέθοδοι έχουν επίσης αναπτυχθεί για δύο κοντινά προβλήματα, την ενδοσυμπλήρωση (inpainting) [16, 26, 30] και παρεμβολή διάσπαρτων (scattered) δεδομένων [138].

Σε αυτό το κεφάλαιο, προτείνουμε ένα μη-γραμμικό μοντέλο για την παρεμβολή διανυσματικών εικόνων. Το μοντέλο αυτό βασίζεται σε μία Μερική Διαφορική Εξίσωση ανισοτροπικής διάχυσης και εκτελεί μία παρεμβολή η οποία είναι αντιστρέψιμη. Η λύση της παρεμβολής περιορίζεται στον υποχώρο των συναρτήσεων από τις οποίες μπορεί να ανα-

κτηθεί η διακριτή εικόνα εισόδου, ύστερα από κατάλληλη εξομάλυνση ακολουθούμενη από δειγματοληψία. Η προτεινόμενη ροή διάχυσης κείται σε αυτόν τον υποχώρο ενώ παράλληλα η ένταση και η ανισοτροπικότητά της προσαρμόζονται στις τοπικές μεταβολές και την γεωμετρία των δομών της εικόνας. Η προκύπτουσα μέθοδος ανακατασκευάζει αποτελεσματικά τις πραγματικές δομές της εικόνας και παράγει ικανοποιητικά αποτελέσματα παρεμβολής. Σε σύγκριση με τις κλασικές και άλλες υπάρχουσες μεθόδους παρεμβολής με ΜΔΕ, η μέθοδός μας φαίνεται να αυξάνει την ακρίβεια του αποτελέσματος και να ελαττώνει τα ανεπιθύμητα ψεύδερα, όπως το θόλωμα και την διατάραξη των ακμών. Παρουσιάζουμε εκτεταμένα πειραματικά αποτελέσματα που επιδεικνύουν τις δυνατότητες της μεθόδου, όταν εφαρμόζεται σε γκριζας κλίμακας και έγχρωμες εικόνες.

Στην Ενότητα 3.2, παρουσιάζονται κάποιες μέθοδοι παρεμβολής που συνδέονται με την προτεινόμενη μέθοδο. Η Ενότητα 3.3 είναι αφιερωμένη στην νέα προτεινόμενη μέθοδο παρεμβολής. Τέλος, στην Ενότητα 3.4, περιγράφονται και αναλύονται τα αποτελέσματα από διάφορα πειράματα παρεμβολής, τα οποία αναδεικνύουν την αποτελεσματικότητα της νέας μεθόδου.

3.2 Προκαταρκτικά Θέματα και Υπόβαθρο

3.2.1 Συνθήκη Αντιστρεψιμότητας για την Παρεμβολή

Η σκοπιά υπό την οποία βλέπουμε το πρόβλημα της παρεμβολής εικόνων είναι αντίστοιχη με αυτή των [56, 76]. Η συνεχής λύση $u(x, y)$ της παρεμβολής θα πρέπει να αποδίδει την δεδομένη διακριτή εικόνα χαμηλής ευκρίνειας $z[i_1, i_2]$, ύστερα από ένα βαθυπερατό φιλτράρισμα ακολουθούμενο από δειγματοληψία. Επομένως, απαιτούμε η πράξη της παρεμβολής να είναι αντιστρέψιμη.

Πριν διατυπώσουμε αυτή την *συνθήκη αντιστρεψιμότητας* πιο επίσημα, ας υποθέσουμε ότι η $z[i_1, i_2]$ λαμβάνει πραγματικές τιμές και είναι ορισμένη σε ένα ορθογωνικό πλέγμα $N_x \times N_y$ σημείων με κατακόρυφο και οριζόντιο βήμα h_x και h_y αντίστοιχα. Επίσης, έστω ότι η $u(\mathbf{x})$, όπου $\mathbf{x}=(x, y)$, ορίζεται σε ένα ορθογωνικό χωρίο $\Omega=h_x[\frac{1}{2}, N_x+\frac{1}{2}] \times h_y[\frac{1}{2}, N_y+\frac{1}{2}]$, το οποίο περιέχει τα σημεία του πλέγματος.

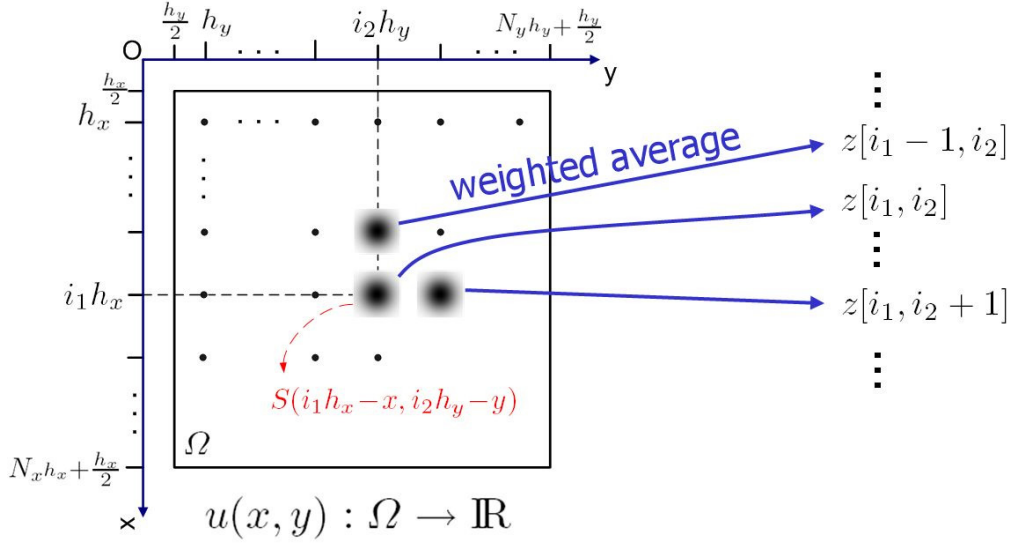
Έτσι, η συνθήκη αντιστρεψιμότητας για την λύση $u(\mathbf{x})$ μπορεί να γραφεί ως εξής (βλ. επίσης Σχήμα 3.1):

$$(S * u)(i_1 h_x, i_2 h_y) = z[i_1, i_2], \quad \forall (i_1, i_2) \in \{1, \dots, N_x\} \times \{1, \dots, N_y\}, \quad (3.1)$$

όπου το $*$ συμβολίζει 2D συνέλιξη και εφεξής θέτουμε $h_x=h_y=1$. Επίσης, ο $S(\mathbf{x}) : \mathbb{R}^2 \rightarrow \mathbb{R}$ είναι ένας πυρήνας εξομάλυνσης που επιτελεί το βαθυπερατό φιλτράρισμα. Έστω ότι ο 2D μετασχηματισμός Fourier του $S(\mathbf{x})$ είναι $\hat{S}(\omega_1, \omega_2) : \mathbb{R}^2 \rightarrow \mathbb{C}$. Υποθέτουμε ότι ο $\hat{S}(\omega_1, \omega_2)$ λαμβάνει μη μηδενικές τιμές για όλες τις συχνότητες της βασικής ζώνης, δηλαδή:

$$\hat{S}(\omega_1, \omega_2) \neq 0, \quad \forall (\omega_1, \omega_2) \in [-\pi, \pi]^2. \quad (3.2)$$

Για παράδειγμα, ο $S(\mathbf{x})$ θα μπορούσε να είναι ο *πυρήνας μέσης τιμής* (mean kernel), δηλαδή $S(\mathbf{x})=\mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]^2}(\mathbf{x})$, όπου το $\mathbf{1}_B(\mathbf{x})$ συμβολίζει την δείκτρια συνάρτηση (indicator function) ενός συνόλου $B \subset \mathbb{R}^2$. Ένα άλλο παράδειγμα για τον $S(\mathbf{x})$ είναι ένας ιστροπικός γκαουσιανός πυρήνας με κατάλληλη διακύμανση.



Σχήμα 3.1: Συνθήκη αντιστρεψιμότητας για την άγνωστη λύση $u(\mathbf{x})$ της παρεμβολής.

Παρόλο που οι τιμές της συνέλιξης στην Εξ. (3.1) επηρεάζονται κυρίως από τις τιμές της $u(\mathbf{x})$ εντός του Ω , πρέπει να υιοθετήσουμε έναν τρόπο επέκτασης της $u(\mathbf{x})$ σε όλο το \mathbb{R}^2 , έτσι ώστε η $S * u$ να είναι καλά ορισμένη. Στην ακόλουθη παρουσίαση, θα θεωρήσουμε δύο εναλλακτικές για αυτήν την επέκταση:

- **Επέκταση Α:** *Απλή Περιοδικοποίηση*. Αυτή η επιλογή γίνεται σιωπηρά όταν εφαρμόζεται ανάλυση Fourier τέτοιων σημάτων [48, 92], αλλά δυστυχώς φέρνει κοντά τιμές της εικόνας από μακρινές περιοχές του Ω .
- **Επέκταση Β:** *Κατοπτρισμός και Ύστερα Περιοδικοποίηση*¹. Αυτή είναι μία τυπική επιλογή στην ανάλυση εικόνων, η οποία αποφεύγει το μειονέκτημα της απλής περιοδικοποίησης, αφού η $u(\mathbf{x})$ είναι συνεχής στα σύνορα της εικόνας. Όταν χρησιμοποιείται για την λύση μιας ΜΔΕ, αυτή η επιλογή είναι ισοδύναμη με την επιβολή μηδενικών συνοριακών συνθηκών τύπου Neumann. Αυτή η επέκταση θα χρησιμοποιηθεί από την μέθοδό μας.

Επομένως, η επεκταμένη εικόνα $u(\mathbf{x})$ είναι $(\tilde{N}_x, \tilde{N}_y)$ -περιοδική, όπου $(\tilde{N}_x, \tilde{N}_y) = (N_x, N_y)$ στην περίπτωση της Επέκτασης Α ενώ $(\tilde{N}_x, \tilde{N}_y) = (2N_x, 2N_y)$ στην περίπτωση της Επέκτασης Β. Επιπρόσθετα, θεωρούμε ότι η διακριτή είσοδος $z[i_1, i_2]$ επεκτείνεται στο \mathbb{Z}^2 χρησιμοποιώντας έναν τύπο επέκτασης παρόμοιο με αυτόν της $u(\mathbf{x})$. Έτσι, η Εξ. (3.1) είναι πλέον έγκυρη για όλα τα $(i_1, i_2) \in \mathbb{Z}^2$.

¹ Αρχικά, η $u(\mathbf{x})$ επεκτείνεται στο $\tilde{\Omega} = [-N_x + \frac{1}{2}, N_x + \frac{1}{2}] \times [-N_y + \frac{1}{2}, N_y + \frac{1}{2}]$ με κατοπτρισμό ως προς τις γραμμές των συνόρων της εικόνας $\{(x, y) : x = 0.5\}$ και $\{(x, y) : y = 0.5\}$:

$$u(0.5 \pm x, 0.5 \pm y) = u(0.5 + x, 0.5 + y), \quad \forall \mathbf{x} \in [0, N_x] \times [0, N_y].$$

Έπειτα, η $u(\mathbf{x})$ επεκτείνεται στο \mathbb{R}^2 με περιοδικοποίηση:

$$u(x + k_1 \cdot 2N_x, y + k_2 \cdot 2N_y) = u(\mathbf{x}), \quad \text{για όλα τα } (k_1, k_2) \in \mathbb{Z}^2.$$

Σε αυτό το σημείο, αξίζει να σημειωθεί ότι η Εξ. (3.1) εκφυλίζεται στην συνθήκη ακριβούς παρεμβολής όταν $S(\mathbf{x})=\delta(\mathbf{x})$ (2Δ συνάρτηση Dirac). Όμως η συνθήκη (3.1) με έναν κατάλληλο πυρήνα εξομάλυνσης, π.χ. έναν γκαουσιανό πυρήνα, μπορεί να γίνει πιο ρεαλιστική, μιας και μοντελοποιεί καλύτερα την διαδικασία ψηφιοποίησης, η οποία είναι το τελικό στάδιο των συστημάτων καταγραφής εικόνων [4]. Επίσης, το αντίστοιχο βαθυπερατό φίλτράρισμα είναι επιθυμητό, διότι περιορίζει τα φαινόμενα φασματικής αναδίπλωσης (aliasing effects) στην καταγραφόμενη εικόνα.

3.2.2 Συνθήκη Αντιστρεψιμότητας στο Πεδίο των Συχνοτήτων

Εδώ, θα εξάγουμε μία απλή έκφραση για την συνθήκη αντιστρεψιμότητας στο πεδίο των συχνοτήτων, η οποία θα είναι χρήσιμη στην ανάλυση των ακόλουθων εννοιών. Έστω λοιπόν $\hat{u}_{m_1, m_2} : \mathbb{Z}^2 \rightarrow \mathbb{C}$ οι συντελεστές της 2Δ σειράς Fourier της $(\tilde{N}_x, \tilde{N}_y)$ -περιοδικής επέκτασης της $u(\mathbf{x})$ (χρησιμοποιώντας είτε την Επέκταση Α είτε την Επέκταση Β). Έστω επίσης $\hat{z}_{n_1, n_2} : \mathbb{Z}^2 \rightarrow \mathbb{C}$ οι συντελεστές της διακριτής 2Δ σειράς Fourier της επεκταμένης $z[i_1, i_2]$, χρησιμοποιώντας την ίδια μέθοδο επέκτασης με την $u(\mathbf{x})$. Ας σημειωθεί ότι, από τη στιγμή που η $z[i_1, i_2]$ είναι διακριτή, η \hat{z}_{n_1, n_2} είναι $(\tilde{N}_x, \tilde{N}_y)$ -περιοδική.

Κάνοντας χρήση της ιδιότητας συνέλιξης του 2Δ μετασχηματισμού Fourier [48, 92], μπορεί να δειχθεί ότι και το $S * u(\mathbf{x})$ είναι $(\tilde{N}_x, \tilde{N}_y)$ -περιοδικό και οι συντελεστές Fourier του είναι $\hat{S}\left(\frac{2\pi m_1}{\tilde{N}_x}, \frac{2\pi m_2}{\tilde{N}_y}\right) \cdot \hat{u}_{m_1, m_2}$. Η επεκταμένη $z[i_1, i_2]$ προκύπτει από τη δειγματοληψία της $S * u(\mathbf{x})$, επομένως μπορούμε να καταλήξουμε [48, 92] στο ότι η συνθήκη αντιστρεψιμότητας (3.1) είναι ισοδύναμη με την ακόλουθη σχέση:

$$\begin{aligned} \sum_{(k_1, k_2) \in \mathbb{Z}^2} \hat{S}\left(\frac{2\pi}{\tilde{N}_x}(n_1 + k_1 \tilde{N}_x), \frac{2\pi}{\tilde{N}_y}(n_2 + k_2 \tilde{N}_y)\right) \cdot \hat{u}_{n_1 + k_1 \tilde{N}_x, n_2 + k_2 \tilde{N}_y} &= \\ &= \hat{z}_{n_1, n_2}, \quad \forall (n_1, n_2) \in \mathbb{Z}^2 \end{aligned} \quad (3.3)$$

Σημειώστε ότι, εξαιτίας της περιοδικότητας και των δύο μελών της συνθήκης (3.3), για να ισχύει η παραπάνω συνθήκη είναι επαρκές να ικανοποιείται μόνο για τα (n_1, n_2) που ανήκουν σε οποιοδήποτε $\tilde{N}_x \times \tilde{N}_y$ ορθογωνικό πλέγμα. Επιπρόσθετα, εάν η άγνωστη συνάρτηση $u(\mathbf{x})$ στην συνθήκη (3.1) θεωρείται πραγματική, πράγμα το οποίο αποτελεί έναν επιθυμητό περιορισμό για την λύση της παρεμβολής, τότε το \hat{u}_{m_1, m_2} πρέπει επίσης να ικανοποιεί:

$$\hat{u}_{-m_1, -m_2} = \overline{\hat{u}_{m_1, m_2}}, \quad \forall (m_1, m_2) \in \mathbb{Z}^2,$$

όπου με \bar{c} συμβολίζεται το συζυγές ενός μιγαδικού c .

Το πρόβλημα της εύρεσης της $u(\mathbf{x})$ στην (3.1) ή ισοδύναμα στην (3.3) είναι ασθενώς ορισμένο, μιας και η συνθήκη αντιστρεψιμότητας ικανοποιείται από απείρως πολλές συναρτήσεις. Έστω $\mathcal{U}_{z, S}$ το σύνολο αυτών των συναρτήσεων. Γίνεται λοιπόν σαφές ότι πρέπει να διατυπωθεί κάποιο επιπρόσθετο κριτήριο επιλογής ανάμεσα στις συναρτήσεις του $\mathcal{U}_{z, S}$.

3.2.3 Παρεμβολή Μηδενικού Γεμίσιματος Συχνοτήτων

Μία απλή γραμμική μέθοδος παρεμβολής ανακλύπει θέτοντας τον επιπρόσθετο περιορισμό η $u(\mathbf{x})$ να είναι ένα βαθυπερατό 2Δ σήμα, σε αντιστοιχία με την θεωρία του

Shannon. Σε αυτή την περίπτωση, η λύση της (3.1) την οποία αποκαλούμε *παρεμβολή μηδενικού γεμίματος συχνοτήτων* (frequency zero-padding interpolation), είναι μοναδική και μπορεί να βρεθεί εύκολα χρησιμοποιώντας την έκφραση (3.3).

Ας θεωρήσουμε μόνο την περίπτωση όπου η $u(\mathbf{x})$ επεκτείνεται στο \mathbb{R}^2 χρησιμοποιώντας την Επέκταση B. Τότε, ο βαθυπερατός περιορισμός μπορεί να γραφεί ως εξής:

$$\hat{u}_{m_1, m_2} = 0, \quad \forall (m_1, m_2) \notin \{-N_x, \dots, N_x\} \times \{-N_y, \dots, N_y\} \quad (3.4)$$

Επίσης, εξαιτίας του κατοπτρισμού που εφαρμόζεται στην $z[i_1, i_2]$, συνεπάγεται ότι $\hat{z}_{\pm N_x, n_2} = 0, \forall n_2$ και $\hat{z}_{n_1, \pm N_y} = 0, \forall n_1$. Συνεπώς, μπορούμε επίσης να ισχυριστούμε ότι $\hat{u}_{\pm N_x, m_2} = 0, \forall m_2$ και $\hat{u}_{m_1, \pm N_y} = 0, \forall m_1$. Άρα, η \hat{u}_{m_1, m_2} περιορίζεται τελικά να παίρνει μη-μηδενικές τιμές μόνο στο $B_1 = \{-N_x + 1, \dots, N_x - 1\} \times \{-N_y + 1, \dots, N_y - 1\}$.

Επομένως, εάν αποτιμήσουμε την συνθήκη (3.3) στα σημεία $(n_1, n_2) \in B_1$, μόνο ο όρος του αθροίσματος που αντιστοιχεί στα $(k_1, k_2) = (0, 0)$ είναι μη μηδενικός. Συμπερασματικά, οι συντελεστές Fourier της παρεμβολής μηδενικού γεμίματος συχνοτήτων δίνονται από την σχέση:

$$\hat{u}_{m_1, m_2} = \begin{cases} \frac{\hat{z}_{m_1, m_2}}{\hat{S}(\frac{2\pi m_1}{N_x}, \frac{2\pi m_2}{N_y})} & , \text{ αν } (m_1, m_2) \in B_1 \\ 0 & , \text{ αν } (m_1, m_2) \in \mathbb{Z}^2 \setminus B_1 \end{cases}, \quad (3.5)$$

μιας και οι τιμές στον παρονομαστή $\hat{S}(\frac{2\pi m_1}{N_x}, \frac{2\pi m_2}{N_y}) \neq 0$, εξαιτίας της Εξ. (3.2).

Στην περίπτωση της Επέκτασης A, μπορεί να εξαχθεί μία έκφραση πολύ παρόμοια με την Εξ. (3.5). Η μόνη διαφορά έγκειται στο ότι, εάν το N_x ή το N_y είναι άρτιο, τότε ο ορισμός των τιμών \hat{u}_{m_1, m_2} στα σύνορα της βαθυπερατής ζώνης είναι πιο περίπλοκος.

Η παρεμβολή μηδενικού γεμίματος συχνοτήτων ανακατασκευάζει τις ακμές της εικόνας χωρίς να τις θολώνει ή να τις αλλοιώνει σημαντικά, αλλά συνήθως εισάγει έντονες τριαντώσεις της φωτεινότητας γύρω από τις ακμές [76]. Η αποκοπή των υψηλών συχνοτήτων που γίνεται από την μέθοδο αυτή είναι ανεπιθύμητη, μιας και η υπόθεση βαθυπερατότητας δεν είναι αληθής για τις περισσότερες εικόνες του πραγματικού κόσμου. Συνεπώς, απαιτείται μία πιο κατάλληλη μέθοδος για την επιλογή ανάμεσα στις συναρτήσεις του $\mathcal{U}_{z, S}$. Τέτοιου είδους μέθοδοι θα παρουσιαστούν στις ακόλουθες ενότητες.

3.2.4 Παρεμβολή Βασισμένη στην Ολική Μεταβολή (OM)

Οι Guichard και Malgouyres [56, 76] πρότειναν σαν κριτήριο για την επιλογή της λύσης της παρεμβολής την ελαχιστοποίηση της Ολικής Μεταβολής - OM (Total Variation - TV):

$$E[u] = \iint_{\Omega} \|\nabla u(\mathbf{x})\| \, d\mathbf{x} \quad (3.6)$$

υπό τον περιορισμό $u \in \mathcal{U}_{z, S}$. Αυτό το πρόβλημα ελαχιστοποίησης λύνεται στο [56] εφαρμόζοντας μία ροή καθοδικής κλίσης υπό περιορισμούς, η οποία περιγράφεται από την ακόλουθη ΜΔΕ:

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} = P_{\mathcal{U}_{0, S}} \left\{ \operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right) \right\}, \quad (3.7)$$

συμπληρωμένη με την αρχική συνθήκη ότι η $u(\mathbf{x}, 0)$ είναι η παρεμβολή μηδενικού γεμίματος συχνοτήτων της $z[i_1, i_2]$. Σημειώνουμε ότι αυτή η μέθοδος, για να χρησιμοποιήσει

την συνθήκη αντιστρεψιμότητας, κάνει την σιωπηρή υπόθεση ότι η $u(\mathbf{x})$ επεκτείνεται στο \mathbb{R}^2 χρησιμοποιώντας την Επέκταση A. Το $P_{\mathcal{U}_{0,S}}\{\cdot\}$ συμβολίζει τον τελεστή της ορθογωνικής προβολής στον υποχώρο $\mathcal{U}_{0,S}$, ο οποίος αντιστοιχεί στην συνθήκη (3.1) με $z[i_1, i_2] = 0$ για όλα τα (i_1, i_2) . Αυτή η προβολή εξασφαλίζει ότι $u(\mathbf{x}, t) \in \mathcal{U}_{z,S}, \forall t > 0$, από τη στιγμή που $u(\mathbf{x}, 0) \in \mathcal{U}_{z,S}$. Οι συγγραφείς προτείνουν δύο εναλλακτικές για τον πυρήνα εξομάλυνσης της συνθήκης (3.1): (α) τον *πυρήνα μέσης τιμής* $\mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]^2}(\mathbf{x})$ και (β) τον *πυρήνα sinc*, ο οποίος παρέχει ένα ιδανικό βαθυπερατό φιλτράρισμα, αφού ο Fourier μετασχηματισμός του είναι $\mathbf{1}_{[-\pi, \pi]^2}(\omega_1, \omega_2)$.

Αυτή η μέθοδος οδηγεί σε ανακατασκευασμένες εικόνες χωρίς αξιοσημείωτα φαινόμενα θολώματος, αφού επιτρέπει ασυνέχειες και συντηρεί τις 1Δ δομές της εικόνας. Παρόλα αυτά, η ελαχιστοποίηση της OM βασίζεται στην υπόθεση ότι η εικόνα είναι ανά τμήματα σχεδόν σταθερή, υπόθεση η οποία έχει ως αποτέλεσμα την υπερβολική εξομάλυνση των ομοιογενών περιοχών. Επίσης, η διάχυση (3.7) ελέγχεται από τον απλό συντελεστή $1/\|\nabla u\|$ και για αυτό δεν μπορεί να απομακρύνει τα μπλοκ ψεύδεργα (block artifacts), ιδιαίτερα στις περιοχές της εικόνας με υψηλή διακύμανση. Τέλος, ο πυρήνας μέσης τιμής μηδενίζεται υπερβολικά απότομα, οπότε η αντίστοιχη προβολή $P_{\mathcal{U}_{0,S}}\{\cdot\}$ ξαναεισάγει μπλοκ ψεύδεργα ενώ ο πυρήνας sinc επεκτείνεται υπερβολικά στον χώρο και περιέχει ταλαντώσεις, με αποτέλεσμα η $P_{\mathcal{U}_{0,S}}\{\cdot\}$ να προκαλεί ταλαντώσεις της φωτεινότητας γύρω από τις ανακατασκευασμένες ακμές.

3.2.5 Μέθοδος Belahmidi-Guichard (BG)

Οι Belahmidi Guichard [14] προσπάθησαν να βελτιώσουν την μέθοδο παρεμβολής της Ολικής Μεταβολής αναπτύσσοντας μία μη-γραμμική ανισοτροπική ΜΔΕ, στην οποία θα αναφερόμαστε εφεξής ως *μέθοδος BG*. Με σκοπό την βελτίωση της ανακατασκευής των ακμών, αυτή η ΜΔΕ εφαρμόζει μία διάχυση της οποίας η ένταση και ο προσανατολισμός προσαρμόζονται στις δομές της εικόνας. Η ΜΔΕ αυτή αρχικοποιείται με την απλή *συγκράτηση μηδενικής τάξης* (zero order hold - ZOH), δηλαδή την επανάληψη των τιμών των εικονοστοιχείων, η οποία επιφέρει έντονα μπλοκ ψεύδεργα. Η συνθήκη αντιστρεψιμότητας (3.1) (με την επιλογή του πυρήνα μέσης τιμής για το $S(\mathbf{x})$) λαμβάνεται υπόψιν προσθέτοντας στην ΜΔΕ έναν κατάλληλο όρο πιστότητας (fidelity term) έτσι ώστε η ροή $u(\mathbf{x}, t)$ να παραμένει κοντά στον υποχώρο $\mathcal{U}_{z,S}$ (βλ. [14] για λεπτομέρειες). Η μέθοδος αυτή ισορροπεί μεταξύ της γραμμικής παρεμβολής στις ομοιογενείς περιοχές και της ανισοτροπικής διάχυσης κοντά στις ακμές, προσπαθώντας να συνδυάσει τα πλεονεκτήματα των δύο αυτών διαδικασιών. Παρόλα αυτά, η διάχυση δεν προσαρμόζεται πάντα με έναν επιθυμητό τρόπο στις πραγματικές δομές της εικόνας και το γεγονός ότι η χρονική της εξέλιξη δεν περιορίζεται στον υποχώρο $\mathcal{U}_{z,S}$ ελαττώνει την ακρίβεια του αποτελέσματος.

3.2.6 Μοντέλο ΜΔΕ των Tschumperlé-Deriche (TD)

Στα άρθρα [122, 124], προτείνεται μία αποτελεσματική μέθοδος ΜΔΕ για την βελτίωση της ποιότητας διανυσματικών εικόνων. Αυτή η μέθοδος, στην οποία θα αναφερόμαστε ως *ΜΔΕ TD*, είναι σχεδιασμένη κυρίως για εφαρμογές αποθορυβοποίησης εικόνων, αλλά παρουσιάζεται εδώ γιατί την χρησιμοποιούμε στην σχεδίαση της δικιάς μας μεθόδου παρεμβολής (Ενότητα 3.3). Το μοντέλο αυτό είναι μία ανισοτροπική διάχυση που χρησιμοποιεί τανυστές για να προσαρμόζεται στις δομές της εικόνας. Έστω $\mathbf{u}(\mathbf{x}, t) = [u_1, \dots, u_M]^T$

η διανυσματική εικόνα εξόδου στον (τεχνητό) χρόνο t και M ο αριθμός των διανυσματικών συνιστωσών (καναλιών) της εικόνας. Τότε, η ΜΔΕ TD μπορεί να περιγραφεί από το ακόλουθο σύνολο συζευγμένων ΜΔΕ :

$$\frac{\partial u_m(\mathbf{x}, t)}{\partial t} = \text{trace} \left(T(J_\rho(\nabla \mathbf{u}_\sigma)) D^2 u_m \right), \quad m=1, \dots, M, \quad (3.8)$$

με αρχική συνθήκη ότι η $\mathbf{u}(\mathbf{x}, 0)$ είναι η διανυσματική εικόνα εισόδου. Το $D^2 u_m$ συμβολίζει την χωρική μήτρα Hessian του καναλιού $u_m(\mathbf{x}, t)$ και T είναι ο 2×2 *τανυστής διάχυσης*:

$$T(J_\rho(\nabla \mathbf{u}_\sigma)) = [1 + (\mathcal{N}/K)^2]^{-\frac{1}{2}} \mathbf{w}_- \mathbf{w}_-^T + [1 + (\mathcal{N}/K)^2]^{-1} \mathbf{w}_+ \mathbf{w}_+^T, \quad (3.9)$$

όπου $\mathcal{N} = \sqrt{\lambda_+ + \lambda_-}$ και K είναι μία σταθερά κατωφλίου παρόμοια με την σταθερά του συντελεστή διάχυσης του [95]². Επίσης, $\lambda_- \leq \lambda_+$ και $\mathbf{w}_-, \mathbf{w}_+$ είναι οι ιδιοτιμές και τα μοναδιαία ιδιοδιανύσματα του 2×2 *δομικού τανυστή* :

$$J_\rho(\nabla \mathbf{u}_\sigma) = G_\rho * \sum_{m=1}^M \nabla(G_\sigma * u_m) (\nabla(G_\sigma * u_m))^T, \quad (3.10)$$

όπου η συνέλιξη του πυρήνα G_ρ με τον αντίστοιχο 2×2 πίνακα ορίζεται μέσω της συνέλιξης του G_ρ με κάθε στοιχείο του πίνακα χωριστά. Οι 2Δ ισοτροπικοί γκαουσιανοί πυρήνες $G_\sigma(\mathbf{x})$ και $G_\rho(\mathbf{x})$ έχουν τυπική απόκλιση σ και ρ αντίστοιχα³ :

$$G_\sigma(\mathbf{x}) \triangleq \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).$$

Ο δομικός τανυστής $J_\rho(\nabla \mathbf{u}_\sigma)$ μετράει την τοπική γεωμετρία των δομών της εικόνας [135]. Χάρη στις συνελίξεις με τους G_σ και G_ρ , ο $J_\rho(\nabla \mathbf{u}_\sigma)$ δεν επηρεάζεται από λεπτομέρειες της εικόνας που είναι μικρότερες από $O(\sigma)$ ενώ λαμβάνει υπόψιν τις μεταβολές της εικόνας εντός μιας γειτονιάς μεγέθους $O(\rho)$ γύρω από κάθε σημείο. Τα ιδιοδιανύσματα \mathbf{w}_- και \mathbf{w}_+ περιγράφουν την κατεύθυνση ελάχιστης και μέγιστης διανυσματικής μεταβολής της \mathbf{u} και οι ιδιοτιμές λ_- και λ_+ παρέχουν εκτιμήσεις των μεταβολών αυτών. Ο όρος \mathcal{N} που εξαρτάται από τις ιδιοτιμές είναι ένας εκτιμητής των ακμών και της έντασής τους, ο οποίος γενικεύει αποτελεσματικά το μέτρο της κλίσης $\|\nabla u\|$. Έτσι, η διάχυση είναι ισχυρή και ισοτροπική στις ομοιογενείς περιοχές (μικρό \mathcal{N}), αλλά ασθενέστερη και ανισοτροπικά προσανατολισμένη από τις δομές της εικόνας κοντά στις ακμές (μεγάλο \mathcal{N}). Συμπερασματικά, αυτή η μέθοδος προσφέρει έναν αρκετά ευέλικτο και αποτελεσματικό έλεγχο στην διαδικασία της διάχυσης (βλ. [124] για περισσότερες λεπτομέρειες).

Ανάμεσα σε διάφορες άλλες εφαρμογές, οι συγγραφείς των [122, 124] εφαρμόζουν το γενικό τους μοντέλο ΜΔΕ και στην παρεμβολή εικόνων (θα αναφερόμαστε στην προκύπτουσα μέθοδο παρεμβολής ως *παρεμβολή TD*). Αυτή η μέθοδος αντιμετωπίζει την παρεμβολή σαν μία ειδική περίπτωση του προβλήματος της ενδοσυμπλήρωσης εικόνων [16].

²στην παρουσίαση μας αυτή υιοθετούμε μία ελαφρώς πιο γενική εκδοχή σε σχέση με την πρωτότυπη εκδοχή των [122, 124], όπου $K = 1$.

³η πρωτότυπη εκδοχή του μοντέλου αντιστοιχεί στην περίπτωση $\sigma=0$, αλλά εδώ χρησιμοποιούμε για τον δομικού τανυστή την γενικότερη εκδοχή τού [135].

Θέτει τον περιορισμό ότι η λύση πρέπει να συμπίπτει με την είσοδο στα κατάλληλα εικονοστοιχεία του νέου πυκνότερου πλέγματος (συνθήκη ακριβούς παρεμβολής). Επομένως, το χωρίο ενδοσυμπλήρωσης (δηλ. το χωρίο όπου οι τιμές της εικόνας θεωρούνται άγνωστες) αποτελείται από τα υπόλοιπα εικονοστοιχεία. Οι τιμές της εικόνας σε αυτό το χωρίο επεξεργάζονται σύμφωνα με την ΜΔΕ (3.8), αλλά με τον ακόλουθο τροποποιημένο τανυστή διάχυσης [122]:

$$T(J_\rho(\nabla \mathbf{u}_\sigma)) = [1 + (\mathcal{N}/K)^2]^{-\frac{1}{2}} \mathbf{w}_- \mathbf{w}_-^T . \quad (3.11)$$

Σαν αρχική συνθήκη $\mathbf{u}(\mathbf{x}, 0)$ επιλέγεται η διγραμμική παρεμβολή της εικόνας εισόδου και η τελική λύση εξάγεται από την κατάσταση ισορροπίας της ΜΔΕ. Σε αντίθεση με την αποτελεσματικότητα του μοντέλου ΜΔΕ TD για την αποθορυβοποίηση, η προκύπτουσα μέθοδος παρεμβολής έχει κάποιες κρίσιμες ανεπάρκειες. Η αρχικοποίηση από την διγραμμική παρεμβολή περιέχει ακμές με σημαντικό θόλωμα. Επίσης, η πληροφορία της κάθε τιμής εισόδου $z[i_1, i_2]$ δεν απλώνεται σε όλα τα αντίστοιχα εικονοστοιχεία του πυκνότερου πλέγματος: Κατά τη διάρκεια της εξέλιξης της ΜΔΕ, οι τιμές της \mathbf{u} παραμένουν σταθερές στα εικονοστοιχεία του αρχικού αραιότερου πλέγματος αλλά μεταβάλλονται χωρίς κανένα περιορισμό στα υπόλοιπα. Επιπλέον, ο τανυστής διάχυσης (3.11) είναι πλήρως ανισοτροπικός ακόμα και σε περιοχές με μικρές μεταβολές της εικόνας, επομένως διαστρεβλώνει τις δομές της εικόνας και δημιουργεί ψεύτικες ακμές.

3.3 Προτεινόμενη ΜΔΕ Ανισοτροπικής Διάχυσης-Προβολής

Οι προαναφερθείσες μέθοδοι παρεμβολής με ΜΔΕ έχουν συνήθως καλύτερη απόδοση από τις γραμμικές μεθόδους, μιας και ανακατασκευάζουν τις ακμές χωρίς να τις θολώνουν σημαντικά. Ωστόσο, σε μερικές περιπτώσεις, καταλήγουν και αυτές σε ψεύδεργα, όπως υπερβολική εξομάλυνση των ομοιογενών περιοχών, μπλοκ ψεύδεργα ή διαστρέβλωση ακμών. Με σκοπό την περαιτέρω βελτίωση της ποιότητας της παρεμβολής, προτείνουμε ένα νέο μοντέλο ΜΔΕ το οποίο επιτελεί μη-γραμμική παρεμβολή βασισμένο σε έναν συνδυασμό της μοντελοποίησης βάσει της συνθήκης αντιστρεψιμότητας και της ανισοτροπικής διάχυσης. Το μοντέλο είναι σχεδιασμένο για γενικές διανυσματικές εικόνες και επεξεργάζεται τα διαφορετικά κανάλια της εικόνας με έναν συζευγμένο τρόπο.

Πιο συγκεκριμένα, η σχεδίαση της μεθόδου μας έχει βασιστεί στην παρατήρηση ότι η ΜΔΕ της παρεμβολής OM (3.7) μπορεί να εξαχθεί μέσω μιας μη-μεταβολικής οπτικής γωνίας: πρόκειται στην ουσία για μία τροποποίηση της ακόλουθης ΜΔΕ Ολικής Μεταβολής με μηδενικό συντελεστή ($\lambda=0$) για τον όρο πιστότητας (fidelity term) [106]:

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} = \operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right) , \quad (3.12)$$

η οποία μπορεί να ιδωθεί σαν μία ειδική περίπτωση της γενικής μη-γραμμικής διάχυσης των Perona-Malik [95]. Αυτή η τροποποίηση γίνεται αντικαθιστώντας το δεξί μέλος της ΜΔΕ με την προβολή του στον υποχώρο $\mathcal{U}_{0,S}$. Χάρη στην προβολή αυτή, ολόκληρη η ροή παραμένει στον $\mathcal{U}_{z,S}$, αρκεί αρχικά να έχουμε $u(\mathbf{x}, 0) \in \mathcal{U}_{z,S}$.

Ακολουθούμε μία παρόμοια προσέγγιση στο πρόβλημα, χρησιμοποιώντας το δεξί μέλος κάποιας άλλης ΜΔΕ, πιο αποτελεσματικής για την παρούσα εφαρμογή, αντί της ΜΔΕ

ΟΜ (3.12). Στο [101] χρησιμοποιήσαμε το δεξί μέλος της ΜΔΕ TD, η οποία εφαρμόζει μια πιο ευέλικτη εξομάλυνση. Όμως, στο πιο πρόσφατο άρθρο [102], χρησιμοποιούμε το δεξί μέλος της ακόλουθης τροποποιημένης εκδοχής της ΜΔΕ TD:

$$\partial u_m(\mathbf{x}, t)/\partial t = \operatorname{div} \left(T(J_\rho(\nabla \mathbf{u}_\sigma)) \nabla u_m \right), \quad m=1, \dots, M, \quad (3.13)$$

με τον τανυστή $T(J_\rho(\nabla \mathbf{u}_\sigma))$ να δίνεται ξανά από την Εξ. (3.9). Αυτή η τροποποίηση βασίζεται στο γενικό μοντέλο ανισοτροπικής διάχυσης του [135]. Όπως αποκαλύφθηκε από τα πειράματά μας (βλ. Ενότητα 3.4), η έκφραση της διάχυσης με την χρήση του τελεστή div αντί του τελεστή trace οδηγεί σε ελαφριά βελτίωση του αποτελέσματος της παρεμβολής, ειδικά σε μικρές περιοχές της εικόνας όπου ο τανυστής T έχει μία έντονη χωρική μεταβολή. Επομένως, στην περίπτωση της χρήσης του τελεστή div , η προσαρμογή της διάχυσης στις δομές της εικόνας είναι πιο επιθυμητή.

Σε αυτό το σημείο πρέπει να σημειωθεί ότι η χρήση του δεξιού μέλους της ΜΔΕ (3.13) στον σχεδιασμό της προτεινόμενης μεθόδου αποτελεί απλά μία από τις πολλές πιθανές επιλογές. Οποιαδήποτε άλλη ΜΔΕ εξομάλυνσης θα μπορούσε να χρησιμοποιηθεί και η επακόλουθη ανάλυση της μεθόδου θα ήταν πολύ παρόμοια.

3.3.1 Συνθήκη Αντιστρεψιμότητας για το Προτεινόμενο Μοντέλο

Χρησιμοποιούμε μία άμεση γενίκευση της συνθήκης (3.1) στις διανυσματικές εικόνες: Κάθε κανάλι της λύσης $\mathbf{u}(\mathbf{x})$ της παρεμβολής περιορίζεται να ικανοποιεί ανεξάρτητα την Εξ. (3.1). Έτσι, αυτή η γενικευμένη συνθήκη αντιστρεψιμότητας μπορεί να γραφθεί ως εξής:

$$(S * u_m)(i_1, i_2) = z_m[i_1, i_2], \quad (3.14)$$

για όλα τα σημεία $(i_1, i_2) \in \{1, \dots, N_x\} \times \{1, \dots, N_y\}$ και τα κανάλια $m \in \{1, \dots, M\}$. $z_m[i_1, i_2]$ και $u_m(\mathbf{x})$ είναι η m -οστή από τις M συνιστώσες της διακριτής εισόδου και του αποτελέσματος της παρεμβολής αντίστοιχα. Έστω $\mathbf{U}_{z,S}$ το σύνολο των διανυσματικών εικόνων $\mathbf{u}(\mathbf{x})$ που ικανοποιούν την γενικευμένη συνθήκη αντιστρεψιμότητας (3.14).

Ας σημειωθεί ότι θεωρούμε ότι η $\mathbf{u}(\mathbf{x})$ επεκτείνεται στο \mathbb{R}^2 χρησιμοποιώντας την Επέκταση B. Επίσης, για την συνθήκη αντιστρεψιμότητας (3.14), επιλέγουμε έναν γκαουσιανό πυρήνα εξομάλυνσης $S(\mathbf{x}) = G_r(\mathbf{x})$ με τυπική απόκλιση r . Με αυτήν την επιλογή, η διαδικασία ψηφιοποίησης κατά τη διάρκεια της καταγραφής της εικόνας μπορεί να μοντελοποιηθεί αξιόπιστα. Η τυπική απόκλιση r δεν πρέπει να είναι ούτε πολύ μικρή αλλά ούτε πολύ μεγάλη. Εάν η r είναι υπερβολικά μικρή, ο $S(\mathbf{x})$ είναι υπερβολικά συγκεντρωμένος χωρικά και η πληροφορία της κάθε τιμής εισόδου $z[i_1, i_2]$ δεν εξαπλώνεται κατάλληλα σε όλη τη γειτονιά του σημείου (i_1, i_2) . Από την άλλη, εάν η r είναι υπερβολικά μεγάλη, ο $S(\mathbf{x})$ εκτείνεται υπερβολικά και επιτελεί ένα βαθυπερατό φιλτράρισμα με πολύ μικρό εύρος ζώνης και επομένως η συνθήκη αντιστρεψιμότητας δεν είναι ρεαλιστική.

Ο γκαουσιανός πυρήνας $G_r(\mathbf{x})$, συγκρινόμενος με τον πυρήνα μέσης τιμής και τον sinc πυρήνα των μεθόδων παρεμβολής ΟΜ και ΒG, οδηγεί σε μία πιο αξιόπιστη εκδοχή της συνθήκης αντιστρεψιμότητας, όταν επιλέγεται ένα κατάλληλο r . Πιο συγκεκριμένα, αποφεύγει την απότομη αποκοπή είτε στον τομέα του χώρου (όπως στον πυρήνα μέσης τιμής) είτε στον τομέα των συχνοτήτων (όπως στον πυρήνα sinc), και μηδενίζεται ομαλά ενώ παράλληλα είναι καλά εντοπισμένος στον χώρο.

Η χρήση της γκαουσιανής για τον πυρήνα εξομάλυνσης αποτελεί μία ελαφριά διαφορά της πιο πρόσφατης εκδοχής της προτεινόμενης μεθόδου [102] σε σχέση με την παλαιότερη [101]: Στο [101], είχαμε χρησιμοποιήσει έναν *περικομμένο* (truncated) γκαουσιανό πυρήνα, ο οποίος είχε περιοχή υποστήριξης (support) μεγέθους $h_x \times h_y = 1 \times 1$ στο συνεχή χώρο (ίδιο με το μέγεθος των εικονοστοιχείων της διακριτής εικόνας εισόδου). Η χρήση γκαουσιανής χωρίς περικοπή αποτελεί μία βελτίωση, μιας και η περικοπή εισάγει έναν σχετικά απότομο μηδενισμό στον πυρήνα (αν και λιγότερο απότομο σε σχέση με τον πυρήνα μέσης τιμής). Αυτή η περικοπή είχε γίνει στο [101] με σκοπό την εξαγωγή μίας απλής έκφρασης του τελεστή προβολής στο πεδίο του χώρου. Στο [102] ωστόσο, εξάγαμε μια γενική έκφραση για την προβολή (που δεν χρειάζεται καμία παραδοχή σχετικά με την περιοχή υποστήριξης του πυρήνα), χρησιμοποιώντας μία πιο πολύπλοκη ανάλυση στο πεδίο της συχνότητας (βλ. Ενότητες 3.3.3 και 3.5).

3.3.2 ΜΔΕ του Προτεινόμενου Μοντέλου

Προτείνουμε λοιπόν το αποτέλεσμα της παρεμβολής να εξάγεται από την λύση ισορροπίας του ακόλουθου συστήματος συζευγμένων ΜΔΕ (το οποίο ονομάζουμε *ΜΔΕ Ανισοτροπικής Διάχυσης-Προβολής*):

$$\frac{\partial u_m(\mathbf{x}, t)}{\partial t} = P_{\mathcal{U}_{0,S}} \left\{ \operatorname{div} \left(T(J_\rho(\nabla \mathbf{u}_\sigma)) \nabla u_m \right) \right\}, \quad m=1, \dots, M, \quad (3.15)$$

όπου το $P_{\mathcal{U}_{0,S}}\{\cdot\}$ συμβολίζει τον τελεστή της ορθογωνικής προβολής στον υποχώρο $\mathcal{U}_{0,S}$ και οι τανυστές $T(J_\rho(\nabla \mathbf{u}_\sigma))$ και $J_\rho(\nabla \mathbf{u}_\sigma)$ δίνονται ξανά από τις Εξισώσεις (3.9) και (3.10) αντίστοιχα. Διαλέξαμε τις ακόλουθες αρχικές συνθήκες για την Εξ. (3.15): κάθε $u_m(\mathbf{x}, 0)$ εξάγεται από την παρεμβολή μηδενικού γεμίσματος συχνοτήτων της $z_m[i_1, i_2]$ (3.5). Σημειώστε ότι $\mathbf{u}(\mathbf{x}, 0) \in \mathcal{U}_{z,S}$, οπότε η προβολή $P_{\mathcal{U}_{0,S}}\{\cdot\}$ εξασφαλίζει ότι $\mathbf{u}(\mathbf{x}, t) \in \mathcal{U}_{z,S}$, $\forall t > 0$.

Η παραπάνω αρχικοποίηση, η οποία είναι παρόμοια με αυτήν της ΜΔΕ (3.7) του [56], μπορεί εύκολα να υπολογιστεί αριθμητικά και περιέχει μία αρκετά αξιόπιστη ανακατασκευή των ακμών της εικόνας (βλ. επίσης την συζήτηση της Ενότητας 3.3.4). Σε αντίθεση όμως με το [56], υποθέτουμε, όπως έχει προαναφερθεί, ότι η $u(\mathbf{x})$ επεκτείνεται στο \mathbb{R}^2 κάνοντας χρήση της Επέκτασης Β αντί της Επέκτασης Α. Ο κατοπτρισμός που έχει προστεθεί στην Επέκταση Β προσφέρει μία ελαφριά βελτίωση στην παρεμβολή μηδενικού γεμίσματος συχνοτήτων, μιας και περιορίζει τα ψεύδεργα κωδωνισμού (ringing effects - ταλαντώσεις της φωτεινότητας) κοντά στα σύνορα της εικόνας.

3.3.3 Έκφραση του Τελεστή Προβολής

Είναι φανερό ότι χρειάζεται μία έκφραση για τον τελεστή προβολής στην (3.15). Όπως δείχνουμε στην Ενότητα 3.5, η προβολή $P_{\mathcal{U}_{0,S}}\{v\}$ μιας (N_x, N_y) -περιοδικής συνάρτησης $v(\mathbf{x})$ στον υποχώρο $\mathcal{U}_{0,S}$ μπορεί να προκύψει από τις ακόλουθες εξισώσεις:

$$P_{\mathcal{U}_{0,S}}\{v\} = v(\mathbf{x}) - w(\mathbf{x}), \quad (3.16)$$

όπου η $w(\mathbf{x})$ είναι και αυτή μία $(\tilde{N}_x, \tilde{N}_y)$ -περιοδική συνάρτηση με 2Δ συντελεστές Fourier:

$$\hat{w}_{m_1, m_2} = \left\{ \sum_{(k_1, k_2) \in \mathbb{Z}^2} \hat{\phi} \left(\frac{2\pi m_1}{\tilde{N}_x} + k_1 2\pi, \frac{2\pi m_2}{\tilde{N}_y} + k_2 2\pi \right) \cdot \hat{v}_{m_1 + k_1 \tilde{N}_x, m_2 + k_2 \tilde{N}_y} \right\} \cdot \hat{\phi} \left(\frac{2\pi m_1}{\tilde{N}_x}, \frac{2\pi m_2}{\tilde{N}_y} \right), \quad (3.17)$$

και $\phi(\mathbf{x})$ είναι μία συνάρτηση με μετασχηματισμό Fourier:

$$\hat{\phi}(\omega_1, \omega_2) = \left\{ \sum_{(k_1, k_2) \in \mathbb{Z}^2} \left| \hat{S}(\omega_1 + k_1 2\pi, \omega_2 + k_2 2\pi) \right|^2 \right\}^{-\frac{1}{2}} \cdot \overline{\hat{S}(\omega_1, \omega_2)}. \quad (3.18)$$

Σημειώστε ότι η φασματική ενέργεια της $\phi(\mathbf{x})$ είναι συγκεντρωμένη στις χαμηλές συχνότητες. Στην πράξη, ο προσδιορισμός (3.17) της $w(\mathbf{x})$ στο πεδίο της συχνότητας είναι πιο κατάλληλος για την υλοποίηση της προβολής. Για να αποκτήσουμε όμως περισσότερη διαίσθηση, μπορούμε να την εκφράσουμε επίσης στο πεδίο του χώρου. Έτσι, χρησιμοποιώντας την ιδιότητα της συνέλιξης του μετασχηματισμού Fourier και την σχέση ανάμεσα στους 2Δ μετασχηματισμούς Fourier ενός σήματος και της δειγματοληψίας του [48, 92], μπορεί να δειχθεί ότι η Εξ. (3.17) ισοδυναμεί με την έκφραση:

$$w(\mathbf{x}) = (\text{III} \cdot (\phi * v)) * \phi(\mathbf{x}), \quad (3.19)$$

όπου $\text{III}(x, y) = \sum_{(i, j)} \delta(x - i, y - j)$ είναι ένα 2Δ κρουστικό τρένο (impulse train) για την δειγματοληψία σε ένα ορθογωνικό πλέγμα με μοναδιαία οριζόντια και κάθετη περίοδο δειγματοληψίας. Σαν ένα απλό παράδειγμα, αν θεωρήσουμε ότι ο $S(\mathbf{x})$ είναι ο πυρήνας sinc, τότε $\phi(\mathbf{x}) = S(\mathbf{x})$ και η παραπάνω εξίσωση απλοποιείται σε $w = \phi * v$, δηλαδή το w γίνεται μία βαθυπερατή εκδοχή της v και το $P_{U_{0,S}}\{v\}$ περιέχει μόνο το υψιπερατό μέρος της v .

Από την Εξ. (3.18) προκύπτει ότι η ϕ είναι πραγματική. Επομένως, αν η v είναι πραγματική, τότε τα $w(\mathbf{x})$ και $P_{U_{0,S}}\{v\}$ είναι επίσης πραγματικά, εξαιτίας των (3.19) και (3.16). Αυτό αποτελεί μία ιδιότητα επιθυμητή για την εφαρμογή που μελετάμε εδώ, μιας και ασχολούμαστε με εικόνες που λαμβάνουν πραγματικές τιμές και θέλουμε και το αποτέλεσμα της παρεμβολής να είναι πραγματικό.

Στην ειδική περίπτωση όπου ο πυρήνας εξομάλυνσης $S(\mathbf{x})$ είναι ένας γκαουσιανός πυρήνας $G_r(\mathbf{x})$, όπως έχουμε επιλέξει στην μέθοδό μας, ισχύει ότι $\hat{S}(\omega_1, \omega_2) = \frac{2\pi}{r^2} G_{1/r}(\omega_1, \omega_2)$. Από τη σχέση αυτή, εξάγουμε μία αναλυτική έκφραση για το $\hat{\phi}(\omega_1, \omega_2)$ και ύστερα την χρησιμοποιούμε στην Εξ. (3.17).

3.3.4 Ιδιότητες του Μοντέλου

Όπως έχει ήδη προαναφερθεί, η παρεμβολή μηδενικού γεμίσματος συχνοτήτων, την οποία χρησιμοποιούμε ως αρχικοποίηση της ΜΔΕ (3.15), ανακατασκευάζει τις ακμές της εικόνας χωρίς να τις θολώνει ή να τις αλλοιώνει σημαντικά, αλλά εισάγει ισχυρές ταλαντώσεις γύρω από τις ακμές (βλ. [76]). Μπορεί λοιπόν να ιδωθεί σαν ένα επιθυμητό αποτέλεσμα παρεμβολής διαταραγμένο από σημαντικό ποσό θορύβου. Σκοπός της προτεινόμενης ΜΔΕ (3.15) είναι να βελτιώσει την ποιότητα της εικόνας $\mathbf{u}(\mathbf{x}, 0)$ απομακρύνοντας τις ταλαντώσεις αυτές.

Ας σημειωθεί επίσης ότι, όπως φαίνεται από την Εξ. (3.16), η προβολή $P_{\mathcal{U}_{0,S}}\{v\}$ αφαιρεί την συνιστώσα της v που δεν συμμορφώνεται με την συνθήκη αντιστρεψιμότητας. Αυτή η αφαίρεση δεν επηρεάζει τα βασικά χαρακτηριστικά της εξομάλυνσης την οποία η ταχύτητα $v_m = \text{div}(T \nabla u_m)$ τείνει να εφαρμόσει στην εικόνα. Συνεπώς, παρόμοια με την ΜΔΕ TD (3.8), ο ανισοτροπικός τανυστής T ωθεί την διάχυση να είναι ισχυρή και ισοτροπική στις ομοιογενείς περιοχές, αλλά σχετικά ασθενής και ανισοτροπικά προσανατολισμένη από τις δομές της εικόνας κοντά στις ακμές. Με αυτόν τον τρόπο, η προτεινόμενη ΜΔΕ απομακρύνει τις ανεπιθύμητες ταλαντώσεις και ταυτόχρονα συντηρεί τις σημαντικές δομές της εικόνας. Με άλλα λόγια, η προτεινόμενη ΜΔΕ μπορεί να θεωρηθεί σαν μία ροή προς στοιχεία του $\mathcal{U}_{z,S}$ με «καλύτερη» οπτική ποιότητα.

Επιπρόσθετα, έχουμε διαπιστώσει πειραματικά ότι η βέλτιστη εξομαλυμένη εικόνα επιτυγχάνεται σε πεπερασμένο χρόνο, όπου στην πράξη η ροή φτάνει σε ισορροπία (βλ. Σχήμα 3.8(a) για ένα παράδειγμα), χάρη στον όρο που η $P_{\mathcal{U}_{0,S}}\{v\}$ αφαιρεί από την ταχύτητα v . Επομένως, δεν χρειάζεται να προσδιοριστεί σαν παράμετρος της μεθόδου ο χρόνος παύσης της διάχυσης.

3.3.5 Αριθμητική Υλοποίηση

Στην αριθμητική υλοποίηση, το συνεχές αποτέλεσμα $\mathbf{u}(\mathbf{x})$ της προτεινόμενης μεθόδου προσεγγίζεται από μία διακριτή εικόνα $\mathbf{u}[i'_1, i'_2]$, η οποία ορίζεται σε ένα πλέγμα πυκνότερο από αυτό της εικόνας εισόδου $\mathbf{z}[i_1, i_2]$. Άρα, η $\mathbf{u}[i'_1, i'_2]$ προσφέρει μία διακριτή παρεμβολή (δειγματοληψία προς τα πάνω) της εισόδου $\mathbf{z}[i_1, i_2]$. Θεωρούμε εδώ μόνο την περίπτωση όπου το οριζόντιο και κατακόρυφο βήμα του πλέγματος της $\mathbf{z}[i_1, i_2]$ είναι πολλαπλάσιο του βήματος του πλέγματος της $\mathbf{u}[i'_1, i'_2]$ κατά έναν ακέραιο παράγοντα d , τον οποίο αποκαλούμε *παράγοντα μεγέθυνσης* (zoom factor). Αυτό σημαίνει ότι μεγεθύνουμε την εικόνα εισόδου $d \times d$ φορές. Από εδώ και στο εξής, υποθέτουμε, για χάρη της απλότητας, ότι το πυκνότερο πλέγμα της $\mathbf{u}[i'_1, i'_2]$ έχει μοναδιαία βήματα και άρα το πλέγμα της εισόδου $\mathbf{z}[i_1, i_2]$ έχει βήματα ίσα με d .

Για να διακριτοποιήσουμε τη ΜΔΕ (3.15), κάνουμε χρήση ενός *άμεσου* (explicit) αριθμητικού σχήματος με πεπερασμένες διαφορές, παρόμοια με ό,τι περιγράφεται στα [99, 122]. Για λόγους ευστάθειας, το διακριτό χρονικό βήμα δt επιλέγεται κατάλληλα μικρό (χρησιμοποιείται η τυπική τιμή $\delta t=0.2$). Λόγω του ότι η εικόνα εξόδου δίνεται από την κατάσταση ισορροπίας, σταματάμε την επαναληπτική διαδικασία όταν η \mathbf{u}^{n+1} διαφέρει από την \mathbf{u}^n κατά μία μικρή σταθερά, ως προς μία κατάλληλη νόρμα διαφοράς, π.χ. την ενεργό τιμή (root mean square - RMS) λαμβάνοντας όλα τα σημεία του πλέγματος και τα κανάλια των εικόνων.

Αξίζει τέλος να σημειωθεί ότι, μιας και στην αριθμητική υλοποίηση της προτεινόμενης ΜΔΕ η εξαγόμενη λύση είναι διακριτή αντί για συνεχής, μπορούμε να αναπτύξουμε ένα εξολοκλήρου διακριτό πλαίσιο εργασίας για την συνθήκη αντιστρεψιμότητας, την παρεμβολή μηδενικού γεμίσματος συχνοτήτων και τον τελεστή προβολής, πλαίσιο το οποίο είναι πολύ παρόμοιο με το συνεχές πλαίσιο που παρουσιάστηκε στις Ενότητες 3.2.1, 3.2.3 και 3.3.3. Οι διακριτές εκφράσεις που εξαγονται από αυτό το πλαίσιο είναι ακριβώς αντίστοιχες προς τις συνεχείς εκφράσεις που παρουσιάσαμε. Επομένως, στα πειράματά μας, υλοποιούμε την παρεμβολή μηδενικού γεμίσματος συχνοτήτων και τον τελεστή προβολής βασισμένοι σε αυτό το διακριτό πλαίσιο και κάνοντας χρήση του 2D Ταχέος Μετασχηματισμού Fourier (2D Fast Fourier Transform).

3.4 Πειραματικά Αποτελέσματα και Συγκρίσεις

3.4.1 Πειραματικό Πλαίσιο

Για να συγκρίνουμε τις μεθόδους παρεμβολής και να εξάγουμε μετρήσεις απόδοσης, υιοθετούμε το ακόλουθο πλαίσιο: Διαλέγουμε μία εικόνα αναφοράς με μία σχετικά καλή ανάλυση και αμελητέο θόρυβο. Ελαττώνουμε τις διαστάσεις αυτής της εικόνας κατά έναν ακέραιο παράγοντα d (δηλ. η εικόνα συρρικνώνεται στο $\frac{1}{d} \times \frac{1}{d}$ του αρχικού της μεγέθους), χρησιμοποιώντας μία διαδικασία αποδεκατισμού (decimation), δηλαδή βαθυπερατό φίλτράρισμα ακολουθούμενο από διακριτή δειγματοληψία. Υλοποιούμε το βαθυπερατό φίλτράρισμα μέσω της συνέλιξης με ένα κυβικό spline, το οποίο οδηγεί σε μία αξιόπιστη διαδικασία αποδεκατισμού. Τέλος, εφαρμόζουμε τις μεθόδους παρεμβολής για να μεγεθύνουμε την αποδεκατισμένη εικόνα κατά τον παράγοντα μεγέθυνσης d , έτσι ώστε οι εικόνες εξόδου να έχουν τις ίδιες διαστάσεις με την εικόνα αναφοράς. Σημειώστε ότι επιλέξαμε το εύρος $[0, 255]$ για τις τιμές των εικόνων και στην περίπτωση έγχρωμων εικόνων, εφαρμόσαμε τις μεθόδους αναπαριστώντας τις εικόνες στον χώρο χρώματος RGB. Η εικόνα αναφοράς μπορεί να θεωρηθεί το ιδανικό αποτέλεσμα παρεμβολής, μιας και δεν περιέχει θόρυβο. Συνεπώς, η διαφορά μεταξύ της εικόνας αναφοράς $r[i_1, i_2]$ και της εξόδου $u[i_1, i_2]$ μίας μεθόδου μπορεί να ιδωθεί σαν σφάλμα ανακατασκευής και είναι ενδεικτική της απόδοσης της μεθόδου.

Χρησιμοποιούμε δύο μετρικά για αυτό το σφάλμα, τον κλασικό λόγο κορυφής σήματος προς θόρυβο (Peak Signal to Noise Ratio - PSNR) ⁴ και τον δείκτη μέσης δομικής ομοιότητας (Mean Structural Similarity - MSSIM) [131]. Ο δείκτης MSSIM, συγκρινόμενος με το PSNR ή διάφορα άλλα μετρικά, φαίνεται να προσεγγίζει καλύτερα την ποιότητα μιας εικόνας με τον τρόπο που την αντιλαμβάνεται οπτικά ένας άνθρωπος. Αυτός ο δείκτης λαμβάνει τιμές από το διάστημα $[0, 1]$ και αυξάνεται με την αύξηση της ποιότητας ⁵. Στην περίπτωση έγχρωμων εικόνων, επεκτείνουμε τον MSSIM με τον απλούστερο τρόπο: υπολογίζουμε τον δείκτη MSSIM κάθε RGB καναλιού χωριστά και μετά βρίσκουμε την μέση τιμή.

Επαναλαμβάνουμε την παραπάνω διαδικασία για διαφορετικές εικόνες αναφοράς από ένα σύνολο εικόνων και για παράγοντες μεγέθυνσης $d=2, 3$ και 4. Για κάθε παράγοντα μεγέθυνσης και μέθοδο παρεμβολής, σχηματίζουμε τα σύνολα των μετρήσεων PSNR και MSSIM για όλες τις εικόνες και ύστερα υπολογίζουμε τις μέσες τιμές αυτών των δύο συνόλων, τις οποίες θεωρούμε ως τελικές μετρήσεις απόδοσης.

Ακολουθήσαμε το προαναφερθέν πειραματικό πλαίσιο χρησιμοποιώντας ένα σύνολο από 23 φωτογραφίες μεγέθους 768×512 εικονοστοιχεία ⁶ (βλ. Σχήματα 3.2 και 3.3). Τρέξαμε δύο σειρές πειραμάτων: (1) για τις ασπρόμαυρες εκδοχές των εικόνων, όπου εφαρμόσαμε τις μεθόδους bicubic, Ολικής Μεταβολής, BG, την προηγούμενη [101] και τρέχουσα [102] εκδοχή της μεθόδου μας ⁷ και (2) για τις έγχρωμες εκδοχές των εικόνων

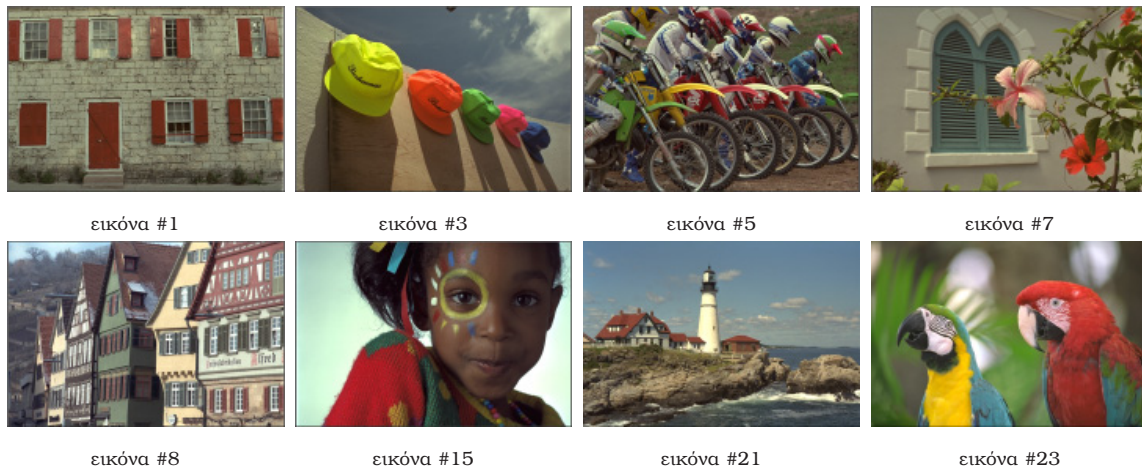
⁴Κάνουμε χρήση του ορισμού $PSNR = 10 \log_{10} (255^2 M / \text{var} \{ \|u[i_1, i_2] - r[i_1, i_2]\| \})$, όπου το $\|\cdot\|$ συμβολίζει εδώ την ευκλείδεια νόρμα διανυσμάτων με M συνιστώσες.

⁵Υπολογίζουμε τον δείκτη MSSIM βασισμένοι στον κώδικα που είναι διαθέσιμος στην διεύθυνση: <http://www.cns.nyu.edu/~lcv/ssim/> και χρησιμοποιώντας τις προεπιλεγμένες παραμέτρους του.

⁶Πήραμε αυτό το σύνολο εικόνων από τη δημόσια διαθέσιμη βάση: <http://www.cipr.rpi.edu/resource/stills/kodak.html>.

⁷Υπενθυμίζουμε ότι η τρέχουσα εκδοχή [102] της μεθόδου μας έχει δύο ελαφρές διαφορές από την προηγούμενη εκδοχή [101]: (α) χρησιμοποιεί έναν γκαουσιανό πυρήνα εξομάλυνσης αντί της περικομμένης

του συνόλου, όπου εφαρμόσαμε τις μεθόδους bicubic, TD και ξανά τις δύο εκδοχές [101] και [102] της προτεινόμενης μεθόδου. Στην δεύτερη σειρά πειραμάτων δεν εφαρμόσαμε τις μεθόδους OM και BG διότι είναι σχεδιασμένες για την παρεμβολή εικόνων γκριζας κλίμακας μόνο.



Σχήμα 3.2: Έγχρωμες εκδοχές 8 από τις 23 εικόνες αναφοράς από το χρησιμοποιούμενο σύνολο εικόνων (768×512 εικονοστοιχεία ανά εικόνα).



Σχήμα 3.3: Λεπτομέρειες των εικόνων αναφοράς που χρησιμοποιούνται στα παραδείγματα των ακόλουθων σχημάτων. (α) Λεπτομέρεια της εικόνας #7 (ασπρόμαυρη εκδοχή). (β) Λεπτομέρεια της εικόνας #5 (έγχρωμη εκδοχή).

Σημειώστε ότι υλοποιήσαμε τις υπόλοιπες μεθόδους ΜΔΕ παρόμοια με την υλοποίηση της προτεινόμενης μεθόδου, όπως αυτή περιγράφηκε με συντομία στην Ενότητα 3.3. Για τις μεθόδους που χρειάζονταν την επιλογή παραμέτρων, χρησιμοποιήσαμε τιμές σταθερές σε όλες τις εικόνες του συνόλου, τις οποίες προσδιορίσαμε εμπειρικά, βασισμένοι στην ποιότητα των αποτελεσμάτων όπως αυτά προσλαμβάνονται οπτικά από έναν άνθρωπο. Έτσι, επιλέξαμε τις παραμέτρους $\sigma=0.3d$, $\rho=0.4d$, $r=0.35d$ και $K=1$ για αμφότερες τις

γκασουσιανής και (β) η ΜΔΕ του μοντέλου περιέχει έναν όρο διάχυσης του οποίου η έκφραση βασίζεται στον τελεστή div αντί του τελεστή trace.

εκδοχές της μεθόδου μας. Επίσης, στην μέθοδο TD χρησιμοποιήσαμε τις ίδιες τιμές $\rho=0.4d$, $K=1$ και στην μέθοδο BG διαλέξαμε $K=3$ για την σχετική σταθερά κατωφλίου.

Μία πλήρης επίδειξη των πειραματικών αποτελεσμάτων για όλες τις εικόνες του συνόλου μπορεί να βρεθεί στην ιστοσελίδα :

<http://cvsp.cs.ntua.gr/~tassos/PDEinterp/ssvm07res>.

3.4.2 Παραδείγματα από τα Πειραματικά Αποτελέσματα

Το Σχήμα 3.4 περιέχει ένα παράδειγμα των αποτελεσμάτων παρεμβολής ασπρόμαυρων εικόνων. Στο Σχ. 3.4(a), παρουσιάζεται η εικόνα εισόδου, η οποία αποτελεί έναν 4×4 αποδεκατισμό της εικόνας αναφοράς του Σχ. 3.3(a). Για τους σκοπούς της παρουσίας έχει μεγεθυνθεί με την απλή συγκράτηση μηδενικής τάξης (zero order hold -ZOH), της οποίας οι μετρήσεις σφάλματος παρέχονται κάπως από αυτή την εικόνα. Παρατηρούμε ότι η bicubic παρεμβολή θολώνει σημαντικά τις ακμές (π.χ. προσέξτε το σύνορο του λουλουδιού στο Σχ. 3.4(β)). Η παρεμβολή OM εξομαλύνει υπερβολικά τις ομοιογενείς περιοχές (π.χ. το εσωτερικό του λουλουδιού στο Σχ. 3.4(γ)), δημιουργεί μπλοκ ψεύδεργα (π.χ. το λεπτό μάρμο κλαδί στο πάνω δεξί μέρος του Σχ. 3.4(δ)) και ταλαντώσεις στις ανακατασκευασμένες ακμές (π.χ. το παντζούρι πίσω από το λουλούδι στο Σχ. 3.4(ε)). Στο αποτέλεσμα της παρεμβολής BG, τα μπλοκ ψεύδεργα έχουν μειωθεί αλλά υπάρχουν ακόμα σε κάποιες περιοχές (π.χ. στο κλαδί στο πάνω δεξί μέρος του Σχ. 3.4(ε)). Επίσης, το αποτέλεσμα αυτό είναι περισσότερο θολό από τα αποτελέσματα της μεθόδου OM. Τέλος, όπως φαίνεται στο Σχ. 3.4(ζ), η προτεινόμενη μέθοδος καταλήγει στην πιο αποτελεσματική ανακατασκευή των δομών της εικόνας, με τα λιγότερο έντονα ψεύδεργα μπλοκ και θολώματος. Για παράδειγμα, παρατηρούμε ότι το παντζούρι μεγεθύνεται με τον πιο ικανοποιητικό τρόπο από την προτεινόμενη μέθοδο. Επιπρόσθετα, από ποσοτικής άποψης, η μέθοδος αυτή αποδίδει τις καλύτερες μετρήσεις PSNR και MSSIM.

Ας σημειωθεί ότι εάν στο παράδειγμα του Σχ. 3.4 είχαμε παρουσιάσει και την παλαιότερη εκδοχή [101] της μεθόδου μας, τότε το αποτέλεσμα θα φαινόταν σχεδόν πανομοιότυπο με αυτό της τρέχουσας εκδοχής, μιας και οι διαφορές τους είναι μικρές. Όμως, στο Σχ. 3.5 δείχνουμε μία ακόμα μικρότερη λεπτομέρεια από τα αποτελέσματα και έτσι γίνονται ορατές οι διαφορές των δύο αυτών εκδοχών. Παρατηρούμε λοιπόν ότι, στην τρέχουσα εκδοχή της μεθόδου, οι ανακατασκευασμένες ακμές είναι ακόμα λιγότερο θολωμένες (π.χ. συγκρίνετε τις μεγεθύνσεις των παράλληλων ευθύγραμμων τμημάτων του παντζουριού). Αυτή η βελτίωση οφείλεται τόσο στον πιο ομαλό μηδενισμό του πυρήνα εξομάλυνσης όσο και στο γεγονός ότι ο τελεστής d_{in} κάνει τον όρο διάχυσης να προσαρμόζεται πιο αποτελεσματικά στις δομές της εικόνας από ό,τι ο τελεστής trace.

Το Σχ. 3.6 παρουσιάζει ένα παράδειγμα από τα αποτελέσματα παρεμβολής έγχρωμων εικόνων. Παρατηρούμε ότι η παρεμβολή bicubic δίνει και εδώ ένα αποτέλεσμα με σημαντικά ψεύδεργα θολώματος και κλιμάκωσης (staircase). Η παρεμβολή TD, Σχ. 3.6(γ), επιφέρει στο αποτέλεσμα μία υπερβολικά συνθετική εμφάνιση, αφού έχει διαταράξει τις ακμές της εικόνας και έχει δημιουργήσει ψεύτικες λεπτές ακμές κοντά στις πραγματικές. Το αποτέλεσμα της προτεινόμενης μεθόδου, Σχ. 3.6(δ), είναι ξανά το πιο ικανοποιητικό αισθητικά και αυτό με τις υψηλότερες μετρήσεις απόδοσης. Περιέχει λιγότερο θολωμένες ακμές από την παρεμβολή bicubic (π.χ. προσέξτε την πιο αποτελεσματική ανακατασκευή των ακμών του κράνους) και φαίνεται πολύ πιο φυσικό από το αποτέλεσμα της παρεμβολής TD.



(α) Είσοδος (μεγεθυμένη με ZOH)
PSNR=25.58, MSSIM=0.758



(β) Παρεμβολή bicubic
PSNR=26.95, MSSIM=0.815



(γ) Ολική Μεταβολή (OM) [56, 76], πυρήνας sinc
PSNR=27.92, MSSIM=0.846



(δ) Ολική Μεταβολή (OM) [56, 76], πυρήνας μέσης τιμής
PSNR=27.27, MSSIM=0.831



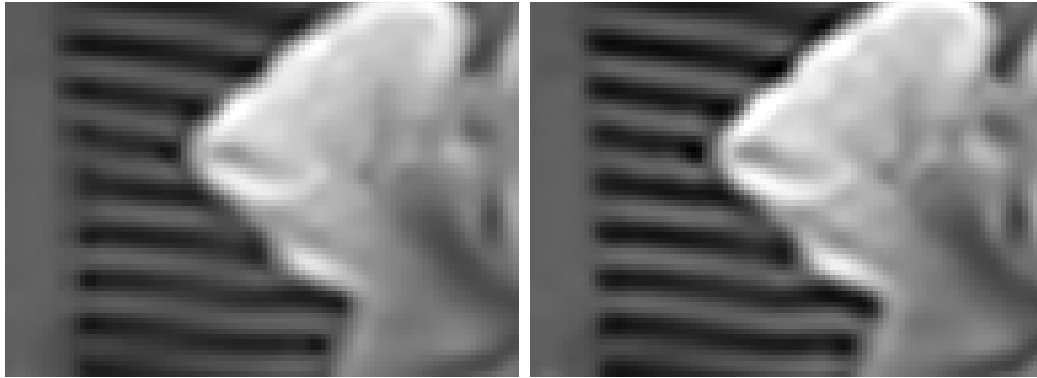
(ε) Παρεμβολή Belahmidi-Guichard (BG) [14]
PSNR=26.89, MSSIM=0.818



(ζ) Προτεινόμενη μέθοδος
PSNR=28.54, MSSIM=0.868

Σχήμα 3.4: Οπτικές λεπτομέρειες και μετρήσεις σφάλματος για τα αποτελέσματα 4×4 παρεμβολής χρησιμοποιώντας την εικόνα #7 σαν αναφορά (βλ. Σχ. 3.3(α)).

Ας δώσουμε τώρα περισσότερες λεπτομέρειες σχετικά με την εφαρμογή της προτεινόμενης μεθόδου στο παραπάνω πείραμα, οι οποίες θα εξηγήσουν καλύτερα πώς η μέθοδος αυτή παράγει ένα επιθυμητό αποτέλεσμα. Το Σχ. 3.7 παρουσιάζει την εξέλιξη της ροής της προτεινόμενης ΜΔΕ (3.15). Στην παρεμβολή μηδενικού γεμίσματος συχνότητας (Σχ. 3.7(β)), που χρησιμοποιείται σαν αρχικοποίηση, οι ανακατασκευασμένες ακμές είναι αρκετά καθαρές και έχουν αληθοφανή σχήματα. Αλλά εμφανίζονται επίσης ισχυρές τα-



(α) Προτεινόμενη μέθοδος, προηγούμενη εκδοχή [101]
PSNR=28.08, MSSIM=0.852

(β) Προτεινόμενη μέθοδος, τρέχουσα εκδοχή [102]
PSNR=28.54, MSSIM=0.868

Σχήμα 3.5: Σύγκριση των δύο εκδοχών της προτεινόμενης μεθόδου παρεμβολής, για το παράδειγμα του Σχ. 3.4 (μικρές λεπτομέρειες 95×70 εικονοστοιχείων).

λαντώσεις λόγω του ότι όλες οι υψηλές συχνότητες της εικόνας έχουν λάβει μηδενική τιμή. Στο ενδιάμεσο αποτέλεσμα του Σχ. 3.7(γ), οι ταλαντώσεις αυτές έχουν απομακρυνθεί από κάποια μέρη της εικόνας, αλλά είναι ακόμα παρούσες κοντά στις ισχυρές ακμές. Στο τε-



(α) Είσοδος (μεγεθυμένη με ZOH)
PSNR=20.87, MSSIM=0.523

(β) Παρεμβολή bicubic
PSNR=21.85, MSSIM=0.579



(γ) Παρεμβολή Tschumperlé-Deriche (TD) [122, 124]
PSNR=19.89, MSSIM=0.458



(δ) Προτεινόμενη μέθοδος
PSNR=22.63, MSSIM=0.652

Σχήμα 3.6: Οπτικές λεπτομέρειες και μετρήσεις σφάλματος για τα αποτελέσματα έγχρωμης παρεμβολής 4×4 χρησιμοποιώντας την εικόνα #5 σαν αναφορά (βλ. Σχ. 3.3(β)).

λικό αποτέλεσμα του Σχ. 3.7(δ), φαίνεται να έχουν πλήρως απομακρυνθεί από την εικόνα. Αυτό γίνεται χάρη στον όρο ανισοτροπικής διάχυσης της ΜΔΕ. Παρατηρούμε επίσης ότι οι ανακατασκευασμένες ακμές της παρεμβολής μηδενικού γεμίσματος συχνοτήτων συντηρούνται καθ' όλη τη διάρκεια της εξέλιξης της ΜΔΕ, χωρίς κανένα ανεπιθύμητο θόλωμα ή διαταραχή (π.χ. προσέξτε την ομοιότητα ανάμεσα στα περιγράμματα του κράνους των Σχ. 3.7(β),(γ) και (δ)). Αυτό επιτυγχάνεται χάρη στην ευέλικτη προσαρμογή του τανυστή διάχυσης στις δομές της εικόνας, καθώς επίσης και στον τελεστή προβολής της ΜΔΕ, ο οποίος αυξάνει την ακρίβεια του αποτελέσματος.

Στο Σχ. 3.8, μπορούμε να δούμε την μεταβολή 3 μετρήσεων κατά τη διάρκεια της εξέλιξης της προτεινόμενης ΜΔΕ, όταν αυτή εφαρμόζεται στο προαναφερθέν παράδειγμα. Στο Σχ. 3.8(α) έχει σχεδιαστεί η RMS τιμή του ρυθμού μεταβολής $\frac{\partial u}{\partial t}(x, y, n\delta t)$ της λύσης της ΜΔΕ (η οποία υπολογίζεται από την διακριτοποίηση του δεξιού μέλους της ΜΔΕ). Παρατηρούμε ότι η ποσότητα αυτή ελαττώνεται κατά τη διάρκεια της εξέλιξης και τελικά τείνει στο μηδέν, πράγμα το υποδεικνύει ότι η ροή ισορροπεί όταν $t \rightarrow \infty$. Πρόκειται για μία επιθυμητή ιδιότητα, μιας και παρέχει μία ένδειξη σχετικά με την ευστάθεια της προτεινόμενης ΜΔΕ και της αριθμητικής της υλοποίησης. Τα Σχ. 3.8(β) και (γ) φανερώνουν ότι οι μετρήσεις PSNR και MSSIM είναι και οι δύο αύξουσες ως προς τον χρόνο διάχυσης. Αυτή η ιδιότητα είναι επίσης επιθυμητή, διότι σημαίνει ότι η εικόνα με την καλύτερη ποιότητα παρέχεται από την κατάσταση ισορροπίας, η οποία αποτελεί το τελικό αποτέλεσμα της μεθόδου μας. Η βελτίωση στην ποιότητα της εικόνας μπορεί να αποδοθεί στο γεγονός ότι η ροή της ΜΔΕ σταδιακά απομακρύνει τις ταλαντώσεις χωρίς να καταστρέφει τις δομές της εικόνας, οπότε φέρνει την εικόνα όλο και πιο κοντά στην εικόνα αναφοράς.

3.4.3 Συγκεντρωτικές Μετρήσεις Απόδοσης

Το Σχ. 3.9 περιέχει τις συγκεντρωτικές μετρήσεις απόδοσης των μεθόδων παρεμβολής, για τις δύο σειρές πειραμάτων στο σύνολο των 23 εικόνων. Παρατηρούμε στο Σχ. 3.9(α) ότι οι χειρότερες μετρήσεις στα πειράματα με τις γκριζες εικόνες αντιστοιχούν στις μεθόδους bicubic και BG. Η παρεμβολή bicubic επιφέρει καλύτερες μετρήσεις για τον παράγοντα μεγέθυνσης $d=2$, αλλά αυτή η σχέση αντιστρέφεται όταν $d=4$. Αυτό οφείλεται στο γεγονός ότι η μέθοδος bicubic εισάγει ψεύδεργα θολώματος, τα οποία γίνονται πιο έντονα καθώς ο d αυξάνεται. Από την άλλη μεριά, η μέθοδος BG εξομαλύνει υπερβολικά τις ομοιογενείς περιοχές της εικόνας ακόμα κι όταν ο d είναι μικρός, αλλά οι αντίστοιχες μετρήσεις σφάλματος ελαττώνονται με πιο αργό ρυθμό, μιας και εφαρμόζεται μία προσαρμοστική επεξεργασία. Επίσης, η μέθοδος BG επιφέρει χειρότερες μετρήσεις σφάλματος από τις άλλες μεθόδους ΜΔΕ διότι η ροή διάχυσης αυτής της μεθόδου δεν περιορίζεται στο να ικανοποιεί ακριβώς την συνθήκη αντιστρεψιμότητας, επομένως εισάγει ψεύδεργα υπερβολικής εξομάλυνσης και ελαττώνει την ακρίβεια του αποτελέσματος. Επιπρόσθετα, η μέθοδος παρεμβολής βασισμένης στην OM επιτυγχάνει σε όλες τις περιπτώσεις καλύτερες μετρήσεις σφάλματος από τις μεθόδους bicubic και BG. Οι μετρήσεις αυτές είναι αρκετά παρόμοιες για τις δύο επιλογές του πυρήνα εξομάλυνσης (sinc και μέσης τιμής). Παρατηρούμε τέλος ότι και οι δύο εκδοχές της προτεινόμενης μεθόδου επιφέρουν καλύτερες μέσες μετρήσεις PSNR και MSSIM από όλες τις άλλες μεθόδους. Αυτή η βελτίωση της απόδοσης μπορεί να αποδοθεί στο γεγονός ότι η μεθόδός μας εφαρμόζει μία πιο ευέλικτη προσαρμοστική εξομάλυνση και εκμεταλλεύεται αξιόπιστα την πληροφορία της εικόνας εισόδου ώστε να αυξήσει την ακρίβεια του αποτελέσματος. Σημειώστε επίσης ότι η τρέ-

χουσα εκδοχή της μεθόδου έχει βελτιώσει περαιτέρω, σχεδόν σε όλες τις περιπτώσεις, τις μέσες τιμές των μετρήσεων PSNR και MSSIM. Αυτή η περαιτέρω βελτίωση οφείλεται στις δύο τροποποιήσεις που έχουν γίνει στην τρέχουσα μέθοδο.

Το Σχ. 3.9(β) επιδεικνύει ότι στα πειράματα με τις έγχρωμες εικόνες, η προτεινόμενη μέθοδος επιφέρει ξανά τις καλύτερες συγκεντρωτικές μετρήσεις PSNR και MSSIM. Επιπρόσθετα, βλέπουμε ότι οι μετρήσεις για τη μέθοδο TD είναι πολύ χειρότερες ακόμα και από την μέθοδο bicubic. Αυτό οφείλεται στην υπερβολικά συνθετική όψη των αποτελεσμάτων της μεθόδου αυτής.

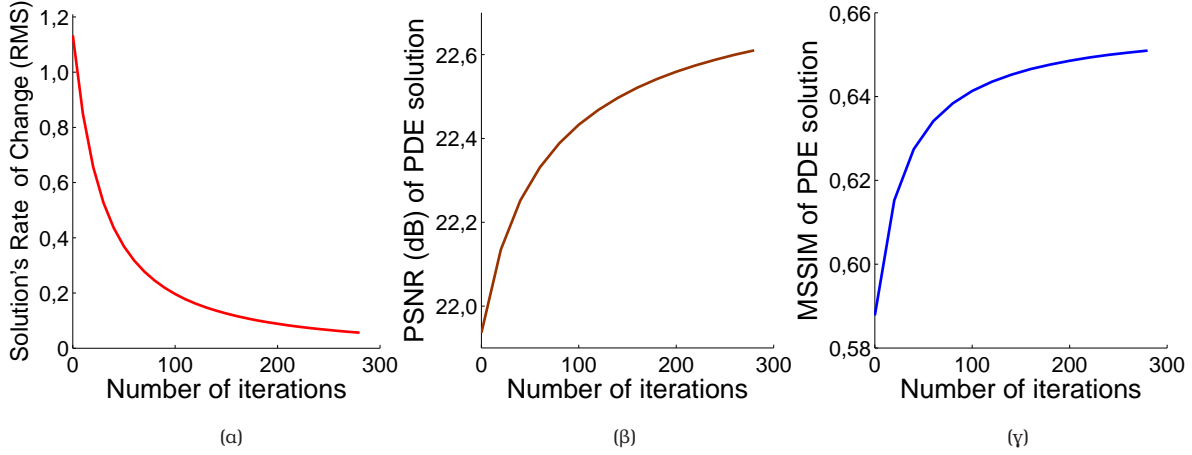
3.4.4 Παρεμβολή Βιοϊατρικών Εικόνων

Εφαρμόσαμε επίσης την προτεινόμενη μέθοδο παρεμβολής σε μία ειδική περίπτωση βιοϊατρικών δεδομένων τα οποία είναι εικόνες μαγνητικής τομογραφίας (MRI) της φωνητικής οδού ομιλητή. Όπως έχει παρουσιαστεί στην Ενότητα 1.2.1, δεδομένα εικόνων τέτοιου τύπου είναι σημαντικά για την ανάλυση και μοντελοποίηση του ανθρώπινου συστήματος παραγωγής φωνής. Συχνά οι εικόνες αυτές έχουν χαμηλή χωρική ευκρίνεια, εξαιτίας περιορισμών στα συστήματα καταγραφής εικόνων. Επομένως, η εφαρμογή μιας αποτελεσματικής μεθόδου παρεμβολής σαν ένα βήμα προεπεξεργασίας θα μπορούσε να είναι κρίσιμη για την επιτυχία των ακόλουθων βημάτων επεξεργασίας.

Το Σχ. 3.10 παρουσιάζει ένα παράδειγμα χρησιμοποιώντας μία μεσο-οβελιαία τομή



Σχήμα 3.7: Λεπτομέρειες από την εξέλιξη της προτεινόμενης ΜΔΕ (3.15), όταν αυτή εφαρμόζεται στην 4×4 παρεμβολή χρησιμοποιώντας την έγχρωμη εικόνα #5 σαν αναφορά (χρονικό βήμα $\delta t = 0.2$).



Σχήμα 3.8: Μετρήσεις κατά τη διάρκεια της εξέλιξης της προτεινόμενης ΜΔΕ (3.15), για το παράδειγμα των Εικόνων 3.6 και 3.7. **(α)** RMS τιμή του ρυθμού μεταβολής της λύσης της ΜΔΕ $\partial \mathbf{u}(x, y, n\delta t)/\partial t$. **(β)** PSNR και **(γ)** MSSIM μετρικά σφάλματος.

της απεικόνιση MRI της φωνητικής οδού⁸. Το Σχ. 3.10(a) είναι μία λεπτομέρεια μίας αποθορυβοποιημένης εκδοχής (εφαρμόζοντας ανισοτροπική διάχυση) της εικόνας αυτής. Παρόμοια με τα παραπάνω πειράματα, χρησιμοποιήσαμε την εικόνα αυτή σαν αναφορά και την σμικρύναμε στο $\frac{1}{4} \times \frac{1}{4}$ του μεγέθους της (βλ. Σχ. 3.10(β)). Εφαρμόσαμε τέλος τις μεθόδους παρεμβολής bicubic, OM, BG και την προτεινόμενη για να μεγεθύνουμε 4×4 φορές την αποδεκατισμένη εικόνα (Σχήματα 3.10 (γ)-(ζ)). Παρατηρώντας τα αποτελέσματα αυτά, οδηγούμαστε σε συμπεράσματα αντίστοιχα με τα συμπεράσματα από τα αποτελέσματα του Σχ. 3.4. Η προτεινόμενη μέθοδος φαίνεται να παράγει ξανά το πιο ικανοποιητικό αποτέλεσμα, με την καλύτερη ανακατασκευή της φωνητικής οδού, ενώ παράλληλα επιτυγχάνει τις καλύτερες μετρήσεις PSNR και MSSIM. Αυτό το αποτέλεσμα αναδεικνύει ότι η προτεινόμενη μέθοδος μπορεί επίσης να εφαρμοστεί για την αξιόπιστη βελτίωση της ευκρίνειας βιοϊατρικών εικόνων της φωνητικής οδού.

3.5 Θεωρητικές Λεπτομέρειες

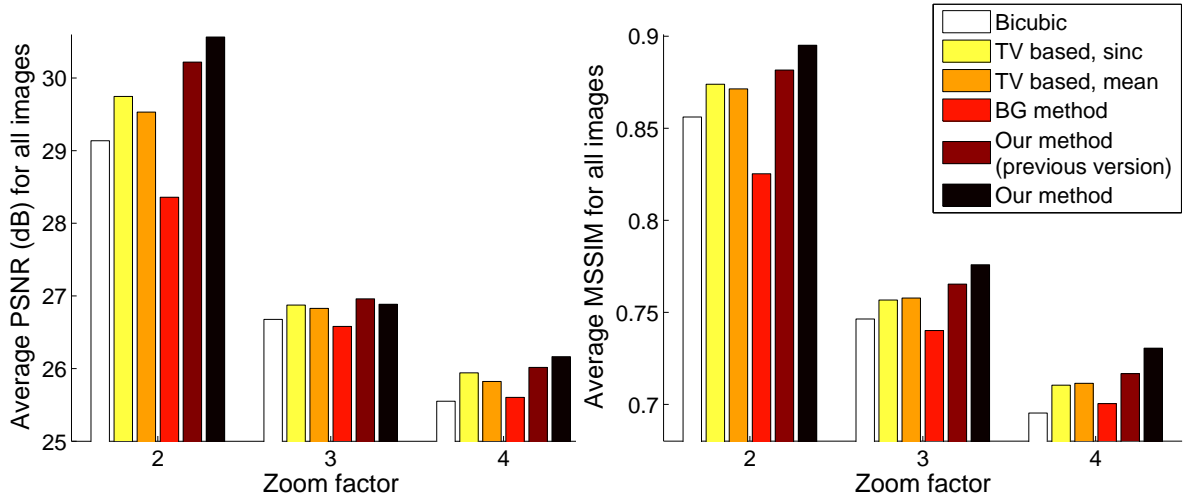
Σε αυτήν την ενότητα, δείχνουμε την εγκυρότητα των εκφράσεων (3.16), (3.17) και (3.18), οι οποίες περιγράφουν την προβολή $P_{\mathcal{U}_{0,S}}\{\cdot\}$ στον υποχώρο $\mathcal{U}_{0,S}$. Υποθέτουμε ότι στην προτεινόμενη ΜΔΕ, όλες οι συναρτήσεις εισόδου της προβολής $P_{\mathcal{U}_{0,S}}\{\cdot\}$ ανήκουν σε ένα χώρο Hilbert.

Ο υποχώρος $\mathcal{U}_{0,S}$ μπορεί να οριστεί σαν το σύνολο των συναρτήσεων $v(\mathbf{x})$ που ικανοποιούν την Εξ. (3.1) με $z[i_1, i_2] = 0$ παντού, ή ισοδύναμα (3.3) με $\hat{z}_{n_1, n_2} = 0$ παντού, δηλαδή:

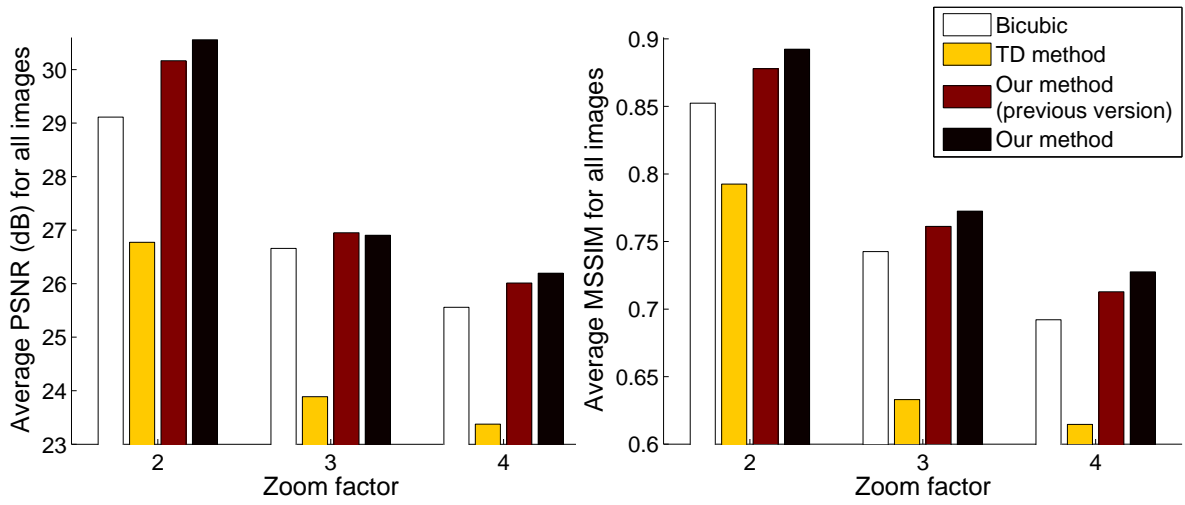
$$\sum_{(k_1, k_2) \in \mathbb{Z}^2} \hat{S} \left(\frac{2\pi}{\tilde{N}_x} (n_1 + k_1 \tilde{N}_x), \frac{2\pi}{\tilde{N}_y} (n_2 + k_2 \tilde{N}_y) \right) \cdot \hat{v}_{n_1 + k_1 \tilde{N}_x, n_2 + k_2 \tilde{N}_y} = 0, \quad (3.20)$$

για όλα τα $(n_1, n_2) \in \mathbb{Z}^2$. Όπως έχει προαναφερθεί, είναι αρκετό για το \hat{v}_{m_1, m_2} να ικανοποιεί την (3.20) μόνο για τα (n_1, n_2) που ανήκουν σε οποιοδήποτε $\tilde{N}_x \times \tilde{N}_y$ ορθογωνικό

⁸Ευχαριστούμε τον δρ. Olov Engwall για την παροχή της εικόνας αυτής.



(α) Πειράματα με εικόνες γκριζας κλίμακας



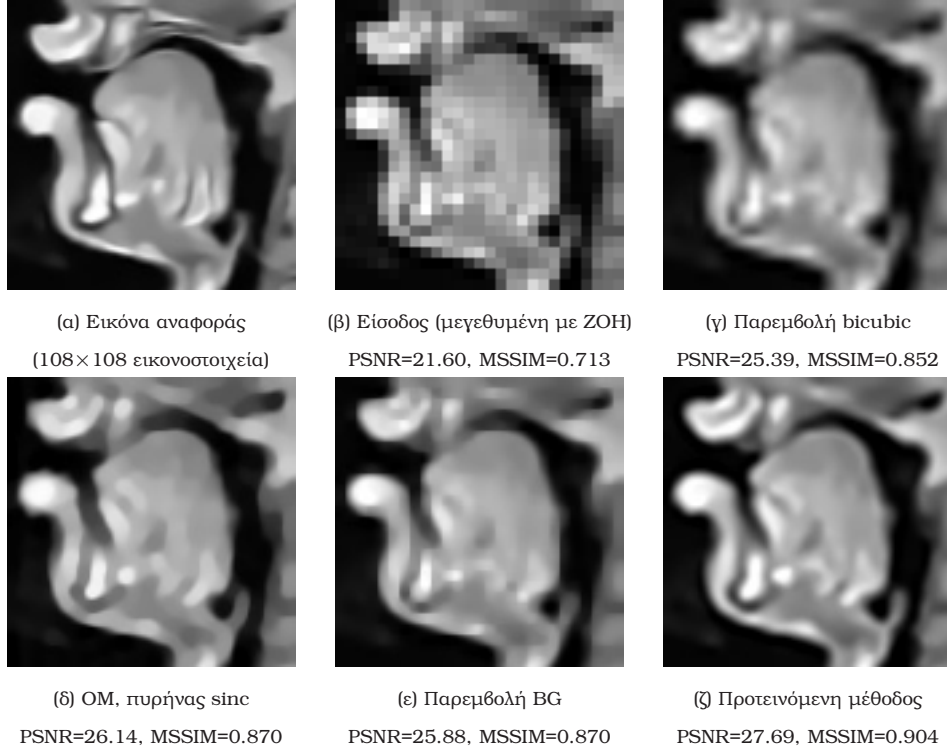
(β) Πειράματα με έγχρωμες εικόνες

Σχήμα 3.9: Μέσες τιμές μετρήσεων σφάλματος σε όλα τα αποτελέσματα χρησιμοποιώντας τις 23 εικόνες, για διαφορετικούς παράγοντες μεγέθυνσης. Συγκρινόμενες μέθοδοι: Δικυβική παρεμβολή (*Bicubic*). Παρεμβολή βασισμένη στην Ολική Μεταβολή [56, 76] με πυρήνα sinc και πυρήνα μέσης τιμής (*TV based, sinc* και *TV based, mean* αντίστοιχα). Παρεμβολή Belahmidi-Guichard [14] (*BG method*). Παρεμβολή Tschumperlé-Deriche [122, 124] (*TD method*). Προτεινόμενη μέθοδος, παλαιότερη εκδοχή [101] (*Our method - previous version*). Προτεινόμενη μέθοδος [102] (*Our method*).

πλέγμα. Επιλέγουμε το πλέγμα $B_2 = \{-N_x, \dots, N_x - 1\} \times \{-N_y, \dots, N_y - 1\}$ που περιβάλλει την αρχή των αξόνων.

Για κάθε $(n_1, n_2) \in B_2$, έστω $p^{(n_1, n_2)}(\mathbf{x})$ η $(\tilde{N}_x, \tilde{N}_y)$ -περιοδική μιγαδική συνάρτηση, της οποίας οι συντελεστές Fourier είναι (βλ. Σχ. 3.11):

$$\hat{p}_{m_1, m_2}^{(n_1, n_2)} = \begin{cases} \hat{S} \left(\frac{2\pi m_1}{\tilde{N}_x}, \frac{2\pi m_2}{\tilde{N}_y} \right) & , \text{αν } \exists (k_1, k_2) \in \mathbb{Z}^2 \text{ τ.ω.:} \\ & (m_1, m_2) = (n_1 + k_1 \tilde{N}_x, n_2 + k_2 \tilde{N}_y) \\ 0 & , \text{διαφορετικά} \end{cases} \quad (3.21)$$



Σχήμα 3.10: Παρεμβολή (4×4) μιας βιοϊατρικής εικόνας της φωνητικής οδού.

για όλα τα $(m_1, m_2) \in \mathbb{Z}^2$. Τότε, η συνθήκη (3.20) είναι ισοδύναμη με:

$$\forall (n_1, n_2) \in B_2 : \langle \hat{v}, \hat{p}^{(n_1, n_2)} \rangle_{\ell^2} = 0 \Leftrightarrow \langle v, p^{(n_1, n_2)} \rangle_{L^2} = 0, \quad (3.22)$$

όπου τα $\langle \cdot, \cdot \rangle_{\ell^2}$ και $\langle \cdot, \cdot \rangle_{L^2}$ συμβολίζουν τα εσωτερικά γινόμενα των $\ell^2(\mathbb{Z}^2)$ και $L^2(\tilde{\Omega})$ αντίστοιχα⁹. Η ισοδυναμία στην Εξ. (3.22) οφείλεται στην ισότητα Parseval [90]:

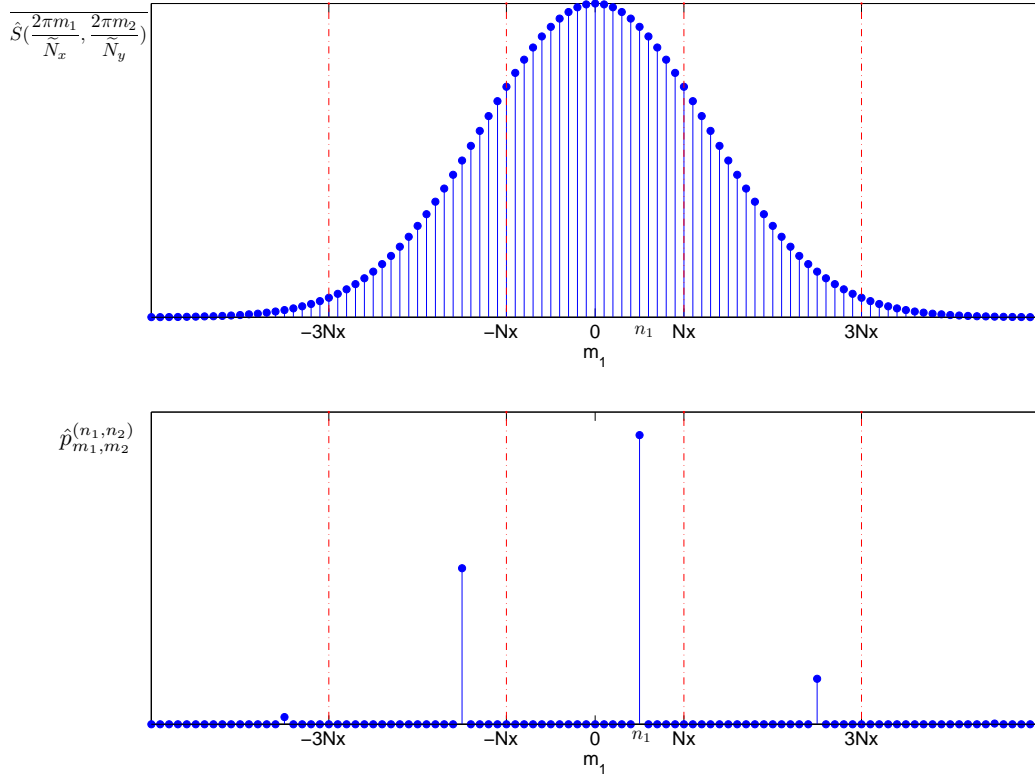
$$\langle \hat{v}, \hat{p}^{(n_1, n_2)} \rangle_{\ell^2} = (\tilde{N}_x \tilde{N}_y)^{-1} \langle v, p^{(n_1, n_2)} \rangle_{L^2}. \quad (3.23)$$

Από τη στιγμή που η Εξ. (3.20), η οποία ορίζει τον υποχώρο $\mathcal{U}_{0,S}$, είναι ισοδύναμη με την Εξ. (3.22), ο $\mathcal{U}_{0,S}$ είναι στην πραγματικότητα το σύνολο των συναρτήσεων που είναι κάθετα σε όλα τα $p^{(n_1, n_2)}$. Επομένως, μπορεί ναδειχθεί ότι το σύνολο $\{p^{(n_1, n_2)} : (n_1, n_2) \in B_2\}$ αποτελεί μία βάση του $\mathcal{U}_{0,S}^\perp$, δηλαδή του ορθογωνικού συμπληρώματος του $\mathcal{U}_{0,S}$. Αυτή η βάση είναι η ορθογώνια διότι:

$$\begin{aligned} \langle p^{(n_1, n_2)}, p^{(n'_1, n'_2)} \rangle_{L^2} &= (\tilde{N}_x \tilde{N}_y) \langle \hat{p}^{(n_1, n_2)}, \hat{p}^{(n'_1, n'_2)} \rangle_{\ell^2} \\ &= \|p^{(n_1, n_2)}\|_{L^2}^2 \delta_{n_1 - n'_1, n_2 - n'_2}, \quad \forall (n_1, n_2) \in B_2, \forall (n'_1, n'_2) \in B_2, \end{aligned} \quad (3.24)$$

μιας και, εάν $(n'_1, n'_2) \neq (n_1, n_2)$, το $\hat{p}^{(n'_1, n'_2)}$ λαμβάνει μη μηδενικές τιμές σε διαφορετικά (m_1, m_2) από το $p^{(n_1, n_2)}$ (βλ. Εξ. (3.21) και Σχ. 3.11). Η νόρμα $\|p^{(n_1, n_2)}\|_{L^2}$ στην Εξ. (3.24)

⁹ $\langle f, g \rangle_{\ell^2} = \sum_{(m_1, m_2) \in \mathbb{Z}^2} f_{m_1, m_2} \overline{g_{m_1, m_2}}$, $\langle f, g \rangle_{L^2} = \iint_{\tilde{\Omega}} f(\mathbf{x}) \overline{g(\mathbf{x})} d\mathbf{x}$.



Σχήμα 3.11: Εξαγωγή του $\hat{p}_{m_1, m_2}^{(n_1, n_2)}$ (για χάρη της παρουσίασης, το m_2 είναι σταθερό στην τιμή $n_2 + k_2 \tilde{N}_y$, για κάποιο $k_2 \in \mathbb{Z}$).

δίνεται από:

$$\begin{aligned} \|p^{(n_1, n_2)}\|_{L^2} &= (\tilde{N}_x \tilde{N}_y)^{\frac{1}{2}} \cdot \|\hat{p}^{(n_1, n_2)}\|_{\ell^2} \\ &= \left\{ (\tilde{N}_x \tilde{N}_y)^{-\frac{1}{2}} \cdot \sum_{(k_1, k_2) \in \mathbb{Z}^2} \left| \hat{S} \left(\frac{2\pi}{\tilde{N}_x} (n_1 + k_1 \tilde{N}_x), \frac{2\pi}{\tilde{N}_y} (n_2 + k_2 \tilde{N}_y) \right) \right|^2 \right\}^{-\frac{1}{2}}. \end{aligned} \quad (3.25)$$

Σχηματίζουμε το σύνολο $\{q^{(n_1, n_2)} : (n_1, n_2) \in B_2\}$, το οποίο αποτελεί μία ορθογωνική βάση του $\mathcal{U}_{0, S}^\perp$, κανονικοποιώντας κάθε ένα από τα $p^{(n_1, n_2)}$: Για κάθε $(n_1, n_2) \in B_2$, το $q^{(n_1, n_2)}$ ορίζεται από τη σχέση:

$$q^{(n_1, n_2)}(\mathbf{x}) = \frac{1}{\|p^{(n_1, n_2)}\|_{L^2}} \cdot p^{(n_1, n_2)}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^2. \quad (3.26)$$

Μπορούμε τώρα εύκολα να υπολογίσουμε την προβολή στον $\mathcal{U}_{0, S}^\perp$ [90]:

$$w(\mathbf{x}) \triangleq P_{\mathcal{U}_{0, S}^\perp} \{v\} = \sum_{(n_1, n_2) \in B_2} \langle v, q^{(n_1, n_2)} \rangle_{L^2} \cdot q^{(n_1, n_2)}(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^2,$$

η οποία είναι επίσης $(\tilde{N}_x, \tilde{N}_y)$ -περιοδική, όπως όλες οι συναρτήσεις $q^{(n_1, n_2)}(\mathbf{x})$. Άρα, οι συντελεστές Fourier της w δίνονται από:

$$\hat{w}_{m_1, m_2} = (\tilde{N}_x \tilde{N}_y) \sum_{(n_1, n_2) \in B_2} \langle \hat{v}, \hat{q}^{(n_1, n_2)} \rangle_{\ell^2} \cdot \hat{q}_{m_1, m_2}^{(n_1, n_2)}, \quad (3.27)$$

για όλα τα $(m_1, m_2) \in \mathbb{Z}^2$.

Όπως είπαμε, οι συντελεστές Fourier κάθε στοιχείου του $\{p^{(n_1, n_2)} : (n_1, n_2) \in B_2\}$ είναι μη-μηδενικοί σε διαφορετικά σημεία (m_1, m_2) από όλα τα υπόλοιπα στοιχεία. Η ίδια ιδιότητα ισχύει επίσης για την βάση $\{q^{(n_1, n_2)} : (n_1, n_2) \in B_2\}$, εξαιτίας της Εξ. (3.26). Αυτό σημαίνει ότι, για κάθε $(m_1, m_2) \in \mathbb{Z}^2$, μόνο ένας όρος στο άθροισμα της (3.27) είναι μη μηδενικός: αυτός που αντιστοιχεί στο ζευγάρι (3.27) το οποίο μπορεί να γραφθεί ως $(n_1, n_2) = (m_1 - k_1 \tilde{N}_x, m_2 - k_2 \tilde{N}_y)$, για κάποιο $(k_1, k_2) \in \mathbb{Z}^2$. Χρησιμοποιώντας την σχέση αυτή και τις Εξ. (3.21), (3.25), (3.26), μπορούμε να γράψουμε την Εξ. (3.27) ως ακολούθως:

$$\hat{w}_{m_1, m_2} = \left\{ \sum_{(k_1, k_2) \in \mathbb{Z}^2} \hat{\phi} \left(\frac{2\pi m_1}{\tilde{N}_x} + k_1 2\pi, \frac{2\pi m_2}{\tilde{N}_y} + k_2 2\pi \right) \cdot \hat{v}_{m_1 + k_1 \tilde{N}_x, m_2 + k_2 \tilde{N}_y} \right\} \cdot \hat{\phi} \left(\frac{2\pi m_1}{\tilde{N}_x}, \frac{2\pi m_2}{\tilde{N}_y} \right), \forall (m_1, m_2) \in \mathbb{Z}^2,$$

όπου έχει εισαχθεί η συνάρτηση $\phi(\mathbf{x})$ της οποίας ο μετασχηματισμός Fourier είναι:

$$\hat{\phi}(\omega_1, \omega_2) = \left\{ \sum_{(k_1, k_2) \in \mathbb{Z}^2} \left| \hat{S}(\omega_1 + k_1 2\pi, \omega_2 + k_2 2\pi) \right|^2 \right\}^{-\frac{1}{2}} \cdot \overline{\hat{S}(\omega_1, \omega_2)}.$$

Τέλος, η προβολή $P_{\mathcal{U}_{0,S}}\{v\}$ μπορεί να υπολογιστεί κάνοντας χρήση της έκφρασης για την $P_{\mathcal{U}_{0,S}^\perp}\{v\}$: Από το θεώρημα προβολής [90], συνεπάγεται ότι:

$$P_{\mathcal{U}_{0,S}}\{v\} = v - P_{\mathcal{U}_{0,S}^\perp}\{v\} = v - w.$$

3.6 Συμπεράσματα

Σε αυτό το κεφάλαιο, προτείναμε ένα μοντέλο για την παρεμβολή διανυσματικών εικόνων, βασισμένο σε μία ΜΔΕ ανισοτροπικής διάχυσης. Η κύρια συνεισφορά μας έγκειται σε έναν αποτελεσματικό συνδυασμό του πλαισίου της συνθήκης αντιστρεψιμότητας [56] με την μη-γραμμική ανισοτροπική διάχυση [124, 135]. Για τον σκοπό αυτό, εισάγαμε έναν κατάλληλο τελεστή προβολής και καταλήξαμε σε μία έκφραση για αυτόν χρησιμοποιώντας τον μετασχηματισμό Fourier. Το προτεινόμενο μοντέλο περιορίζει τα ανεπιθύμητα ψεύδεργα των κλασικών γραμμικών καθώς και σχετικών μεθόδων παρεμβολής που βασίζονται σε ΜΔΕ. Διεξοδικά πειραματικά αποτελέσματα ανέδειξαν την δυναμική της μεθόδου μας όταν εφαρμόζεται σε γκρίζας κλίμακας και έγχρωμων εικόνων.

Ας σημειωθεί ότι σημαντικά θεωρητικά ζητήματα, όπως ύπαρξη, μοναδικότητα και ευστάθεια της λύσης της προτεινόμενης ΜΔΕ παραμένουν ακόμα ανοικτά. Από την άλλη μεριά όμως, τα εκτενή πειράματα αναδεικνύουν ότι η διακριτοποίηση της ΜΔΕ είναι ευσταθής και στην πράξη καταλήγει σε ισορροπία εντός πεπερασμένων επαναλήψεων. Σημειώστε τέλος ότι η προτεινόμενη μέθοδος υποθέτει ότι η εικόνα εισόδου δεν περιέχει θόρυβο. Θα μπορούσε όμως να τροποποιηθεί για να χειρίζεται εικόνες με μη αμελητέο θόρυβο.

Κεφάλαιο 4

Μέθοδοι Διάχυσης με Τανυστές Εξαγμένες από Μεταβολικά Πλαίσια

4.1 Εισαγωγή

Όπως παρουσιάστηκε στην Ενότητα 2.1, κατά τα τελευταία 20 χρόνια έχουν προταθεί διάφορες μέθοδοι μη-γραμμικής διάχυσης για την Όραση Υπολογιστών, οι οποίες έχουν αποδειχθεί ιδιαίτερα χρήσιμες στην αποθορυβοποίηση εικόνων, στην δημιουργία χώρων κλίμακας καθώς και σε αρκετές άλλες εφαρμογές. Για κάποιες από αυτές τις μεθόδους, προηγούμενες εργασίες έχουν δείξει ότι μπορούν να προκύψουν από την ελαχιστοποίηση ενός συναρτησιακού της εικόνας, όπως για παράδειγμα στην περίπτωση της Ολικής Μεταβολής. Όμως, για αρκετούς άλλους τύπους διάχυσης, όπως η κανονικοποιημένη μη-γραμμική διάχυση [27] ή η μη-γραμμική και ανισοτροπική διάχυση [135], οι αντίστοιχες ΜΔΕ έχουν σχεδιαστεί απευθείας, χωρίς κάποια μεταβολική ερμηνεία.

Συγκρίνοντας λοιπόν τις μεταβολικές μεθόδους με τις υπόλοιπες μεθόδους διάχυσης, μπορούμε να σημειώσουμε ότι η ύπαρξη μιας σύνδεσης ανάμεσα σε μία μέθοδο διάχυσης και μία ελαχιστοποίηση συναρτησιακού προσφέρει διάφορα πλεονεκτήματα. Πρώτα από όλα, προσφέρει καλύτερη διαισθητική ερμηνεία και κάνει την μέθοδο λιγότερο ευρετική. Επίσης, διευκολύνει την χρησιμοποίηση της μεθόδου σε πληθώρα προβλημάτων Επεξεργασίας Εικόνας και Όρασης Υπολογιστών, όπως αποκατάσταση εικόνων, ανακατασκευή χαμένων κομματιών της εικόνας (*inpainting*) και παρεμβολή, τα οποία μπορούν να διατυπωθούν σαν ελαχιστοποίηση κάποιου κατάλληλου συναρτησιακού με περιορισμούς που επιβάλλονται από το πρόβλημα. Επιπρόσθετα, είναι ευκολότερο να εξάγει κανείς μεθόδους με ελαττωμένο αριθμό παραμέτρων: για παράδειγμα η ροή ταχύτερης καθόδου της Ολικής Μεταβολής δεν έχει καθόλου παραμέτρους. Τέλος, η μεταβολική ερμηνεία μιας μεθόδου ανοίγει το δρόμο σε αποδοτικές υλοποιήσεις που βασίζονται σε τεχνικές βελτιστοποίησης.

Έχοντας τα παραπάνω σαν κίνητρο, σε αυτό το κεφάλαιο εισάγουμε ένα νέο συναρτησιακό για διανυσματικές εικόνες που γενικεύει διάφορες υπάρχουσες μεταβολικές μεθόδους διάχυσης, όπως είναι η Ολική Μεταβολή και το Συναρτησιακό Beltrami. Το συναρτησιακό αυτό βασίζεται στον *δομικό τανυστή*, ο οποίος περιγράφει την γεωμετρία των δομών της εικόνας στην γειτονιά του κάθε σημείου. Αρχικά γενικεύουμε το Συναρτησιακό Beltrami βασιζόμενοι στα *τεμάχια της εικόνας* και χρησιμοποιώντας εμβυθίσεις (*embeddings*) σε χώρους υψηλών διαστάσεων. Προχωρώντας στην γενικότερη μορφή του

συναρτησιακού, αποδεικνύουμε ότι η ελαχιστοποίησή του οδηγεί σε μία μη-γραμμική και ανισοτροπική διάχυση η οποία είναι κανονικοποιημένη, υπό την έννοια ότι ο τανυστής διάχυσης περιλαμβάνει συνελίξεις με έναν πυρήνα. Βασισμένοι στο αποτέλεσμα αυτό, προτείναμε δύο νέες μεθόδους διάχυσης, την *Γενικευμένη Ροή Beltrami* και την *Τανυστική Ολική Μεταβολή*. Αυτές οι μέθοδοι συνδυάζουν τα πλεονεκτήματα των μεταβολικών μεθόδων και των μεθόδων που βασίζονται σε τανυστές. Τα πειράματα που πραγματοποιήθηκαν δείχνουν την προοπτική των προτεινόμενων μεθόδων για τα προβλήματα της βελτίωσης της ποιότητας των εικόνων.

4.2 Γενίκευση του Συναρτησιακού Beltrami Βασισμένη στα Τεμάχια Εικόνας

Ο Sochen και οι συνεργάτες του [112] ερμηνεύουν μία διανυσματική εικόνα \mathbf{u} με N κανάλια σαν μια 2D επιφάνεια (\mathbf{x}, \mathbf{u}) εμβυθισμένη (embedded) σε έναν χώρο $N+2$ διαστάσεων. Προτείνουν λοιπόν την εξέλιξη της εικόνας σύμφωνα με την *Ροή Beltrami* (Beltrami flow), η οποία είναι μία ροή ανισοτροπικής διάχυσης προς την ελαχιστοποίηση του εμβαδού της επιφάνειας αυτής. Το *Συναρτησιακό Beltrami* αυτής της επιφάνειας προσφέρει έναν αξιόπιστο και κομψό τρόπο να συζευχθούν τα διαφορετικά κανάλια της εικόνας και να επεκταθούν οι ιδιότητες της Ολικής Μεταβολής στις διανυσματικές εικόνες. Το πεδίο των τανυστών που ελέγχει την Ροή Beltrami συνδυάζει την πληροφορία από όλα τα κανάλια αλλά ανεξάρτητα σε κάθε σημείο της εικόνας, χωρίς να λαμβάνει υπόψιν του την μεταβολή της εικόνας στην γειτονιά του κάθε σημείου. Αυτή η έλλειψη χωρικής επέκτασης έχει αρνητικά αποτελέσματα στην σθεναρότητα ενάντια στον θόρυβο και στην βελτίωση των ακμών της εικόνας.

Με σκοπό να ξεπεραστούν αυτοί οι περιορισμοί, γενικεύουμε το συναρτησιακό Beltrami χρησιμοποιώντας απεικονίσεις μεγαλύτερων διαστάσεων της μορφής $(\mathbf{x}, \mathcal{P}^u(\mathbf{x}))$ όπου $\mathcal{P}^u(\mathbf{x})$ είναι ένα *τεμάχιο της εικόνας* (image patch) [123] το οποίο περιέχει σταθμισμένες τιμές της εικόνας όχι μόνο στο σημείο \mathbf{x} αλλά και σε σημεία σε ένα παράθυρο που το περιβάλλει. Με αυτόν τον τρόπο, κάθε σημείο συνεισφέρει στο εμβαδόν της εμβυθισμένης επιφάνειας λαμβάνοντας υπόψιν την μεταβολή της εικόνας στην γειτονιά του. Όπως θα δειχθεί στην Ενότητα 4.6.1, εάν αφήσουμε τα τεμάχια να δειγματοληπτήσουν την εικόνα με βήμα που τείνει στο 0, το εμβαδόν της εμβυθισμένης επιφάνειας τείνει στο ακόλουθο συναρτησιακό:

$$A[\mathbf{u}] = \int_{\Omega} \sqrt{(\alpha^2 + \lambda_1)(\alpha^2 + \lambda_2)} d\mathbf{x} \quad (4.1)$$

όπου α είναι μία σταθερά, $\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}) \cdots u_N(\mathbf{x})] : \Omega \rightarrow \mathbb{R}^N$ είναι η διανυσματική εικόνα και $\lambda_i = \lambda_i(J_K(\nabla \mathbf{u}))$, $\lambda_1 \geq \lambda_2$, είναι οι ιδιοτιμές του δομικού τανυστή:

$$J_K(\nabla \mathbf{u}) = K * \sum_{i=1}^N \nabla u_i \otimes \nabla u_i, \quad (4.2)$$

όπου K είναι ένας πυρήνας συνέλιξης (π.χ. μία 2D γκαουσιανή) και το \otimes συμβολίζει τανυστικό γινόμενο. Αυτός ο τανυστής μετράει την γεωμετρία των δομών της εικόνας στην γειτονιά του κάθε σημείου [135]. Ας σημειωθεί ότι το κλασικό συναρτησιακό Beltrami αντιστοιχεί στην ειδική περίπτωση που ο πυρήνας K είναι η 2D συνάρτηση Dirac

δέλτα $\delta(\mathbf{x})$. Το γενικευμένο συναρτησιακό Beltrami είναι μία ειδική περίπτωση του πιο γενικού συναρτησιακού που θα προταθεί στην επόμενη ενότητα, όπου παρέχουμε ένα αποτέλεσμα σχετικά με την αντίστοιχη συναρτησιακή ελαχιστοποίηση. Στην Ενότητα 4.4.2 εφαρμόζουμε το αποτέλεσμα αυτό για να καταλήξουμε, εκτός των άλλων, στην μέθοδο της *Γενικευμένης Ροής Beltrami*.

4.3 Γενικευμένο Συναρτησιακό Βασισμένο στον Δομικό Τανυστή

Έχοντας κατά νου την ανάλυση των προηγούμενων εννοιών, προτείνουμε το ακόλουθο γενικευμένο συναρτησιακό:

$$E[\mathbf{u}] = \int_{\Omega} \psi(\lambda_1(J_K(\nabla \mathbf{u})), \lambda_2(J_K(\nabla \mathbf{u}))) \, d\mathbf{x} \quad (4.3)$$

όπου $\psi(\lambda_1, \lambda_2) : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$ είναι μία συνάρτηση κόστους που είναι αύξουσα ως προς αμφότερα τα ορίσματά της. Οι ιδιοτιμές λ_i που χρησιμοποιούνται στην (4.3) μετράνε την μέγιστη και ελάχιστη διανυσματική μεταβολή της εικόνας στην γειτονιά του κάθε σημείου.

Η δυσκολία στην θεωρητική ανάλυση του συναρτησιακού (4.3) έγκειται στο ότι, σε αντίθεση με τις περισσότερες σχετικές μεταβολικές μεθόδους (π.χ. [106, 112, 124]), ολοκληρώνει ποσότητες που δεν εξαρτώνται απευθείας από παραγώγους της εικόνας, αλλά από συνελίξεις τους με έναν πυρήνα. Επομένως, οι συναρτησιακές κλίσεις (functional gradients) που χρειάζονται για την ελαχιστοποίηση δεν μπορούν να εξαχθούν χρησιμοποιώντας τις εξισώσεις Euler-Lagrange. Παρόλα αυτά, κάνοντας χρήση του ορισμού της συναρτησιακής κλίσης και πιο περίπλοκο λογισμό, αποδείξαμε το ακόλουθο θεώρημα (βλ. Ενότητα 4.6.2 για ένα περίγραμμα της απόδειξης):

Θεώρημα 4.3.1. *Η συναρτησιακή κλίση του $E[\mathbf{u}]$ (4.3) ως προς καθένα από τα κανάλια $u_i(\mathbf{x})$ της εικόνας είναι:*

$$\delta E[\mathbf{u}]/\delta u_i = -\operatorname{div}(D_K \nabla u_i), \quad i = 1, \dots, N, \quad (4.4)$$

$$\text{όπου: } D_K = K * \left(2 \frac{\partial \psi}{\partial \lambda_1} \boldsymbol{\theta}_1 \otimes \boldsymbol{\theta}_1 + 2 \frac{\partial \psi}{\partial \lambda_2} \boldsymbol{\theta}_2 \otimes \boldsymbol{\theta}_2 \right) \quad (4.5)$$

και $\boldsymbol{\theta}_{1,2}(\mathbf{x})$ είναι τα ιδιοδιανύσματα του $J_K(\nabla \mathbf{u})$.

Συνεπώς, η ροή ταχύτερης καθόδου προς την ελαχιστοποίηση του προτεινόμενου συναρτησιακού (4.3), $\partial u_i / \partial t = -\delta E[\mathbf{u}] / \delta u_i$, είναι ένας νέος γενικός τύπος ανισοτροπικής διάχυσης για διανυσματικές εικόνες. Ας σημειωθεί ότι ο τανυστής D_K της διάχυσης αυτής δημιουργείται χρησιμοποιώντας συνελίξεις με τον πυρήνα K σε δύο φάσεις: **1**) στον σχηματισμό του δομικού τανυστή $J_K(\nabla \mathbf{u})$, από όπου οι ιδιοτιμές λ_i και τα ιδιοδιανύσματα $\boldsymbol{\theta}_i$ υπολογίζονται και **2**) στην συνέλιξη της Εξ. (4.5) που τελικά παράγει τον τανυστή διάχυσης. Τα ιδιοδιανύσματα $\boldsymbol{\theta}_1(\mathbf{x})$ και $\boldsymbol{\theta}_2(\mathbf{x})$ αντιστοιχούν στις κατευθύνσεις της μέγιστης και ελάχιστης διανυσματικής μεταβολής της εικόνας, όπως αυτή η μεταβολή φαίνεται από ένα παράθυρο γύρω από το \mathbf{x} . Στην ακόλουθη ενότητα, το παραπάνω αποτέλεσμα εφαρμόζεται σε μερικές ενδιαφέρουσες ειδικές περιπτώσεις του συναρτησιακού (4.3), καταλήγοντας σε νέες μεθόδους μη γραμμικής διάχυσης.

4.4 Νέες Μέθοδοι σαν Ειδικές Περιπτώσεις του Προτεινόμενου Συναρτησιακού

4.4.1 Τανυστική Ολική Μεταβολή

Μία αποτελεσματική επιλογή για την συνάρτηση κόστους του γενικού συναρτησιακού (4.3) είναι η εξής: $\psi(\lambda_1, \lambda_2) = \sqrt{\lambda_1} + \sqrt{\lambda_2}$. Εφαρμόζοντας το Θεώρημα 4.3.1, καταλήγουμε ότι η ροή ταχύτερης καθόδου για την ελαχιστοποίηση του συναρτησιακού δίνεται από (για κάθε $i = 1, \dots, N$):

$$\frac{\partial u_i}{\partial t} = \operatorname{div} \left(\left[K * \left(\frac{1}{\sqrt{\lambda_1}} \boldsymbol{\theta}_1 \otimes \boldsymbol{\theta}_1 + \frac{1}{\sqrt{\lambda_2}} \boldsymbol{\theta}_2 \otimes \boldsymbol{\theta}_2 \right) \right] \nabla u_i \right) \quad (4.6)$$

Ονομάζουμε την προκύπτουσα μέθοδο *Τανυστική Ολική Μεταβολή*, μιας και γενικεύει την Ολική Μεταβολή και τόσο το συναρτησιακό όσο και η διάχυση βασίζονται σε τανυστές. Η κλασική ΟΜ αντιστοιχεί στην ειδική περίπτωση όπου $N=1$ (εικόνες γκριζας κλίμακας) και $K = \delta(\mathbf{x})$. Σε αυτή την περίπτωση, $\lambda_1 = \|\nabla u\|^2$ και $\lambda_2 = 0$, επομένως $\psi(\lambda_1, \lambda_2) = \|\nabla u\|$.

Η ΜΔΕ (4.6) εξομαλύνει προσαρμοστικά μια διανυσματική εικόνα με τον ακόλουθο τρόπο. Η διάχυση είναι ισχυρή και ισοτροπική στις ομοιογενείς περιοχές της εικόνας (μικρά λ_1 και λ_2), αλλά ασθενέστερη και ανισοτροπικά προσανατολισμένη από τις δομές της εικόνας κοντά στις ακμές (μεγάλο λ_1 , μικρό λ_2). Τέλος, στην γειτονιά των γωνιών της εικόνας (μεγάλα λ_1 και λ_2), η διάχυση είναι ακόμα ασθενέστερη. Η προτεινόμενη Τανυστική Ολική Μεταβολή συνδυάζει τα πλεονεκτήματα της μεθόδου ΟΜ και των μεθόδων διάχυσης με τανυστές: Η διάχυση από τη μια προέρχεται από ένα μεταβολικό πλαίσιο και από την άλλη ελέγχεται από έναν τανυστή που προσαρμόζεται ευέλικτα στις δομές της εικόνας.

4.4.2 Γενικευμένη Ροή Beltrami

Το Γενικευμένο Συναρτησιακό Beltrami (4.1) μπορεί να προκύψει από το συναρτησιακό (4.3) θέτοντας $\psi(\lambda_1, \lambda_2) = \sqrt{(\alpha^2 + \lambda_1)(\alpha^2 + \lambda_2)}$. Το Θεώρημα 4.3.1 συνεπάγεται ότι η ροή καθόδου κλίσης για την ελαχιστοποίηση του συναρτησιακού αυτού δίνεται από (για $i = 1, \dots, N$):

$$\frac{\partial u_i}{\partial t} = \operatorname{div} \left(\left[K * \left(\sqrt{\frac{\alpha^2 + \lambda_2}{\alpha^2 + \lambda_1}} \boldsymbol{\theta}_1 \otimes \boldsymbol{\theta}_1 + \sqrt{\frac{\alpha^2 + \lambda_1}{\alpha^2 + \lambda_2}} \boldsymbol{\theta}_2 \otimes \boldsymbol{\theta}_2 \right) \right] \nabla u_i \right)$$

Ας σημειώσουμε ότι η Ροή Beltrami του άρθρου [112] προκύπτει εάν στην παραπάνω εξίσωση θέσουμε $K = \delta(\mathbf{x})$ αλλά επίσης προσθέσουμε στο δεξί της μέλος τον πολλαπλασιαστικό όρο $1/\psi(\lambda_1, \lambda_2)$. Αυτή η επιπρόσθετη στάθμιση οφείλεται στο γεγονός ότι στο [112], η κάθοδος κλίσης εφαρμόζεται στον χώρο των εμβυθίσεων αντί του χώρου των εικόνων. Όμως, εξαιτίας αυτής της στάθμισης, η διάχυση ελαττώνεται υπερβολικά κοντά στις ακμές και δεν μπορεί να τις βελτιώσει αποτελεσματικά.

4.4.3 Άλλες Ενδιαφέρουσες Περιπτώσεις και Σχέσεις με Προηγούμενες Εργασίες

Εάν στο συναρτησιακό (4.3) επιλέξουμε $\psi(\lambda_1, \lambda_2) = \phi(\lambda_1 + \lambda_2)$ για κάποια συνάρτηση ϕ , τότε η καθοδική κλίση περιγράφεται από μία διάχυση που είναι μη-γραμμική αλλά ισοτροπική:

$$\partial u_i / \partial t = \operatorname{div} (2 [K * \varphi'(K * \|\nabla \mathbf{u}\|^2)] \nabla u_i) \quad (4.7)$$

όπου $\|\nabla \mathbf{u}\|^2 = \sum_{i=1}^N \|\nabla u_i\|^2$. Η περίπτωση αυτή προσφέρει την ακόλουθη εναλλακτική στην κλασική ομαλοποίηση (regularization) του μοντέλου Perona-Malik από τους Catté et al. [27]: Ο συντελεστής διάχυσης $g(\|\nabla \mathbf{u}\|^2)$ αντικαθίσταται από τον όρο $K * g(K * \|\nabla \mathbf{u}\|^2)$ αντί να αντικατασταθεί από τον όρο $g(\|\nabla K * u\|^2)$. Το πλεονέκτημα αυτής της νέας ομαλοποίησης είναι ότι έχει μεταβολική ερμηνεία.

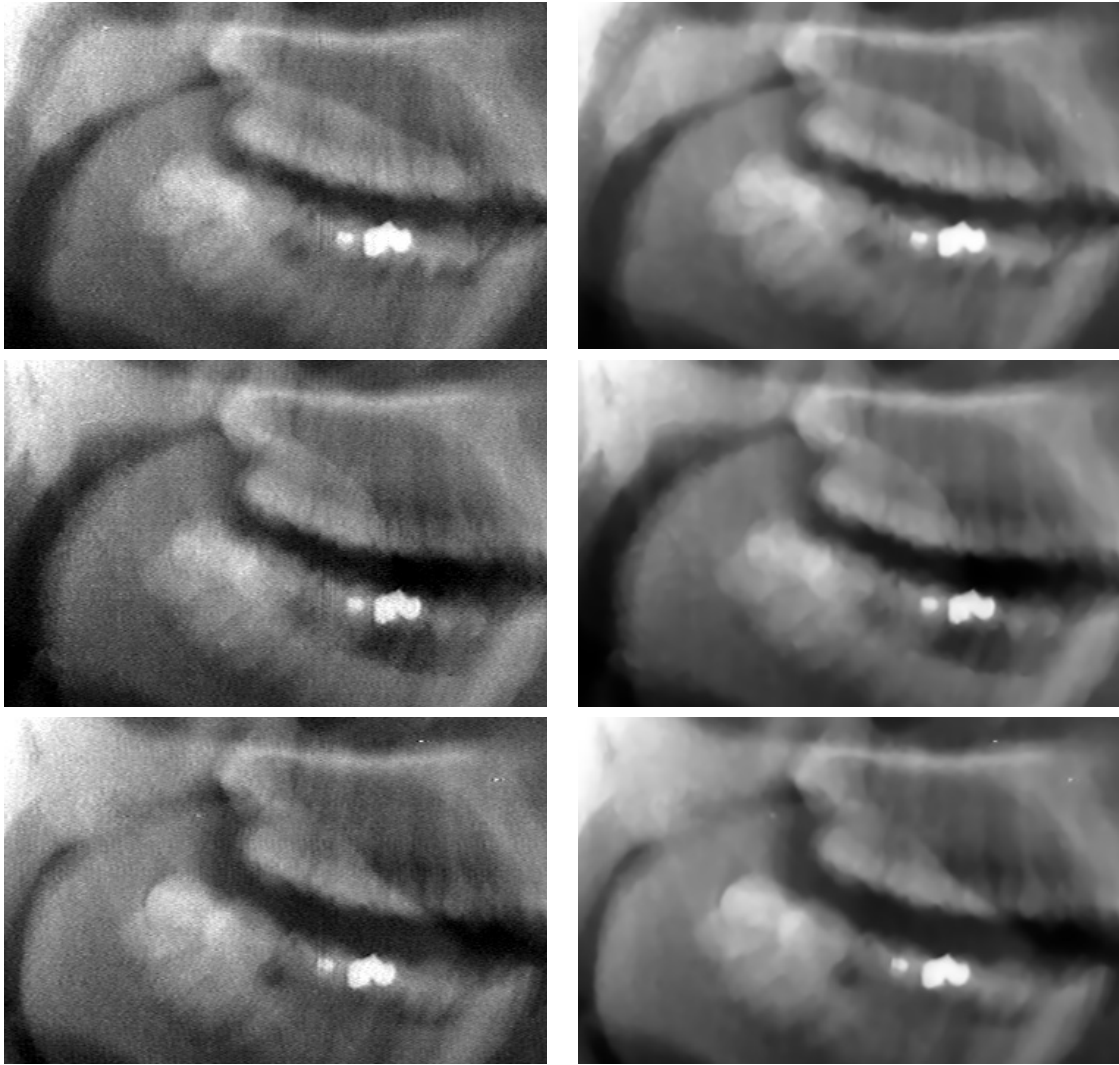
Τέλος, αξίζει να σημειωθεί ότι μία απλούστερη μορφή του προτεινόμενου συναρτησιακού, η οποία δεν περιέχει συνέλιξη με πυρήνα (ή ισοδύναμα περιέχει συνέλιξη με τον πυρήνα $K = \delta(\mathbf{x})$) έχει ήδη μελετηθεί στο [124]. Σε αυτή την περίπτωση, η αντίστοιχη διάχυση είναι ανισοτροπική αλλά μόνον εάν ο αριθμός των καναλιών της εικόνας είναι $N \geq 2$. Επίσης, ο τανυστής διάχυσής της δεν ενσωματώνει πληροφορία σχετικά με την μεταβολή της εικόνας στην γειτονιά του κάθε σημείου.

4.5 Πειραματικά Αποτελέσματα και Συγκρίσεις

Στο Σχ. 4.1 η Τανυστική ΟΜ εφαρμόζεται σε μία ακολουθία εικόνων ακτίνων X, στις οποίες απεικονίζεται η φωνητική οδός ομιλητή. Όπως έχει προαναφερθεί στο Κεφάλαιο 1, πολυάριθμα δεδομένα τέτοιου τύπου είναι διαθέσιμα από καταγραφές παλαιότερων δεκαετιών, όμως συχνά, όπως στο παράδειγμα αυτό, είναι διαταραγμένα από σημαντικό θόρυβο. Παρατηρούμε ότι η προτεινόμενη μέθοδος καταφέρνει να απομακρύνει αποτελεσματικά τον θόρυβο χωρίς να εισάγει θόλωμα ή να αλλοιώσει τις δομές της εικόνας.

Επιπρόσθετα, ακολουθώντας παρόμοια φιλοσοφία με την Ενότητα 3.4, εκτελούμε ποσοτικές αξιολογήσεις σε πειράματα αποθορυβοποίησης, υιοθετώντας το ακόλουθο πλαίσιο: Κάθε φορά, χρησιμοποιούμε μία εικόνα αναφοράς με αμελητέο θόρυβο και προσθέτουμε γκαουσιανό θόρυβο. Η προκύπτουσα θορυβώδης εικόνα χρησιμοποιείται σαν είσοδος στις μεθόδους διάχυσης που συγκρίνονται. Αφού η αρχική καθαρή εικόνα είναι διαθέσιμη, υπολογίζουμε τους λόγους κορυφής σήματος προς θόρυβο (Peak Signal to Noise Ratio - PSNR) κατά τη διάρκεια των εξελίξεων των αντίστοιχων ΜΔΕ και θεωρούμε σαν έξοδο κάθε μεθόδου την εικόνα από την αντίστοιχη ροή που επιτυγχάνει μέγιστο PSNR. Φυσικά, στην πράξη η καθαρή εικόνα δεν είναι διαθέσιμη και ο χρόνος παύσης της διάχυσης πρέπει να εκτιμηθεί, αλλά η επιλογή που κάνουμε εδώ αντιστοιχεί στην καλύτερη δυνατή εκτίμηση για κάθε μέθοδο. Σε όλα τα πειράματα, ο πυρήνας K που χρησιμοποιείται στις προτεινόμενες μεθόδους επιλέχθηκε να είναι μία 2Δ γκαουσιανή με σταθερή τυπική απόκλιση $\rho=0.5$ εικονοστοιχεία.

Στο παράδειγμα του Σχ. 4.2, η προτεινόμενη Τανυστική ΟΜ και η κλασική ΟΜ έχουν εφαρμοστεί σε μία εικόνα γκριζας κλίμακας η οποία έχει αλλοιωθεί με γκαουσιανό θόρυβο. Παρατηρούμε ότι, σε σύγκριση με την κλασική ΟΜ, η προτεινόμενη μέθοδος βελτιώνει τις ακμές πιο αποτελεσματικά και γενικά αποδίδει ένα πιο ικανοποιητικό αποτέλεσμα. Επιπρόσθετα, όπως διαφαίνεται από τα διαγράμματα των PSNR, επιτυγχάνει



Ακολουθία εισόδου

Ακολουθία εξόδου

Σχήμα 4.1: Εφαρμογή της Τανυστικής ΟΜ σε μια ακολουθία εικόνων ακτίνων Χ της φωνητικής οδού ομιλητή. *Αριστερά:* λεπτομέρειες πλαισίων (# 460,560,660) από την ακολουθία εισόδου. *Δεξιά:* αντίστοιχες εξόδου, μετά την εφαρμογή σε κάθε πλαίσιο της Τανυστικής ΟΜ, χρησιμοποιώντας γκαουσιανό πυρήνα τυπικής απόκλισης $\rho=1$, για χρόνο διάχυσης $t=8$.

ένα υψηλότερο μέγιστο PSNR κατά τη διάρκεια της εξέλιξης της ΜΔΕ. Αυτό το μέγιστο αντιστοιχεί σε μικρότερο χρόνο διάχυσης (λιγότερες επαναλήψεις), πράγμα το οποίο αντικατοπτρίζει το ότι η Τανυστική ΟΜ εκμεταλλεύεται τις συνελίξεις με τον πυρήνα K για να προσαρμοστεί με ευελιξία στην εικόνα και να βελτιώσει την σθεναρότητα έναντι του θορύβου.

Το Σχήμα 4.3 δείχνει την εφαρμογή της κλασικής και της γενικευμένης Ροής Beltrami σε ένα παράδειγμα αποθορυβοποίησης έγχρωμων εικόνων. Η προτεινόμενη γενίκευση φαίνεται να εξομαλύνει τις διάφορες περιοχές της εικόνας με έναν πιο ισορροπημένο τρόπο και να απομακρύνει τον θόρυβο παραμορφώνοντας λιγότερο την εικόνα. Επιπρόσθετα,



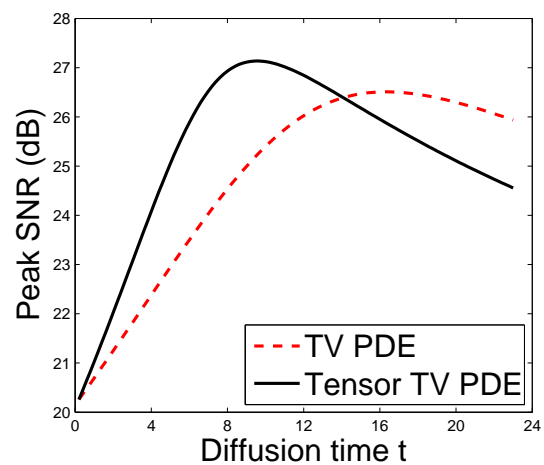
(α') Θορυβώδης Είσοδος (PSNR= 20.1 dB)



(β') ΜΔΕ Ολικής Μεταβολής (*TV PDE*) (PSNR= 26.5 dB)



(γ') ΜΔΕ Τανυστικής Ολικής Μεταβολής (*Tensor TV PDE*) (PSNR= 27.1 dB)



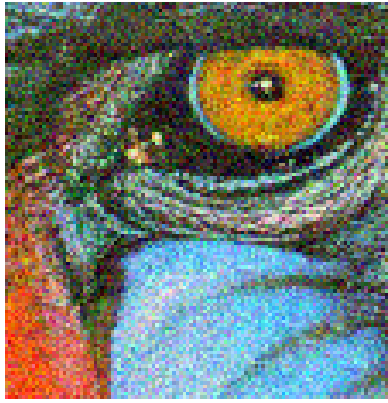
(δ') PSNRs κατά την εξέλιξη των ροών

Σχήμα 4.2: Αποθορυβοποίηση χρησιμοποιώντας την OM και την Τανυστική OM. Τα αποτελέσματα (β) και (γ) αντιστοιχούν στα μέγιστα των PSNR καμπυλών (δ).

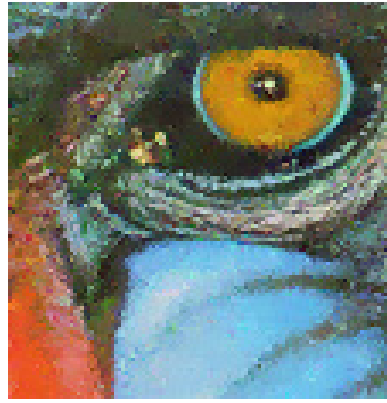
καταλήγει σε ένα βελτιωμένο PSNR.

Τέλος, ακολουθήσαμε το προαναφερθέν πειραματικό πλαίσιο χρησιμοποιώντας 23 φωτογραφίες¹. Τρέξαμε δύο σειρές πειραμάτων, βασισμένοι στις ασπρόμαυρες και έγχρωμες εκδοχές των εικόνων αυτών. Το Σχήμα 4.4 δείχνει συγκεντρωτικές μετρήσεις PSNR για κάθε μέθοδο που δοκιμάστηκε και επίπεδο θορύβου των εικόνων εισόδου. Παρατηρούμε ότι η προτεινόμενη Τανυστική OM αποδίδει σε όλες τις περιπτώσεις τις καλύτερες συγκεντρωτικές μετρήσεις απόδοσης. Η δεύτερη προτεινόμενη μέθοδος, η Γενικευμένη Ροή Beltrami, αποδίδει χειρότερες μετρήσεις από την ΜΔΕ Ολικής Μεταβολής στα πειράματα με τις εικόνες γκριζας κλίμακας. Από την άλλη μεριά όμως, συγκρινόμενη με την Ροή Beltrami στην περίπτωση των έγχρωμων εικόνων, παρουσιάζει μία σημαντικά βελτιωμένη απόδοση όταν τα επίπεδα του θορύβου είναι σχετικά υψηλά (χαμηλά PSNR).

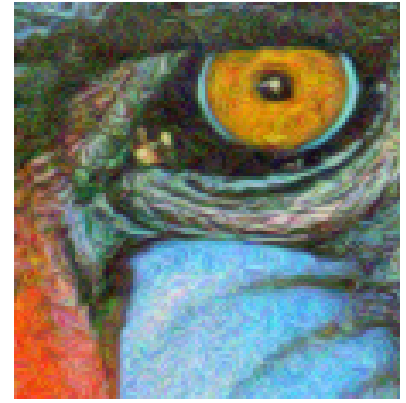
¹Πήραμε αυτό το σύνολο εικόνων από τη δημόσια διαθέσιμη βάση: <http://www.cipr.rpi.edu/resource/stills/kodak.html>.



(α) Θορυβώδης Είσοδος
(PSNR=19.9 dB)

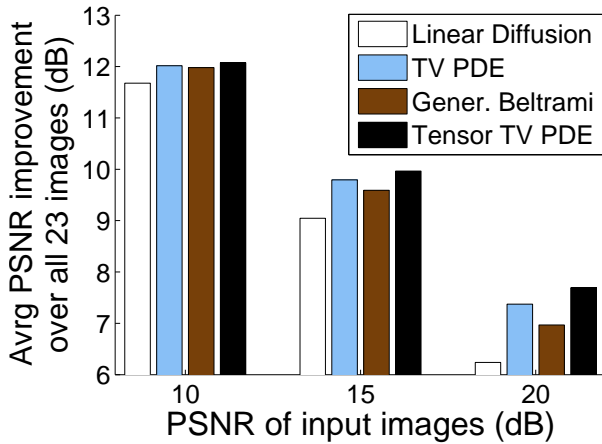


(β) Ροή Beltrami
(PSNR=23.4 dB)

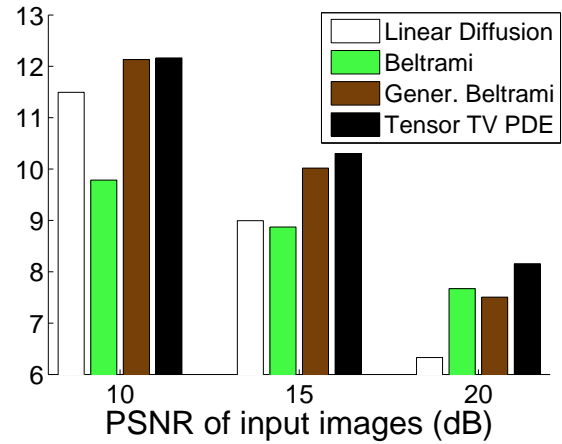


(γ) Γενικευμένη Ροή Beltrami
(PSNR=24.0 dB)

Σχήμα 4.3: Αποθρομβοποίηση μιας έγχρωμης εικόνας. Τα αποτελέσματα (β) και (γ) αντιστοιχούν στα μέγιστα PSNR που επιτεύχθηκαν κατά τη διάρκεια κάθε ροής.



(α) Πειράματα με εικόνες γκριζας κλίμακας



(β) Πειράματα με έγχρωμες εικόνες

Σχήμα 4.4: Συγκεντρωτικές μετρήσεις PSNR για διαφορετικές μεθόδους διάχυσης.

4.6 Θεωρητικές Λεπτομέρειες

4.6.1 Εξαγωγή του Γενικευμένου Συναρτησιακού Beltrami

Σε αυτήν την ενότητα, παρουσιάζεται με περισσότερες θεωρητικές λεπτομέρειες η εξαγωγή του Γενικευμένου Συναρτησιακού Beltrami (4.1) σαν ένα όριο του εμβαδού των εμβυθίσεων που βασίζονται στα τεμάχια εικόνας. Για κάθε σημείο $x \in \Omega$, ορίζουμε μία ακολουθία $\{\mathcal{P}_n^u(x)\}_{n=1,2,\dots}$ από τεμάχια εικόνας γύρω από το x δειγματοληπώντας όλα τα κανάλια της εικόνας εντός του τετραγώνου $S_x = [x-L, x+L] \times [y-L, y+L]$, για κάποιο $L > 0$, με ομοιόμορφο βήμα $\epsilon_n = L/n$ και στη συνέχεια σταθμίζοντας τα δείγματα με μία

Χρησιμοποιώντας το μετρικό \mathcal{G}_n , μπορούμε να εκφράσουμε το εμβαδόν της κάθε εμβυθισμένης επιφάνειας \mathbf{F}_n σαν $A_n = \int_{\Omega} \sqrt{\det(\mathcal{G}_n)}$ [112]. Άρα:

$$A[\mathbf{u}] = \lim_{n \rightarrow \infty} A_n = \int_{\Omega} \sqrt{\det(\mathcal{G})} \stackrel{(4.9)}{=} \int_{\Omega} \sqrt{(\alpha^2 + \lambda_1)(\alpha^2 + \lambda_2)},$$

μιας και, όπως είναι δυνατό να δειχθεί, το παραπάνω όριο μπορεί να αλλάξει θέση με την ολοκλήρωση.

4.6.2 Περίγραμμα της Απόδειξης του Θεωρήματος 4.3.1

Εστω $\mathbf{v}(\mathbf{x}) = [v_1, \dots, v_N]$ μία αυθαίρετη διανυσματική συνάρτηση δοκιμής που μηδενίζεται στο $\partial\Omega$, $V(\mathbf{x}) = \text{diag}(\mathbf{v})$ και $\boldsymbol{\varepsilon} \in \mathbb{R}^N$. Εάν ορίσουμε $\Phi(\boldsymbol{\varepsilon}) = E[\mathbf{u} + V\boldsymbol{\varepsilon}]$, τότε η συναρτησιακή κλίση του $E[u]$ ως προς κάθε $u_i(\mathbf{x})$ πρέπει εξ ορισμού να ικανοποιεί:

$$\left. \frac{\partial \Phi(\boldsymbol{\varepsilon})}{\partial \varepsilon_i} \right|_{\boldsymbol{\varepsilon}=0} = \langle v_i, \delta E[\mathbf{u}] / \delta u_i \rangle_{L^2(\Omega)} \quad (4.10)$$

για οποιαδήποτε συνάρτηση δοκιμής $\mathbf{v}(\mathbf{x})$. Από την άλλη μεριά, αυτή η πρώτη μεταβολή του $E[u]$ μπορεί να γραφθεί επίσης ως εξής (χρησιμοποιώντας την Εξ. (4.3)):

$$\left. \frac{\partial \Phi(\boldsymbol{\varepsilon})}{\partial \varepsilon_i} \right|_{\boldsymbol{\varepsilon}=0} = \int_{\Omega} d\mathbf{x} \sum_{j=1}^2 \frac{\partial \psi}{\partial \lambda_j} \frac{\partial \lambda_j(J_K(\mathbf{u} + V\boldsymbol{\varepsilon}))}{\partial \varepsilon_i} \Big|_{\boldsymbol{\varepsilon}=0} \quad (4.11)$$

ξανά για οποιαδήποτε συνάρτηση δοκιμής $\mathbf{v}(\mathbf{x})$. Εάν $a_{k\ell}(\mathbf{u})$ είναι τα στοιχεία του δομικού τανυστή $J_K(\mathbf{u})$ (4.2), δηλαδή εάν:

$$J_K(\mathbf{u}) = \begin{pmatrix} a_{11}(\mathbf{u}) & a_{12}(\mathbf{u}) \\ a_{12}(\mathbf{u}) & a_{22}(\mathbf{u}) \end{pmatrix} \quad (4.12)$$

τότε οι ιδιοτιμές του τανυστή δίνονται από την σχέση [135]:

$$\lambda_j(J_K(\mathbf{u})) = (a_{11} + a_{22} + C_j \sqrt{(a_{11} - a_{22})^2 + 4a_{12}^2})/2, \quad (4.13)$$

$$\text{όπου } C_j = \begin{cases} 1, & \text{εάν } j = 1 \\ -1, & \text{εάν } j = 2 \end{cases} \quad (4.14)$$

Χρησιμοποιώντας τον ορισμό (4.2), μπορεί να δειχθεί ότι (για $k, \ell=1,2$ και $i=1, \dots, N$):

$$\left. \frac{\partial a_{k\ell}(\mathbf{u} + V\boldsymbol{\varepsilon})}{\partial \varepsilon_i} \right|_{\boldsymbol{\varepsilon}=0} = K * \left(\frac{\partial u_i}{\partial x_k} \frac{\partial v_i}{\partial x_\ell} + \frac{\partial u_i}{\partial x_\ell} \frac{\partial v_i}{\partial x_k} \right). \quad (4.15)$$

Καταλήγουμε λοιπόν στην ακόλουθη σχέση (για $j=1,2$):

$$\left. \frac{\partial \lambda_j(J_K(\mathbf{u} + V\boldsymbol{\varepsilon}))}{\partial \varepsilon_i} \right|_{\boldsymbol{\varepsilon}=0} = [(a_{11} - a_{22})K * \langle M_1 \nabla u_i, \nabla v_i \rangle + 2a_{12}K * \langle M_2 \nabla u_i, \nabla v_i \rangle] \frac{C_j}{\lambda_1 - \lambda_2} + K * \langle \nabla u_i, \nabla v_i \rangle \quad (4.16)$$

όπου $M_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ και $M_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Βασισμένοι στην τελευταία εξίσωση, παρατηρούμε ότι το δεξί μέλος της Εξ. (4.11) μπορεί να εκφραστεί σαν ένα άθροισμα 6 ολοκληρωμάτων (3 για κάθε j) τα οποία έχουν όλα την ακόλουθη μορφή:

$$I = \int_{\Omega} f(\mathbf{x}) K * \langle M \nabla u_i, \nabla v_i \rangle d\mathbf{x}, \quad (4.17)$$

για κατάλληλες επιλογές στη συνάρτηση $f(\mathbf{x})$ και στην 2×2 μήτρα M . Στο σημείο αυτό, απαιτείται ένας πιο αυστηρός ορισμός των συνελιξων με τον πυρήνα K . Με αναφορά στον ορισμό του δομικού τανυστή (4.2), όπου οι συνελιξεις αυτές χρησιμοποιούνται αρχικά, πρώτα επεκτείνουμε την εικόνα \mathbf{u} στο \mathbb{R}^2 με κατοπτρισμό ακολουθούμενο από περιοδικοποίηση και στη συνέχεια κάνουμε χρήση της κλασικής συνεχούς 2Δ συνέλιξης. Μπορεί να δειχθεί ότι η διαδικασία αυτή είναι ισοδύναμη με έναν ολοκληρωτικό μετασχηματισμό με πεδίο ολοκλήρωσης το Ω και του οποίου ο πυρήνας ολοκλήρωσης $\hat{K}(\boldsymbol{\tau}; \mathbf{x})$ προκύπτει από τον $K(\mathbf{x} - \boldsymbol{\tau})$ ύστερα από κατάλληλες αναδιπλώσεις και χωρικές αντιγραφές.

Υποθέτοντας ότι ο πυρήνας K ικανοποιεί την ιδιότητα συμμετρίας $K(x, y) = K(|x|, |y|)$ δεν είναι δύσκολο να δειχθεί ότι $\hat{K}(\boldsymbol{\tau}; \mathbf{x}) = \hat{K}(\mathbf{x}; \boldsymbol{\tau})$. Έχοντας αυτά κατά νου και χρησιμοποιώντας ολοκλήρωση κατά παράγοντες και το γεγονός ότι τα v_i μηδενίζονται στο $\partial\Omega$, μπορούμε να γράψουμε το ολοκλήρωμα I ως ακολούθως:

$$\begin{aligned} I &= \int_{\Omega} f(\mathbf{x}) \left\{ \int_{\Omega} \langle M \nabla u_i(\boldsymbol{\tau}), \nabla v_i(\boldsymbol{\tau}) \rangle \hat{K}(\boldsymbol{\tau}; \mathbf{x}) d\boldsymbol{\tau} \right\} d\mathbf{x} \\ &= -\langle v_i, \operatorname{div}((K * f) M \nabla u_i) \rangle_{L^2(\Omega)} \end{aligned}$$

Χρησιμοποιώντας τις Εξ. (4.10), (4.11), (4.16), το παραπάνω αποτέλεσμα και το θεμελιώδες λήμμα του λογισμού των μεταβολών, καταληγούμε ότι (για $i = 1, \dots, N$):

$$\delta E[\mathbf{u}] / \delta u_i = -\operatorname{div}(D \nabla u_i), \quad \text{όπου:} \quad (4.18)$$

$$\begin{aligned} D &= \sum_{j=1}^2 \left\{ (K * \frac{\partial \psi}{\partial \lambda_j} \frac{(a_{11} - a_{22}) C_j}{\lambda_1 - \lambda_2}) M_1 + (K * \frac{\partial \psi}{\partial \lambda_j} \frac{2a_{12} C_j}{\lambda_1 - \lambda_2}) M_2 + (K * \frac{\partial \psi}{\partial \lambda_j}) I_2 \right\} \\ &= K * \sum_{j=1}^2 \frac{\partial \psi}{\partial \lambda_j} \underbrace{\left(I_2 + \frac{a_{11} - a_{22}}{\lambda_1 - \lambda_2} C_j M_1 + \frac{2a_{12}}{\lambda_1 - \lambda_2} C_j M_2 \right)}_{B_j} \end{aligned}$$

Μετά από κάποιες αλγεβρικές πράξεις, βρίσκουμε ότι οι ιδιοτιμές της μήτρας B_j , $j=1,2$, ισούνται με 2 και 0 και το μοναδιαίο ιδιοδιάνυσμα που αντιστοιχεί στην μη μηδενική ιδιοτιμή είναι $\boldsymbol{\theta}_j$. Επομένως $B_j = 2\boldsymbol{\theta}_j \otimes \boldsymbol{\theta}_j$. Αυτό ολοκληρώνει την απόδειξη.

4.7 Συμπεράσματα

Εισάγαμε ένα γενικό συναρτησιακό για διανυσματικές εικόνες και αποδείξαμε ότι η ελαχιστοποίησή του αποφέρει μία διαδικασία διάχυσης βασισμένη σε τανυστές. Χρησιμοποιώντας αυτό το αποτέλεσμα, προτείναμε δύο νέες μεθόδους ανισοτροπικής διάχυσης.

Τα αποτελέσματα που παρουσιάστηκαν παρέχουν ενδείξεις σχετικά με την αποτελεσματικότητα των μεθόδων αυτών όταν εφαρμόζονται στην αποθρομβοποίηση εικόνων. Χάρη στη μεταβολική διατύπωση του ακολουθούμενου πλαισίου, τέτοιες κανονικοποιημένες διαχύσεις με τανυστές μπορούν άμεσα να εφαρμοστούν και σε διάφορα άλλα προβλήματα επεξεργασίας εικόνας, όπως αποκατάσταση, παρεμβολή και ενδοσυμπλήρωση (inpainting) εικόνων. Αυτές αποτελούν κατευθύνσεις μελλοντικής έρευνας.

Κεφάλαιο 5

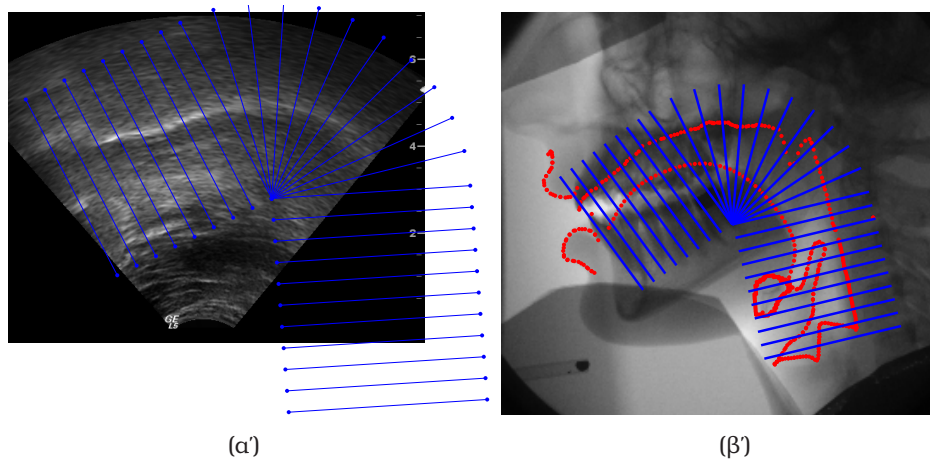
Παρακολούθηση της Γλώσσας σε Εικόνες Υπερήχων με Ενεργά Μοντέλα Εμφάνισης

5.1 Εισαγωγή

Το σχήμα και η δυναμική της ανθρώπινης γλώσσας κατά τη διάρκεια της ομιλίας είναι σημαντικά στην ανάλυση και μοντελοποίηση του ανθρώπινου συστήματος παραγωγής φωνής. Η απεικόνιση υπερήχων αποτελεί έναν από τους πιο αξιόπιστους τρόπους καταγραφής τέτοιου είδους πληροφορίας. Μπορεί να χρησιμοποιηθεί σχετικά εύκολα, δεν εκθέτει το υποκείμενο σε ακτινοβολία και επιτυγχάνει ρυθμούς δειγματοληψίας που μπορούν να καταγράψουν την γρήγορη κίνηση της γλώσσας. Δεδομένου του ότι ακόμα και λίγα λεπτά ομιλίας αντιστοιχούν σε χιλιάδες πλαίσια υπερήχων (ultrasound frames), γίνεται φανερό ότι η αυτόματη εξαγωγή της καμπύλης της γλώσσας μπορεί να προσφέρει σημαντική βοήθεια. Πρόκειται όμως για ένα αρκετά δύσκολο πρόβλημα, λόγω του ότι οι εικόνες υπερήχων περιέχουν υψηλές ποσότητες θορύβου τύπου speckle, κάποια μέρη της καμπύλης της γλώσσας δεν είναι ορατά ενώ τα υπόλοιπα είναι μόνο ασθενώς ορατά.

Σε αυτό το κεφάλαιο, προτείνουμε μία νέα μέθοδο παρακολούθησης η οποία αξιοποιεί πρότερη πληροφορία σχετικά με την μεταβολή του σχήματος της καμπύλης της γλώσσας. Αυτή η μέθοδος είναι εύρωστη ακόμα και σε περιπτώσεις χαμηλής ορατότητας της γλώσσας. Επίσης, εκτός του ότι εξαγει τα ορατά τμήματα της καμπύλης της γλώσσας σε κάθε πλαίσιο, παράλληλα παρέχει *εκτίμηση της επέκτασης* (extrapolation) της καμπύλης στα αόρατα τμήματα, χάρη στο μοντέλο για την μεταβολή του σχήματος. Χρησιμοποιείται η μεθοδολογία των Ενεργών Μοντέλων Εμφάνισης (Active Appearance Models - AAM) [35, 83, 94], ειδικά προσαρμοσμένη για την παρούσα εφαρμογή. Το μοντέλο μεταβολής του σχήματος της γλώσσας εκπαιδεύεται χρησιμοποιώντας επισημειωμένα βίντεο ακτίνων X του κεφαλιού του ομιλητή κατά τη διάρκεια της ομιλίας. Το μοντέλο υψής, δηλ. το μοντέλο των φωτεινοτήτων των εικόνων υπερήχων γύρω από την γλωσσική καμπύλη¹ εκπαιδεύεται σε επισημειωμένα πλαίσια υπερήχων. Το πρόβλημα παρακολούθησης διατυπώνεται σε ένα μπεϋζιανό (bayesian) πλαίσιο εργασίας, περιλαμβάνοντας

¹Όπως συνηθίζεται στα AAM, ο όρος ‘υψή’ χρησιμοποιείται με την έννοια που έχει στην γραφική υπολογιστών.



Σχήμα 5.1: Εικόνες υπερήχων και ακτίνων X του ομιλητή, στις οποίες έχει υπερτεθεί το ευθυγραμμισμένο πλέγμα της φωνητικής οδού (μπλε γραμμές).

εκτίμηση των παραμέτρων του μοντέλου σε κάθε πλαίσιο. Τα πειραματικά αποτελέσματα επιδεικνύουν την αποτελεσματικότητα της προτεινόμενης μεθόδου και την βελτιωμένη της απόδοση, σε σύγκριση με τις μεθόδους των [71] και [7].

5.2 Προκαταρκτικά

5.2.1 Σύστημα Καταγραφής Δεδομένων

Το σύστημα καταγραφής δεδομένων που χρησιμοποιούμε είναι αυτό που περιγράφεται στα [7, 8]. Μεταξύ άλλων, περιλαμβάνει απεικόνιση υπερήχων της γλώσσας του ομιλητή στα 66 Hz. Το Σχ. 5.1(α) παρουσιάζει ένα παράδειγμα ενός καταγεγραμμένου πλαισίου υπερήχων. Μόνο ένα μέρος της τομής του κατώτερου συνόρου της φωνητικής οδού με το midsagittal επίπεδο είναι ορατό. Αναφερόμαστε σε αυτή την τομή ως *γλωσσική καμπύλη* μας και το μεγαλύτερο της κομμάτι αντιστοιχεί στη γλώσσα.

5.2.2 Χρήση των Ακτίνων X για την Μοντελοποίηση Σχήματος

Για να μοντελοποιήσουμε την μεταβολή του σχήματος της γλωσσικής καμπύλης, χρησιμοποιούμε βίντεο ακτίνων X του ίδιου ομιλητή κατά τη διάρκεια της ομιλίας, Σχ. 5.1(β). Σε αντίθεση με τις εικόνες υπερήχων, ολόκληρη η γλωσσική καμπύλη είναι ορατή στις εικόνες ακτίνων X και αυτό τις κάνει πιο κατάλληλες για την επιθυμητή μοντελοποίηση σχήματος. Όπως θα γίνει φανερό στην Ενότητα 5.4, η μεθοδολογία της μοντελοποίησης σχήματος που ακολουθούμε προϋποθέτει ότι το σχήμα αναπαρίσταται από έναν σταθερό αριθμό ενδεικτικών σημείων (landmark points). Αλλά η κύρια δυσκολία στην μοντελοποίηση σχήματος της γλώσσας έγκειται στο γεγονός ότι, σε αντίθεση με άλλες εφαρμογές (π.χ. μοντελοποίηση προσώπου) η αντιστοιχία των ενδεικτικών σημείων από πλαίσιο σε πλαίσιο δεν είναι προφανής, μιας και δεν είναι δυνατόν να επισημειωθούν από κάποιο χρήστη το ίδιο σύνολο σημείων της γλώσσας σε διαφορετικές εικόνες υπερήχων. Για αυτόν τον λόγο χρησιμοποιούμε, παρόμοια με το [75], ένα **πλέγμα φωνητικής οδού** το οποίο έχει ένα σταθερό σχήμα και του οποίου η *πόζα* (δηλ. η θέση, ο προσανατολισμός και η

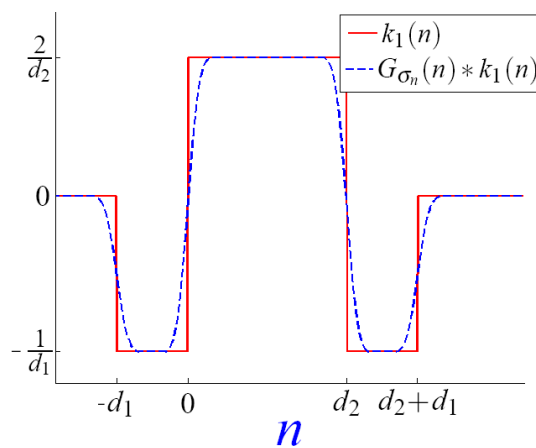
κλίμακα) είναι σταθερή ως προς τον ουρανίσκο, Σχ. 5.1(β). Οι τομές της γλωσσικής καμπύλης με τις γραμμές αυτού του πλέγματος σχηματίζουν τα σημεία που αναπαριστούν το σχήμα της γλώσσας.

Ας σημειωθεί εδώ ότι το μοντέλο σχήματος της γλώσσας θα μπορούσε να βασιστεί σε άλλου τύπου οπτικά δεδομένα που είναι ευκολότερα στην καταγραφή από τις ακτίνες X, όπως στις εικόνες MRI. Επίσης το ίδιο μοντέλο θα μπορούσε να χρησιμοποιηθεί και για διαφορετικό ομιλητή, μετά από μία κατάλληλη προσαρμογή.

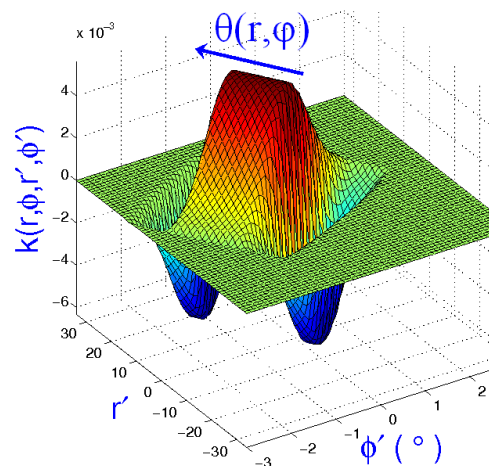
5.2.3 Εκτίμηση της Πόζας του Πλέγματος Φωνητικής Οδού

Για την κατασκευή του μοντέλου σχήματος και την μετέπειτα χρησιμοποίησή του στην παρακολούθηση, είναι απαραίτητο η πόζα του πλέγματος φωνητικής οδού να εκτιμάται σε κάθε πλαίσιο ακτίνων X και υπερήχων, Σχ. 5.1. Αυτό μπορεί να επιτευχθεί εάν γνωρίζουμε την θέση του ουρανίσκου σε κάθε πλαίσιο. Για τις ακτίνες X, αυτό είναι άμεσο, διότι ο ουρανίσκος είναι ορατός. Όμως, τα πράγματα δεν είναι τόσο απλά στην περίπτωση των εικόνων υπερήχων. Για τον λόγο αυτό, εικόνες MRI του κεφαλιού του ομιλητή, στις οποίες ο ουρανίσκος είναι επίσης ορατός, αντιστοιχίζονται γεωμετρικά με κάθε πλαίσιο υπερήχων. Αυτή η αντιστοίχιση διεκπεραιώνεται με τη μέθοδο που περιγράφεται στο [8], κάνοντας χρήση δεδομένων θέσης των ηλεκτρομαγνητικών αισθητήρων στο probe υπερήχων και το κεφάλι.

5.3 Προεπεξεργασία των Πλαισίων Υπερήχων



(α)



(β)

Σχήμα 5.2: Κατασκευή του πυρήνα $k(r, \phi; r', \phi')$, ο οποίος συσχετίζεται με την εικόνα υπερήχων.

Για να βελτιώσουμε την ορατότητα της γλωσσικής καμπύλης, φιλτράρουμε πρώτα τις εικόνες υπερήχων χρησιμοποιώντας την μέθοδο που προτείναμε στο [7]. Αυτή η μέθοδος περιορίζει τα πρότυπα speckle των υπερήχων και κάνει την παρακολούθηση της γλώσσας πιο εύρωστη.

Η μέθοδος εκμεταλλεύεται το γεγονός ότι το ορατό μέρος της γλώσσας στις εικόνες υπερήχων αντιστοιχεί στο κατώτερο σύνορο μιας ζώνης σχετικά υψηλής φωτεινότητας. Επεξεργάζεται ανεξάρτητα κάθε πλαίσιο $u(x, y)$ της ακολουθίας υπερήχων και αποτελείται από τα ακόλουθα βήματα:

1. Η $u(x, y)$ μετατρέπεται στην $u(r, \phi)$, η οποία είναι η αναπαράσταση της εικόνας σε ένα σύστημα πολικών συντεταγμένων με αρχή το σημείο τομής των δεσμών υπερήχων. Αυτή η αναπαράσταση είναι πιο βολική διότι η κατεύθυνση της μεταβλητής r είναι ακριβώς η κατεύθυνση της δέσμης των υπερήχων.

2. Σε κάθε σημείο, γίνεται μία εύρωστη εκτίμηση του προσανατολισμού $\theta(r, \phi)$ που είναι κάθετος στις ακμές της $u(r, \phi)$. Αυτό γίνεται υπολογίζοντας πρώτα τον *δομικό τανυστή* της $u(r, \phi)$ [135] (βλ. επίσης Ενότητα 3.2.6):

$$\mathbf{J}(r, \phi) = G_\rho * (\nabla(G_\sigma * u) \otimes \nabla(G_\sigma * u)) , \quad (5.1)$$

όπου το ' \otimes ' συμβολίζει τανυστικό γινόμενο. Στη συνέχεια, ο $\theta(r, \phi) \in [0, \pi)$ εξάγεται σαν ο προσανατολισμός του ιδιοδιανύσματος που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή του $\mathbf{J}(r, \phi)$. Όπως έχει προαναφερθεί, χάρη στις συνελίξεις στην Εξ. (5.1), ο \mathbf{J} δεν επηρεάζεται από λεπτομέρειες της εικόνας μικρότερες από $O(\sigma)$ ενώ παράλληλα λαμβάνει υπόψιν την μεταβολή της εικόνας εντός μιας γειτονιάς μεγέθους $O(\rho)$.

Επομένως, η παράμετρος σ πρέπει να είναι σχετικά μικρή, της τάξης του χαρακτηριστικού μεγέθους των προτύπων speckle, ενώ η ρ πρέπει να είναι μεγαλύτερη από σ , συγκρίσιμου μεγέθους με το πλάτος της φωτεινής ζώνης της γλώσσας στην εικόνα $u(r, \phi)$.

3. Εφαρμόζεται συσχέτιση μεταξύ της $u(r, \phi)$ και ενός χωρικά μεταβαλλόμενου πυρρήνα $k(r, \phi; r', \phi')$ (βλ. Σχήμα 5.2β), ο οποίος ευθυγραμμίζεται σε κάθε σημείο με την κατεύθυνση της $\theta(r, \phi)$:

$$f_1(r, \phi) = \iint u(r', \phi') k(r, \phi; r + r', \phi + \phi') dr' d\phi' . \quad (5.2)$$

Η απόκριση $f_1(r, \phi)$ λαμβάνει υψηλές τιμές κυρίως στο κατώτερο σύνορο των φωτεινών ζωνών. Πιο συγκεκριμένα, ο πυρρήνας k κατασκευάζεται σύμφωνα με τα ακόλουθα:

Αρχικά, θεωρούμε τον τμηματικά συνεχή 1D πυρρήνα $k_1(n)$ που έχει σχεδιαστεί στο Σχ. 5.2. Αυτός ο πυρρήνας αντιστοιχεί στην διακύμανση μιας ιδανικής φωτεινής ζώνης κατά την κατεύθυνση n που την τέμνει κάθετα (το σημείο $n = 0$ αντιστοιχεί στο κατώτερο σύνορο αυτής της ζώνης). Επομένως, η παράμετρος d_2 πρέπει να προσεγγίζει το τυπικό πλάτος της ζώνης της γλώσσας στην εικόνα $u(r, \phi)$. Επίσης, επιλέγουμε $d_1 = 0.44d_2$.

Στη συνέχεια, κατασκευάζεται η ακόλουθη κανονικοποιημένη και επεκταμένη στις δύο διαστάσεις εκδοχή του $k_1(n)$:

$$k_2(n, \xi) = (G_{\sigma_n}(n) * k_1(n)) G_{\sigma_\xi}(\xi)$$

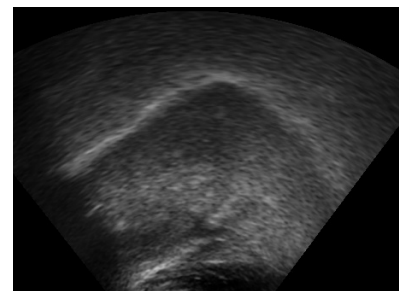
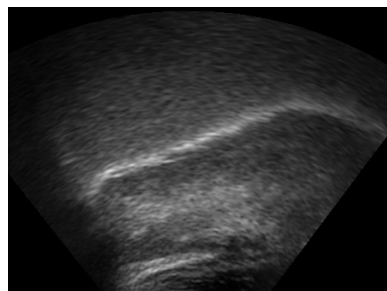
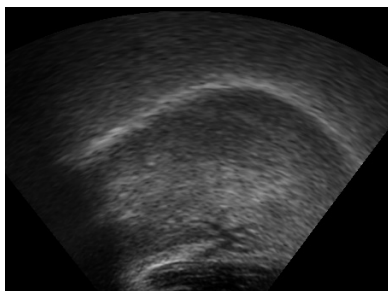
όπου επιλέγουμε $\sigma_n = d_1/8$ και $\sigma_\xi = d_1$ (βλ. επίσης το Σχ. 5.2.α). Η συνέλιξη με τον G_{σ_n} εξομαλύνει τις μεταβάσεις μεταξύ των διαφορετικών περιοχών. Επίσης, η επέκταση στην κατεύθυνση ξ κάνει την απόκριση f_1 περισσότερο εύρωστη απέναντι στο θόρυβο speckle. Η παράμετρος σ_ξ διαλέγεται σχετικά μεγάλη, μιας και κοντά στην γλώσσα οι μεταβολές της εικόνας στην κατεύθυνση ξ προέρχονται μόνο από τα πρότυπα speckle. Συνοψίζοντας, οι γκαουσιανές G_{σ_n} και G_{σ_ξ} προσφέρουν μία προσαρμοστική ανισοτροπική εξομάλυνση, ευθυγραμμισμένη με τις ακμές της $u(r, \phi)$.

Τελικά, ο πυρήνας $k(r, \phi; r', \phi')$ (Σχ. 5.2.6) εξαγεται μέσω της περιστροφής της συνάρτησης $k_2(r', \phi')$ κατά τη γωνία $\theta(r, \phi)$.

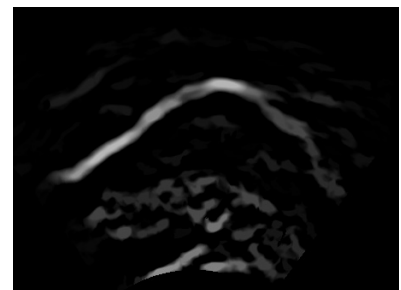
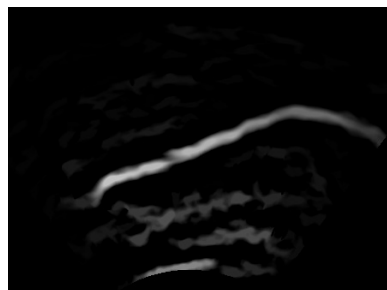
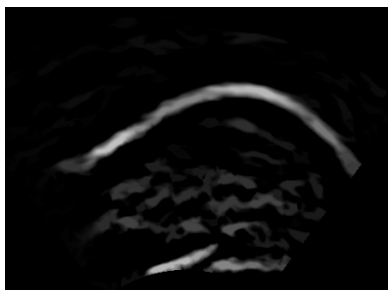
4. Υπολογίζουμε την συνάρτηση $f_2(r, \phi)$ από την $f_1(r, \phi)$ μηδενίζοντας όλες τις αρνητικές τιμές της f_1 . Αυτό γίνεται διότι στα σημεία όπου $f_1 < 0$, η μεταβολή της εικόνας είναι πιο κοντά στο πρότυπο μιας σκοτεινής αντί μιας φωτεινής ζώνης.

5. Η $f_2(r, \phi)$ μετασχηματίζεται σε $f_2(x, y)$. Ύστερα, υπολογίζεται η $f_3(x, y)$ σαν το μορφολογικό area opening της $f_2(x, y)$ για μία σχετικά μικρή κλίμακα A [128]. Αυτή η πράξη περιορίζει τις φωτεινές περιοχές με εμβαδόν μικρότερο από A . Συνήθως, αυτές οι περιοχές δημιουργούνται από τον θόρυβο speckle, μιας και οι ζώνες της γλώσσας έχουν πολύ μεγαλύτερο εμβαδόν.

Στο Σχ. 5.3. παρουσιάζονται μερικά παραδείγματα εφαρμογής αυτής της μεθόδου, όπου παρατηρούμε ότι βελτιώνει σημαντικά την ορατότητα της γλώσσας ενώ ελαχιστοποιεί τον θόρυβο speckle.



Ακολουθία εισόδου



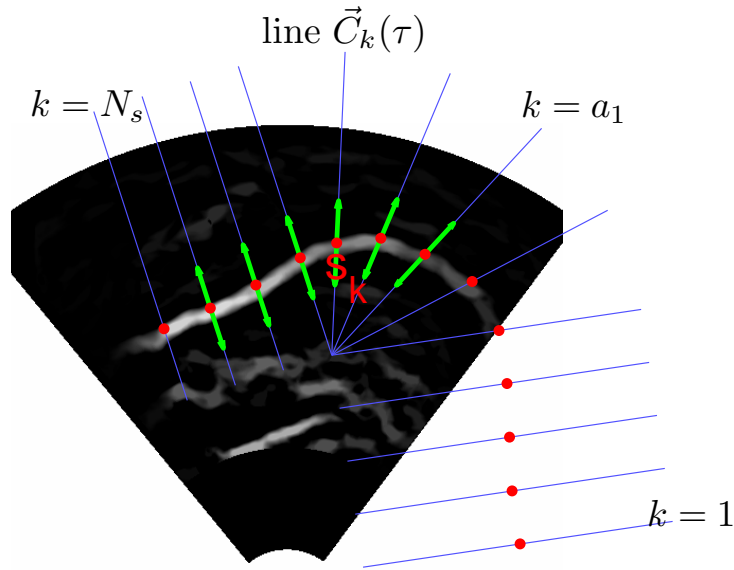
Ακολουθία εξόδου

Πλαίσιο #50

Πλαίσιο #100

Πλαίσιο #150

Σχήμα 5.3: Προσαρμοστικό φιλτράρισμα ακολουθίας εικόνων υπερήχων με την προτεινόμενη μέθοδο. Η ορατότητα της γλώσσας βελτιώνεται σημαντικά ενώ ο θόρυβος speckle ελαχιστοποιείται.



Σχήμα 5.4: Φιλτραρισμένο πλαίσιο υπερήχων και αναπαράσταση της εμφάνισης της γλωσσικής καμπύλης. Οι κουκκίδες πάνω στο πλέγμα φωνητικής οδού αναπαριστούν το σχήμα της καμπύλης. Τα διανύσματα βρίσκονται πάνω στις γραμμές πλέγματος ενεργού υφής και δείχνουν το παράθυρο στο οποίο η εικόνα δειγματοληπτείται για να σχηματιστεί το διάνυσμα υφής.

5.4 Ενεργό Μοντέλο Παρακολούθησης της Γλώσσας

5.4.1 Αναπαράσταση της Εμφάνισης της Γλώσσας στα Πλαίσια Υπερήχων

Χρησιμοποιώντας ένα πλαίσιο εργασίας αντίστοιχο με αυτό των AAM [35, 83], αναπαριστούμε την *εμφάνιση* της γλώσσας στα φιλτραρισμένα πλαίσια υπερήχων. Αυτή αποτελείται από το *σχήμα* της γλωσσικής καμπύλης και την *υφή*, δηλαδή τις φωτεινότητες γύρω από τα ορατά μέρη της καμπύλης αυτής.

Έστω ότι το γεωμετρικά αντιστοιχισμένο πλέγμα φωνητικής οδού αποτελείται από N_s γραμμές. Το σχήμα της γλώσσας αναπαρίσταται από το διάνυσμα σχήματος $\mathbf{s} = [s_1, \dots, s_{N_s}]^T$, το οποίο περιέχει τις βαθμωτές τιμές s_k που καθορίζουν τα σημεία τομής των N_s γραμμών του πλέγματος φωνητικής οδού με την καμπύλη της γλώσσας. Κάθε s_k είναι η απόσταση του σημείου τομής από το σημείο εκκίνησης της αντίστοιχης γραμμής πλέγματος, βλ. Σχ. 5.4.

Το διάνυσμα υφής $\mathbf{g}(\mathbf{s})$ αναπαριστάνει την υφή γύρω από κάποια μέρη της γλώσσας, όπως αυτά ορίζονται από το υποσύνολο των γραμμών πλέγματος $G_{act} = \{a_1, \dots, a_{N_a}\} \subset \{1, \dots, N_s\}$, τις οποίες αποκαλούμε *γραμμές πλέγματος ενεργού υφής* (βλ. Ενότητα 5.4.3 για το πώς διαλέγουμε αυτό το υποσύνολο). Ο λόγος για τον οποίο χρησιμοποιούμε μόνο ένα υποσύνολο των γραμμών πλέγματος είναι ότι κάποια τμήματα της γλωσσικής καμπύλης δεν είναι ποτέ ή είναι σπάνια ορατά στις εικόνες υπερήχων. Το N_g -διάστατο διάνυσμα $\mathbf{g}(\mathbf{s})$ σχηματίζεται δειγματοληπτώντας τα φιλτραρισμένα πλαίσια $u(x, y)$ σε 1Δ παράθυρα πάνω στις γραμμές πλέγματος $k \in G_{act}$ και εκατέρωθεν από τα

αντίστοιχα χαρακτηριστικά σημεία της γλώσσας s_k (βλ. Σχ. 5.4):

$$\mathbf{g}(\mathbf{s}) = \left[\underbrace{[u_{a_1}(s_{a_1}+t)]_{t \in W}^T}_{1 \times N_W} \cdots \underbrace{[u_{a_{N_a}}(s_{a_{N_a}}+t)]_{t \in W}^T}_{1 \times N_W} \right]^T$$

όπου $u_k(\tau) = u(\vec{C}_k(\tau))$ είναι ο περιορισμός της $u(x, y)$ στην k -οστή γραμμή πλέγματος $\vec{C}_k(\tau)$ και $W = \{-d, -d+1, \dots, d\} \cdot \delta l$ είναι το παράθυρο δειγματοληψίας βήματος δl και $N_W = 2d+1$ δειγμάτων².

5.4.1.1 Διαφορές από τα Κλασικά AAM

Σε σύγκριση με τα συμβατικά AAM [35, 83], η προτεινόμενη αναπαράσταση περιέχει διάφορες τροποποιήσεις, έτσι ώστε να εκμεταλλευτούμε τις ιδιαιτερότητες της συγκεκριμένης εφαρμογής. Η εκ των προτέρων γνώση της πόζας του πλέγματος φωνητικής οδού σε κάθε πλαίσιο μας επιτρέπει να ελαττώσουμε την πολυπλοκότητα της αναπαράστασης και μοντελοποίησης της εμφάνισης της γλώσσας. Πρώτα από όλα, τα σημεία του σχήματος αναπαρίστανται από βαθμωτές τιμές s_k αντί 2Δ συντεταγμένων, το οποίο συνεπάγεται N_s αντί $2N_s$ μετρήσεις αντίστοιχα. Επίσης, δεν υπάρχει ανάγκη να λάβουμε ξεχωριστά υπόψη την γεωμετρική αντιστοίχιση των σχημάτων μέσω μετασχηματισμών ομοιότητας· χάρη στη χρήση του πλέγματος φωνητικής οδού, η μεταβολή που προκαλείται από την κίνηση του ουρανίσκου του ομιλητή δεν επηρεάζει το μοντέλο σχήματος. Επιπρόσθετα, το διάνυσμα υψής σχηματίζεται μέσω της σάρωσης των γραμμών πλέγματος γύρω από τα σημεία σχήματος αντί της σάρωσης ολόκληρων 2Δ περιοχών και της γεωμετρικής στρέβλωσης (warping) σε ένα μέσο σχήμα, επομένως οι διαστάσεις του διανύσματος υψής μειώνονται σημαντικά.

Χάρη στις παραπάνω απλοποιήσεις, το ταίριασμα AAM εμπλέκει απλούστερες αναλυτικές εκφράσεις και μπορεί να πραγματοποιηθεί λύνοντας ένα πιο ελαφρύ και πιο αξιόπιστο πρόβλημα βελτιστοποίησης από ό,τι στην κλασική προσέγγιση (βλ. Ενότητα 5.4.4). Αξίζει να σημειωθεί ότι η χρήση ενός πλέγματος που είναι σχεδιασμένο για ένα συγκεκριμένο παραμορφώσιμο σχήμα θα μπορούσε να υιοθετηθεί και σε άλλες εφαρμογές εκτός της ανίχνευσης της γλώσσας. Εάν οι παράμετροι πόζας ενός τέτοιου πλέγματος δεν είναι εκ των προτέρων γνωστές όπως εδώ, θα μπορούσαν να εκτιμηθούν μέσω της εισαγωγής τους στην διαδικασία βελτιστοποίησης του ταίριασματος AAM, ενσωματώνοντας επίσης περιορισμούς σχετικά με την δυναμική τους.

5.4.2 Μοντελοποίηση της Μεταβολής του Σχήματος

Για το μοντέλο σχήματος της γλώσσας, εκμεταλλευόμαστε τις γλωσσικές καμπύλες που έχουν εξαχθεί «με το χέρι» από N_x πλαίσια ακτίνων X του ίδιου ομιλητή κατά τη διάρκεια της ομιλίας. Χρησιμοποιώντας την πληροφορία σχετικά με την θέση του πλέγματος φωνητικής οδού σε κάθε πλαίσιο, εξάγονται N_x διανύσματα εκπαίδευσης του σχήματος. Αντίστοιχα με το [35], θεωρούμε ένα γραμμικό στατιστικό μοντέλο για το διάνυσμα σχήματος:

$$\mathbf{s} \approx \mathbf{s}_0 + \mathbf{Q}_s \mathbf{b}, \quad (5.3)$$

²Δεν υιοθετούμε καμία καθολική γραμμική απεικόνιση των φωτεινότητων της εικόνας, μιας και όλα τα πλαίσια υπερήχων καταγράφονται υπό παρόμοιες συνθήκες.

όπου \mathbf{s}_0 είναι ένα σχήμα βάσης, \mathbf{b} είναι το διάνυσμα που περιέχει τις N_b παραμέτρους σχήματος, με $N_b < N_s$, και \mathbf{Q}_s είναι μία $N_s \times N_b$ μήτρα της οποίας οι στήλες ορίζουν τα *ιδιοσχήματα* (eigenshapes), δηλαδή τους *τρόπους* (modes) της παραμόρφωσης του σχήματος.

Τα \mathbf{Q}_s και \mathbf{s}_0 μαθαίνονται στατιστικά από τα N_x διανύσματα εκπαίδευσης χρησιμοποιώντας *Ανάλυση Πρωτευουσών Συνιστωσών* (Principal Component Analysis - PCA): \mathbf{s}_0 είναι το μέσο σχήμα και οι στήλες του \mathbf{Q}_s περιέχουν τα ιδιοδιανύσματα που αντιστοιχούν στις N_b μεγαλύτερες ιδιοτιμές της μήτρας συμμεταβλητότητας του συνόλου εκπαίδευσης. Η κλιμάκωση κάθε ενός από αυτά τα ιδιοδιανύσματα επιλέγεται έτσι ώστε το \mathbf{b} να έχει μοναδιαία διακύμανση. Επομένως, υποθέτοντας ότι το \mathbf{b} ακολουθεί γκαουσιανή κατανομή, η βέλτιστη προσέγγιση της συνάρτησης πυκνότητας πιθανότητάς του (σπιπ) είναι $p(\mathbf{b}) = \mathcal{N}(\mathbf{b} | \mathbf{0}, \mathbf{I}_{N_b})$, όπου το \mathbf{I}_M συμβολίζει τον $M \times M$ μοναδιαίο πίνακα και το $\mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \Sigma)$ αναπαριστά την πολυμεταβλητή γκαουσιανή σπιπ του \mathbf{x} με μέσο $\boldsymbol{\mu}$ και μήτρα συμμεταβλητότητας Σ .

5.4.3 Μοντελοποίηση της Μεταβολής της Υφής

Εδώ χρησιμοποιούμε επισημειώσεις από N_{us} πλαίσια υπερήχων, στις οποίες μόνο τα μέρη των καμπυλών της γλώσσας που είναι ορατά έχουν σημειωθεί. Ταξινομούνται σαν γραμμές ενεργού υφής εκείνες οι γραμμές πλέγματος φωνητικής οδού που τέμνουν τα επισημειωμένα ορατά μέρη της γλώσσας σε πάνω από $\pi_{act} N_{us}$ πλαίσια εκπαίδευσης. Από την στιγμή που επιλέγουμε $\pi_{act} < 1$, σε κάποια πλαίσια εκπαίδευσης οι συντεταγμένες σχήματος για κάποιες από τις γραμμές πλέγματος ενεργού υφής μπορεί να λείπουν, λόγω του ότι η ορατότητα της γλώσσας σε αυτές τις γραμμές είναι είτε ανύπαρκτη είτε υπερβολικά χαμηλή για τον άνθρωπο. Σε αυτές τις περιπτώσεις, το διάνυσμα σχήματος επεκτείνεται στα άγνωστα μέρη μέσω εκτίμησης με το maximum a posteriori (MAP) κριτήριο των παραμέτρων σχήματος, χρησιμοποιώντας το εξαγμένο μοντέλο της μεταβολής σχήματος. Ύστερα από αυτό, τα N_{us} εκπαιδευτικά διανύσματα υφής μπορούν πλέον να σχηματιστούν. Αυτό το σύνολο εκπαίδευσης χωρίζεται σε 2 υποσύνολα, T_1 και T_2 . Σε αντιστοιχία με την προηγηθείσα μοντελοποίηση σχήματος, ένα γραμμικό μοντέλο χρησιμοποιείται για το διάνυσμα υφής:

$$\mathbf{g} = \mathbf{g}_0 + \mathbf{Q}_g \boldsymbol{\lambda} + \boldsymbol{\varepsilon}, \quad (5.4)$$

όπου $\boldsymbol{\lambda}$ είναι οι παράμετροι υφής διάστασης $N_\lambda < N_g$ και $\boldsymbol{\varepsilon}$ είναι το σφάλμα της ανακατασκευής του διανύσματος υφής \mathbf{g} μέσω του μοντέλου³. Το σύνολο εκπαίδευσης T_1 χρησιμοποιείται για την στατιστική εκμάθηση των \mathbf{g}_0 και \mathbf{Q}_g , κάνοντας χρήση μιας διαδικασίας ίδιας με αυτή που περιγράφηκε στην Ενότητα 5.4.2. Υποθέτοντας μία γκαουσιανή κατανομή για το $\boldsymbol{\lambda}$, έχουμε $p(\boldsymbol{\lambda}) = \mathcal{N}(\boldsymbol{\lambda} | \mathbf{0}, \mathbf{I}_{N_\lambda})$. Θεωρούμε επίσης ότι το $\boldsymbol{\varepsilon}$ ακολουθεί μία γκαουσιανή κατανομή μηδενικής μέσης τιμής με μήτρα συμμεταβλητότητας της μορφής $\Sigma_\varepsilon = \tilde{\mathbf{Q}}_g \text{diag}(\rho_1, \dots, \rho_{N_g}) \tilde{\mathbf{Q}}_g^T$, όπου οι στήλες της $\tilde{\mathbf{Q}}_g$ περιέχουν όλα τα ορθοκανονικά ιδιοδιανύσματα της μήτρας συμμεταβλητότητας της \mathbf{g} , όπως αυτά προκύπτουν από το σύνολο εκπαίδευσης T_1 . Επομένως, για το $\boldsymbol{\varepsilon}$ έχουμε ότι $p(\boldsymbol{\varepsilon}) = \mathcal{N}(\boldsymbol{\varepsilon} | \mathbf{0}, \Sigma_\varepsilon)$. Οι βέλτιστες παράμετροι $\rho_1, \dots, \rho_{N_g}$ μαθαίνονται στατιστικά χρησιμοποιώντας τα σφάλματα ανακατασκευής $\{\varepsilon_i\}$ του συνόλου εκπαίδευσης T_2 .

Σημειώστε ότι το σχήμα και η υφή της γλώσσας θεωρούνται στατιστικά ανεξάρτητα

³Στο μοντέλο σχήματος, δεν περιλαμβάνουμε τέτοιο σφάλμα διότι η γλωσσική καμπύλη δεν είναι άμεσα παρατηρήσιμη στα πλαίσια υπερήχων.

(ανεξάρτητα AAM στην ορολογία του [83]). Ο λόγος για αυτό είναι ότι στα πλαίσια υπερήχων δεν είναι ορατή ολόκληρη η καμπύλη της γλώσσας και άρα το μοντέλο σχήματος έπρεπε να εκπαιδευθεί από οπτικά δεδομένα (εικόνες ακτίνων X) διαφορετικού τύπου από αυτά που χρησιμοποιήθηκαν στο μοντέλο σχήματος.

5.4.4 Παρακολούθηση μέσω του Ταιριάσματος του Μοντέλου

Στο πλαίσιο εργασίας που περιγράφουμε, η εξαγωγή της καμπύλης της γλώσσας από τα πλαίσια υπερήχων μπορεί να επιτευχθεί μέσω του ταιριάσματος του μοντέλου εμφάνισης σε κάθε εικόνα. Για αυτό, χρησιμοποιείται εκτίμηση με κριτήριο MAP των \mathbf{b} και λ . Ο στόχος μας είναι η μεγιστοποίηση της ύστερης πιθανότητας:

$$p(\mathbf{b}, \lambda | u(x, y)) \propto p(u | \mathbf{b}, \lambda) p(\mathbf{b}, \lambda) = p(\boldsymbol{\varepsilon}) p(\mathbf{b}) p(\lambda)$$

όπου $u(x, y)$ είναι το φιλτραρισμένο πλαίσιο υπερήχων και $\boldsymbol{\varepsilon} = \mathbf{g}(s(\mathbf{b})) - \mathbf{g}_0 - \mathbf{Q}_g \lambda$ είναι το σφάλμα ανακατασκευής της υφής. Η παραπάνω μεγιστοποίηση είναι ισοδύναμη με την ελαχιστοποίηση της ακόλουθης ενέργειας:

$$E(\mathbf{b}, \lambda) = -\ln p(\mathbf{b}, \lambda | u) = C + \frac{1}{2} \{ \|\mathbf{b}\|^2 + \|\lambda\|^2 + \boldsymbol{\varepsilon}^T \Sigma_{\boldsymbol{\varepsilon}}^{-1} \boldsymbol{\varepsilon} \}$$

όπου C είναι μία σταθερά. Οι κλίσεις της $E(\mathbf{b}, \lambda)$ μπορούν εύκολα να εξαχθούν χρησιμοποιώντας τον κανόνα αλυσίδας:

$$\begin{aligned} \nabla_{\mathbf{b}} E &= \mathbf{b} + \mathbf{Q}_s^T (\partial \mathbf{g} / \partial \mathbf{s})^T \Sigma_{\boldsymbol{\varepsilon}}^{-1} \boldsymbol{\varepsilon} \\ \nabla_{\lambda} E &= \lambda - \mathbf{Q}_g^T \Sigma_{\boldsymbol{\varepsilon}}^{-1} \boldsymbol{\varepsilon} \end{aligned} \quad (5.5)$$

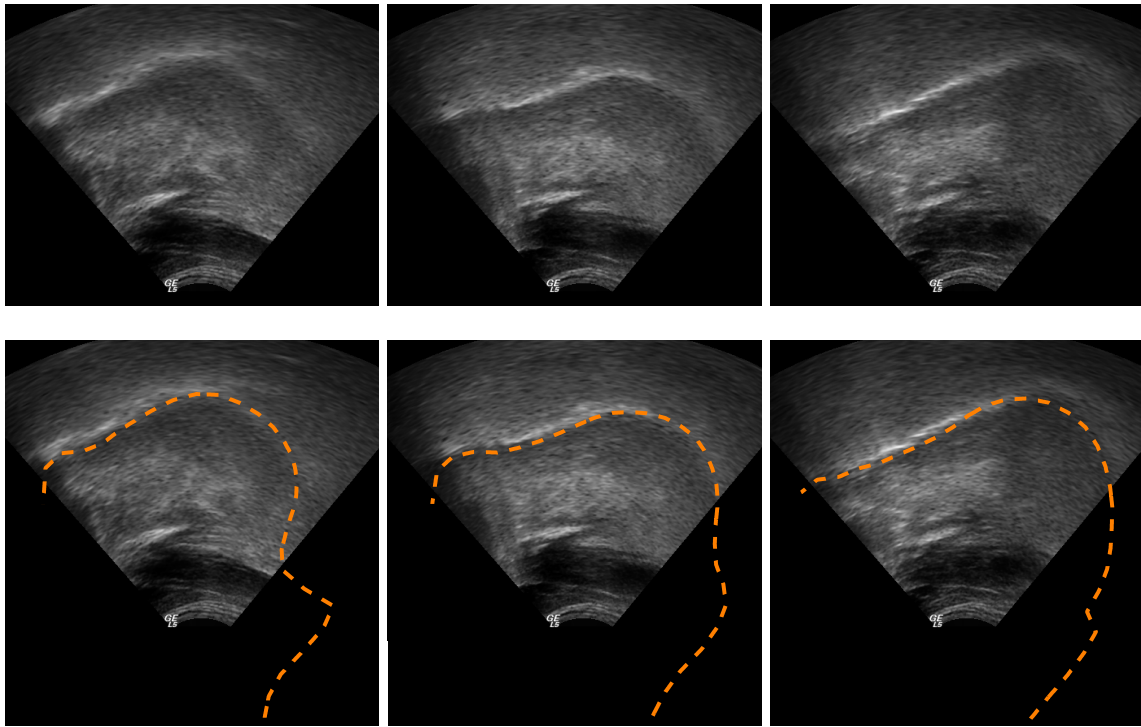
όπου η k -στη στήλη, $1 \leq k \leq N_s$, της ιακωβιανής $\partial \mathbf{g} / \partial \mathbf{s}$ είναι:

$$\frac{\partial \mathbf{g}}{\partial s_k} = \begin{cases} [0 \cdots \cdots 0]^T, & \text{εάν } k \notin G_{act} \\ \left[\underbrace{0 \cdots \cdots 0}_{(k-1)N_W} [u'_k(s_k + t)]_{t \in W}^T \underbrace{0 \cdots \cdots 0}_{(N_s - k)N_W} \right]^T, & \text{εάν } k \in G_{act} \end{cases}$$

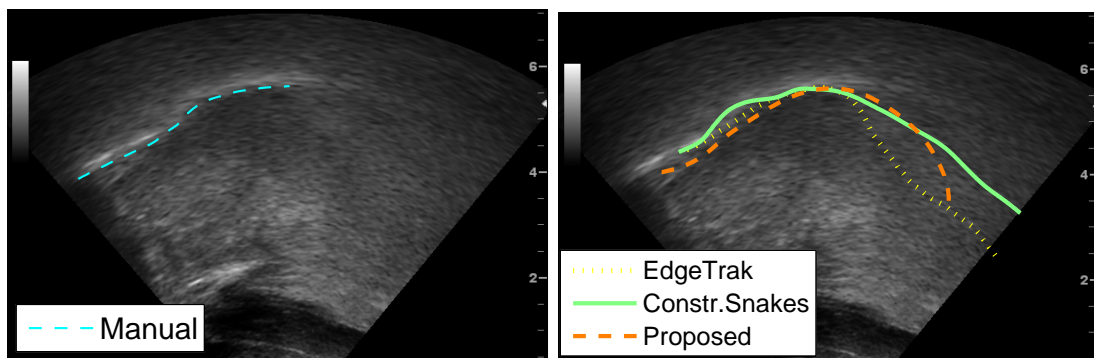
Χρησιμοποιούμε μία απλή κάθοδο κλίσης για την ελαχιστοποίηση της E , βασισμένοι στην Εξ. (5.5). Σε κάθε πλαίσιο, αρχικοποιούμε το \mathbf{b} από το αποτέλεσμα του προηγούμενου πλαισίου και το λ από το αποτέλεσμα της μεγιστοποίησης της ύστερης πιθανότητας $p(\lambda | \mathbf{g}(s(\mathbf{b}_0)))$. Ας σημειωθεί όμως ότι πιο αποδοτικές μέθοδοι ελαχιστοποίησης μπορούν εύκολα να εφαρμοστούν, μιας και οι μερικές παράγωγοι 2ης τάξης της E έχουν και αυτές σχετικά απλές αναλυτικές μορφές.

5.5 Πειραματικά Αποτελέσματα

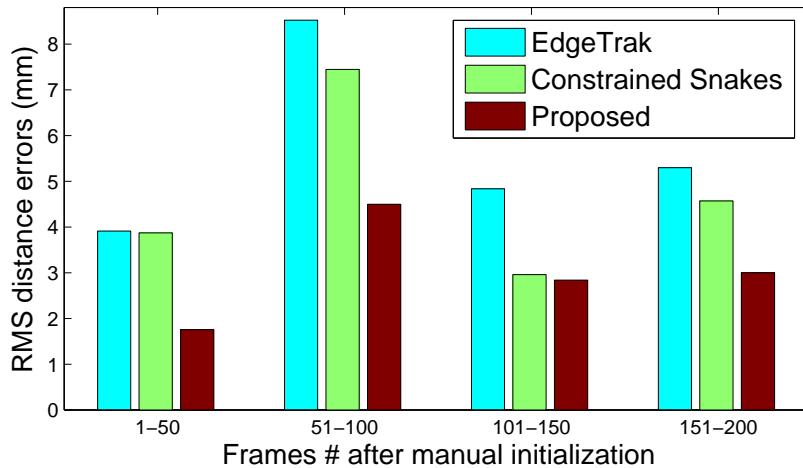
Στα πειράματα που διεξήχθησαν, η διάσταση του διανύσματος σχήματος είναι $N_s=30$ και το μοντέλο σχήματος εκπαιδεύτηκε χρησιμοποιώντας $N_x=700$ πλαίσια ακτίνων X και $N_b=6$ παραμέτρους, έτσι ώστε το μοντέλο να εξηγεί το 96% της μεταβλητότητας του διανύσματος σχήματος. Επίσης, η διάσταση του διανύσματος υφής είναι $N_g=1215$ και στην εκπαίδευση του μοντέλου υφής, χρησιμοποιήσαμε $N_{us}=400$ πλαίσια υπερήχων και θέσαμε $\pi_{act}=50\%$ και $N_{\lambda}=35$, το οποίο εξηγεί το 93% της μεταβλητότητας των διανυσμάτων υφής.



Σχήμα 5.5: Εφαρμογή της προτεινόμενης μεθόδου για την ανίχνευση της γλώσσας και επεκτατική εκτίμηση όλου του κατώτερου συνόρου της φωνητικής οδού σε μία ακολουθία εικόνων υπερήχων. *Πρώτη σειρά:* Τρία πλαίσια της ακολουθίας εισόδου. *Δεύτερη σειρά:* Εκτίμηση των αντίστοιχων καμπυλών.



Σχήμα 5.6: Πλαίσιο από μία ακολουθία όπου οι μέθοδοι ανίχνευσης γλώσσας έχουν εφαρμοστεί. *Αριστερά:* επισημειωμένη καμπύλη. *Δεξιά:* σύγκριση των αποτελεσμάτων των μεθόδων.



Σχήμα 5.7: Αξιολόγηση της προτεινόμενης μεθόδου σε σύγκριση με άλλες μεθόδους παρούσας επιστημονικής στάθμης για την ανίχνευση της γλώσσας ομιλητή. Επισημειωμένες καμπύλες της γλώσσας χρησιμοποιήθηκαν σαν αναφορά.

Το Σχ. 5.5 παρουσιάζει αποτελέσματα της προτεινόμενης μεθόδου παρακολούθησης. Παρατηρούμε ότι το ορατό μέρος της καμπύλης της γλώσσας έχει εντοπιστεί με ακρίβεια και έχει επεκταθεί με έναν αρκετά αληθοφανή τρόπο. Για λόγους αξιολόγησης, εφαρμόζουμε την προτεινόμενη μέθοδο αλλά και τις μεθόδους EdgeTrak [71] (χωρίς καθόλου διορθώσεις από τον χρήστη) και Περιορισμένου Μοντέλου Φιδιού (Constrained Snakes) [7] σε μία ακολουθία 200 πλαισίων υπερήχων για τα οποία είναι διαθέσιμες επισημειωμένες καμπύλες της γλώσσας. Στο Σχ. 5.6, φαίνεται ένα παράδειγμα από τα αποτελέσματα αυτά (για την προτεινόμενη μέθοδο, λαμβάνονται εδώ υπόψιν μόνο τα εντοπισμένα σημεία στις γραμμές πλέγματος ενεργούς υφής). Η μέθοδος EdgeTrak επεκτείνει εσφαλμένα την καμπύλη της γλώσσας ενώ το Περιορισμένο Μοντέλο Φιδιού παρουσιάζει βελτιωμένη απόδοση. Τέλος, η προτεινόμενη μέθοδος, η οποία είναι η μόνη που ενσωματώνει πρότερη πληροφορία σχήματος, αποδίδει το περισσότερο αληθοφάνες αποτέλεσμα.

Στο Σχ. 5.7, παρουσιάζονται ποσοτικές μετρήσεις, χρησιμοποιώντας την συμμετρική RMS απόσταση σφάλματος $e_d = \sqrt{(d_{om}^2 + d_{mo}^2)/2}$, όπου d_{om} (d_{mo}) είναι η RMS απόσταση των σημείων της ανιχνευμένης (επισημειωμένης) από την επισημειωμένη (ανιχνευμένη) καμπύλη. Βλέπουμε ότι η προτεινόμενη μέθοδος επιφέρει καλύτερες ποσοτικές μετρήσεις. Τα παραπάνω αποτελέσματα επιδεικνύουν την δυναμική αυτής της νέας προσέγγισης στην παρακολούθηση της γλώσσας.

Κεφάλαιο 6

Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας

6.1 Εισαγωγή

Ο εντοπισμός και η παρακολούθηση των χεριών του νοηματιστή καθώς και η εξαγωγή χαρακτηριστικών που να περιγράφουν αξιόπιστα την πόζα και την διαμόρφωσή τους είναι κρίσιμα για την τελική επιτυχία ενός συστήματος Αυτόματης Αναγνώρισης Νοηματικής Γλώσσας (AANΓ). Τα προβλήματα όμως αυτά είναι ιδιαίτερα απαιτητικά, κυρίως εξαιτίας της γρήγορης κίνησης των χεριών, των αλληλεπικαλύψεων καθώς και της μεγάλης μεταβλητότητας στην 3D πόζα και το σχήμα του χεριού.

Σε αυτό το κεφάλαιο, προτείνουμε μία δυναμική, αφινικά αναλλοίωτη μοντελοποίηση των χειρομορφών νοηματικής γλώσσας, η οποία προσφέρει μία συμπαγή και περιγραφική αναπαράσταση των διαμορφώσεων του χεριού. Επίσης, προσφέρει μία ομαλοποιημένη (regularized) διαδικασία ταιριάσματος του μοντέλου, η οποία παρακολουθεί το κυρίαρχο χέρι του νοηματιστή και ταυτόχρονα εξάγει χαρακτηριστικά της διαμόρφωσης του χεριού. Χάρη στην ενσωμάτωση στατικής και δυναμικής πρότερης γνώσης σχετικά με το χέρι, η διαδικασία αυτή είναι εύρωστη (robust) ακόμα και σε περιπτώσεις αλληλεπικαλύψεων με σχετικά μεγάλη διάρκεια. Επίσης, στο προτεινόμενο σύστημα, χρησιμοποιούμε πρωτότυπες τεχνικές κατάτμησης και παρακολούθησης των χεριών και του κεφαλιού, τεχνικές οι οποίες σχετίζονται με την πιθανοτική και μορφολογική προεπεξεργασία των βίντεο συνεχούς νοηματισμού. Οι συνεισφορές μας περιλαμβάνουν τα εξής:

- Με σκοπό να εκτιμήσουμε τις περιοχές δέρματος που μας ενδιαφέρουν (χέρια και κεφάλι), συνδυάζουμε ένα στοιχειώδες πιθανοτικό μοντέλο χρώματος δέρματος με καινοτόμους αλγορίθμους φιλτραρίσματος σχήματος τους οποίους σχεδιάσαμε βασισμένοι στην μαθηματική μορφολογία [57, 79, 109].
- Παρακολουθούμε τα χέρια και το κεφάλι χρησιμοποιώντας μία διαδικασία χρονικά αμφίδρομης πρόβλεψης (forward-backward prediction) και πρότερη στατιστική πληροφορία οργανωμένη σε μορφή απλών κανόνων.
- Για την αναπαράσταση των διαμορφώσεων του χεριού, υιοθετούμε εικόνες Σχήματος-Εμφάνισης. Κατασκευάζουμε το λεγόμενο Μοντέλο Σχήματος-Εμφάνισης (ΜΣΕ), στο οποίο οι εικόνες αυτές μοντελοποιούνται με έναν γραμμικό συνδυασμό

ιδιοεικόνων ακολουθούμενο από έναν αφινικό μετασχηματισμό, ο οποίος προσεγγίζει διάφορες αλλαγές στην 3Δ πόζα του χεριού και κάνει πιο συμπαγές το μοντέλο.

- Με σκοπό την ευρωστία απέναντι σε αλληλεπικαλύψεις, σχεδιάζουμε ένα ομαλοποιημένο ταίριασμα του ΜΣΕ, το οποίο αξιοποιεί πρότερη πληροφορία σχετικά με τη μορφή και τη δυναμική του χεριού. Η διαδικασία αυτή προσφέρει ένα ακριβές αποτέλεσμα παρακολούθησης καθώς και αποτελεσματικά χαρακτηριστικά χειρομορφών.
- Αξιολογούμε την εξαγωγή χαρακτηριστικών με βάση το προτεινόμενο ΜΣΕ σε εκτενή πειράματα ταξινόμησης χειρομορφών, στα οποία συγκρίνουμε τη μέθοδό μας με βασικές μεθόδους της βιβλιογραφίας. Η τουλάχιστον 10.5% απόλυτη βελτίωση που παρατηρείται στα πειράματα αυτά καταδεικνύει την αποτελεσματικότητα της προτεινόμενης μεθόδου για το πρόβλημα της ταξινόμησης χειρομορφών.

6.2 Αρχική Παρακολούθηση Χεριών και Κεφαλιού Βασισμένη σε Οπτική Προεπεξεργασία

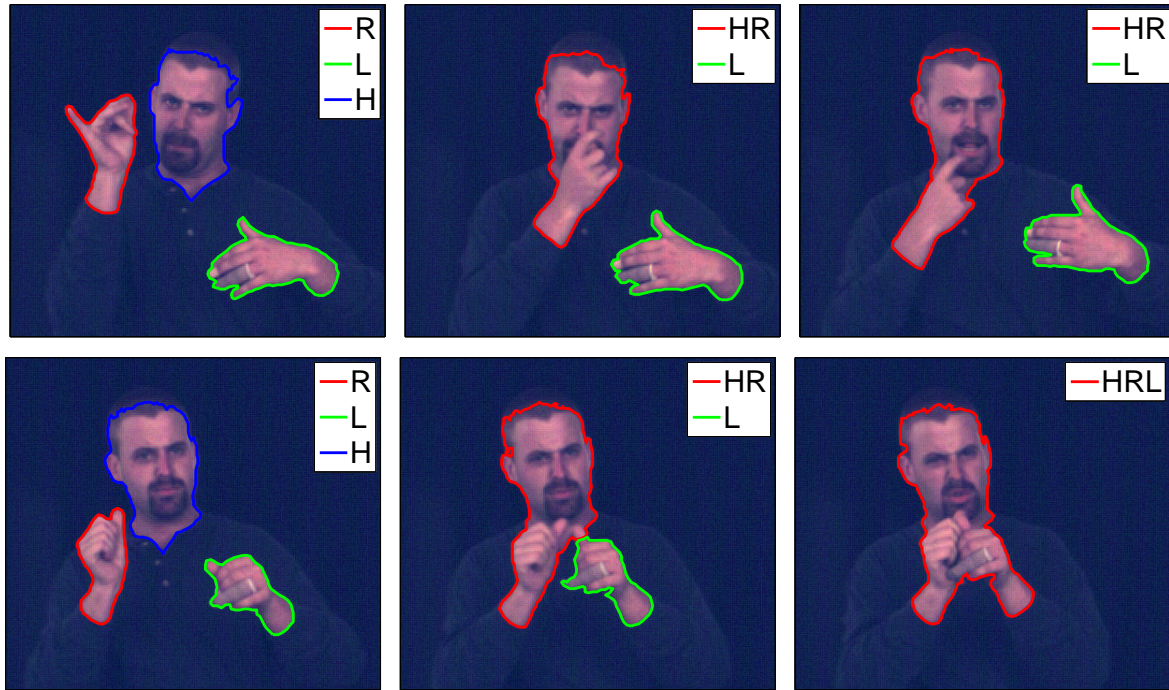
Σαν ένα αρχικό στάδιο επεξεργασίας, χρησιμοποιούμε το σύστημα που έχουμε περιγράψει στο [105], το οποίο όπως θα δούμε αποτελείται από 1) εξαγωγή μάσκας δέρματος βασισμένη στο χρώμα, 2) μορφολογικό φιλτράρισμα της μάσκας αυτής και 3) παρακολούθηση των χεριών και του κεφαλιού βασισμένη σε χρονικά αμφίδρομη πρόβλεψη (forward-backward prediction) και πρότερη στατιστική πληροφορία οργανωμένη σε μορφή απλών κανόνων. Στην τρέχουσα μορφή της εργασίας μας, θεωρούμε ότι η έξοδος αυτού του υποσυστήματος σε κάθε πλαίσιο είναι ένα σύνολο από μάσκες περιοχών δέρματος συνοδευόμενες από μία ή περισσότερες από τις ετικέτες H (head - κεφάλι), L (left hand - αριστερό χέρι) και R (right hand - δεξί χέρι), βλ. Σχ. .6.1.

Αναφερόμενοι για παράδειγμα στο *δεξί χέρι*, μπορούμε να διακρίνουμε δύο κύρια ενδεχόμενα:

1. σε κάποια πλαίσια, το υποσύστημα εξάγει μία μάσκα ειδικά για αυτό το χέρι (η μάσκα έχει ως μοναδική ετικέτα την R) και
2. σε διάφορα άλλα πλαίσια, υπάρχει μία μάσκα εξόδου που αντιστοιχεί στο δεξί χέρι καθώς και σε άλλες σωματικές περιοχές ενδιαφέροντος (δηλαδή η μάσκα έχει τις ετικέτες HR ή RL ή HRL).

Για την κατάσταση της δεύτερης περίπτωσης, χρησιμοποιούμε την ορολογία (*αλληλεπικάλυψη*) (occlusion) και λέμε επίσης ότι το δεξί χέρι είναι *καλυπτόμενο*.

Εξαιτίας των γρήγορων κινήσεων και της συχνής εμφάνισης αλληλεπικαλύψεων, το πρόβλημα εξαγωγής τέτοιων μασκών με σωστές ετικέτες είναι ιδιαίτερα δύσκολο. Παρόλα αυτά, το σύστημα που έχουμε αναπτύξει δημιουργεί τις περισσότερες φορές ακριβή αποτελέσματα. Η έξοδος αυτού του υποσυστήματος μπορεί να ιδωθεί σαν μία πρωταρχική παρακολούθηση των χεριών και του κεφαλιού. Όπως θα παρουσιαστεί αργότερα στην Ενότητα 6.3, η προτεινόμενη μεθοδολογία του Μοντέλου Σχήματος-Εμφάνισης βασίζεται στην παρακολούθηση αυτή για να διαμορφώσει ένα πιο ακριβές αποτέλεσμα παρακολούθησης και παράλληλα να εξάγει περιγραφικά χαρακτηριστικά χειρομορφής.



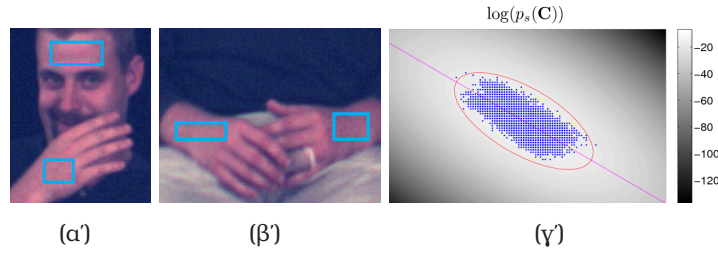
Σχήμα 6.1: Επισκόπηση της εξόδου του αρχικού υποσυστήματος παρακολούθησης των χεριών και του κεφαλιού. Ενδεικτικά πλαίσια με εξαγμένες μάσκες περιοχών δέρματος και ετικέτες σωματικών μερών H (head - κεφάλι), L (left hand - αριστερό χέρι) και R (right hand - δεξί χέρι).

6.2.1 Κατάτμηση και Εντοπισμός Δέρματος

6.2.1.1 Πιθανοτική Μοντελοποίηση Χρώματος Δέρματος

Μία προκαταρκτική εκτίμηση των θέσεων των χεριών και του κεφαλιού αποκομίζεται από το κριτήριο χρώματος (color cue), παρόμοια με πολλές άλλες εργασίες [5, 21, 24]. Πρέπει να υποθέσουμε εδώ ότι ο νοηματιστής φοράει ρούχα με μακριά μανίκια και τα χρώματα στο παρασκήνιο διαφέρουν από το χρώμα του δέρματος. Έτσι, χρησιμοποιούμε δεδομένα βίντεο νοηματικής γλώσσας που να πληρούν αυτούς τους περιορισμούς. Κατασκευάζουμε λοιπόν ένα απλό μοντέλο χρώματος δέρματος χρησιμοποιώντας την αναπαράσταση χρώματος στον χρωματικό χώρο $CIE-Lab$ και κρατώντας μόνο τις δύο συνιστώσες χρωματικότητας a^*, b^* . Με τον τρόπο αυτό, καταφέρνουμε κάποιον βαθμό ευρωστίας σε αλλαγές φωτισμού [63, 142].

Υποθέτουμε ότι οι τιμές χρωματικότητας $C(\vec{x}) = (a^*, b^*)$ των εικονοστοιχείων δέρματος ακολουθούν μία διμεταβλητή γκαουσιανή κατανομή. Η κατανομή αυτή μαθαίνεται χρησιμοποιώντας ένα σύνολο εκπαίδευσης από δείγματα χρώματος δέρματος, τα οποία προκύπτουν από επισημειωμένες από τον χρήστη περιοχές δέρματος σε πλαίσια βίντεο του συγκεκριμένου νοηματιστή, Σχ. 6.2. Μία πρώτη εκτίμηση της μάσκας S_0 με όλες τις περιοχές δέρματος προκύπτει λοιπόν από την εφαρμογή κατωφλίου στην $p_s(C(\vec{x}))$, Σχήματα 6.2(γ) και 6.3(β).



Σχήμα 6.2: Μοντελοποίηση χρώματος δέρματος. (α,β) Παραδείγματα επισημείωσης από τον χρήστη περιοχών δέρματος (ορθογώνια), οι οποίες προσφέρουν δείγματα εκπαίδευσης. (γ) Δείγματα εκπαίδευσης στον χώρο a^*-b^* και ταιριασμένη συνάρτηση πυκνότητας πιθανότητας $p_s(C)$. Η έλλειψη οριοθετεί τα χρώματα που ταξινομούνται ως δέρμα, σύμφωνα με την επιβολή κατωφλίου στην $p_s(C(\vec{x}))$. Η γραμμή αντιστοιχεί στην προβολή που ορίζει την απεικόνιση g που χρησιμοποιείται στον σχηματισμό των εικόνων Σχήματος-Εμφάνισης.

6.2.1.2 Μορφολογική Βελτίωση της Μάσκας Δέρματος

Σε κάθε πλαίσιο η εξαγμένη μάσκα δέρματος S_0 μπορεί να περιέχει κάποιες μικρές περιοχές λόγω θορύβου καθώς και κάποιες τρύπες στο εσωτερικό της περιοχής του κεφαλιού εξαιτίας των ματιών ή του ενδεχόμενου μουσιού του νοηματιστή. Για τους λόγους αυτούς, προτείνουμε ένα νέο μορφολογικό αλγόριθμο φιλτραρίσματος του συνόλου S_0 : Χρησιμοποιούμε την έννοια των *μορφολογικών τρυπών* $\mathcal{H}(S)$ μιας δυαδικής εικόνας, οι οποίες ορίζονται ως το σύνολο των συνιστωσών παρασκηνίου που δεν είναι συνδεδεμένες με το περίγραμμα της εικόνας [79, 113]. Με σκοπό να γεμίσουμε κάποιες περιοχές παρασκηνίου που είναι τρύπες ή γενικευμένες τρύπες, υπό την έννοια ότι συνδέονται με το περίγραμμα της εικόνας αλλά περνώντας από ένα στενό «κανάλι», εφαρμόζουμε το ακόλουθο *γενικευμένο* γέμισμα τρυπών, το οποίο αποδίδει μία βελτιωμένη εκτίμηση S_1 της μάσκας δέρματος:

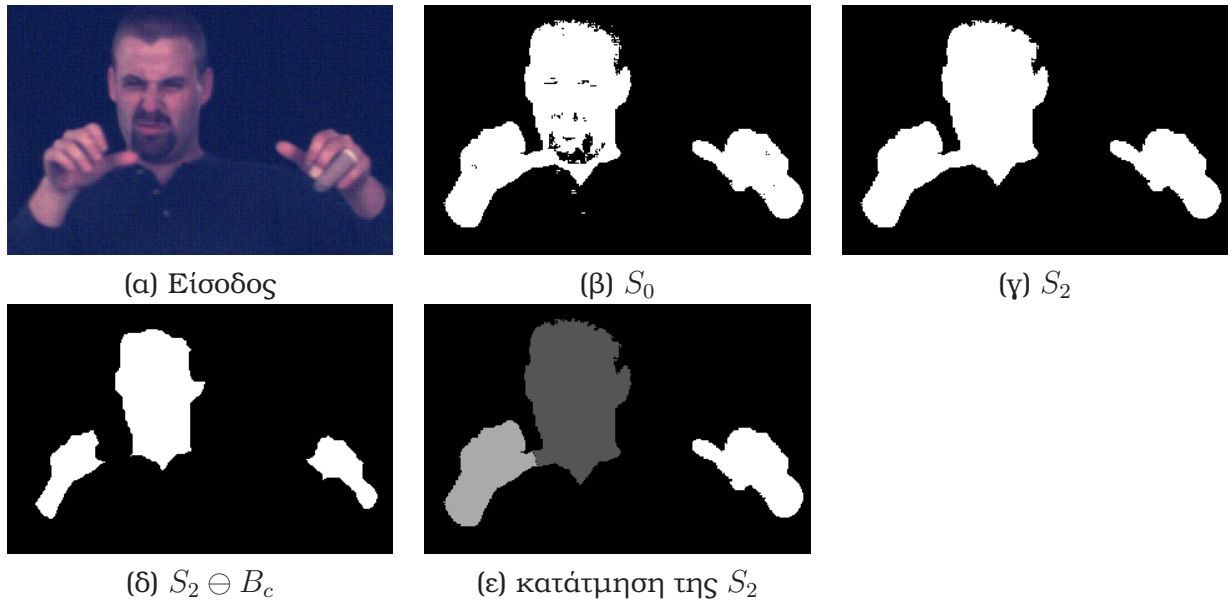
$$S_1 = S_0 \cup \mathcal{H}(S_0) \cup \{\mathcal{H}(S_0 \bullet B) \oplus B\} \quad (6.1)$$

όπου B είναι ένα δομικό στοιχείο (structuring element) σχετικά μικρού μεγέθους (π.χ. ένας δίσκος ακτίνας 3 εικονοστοιχείων) ενώ τα \oplus και \bullet συμβολίζουν Minkowski dilation και closing αντίστοιχα.

Έπειτα, με σκοπό την απομάκρυνση περιοχών που έχουν σχηματιστεί λόγω θορύβου ή αδιάφορης οπτικής πληροφορίας, αξιοποιούμε πρότερη γνώση. Οι **συνεκτικές συνιστώσες (ΣΣ)** των περιοχών δέρματος ενδιαφέροντος 1) μπορεί να είναι το πολύ τρεις, σχηματιζόμενες από τα χέρια και το κεφάλι, και 2) δεν μπορούν να έχουν εμβαδόν μικρότερο από ένα κατώφλι A_{min} . Επομένως, εφαρμόζουμε μορφολογικό area opening [128] με μία μεταβαλλόμενη τιμή κατωφλίου: Βρίσκουμε όλες τις ΣΣ της S_1 , υπολογίζουμε τα εμβαδά τους και τελικά απορρίπτουμε όλες τις συνιστώσες των οποίων το εμβαδόν δεν ανήκει στα 3 μεγαλύτερα εμβαδά ή είναι μικρότερο από A_{min} . Αυτή η διαδικασία δίνει ως αποτέλεσμα την τελική εκτίμηση S_2 της μάσκας δέρματος, Σχ. 6.3(γ).

6.2.1.3 Μορφολογική Κατάτμηση της Μάσκας Δέρματος

Στα πλαίσια όπου η μάσκα δέρματος S_2 περιέχει 3 ΣΣ, αυτές οι συνιστώσες αποδίδουν κατευθείαν μία επαρκή κατάτμηση της μάσκας. Οι περιοχές δέρματος ενδέχεται όμως να αλληλεπικαλύπτονται, το οποίο κάνει την μάσκα S_2 να έχει λιγότερες από 3 ΣΣ. Παρόλα



Σχήμα 6.3: Ενδεικτικά αποτελέσματα του συστήματος εξαγωγής μάσκας δέρματος και μορφολογικής κατάτμησής της.

αυτά, σε αρκετές τέτοιες περιπτώσεις, οι επικαλύψεις μεταξύ περιοχών του δέρματος δεν είναι ουσιώδεις: Διαφορετικές περιοχές μπορεί να συνδέονται μέσω στενών «γεφυρών», Σχ. 6.3(γ), για παράδειγμα όταν ένα δάκτυλο από το ένα χέρι αγγίζει ή βρίσκεται οριακά μπροστά από το άλλο χέρι ή το κεφάλι. Επομένως, μπορούμε να μειώσουμε το σύνολο των πλαισίων με αλληλεπικαλύψεις κατατέμνοντας περαιτέρω κάποιες περιοχές βασισμένοι σε μορφολογικούς τελεστές ως εξής:

Αν η S_2 περιέχει N_{cc} συνεκτικές συνιστώσες και $N_{cc} < 3$, βρίσκουμε τις $\Sigma\Sigma$ του συνόλου $S_2 \ominus B_c$ (π.χ. Σχ. 6.3(δ)) για ένα δομικό στοιχείο B_c μικρού μεγέθους (π.χ. ένας δίσκος ακτίνας 3 εικονοστοιχείων) και απορρίπτουμε αυτές τις $\Sigma\Sigma$ που έχουν εμβαδόν (ύστερα από ένα dilation με το B_c) μικρότερο από A_{min} . Εάν οι $\Sigma\Sigma$ που έμειναν είναι περισσότερες από N_{cc} , αυτό συνεπάγεται την απουσία κάποιας λεπτής σύνδεσης, επομένως η διαδικασία αυτή δεν επιφέρει κανέναν διαχωρισμό επικαλύψεων. Διαφορετικά όμως, χρησιμοποιούμε κάθε μία από αυτές τις $\Sigma\Sigma$ σαν τον σπόρο ενός διαφορετικού τμήματος και επεκτείνουμε τους σπόρους μέχρι να καλύψουν όλη την περιοχή του S_2 . Για τον σκοπό αυτό, προτείνουμε μία διαδικασία που ονομάσαμε *ανταγωνιστικό opening ανακατασκευής* (competitive reconstruction opening), βλ. Σχ. 6.3(ε), το οποίο είναι το αποτέλεσμα ενός επαναληπτικού αλγορίθμου, κατά τον οποίο σε κάθε βήμα 1) επεκτείνουμε κάθε εξελισσόμενο τμήμα εφαρμόζοντάς του το υπό συνθήκη dilation με έναν 3×3 σταυρό, με αναφορά τη μάσκα S_2 , 2) προσδιορίζουμε τα εικονοστοιχεία που ανήκουν σε περισσότερα του ενός τμήματα και τα αφαιρούμε από όλα τα τμήματα. Αυτό σημαίνει ότι τα τμήματα επεκτείνονται μέσα στην S_2 αλλά η επέκτασή τους σταματά όταν και όπου συναντήσουν άλλα τμήματα. Η διαδικασία αυτή συγκλίνει, μιας και μετά από κάποιες επαναλήψεις όλα τα τμήματα παραμένουν αμετάβλητα.

6.2.2 Παρακολούθηση και Διαχείριση Αλληλεπικαλύψεων

Μετά την κατάτμηση της μάσκας δέρματος S_2 , εφαρμόζουμε την πρωταρχική παρακολούθηση των χεριών και κεφαλιού, η οποία σε κάθε πλαίσιο περιλαμβάνει 1) την απονομή μίας ή περισσότερων ετικετών μερών σώματος H, R, L σε κάθε τμήμα της S_2 και 2) την εκτίμηση ελλείψεων σε τμήματα με πολλαπλές ετικέτες (δηλ. επικαλυπτόμενα τμήματα). Οι ελλείψεις αυτές προσφέρουν μία πρόχειρη εκτίμηση των σχημάτων των μερών σώματος που υπεισέρχονται στην επικάλυψη και δεν θεωρούνται έξοδοι αυτού του συστήματος πρωταρχικής παρακολούθησης. Συνεισφέρουν όμως στην σωστή απονομή ετικετών μετά το τέλος κάθε επικάλυψης.

Διακρίνουμε λοιπόν δύο περιπτώσεις: η κατάτμηση της S_2 επέφερε α) 3 τμήματα και β) 1 ή 2 τμήματα. Η πρώτη περίπτωση αντιστοιχεί σε *μη-επικάλυψη* ενώ η δεύτερη σε *επικάλυψη* μεταξύ μερών σώματος.

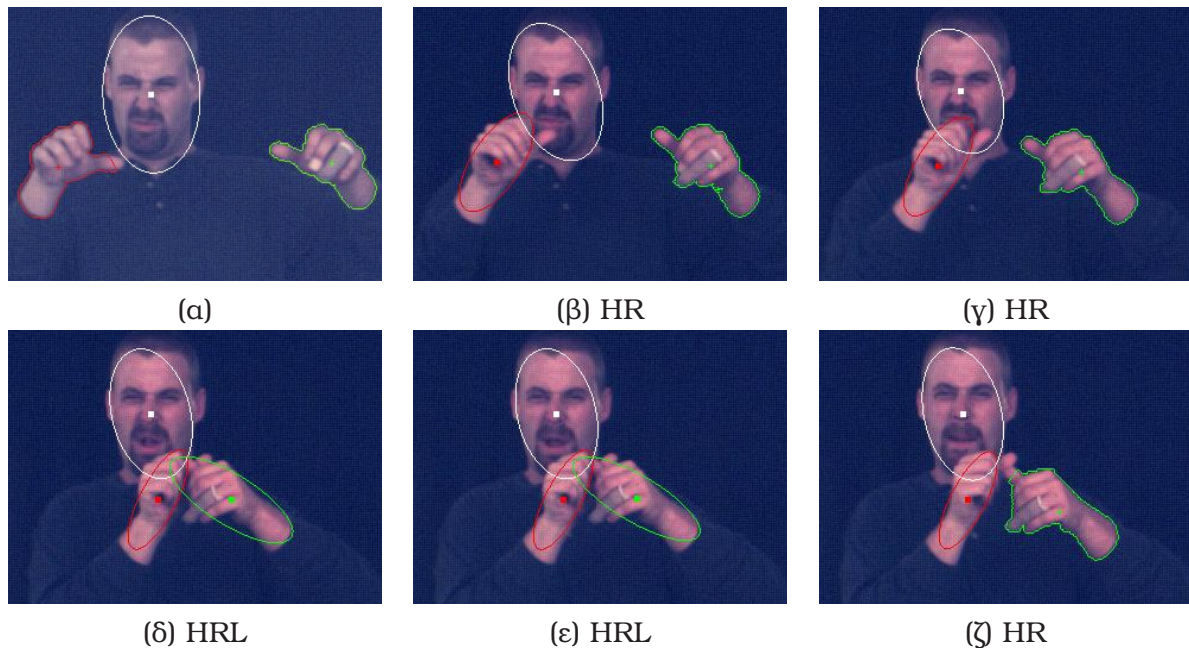
Περίπτωση μη-επικάλυψης. Το τμήμα με το μεγαλύτερο εμβαδόν λαμβάνει την ετικέτα H (κεφάλι) υποθέτοντας ότι το εμβαδόν του κεφαλιού είναι πάντα μεγαλύτερο από τα εμβαδά των χεριών. Για τις ετικέτες των χεριών, δεδομένου ότι έχουν ήδη απονεμηθεί σε προηγούμενα πλαίσια, υιοθετούμε μία γραμμική πρόβλεψη της θέσης του κεντροειδούς λαμβάνοντας υπόψιν τα 3 προηγούμενα πλαίσια (απλό μοντέλο σταθερής επιτάχυνσης). Έπειτα, απονέμουμε τις ετικέτες βασισμένοι στις αποστάσεις μεταξύ των προβλεφθεισών θέσεων και των θέσεων των κεντροειδών των τμημάτων. Ταιριάζουμε επίσης μία έλλειψη στο δυαδικό σχήμα κάθε τμήματος [5].

Περίπτωση επικάλυψης. Αξιοποιώντας τις παραμέτρους των ελλείψεων των σωματικών μερών που έχουν ήδη υπολογιστεί στα 3 προηγούμενα πλαίσια, υιοθετούμε και εδώ παρόμοια γραμμική πρόβλεψη για όλες τις παραμέτρους των ελλείψεων για το τρέχον πλαίσιο. Λαμβάνουμε επίσης μία βοηθητική εκτίμηση του κεντροειδούς του κάθε μέρους του σώματος μέσω ταιριάσματος τεμπλέτας (template matching) της αντίστοιχης περιοχής της εικόνας μεταξύ διαδοχικών πλαισίων. Έπειτα, επαναλαμβάνουμε τις προβλέψεις και τα ταιριάσματα τεμπλέτας με χρονική κατεύθυνση προς τα πίσω. Οι εκτιμήσεις με κατεύθυνση προς τα εμπρός και προς τα πίσω συνδυάζονται οδηγώντας σε μία τελική εκτίμηση των παραμέτρων των ελλείψεων των χεριών και του κεφαλιού του νοηματοιστή.

Τα Σχήματα 6.1 και 6.4 δείχνουν το αποτέλεσμα παρακολούθησης σε μία ακολουθία πλαισίων με περιπτώσεις μη-επικάλυψης και επικάλυψης. Παρατηρούμε ότι το υποσύστημα αυτό αποδίδει ένα σχετικά ακριβές αποτέλεσμα ακόμα και σε περιπτώσεις επικαλύψεων.

6.2.3 Στατιστικός Προσδιορισμός των Παραμέτρων.

Το παραπάνω σύστημα επεξεργασίας των βίντεο εμπλέκει διάφορες παραμέτρους. Οι περισσότερες από αυτές προσδιορίζονται αυτόματα μέσω μιας προκαταρκτικής επεξεργασίας κάποιων πλαισίων από τα βίντεο του συγκεκριμένου νοηματοιστή, κατά την οποία θεωρούμε μόνο πλαίσια χωρίς επικαλύψεις και υπολογίζουμε τα ακόλουθα στατιστικά. Υιοθετώντας γκαουσιανά μοντέλα, εκπαιδεύουμε τις συναρτήσεις πυκνότητας πιθανότητας p_H, p_{RL} των εμβαδών του κεφαλιού και του αριστερού/δεξιού χεριού του νοηματοιστή αντίστοιχα. Υπολογίζουμε επίσης την μέγιστη μετατόπιση d_{max} του κεντροειδούς του χεριού μεταξύ δύο διαδοχικών πλαισίων καθώς και το ελάχιστο εμβαδόν A_{min} του χεριού.



Σχήμα 6.4: Αρχική παρακολούθηση των χεριών και του κεφαλιού σε μια ακολουθία πλαισίων στην οποία συμβαίνουν επικαλύψεις (β-ζ), μεταξύ κεφαλιού (H, λευκή καμπύλη), δεξιού χεριού (R, κόκκινη καμπύλη) και αριστερού χεριού (L, πράσινη καμπύλη). Αναγράφεται κάθε φορά η ετικέτα της περιοχής δέρματος που περιέχει το δεξί (κυρίαρχο) χέρι. Οι εκτιμηθείσες ελλείψεις δεν θεωρούνται έξοδοι, αλλά είναι απαραίτητες για την σωστή απονομή ετικετών μετά το τέλος της επικάλυψης.

6.3 Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος-Εμφάνισης

Στην παρούσα Ενότητα, προτείνουμε την μεθοδολογία της δυναμικής, αφινικά αναλλοίωτης μοντελοποίησης των χειρομορφών νοηματικής γλώσσας, η οποία προσφέρει μία περιγραφική αναπαράσταση των διαμορφώσεων του χεριού. Επίσης προσφέρει μία διαδικασία παρακολούθησης του χεριού και ταυτόχρονης εξαγωγής χαρακτηριστικών χειρομορφής, η οποία είναι εύρωστη και αρκετά ακριβής ακόμα και σε περιπτώσεις αλληλεπικαλύψεων χεριών/κεφαλιού με μεγάλη διάρκεια. Η μεθοδολογία αυτή εφαρμόζεται εδώ στο *κυρίαρχο χέρι*¹ του νοηματιστή, αλλά μπορεί άμεσα να εφαρμοστεί και στο άλλο του χέρι ή στο κεφάλι του. Επιπρόσθετα, θα μπορούσε να χρησιμοποιηθεί για να μοντελοποιήσει και να παρακολουθήσει και άλλα παραμορφώσιμα και γρήγορα κινούμενα αντικείμενα, για άλλες εφαρμογές.

6.3.1 Αναπαράσταση Χεριών με Εικόνες Σχήματος-Εμφάνισης

Στόχος μας είναι να αναπαραστήσουμε όλες τις δυνατές διαμορφώσεις του κυρίαρχου χεριού κατά τη διάρκεια του νοηματισμού, χρησιμοποιώντας τις 2D εικόνες που λαμβά-

¹Για έναν δεξιόχειρα, το *κυρίαρχο* (dominant) χέρι είναι το δεξί, ενώ για έναν αριστερόχειρα το αριστερό. Στην τρέχουσα εκδοχή του συστήματός μας, ο προσδιορισμός του ποιο από τα δύο χέρια είναι το κυρίαρχο για κάθε νοηματιστή γίνεται χειροκίνητα.

νονται από την κάμερα. Αυτές οι εικόνες παρουσιάζουν μεγάλη ποικιλία εξαιτίας των διακυμάνσεων στην διαμόρφωση και την 3D πόζα του χεριού. Επιπρόσθετα, το σύνολο των επιφανειακών σημείων του χεριού που είναι ορατά από την κάμερα μεταβάλλεται σημαντικά. Για τους λόγους αυτούς, είναι πιο αξιόπιστο για την παρούσα εφαρμογή να αναπαριστούμε το 2D σχήμα του χεριού χωρίς την χρησιμοποίηση ενδεικτικών σημείων (landmark points), όπως γίνεται σε άλλες εργασίες της βιβλιογραφίας, π.χ. [36,83]. Έτσι, για την αναπαράσταση της χειρομορφής, χρησιμοποιούμε έμμεσα την δυαδική της μάσκα M , η οποία έχει εξαχθεί κατά την διάρκεια των σταδίων της κατάτμησης και ιχνηλάτησης των χεριών. Παράλληλα, συνδυάζουμε και την *εμφάνιση* του χεριού, δηλαδή τις έγχρωμες τιμές στο εσωτερικό της μάσκας M . Αυτές οι τιμές εξαρτώνται από την υφή του χεριού και την σκίαση, επομένως προσφέρουν σημαντική πληροφορία σχετικά με το 3D σχήμα του χεριού.

Πιο συγκεκριμένα, κόβουμε ένα κομμάτι $I(\mathbf{x})$ του τρέχοντος έγχρωμου frame γύρω από την μάσκα M και ύστερα σχηματίζουμε την ακόλουθη *Εικόνα Σχήματος-Εμφάνισης (ΣΕ)*:

$$f(\mathbf{x}) = \begin{cases} g(I(\mathbf{x})), & \text{εάν } \mathbf{x} \in M \\ -c_b, & \text{διαφορετικά} \end{cases} \quad (6.2)$$

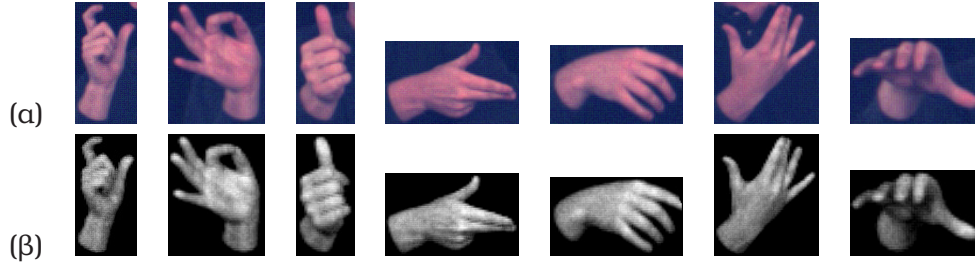
Η συνάρτηση $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ απεικονίζει τις έγχρωμες τιμές των εικονοστοιχείων δέρματος σε τιμές που είναι πιο κατάλληλες για την αναπαράσταση της εμφάνισης του χεριού. Προϋποθέτουμε ότι αυτή η συνάρτηση είναι κανονικοποιημένη έτσι ώστε η τυχαία μεταβλητή $g(C_s)$, όπου C_s είναι το τυχαίο διάνυσμα του χρώματος του δέρματος του συγκεκριμένου νοσηματιστή, να έχει μηδενική μέση τιμή και μοναδιαία διακύμανση. Η σταθερά $c_b > 1$ είναι μια σταθερά παρασκηνίου η οποία ελέγχει την ισορροπία μεταξύ σχήματος και εμφάνισης και η οποία είναι μία βασική παράμετρος του συστήματος αναπαράστασης και μοντελοποίησης χεριών που αναπτύξαμε. Καθώς η c_b γίνεται μεγαλύτερη, η διακύμανση της εμφάνισης λαμβάνει μικρότερο βάρος. Στην ακραία περίπτωση όπου $c_b \rightarrow \infty$, η εικόνα Σχήματος-Εμφάνισης είναι ισοδύναμη με μια δυαδική εικόνα και μόνο η πληροφορία σχήματος λαμβάνεται υπόψιν.

Στη συνέχεια, κατασκευάζουμε την συνάρτηση $g(C_s)$ ακολουθώντας τα εξής βήματα: Πρώτα, μετασχηματίζουμε κάθε έγχρωμη τιμή C_s στον χώρο χρώματος *CIE-Lab*, έπειτα κρατάμε μόνο τις συνιστώσες χρώματος a^*, b^* με στόχο την ανεξαρτησία από τον φωτισμό και στο τέλος παράγουμε σαν έξοδο έναν γραμμικό συνδυασμό αυτών των συνιστωσών. Τα βάρη αυτού του συνδυασμού αντιστοιχούν στην προβολή του σημείου (a^*, b^*) στον πρωτεύοντα άξονα της διμεταβλητής γκαουσιανής που μοντελοποιεί την πιθανότητα χρώματος δέρματος (βλ. Ενότητα 6.2.1.1 και Σχήμα 6.2(γ)). Στο Σχήμα 6.5, παρουσιάζονται μερικά παραδείγματα από την κατασκευή των εικόνων Σχήματος-Εμφάνισης του χεριού.

6.3.2 Μοντελοποίηση των Εικόνων Σχήματος-Εμφάνισης

Μοντελοποιούμε τις εικόνες Σχήματος-Εμφάνισης του χεριού, $f(\mathbf{x})$, με έναν γραμμικό συνδυασμό προκαθορισμένων εικόνων μεταβολής, ακολουθούμενο από έναν αφινικό μετασχηματισμό:

$$f(W_p(\mathbf{x})) \approx A_0(\mathbf{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\mathbf{x}), \quad \mathbf{x} \in \Omega_M \quad (6.3)$$



Σχήμα 6.5: (α) Κομμένες εικόνες του χεριού $I_k(\mathbf{x})$, για μερικά πλαίσια k τα οποία περιέχονται στο σύνολο των 200 δειγμάτων του συνόλου εκπαίδευσης του ΜΣΕ. (β) Αντίστοιχες εικόνες Σχήματος-Εμφάνισης $f_k(\mathbf{x})$.

όπου $A_0(\mathbf{x})$ είναι η εικόνα βάσης και $A_i(\mathbf{x})$ είναι οι N_c ιδιοεικόνες (eigenimages) που μοντελοποιούν την γραμμική μεταβολή. Αυτές οι εικόνες έχουν όλες ως πεδίο ορισμού το $\Omega_M \subset \mathbb{R}^2$. Επίσης, $\boldsymbol{\lambda} = (\lambda_1 \cdots \lambda_{N_c})$ είναι τα βάρη του γραμμικού συνδυασμού και W_p είναι ένας αφινικός μετασχηματισμός με παραμέτρους $\mathbf{p} = (p_1 \cdots p_6)$ που ορίζεται ως εξής:

$$W_p(x, y) = \begin{pmatrix} 1 + p_1 & p_3 & p_5 \\ p_2 & 1 + p_4 & p_6 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (6.4)$$

Το προτεινόμενο μοντέλο, το οποίο εφεξής θα αποκαλούμε **Μοντέλο Σχήματος-Εμφάνισης** (ΜΣΕ), είναι παρόμοιο με την γενική διατύπωση των Ενεργών Μοντέλων Εμφάνισης από τους Matthews και Baker [83], αλλά διαφέρει κυρίως στο ότι οι εικόνες είναι εικόνες Σχήματος-Εμφάνισης και ότι η στρέβλωση (warping) δεν ελέγχεται από τα χαρακτηριστικά σημεία σχήματος (shape landmarks) αλλά απλούστερα από τις 6 παραμέτρους του αφινικού μετασχηματισμού. Ο αφινικός μετασχηματισμός μπορεί να μοντελοποιήσει μετασχηματισμούς ομοιότητας της εικόνας καθώς επίσης και σχετικά μικρές αλλαγές στην 3D πόζα του χεριού. Έχει ένα μη γραμμικό αντίκτυπο στις εικόνες Σχήματος-Εμφάνισης και μειώνει την μεταβλητότητα των μετασχηματισμένων εικόνων του χεριού, $f(\mathbf{x})$, σε σύγκριση με άλλες μεθόδους που βασίζονται στην εμφάνιση, π.χ. [39], οι οποίες χρησιμοποιούν γραμμικά μοντέλα κατευθείαν στις αρχικές εικόνες. Ο γραμμικός συνδυασμός της εξίσωσης (6.3) μοντελοποιεί τις μεταβολές στην διαμόρφωση του χεριού και τις μεταβολές στον 3D προσανατολισμό του που δεν μπορούν να μοντελοποιηθούν από τον αφινικό μετασχηματισμό.

Ένα συγκεκριμένο μοντέλο εικόνων Σχήματος-Εμφάνισης του χεριού καθορίζεται από την εικόνα βάσης $A_0(\mathbf{x})$, τις ιδιοεικόνες $A_i(\mathbf{x})$ του γραμμικού συνδυασμού και τον αριθμό τους N_c . Όλες οι ποσότητες αυτές μαθαίνονται στατιστικά από δεδομένα εκπαίδευσης, όπως θα παρουσιαστεί στην ακόλουθη ενότητα. Τα διανύσματα \mathbf{p} και $\boldsymbol{\lambda}$ είναι οι παράμετροι του μοντέλου που ταιριάζουν το μοντέλο σε μια δεδομένη εικόνα ΣΕ. Λαμβάνουμε αυτές τις παραμέτρους σαν χαρακτηριστικά πόζας και σχήματος του χεριού αντίστοιχα.

Αξίζει εδώ να σημειωθεί ότι οι συνιστώσες του σχήματος και της εμφάνισης του χεριού συνδυάζονται επίσης σε εργασίες που χρησιμοποιούν το λεγόμενο *Ιστόγραμμα Κατευθυνόμενων Κλίσεων* (Histogram of Oriented Gradients - HOG), π.χ. [24, 73]. Σε αντίθεση με αυτές τις προσεγγίσεις όμως, η μοντελοποίηση που προτείνουμε προσφέρει έναν άμεσο έλεγχο της ισορροπίας μεταξύ σχήματος και εμφάνισης καθώς επίσης είναι προσαρμοσμένη στον συγκεκριμένο τύπο αντικειμένων λόγω του ότι είναι αναγεννητική. Επίσης,

διάφορες άλλες εργασίες, π.χ. [17, 21, 24, 140], χρησιμοποιούν την εμφάνιση του χεριού σε κομμένες εικόνες για την εξαγωγή χαρακτηριστικών, αφού προηγηθεί κάποια επεξεργασία. Μία σημαντική διαφορά της μεθόδου μας από τις εργασίες των [24, 140] έγκειται στο ότι χρησιμοποιούμε χαρακτηριστικά χειρομορφής λ τα οποία είναι αναλλοίωτα σε μετατόπιση, κλιμάκωση και περιστροφή στο επίπεδο της εικόνας. Αυτή η ιδιότητα ισχύει και για τις μεθόδους των [17, 21], αλλά στην μοντελοποίηση που προτείνουμε τα χαρακτηριστικά χειρομορφών μπορούν να θεωρηθούν αναλλοίωτα σε μία ευρύτερη τάξη αλλαγών στην 3D πόζα του χεριού: όλες οι αλλαγές που μπορούν να προσεγγισθούν ικανοποιητικά από έναν 2D αφινικό μετασχηματισμό επηρεάζουν στην πράξη μόνο τις παραμέτρους μετασχηματισμού p και όχι τις παραμέτρους λ .

6.3.3 Εκπαίδευση του Γραμμικού Συνδυασμού του ΜΣΕ

Για να εκπαιδύσουμε το μοντέλο των εικόνων Σχήματος-Εμφάνισης, χρησιμοποιούμε ένα αντιπροσωπευτικό σύνολο από εικόνες χειρομορφών από πλαίσια στα οποία το χέρι που μοντελοποιείται είναι πλήρως ορατό και χωρίς αλληλεπικαλύψεις με το άλλο χέρι ή το κεφάλι, Σχήμα 6.5(α). Επί του παρόντος, αυτό το σύνολο κατασκευάζεται μέσω τυχαίας δειγματοληψίας 500 τέτοιων εικόνων από πλαίσια ενός βίντεο. Με βάση αυτές τις διαλεγμένες εικόνες, το σύνολο εκπαίδευσης κατασκευάζεται από τις αντίστοιχες εικόνες ΣΕ $f_1 \cdots f_N$.

Με σκοπό την «αφαίρεση» της μεταβλητότητας του συνόλου που μπορεί να εξηγηθεί από το κομμάτι του αφινικού μετασχηματισμού του μοντέλου, εφαρμόζουμε μία ημιαυτόματη διαδικασία αφινικής ευθυγράμμισης του συνόλου εκπαίδευσης, Ενότητα 6.3.3.1. Έπειτα, οι εικόνες του γραμμικού συνδυασμού του μοντέλου μαθαίνονται εφαρμόζοντας PCA στο ευθυγραμμισμένο σύνολο, Ενότητα 6.3.3.2.

6.3.3.1 Αφινική Ευθυγράμμιση του Συνόλου Εκπαίδευσης

Σχεδιάσαμε ένα σύστημα αφινικής ευθυγράμμισης του συνόλου εκπαίδευσης ακολουθώντας μία προσέγγιση συστημικής σύνθεσης τύπου *άνωθεν* (bottom-up). Αυτό το σύστημα αποτελείται κυρίως από τα 3 επίπεδα που περιγράφονται στη συνέχεια.

6.3.3.1.1 Επίπεδο 1: 1-1 ευθυγράμμιση. Πρώτα αντιμετωπίζουμε το βασικό πρόβλημα της αφινικής ευθυγράμμισης μιας εικόνας Σχήματος-Εμφάνισης με μια άλλη. Για αυτό το πρόβλημα, χρησιμοποιούμε τον Αντίστροφο Συνθετικό (Inverse-Compositional - IC) Αλγόριθμο [55]. Αυτός ο αλγόριθμος βασίζεται στην ίδια διατύπωση του προβλήματος με τον κλασικό αλγόριθμο των Lucas-Kanade [74], αλλά βελτιώνει δραστικά τον υπολογιστικό χρόνο. Έστω λοιπόν ότι $f(x)$ είναι η μία εικόνα εισόδου και $g(x)$ η δεύτερη, η οποία αντιμετωπίζεται ως η εικόνα προτύπου που πρέπει να ευθυγραμμιστεί με την $f(x)$. Ο στόχος είναι να ελαχιστοποιηθεί το εξής άθροισμα τετραγώνων:

$$\sum_x \{g(W_p(x)) - f(x)\}^2 \quad (6.5)$$

ως προς τις παραμέτρους p του αφινικού μετασχηματισμού (στρέβλωσης) W_p (6.4). Ας σημειωθεί ότι το άθροισμα γίνεται σε όλα τα εικονοστοιχεία x του πεδίου ορισμού της

$g(\mathbf{x})$. Ο αλγόριθμος IC ελαχιστοποιεί την ενέργεια (6.5) μέσω της επαναληπτικής ελαχιστοποίησης του ακόλουθου αθροίσματος:

$$\sum_{\mathbf{x}} \{f(W_{\Delta p}(\mathbf{x})) - g(W_p(\mathbf{x}))\}^2 \quad (6.6)$$

ως προς το Δp . Αυτό γίνεται χρησιμοποιώντας μία προσέγγιση Taylor 1ης τάξης της ενέργειας (6.6) και λύνοντας το προκύπτον πρόβλημα ελαχίστων τετραγώνων. Στη συνέχεια, ο μετασχηματισμός ανανεώνεται ως ακολούθως:

$$W_p \leftarrow W_p \circ (W_{\Delta p})^{-1} \quad (6.7)$$

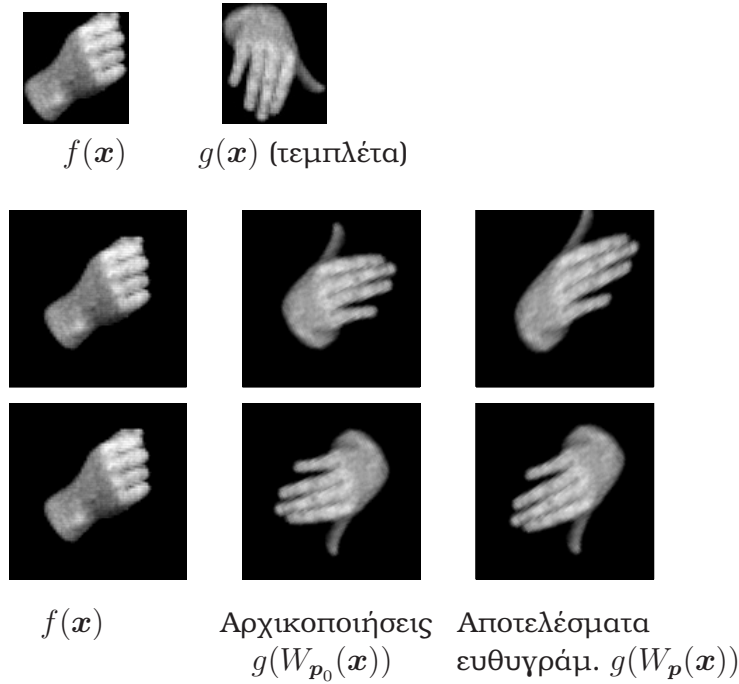
όπου το ‘ο’ συμβολίζει την σύνθεση δύο μετασχηματισμών.

Το αποτέλεσμα του αλγορίθμου IC εξαρτάται από την αρχικοποίηση του αφινικού μετασχηματισμού, μιας και ο αλγόριθμος συγκλίνει σε ένα τοπικό, αλλά όχι απαραίτητα ολικό βέλτιστο. Για τον λόγο αυτό, δίνουμε ιδιαίτερη προσοχή στον καθορισμό της αρχικοποίησης και, όταν χρειάζεται δοκιμάζουμε πολλαπλές αρχικοποιήσεις. Πιο συγκεκριμένα, υιοθετούμε δύο τύπους αρχικοποιήσεων (οι οποίοι χρησιμοποιούνται σε δύο διαφορετικές φάσεις του αλγορίθμου ευθυγράμμισης συνόλου του Επίπεδου 2):

- *Αρχικοποίηση από προηγούμενο αποτέλεσμα.* Χρησιμοποιούμε αυτό τον τύπο αρχικοποίησης όταν ένα «παρόμοιο» πρόβλημα 1-1 ευθυγράμμισης έχει ήδη λυθεί και θεωρούμε το σχετικό αποτέλεσμα αξιόπιστο. Το αποτέλεσμα αυτό χρησιμοποιείται σαν η μοναδική αρχικοποίηση.
- *Αρχικοποιήσεις βασισμένες σε γεωμετρικές μετρήσεις της δυαδικής μάσκας του χειριού.* Αυτός ο τύπος χρησιμοποιείται όταν καμία λύση σε παρόμοιο πρόβλημα δεν είναι διαθέσιμη. Στην περίπτωση αυτή, βασισμένοι στις μάσκες M των εικονοστοιχείων προσκηνίου των εικόνων ΣΕ, αρχικοποιούμε με τους δύο μετασχηματισμούς ομοιότητας οι οποίοι όταν εφαρμοστούν στην μία εικόνα ΣΕ κάνουν την μάσκα της να έχει κεντροειδές, εμβαδόν και προσανατολισμό ίδια με αυτά της μάσκας της άλλης εικόνας. Η ύπαρξη δύο τέτοιων μετασχηματισμών είναι λόγω της modulo- π ασάφειας του προσανατολισμού (βλ. Σχ. 6.6).

Τελικά, μπορούμε να διαλέξουμε την αρχικοποίηση που οδηγεί στην μικρότερη τιμή της αντικειμενικής συνάρτησης (6.5). Όμως, σε κάποιες περιπτώσεις, αυτή η επιλογή δεν είναι αληθοφανής. Για αυτό, ένας μικρός αριθμός σημείων αντιστοιχισμένων μεταξύ των δύο εικόνων ενδέχεται να εισαχθεί από τον χρήστη (βλ. Επίπεδο 3). Σε αυτή την περίπτωση, επιλέγεται το αποτέλεσμα που επιφέρει το μικρότερο άθροισμα τετραγώνων αποστάσεων μεταξύ των αντιστοιχισμένων σημείων.

6.3.3.1.2 Επίπεδο 2: Ευθυγράμμιση του συνόλου εκπαίδευσης. Έχοντας εξοπλιστεί με την προαναφερθείσα μέθοδο αφινικής ευθυγράμμισης δύο εικόνων, ευθυγραμμίζουμε ολόκληρο το σύνολο εκπαίδευσης γενικεύοντας την *προκρούστεια ανάλυση* (procrustes analysis) [36, 47]. Η γενίκευση έγκειται στο ότι θεωρούμε αφινικούς μετασχηματισμούς αντί για μετασχηματισμούς ομοιότητας. Επιπρόσθετα, στην περίπτωσή μας οι μετασχηματισμοί εφαρμόζονται σε εικόνες, σε αντίθεση με το [36] όπου οι μετασχηματισμοί εφαρμόζονται σε σύνολα χαρακτηριστικών σημείων που αναπαριστούν τα σχήματα.



Σχήμα 6.6: 1-1 αφινική ευθυγράμμιση δύο εικόνων ΣΕ $f(x)$ και $g(x)$. Το αντίστοιχο πρόβλημα βελτιστοποίησης λύνεται δύο φορές, χρησιμοποιώντας τις δύο αρχικοποιήσεις που κάνουν την μάσκα χεριού της $g(x)$ να έχει κεντροειδές, εμβαδόν και προσανατολισμό ίδια με αυτά της μάσκας χεριού της $f(x)$ (Πάνω σειρά) Εικόνες ΣΕ εισόδου. (Μεσαία σειρά) Αφινική ευθυγράμμιση για την 1^η αρχικοποίηση. (Κάτω σειρά) Αφινική ευθυγράμμιση για την 2^η αρχικοποίηση, η οποία προκύπτει από την 1^η μετά από περιστροφή 180°.

Ο αλγόριθμος που σχεδιάσαμε λαμβάνει σαν εισόδους ένα σύνολο από εικόνες εκπαίδευσης (με μία εξ αυτών σημειωμένη σαν εικόνα αναφοράς f_{ref}) και πιθανόν ένα σύνολο από επισημειώσεις αντιστοιχισμένων σημείων για κάποιες από τις εικόνες του συνόλου εκπαίδευσης. Ο αλγόριθμος αυτός αποτελείται από τα ακόλουθα βήματα :

1. Όρισε την εικόνα αναφοράς f_{ref} σαν αρχική εκτίμηση της A_0 και θέσε το χωρίο Ω_M να είναι ίσο με το πεδίο ορισμού της f_{ref} .
2. Εφάρμοσε 1-1 αφινική ευθυγράμμιση (Επίπεδο 1) μεταξύ κάθε εικόνας εκπαίδευσης f_k , $k = 1, \dots, N_t$ και της A_0 .
3. Ανανέωσε την εκτίμηση της A_0 σαν την μέση τιμή των ευθυγραμμισμένων εικόνων.
4. Εάν η διαφορά μεταξύ της τρέχουσας και της προηγούμενης εκτίμησης της A_0 είναι αμελητέα, θεώρησε σύγκλιση και τερμάτισε τον αλγόριθμο. Αλλιώς επέστρεψε στο βήμα 2.

6.3.3.1.3 Επίπεδο 3: Επαναληπτική ανάδραση από τον χρήστη. Με σκοπό την αποφυγή μη-αληθοφανών αποτελεσμάτων ευθυγράμμισης στην φάση της εκπαίδευσης του μοντέλου, προσθέτουμε στο σύστημα ευθυγράμμισης ένα τρίτο επίπεδο που υλοποιεί μία ημιαυτόματη επαναληπτική διαδικασία. Σε κάθε επανάληψη, εφαρμόζεται η μέθοδος του Επιπέδου 2 (ευθυγράμμιση συνόλου εικόνων ΣΕ) και ύστερα ο χρήστης εισάγει ανάδραση

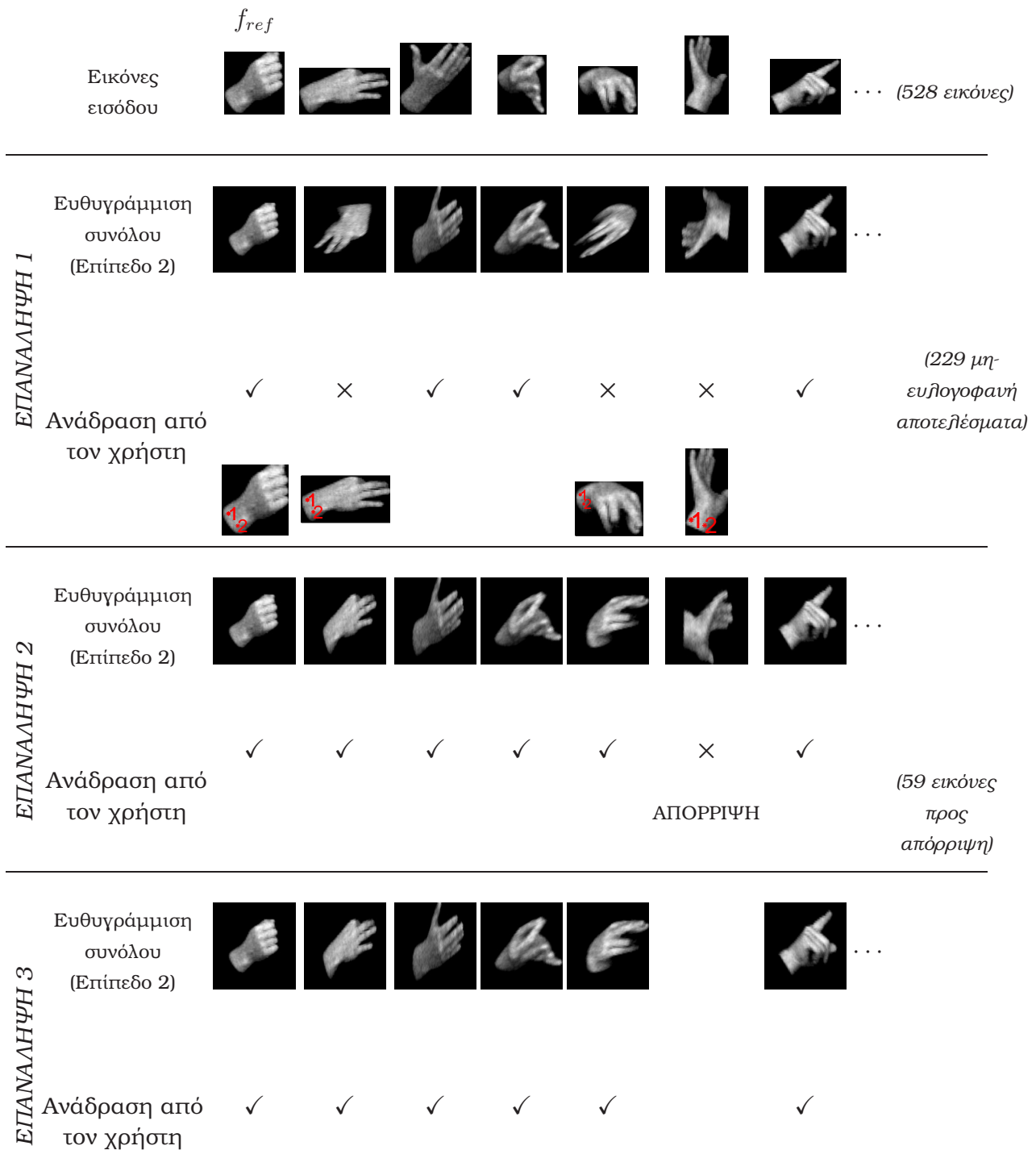
σχετικά με το αποτέλεσμα : Για κάθε εκπαιδευτική εικόνα ΣΕ για την οποία η ευθυγράμμιση συνόλου δεν είναι ευλογοφανής, ο χρήστης είτε επισημειώνει αντιστοιχισμένα σημεία ώστε να βοηθήσει την ευθυγράμμιση συνόλου της επόμενης επανάληψης, είτε, αν αυτό έχει ήδη γίνει σε προηγούμενη επανάληψη, απορρίπτει την εικόνα από το σύνολο εκπαίδευσης. Έπειτα, στην επόμενη επανάληψη, η ευθυγράμμιση συνόλου εφαρμόζεται ξανά λαμβάνοντας υπόψιν την ανάδραση από τον χρήστη και ούτω καθεξής. Τα βασικά βήματα της διαδικασίας αυτής φαίνονται στο Σχήμα 6.7.

Ολοκληρώνοντας την παρουσίαση του συστήματος αφινικής ευθυγράμμισης του συνόλου εκπαίδευσης, αξίζει να τονιστεί ότι η επισημείωση από τον χρήστη αντιστοιχισμένων σημείων χρειάζεται μόνο κατά τη διάρκεια της εκπαίδευσης του Μοντέλου Σχήματος-Εμφάνισης και όχι κατά τη διάρκεια του ταιριάσματος του μοντέλου αυτού. Επίσης, σε αντίθεση με μεθόδους που χρησιμοποιούν ενδεικτικά σημεία για την μοντελοποίηση του σχήματος, π.χ. [36, 83], εδώ αρκεί απλά: 1) να επισημειώσουμε πολύ λίγα σημεία (προς το παρόν χρησιμοποιούμε μόνο 2 σημεία), 2) τα σημεία αυτά να έχουν μόνο μία χαλαρή αντιστοίχιση μεταξύ διαφορετικών πλαισίων και 3) να κάνουμε τέτοιες επισημειώσεις σε ορισμένες μόνο εικόνες από το σύνολο εκπαίδευσης (π.χ. στα τρέχοντα πειράματά μας αυτές οι επισημειώσεις χρειάστηκαν κατά μέσο όρο να γίνουν στο 40% των εικόνων εκπαίδευσης).

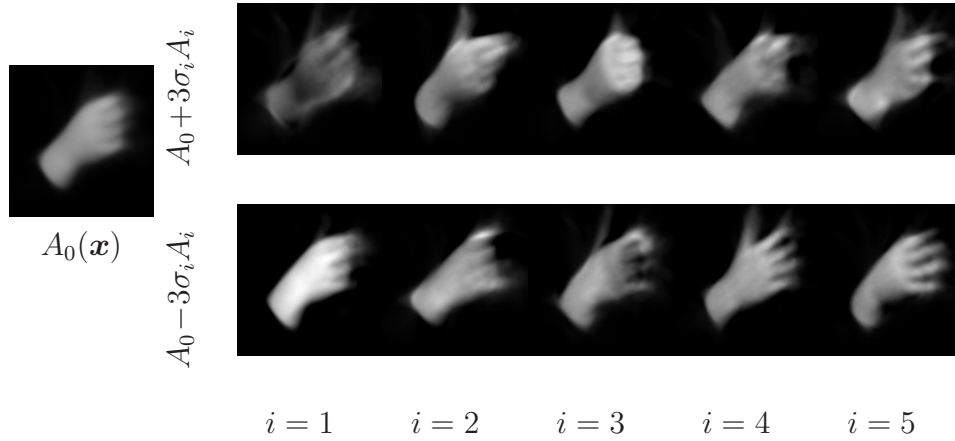
6.3.3.2 Εφαρμογή PCA στο Ευθυγραμμισμένο Σύνολο Εκπαίδευσης

Κάθε ευθυγραμμισμένη εικόνα του συνόλου εκπαίδευσης ορίζεται στο ίδιο ορθογωνικό χωρίο Ω_M , επομένως έχει τις ίδιες διαστάσεις, έστω $\mu \times \nu$ εικονοστοιχεία. Σαρώνοντας όλες αυτές τις εικόνες με ένα προκαθορισμένο μοτίβο, π.χ. από αριστερά προς τα δεξιά και από πάνω προς τα κάτω, σχηματίζουμε ένα σύνολο \mathcal{F} από διανύσματα εκπαίδευσης τα οποία ανήκουν στον $\mathbb{R}^{\mu\nu}$. Έπειτα, οι εικόνες A_i του γραμμικού συνδυασμού του μοντέλου Σχήματος-Εμφάνισης μαθαίνονται στατιστικά μέσω της ανάλυσης PCA αυτού του συνόλου: Η εικόνα βάσης προκύπτει από το μέσο διάνυσμα του συνόλου \mathcal{F} και οι εικόνες A_1, \dots, A_{N_c} προκύπτουν από τα μοναδιαία ιδιοδιανύσματα που αντιστοιχούν στις N_c μεγαλύτερες ιδιοτιμές $\ell_1, \dots, \ell_{N_c}$ της μήτρας συμμεταβλητότητας (covariance matrix) του \mathcal{F} . Αυτός είναι και ο λόγος για τον οποίο χρησιμοποιούμε τον όρο *ιδιοεικόνες* (eigenimages) για τις εικόνες A_1, \dots, A_{N_c} .

Ας σημειωθεί ότι ο αριθμός N_c των ιδιοεικόνων που κρατάμε είναι μία βασική παράμετρος του Μοντέλου Σχήματος-Εμφάνισης. Χρησιμοποιώντας μεγαλύτερο N_c , το μοντέλο μπορεί να διακρίνει καλύτερα μεταξύ διαφορετικών διαμορφώσεων του χεριού. Από την άλλη, εάν το N_c γίνει υπερβολικά μεγάλο, υπάρχει ο κίνδυνος το μοντέλο να μην γενικεύει ικανοποιητικά, υπό την έννοια ότι θα προσπαθεί να 'εξηγή' και διακυμάνσεις των εικόνων που οφείλονται σε θόρυβο ή οπτική πληροφορία που είναι αδιάφορη για την συγκεκριμένη εφαρμογή. Επιπρόσθετα, υψηλότερο N_c συνεπάγεται μεγαλύτερο υπολογιστικό χρόνο για το ταιρίασμα του μοντέλου ΣΕ σε κάθε frame. Στην τρέχουσα εκδοχή των πειραμάτων μας, καταλήξαμε ότι η τιμή $N_c = 35$ είναι αρκετά ικανοποιητική. Στην Εικόνα 6.8 φαίνονται αποτελέσματα της εφαρμογής της PCA στο ευθυγραμμισμένο σύνολο εικόνων ΣΕ. Παρατηρούμε ότι η επίδραση κάθε ιδιοεικόνας στην εικόνα ΣΕ του μοντέλου είναι αρκετά αισθητική.



Σχήμα 6.7: Ημιαυτόματη διαδικασία αφινικής ευθυγράμμισης ενός συνόλου εκπαιδευσης εικόνων ΣΕ (Επίπεδο 3). Σε κάθε επανάληψη, εφαρμόζεται η μέθοδος ευθυγράμμισης του Επιπέδου 2 και ύστερα ο χρήστης εισάγει ανάδραση σχετικά με το αποτέλεσμα. Στην επόμενη επανάληψη, η ευθυγράμμιση συνόλου εφαρμόζεται ξανά λαμβάνοντας υπόψιν την ανάδραση αυτή. Η διαδικασία αυτή τελειώνει όταν ο χρήστης επισημειώσει ότι όλες οι ευθυγραμμίσεις εικόνων είναι αληθοφανείς.



Σχήμα 6.8: Εκμάθηση των εικόνων του γραμμικού συνδυασμού του ΜΣΕ (6.3) μέσω της εφαρμογής PCA στο ευθυγραμμισμένο σύνολο εκπαίδευσης: Μέση εικόνα $A_0(\mathbf{x})$ και μεταβολές στις κατευθύνσεις των 5 πρώτων ιδιοεικόνων.

6.3.4 Άμεσο Ταίριασμα του Μοντέλου

Έχοντας κατασκευάσει το Μοντέλο Σχήματος-Εμφάνισης, μπορούμε τώρα να δοκιμάσουμε να το ταιριάξουμε στα πλαίσια ενός βίντεο νοηματικής γλώσσας, με σκοπό την παρακολούθηση του χεριού και την εξαγωγή χαρακτηριστικών χειρομορφής. Στην ενότητα αυτή παρουσιάζουμε μία μέθοδο ταιριάσματος η οποία είναι *άμεση*, υπό την έννοια ότι ελαχιστοποιεί, ως προς τις παραμέτρους του μοντέλου, μία απλή L_2 νόρμα του σφάλματος ανακατασκευής βασισμένης στο μοντέλο. Η μέθοδος αυτή είναι αρκετά αποτελεσματική στην περίπτωση που το χέρι δεν αλληλεπικαλύπτεται με το άλλο χέρι ή το κεφάλι, αλλά δεν είναι εύρωστη σε περιπτώσεις αλληλεπικάλυψης. Για τον λόγο αυτό, όπως θα περιγραφεί στην Ενότητα 6.3.5, η μέθοδος ταιριάσματος που τελικά χρησιμοποιούμε στο σύστημά μας είναι μία *ομαλοποιημένη* (regularized) μέθοδος που ενσωματώνει πρότερη στατική και δυναμική πληροφορία για τις παραμέτρους του μοντέλου. Η άμεση μέθοδος της παρούσας ενότητας χρησιμοποιείται μόνο στην διαδικασία της εκπαίδευσης των πρότερων πληροφοριών των παραμέτρων του ΜΣΕ.

Πιο συγκεκριμένα, στο άμεσο ταίριασμα του ΜΣΕ, στόχος μας είναι να βρούμε τις παραμέτρους \mathbf{p} και $\boldsymbol{\lambda}$, έτσι ώστε το μοντέλο να περιγράφει την αντίστοιχη εικόνα ΣΕ $f(\mathbf{x})$ όσο το δυνατόν ακριβέστερα. Για τον σκοπό αυτό, ελαχιστοποιούμε το μέσο τετραγωνικό σφάλμα ανακατασκευής (μετρούμενο στο πεδίο του μοντέλου ΣΕ Ω_M):

$$E_{rec}(\boldsymbol{\lambda}, \mathbf{p}) = \frac{1}{N_M} \sum_{\mathbf{x}} \left\{ A_0(\mathbf{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\mathbf{x}) - f(W_{\mathbf{p}}(\mathbf{x})) \right\}^2, \quad (6.8)$$

ταυτόχρονα ως προς τα $\boldsymbol{\lambda}$ και \mathbf{p} . Σημειώστε ότι η παραπάνω άθροιση αφορά όλα τα εικονοστοιχεία \mathbf{x} του πεδίου Ω_M . Υλοποιούμε αυτή την ελαχιστοποίηση χρησιμοποιώντας τον *Ταυτόχρονο Αντίστροφο Συνθετικό* (Simultaneous Inverse Compositional - SIC) αλγόριθμο του άρθρου [55], ο οποίος γενικεύει τον αλγόριθμο IC που χρησιμοποιήσαμε για την 1-1 ευθυγράμμιση δύο εικόνων ΣΕ (βλ. Επίπεδο 1, Ενότητα 6.3.3). Ο αλγόριθμος SIC εφαρμόζει μία βελτιστοποίηση τύπου Gauss-Newton ταυτόχρονα στα \mathbf{p} και $\boldsymbol{\lambda}$. Όσον αφορά την αρχικοποίηση του αλγορίθμου, ακολουθούμε παρόμοια φιλοσοφία όπως στην 1-1 ευθυγράμμιση: Σε όλα τα πλαίσια μιας ακολουθίας εκτός του αρχικού, αρχικοποιούμε

χρησιμοποιώντας το αποτέλεσμα από το προηγούμενο πλαίσιο. Για το αρχικό πλαίσιο, δοκιμάζουμε τους δύο μετασχηματισμούς ομοιότητας που όταν εφαρμόζονται στην μέση εικόνα A_0 κάνουν την μάσκα της να έχει το ίδιο κεντροειδές, εμβαδόν και προσανατολισμό με την μάσκα της εικόνας εισόδου. Τελικά, επιλέγουμε την αρχικοποίηση που επιφέρει το μικρότερο μέσο τετραγωνικό σφάλμα (6.8).

Μερικά παραδείγματα από την παραπάνω διαδικασία άμεσου ταιριάσματος εμφανίζονται στο Σχήμα 6.9. Όταν το χέρι δεν αλληλεπικαλύπτεται, τα αποτελέσματα φαίνονται αρκετά λογικοφανή και οι ανακατασκευές βασισμένες στο μοντέλο είναι αρκετά ακριβείς, παρά τον σχετικά μικρό αριθμό ($N_c = 35$) ιδιοεικόνων. Ακόμα, οι βέλτιστοι παράμετροι του αφινικού μετασχηματισμού εντοπίζονται επιτυχώς μεταβολές στην 3D πόζα. Από την άλλη, στις περισσότερες περιπτώσεις αλληλεπικαλύψεων, όπου η εικόνα ΣΕ δεν περιλαμβάνει μόνο το μοντελοποιούμενο χέρι, ο αλγόριθμος δεν είναι εύρωστος διότι επεκτείνει υπερβολικά το ανακατασκευασμένο χέρι και το κάνει να καλύπτει ολόκληρη την περιοχή με χρώμα δέρματος.

6.3.5 Ομαλοποιημένο Ταίριασμα του ΜΣΕ με Στατική και Δυναμική Πρότερη Πληροφορία

Παρακινούμενοι από την ανάγκη να επιτύχουμε ευρωστία σε αλληλεπικαλύψεις, έχουμε σχεδιάσει μία *ομαλοποιημένη* (regularized) μέθοδο ταιριάσματος του ΜΣΕ. Προσθέτουμε στο μέσο τετραγωνικό σφάλμα ανακατασκευής E_{rec} (6.8), το οποίο χρησιμοποιείται στο Άμεσο Ταίριασμα, δύο όρους που αντιστοιχούν σε πρότερη στατική και δυναμική πληροφορία για τις παραμέτρους λ και \mathbf{p} του μοντέλου. Σε κάθε πλαίσιο n του βίντεο λοιπόν, βρίσκουμε τα βέλτιστα $\lambda = \lambda[n]$ και $\mathbf{p} = \mathbf{p}[n]$ που ελαχιστοποιούν την ολική ενέργεια :

$$E(\lambda, \mathbf{p}) = E_{rec}(\lambda, \mathbf{p}) + w_S E_S(\lambda, \mathbf{p}) + w_D E_D(\lambda, \mathbf{p}) , \quad (6.9)$$

όπου w_S, w_D είναι θετικά βάρη οι οποίες ελέγχουν την ισορροπία μεταξύ των 3 όρων. Ο όρος $E_S(\lambda, \mathbf{p})$ αντιστοιχεί σε στατική πρότερη πληροφορία και διασφαλίζει ότι η λύση μένει σχετικά κοντά στις μέσες τιμές λ_0, \mathbf{p}_0 των παραμέτρων :

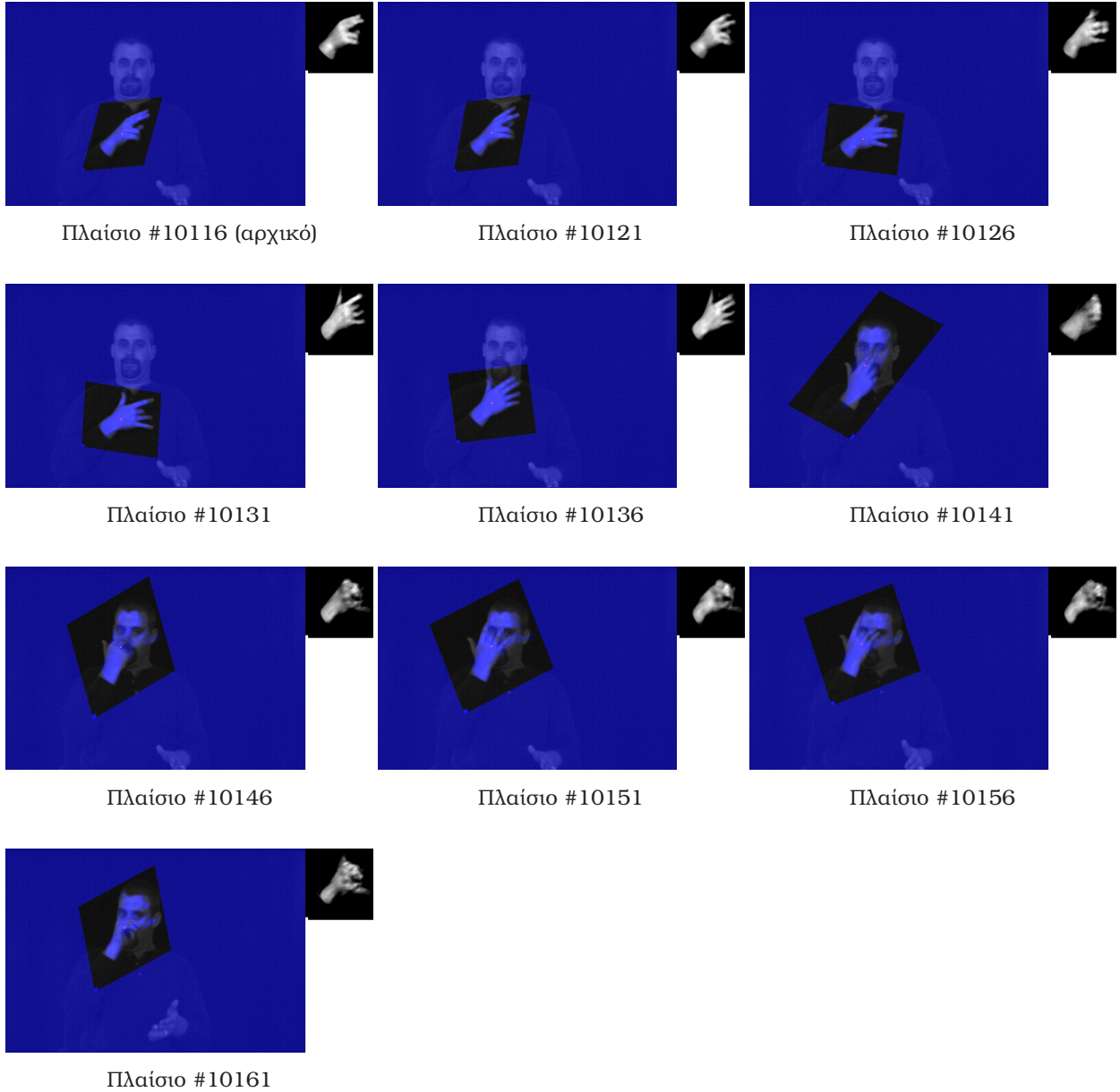
$$E_S(\lambda, \mathbf{p}) = \frac{1}{N_c} \|\lambda - \lambda_0\|_{\Sigma_\lambda}^2 + \frac{1}{N_p} \|\mathbf{p} - \mathbf{p}_0\|_{\Sigma_p}^2 , \quad (6.10)$$

όπου N_c, N_p είναι οι διαστάσεις των λ, \mathbf{p} ($N_p=6$ λόγω του ότι μοντελοποιούμε αφινικούς μετασχηματισμούς). Επιπρόσθετα, ο όρος $E_D(\lambda, \mathbf{p})$ αντιστοιχεί σε δυναμική πρότερη πληροφορία και κάνει τη λύση να παραμένει κοντά στις προβλέψεις $\lambda^e = \lambda^e[n], \mathbf{p}^e = \mathbf{p}^e[n]$ των παραμέτρων. Οι προβλέψεις αυτές βασίζονται σε χρονικά γειτονικές τιμές από πλαίσια στα οποία το ταίριασμα του μοντέλου έχει ήδη γίνει (για το πώς εξάγονται οι εκτιμήσεις αυτές, βλ. Ενότητα 6.3.5.1):

$$E_D(\lambda, \mathbf{p}) = \frac{1}{N_c} \|\lambda - \lambda^e\|_{\Sigma_{\epsilon_\lambda}}^2 + \frac{1}{N_p} \|\mathbf{p} - \mathbf{p}^e\|_{\Sigma_{\epsilon_p}}^2 . \quad (6.11)$$

Ας σημειωθεί ότι στα παραπάνω έχουμε συμβολίσει με $\|\mathbf{y}\|_A$, με το A να είναι ένας $N \times N$ συμμετρικός θετικά ορισμένος πίνακας και $\mathbf{y} \in \mathbb{R}^N$, την ακόλουθη απόσταση Mahalanobis από το $\mathbf{0}$:

$$\|\mathbf{y}\|_A \triangleq \sqrt{\mathbf{y}^T A^{-1} \mathbf{y}} \quad (6.12)$$



Σχήμα 6.9: Άμεσο ταίριασμα του μοντέλου Σχήματος-Εμφάνιση σε ένα τμήμα διαδοχικών πλαισίων ενός βίντεο. Σε κάθε αρχική εικόνα πλαισίου, υπερθέτουμε την ανακατασκευή βασισμένη στο μοντέλο, $A_0(W_p^{-1}(\mathbf{x})) + \sum \lambda_i A_i(W_p^{-1}(\mathbf{x}))$. Στην πάνω δεξιά γωνία δείχνουμε κάθε φορά την ίδια ανακατασκευή, αλλά στον χώρο του μοντέλου $\Sigma \in \Omega_M$, $A_0(\mathbf{x}) + \sum \lambda_i A_i(\mathbf{x})$, η οποία καθορίζει τα βέλτιστα βάρη λ .

Επίσης, Σ_λ και Σ_p είναι οι μήτρες συμμεταβλητότητας των λ και p αντίστοιχα, ενώ $\Sigma_{\epsilon_\lambda}$ και Σ_{ϵ_p} είναι οι μήτρες συμμεταβλητότητας των σφαλμάτων πρόβλεψης των λ και p αντίστοιχα (για την εκμάθηση αυτών των ποσοτήτων, βλ. Ενότητα 6.3.5.2). Επομένως, οι όροι $E_S(\lambda, p)$ και $E_D(\lambda, p)$ επιβάλλουν ποινή στην απόκλιση από τις μέσες και προβλεφθείσες τιμές, αλλά με έναν σταθμισμένο τρόπο, σύμφωνα με τις κατάλληλες μήτρες συμμεταβλητότητας.

Επιπρόσθετα, έχουμε υποθέσει ότι οι παράμετροι λ και p είναι στατιστικά ανεξάρτητες.

Μιας και τα διανύσματα αυτά έχουν διαφορετικές διαστάσεις, τα βάρη $1/N_c$ και $1/N_p$ στις Εξισώσεις (6.10),(6.11) λειτουργούν σαν κανονικοποίηση και εξασφαλίζουν παρόμοια ισχύ στους όρους ποινής του λ και του p .

Από τη στιγμή που, όπως περιγράφηκε στην Ενότητα 6.3.2, οι παράμετροι λ είναι τα βάρη των ιδιοεικόνων που προέρχονται από την εφαρμογή PCA στις ευθυγραμμισμένες εικόνες ΣΕ, υποθέτουμε επίσης ότι η μέση τιμή τους λ_0 είναι μηδενική και ότι η μήτρα συμμεταβλητότητας Σ_λ είναι διαγώνια, το οποίο σημαίνει ότι κάθε συνιστώσα του λ είναι ανεξάρτητη από όλες τις υπόλοιπες.

Ας σημειωθεί τέλος ότι τα βάρη w_S, w_D δεν είναι χρονικά σταθερά, αλλά εξαρτώνται από το εάν το μοντελοποιούμενο χέρι στο τρέχον πλαίσιο βρίσκεται ή όχι σε αλληλεπικάλυψη. Στις περιπτώσεις αλληλεπικαλύψεων, μπορούμε να εμπιστευθούμε λιγότερο την εικόνα εισόδου $f(x)$, η οποία υπεισέρχεται στον όρο $E_{rec}(\lambda, p)$. Επομένως, στις περιπτώσεις αυτές αυξάνουμε τα βάρη w_S, w_D για να επιτύχουμε περισσότερη ευρωστία. Παράλληλα, ελαττώνουμε το σχετικό βάρος $\frac{w_D}{w_S+w_D}$ του όρου δυναμικής πληροφορίας, έτσι ώστε να αποτρέψουμε πιθανή συσσώρευση σφάλματος, η οποία, σε περιπτώσεις αλληλεπικαλύψεων μεγάλης διάρκειας, θα μπορούσε να διαδοθεί μέσω των προβλέψεων λ^e, p^e . Στην τρέχουσα οργάνωση των πειραμάτων μας για παράδειγμα, έχουμε επιλέξει 1) $w_S = 0.98, w_D = 0.42$ για τις περιπτώσεις αλληλεπικαλύψεων και 2) $w_S = 0.0695, w_D = 0.0695$ για τις υπόλοιπες περιπτώσεις.

6.3.5.1 Δυναμικά Μοντέλα για την Πρόβλεψη των Παραμέτρων

Με σκοπό την εξαγωγή εκτιμήσεων λ^e, p^e οι οποίες χρησιμοποιούνται στον όρο δυναμικής πρότερης πληροφορίας E_D (6.11), χρησιμοποιούμε απλά γραμμικά μοντέλα πρόβλεψης.

Πρώτα από όλα, ας σημειώσουμε ότι το ομαλοποιημένο ταίριασμα του ΜΣΕ σε ένα πλαίσιο μίας συγκεκριμένης χρονικής στιγμής n είναι μέρος του ακολουθιακού ταιριάσματος σε ένα ολόκληρο τμήμα από διαδοχικά πλαίσια ενός βίντεο. Ανάλογα με το αποτέλεσμα της διαδικασίας κατάτμησης ενός βίντεο που θα περιγραφεί στην Ενότητα 6.3.5.4, η χρονική κατεύθυνση ταιριάσματος που ακολουθείται στο τρέχον τμήμα μπορεί να είναι είτε προς τα εμπρός είτε προς τα πίσω.

Τα πλαίσια που θα χρησιμοποιηθούν στις προβλέψεις των παραμέτρων την στιγμή n εξαρτώνται από μία τιμή $K = K(n)$ για την οποία κάνουμε την ακόλουθη σύμβαση: Εάν $K > 0$, το ταίριασμα κατευθύνεται προς τα εμπρός και η πρόβλεψη θα χρησιμοποιήσει τα ήδη υπάρχοντα αποτελέσματα ταιριάσματος από το παράθυρο των K αμέσως προηγούμενων πλαισίων. Εάν $K < 0$, το ταίριασμα κατευθύνεται προς τα πίσω και η πρόβλεψη θα χρησιμοποιήσει τα ήδη υπάρχοντα αποτελέσματα των $|K|$ αμέσως επόμενων πλαισίων.

Μία παράμετρος $K_m \geq 1$, η οποία προσδιορίζεται χειροκίνητα (επί του παρόντος έχουμε θέσει $K_m=8$), καθορίζει τον μέγιστο αριθμό προηγούμενων ή επόμενων πλαισίων που επιτρέπεται να χρησιμοποιηθούν, δηλαδή την μέγιστη τάξη του διακριτού μοντέλου πρόβλεψης. Συνεπώς, εάν K_{av} είναι ο αριθμός των συνεχόμενων πλαισίων του τρέχοντος τμήματος με ήδη διαθέσιμες παραμέτρους ΜΣΕ, η τιμή K προσδιορίζεται ως εξής: Το πρόσημό της τίθεται θετικό στην περίπτωση ταιριάσματος προς τα εμπρός και αρνητικό διαφορετικά, ενώ η απόλυτη τιμή της λαμβάνει την τιμή $|K| = \min(K_{av}, K_m)$.

Ας σημειωθεί ότι το K μπορεί να λάβει τις τιμές $\{-K_m, \dots, K_m\}$. Εάν $K = 0$, βρισκόμαστε στο αρχικό πλαίσιο μιας διαδικασίας ακολουθιακού ταιριάσματος, οπότε δεν

μπορεί να γίνει κάποια πρόβλεψη βασισμένη σε αποτελέσματα από άλλα πλαίσια. Στην περίπτωση αυτή λοιπόν, η οποία είναι εκφυλισμένη για την γραμμική πρόβλεψη, θεωρούμε ότι οι προβλέψεις προκύπτουν από τις προτέρες μέσες τιμές $\boldsymbol{\lambda}^e = \boldsymbol{\lambda}_0$, $\boldsymbol{p}^e = \boldsymbol{p}_0$ και επίσης ότι $\Sigma_{\epsilon_\lambda} = \Sigma_\epsilon$, $\Sigma_{\epsilon_p} = \Sigma_p$, το οποίο έχει ως αποτέλεσμα $E_D(\boldsymbol{\lambda}, \boldsymbol{p}) = E_S(\boldsymbol{\lambda}, \boldsymbol{p})$. Σε όλες τις υπόλοιπες περιπτώσεις, εφαρμόζουμε την μεθοδολογία που περιγράφεται στη συνέχεια.

Δεδομένης της τιμής παραθύρου πρόβλεψης K , οι παράμετροι $\boldsymbol{\lambda}$ προβλέπονται σύμφωνα με το ακόλουθο μοντέλο αυτοπαλινδρόμησης (autoregressive model):

$$\boldsymbol{\lambda}^e[n] = \sum_{\nu \in W(K)} A_{K,\nu} \boldsymbol{\lambda}[n - \nu], \quad (6.13)$$

όπου το $W(K)$ είναι το σύνολο από τους σχετικούς δείκτες ν για τους οποίους τα $n - \nu$ αντιστοιχούν στις χρονικές στιγμές με υπάρχοντα αποτελέσματα που χρησιμοποιούνται στην πρόβλεψη:

$$W(K) = \begin{cases} \{1, 2, \dots, K\}, & \text{αν } K > 0 \\ \{-1, -2, \dots, K\}, & \text{αν } K < 0 \end{cases} \quad (6.14)$$

Επίσης, $A_{K,\nu}$, $n \in W(K)$, είναι οι $|K|$ μήτρες στάθμισης (weight matrices) με διαστάσεις $N_c \times N_c$. Σημειώστε ότι για κάθε K , χρησιμοποιούμε διαφορετικές μήτρες στάθμισης, οι οποίες προκύπτουν από μία ξεχωριστή εκπαίδευση. Το κίνητρο για αυτό είναι ότι θέλουμε να βελτιστοποιήσουμε την ακρίβεια των προβλέψεων μόνο για την συγκεκριμένη κάθε φορά περίπτωση παραθύρου πρόβλεψης (βλ. Ενότητα 6.3.5.2). Το Σχήμα 6.10 δείχνει τις μήτρες στάθμισης που χρησιμοποιούνται για κάθε K καθώς επίσης τις μετατοπισμένες χρονικές στιγμές στις οποίες εφαρμόζονται πάνω στις αντίστοιχες παραμέτρους. Από τη στιγμή που οι συνιστώσες του $\boldsymbol{\lambda}$ θεωρούνται αμοιβαία ανεξάρτητες, είναι εύλογο να θεωρήσουμε ότι όλες οι μήτρες $A_{K,\nu}$ είναι διαγώνιες, το οποίο σημαίνει ότι κάθε συνιστώσα έχει ένα ανεξάρτητο μοντέλο πρόβλεψης.

Όσον αφορά τώρα τις παραμέτρους \boldsymbol{p} , αυτές δεν έχουν μηδενική μέση τιμή, ούτε μπορούμε να θεωρήσουμε τις συνιστώσες του \boldsymbol{p} σαν ανεξάρτητες. Επομένως, για να εφαρμόσουμε την ίδια μεθοδολογία με αυτήν που προηγήθηκε για το $\boldsymbol{\lambda}$, παράγουμε πρώτα μία νέα παραμετροποίηση του \boldsymbol{p} εφαρμόζοντας PCA και θεωρώντας όλες τις εξαχθείσες συνιστώσες πρωτεύουσες:

Εστω $\Sigma_{\tilde{\boldsymbol{p}}}$ ο διαγώνιος πίνακας που περιέχει τις ιδιοτιμές του Σ_p και έστω U_p η μήτρα της οποίας οι στήλες περιέχουν τα αντίστοιχα ιδιοδιανύσματα του Σ_p . Τότε:

$$\Sigma_p = U_p \Sigma_{\tilde{\boldsymbol{p}}} U_p^T \quad (6.15)$$

και ο μετασχηματισμός:

$$\tilde{\boldsymbol{p}} = U_p^T (\boldsymbol{p} - \boldsymbol{p}_0) \quad (6.16)$$

εξάγει ένα διάνυσμα νέων παραμέτρων $\tilde{\boldsymbol{p}}$, το οποίο έχει την ίδια διάσταση με το \boldsymbol{p} και επιπλέον είναι κανονικοποιημένο, υπό την έννοια ότι έχει μηδενική μέση τιμή και διαγώνια μήτρα συμμεταβλητότητας $\Sigma_{\tilde{\boldsymbol{p}}}$ (το οποίο σημαίνει ότι οι συνιστώσες είναι αμοιβαία ανεξάρτητες).

Οι κανονικοποιημένες παράμετροι $\tilde{\boldsymbol{p}}$ προβλέπονται λοιπόν σύμφωνα με τα ακόλουθα μοντέλα:

$$\tilde{\boldsymbol{p}}^e[n] = \sum_{\nu \in W(K)} B_{K,\nu} \tilde{\boldsymbol{p}}[n - \nu], \quad (6.17)$$

	$\lambda[n + K_m]$	\cdots	$\lambda[n + 3]$	$\lambda[n + 2]$	$\lambda[n + 1]$	$\lambda[n - 1]$	$\lambda[n - 2]$	$\lambda[n - 3]$	\cdots	$\lambda[n - K_m]$
$K = -K_m$	$A_{-K_m, -K_m}$	\cdots	$A_{-K_m, -3}$	$A_{-K_m, -2}$	$A_{-K_m, -1}$					
\vdots		\ddots	\vdots	\vdots	\vdots					
$K = -3$			$A_{-3, -3}$	$A_{-3, -2}$	$A_{-3, -1}$					
$K = -2$				$A_{-2, -2}$	$A_{-2, -1}$					
$K = -1$					$A_{-1, -1}$					
$K = 1$						$A_{1, 1}$				
$K = 2$						$A_{2, 1}$	$A_{2, 2}$			
$K = 3$						$A_{3, 1}$	$A_{3, 2}$	$A_{3, 2}$		
\vdots						\vdots	\vdots	\vdots	\ddots	
$K = K_m$						$A_{K_m, 1}$	$A_{K_m, 2}$	$A_{K_m, 2}$	\cdots	A_{K_m, K_m}

Σχήμα 6.10: Μήτρες στάθμισης που χρησιμοποιούνται για την πρόβλεψη του $\lambda[n]$ και αντίστοιχες μετατοπισμένες τιμές $\lambda[n - \nu]$ τις οποίες πολλαπλασιάζουν οι μήτρες. Κάθε γραμμή αντιστοιχεί σε μία διαφορετική τιμή παραθύρου πρόβλεψης K . Η περίπτωση $K > 0$ συμβολίζει πρόβλεψη προς τα εμπρός ενώ η περίπτωση $K < 0$ συμβολίζει πρόβλεψη προς τα πίσω. Αντίστοιχη οπτικοποίηση ισχύει και για τις παραμέτρους \tilde{p} και μήτρες $B_{K, \nu}$.

όπου $B_{K, \nu}$ είναι οι αντίστοιχες μήτρες στάθμισης, οι οποίες θεωρούνται και αυτές διαγώνιες.

6.3.5.2 Εκπαίδευση των Στατικών και Δυναμικών Πρότερων Πληροφοριών

Για να μπορέσουμε να εφαρμόσουμε το ομαλοποιημένο ταίριασμα του μοντέλου ΣΕ, πρέπει πρώτα να μάθουμε στατιστικά τις πρότερες στατικές και δυναμικές πληροφορίες για τις παραμέτρους λ και p . Με άλλα λόγια, πρέπει να προσδιορίσουμε τα εξής:

- μέσες τιμές λ_0, p_0 ,
- μήτρες συμμεταβλητότητας $\Sigma_\lambda, \Sigma_p, \Sigma_{\epsilon_\lambda}, \Sigma_{\epsilon_p}$ και
- μήτρες στάθμισης $A_{K, \nu}, B_{K, \nu}$ των μοντέλων πρόβλεψης.

Αυτός ο προσδιορισμός γίνεται χρησιμοποιώντας υπακολουθίες εκπαίδευσης στις οποίες το μοντελοποιούμενο χέρι δεν βρίσκεται σε αλληλεπικάλυψη.

Πιο συγκεκριμένα, εφαρμόζεται αρχικά μία τυχαία επιλογή τέτοιου είδους υπακολουθιών από κάποια βίντεο εκπαίδευσης του ίδιου νοηματιστή. Στα τρέχοντα πειράματα, έχουν χρησιμοποιηθεί 3 βίντεο εκπαίδευσης και έχουν επιλεγθεί τυχαία 120 υπακολουθίες οι οποίες περιέχουν 2882 πλαίσια. Στη συνέχεια, το άμεσο ταίριασμα του ΜΣΕ που περιγράψαμε στην Ενότητα 6.3.4 εφαρμόζεται σε όλες τις υπακολουθίες εκπαίδευσης και εξάγει βέλτιστες παραμέτρους λ, p για όλα τα πλαίσια εκπαίδευσης. Αυτές χρησιμοποιούνται για την εκπαίδευση των Στατικών και Δυναμικών Πρότερων Πληροφοριών.

6.3.5.2.1 Στατική Πρότερη Πληροφορία. Στην περίπτωση αυτή, οι παράμετροι εκπαίδευσης λ, p χρησιμοποιούνται σαν δείγματα της ίδιας πολυμεταβλητής κατανομής,

χωρίς να λαμβάνεται υπόψιν οποιαδήποτε ιδιότητα διαδοχικότητας μέσα στις υπακολουθίες εκπαίδευσης. Με τον τρόπο αυτό, σχηματίζουμε τα σύνολα εκπαίδευσης T_λ και T_p , τα οποία αντιστοιχούν στα λ και p αντίστοιχα.

Όσον αφορά το διάνυσμα παραμέτρων λ , έχουμε ήδη υποθέσει ότι $\lambda_0 = 0$ και ότι η μήτρα συμμεταβλητότητας Σ_λ είναι διαγώνια. Επομένως, μόνο οι διακυμάνσεις $\sigma_{\lambda_i}^2$ στην διαγώνιο της Σ_λ μένουν να προσδιοριστούν. Η εκτίμηση της κάθε μιας διακύμανσης $\sigma_{\lambda_i}^2$ γίνεται από την εμπειρική διακύμανση της αντίστοιχης συνιστώσας λ_i στο σύνολο εκπαίδευσης T_λ .

Όσον αφορά τώρα τις παραμέτρους p , υπολογίζουμε τα p_0 και Σ_p από τον εμπειρικό μέσο και την εμπειρική μήτρα συμμεταβλητότητας του συνόλου T_p . Όπως έχει ήδη περιγραφεί στην Ενότητα 6.3.5.1, αυτές οι ποσότητες χρησιμοποιούνται στην εφαρμογή PCA στο p , η οποία οδηγεί στην κανονικοποιημένη παραμετροποίηση \tilde{p} .

6.3.5.2.2 Δυναμική Πρότερη Πληροφορία. Όπως έχουμε ήδη σημειώσει, για κάθε συγκεκριμένη τιμή K , $K \in \{-K_m, \dots, -1, 1, \dots, K_m\}$, θεωρούμε ένα διαφορετικό μοντέλο πρόβλεψης. Για να εκπαιδεύσουμε κάθε τέτοιο μοντέλο ακολουθούμε για κάθε K την ακόλουθη διαδικασία:

Σε κάθε υπακολουθία εκπαίδευσης, η πρόβλεψη για το τρέχον K δοκιμάζεται μόνο στα πλαίσια τα οποία έχουν τουλάχιστον 1) K προηγούμενα (αν $K > 0$) ή 2) $|K|$ επόμενα (αν $K < 0$), πλαίσια στην ίδια υπακολουθία. Ονομάζουμε τα πλαίσια αυτά *πλαίσια δοκιμής πρόβλεψης*. Δεδομένων αυθαίρετων τιμών για τις μήτρες $A_{K,\nu}$ και $B_{K,\nu}$, μπορούμε σε κάθε πλαίσιο δοκιμής πρόβλεψης να συγκρίνουμε τις τιμές των παραμέτρων λ , \tilde{p} από τα δεδομένα εκπαίδευσης με τις αντίστοιχες τιμές λ^e , \tilde{p}^e που προκύπτουν από τα μοντέλα πρόβλεψης (6.13), (6.17). Οι διαφορές μεταξύ των τιμών αυτών σχηματίζουν τα σφάλματα πρόβλεψης ϵ_λ , $\epsilon_{\tilde{p}}$, βλέπε Σχ. 6.11 για ένα ενδεικτικό απλουστευτικό παράδειγμα.

	$\lambda[28]$	$\lambda[29]$	$\lambda[30]$	$\lambda[31]$	$\lambda[110]$	$\lambda[111]$	$\lambda[112]$	$\lambda[113]$	$\lambda[114]$
$K = -3$	✓				✓	✓			
$K = -2$	✓	✓			✓	✓	✓		
$K = -1$	✓	✓	✓		✓	✓	✓	✓	
$K = 1$		✓	✓	✓		✓	✓	✓	✓
$K = 2$			✓	✓			✓	✓	✓
$K = 3$				✓				✓	✓

Σχήμα 6.11: Πλαίσια δοκιμής πρόβλεψης για την εκπαίδευση των πρότερων δυναμικών πληροφοριών σχετικά με το λ σε ένα ενδεικτικό απλουστευτικό παράδειγμα με 2 υπακολουθίες εκπαίδευσης και $K_m = 3$. Κάθε σειρά αντιστοιχεί σε μία διαφορετική τιμή παραθύρου πρόβλεψης K και τα σύμβολα '✓' αντιστοιχούν σε πλαίσια στα οποία δοκιμάζεται η πρόβλεψη με το μοντέλο πρόβλεψης για το τρέχον K . Για παράδειγμα, όταν $K = 2$, τα σφάλματα πρόβλεψης για όλες τις συνιστώσες του $\lambda[30]$ είναι $\lambda[30] - A_{2,1}\lambda[29] - A_{2,2}\lambda[28]$.

Παρόμοια με την Ανάλυση Γραμμικής Πρόβλεψης [96] καθώς και άλλες μεθόδους παρακολούθησης αντικειμένων που χρησιμοποιούν δυναμική πληροφορία, π.χ. [18], μαθαίνουμε τις μήτρες στάθμισης $A_{K,\nu}$, $B_{K,\nu}$ μέσω της ελαχιστοποίησης του μέσου τετραγωνικού σφάλματος πρόβλεψης ως προς όλα τα πλαίσια δοκιμής πρόβλεψης. Από τη στιγμή που οι μήτρες $A_{K,\nu}$, $B_{K,\nu}$ είναι διαγώνιες, αυτή η βελτιστοποίηση γίνεται ανεξάρτητα για

κάθε συνιστώσα των λ και p , αντιμετωπίζοντάς την ως 1Δ σήμα. Οι συντελεστές πρόβλεψης για κάθε συνιστώσα προκύπτουν λοιπόν από τη επίλυση ενός τυπικού προβλήματος ελαχίστων τετραγώνων.

Οι βέλτιστες τιμές των μέσων τετραγωνικών σφαλμάτων καθορίζουν τις εκτιμήσεις των διακυμάνσεων $\sigma_{\epsilon_{\lambda,i}}^2(K)$ ($i = 1, \dots, N_c$) και $\sigma_{\epsilon_{\bar{p},i}}^2(K)$ ($i = 1, \dots, N_p$) των συνιστωσών των σφαλμάτων πρόβλεψης ϵ_{λ} και $\epsilon_{\bar{p}}$. Υποθέτουμε ότι και οι μήτρες συμμεταβλητότητας $\Sigma_{\epsilon_{\lambda}}$ και $\Sigma_{\epsilon_{\bar{p}}}$ είναι διαγώνιες και τα στοιχεία των διαγώνιων τους προσδιορίζονται από τις διακυμάνσεις $\sigma_{\epsilon_{\lambda,i}}^2$ και $\sigma_{\epsilon_{\bar{p},i}}^2$ αντίστοιχα.

6.3.5.3 Αλγόριθμος Ομαλοποιημένου Ταίριασματος

Η ενέργεια $E(\lambda, p)$ (6.9) του προτεινόμενου ταίριασματος του μοντέλου ΣΕ είναι μία ειδική περίπτωση της γενικής αντικειμενικής συνάρτησης που ελαχιστοποιείται από τον *Ταυτόχρονο Αντίστροφο Συνθετικό με Πρότερη Πληροφορία* (Simultaneous Inverse Compositional with a Prior - SICP) αλγόριθμο της αναφοράς [11]. Συνεπώς, για να ελαχιστοποιήσουμε την $E(\lambda, p)$, εξειδικεύουμε τον αλγόριθμο αυτό στην περίπτωση των συγκεκριμένων τύπων όρων πρότερου σχήματος που έχουμε κατασκευάσει.

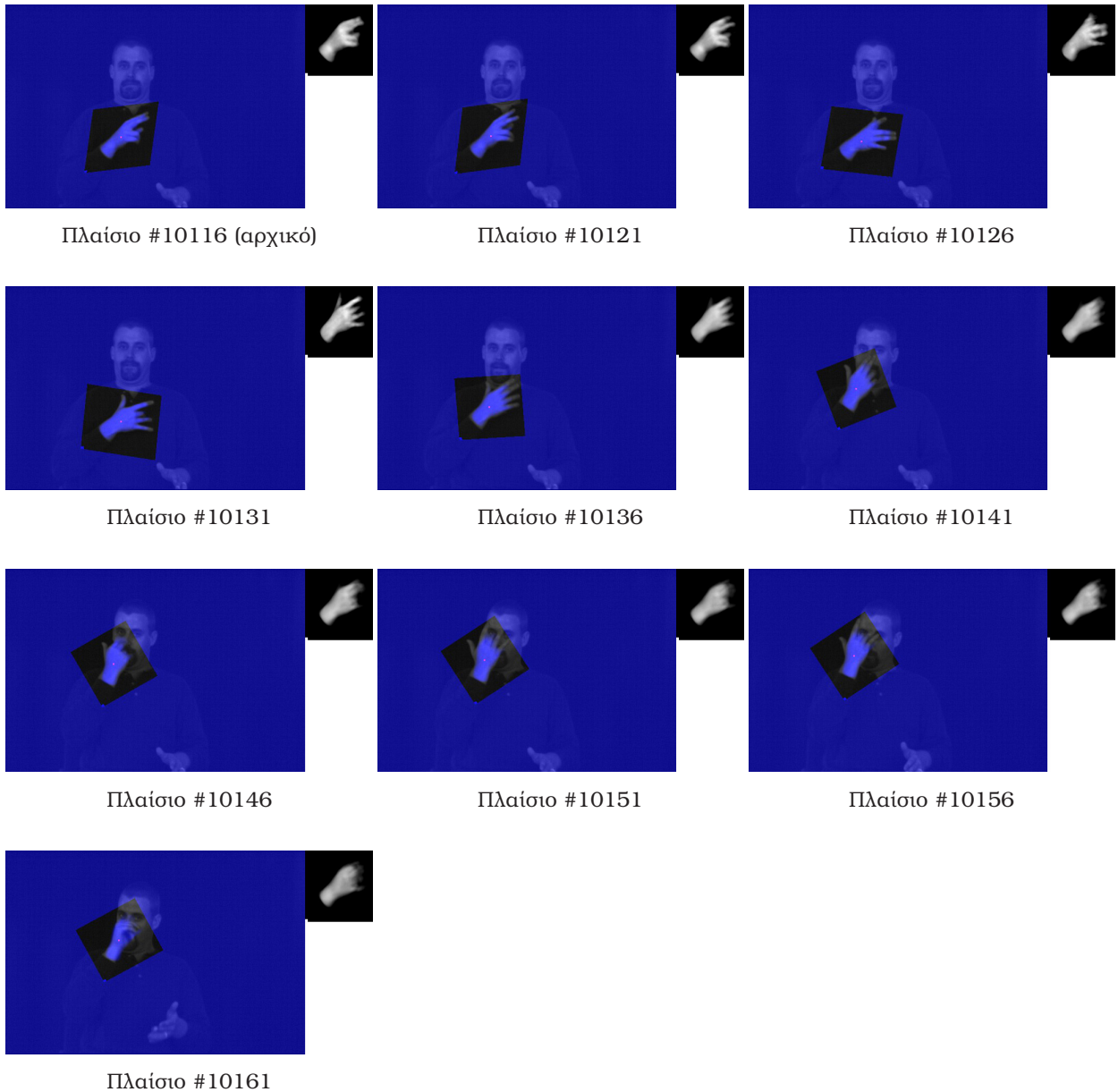
Σε κάθε πλαίσιο n ενός τμήματος βίντεο, ο αλγόριθμος ταίριασματος αρχικοποιείται ως ακολούθως. Εάν το τρέχον πλαίσιο δεν είναι το πλαίσιο εκκίνησης του ακολουθιακού ταίριασματος (δηλαδή $K(n) \neq 0$), τότε οι παράμετροι λ, p αρχικοποιούνται με τις προβλέψεις λ^e, p^e που προκύπτουν από τα μοντέλα πρόβλεψης για το τρέχον K . Διαφορετικά, εάν $K(n) = 0$, υιοθετούμε μία στρατηγική παρόμοια με τις περιπτώσεις 1-1 αφινικής ευθυγράμμισης (Επίπεδο 1, Ενότητα 6.3.3) και άμεσου ταίριασματος ΜΣΕ (Ενότητα 6.3.4). Δοκιμάζουμε σαν αρχικοποιήσεις τους δύο μετασχηματισμούς ομοιότητας που όταν εφαρμόζονται στην μέση εικόνα A_0 του ΜΣΕ κάνουν την μάσκα της να έχει το ίδιο κεντροειδές, εμβαδόν και προσανατολισμό με την μάσκα της εικόνας ΣΕ εισόδου. Εφαρμόζουμε δύο φορές τον αλγόριθμο SICP χρησιμοποιώντας αυτές τις δύο αρχικοποιήσεις και τελικά επιλέγουμε την αρχικοποίηση που επιφέρει την μικρότερη ομαλοποιημένη ενέργεια $E(\lambda, p)$.

Το Σχήμα 6.12 παρουσιάζει παραδείγματα από το ομαλοποιημένο ταίριασμα του μοντέλου ΣΕ του κυρίαρχου χεριού στο ίδιο τμήμα βίντεο με το παράδειγμα του Σχήματος 6.9. Παρατηρούμε ότι σε πλαίσια χωρίς αλληλεπικαλύψεις του χεριού αυτού, αυτή η ομαλοποιημένη μέθοδος συνεχίζει να είναι αποτελεσματική και να παρακολουθεί με ακρίβεια τη χειρομορφή, όπως και η άμεση μέθοδος. Παρουσιάζει όμως το επιπλέον θετικό ότι, σε περιπτώσεις αλληλεπικαλύψεων, ακόμα και ύστερα από αρκετά πλαίσια, το αποτέλεσμα είναι εύρωστο. Από την άλλη όμως, σε σύγκριση με τις περιπτώσεις χωρίς αλληλεπικαλύψεις, η ακρίβεια της εξαγμένης χειρομορφής είναι μικρότερη, διότι οι όροι πρότερων πληροφοριών κάνουν το αποτέλεσμα να παραμένει πιο κοντά στην μέση εικόνα A_0 του ΜΣΕ.

6.3.5.4 Κατάτμηση Βίντεο για το Παράλληλο Ταίριασμα του ΜΣΕ

Παρουσιάζουμε εδώ μία διαδικασία για την διαίρεση ενός βίντεο εισόδου σε τμήματα, έτσι ώστε στη συνέχεια το ομαλοποιημένο ταίριασμα του ΜΣΕ να εφαρμόζεται αποτελεσματικά και αποδοτικά σε όλα τα πλαίσια του βίντεο. Για να σχεδιάσουμε την διαδικασία αυτή, θέτουμε κάποιες προδιαγραφές. Πιο συγκεκριμένα θέλουμε:

- τα τμήματα να ξεκινούν από πλαίσια στα οποία το μοντελοποιούμενο χέρι να μην



Σχήμα 6.12: Ομαλοποιημένο ταίριασμα του μοντέλου ΣΕ στο ίδιο τμήμα βίντεο με το παράδειγμα του Σχήματος 6.9. Σε κάθε αρχική εικόνα πλαισίου, υπερθέτουμε την ανακατασκευή βασισμένη στο μοντέλο, $A_0(W_p^{-1}(\mathbf{x})) + \sum \lambda_i A_i(W_p^{-1}(\mathbf{x}))$. Στην πάνω δεξιά γωνία δείχνουμε κάθε φορά την ίδια ανακατασκευή, αλλά στον χώρο του μοντέλου ΣΕ Ω_M , $A_0(\mathbf{x}) + \sum \lambda_i A_i(\mathbf{x})$, η οποία καθορίζει τα βέλτιστα βάρη λ . Το αποτέλεσμα είναι εύρωστο ακόμα και σε περιπτώσεις αλληλεπικαλύψεων σχετικά μεγάλης χρονικής διάρκειας.

βρίσκεται σε αλληλεπικάλυψη, έτσι ώστε να μπορούμε να εμπιστευτούμε την εικόνα ΣΕ εισόδου,

- ο αριθμός των πλαισίων που χρησιμοποιούνται σαν πλαίσια εκκίνησης να είναι όσο το δυνατόν μικρότερος, μιας και σε αυτές τις περιπτώσεις δεν μπορούμε να αξιοποιήσουμε τις πρότερες δυναμικές πληροφορίες (εξαιτίας έλλειψης πληροφορίας από προηγούμενα αποτελέσματα του αντίστοιχου τμήματος βίντεο) και το ταίριασμα του

ΜΣΕ είναι υπολογιστικά πιο απαιτητικό (μιας και η αρχικοποίηση βρίσκεται σχετικά μακριά από την λύση),

- η διάρκεια των αλληλεπικαλύψεων σε κάθε τμήμα να είναι όσο το δυνατόν μικρότερη, μιας και αυτό μειώνει τον κίνδυνο της συσσώρευσης σφάλματος κατά το ακολουθιακό ταίριασμα.

Βασισμένοι στις παραπάνω προδιαγραφές, η κατάτμηση του βίντεο γίνεται ως εξής. Εξάγονται όλα τα τμήματα του βίντεο με συνεχόμενα μη-επικαλυπτόμενα καθώς και όλα τα τμήματα με συνεχόμενα επικαλυπτόμενα πλαίσια και προσδιορίζεται το μεσαίο πλαίσιο κάθε τέτοιου τμήματος. Για κάθε τμήμα μη-επικαλυπτόμενων πλαισίων, ξεκινάμε από το μεσαίο πλαίσιο και βγάζουμε στην έξοδο 1) ένα τμήμα με κατεύθυνση προς τα εμπρός, τερματίζοντας στο μέσο πλαίσιο του επόμενου τμήματος επικαλυπτόμενων πλαισίων και 2) ένα τμήμα με κατεύθυνση προς τα πίσω, τερματίζοντας ακριβώς μετά το μέσο πλαίσιο του προηγούμενου τμήματος επικαλυπτόμενων πλαισίων.

Τα τμήματα του βίντεο με διαδοχικά πλαίσια που κατασκευάζονται με τον παραπάνω αλγόριθμο εισάγονται στην συνέχεια σαν είσοδοι στο ακολουθιακό ομαλοποιημένο ταίριασμα του ΜΣΕ. Αυτό επιτρέπει ένα υψηλό βαθμό παραλληλοποίησης.

6.4 Πειράματα Ταξινόμησης Χειρομορφών

Για να αποτιμηθεί η αποτελεσματικότητα της προτεινόμενης Αφινικής Μοντέλοποίησης Σχήματος-Εμφάνισης στην εξαγωγή χαρακτηριστικών (feature extraction) των χειρομορφών, πραγματοποιούμε πειράματα επιβλεπόμενης ταξινόμησης των χειρομορφών. Ας σημειωθεί ότι τα πειράματα αυτά αποτιμούν την δυνατότητα του προτεινόμενου μοντέλου να εξάγει μία περιγραφική αναπαράσταση των χειρομορφών του νοηματιστή και όχι την αποτελεσματικότητα της αρχικής παρακολούθησης των χειρών ή του ταιριάσματος του SAM μοντέλου σε βίντεο νοηματικής γλώσσας. Υπενθυμίζουμε ότι στην προτεινόμενη μοντελοποίηση, το διάνυμα χαρακτηριστικών χειρομορφής σχηματίζεται από τους συντελεστές λ των ιδιοεικόνων.

6.4.1 Δεδομένα, Επισημείωση και Κλάσεις Χειρομορφών

Όλες οι ακολουθίες εικόνων χειρομορφών που χρησιμοποιούνται προέρχονται από την βάση BU400 [45] η οποία περιέχει καταγραφές με συνεχόμενο νοηματισμό σε Αμερικανική Νοηματική Γλώσσα. Τα αρχικά έγχρωμα βίντεο έχουν ανάλυση 648×484 εικονοστοιχεία και συχνότητα δειγματοληψίας 60 πλαίσια ανά δευτερόλεπτο. Από την βάση αυτή, χρησιμοποιούμε 6 βίντεο με τα εξής ονόματα: `accident`, `biker_buddy`, `boston_la`, `football`, `lapd_story`, `siblings`. Τα βίντεο αυτά περιέχουν ιστορίες τις οποίες διηγείται ένας συγκεκριμένος νοηματιστής.

Από το σύνολο των κατατμημένων χειρομορφών των πλαισίων αυτών των βίντεο, διαλέγουμε, ύστερα από υποκειμενική επισκόπηση, χαρακτηριστικές περιπτώσεις διαμορφώσεων χειρομορφών και καταλήγουμε σε 2223 χειρομορφές.

Επιλέγουμε τις χειρομορφές έτσι ώστε 1) να καλύπτουν ικανοποιητικά την διακύμανση που παρατηρείται στο σύνολο των δεδομένων και 2) η συχνότητα εμφάνισής τους να είναι αρκετά μεγάλη. Στη συνέχεια, χωρίζουμε τυχαία τις επιλεγμένες χειρομορφές σε

υποσύνολα εκπαίδευσης και δοκιμής ώστε να περιέχουν 75% και 25% του συνόλου αντι-στοίχως.

6.4.2 Οργάνωση Πειραμάτων

Εκτελούμε δύο ξεχωριστά σύνολα πειραμάτων, τα οποία διαφέρουν ως προς τις θεωρούμενες τάξεις χειρομορφών. Η δομή των πειραμάτων αυτών έχει ως εξής:

Πείραμα-Α. Αυτό το πείραμα ταξινόμησης περιέχει έναν μικρό αριθμό 14 τάξεων. Αυτές τις επιλέγουμε κρατώντας για κάθε διαφορετική χειρομορφή παραδείγματα από μία μόνο αντιπροσωπευτική πόζα της. Το Πείραμα-Α είναι ένα σχετικά απλό πρόβλημα ταξινόμησης, του οποίου τα αποτελέσματα θέλουμε να αξιοποιήσουμε σαν ένα μέτρο σύγκρισης για τα αποτελέσματα του πιο σύνθετου Πειράματος-Β. Το Σχήμα 6.13 περιέχει μία ενδεικτική δειγματοληψία των 14 τάξεων που θεωρήθηκαν στο πείραμα αυτό.

Πείραμα-Β. Σε αυτό το πιο ρεαλιστικό σενάριο, *επεκτείνουμε* τόσο 1) το σύνολο των τάξεων αυξάνοντας τον αριθμό τους σε 28 όσο και 2) την μεταβλητότητα των περισσότερων τάξεων, προσθέτοντας συμπληρωματικές υποκλάσεις που περιέχουν *μη αμελητέα* διακύμανση στην πόζα της χειρομορφής. Το Σχήμα 6.14 περιέχει μία ενδεικτική δειγματοληψία όλων των 28 τάξεων που θεωρήθηκαν στο Πείραμα-Β. Η επίδειξη πολλαπλών εικόνων στην ίδια τάξη (ίδια στήλη), υποδηλώνει ότι στην συγκεκριμένη τάξη έχουν συμπεριληφθεί πολλαπλές πόζες της ίδιας χειρομορφής.



Σχήμα 6.13: Τάξεις χειρομορφών για το **Πείραμα-Α**: Όνομα και αντιπροσωπευτικό δείγμα κάθε κλάσης.

Ονοματολογία των τάξεων. Ας σημειωθεί ότι τα ονόματα που χρησιμοποιούμε για τις τάξεις των χειρομορφών περιέχουν ένα συνθετικό (π.χ. το Α στο όνομα “ΑΠ2”) το οποίο είναι *παρόμοιο* με την ονοματοδοσία των χειρομορφών της Αμερικανικής Νοηματικής Γλώσσας. Επίσης, περιέχουν ένα δεύτερο συνθετικό (π.χ. το Π2 στο όνομα “ΑΠ2”) το οποίο αποτελεί μία υποκειμενική κωδικοποίηση της πόζας της χειρομορφής. Επομένως, η ονοματολογία που χρησιμοποιούμε αποτελεί απλά μία σύμβαση για τους σκοπούς της παρουσίασης και δεν έχει σχέση με την επίσημη ονοματολογία της Αμερικανικής Νοηματικής Γλώσσας.



Σχήμα 6.14: Τάξεις χειρομορφών για το **Πείραμα-Β**: Όνομα και αντιπροσωπευτικά δείγματα κάθε κλάσης που επιδεικνύουν την μεταβλητότητα στην πόζα.

6.4.3 Συγκρινόμενες Μέθοδοι Εξαγωγής Χαρακτηριστικών Χειρομορφών

Για τον σκοπό των συγκρίσεων, εξάγουμε χαρακτηριστικά από τις χειρομορφές χρησιμοποιώντας την προτεινόμενη αλλά και διάφορες άλλες μεθόδους. Σε όλες τις περιπτώσεις, η ταξινόμηση πραγματοποιείται επιλέγοντας την τάξη που παρέχει την μέγιστη πιθανοφάνεια για κάθε χειρομορφή. Για κάθε μέθοδο, οι πιθανοφάνειες αυτές υπολογίζονται χρησιμοποιώντας μία πολυμεταβλητή γκαουσιανή κατανομή για κάθε κλάση, οι παράμετροι της οποίας προκύπτουν από τα στατιστικά στοιχεία (μέσος και μήτρα συμμεταβλητότητας) των χαρακτηριστικών που εξάγονται από τις χειρομορφές εκπαίδευσης της κάθε κλάσης. Ακολούθως, παρουσιάζονται οι συγκρινόμενες μέθοδοι εξαγωγής χαρακτηριστικών:

- **Αφινική Μοντελοποίηση Σχήματος-Εμφάνισης (Αφφ-ΜΣΕ).** Πρόκειται για την προτεινόμενη μέθοδο, την οποία εφαρμόζουμε δοκιμάζοντας διάφορες τιμές για τις βασικές της παραμέτρους, δηλαδή για την τάξη του PCA μοντέλου (αριθμό ιδιοεικόνων) N_c και για την σταθερά παρασκηνίου C_b που χρησιμοποιείται στις εικόνες ΣΕ.
- **Μοντελοποίηση Σχήματος-Εμφάνισης με Απευθείας Μετασχηματισμούς Ομοιότητας (ΜΣΕ-ΑΜΟ).** Πρόκειται για μία απλοποίηση του προτεινόμενου μο-

ντέλου, στην οποία χρησιμοποιούμε μετασχηματισμούς ομοιότητας (μετατόπιση-κλιμάκωση-περιστροφή) αντί για αφινικούς και οι παράμετροι των μετασχηματισμών αυτών εκτιμώνται απευθείας (χωρίς κάποια διαδικασία βελτιστοποίησης) χρησιμοποιώντας τις ακόλουθες απλές γεωμετρικές ποσότητες της μάσκας του χεριού: το κεντροειδές, το εμβαδόν και τον προσανατολισμό του πρωτεύοντα άξονά της. Όσον αφορά τον τύπο των μετασχηματισμών και τον τρόπο εκτίμησής τους καθώς και την στατιστική μοντελοποίηση μέσω PCA, αυτή η απλοποίηση είναι παρόμοια με την προσέγγιση του άρθρου [17].

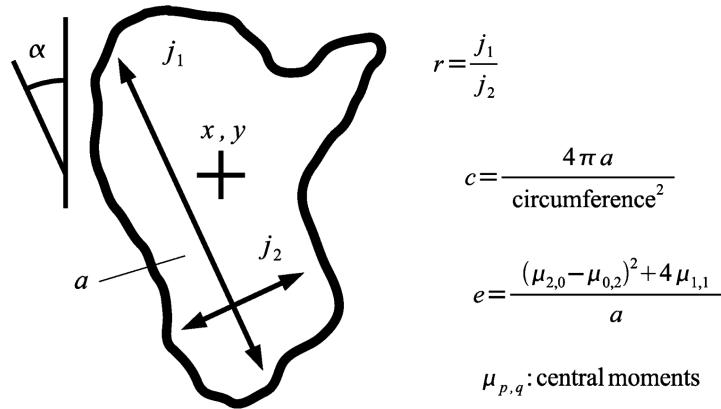
- **Μοντελοποίηση Σχήματος-Εμφάνιση με Απευθείας Μετασχηματισμούς Μετατόπισης & Κλιμάκωσης (ΜΣΕ-ΑΜΜΚ).** Αυτή η μέθοδος αποτελεί μία περαιτέρω απλοποίηση της προτεινόμενης μεθόδου, κατά την οποία μοντελοποιούνται μόνο μετασχηματισμοί μετατόπισης & κλιμάκωσης και οι παράμετροί τους εκτιμώνται απευθείας χρησιμοποιώντας απλώς το τετράγωνο με κατακόρυφες και οριζόντιες πλευρές που εφάπτεται ακριβώς στην δυαδική μάσκα του χεριού. Ως προς τον τύπο και τρόπο χρήσης των μετασχηματισμών και της PCA, αυτή η απλοποίηση παρουσιάζει ομοιότητες με τις μεθόδους των άρθρων [39, 140].

Εφαρμόζουμε επίσης τις ακόλουθες βασικές μεθόδους εξαγωγής χαρακτηριστικών σχήματος που χρησιμοποιούνται συχνά στην Αυτόματη Αναγνώριση Νοηματικής Γλώσσας:

- **Περιγραφητές Fourier (PF),** οι οποίοι προκύπτουν από τους συντελεστές Fourier του περιγράμματος του χεριού μετά από κατάλληλες κανονικοποιήσεις που τους κάνουν αναλλοίωτους σε μετατόπιση, κλιμάκωση και περιστροφή [34]. Αυτή η μέθοδος έχει σαν βασική παράμετρο τον αριθμό N_π των περιγραφητών, οι οποίοι εξάγονται χρησιμοποιώντας τους συντελεστές των αρμονικών με τις μικρότερες συχνότητες. Στα πειράματα, δοκιμάζουμε διαφορετικές τιμές για την παράμετρο αυτή.
- **Ροπές Hu,** οι οποίες είναι 7 μετρήσεις που εξαρτώνται μόνο από τις κεντρικές ροπές ενός δυαδικού σχήματος και είναι αναλλοίωτες ως προς μετασχηματισμούς ομοιότητας του σχήματος [60].
- **Γεωμετρικά χαρακτηριστικά περιοχής (ΓΧΠ) του χεριού,** τα οποία συνίστανται από τις ακόλουθες μετρήσεις της μάσκας του χεριού: εμβαδόν a , εκκεντρότητα e , βαθμός συμπύκνωσης c , μήκη j_1, j_2 του ελάσσονος και μείζονος άξονα του χεριού (βλ. Σχήμα 6.15) [1].

6.4.4 Αποτελέσματα και Συγκρίσεις

Στο Σχήμα 6.16(a) φαίνονται τα αποτελέσματα ταξινόμησης στα Πειράματα A και B χρησιμοποιώντας την προτεινόμενη μέθοδο με διαφορετικές τιμές για τις βασικές της παραμέτρους N_c και C_b . Όπως ήταν αναμενόμενο, η αύξηση της τάξης N_c του PCA μοντέλου οδηγεί σε ακριβέστερη μοντελοποίηση και άρα σε βελτιωμένη απόδοση. Από την άλλη, βλέπουμε ότι η απόδοση δεν εξαρτάται σημαντικά από την σταθερά παρασκηνίου C_b . Παρατηρώντας τα αποτελέσματα της προτεινόμενης μεθόδου στο Πείραμα-B, βλέπουμε

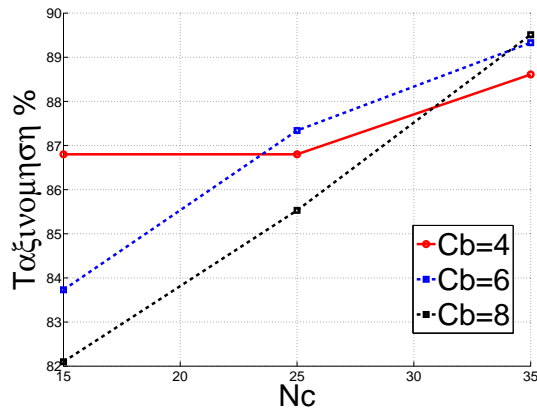


Σχήμα 6.15: Γεωμετρικά χαρακτηριστικά περιοχής (ΓΧΠ) του χεριού [1].

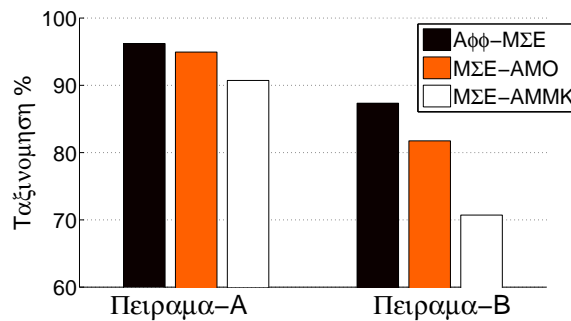
ότι η ταξινόμηση συγχέει περισσότερο κλάσεις που είναι ‘κοντινές’, υπό την έννοια ότι οι αντίστοιχες εικόνες ΣΕ είναι παρόμοιες. Τέτοιου είδους σύνολα τάξεων τα στοιχεία των οποίων η ταξινόμηση τείνει να μπερδεύει είναι τα ακόλουθα: ΙΠ9-ΙΠ2, ΙΠ4-ΒΠ10 και ΑΠ4-ΑΠ5-ΑΠ2.

Στο Σχήμα 6.16(β) συγκρίνονται τα αποτελέσματα ταξινόμησης για την προτεινόμενη (Αφφ-ΜΣΕ) και τις απλοποιημένες εκδοχές (ΜΣΕ-ΑΜΟ, ΜΣΕ-ΑΜΜΚ) της μοντελοποίησης Σχήματος-Εμφάνισης. Παρατηρούμε ότι στο Πείραμα-Β η Αφφ-ΜΣΕ παρουσιάζει μία απόλυτη βελτίωση 5% και 16% έναντι της ΜΣΕ-ΑΜΟ και της ΜΣΕ-ΑΜΜΚ αντίστοιχα, ενώ στο απλούστερο Πείραμα-Α οι βελτιώσεις αυτές είναι 1% και 5%.

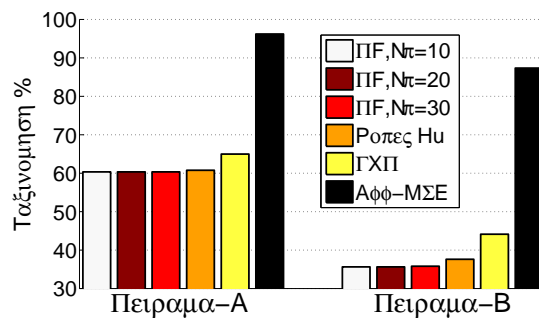
Το Σχήμα 6.16(γ) παρουσιάζει συγκρίσεις ταξινόμησης με την προτεινόμενη μέθοδο και τις βασικές μεθόδους Περιγραφητών Fourier, Ροπών Hu και ΓΧΠ του χεριού. Βλέπουμε ότι η προτεινόμενη μέθοδος προσφέρει βελτιώσεις οι οποίες κατά μέσο όρο αγγίζουν το 50% για το επεκταμένο Πείραμα-Β.



(α) Σύγκριση διαφορετικών παραμέτρων για την προτεινόμενη μέθοδο Αφφινικής Μοντελοποίησης Σχήματος-Εμφάνισης (Αφφ-ΜΣΕ), μεταβάλλοντας την τάξη N_c του PCA μοντέλου και την σταθερά παρασκηνίου C_b στο Πείραμα-B



(β) Σύγκριση της προτεινόμενης Αφφ-ΜΣΕ με τις απλούστερες εκδοχές ΜΣΕ με Απευθείας Μετασχηματισμούς Ομοιότητας (ΜΣΕ-ΑΜΟ) και ΜΣΕ με Απευθείας Μετασχηματισμούς Μετατόπισης & Κλιμάκωσης (ΜΣΕ-ΑΜΜΚ), για τα Πειράματα Α και Β



(γ) Σύγκριση της προτεινόμενης Αφφ-ΜΣΕ με τις βασικές μεθόδους Περιγραφητών Fourier (ΠΦ), Ροπών Hu και Γεωμετρικών Χαρακτηριστικών Περιοχής (ΓΧΠ) του χεριού, για τα Πειράματα Α και Β

Σχήμα 6.16: Αποτελέσματα ταξινόμησης χειρομορφών.

Κεφάλαιο 7

Σύνοψη Προόδου και Κατευθύνσεις Μελλοντικής Έρευνας

7.1 Ερευνητική Συνεισφορά

Η διδακτορική μας έρευνα κινείται στην περιοχή της Όρασης Υπολογιστών και ακολουθεί μεθοδολογίες που βασίζονται στις Μερικές Διαφορικές Εξισώσεις και τα Στατιστικά Μοντέλα Σχήματος. Οι κύριες εφαρμογές της αφορούν την ανάλυση οπτικών δεδομένων αρθρωτών φωνητικού και νοηματικού λόγου, η οποία χρησιμεύει στην έρευνα τόσο του ανθρώπινου συστήματος παραγωγής φωνής όσο και της νοηματικής γλώσσας. Οι ερευνητικές μας συνεισφορές μπορούν να συνοψισθούν στα ακόλουθα σημεία :

- Ανάπτυξη πρωτότυπων μεθόδων μη-γραμμικής διάχυσης για την βελτίωση της ποιότητας εικόνων :
 - Προτείνουμε μία μέθοδο ανισοτροπικής διάχυσης - προβολής για την βελτίωση της ευκρίνειας γενικών διανυσματικών εικόνων, η οποία αποδίδει αποτελέσματα υψηλής ακρίβειας, με περιορισμένα ανεπιθύμητα ψεύδεργα (artifacts).
 - Εισάγαμε ένα θεωρητικό πλαίσιο το οποίο βασίζεται στον δομικό τανυστή της εικόνας και γενικεύει διάφορες μεταβολικές μεθόδους μη-γραμμικής διάχυσης. Βασισμένοι στο πλαίσιο αυτό, προτείνουμε νέες μεθόδους διάχυσης, οι οποίες συνδυάζουν τα πλεονεκτήματα των μεταβολικών μεθόδων και των μεθόδων που βασίζονται σε τανυστές.
- Σχεδίαση στατιστικών μοντέλων σχήματος για την παρακολούθηση και ταξινόμηση αντικειμένων :
 - Προτείνουμε ένα στατιστικό μοντέλο για την καμπύλη της γλώσσας ομιλητή, το οποίο ενσωματώνει πρότερη γνώση σχήματος και μπορεί να εκτιμήσει εύρωστα ολόκληρη την καμπύλη της γλώσσας (ακόμα και τα μη ορατά κομμάτια) σε ακολουθίες εικόνων υπερήχων.
 - Εισάγαμε μία αφινικά αναλλοίωτη μοντελοποίηση των χειρομορφών νοηματικής γλώσσας που βασίζεται στις λεγόμενες εικόνες Σχήματος-Εμφάνισης και

προσφέρει μία συμπαγή αναπαράσταση των διαμορφώσεων του χεριού. Μέσω της μη-γραμμικής βελτιστοποίησης μιας ομαλοποιημένης (regularized) ενέργειας, πετυχαίνουμε εύρωστο ταίριασμα του μοντέλου σε βίντεο συνεχούς νοηματισμού. Το ταίριασμα αυτό οδηγεί σε χαρακτηριστικά χειρομορφής που χρησιμοποιήθηκαν με ιδιαίτερη επιτυχία στην κατηγοριοποίηση χειρομορφών.

Υπό την οπτική γωνία των εφαρμογών, με την παρούσα διδακτορική διατριβή συνεισφέραμε στην έρευνα για την *Αρθρωτή Σύνθεση Φωνής* καθώς και την *Οπτικοακουστική Αντιστροφή Φωνής* μέσω της εφαρμογής καινοτόμων αλγορίθμων επεξεργασίας και ανάλυσης βιοϊατρικών οπτικών δεδομένων της φωνητικής οδού κατά τη διάρκεια της ομιλίας. Επίσης, συνεισφέραμε στην έρευνα για την *Αυτόματη Αναγνώριση Νοηματικής Γλώσσας* μέσω της ανάπτυξης μοντέλων για την εύρωστη παρακολούθηση των χειρών του νοηματοπιστή και την εξαγωγή χαρακτηριστικών (features) χειρομορφών, σε συνθήκες συνεχούς νοηματισμού. Τέλος, οι μέθοδοι μη-γραμμικής διάχυσης για την βελτίωση της ποιότητας εικόνων που αναπτύξαμε εφαρμόζονται με ιδιαίτερη επιτυχία και στην *Επεξεργασία Φωτογραφιών*, όπως διαφάνηκε από εκτενή πειράματα παρεμβολής και αποθορυβοποίησης.

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

7.2 Κατευθύνσεις Μελλοντικής Έρευνας

Με αφορμή την παρούσα διδακτορική έρευνα, διαγράφονται διάφορες πιθανές κατευθύνσεις για μελλοντική έρευνα. Ενδεικτικά αναφέρουμε τα εξής:

- Λόγω της αποτελεσματικότητας της μεθόδου μη-γραμμικής διάχυσης για την παρεμβολή εικόνων που προτείναμε στο Κεφάλαιο 3, ιδιαίτερο ενδιαφέρον παρουσιάζει η ανάπτυξη αποδοτικών αλγορίθμων υλοποίησης της μεθόδου αυτής. Στην παρούσα εργασία, βασιστήκαμε σε ένα άμεσο (explicit) αριθμητικό σχήμα με πεπερασμένες διαφορές, αλλά αντί αυτού θα μπορούσαν να χρησιμοποιηθούν μεθοδολογίες όπως αυτές π.χ. που αναφέρονται στα [93, 137].
- Το γενικευμένο μεταβολικό πλαίσιο που εισάγαμε στο Κεφάλαιο 4 θα μπορούσε να βρει διάφορες εφαρμογές. Προσφέρει ένα θεωρητικό εργαλείο ώστε οι μέθοδοι διάχυσης που περιέχουν συνελίζεις χωρικής ομαλοποίησης να αποκτήσουν πιο γερά θεμέλια και να επεκταθεί η δυνατότητα εφαρμογής τους σε διάφορα προβλήματα της Όρασης Υπολογιστών. Για παράδειγμα, θα μπορούσε να οδηγήσει σε μεταβολικές μεθόδους ανισοτροπικής διάχυσης με ομαλοποίηση σε προβλήματα όπως αποκατάσταση, αναίρεση θολώματος (deblurring), παρεμβολή και ενδοσυμπλήρωση (inpainting) εικόνων, χρησιμοποιώντας την διατύπωση των προβλημάτων αυτών σαν συναρτησιακή βελτιστοποίηση υπό περιορισμούς.
- Οι μεθοδολογίες μη-γραμμικής διάχυσης εικόνων που αναπτύξαμε στα Κεφάλαια 3 και 4 περιέχουν ένα σημαντικό μέρος μαθηματικής μοντελοποίησης. Παρόλα αυτά, διάφορα ενδιαφέροντα θεωρητικά ζητήματα παραμένουν ανοικτά, όπως για παράδειγμα ύπαρξη, μοναδικότητα και ευστάθεια της λύσης της προτεινόμενης ΜΔΕ για την παρεμβολή εικόνων, ή συνθήκες ύπαρξης και μοναδικότητας για το γενικευμένο συναρτησιακό του Κεφαλαίου 4.

- Η μέθοδος παρακολούθησης της γλώσσας σε εικόνες υπερήχων που προτείναμε στο Κεφάλαιο 5 θα μπορούσε να γίνει πιο εύρωστη μέσω της ενσωμάτωσης πρότερης στατικής και δυναμικής πληροφορίας σε ένα πλαίσιο εργασίας αντίστοιχο με αυτό που ακολουθήσαμε για την παρακολούθηση του χεριού του νοηματιστή (βλ. Ενότητα 6.3.5).
- Σχετικά με το πλαίσιο της μοντελοποίησης χειρομορφών νοηματικής γλώσσας που προτείναμε στο Κεφάλαιο 6, υπάρχουν διάφορες ενδιαφέρουσες κατευθύνσεις επέκτασής του. Πρώτα από όλα, θα μπορούσε να τροποποιηθεί έτσι ώστε να επιτυγχάνεται ανεξαρτησία από τον νοηματιστή και τις συνθήκες καταγραφής του βίντεο. Ακόμα, είναι ενδιαφέρουσα η διερεύνηση της εφαρμογής της μοντελοποίησης αυτής εκτός από το κυρίαρχο χέρι, στο άλλο χέρι και στο κεφάλι του νοηματιστή καθώς επίσης η αλληλεπίδραση μεταξύ των τριών αυτών μοντέλων και η συνδυασμένη παρακολούθηση των τριών μερών του σώματος. Επίσης, το πλαίσιο αυτό θα μπορούσε να ενσωματωθεί σε ένα σύστημα αυτόματης αναγνώρισης συνεχούς νοηματικής γλώσσας.

Βιβλιογραφία

- [1] U. Agris, J. Zieren, U. Canzler, B. Bauer, and K. F. Kraiss. Recent developments in visual sign language recognition. *Universal Access in the Information Society*, 6:323–362, 2008.
- [2] T. Ahmad, C.J. Taylor, and T.F. Lanitis, A. Cootes. Tracking and recognising hand gestures, using statistical shape models. *Image and Vision Computing*, 15(5):345–352, 1997.
- [3] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations in image processing. *Arch. Rational Mech. Anal.*, 123:199–257, 1993.
- [4] H. A. Aly and E. Dubois. Image up-sampling using total-variation regularization with a new observation model. *IEEE Trans. on Image Processing*, 14(10):1647–1659, 2005.
- [5] A. Argyros and M. Lourakis. Real time tracking of multiple skin-colored objects with a possibly moving camera. In *Proc. Europ. Conf. on Computer Vision*, 2004.
- [6] M. Aron. *Acquisition et modélisation de données articulatoires dans un contexte multimodal*. PhD thesis, Université Henri Poincaré - Nancy 1, 2009. In french.
- [7] M. Aron, A. Roussos, M.-O. Berger, E. Kerrien, and P. Maragos. Multimodality acquisition of articulatory data and processing. In *Proc. Europ. Signal Processing Conference*, Lausanne, 2008.
- [8] M. Aron, A. Toutios, M.-O. Berger, E. Kerrien, B. Wrobel-Dautcourt, and Y. Laprie. Registration of multimodal data for estimating the parameters of an articulatory model. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2009.
- [9] V. Athitsos and S. Sclaroff. An appearance-based framework for 3d hand shape classification and camera viewpoint estimation. In *Proc. IEEE Int. Conf. on Automatic Face & Gesture Recognition*, pages 45–52, 2002.
- [10] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Math. Sciences*. Springer-Verlag, 2002.
- [11] S. Baker, R. Gross, and I. Matthews. Lucas-kanade 20 years on: A unifying framework: Part 4. Technical report, Carnegie Mellon University, 2004.

- [12] B. Bauer, H. Hienz, and K.-L. Kraiss. Video-based continuous sign language recognition using statistical methods. *Proc. Int. Conf. on Pattern Recognition*, 2000.
- [13] B. Bauer and K. F. Kraiss. Towards an automatic sign language recognition system using subunits. In *Proc. of Int. Gesture Workshop*, volume 2298, pages 64–75, 2001.
- [14] A. Belahmidi and F. Guichard. A partial differential equation approach to image zoom. In *Proc. Int. Conf. on Image Processing*, pages 649–652, 2004.
- [15] M. Bergtholdt, D. Cremers, and C. Schnörr. Variational segmentation with shape priors. In N. Paragios, Y. Chen, and O. Faugeras, editors, *Handbook of Mathematical Models in Computer Vision*. 2005.
- [16] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proc. SIGGRAPH*, pages 417–424, 2000.
- [17] H. Birk, T.B. Moeslund, and C.B. Madsen. Real-time recognition of hand alphabet gestures using principal component analysis. In *Proc. Scandinavian Conf. on Image Analysis*, 1997.
- [18] A. Blake and M. Isard. *Active contours*. Springer, 1998.
- [19] P. Blomgren and T.F. Chan. Color TV: Total Variation methods for restoration of vector-valued images. *IEEE Trans. on Image Processing*, 7(3):304–309, 1998. Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis.
- [20] R. Bowden and M. Sarhadi. A nonlinear model of shape and motion for tracking fingerspelt american sign language. *Image and Vision Computing*, 20:597–607, 2002.
- [21] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady. A linguistic feature vector for the visual interpretation of sign language. In *Proc. Europ. Conf. on Computer Vision*, 2004.
- [22] M.A. Brubaker, L. Sigal, and D.J. Fleet. Video-based people tracking. In H. Nakashima, H. Aghajan, and J.C. Augusto, editors, *Handbook of Ambient Intelligence and Smart Environments*. Springer, 2009.
- [23] P. Buehler, M. Everingham, D.P. Huttenlocher, and A. Zisserman. Long term arm and hand tracking for continuous sign language TV broadcasts. In *Proc. British Machine Vision Conference*, 2008.
- [24] P. Buehler, M. Everingham, and A. Zisserman. Learning sign language by watching TV (using weakly aligned subtitles). In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2961–2968, June 2009.
- [25] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *Int. Journ. of Computer Vision*, 22(1):61–79, 1997.

- [26] V. Caselles, J. M. Morel, and C. Sbert. An axiomatic approach to image interpolation. *IEEE Trans. on Image Processing*, 7(3):376–386, 1998.
- [27] F. Catté, P.L. Lions, J.M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journ. Numer. Anal.*, 29(1):182–193, 1992.
- [28] A. Chambolle and P. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76:167–188, 1997.
- [29] T.F. Chan, S.H. Kang, and J. Shen. Total variation denoising and enhancement of color images based on the CB and HSV color. *Journ. Visual Communication and Image Representation*, 12(4):422–435, 2001.
- [30] T.F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM Journ. on Applied Mathematics*, 62(3):1019–1043, 2002.
- [31] T.F. Chan and C. Wong. Total variation blind deconvolution. *IEEE Trans. on Image Processing*, 7:370–375, 1998.
- [32] P. Charbonnier, G. Aubert, L. Blanc-Féraud, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proc. Int. Conf. on Image Processing*, volume II, pages 168–172, 1994.
- [33] F.-S. Chen, C.-M. Fu, and C.-L. Huang. Hand gesture recognition using a real-time tracking method and hidden markov models. *Image and Vision Computing*, 21(8):745–758, 2003.
- [34] S. Conseil, S. Bourennane, and L. Martin. Comparison of Fourier descriptors and Hu moments for hand posture recognition. In *Proc. Europ. Signal Processing Conference*, 2007.
- [35] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
- [36] T.F. Cootes and C.J. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, 2004.
- [37] T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham. Active shape models-Their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.
- [38] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion snakes: Introducing statistical shape knowledge into the mumford-shah functional. *Int. Journ. of Computer Vision*, 50(3):295–313, 2002.
- [39] Y. Cui and J. Weng. Appearance-based hand sign recognition from intensity image sequences. *Computer Vision and Image Understanding*, 78(2):157–176, 2000.

- [40] N. Dalal and B. Triggs. Histogram of oriented gradients for human detection. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005.
- [41] J.-W. Deng and H.T. Tsui. A novel two-layer PCA/MDA scheme for hand posture recognition. In *Proc. Int. Conf. on Pattern Recognition*, volume 1, pages 283–286, 2002.
- [42] F. Dornaika and F. Davoine. Simultaneous facial action tracking and expression recognition in the presence of head motion. *Int. Journ. of Computer Vision*, 76:257–281, 2008.
- [43] A.C. Downton and H. Drouet. Model-based image analysis for unconstrained human upper-body motion. In *International Conference on Image Processing and Its Applications*, pages 274–277, April 1992.
- [44] P. Dreuw, J. Forster, T. Deselaers, and H. Ney. Efficient approximations to model-based joint tracking and recognition of continuous sign language. In *Proc. IEEE Int. Conf. on Automatic Face & Gesture Recognition*, Amsterdam, The Netherlands, September 2008.
- [45] P. Dreuw, C. Neidle, V. Athitsos, S. Sclaroff, and H. Ney. Benchmark databases for video-based automatic sign language recognition. In *Proc. Int. Conf. on Language Resources and Evaluation (LREC)*, May 2008.
- [46] P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney. Speech recognition techniques for a sign language recognition system. In *Proc. Int. Conf. on Spoken Language Processing (Interspeech)*, pages 2513–2516, Antwerp, Belgium, August 2007.
- [47] I.L. Dryden and K.V. Mardia. *Statistical Shape Analysis*. John Wiley and Sons, 1998.
- [48] D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing*. Prentice-Hall, NJ, 1984.
- [49] O. Engwall. From real-time MRI to 3D tongue movements. In *Proc. Int. Conf. on Spoken Language Processing (Interspeech)*, pages 1109–1112, October 2004.
- [50] O. Engwall. *Tongue Talking - Studies in Intraoral Speech Synthesis*. PhD thesis, KTH, Sweden, 2002.
- [51] V. Ferrari, M. Marin-Jimenez, and A. Zisserman. 2D human pose estimation in TV shows. In *Proc. of the Dagstuhl Seminar on Statistical and Geometrical Approaches to Visual Motion Analysis*, 2009.
- [52] H. Fillbrandt, S. Akyol, and K.-F. Kraiss. Extraction of 3D hand shape and posture from images sequences from sign language recognition. In *Proc. Int. Workshop on Analysis and Modeling of Faces and Gestures*, pages 181–186, 2003.
- [53] J.L. Flanagan. *Speech Analysis Synthesis and Perception*. Springer-Verlag, Berlin, 1972.

- [54] M. Gasser. *How language works*. Indiana University, 3.0 edition, 2006.
- [55] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(12):1080–1093, 2005.
- [56] F. Guichard and F. Malgouyres. Total variation based interpolation. In *Proc. Europ. Signal Processing Conference*, volume 3, pages 1741–1744, 1998.
- [57] H.J.A.M. Heijmans. *Morphological Image Operators*. Acad. Press, Boston, 1994.
- [58] K. Hienz, H. Grobel and G. Offner. Real-time hand-arm motion analysis using a single video camera. In *Proc. Int. Conf. on Automatic Face and Gesture Recognition*, pages 323–327, 1996.
- [59] E.-J. Holden and R. Owens. Visual sign language recognition. In *Proc. Int. Workshop on Theoretical Foundations of Computer Vision*, pages 270–287, 2000.
- [60] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187, February 1962.
- [61] C.-L. Huang and S.-H. Jeng. A model-based hand gesture recognition system. *Machine Vision and Application*, 12(5):243–258, 2001.
- [62] I. Kakadiaris and D. Metaxas. Model-based estimation of 3D human motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12), December 2000.
- [63] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122, March 2007.
- [64] M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. Journ. of Computer Vision*, 1(4):321–331, January 1988.
- [65] A. Katsamanis. *Μη-Γραμμική Υπολογιστική Μοντελοποίηση Φωνής με Στοιχεία Αεροδυναμικής του Φωνητικού Σωλήνα*. Διδακτορική Διατριβή, Σχολή Η.Μ.Μ.Υ., Εθνικό Μετσόβιο Πολυτεχνείο, 2009.
- [66] S. Kichessammany, A. Kumar, PJ Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contours. In *Proc. Int. Conf. on Computer Vision*, 1995.
- [67] J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [68] P. Kornprobst, R. Deriche, and G. Aubert. Image coupling, restoration and enhancement via PDEs. *Proc. Int. Conf. on Image Processing*, pages 458–461, 1997.
- [69] Y. Laprie, P. Maragos, A. Roussos, M.-O. Berger, E. Kerrien, and M. Aron. Deliverable D3.2, Final report on design, acquisition and processing of articulatory data. Technical report, EU FET programme Audiovisual to Articulatory Speech Inversion (ASPI), January 2009.

- [70] M. Leventon, E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2000.
- [71] M. Li, X. Khambhamettu, and M. Stone. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics*, 6(19):545–554, 2005.
- [72] M. Li, X. Khambhamettu, and M. Stone. A level set approach for shape recovery of open contours. In *Proc. Asian Conf. on Computer Vision*, pages 601–611, 2006.
- [73] S. Liwicki and M. Everingham. Automatic recognition of fingerspelled words in British sign language. In *Proc. IEEE Workshop on Computer Vision and Pattern Recognition for Human Communicative Behavior Analysis*, 2009.
- [74] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Int. Joint Conf. on Artificial Intelligence*, pages 674–679, 1981.
- [75] S. Maeda. Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W.J. Hardcastle and A. Marchal, editors, *Speech Production and Speech Modeling*, pages 131–149. Kluwer Academic Publishers, 1990.
- [76] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *SIAM Journ. Numer. Anal.*, 39(1):1–37, 2001.
- [77] P. Maragos. Differential morphology and image-processing. *IEEE Trans. on Image Processing*, 5(6):922–937, June 1996.
- [78] P. Maragos. *Image Analysis and Computer Vision*. book draft, ver. 2005, NTUA, 2005.
- [79] P. Maragos. Morphological filtering for image enhancement and feature detection. In A.C. Bovik, editor, *The Image and Video Processing Handbook*. Elsevier, 2nd edition, 2005.
- [80] P. Maragos. A variational formulation of PDE’s for dilations and levelings. In *Proc. Int. Symposium on Mathematical Morphology*, April 2005.
- [81] P. Maragos, O. Engwall, Y. Laprie, S. Maeda, J. Schoentgen, A. Katsamanis, and A. Roussos. Deliverable D2.2, Final report on speech inversion methods. Technical report, EU FET programme Audiovisual to Articulatory Speech Inversion (ASPI), January 2009.
- [82] P. Maragos and G. Evangelopoulos. Leveling cartoons, texture energy markers, and image decomposition. In *Proc. Int. Symposium on Mathematical Morphology*, volume 1, pages 125–138, October 2007.
- [83] I. Matthews and S. Baker. Active appearance models revisited. *Int. Journ. of Computer Vision*, 60(2):135–164, 2004.

- [84] D. McNeill. Hand and mind: what gestures reveal about thought. *University of Chicago Press*, 1992.
- [85] E. Meijering. A chronology of interpolation: from ancient astronomy to modern signal and image processing. *Proc. IEEE*, 90(3):319–342, 2002.
- [86] F. Meyer and P. Maragos. Multiscale morphological segmentations based on watershed, flooding, and eikonal PDE. In *Lec. Notes Computer Science vol. 1682 (Proc. Scale-Space’99)*, pages 351–362. Springer-Verlag, 1999.
- [87] F. Meyer and P. Maragos. Nonlinear scale-space representation with morphological levelings. *Journ. Visual Communication and Image Representation*, 11:245–265, 2000.
- [88] N. Michael, C. Neidle, and D. Metaxas. Computer-based recognition of facial expressions in ASL: from face tracking to linguistic interpretation. In *Proc. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. LREC, Malta, May 2010*.
- [89] K. G. Munhall, E. Vatikiotis-Bateson, and Y. Tohkura. X-ray film database for speech research. *Journ. Acoustical Society of America*, 98:1222–1224, 1995.
- [90] A. W. Naylor and G. R. Sell. *Linear Operator Theory in Engineering and Science*. Springer-Verlag, 1982.
- [91] E.-J. Ong and R. Bowden. A boosted classifier tree for hand shape detection. In *Int. Conf. on Automatic Face and Gesture Recognition*, pages 889–894, 2004.
- [92] A. V. Oppenheim, A. S. Willsky, and I. T. Young. *Signals and Systems*. Prentice Hall, 1984.
- [93] G. Papandreou. *Ανάλυση Εικόνας και Όραση Υπολογιστών: Θεωρία και Εφαρμογές στην Αποκατάσταση Αρχαίων Τοιχογραφιών*. Διδακτορική Διατριβή, Σχολή Η.Μ.Μ.Υ., Εθνικό Μετσόβιο Πολυτεχνείο, 2009.
- [94] G. Papandreou and P. Maragos. Adaptive and constrained algorithms for inverse compositional active appearance model fitting. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, June 2008.
- [95] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.
- [96] L. R. Rabiner and R.W. Schafer. Introduction to digital speech processing. *Foundations and Trends in Signal Processing*, 1(1-2):1–194, 2007.
- [97] C. Rochette. *Les groupes de consonnes en français*. Les Presses de l’Université Laval, Québec, Canada, 1973. In french.
- [98] M. Rousson and N. Paragios. Prior knowledge, level set representations and visual grouping. *Int. Journ. of Computer Vision*, 76(3):231–243, 2008.

- [99] A. Roussos. *Θεωρία και Εφαρμογές Μερικών Διαφορικών Εξισώσεων στην Όραση Υπολογιστών*. Διπλωματική Εργασία, Σχολή Η.Μ.Μ.Υ., Εθνικό Μετσόβιο Πολυτεχνείο, 2005.
- [100] A. Roussos, A. Katsamanis, and P. Maragos. Tongue tracking in ultrasound images with active appearance models. In *Proc. Int. Conf. on Image Processing*, November 2009.
- [101] A. Roussos and P. Maragos. Vector-valued image interpolation by an anisotropic diffusion-projection PDE. In *Proc. Int. Conf. on Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 104–115. Springer-Verlag, 2007.
- [102] A. Roussos and P. Maragos. Reversible interpolation of vectorial images by an anisotropic diffusion-projection PDE. *Int. Journ. of Computer Vision*, 84(2), August 2009.
- [103] A. Roussos and P. Maragos. Tensor-based image diffusions derived from generalizations of the total variation and beltrami functionals. In *Proc. Int. Conf. on Image Processing*, September 2010.
- [104] A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Affine-invariant modeling of shape-appearance images applied on sign language handshape classification. In *Proc. Int. Conf. on Image Processing*, September 2010.
- [105] A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Hand tracking and affine shape-appearance handshape sub-units in continuous sign language recognition. In *Proc. of Workshop on Sign, Gesture and Activity, 11th ECCV*, September 2010.
- [106] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [107] G. Sapiro. Color snakes. *Comp. Vision and Image Understanding*, 68(2), 1997.
- [108] G. Sapiro and D.L. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Trans. on Image Processing*, 5(11):1582–1585, 1996.
- [109] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, NY, 1982.
- [110] J. A. Sethian. *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, 2nd edition, 1999.
- [111] J. Sherrah and S. Gong. Resolving visual uncertainty and occlusion through probabilistic reasoning. In *Proc. British Machine Vision Conference*, pages 252–261, 2000.

- [112] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Trans. on Image Processing*, 7:310–338, 1998.
- [113] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer, 2004.
- [114] T. Starner, J. Weaver, and A. Pentland. Real-time american sign language recognition using desk and wearable computer based video. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375, December 1998.
- [115] B. Stenger, P.R.S. Mendonca, and R. Cipolla. Model-based 3D tracking of an articulated hand. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.
- [116] B. Stenger, A. Thayananthan, P.H.S Torr, and R. Cipolla. Model-based hand tracking using a hierarchical bayesian filter. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(9):1372–1384, September 2006.
- [117] K.N. Stevens. *Acoustic Phonetics*. The MIT Press, 1998.
- [118] G.J. Sweeney and A.C. Downton. Towards appearance-based multi-channel gesture recognition. In *Proc. Gesture Workshop on Progress in Gestural Interaction*, pages 7–16, 1996.
- [119] N. Tanibata, N. Shimada, and Y. Shirai. Extraction of hand features for recognition of sign language words. In *Proc. Int. Conf. on Vision Interface*, pages 391–398, 2002.
- [120] G.W. Taylor, L. Sigal, D.J. Fleet, and G Hinton. Dynamical binary latent variable models for 3D human pose tracking. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2010.
- [121] J. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *Proc. IEEE Int. Conf. on Automatic Face & Gesture Recognition*, pages 54–61, 2000.
- [122] D. Tschumperlé. *PDE's Based Regularization of Multivalued Images and Applications*. PhD thesis, Université de Nice-Sophia Antipolis, December 2002.
- [123] D. Tschumperlé and L. Brun. Non-local image smoothing by applying anisotropic diffusion pde's in the space of patches. In *Proc. Int. Conf. on Image Processing*, 2009.
- [124] D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDE's : A common framework for different applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(4):506–517, 2005.
- [125] M. Unser and M. Stone. Automated detection of the tongue surface in sequences of ultrasound images. *Journ. Acoustical Society of America*, 91(5):3001–3007, 1992.

- [126] R. Urtasun, D.J. Fleet, and P. Fua. Monocular 3D tracking of the golf swing. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume II, pages 932–938, 2005.
- [127] R. Van den Boomgaard and A. Smeulders. The morphological structure of images: The differential equations of morphological scale-space. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16:1101–1113, November 1994.
- [128] L. Vincent. Grayscale area openings and closings, their efficient implementation and applications. *1st Worksh. on Mathematical Morphology and its Applications to Signal Processing*, pages 22–27, 1993.
- [129] C. Vogler and S. Goldenstein. Analysis of facial expressions in american sign language. In *Proc. 3rd Int. Conf. on Universal Access in Human-Computer Interaction*. Citeseer, 2005.
- [130] C. Vogler and D. Metaxas. Adapting hidden markov models for asl recognition by using three-dimensional computer vision methods. In *Int. Conf. on Systems, Man, Cybernetics*, volume 1, pages 156–161, 1997.
- [131] Z. Wang, AC Bovik, HR Sheikh, and EP Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on Image Processing*, 13(4):600–612, 2004.
- [132] J. Weickert. Multiscale texture enhancement. In V. Hlavác and R. Sára, editors, *Computer Analysis of Images and Patterns*, volume 970, pages 230–237. Lecture Notes in Computer Science, Springer, Berlin, 1995.
- [133] J. Weickert. Coherence-enhancing diffusion of colour images. *7th National Symposium on Pattern Recognition and Image Analysis*, April 1997.
- [134] J. Weickert. A review of nonlinear diffusion filtering. *Scale-Space Theory in Computer Vision, Lecture Notes in Computer Science (Springer, Berlin)*, 1252:3–28, 1997. Invited Paper.
- [135] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner Stuttgart, 1998.
- [136] J. Weickert, S. Ishikawa, and A. Imiya. On the history of gaussian scale-space axiomatics. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian scale-space theory*, pages 45–59. Kluwer, Dordrecht, 1997.
- [137] J. Weickert, B. ter Haar Romeny, and M. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Trans. on Image Processing*, 7(3):398–410, March 1998.
- [138] J. Weickert and M. Welk. Tensor field interpolation with PDEs. In J. Weickert and H. Hagen, editors, *Visualization and Processing of Tensor Fields*, pages 315–325. Springer, Berlin, 2006.
- [139] A.P. Witkin. Scale-space filtering. In *Proc. Int. Joint Conf. on Artificial Intelligence*, pages 1019–1022, 1983.

- [140] Y. Wu and T.S. Huang. View-independent recognition of hand postures. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 88–94, 2000.
- [141] M.-H. Yang, N. Ahuja, and M. Tabb. Extraction of 2d motion trajectories and its application to hand gesture recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(8):1061–1074, August 2002.
- [142] X. Zabulis, H. Baltzakis, and A. Argyros. Vision-based hand gesture recognition for human-computer interaction. In *The Universal Access Handbook*. LEA, 2009.
- [143] J. Zieren, N. Unger, and S. Akyol. Hands tracking from frontal view for vision-based gesture recognition. In *24th DAGM Symposium*, pages 531–539, 2002.

Παράρτημα Α΄

Κατάλογος Δημοσιεύσεων του Συγγραφέα

Αποτελέσματα της διδακτορικής μας έρευνας έχουν δημοσιευθεί σε διεθνώς αναγνωρισμένα περιοδικά και συνέδρια με κριτή. Ακολουθεί πλήρης κατάλογος των σχετικών δημοσιεύσεων. Ηλεκτρονικά ανάτυπα είναι διαθέσιμα από την ιστοσελίδα:

<http://cvsp.cs.ntua.gr/tassos>.

Δημοσιεύσεις σε Διεθνή Περιοδικά με Κριτές

1. A. Roussos and P. Maragos. Reversible interpolation of vectorial images by an anisotropic diffusion-projection PDE. *International Journal of Computer Vision*, 84(2), August 2009.
2. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. A dynamic affine-invariant shape-appearance model for handshape feature extraction and classification in sign language videos. Under preparation to be submitted to the *International Journal of Computer Vision*.

Δημοσιεύσεις σε Διεθνή Συνέδρια με Κριτές

1. A. Roussos and P. Maragos. Vector-valued image interpolation by an anisotropic diffusion-projection PDE. In *Scale Space and Variational Methods in Computer Vision, First International Conference, SSVN-2007 Proceedings*, volume 4485 of *Lecture Notes in Computer Science*, pages 104–115. Springer-Verlag, 2007.
2. M. Aron, A. Roussos, M.-O. Berger, E. Kerrien, and P. Maragos. Multimodality Acquisition of Articulatory Data and Processing. In *Proceedings of the IEEE European Signal Processing Conference (EUSIPCO)*, 2008.
3. A. Katsamanis, A. Roussos, P. Maragos, M. Aron, and M.-O. Berger. Inversion from audiovisual speech to articulatory information by exploiting multimodal data. In *International Seminar on Speech Production*, December 2008.

4. A. Roussos, A. Katsamanis, and P. Maragos. Tongue tracking in ultrasound images with active appearance models. In *Proceedings of the IEEE International Conference on Image Processing*, November 2009.
5. A. Roussos and P. Maragos. Tensor-based image diffusions derived from generalizations of the total variation and beltrami functionals. In *Proceedings of the IEEE International Conference on Image Processing*, September 2010.
6. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Affine-invariant modeling of shape-appearance images applied on sign language handshape classification. In *Proceedings of the IEEE International Conference on Image Processing*, September 2010.
7. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Hand tracking and affine shape-appearance handshape sub-units in continuous sign language recognition. In *Proceedings of the Workshop on Sign, Gesture and Activity, 11th European Conference on Computer Vision*, September 2010.

Τεχνικές Αναφορές (δημόσια διαθέσιμες)

1. Y. Laprie, P. Maragos, A. Roussos, M.-O. Berger, E. Kerrien, and M. Aron. Deliverable D3.2, Final report on design, acquisition and processing of articulatory data. Technical report, EU FET programme Audiovisual to Articulatory Speech Inversion (ASPI), January 2009.
2. P. Maragos, O. Engwall, Y. Laprie, S. Maeda, J. Schoentgen, A. Katsamanis, and A. Roussos. Deliverable D2.2, Final report on speech inversion methods. Technical report, EU FET programme Audiovisual to Articulatory Speech Inversion (ASPI), January 2009.

Βιογραφικό Σημείωμα

Αναστάσιος Ρούσσος

Εκπαίδευση

- 2005–2010 **Δίπλωμα Διδάκτορα Μηχανικού (Ph.D.)**, *Εθνικό Μετσόβιο Πολυτεχνείο*, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών.
Ερευνητικό πεδίο: Ανάλυση Εικόνων και Όραση Υπολογιστών.
Θέμα Διατριβής: Μη-γραμμική Διάχυση στην Όραση Υπολογιστών και Στατιστικά Μοντέλα Σχήματος με Εφαρμογές στην Ανάλυση Εικόνων Αρθρωτών Φωνητικού και Νοηματικού Λόγου
Επιβλέπων: Καθ. Πέτρος Μαραγκός (<http://cvsp.cs.ntua.gr/maragos>)
Μέσος Όρος Μεταπτυχιακών μαθημάτων: 9.83/10
- 2000–2005 **Δίπλωμα/M.Eng Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών**, *Εθνικό Μετσόβιο Πολυτεχνείο*.
Τελικός Βαθμός: 8.42/10
Κατεύθυνση σπουδών: Κατεύθυνση Επικοινωνιών
Ροές μαθημάτων ειδίκευσης: 'Σήματα, Έλεγχος και Ρομποτική', 'Κύματα και Τηλεπικοινωνίες', 'Λογισμικό Η/Υ', 'Επικοινωνίες και Δίκτυα Υπολογιστών'
Διπλωματική εργασία: Θεωρία και Εφαρμογές Μερικών Διαφορικών Εξισώσεων στην Όραση Υπολογιστών (*επιβλέπων*: Καθ. Πέτρος Μαραγκός).

Ερευνητική εμπειρία

- 2010–παρόν **Μεταδιδακτορικός ερευνητής**, *School of Electronic Engineering & Computer Science*, Queen Mary, University of London.
Μέλος του εργαστηρίου *Queen Mary Vision Laboratory*. Συμμετοχή στο ακόλουθο πρόγραμμα του Ευρωπαϊκού Συμβουλίου Έρευνας:
- *Human Motion Analysis from Image Sequences (HUMANIS)*: Ανάπτυξη αλγορίθμων Όρασης Υπολογιστών ικανών να ανακτήσουν λεπτομερή 3D μοντέλα ανθρώπων ή άλλων παραμορφώσιμων και αρθρωτών αντικειμένων με έναν πλήρως αυτόματο τρόπο, χρησιμοποιώντας μόνο την ακολουθία εικόνων εισόδου. Επιστημονική υπεύθυνος: Δρ. Lourdes Agarito (<http://www.eecs.qmul.ac.uk/~lourdes>).
- 2005-2010 **Διδακτορικός ερευνητής**, *Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών*, Εθνικό Μετσόβιο Πολυτεχνείο.
Μέλος της ερευνητικής ομάδας *Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σήματος*. Συμμετοχή στα ακόλουθα ερευνητικά προγράμματα της Ευρωπαϊκής Ένωσης:
- *Audiovisual to Articulatory Speech Inversion (ASPI, <http://aspi.loria.fr>)*: Έρευνα σε καινοτόμους αλγορίθμους βελτίωσης της ποιότητας βιοϊατρικών οπτικών δεδομένων της φωνητικής οδού κατά την ομιλία. Σχεδίαση μοντέλων για την παρακολούθηση της γλώσσας σε εικόνες υπερήχων, την 3D κατάτμηση της φωνητικής οδού σε εικόνες MRI και την αντιστοίχιση εικόνων προερχόμενες από διαφορετικές τεχνικές απεικόνισης της φωνητικής οδού.
- *Sign Language Recognition, Generation and Modelling with Application in Deaf Communication (Dicta-Sign, <http://www.dictasign.eu>)*: Έρευνα στην αυτόματη κατάτμηση και παρακολούθηση του κεφαλιού και των χεριών νοηματιστή κατά τη διάρκεια συνεχούς νοηματισμού. Στατιστική μοντελοποίηση του σχήματος, της εμφάνισης και της δυναμικής των χειρομορφών για την παρακολούθηση των χεριών και την εξαγωγή χαρακτηριστικών.

Τρέχοντα Ερευνητικά Ενδιαφέροντα

- Μεθοδολογίες Μέθοδοι βασισμένες στον Λογισμό των Μεταβολών και στις Μερικές Διαφορικές Εξισώσεις, μη-γραμμική διάχυση, ενεργά περιγράμματα, στατιστικά μοντέλα σχήματος, δυναμικά μοντέλα για την παρακολούθηση αντικειμένων, πολυκλιμακωτή ανάλυση.

Προβλήματα Όρασης Υπολογιστών	Οπτική ροή, απλοποίηση, αποθορυβοποίηση, παρεμβολή και αποκατάσταση εικόνων, ενδο-συμπλήρωση εικόνων, κατάτμηση εικόνων, εντοπισμός και παρακολούθηση αντικειμένων, εξαγωγή χαρακτηριστικών.
Εφαρμογές	3D ανακατασκευή παραμορφώσιμων αντικειμένων από την κίνησή τους, επεξεργασία βιοϊατρικών εικόνων, αυτόματη αναγνώριση νοσηματικής γλώσσας, οπτικοακουστική αντιστροφή φωνής, αρθρωτή σύνθεση φωνής, υπολογιστική φωτογραφία.

Υποτροφίες και βραβεία

- 2007 **First International Conference on Scale Space and Variational Methods in Computer Vision (SSVM-2007)**, Βραβείο Καλύτερης Φοιτητικής Δημοσίευσης του συνεδρίου για την δημοσίευση: A. Roussos and P. Maragos, "Vector-Valued Image Interpolation by an Anisotropic Diffusion-Projection PDE".
- 2007 **Πρώτο Πανελλήνιο Συνέδριο Φοιτητών Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών (ΣΦΗΜΜΥ-2007)**, Βραβείο Καλύτερου Poster του συνεδρίου για την δημοσίευση: Α. Ρούσσοσ και Π. Μαραγκός, "ΜΔΕ Ανισοτροπικής Διάχυσης - Προβολής για την Παρεμβολή Διανυσματικών Εικόνων".
- 2007–2008 **Εθνικό Μετσόβιο Πολυτεχνείο**, Μονοετής ερευνητική υποτροφία για διδακτορικές σπουδές.
- 2000 **Ένωση Ελλήνων Φυσικών**, Έπαινος επίδοσης στον Πανελλήνιο Διαγωνισμό Φυσικής.
- 1997 **Ελληνική Μαθηματική Εταιρεία**, Γ' Βραβείο στον 57^ο Πανελλήνιο Μαθητικό Διαγωνισμό στα Μαθηματικά 'Ο Ευκλείδης'.

Διδακτική Εμπειρία

- 2007–2010 **Μεταπτυχιακός βοηθός μαθήματος**, Όραση Υπολογιστών (μεταπτυχιακό και προπτυχιακό μάθημα δου εξαμήνου), Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Ε.Μ.Π..
Επικουρικό έργο, εργαστήρια, προετοιμασία υλικού μαθήματος, παρουσιάσεις, συνεπίβλεψη και επικοινωνία με τους φοιτητές. Παράδοση σεμιναρίων:
Μέθοδοι Όρασης Υπολογιστών βασισμένες σε Μερικές Διαφορικές Εξισώσεις (2007, 2008, 2010).
Ανίχνευση χαρακτηριστικών εικόνων βασισμένη στην Μαθηματική Μορφολογία (2008).

Κριτής σε διεθνή επιστημονικά περιοδικά

- 2009 **IEEE Transactions on Image Processing.**
- 2007 **IEEE Signal Processing Letters.**

Μέλος επιστημονικών οργανώσεων

Institute of Electrical and Electronics Engineers (IEEE), <http://www.ieee.org>.
Τεχνικό Επιμελητήριο Ελλάδος (TEE), <http://www.e-tee.org.gr>.

Γλώσσες

Αγγλικά	Άριστα	<i>Cambridge ESOL FCE, Grade A</i>
Γαλλικά	Καλά	<i>εμπειρία σε επιστημονική ανάγνωση, γραφή και ομιλία</i>
Ελληνικά	Μητρική	<i>Diplôme de Langue et de Civilization Françaises, Mention Bien</i>