



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ
ΒΙΟΙΑΤΡΙΚΗ

**Εντοπισμός του περιγράμματος μη φυσιολογικών περιοχών σε
ιατρικές εικόνες με προηγμένες μεθόδους τμηματοποίησης**

Νικόλαος Μαραγκός

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
Υπεύθυνος
Δημήτριος Ιακωβίδης
Επίκουρος Καθηγητής

Λαμία, Σεπτέμβριος 2014



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ
ΒΙΟΙΑΤΡΙΚΗ**

**Εντοπισμός του περιγράμματος μη φυσιολογικών περιοχών σε
ιατρικές εικόνες με προηγμένες μεθόδους τμηματοποίησης**

Νικόλαος Μαραγκός

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Επιβλέπων
Δημήτριος Ιακωβίδης
Επίκουρος Καθηγητής**

Λαμία, Σεπτέμβριος 2014

**Εντοπισμός του περιγράμματος μη φυσιολογικών περιοχών σε
ιατρικές εικόνες με προηγμένες μεθόδους τμηματοποίησης**

Νικόλαος Μαραγκός

Τριμελής Επιτροπή:

Ιακωβίδης Δημήτριος, Επίκουρος Καθηγητής

Δελήμπασης Κωνσταντίνος, Επίκουρος Καθηγητής

Πλαγιανάκος Βασίλειος, Αναπληρωτής Καθηγητής

Βεβαιώνω ότι η παρούσα διπλωματική εργασία είναι αποτέλεσμα δικής μου δουλειάς και δεν αποτελεί προϊόν αντιγραφής. Στις δημοσιευμένες ή μη δημοσιευμένες πηγές που αναφέρω έχω χρησιμοποιήσει εισαγωγικά όπου απαιτείται και έχω παραθέσει τις πηγές τους στο τμήμα της βιβλιογραφίας.

Copyright © Πανεπιστήμιο Θεσσαλίας, 2014.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

Περίληψη

Η παρούσα πτυχιακή εργασία επικεντρώνεται στη διαδικασία της τμηματοποίησης βιοϊατρικών εικόνων και αναπτύσσει μία εφαρμογή που εξειδικεύεται στη κατάτμηση γλωττίδας με απώτερο στόχο την διευκόλυνση της διάγνωσης ιατρικών προβλημάτων στον μηχανισμό φώνησης. Για τον υπολογισμό του περιγράμματος της γλωττίδας η εργασία στηρίζεται σε προηγμένες μεθόδους τμηματοποίησης όπως το μοντέλο των Γεωδαιτικών Ενεργών Περιγραμμάτων (GAC) και η θεωρία των επιπεδοσυνόλων όπου η καμπύλη τμηματοποίησης καθοδηγείται από τις ακμές της εικόνας και επιτυγχάνει τη κατάτμηση με έναν ελεύθερο και φυσικό τρόπο απαλλάσσοντας τη διαδικασία από την αυστηρή μαθηματική παραμετροποίηση. Ακόμα προτείνουμε μία νέα προσέγγιση που μας επιτρέπει να εντοπίζουμε τη περιοχή της εικόνας όπου βρίσκεται η γλωττίδα και να σχηματίζουμε ένα αρχικό περίγραμμα το οποίο λειτουργεί ως αφετηρία για την εξέλιξη της καμπύλης. Επίσης επιχειρούμε να βελτιώσουμε τα αποτελέσματα της τμηματοποίησης συνδυάζοντας τα Γεωδαιτικά Ενεργά Περιγράμματα με ένα προγενέστερο σχήμα για τη γλωττίδα με σκοπό να αξιοποιήσουμε την πρότερη πληροφορία που προσφέρει χρησιμοποιώντας τη μέθοδο του Leventon et al. Για την εξαγωγή του προγενέστερου σχήματος βασίζομαστε στην τεχνική της Ανάλυσης Κύριων Συνιστωσών (PCA) και στην τεχνική της Ανάλυσης Κύριων Συνιστωσών Πυρήνα (Kernel PCA) οι οποίες κατασκευάζουν το στατιστικό μοντέλο σχήματος μοντελοποιώντας με αυτό τον τρόπο το σχήμα της γλωττίδας και τις διακυμάνσεις που ακολουθεί. Επιπλέον μέσα στους στόχους της πτυχιακής εργασίας εντάσσεται η μελέτη της εφαρμογής ανάπτυξης κώδικα Insight Toolkit η οποία υλοποιεί πλήθος αλγορίθμων που εστιάζουν στην επεξεργασία και τμηματοποίηση εικόνων.

Για τη διεξαγωγή των πειραμάτων στηρίζομαστε στη τεχνική της διασταυρωμένης επικύρωσης και η αξιολόγηση των αποτελεσμάτων γίνεται με τους δείκτες ομοιότητας Jaccard και Dice. Για να εκτιμήσουμε την απόδοση της εφαρμογής κάνουμε εκτενή πειραματική διερεύνηση χρησιμοποιώντας την PCA και την Kernel PCA για την παραγωγή των προγενέστερων σχημάτων. Η μέθοδος που υλοποιούμε εντοπίζει με επιτυχία σε όλες τις εικόνες τη περιοχή της γλωττίδας και παρέχει ένα αρκετά καλό αρχικό περίγραμμα με το οποίο ξεκινάει η διαδικασία της κατάτμησης. Τα αποτελέσματα της τμηματοποίησης είναι αρκετά ενθαρρυντικά καθώς οι δύο δείκτες δείχνουν ποσοστά επιτυχίας κοντά στο 90%. Η χρήση των μεθόδων PCA και Kernel PCA οδηγεί σε συγκρίσιμα αποτελέσματα ως προς τη βέλτιστη τμηματοποίηση που μπορεί να επιτευχθεί. Ωστόσο παρατηρείται ότι με τη χρήση της μεθόδου Kernel PCA το ενεργό περίγραμμα γίνεται περισσότερο ανθεκτικό στις μεταβολές των παραμέτρων του.

Λέξεις – κλειδιά

Τμηματοποίηση εικόνας, κατάτμηση γλωττίδας, ανάλυση κύριων συνιστωσών, ανάλυση κύριων συνιστωσών πυρήνα, γεωδαιτικά ενεργά περιγράμματα, προγενέστερα σχήματα.

Abstract

The present thesis focuses on the process of biomedical image segmentation and it develops an application specially designed for glottal segmentation with the aim to facilitate the diagnosis of medical problems in phonation mechanism. The computation of glottis' contour is based on advanced segmentation methods such as Geodesic Active Contour (GAC) model and level-set theory therefore the segmentation curve is guided by the edges of the image and obtains the segmentation in a free and natural way while it rids the process from rigorous mathematical parameterization. We propose a new approach which detects the area of the image where glottis is located and creates an initial contour which initializes the evolving curve of Geodesic Active Contour model. Also, thinking that a prior knowledge about the shape of the segmenting object could possibly help the process, we try to improve the effectiveness of the application by combining GACs with a shape prior for glottis through the method of Leventon et al. The extraction of shape prior is based on the techniques of Principal Component Analysis (PCA) and Kernel Principal Component Analysis (Kernel PCA). These methods construct the statistical shape model for glottis and hence we are able to model the shape of glottis and its fluctuations. Additionally this thesis studies the Insight Toolkit development framework which implements various algorithms that concentrate on image processing and image segmentation.

The methodology we use for the experiments is cross-validation and the evaluation is done by the similarity indexes of Jaccard and Dice. The performance of the application is estimated through extensive experimental study using PCA and Kernel PCA for shape prior extraction. The method we implement detects successfully for all images the location of glottis and it provides a sufficient initial contour for evolving curve. The segmentation results are promising since both indexes show a success rate close to 90%. The use of PCA and Kernel PCA methods leads to comparable results concerning the optimal segmentation that can be achieved. However we observe that the use of Kernel PCA makes the active contour more resistant to parameterization.

Keywords

Image segmentation, glottal segmentation, principal component analysis, kernel principal component analysis, geodesic active contours, prior shapes.

Ευχαριστίες

Η επιστήμη της πληροφορικής έχει συμβάλει καθοριστικά στην εξέλιξη της τεχνολογίας και κατ'επέκταση στη διευκόλυνση της καθημερινή μας ζωή. Οι δυνατότητες που προσφέρει με ώθησαν να ασχοληθώ με τον χώρο αυτό καθώς μου επιτρέπει να συνδυάσω τη χαρά της δημιουργίας με το αίσθημα της κοινωνικής προσφοράς μέσα από την υλοποίηση εφαρμογών που αναβαθμίζουν το επίπεδο των παρεχόμενων ιατρικών υπηρεσιών. Φθάνοντας στο τέλος των προπτυχιακών μου σπουδών μπορώ να πω ότι αισθάνομαι τυχερός αλλά και δικαιωμένος με την επιλογή να φοιτήσω στο τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική.

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Δημήτριο Ιακωβίδη για την καθοδήγηση και την εμπιστοσύνη που μου έδειξε με την ανάθεση της πτυχιακής εργασίας καθώς επίσης και τους καθηγητές μου για τις πολύτιμες γνώσεις που αποκόμισα όλο αυτό το διάστημα. Τέλος, δεν θα μπορούσα να μην αναφερθώ στους φίλους μου Ιωάννη Βέρδη, Διαμαντή Λημναίο, Ιωάννη Σιφνιάδη και Δημήτριο Τυχάλα, οι οποίοι μέσα από τις συζητήσεις που είχαμε με στήριξαν και με βοήθησαν να προχωρήσω.

Η εργασία είναι αφιερωμένη στην αδελφή μου Έλενα και στους γονείς μου Σωτηρία και Γιάννη.

‘The great use of a life is to spend it for something that outlasts it’
William James

Περιεχόμενα

Κεφάλαιο 1	Εισαγωγή	13
1.1	Όραση υπολογιστών	13
1.2	Τμηματοποίηση εικόνας	15
1.2.1	Κατωφλίωση	16
1.2.2	Μέθοδος επέκτασης περιοχής	16
1.2.3	Μετασχηματισμός watershed	16
1.2.4	Τμηματοποίηση με τη χρήση γράφων	17
1.2.5	Ενεργά περιγράμματα	17
1.3	Μηχανική μάθηση	17
1.4	Σύστημα φώνησης	18
1.4.1	Ανατομία λάρυγγα	19
1.4.2	Διαδικασία φώνησης	20
1.4.3	Στροβοσκόπηση	21
1.5	Σχετικές εργασίες	22
Κεφάλαιο 2	Τεχνικές Στατιστικής Ανάλυσης Δεδομένων	25
2.1	Ανάλυση κύριων συνιστωσών	25
2.1.1	Υπολογισμός ανάλυσης κύριων συνιστωσών	26
2.1.2	Ανάλυση κύριων συνιστωσών μέσω της αποσύνθεσης μοναδικής τιμής	27
2.2	Ανάλυση κύριων συνιστωσών πυρήνα	30
2.2.1	Υπολογισμός ανάλυσης κύριων συνιστωσών πυρήνα	31
2.2.2	Το τέχνασμα του πυρήνα	33
2.2.3	Κεντράρισμα πίνακα πυρήνα	34
2.2.4	Ανακατασκευή προ-εικόνας	36
Κεφάλαιο 3	Μέθοδοι Τμηματοποίησης	38
3.1	Ενεργά περιγράμματα	38
3.1.1	Μαθηματικό μοντέλο ‘φιδιού’	38
3.1.2	Φίδι ροής κλιμακωτού διανύσματος	40
3.2	Γεωδαιτικά ενεργά περιγράμματα	42
3.2.1	Υπολογισμός γεωδαιτικού ενεργού περιγράμματος	42
3.2.2	Μέθοδος επιπεδοσυνόλων	43
3.3	Τμηματοποίηση με γεωδαιτικά ενεργά περιγράμματα και χρήση προγενέστερου σχήματος	46

Κεφάλαιο 4	Μεθοδολογία	49
4.1	Αλγόριθμοι τμηματοποίησης εικόνων	49
4.1.1	Παραγωγή προγενέστερου σχήματος	50
4.1.2	Τμηματοποίηση - Διαμόρφωση αρχικού περιγράμματος	51
4.1.3	Τμηματοποίηση - Υπολογισμός τελικής καμπύλης	56
4.2	Εργαλεία υλοποίησης	61
4.2.1	Insight Segmentation and Registration Toolkit	61
4.2.2	MLPACK	62
Κεφάλαιο 5	Πειράματα και Αποτελέσματα	63
5.1	Σύνολα δεδομένων	63
5.2	Μεθοδολογία εκπαίδευσης και τμηματοποίησης	63
5.3	Μέτρα αξιολόγησης της τμηματοποίησης	64
5.4	Αποτελέσματα	65
Κεφάλαιο 6	Συμπεράσματα και Μελλοντικές Κατευθύνσεις	69
6.1	Συμπεράσματα	69
6.2	Μελλοντική εργασία και προοπτικές	70
Παράρτημα Α		72
ΒΙΒΛΙΟΓΡΑΦΙΑ		80

Κατάλογος Εικόνων

1.1	Τμήματα της ανατομίας του λάρυγγα	20
1.2	Οι φάσεις του κύκλου ταλάντωσης των φωνητικών χορδών	20
1.3	Η μορφή της γλωττίδας κατά τη διάρκεια λειτουργιών του λάρυγγα	21
2.1	Παράδειγμα για PCA και κύριες συνιστώσες	26
2.2	(α)-(στ) αντιπροσωπευτικές εικόνες γλωττίδας και εικόνες στατιστικού μοντέλου με PCA	29
2.3	(α),(β) διαφορές των κύριων συνιστωσών από PCA και των κύριων συνιστωσών πυρήνα από KPCA	31
2.4	Πίνακας πυρήνα	33
2.5	(α)-(δ) απεικόνιση της μέσης εικόνας και των τριών πρώτων κύριων συνιστωσών πυρήνα με τη μέθοδο KPCA	37
3.1	Αναπαράσταση πεδίου GVF	41
3.2	Μέθοδος των επιπεδοσυνόλων	44
3.3	Δυνάμεις τμηματοποίησης από τον συνδυασμό του μοντέλου GAC και του προγενέστερου σχήματος	48
4.1	Διάγραμμα ροής εφαρμογής	50
4.2	Διάγραμμα ροής για την παραγωγή του προγενέστερου σχήματος	51
4.3	Διάγραμμα ροής αλγορίθμου 1 για υπολογισμό αρχικού περιγράμματος	53
4.4	(α)-(θ), παραδείγματα υπολογισμού αρχικού περιγράμματος	54
4.5	(α)-(στ), εικόνες από βήματα αλγορίθμου 1	55
4.6	(α)-(δ), εικόνες προγενέστερου σχήματος και τελικής καμπύλης	57
4.7	Διάγραμμα ροής αλγορίθμου 2 για τον υπολογισμό του τελικού περιγράμματος	59
4.8	(α)-(ε), στάδια εξέλιξης της καμπύλης	60
5.1	Διαγράμματα αξιολόγησης τμηματοποίησης	65
5.2	(α)-(δ), Παραδείγματα με αποτελέσματα των πειραμάτων	68

Εισαγωγή

Η φωνή παίζει έναν ιδιαίτερα σημαντικό ρόλο στη καθημερινή μας ζωής καθώς ανήκει στα βασικά μέσα επικοινωνίας του ανθρώπου και άρα πλήθος δραστηριοτήτων εξαρτώνται από την εύρυθμη λειτουργία της. Γίνεται φανερό λοιπόν ότι πιθανά προβλήματα στη διαδικασία της φώνησης θα έχουν αντίκτυπο στη ποιότητα ζωής. Από τον χώρο της ιατρικής δίνεται μεγάλη σημασία στη γλωττίδα για τον εντοπισμό τέτοιων προβλημάτων καθώς αποτελεί μέρος του συστήματος φώνησης και η παρατήρησή της μπορεί να δώσει σημαντικές πληροφορίες για τη λειτουργία του μηχανισμού. Η επιστήμη των υπολογιστών μπορεί να συνδράμει σε αυτή τη κατεύθυνση με την ανάπτυξη αλγορίθμων οι οποίοι επικεντρώνονται στη κατάτμηση εικόνων του λάρυγγα προκειμένου να μελετηθεί το σχήμα της γλωττίδας για να εξαχθούν χρήσιμα συμπεράσματα.

Στο κεφάλαιο αυτό δίνεται το θεωρητικό υπόβαθρο της πτυχιακής εργασίας έτσι ώστε να γίνει κατανοητό το πλαίσιο μέσα στο οποίο υλοποιείται η εφαρμογή. Περιγράφεται το πεδίο της όρασης υπολογιστών, της μηχανικής μάθησης και γίνεται μία εισαγωγή στη διαδικασία της κατάτμησης. Επίσης εξετάζεται η λειτουργία του συστήματος φώνησης και γίνεται μία ανασκόπηση των εργασιών που έχουν παρουσιαστεί πάνω στη τμηματοποίηση της γλωττίδας.

1.1 Όραση υπολογιστών

Ένα πλήθος προβλημάτων που καλούνται να λύσουν οι υπολογιστές σήμερα απαιτούν την χρήση εικόνων ή βίντεο και τον συνδυασμό τους με μεθόδους που μας επιτρέπουν την περαιτέρω επεξεργασία για την εξαγωγή συμπερασμάτων. Η ανάγκη αυτή οδήγησε στην ανάπτυξη του πεδίου της όρασης υπολογιστών (computer vision) [45]. Ο κλάδος ξεκίνησε τη δεκαετία του '60 ως κλάδος της τεχνητής νοημοσύνης (artificial intelligence) [46] και θέτει ως στόχο την λήψη, ανάλυση και κατανόηση εικόνων επιχειρώντας να προσομοιώσει τη βιολογική όραση και να αναπαραστήσει αλγοριθμικά τον τρόπο σκέψης και αντίληψης που ακολουθεί ο άνθρωπος για την ερμηνεία των οπτικών ερεθισμάτων.

Η όραση υπολογιστών επικαλύπτεται σε μεγάλο βαθμό από τη μηχανική όραση (machine vision) [47] και την επεξεργασία εικόνων με αποτέλεσμα, παρ'όλο που ο κάθε χώρος εμφανίζει τα δικά του γνωρίσματα, στις περισσότερες εφαρμογές να

παρατηρείται η συμμετοχή και των τριών καθώς λειτουργούν συμπληρωματικά. Η μηχανική όραση συναντάται συχνότερα στα ρομποτικά συστήματα και προσδίδει σε ένα ρομπότ την ικανότητα να επικοινωνεί με το περιβάλλον του για την εκτέλεση εργασιών, μία δυνατότητα που αξιοποιείται κυρίως στο βιομηχανικό τομέα για την αυτοματοποίηση διαδικασιών. Η επεξεργασία εικόνων χρησιμοποιεί αλγορίθμους και φίλτρα που επιτρέπουν στον χρήστη να επιδρά στην εικόνα προκειμένου να εξάγει χαρακτηριστικά που θα χρησιμοποιηθούν σε μετέπειτα στάδια όπως αυτό της ανάλυσης ή να κάνει μετασχηματισμούς για να την φέρει σε μία επιθυμητή μορφή. Επιπλέον, η πολυπλοκότητα των προβλημάτων με τα οποία ασχολείται η όραση υπολογιστών επιβάλλει τη συνεργασία με πολλά ακόμα επιστημονικά πεδία. Για παράδειγμα από το χώρο των μαθηματικών χρησιμοποιείται η γεωμετρία και τεχνικές στατιστικής ανάλυσης και βελτιστοποίησης, ενώ από την φυσική χρησιμοποιείται ο κλάδος της οπτικής και της ρευστομηχανικής. Ένας άλλος τομέας με σημαντική συνεισφορά είναι αυτός της νευροφυσιολογίας καθώς οι μελέτες και τα συμπεράσματα στα οποία καταλήγει χρησιμοποιούνται για την ανάπτυξη μεθόδων που αντιγράφουν τους μηχανισμούς που εφαρμόζονται στα έμβια όντα. Τέλος, η τεχνητή νοημοσύνη προσφέρει εργαλεία όπως οι αλγόριθμοι αναζήτησης και τους κλάδους της μηχανική μάθηση (machine learning) [48] και της αναγνώριση προτύπων (pattern recognition) [49] με αποτέλεσμα να παίζει καθοριστικό ρόλο στη κατανόηση πληροφοριών και στη λήψη αποφάσεων.

Η διεργασία της όρασης υπολογιστών υλοποιείται σε τρία επίπεδα. Το χαμηλότερο επίπεδο εξετάζει τη λήψη της εικόνας και την επιλογή του κατάλληλου υλικού (hardware). Η επιλογή του υλικού εξαρτάται από το είδος και τις απαιτήσεις της εργασίας που πρέπει να εκπονηθεί, οπότε σε ένα εργοστάσιο για παράδειγμα, θα προτιμηθούν μία ή δύο απλές κάμερες για την παρατήρηση των προϊόντων και τη κίνηση κάποιου ρομποτικού βραχίονα ενώ στον ιατρικό τομέα, για την απεικόνιση του εσωτερικού του ανθρώπινου σώματος, κρίνεται απαραίτητο ένα πιο περίπλοκο ανώτερο σύστημα όπως για παράδειγμα ένας αξονικός ή μαγνητικός τομογράφος. Το μεσαίο επίπεδο επικεντρώνεται στην οργάνωση και επεξεργασία της πληροφορίας που αποκτήθηκε. Στο στάδιο αυτό χρησιμοποιούνται τεχνικές βελτίωσης, ομαλοποίησης, αποθορυβοποίησης, ανάλυσης σε πολλαπλές κλίμακες αλλά και τεχνικές εξαγωγής χαρακτηριστικών όπως ιστόγραμμα, ακμές, υφή, χρώματα. Στο υψηλότερο επίπεδο γίνεται η ερμηνεία της εικόνας μέσα από την αξιολόγηση των χαρακτηριστικών, την εύρεση των μεταξύ τους συσχετίσεων και την ανακάλυψη των μοτίβων που ενυπάρχουν μέσα στο σύνολο της πληροφορίας. Ο συνδυασμός των παραπάνω μεθόδων δίνει στην όραση υπολογιστών τη δυνατότητα να εκτελεί πλήθος απλών και σύνθετων εργασιών όπως:

- ❖ Εντοπισμός ακμών των εικόνων
- ❖ Μοντελοποίηση και ανάλυση του χρώματος των εικόνων
- ❖ Ανάλυση και σύνθεση της υφής που παρουσιάζουν οι εικόνες
- ❖ Εντοπισμός των αντικειμένων που περιέχονται στις εικόνες
- ❖ Τμηματοποίηση εικόνων

- ❖ Κατηγοριοποίηση περιοχών των εικόνων
- ❖ Παρακολούθηση και ανάλυση της κίνησης αντικειμένων
- ❖ Αποκατάσταση εικόνων λόγω υψηλού θορύβου
- ❖ Στερέωση και 3-Δ απεικόνιση ενός τοπίου χρησιμοποιώντας ένα σύνολο εικόνων

Η όραση υπολογιστών με τις πολλαπλές δυνατότητες που προσφέρει βρίσκει εφαρμογή σε ένα μεγάλο εύρος περιοχών. Στο χώρο της ιατρικής συνέβαλε στην ανάπτυξη της ρομποτικής χειρουργικής με τη δημιουργία συστημάτων όπως το Ντα Βίντσι που καθιστούν εφικτή την επέμβαση εξ'αποστάσεως και συνάμα η ακρίβεια που προσφέρουν επιτρέπουν τη διεξαγωγή εξαιρετικά πολύπλοκων και εκτεταμένων επεμβάσεων χρησιμοποιώντας μικρότερες τομές στον ασθενή με αποτέλεσμα να ελαχιστοποιούνται ο χρόνος νοσηλείας, τα σημάδια της επέμβασης και ο πόνος που υφίσταται. Ένα άλλο παράδειγμα χρήσης στη τηλεϊατρική είναι το έξυπνο σπίτι όπου είναι δυνατή η έμμεση παρακολούθηση ασθενών με επιβεβαρωμένη υγεία και σε περίπτωση κάποιου επεισοδίου όπως π.χ. πτώση, καλείται άμεσα βοήθεια. Επομένως ο ασθενής διατηρεί τη ποιότητα ζωής του αφού διαμένει στο σπίτι του και παράλληλα συνεχίζει να δέχεται ιατρική υποστήριξη. Στον τομέα των μεταφορών αρχίζουν να εμφανίζονται τα πρώτα πλήρως αυτόνομα οχήματα οπότε υπάρχουν βάσιμες ελπίδες ότι στο μέλλον θα κυκλοφορούν ασφαλή αυτοκίνητα που θα έχουν τη δυνατότητα να οδηγούν χωρίς τη συμμετοχή οδηγού, μία καινοτομία που πιστεύεται ότι θα συμβάλλει στην ελάττωση των τροχαίων ατυχημάτων και τη βελτίωση του χρόνου διαδρομής. Επίσης μπορεί να χρησιμοποιηθεί για τον έλεγχο της κυκλοφορίας σε δρόμους και για την παρακολούθηση της κίνησης πλήθους οπότε μπορεί να βοηθήσει στο σχεδιασμό καλύτερων ρυθμίσεων. Μεγάλο ενδιαφέρον υπάρχει από τη στρατιωτική βιομηχανία όπου ήδη χρησιμοποιείται στα συστήματα καθοδήγησης πυραύλων ενώ στο μέλλον προγραμματίζεται η ανάπτυξη πλήρως αυτόνομων μη επανδρωμένων αεροσκαφών. Στο βιομηχανικό τομέα οδηγεί στην αύξηση της παραγωγής και τη μείωση του κόστους μέσα από την αυτοματοποίηση των διαδικασιών ενώ μπορεί να συνδράμει στη βελτίωση της ποιότητας αφού εργασίες που απαιτούν μεγάλη ακρίβεια εκτελούνται από εξειδικευμένα ρομποτικά συστήματα. Ακόμη, τεχνικές της όρασης υπολογιστών μπορούν να χρησιμοποιηθούν σε εφαρμογές για την αναγνώριση αριθμών και χαρακτήρων, την αναγνώριση προσώπου καθώς επίσης και για τη σύγκριση βιομετρικών χαρακτηριστικών.

1.2 Τμηματοποίηση εικόνων

Η διεργασία της τμηματοποίησης ή κατάτμησης επιτρέπει το διαχωρισμό μίας εικόνας σε περιοχές με ομοιογενή χαρακτηριστικά ενώ σε μετέπειτα στάδιο δίνει την

δυνατότητα εντοπισμού αντικειμένων. Για τον σκοπό αυτό υπάρχουν δύο προσεγγίσεις, η πρώτη στηρίζεται στα όρια των αντικειμένων (boundary-based) όπως για παράδειγμα στις μεθόδους [22],[50] και η δεύτερη στηρίζεται στην ομοιομορφία περιοχών της εικόνας (region-based) όπως στη μέθοδο [51] ενώ οι δύο προσεγγίσεις μπορούν να χρησιμοποιούνται και συνδυαστικά. Συνήθως τα χαρακτηριστικά που εξετάζονται είναι η ένταση των εικονοστοιχείων, η υφή, το χρώμα και οι ακμές, ενώ σε πιο εξελιγμένες μεθόδους μπορεί να χρησιμοποιηθεί αρχική γνώση με τη μορφή των προγενέστερων σχημάτων. Στη συνέχεια περιγράφονται μερικές από τις δημοφιλέστερες μεθόδους τμηματοποίησης.

1.2.1 Κατωφλίωση

Είναι η απλούστερη μέθοδος τμηματοποίησης και στηρίζεται στον ορισμό ενός κατωφλίου ως κριτήριο για τη κατάτμηση της εικόνας. Στην έξοδο παράγεται μία δυαδική εικόνα που διαχωρίζεται στο παρασκήνιο (background) και το προσκήνιο (foreground) το οποίο περιλαμβάνει τα αντικείμενα της εικόνας. Ο ορισμός του κατωφλίου μπορεί να γίνεται χειροκίνητα ή αυτόματα με τη βοήθεια του ιστογράμματος και άλλων εργαλείων. Επίσης, με τη χρήση πολλαπλών κατωφλίων η εικόνα μπορεί να διαχωριστεί σε περισσότερες περιοχές. Μερικές από τις πιο γνωστές τεχνικές αυτής της κατηγορίας είναι η μέθοδος Otsu [32] και η μέθοδος μέγιστης εντροπίας [33].

1.2.2 Μέθοδος επέκτασης περιοχής

Η μέθοδος βασίζεται στο [34] και εδώ η τμηματοποίηση πραγματοποιείται μέσα από την επέκταση και συγχώνευση περιοχών. Αρχικά ορίζονται σε διάφορα σημεία της εικόνας οι σπόροι (seeds) από τους οποίους αρχίζει να επεκτείνεται κάθε περιοχή. Για την επέκταση ελέγχονται τα γειτονικά εικονοστοιχεία και αν εμφανίζουν ομοιότητες ως προς τα υπόλοιπα εικονοστοιχεία της περιοχής ή ξεπερνούν κάποιο κατώφλι τότε προσαρτώνται σε αυτήν. Με τον ίδιο τρόπο μπορεί να αποφασιστεί η συγχώνευση δύο περιοχών.

1.2.3 Μετασηματισμός watershed

Η μέθοδος του watershed [35] κάνει κατάτμηση εικόνων προσομοιώνοντας τη δημιουργία υποδοχέων νερού κατά το πλημμύρισμα μίας ανάγλυφης περιοχής. Αρχικά με τη τεχνική κλίσης εικόνας (image gradient) δημιουργεί ένα ανάγλυφο για την εικόνα στο οποίο οι περιοχές που εμφανίζουν έντονες ακμές σχηματίζουν όρη και οι περιοχές που είναι ομοιομορφες σχηματίζουν κοιλάδες. Συνεπώς τα σημεία της εικόνας που βρίσκονται κοντά σε μεγάλες ακμές τοποθετούνται στα υψηλότερα

υψόμετρα ενώ τα σημεία της εικόνας που βρίσκονται κοντά στο εσωτερικό των ομοιόμορφων περιοχών τοποθετούνται στα χαμηλότερα υψόμετρα. Στη συνέχεια στα χαμηλότερα σημεία των κοιλάδων τοποθετούνται πηγές νερού οι οποίες πλημμυρίζουν τους χώρους της εικόνας σχηματίζοντας έτσι υποδοχείς νερού. Τα όρη που έχουν σχηματιστεί λειτουργούν σαν φράγματα σε αυτή τη διαδικασία διαχωρίζοντας τους υποδοχείς μεταξύ τους. Με αυτόν το τρόπο γίνεται η τμηματοποίηση της εικόνας καθώς στο τέλος της διαδικασίας κάθε υποδοχέας αντιπροσωπεύει και μία περιοχή της εικόνας. Αυξομειώνοντας τη ποσότητα του νερού στις πηγές αλλάζει ο αριθμός των ακμών που λειτουργούν ως φράγματα οπότε ανάλογα μπορούμε να πάρουμε λιγότερες ή περισσότερες περιοχές κατά τη τμηματοποίηση.

1.2.4 Τμηματοποίηση με τη χρήση γράφων

Στη κατηγορία αυτή εντάσσονται οι μέθοδοι που βασίζονται στη θεωρία των γράφων για να επιτελέσουν τη τμηματοποίηση μίας εικόνας. Στο μοντέλο του γράφου που σχηματίζεται, τα εικονοστοιχεία αντιπροσωπεύονται από τους κόμβους ενώ το βάρος των ακμών που συνδέουν τους κόμβους συμβολίζει το βαθμό ομοιογένειας των γειτονικών εικονοστοιχείων. Όταν όλες οι ακμές αποκτήσουν μία τελική τιμή, τα τμήματα του γράφου που εμφανίζουν μεγαλύτερη συνεκτικότητα θεωρείται ότι αποτελούν από μία ξεχωριστή περιοχή στην εικόνα. Στα [36-40] παρουσιάζονται μερικές από αυτές τις μεθόδους.

1.2.5 Ενεργά περιγράμματα

Στη κατηγορία των ενεργών περιγραμμάτων ανήκουν τα φίδια [7], τα γεωδαιτικά ενεργά περιγράμματα [22] και τεχνικές από τη θεωρία των επιπεδοσυνόλων [23]. Συνήθως βασίζονται στην ελαχιστοποίηση ενός συναρτησιακού ενέργειας και επίσης είναι δυνατή η ενσωμάτωση αρχικής γνώσης. Είναι το βασικό μοντέλο που χρησιμοποιείται στην εργασία και σε επόμενο κεφάλαιο γίνεται αναλυτική περιγραφή των μεθόδων και της λειτουργίας τους.

1.3 Μηχανική μάθηση

Η μηχανική μάθηση αποτελεί κλάδο της τεχνητής νοημοσύνης και έχει στόχο να δώσει σε ένα σύστημα την ικανότητα της μάθησης. Σύμφωνα με τον ορισμό του Tom M. Mitchell “Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από μία εμπειρία E σε σχέση με ένα σύνολο εργασιών T και μία μέτρηση της απόδοσης P , εάν η απόδοση στις εργασίες του T , όπως μετρώνται από το P , βελτιώνεται με την εμπειρία

Ε” [52]. Η ικανότητα της μάθησης επιτρέπει την ανάπτυξη δυναμικών εφαρμογών οι οποίες δεν περιορίζονται στον σχεδιασμό του προγραμματιστή αλλά μέσα από την ανάλυση δεδομένων μπορούν να λαμβάνουν πρωτοβουλίες στην εκτέλεση εργασιών και παράλληλα, μέσω της εμπειρίας που αποκομίζουν, να εξελίσσονται.

Παρατηρώντας τη διαδικασία της μάθησης στον άνθρωπο γίνεται αντιληπτό ότι η αφαιρετική σκέψη παίζει κρίσιμο ρόλο καθώς συμβάλλει στη δημιουργία μοντέλων για την εκτίμηση καταστάσεων και οδηγεί στην εξαγωγή συμπερασμάτων. Η μηχανική μάθηση προσπαθεί να αντιγράψει αυτή τη διεργασία με τη χρήση τεχνικών στατιστικής ανάλυσης οι οποίες μέσα από ένα σύνολο δεδομένων έχουν τη δυνατότητα να ανακαλύπτουν συσχετίσεις και να δημιουργούν πρότυπα. Μία από τις πιο συνηθισμένες χρήσεις της σχετίζεται με τη πρόβλεψη αποτελέσματος όπως για παράδειγμα στη κατηγοριοποίηση (classification) και τη συσταδοποίηση (clustering). Ένας συγγενικός κλάδος με τον οποίο παρουσιάζει μεγάλη επικάλυψη είναι η εξόρυξη γνώσης (data mining) όπου αξιοποιούνται πολλές από τις τεχνικές της μηχανικής μάθησης με σκοπό τον εντοπισμό χαρακτηριστικών μέσα από ένα σύνολο δεδομένων.

Οι βασικές προσεγγίσεις που ακολουθούνται είναι η μάθηση με επίβλεψη (supervised) και η μάθηση χωρίς επίβλεψη (unsupervised). Στη πρώτη περίπτωση η εφαρμογή εκπαιδεύεται αρχικά με ένα σύνολο δεδομένων, το σύνολο εκπαίδευσης (training dataset), με τη βοήθεια του χρήστη ο οποίος για κάθε παράδειγμα/εργασία δίνει μία ετικέτα/αποτέλεσμα. Όταν τελειώσει η εκπαίδευση, η εφαρμογή εκτελεί την ίδια εργασία σε ένα νέο σύνολο, το σύνολο εξέτασης (testing dataset), βρίσκοντας μόνη της τις ετικέτες που πρέπει να αναθέσει. Στη δεύτερη περίπτωση ο χρήστης δεν συμμετέχει σε κανένα στάδιο της εκπαίδευσης καθώς η εφαρμογή καλείται μόνη της, μέσα από ένα άγνωστο σύνολο δεδομένων να βρει τις συσχετίσεις που υπάρχουν μεταξύ των παραδειγμάτων και να αναθέσει ετικέτες χωρίς να έχει κάποια προηγούμενη γνώση. Μερικές από τις πιο διαδεδομένες τεχνικές μηχανικής μάθησης είναι οι παρακάτω:

- ❖ Δέντρα απόφασης [41]
- ❖ Τεχνητά νευρωνικά δίκτυα [42]
- ❖ Μηχανές διανυσμάτων στήριξης (support vector machines, SVMs) [43]
- ❖ Δίκτυα Bayes [44]

1.4 Σύστημα φώνησης

Η διαδικασία με την οποία παράγεται η φωνή στον άνθρωπο αποκαλείται φώνηση και στηρίζεται σε τρία συστήματα, τους πνεύμονες, τον λάρυγγα και την φωνητική οδό. Τεχνικά, η φωνή είναι το αποτέλεσμα αλληλουχίας κυμάτων αέρα τα οποία παράγονται στους πνεύμονες, υπόκεινται σε κατάλληλες τροποποιήσεις κατά την διάρκεια της μεταφοράς τους και τελικά εξέρχονται από την στοματική κοιλότητα.

Η φωνητική οδός σχηματίζεται από τη φαρυγγική κοιλότητα, τη ρινοστοματική κοιλότητα, τη γλώσσα και τα χείλη και λειτουργεί σαν αντηχείο στο οποίο διαμορφώνεται ο ήχος που παράγεται από τις φωνητικές χορδές. Η συχνότητα της ανδρικής φωνής κυμαίνεται συνήθως μεταξύ 85 Hz και 180 Hz ενώ η συχνότητα μίας γυναικείας φωνής βρίσκεται στο διάστημα 165 – 255 Hz. Πρωταρχικός στόχος της φωνής είναι η ομιλία, μία διαδικασία που χρησιμεύει σε χιλιάδες πτυχές της καθημερινής ζωής ενός ατόμου και σχετίζεται κυρίως με την ανάγκη της επικοινωνίας. Επιπρόσθετα, η φωνή μπορεί να λειτουργήσει ως βιομετρικό γνώρισμα για την ταυτοποίηση ενός ατόμου καθώς ο συνδυασμός χαρακτηριστικών όπως η συχνότητα, η χροιά και ο τόνος είναι μοναδικός για κάθε άνθρωπο.

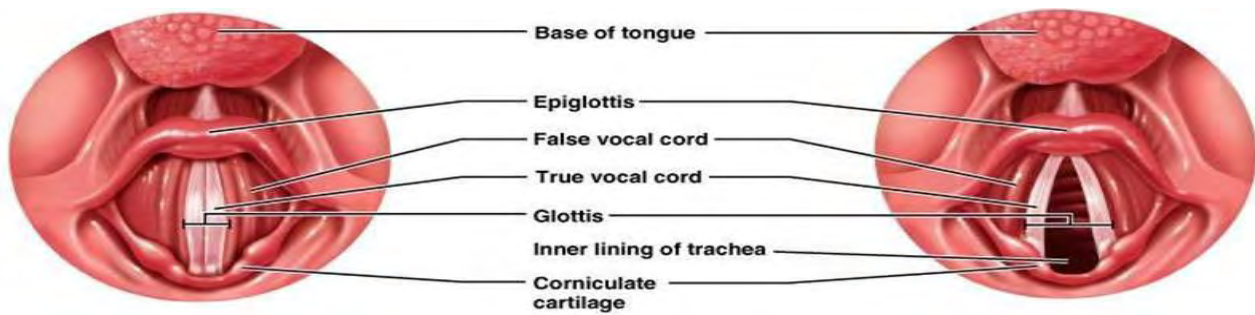
1.4.1 Ανατομία λάρυγγα

Κρίσιμο ρόλο στη λειτουργία της φωνής παίζει ο λάρυγγας. Βρίσκεται στην τραχηλική χώρα στο διάστημα των αυχενικών σπονδύλων και αποτελείται από χόνδρους, μύες και συνδέσμους.

Οι χόνδροι είναι εννέα στο σύνολό τους και διαχωρίζονται σε τρεις μονήρεις και τρεις συνεζευγμένους. Στους μονήρεις χόνδρους ανήκουν ο κρικοειδής, ο θυρεοειδής και η επιγλωττίδα (epiglottis) ενώ στους συνεζευγμένους ανήκουν ο αρυταινοειδής, ο κερατοειδής και ο σφηνοειδής. Η επιγλωττίδα σχηματίζει έναν ελαστικό χόνδρο σε σχήμα φύλλου που βρίσκεται στη βάση της γλώσσας και κλείνει το στόμιο της φαρυγγικής μοίρας του λάρυγγα. Ο θυρεοειδής αποτελείται από 2 τετράπλευρα πέταλα τα οποία προστατεύουν τις φωνητικές χορδές και ενώνονται προς τα εμπρός σχηματίζοντας το λεγόμενο λάρυγγικό έπαρμα ή μήλο του Αδάμ.

Οι μύες αποτελούνται από δύο μεγάλες ομάδες, τους ετερόχθονες και τους αυτόχθονες. Οι μύες της πρώτης ομάδας προσαρμόζουν το λάρυγγα στη θέση του και τον κινούν ως ενιαίο σώμα ενώ οι μύες της δεύτερης ομάδας έχουν πιο εξειδικευμένη λειτουργία και μεταξύ άλλων αναλαμβάνουν τη κίνηση των φωνητικών χορδών. Οι ετερόχθονες περιλαμβάνουν μύες όπως ο στερνοθυρεοειδής, ο στερνοϋοειδής και ο ωμοϋοειδής και στους αυτόχθονες ανήκουν ο κρικοθυρεοειδής, ο εγκάρσιος αρυταινοειδής, ο λοξός αρυταινοειδής, ο έσω θυρεοαρυταινοειδής και ο έξω θυρεοαρυταινοειδής [53].

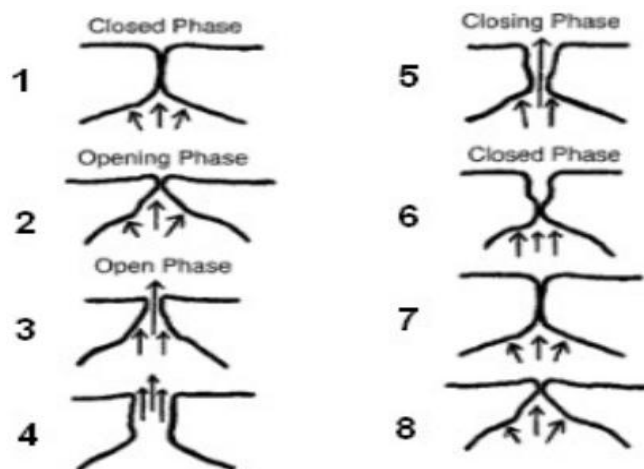
Επίσης στο λάρυγγα εμφανίζονται δύο ζεύγη πτυχών, το ζεύγος των φωνητικών χορδών (true vocal cords) που βρίσκεται στο κέντρο του λάρυγγα και το ζεύγος των ψευδών ή νόθων φωνητικών χορδών (false vocal cords) όπου οι πτυχές τοποθετούνται εκατέρωθεν των φωνητικών χορδών. Η περιοχή μεταξύ των δύο φωνητικών χορδών ονομάζεται γλωττίδα (glottis) και αποτελεί την οπή μέσω της οποίας διέρχεται ο αέρας που δημιουργεί τον ήχο της φωνής.



Εικόνα 1.1: Τμήματα της ανατομίας του λάρυγγα [54].

1.4.2 Διαδικασία φώνησης

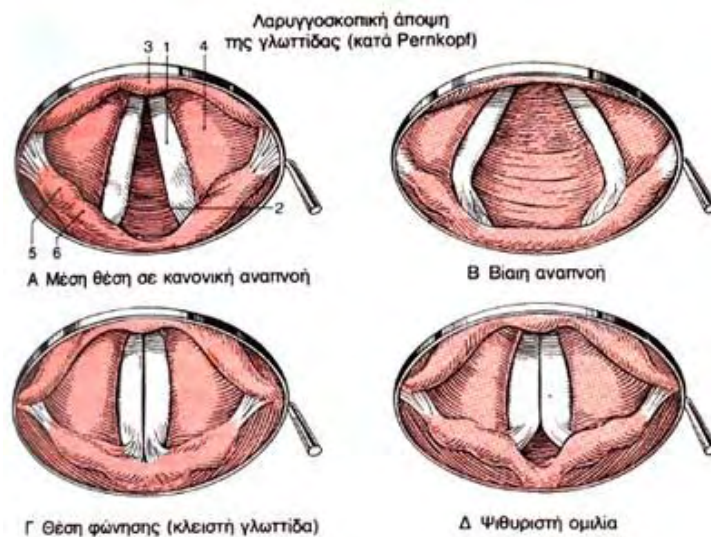
Το δημοφιλέστερο μοντέλο που ερμηνεύει τη διεργασία της φώνησης είναι η μυοελαστική αεροδυναμική θεωρία η οποία παρουσιάστηκε από τον Van den Berg [1] και βελτιώθηκε από τον Titze [2]. Το μοντέλο βασίζεται στις ελαστικές ιδιότητες των φωνητικών χορδών και στο φαινόμενο Bernoulli. Αρχικά οι φωνητικές χορδές βρίσκονται σε ισορροπία και η γλωττίδα είναι κλειστή. Ο εκπνεόμενος από τους πνεύμονες αέρας αυξάνει την υπογλωττιδική πίεση οπότε κάποια στιγμή κάμπτεται η αντίσταση των φωνητικών χορδών με αποτέλεσμα την απώθηση των δύο χορδών και το άνοιγμα της γλωττίδας. Καθώς ο αέρας περνά από τη στενότερη περιοχή της γλωττίδας παρατηρείται πτώση της πίεσης κάθετα προς τη κατεύθυνση ροής του, απόρροια του φαινομένου Bernoulli. Η πτώση της πίεσης σε συνδυασμό με την ελαστικότητα των φωνητικών χορδών οδηγούν στην επαναφορά των φωνητικών χορδών στην αρχική θέση ισορροπίας κλείνοντας έτσι έναν κύκλο ταλάντωσης. Η ελαστικότητα των φωνητικών χορδών επιτρέπουν σε αυτές να ταλαντώνονται καθώς ο αέρας διέρχεται από την γλωττίδα με αποτέλεσμα η φωνή να αποκτά χαρακτηριστικά όπως η συχνότητα και η ποιότητα.



Εικόνα 1.2: Οι φάσεις του κύκλου ταλάντωσης των φωνητικών χορδών. Απεικονίζεται η υπογλωττιδική πίεση που ασκείται κατά την κλειστή φάση (closed phase), την φάση ανοίγματος (opening phase), την ανοιχτή φάση (open phase), και την φάση κλεισίματος (closing phase) [55].

1.4.3 Στροβοσκόπηση

Ασθένειες όπως η χρόνια λαρυγγίτιδα, η οξεία λαρυγγίτιδα, η λαρυγγική παλινδρόμηση, ο καρκίνος του λάρυγγα και η εμφάνιση κυστών ή πολύποδων μπορούν να οδηγήσουν σε δυσλειτουργίες της φώνησης. Η λαρυγγοσκόπηση με στροβοσκόπηση είναι μία συνηθισμένη και αποτελεσματική μέθοδος που εφαρμόζεται για την εξέταση του λάρυγγα καθώς επιτρέπει την οπτική παρακολούθηση της λειτουργίας των φωνητικών χορδών και της περιοχής του λάρυγγα. Όπως επισημάνθηκε προηγουμένως, η ανθρώπινη φωνή έχει υψηλή συχνότητα καθώς στους άνδρες είναι περίπου 100 Hz και στις γυναίκες 220 Hz. Συνεπώς η δόνηση των φωνητικών χορδών είναι τόσο μεγάλη που είναι αδύνατον να γίνει αντιληπτή από το ανθρώπινο μάτι. Το πρόβλημα αυτό μπορεί να ξεπεραστεί με τη μέθοδο της στροβοσκόπησης. Η συγκεκριμένη μέθοδος περιλαμβάνει τη χρήση ενός λαρυγγοσκοπίου το οποίο όμως δεν παράγει συνεχόμενο φως αλλά διακοπτόμενο (flashing light) με συχνότητα περίπου 2 Hz. Επομένως το στροβοσκοπικό φως φωτίζει κάθε φορά διαφορετικά σημεία του κύκλου ταλάντωσης των φωνητικών χορδών επιτρέποντας έτσι στον ιατρό να παρακολουθήσει τη κίνησή τους. Κατά τη στροβοσκόπηση ο ιατρός δίνει μεγάλη σημασία στις εναλλαγές του σχήματος της γλωττίδας για να εντοπίσει πιθανά προβλήματα και να εξάγει συμπεράσματα όσον αφορά την λειτουργία του μηχανισμού.



Εικόνα 1.3: Η μορφή της γλωττίδας κατά τη διάρκεια λειτουργιών του λάρυγγα [56].

1.5 Σχετικές εργασίες

Η παρατήρηση του σχήματος της γλωττίδας μπορεί να δώσει πολύτιμες πληροφορίες σχετικά με το σύστημα φώνησης. Για αυτόν τον λόγο υπάρχει έντονο ενδιαφέρον για την ανάπτυξη μεθόδων που τμηματοποιούν τη γλωττίδα με απώτερο στόχο την πλήρη αυτοματοποίηση της διαδικασίας και τη διευκόλυνση της ιατρικής διάγνωσης.

Οι Karakozoglou et al. [3] προτείνουν μία μέθοδο για αυτοματοποιημένη κατάτμηση γλωττίδας σε μία αλληλουχία εικόνων η οποία βασίζεται στα τοπικά ενεργά περιγράμματα. Αρχικά αποφασίζεται η περιοχή ενδιαφέροντος (Region of Interest, ROI) εξετάζοντας τις εικόνες από έναν κύκλο ταλάντωσης των φωνητικών χορδών. Επομένως περιοχή της εικόνας όπου παρατηρείται η μέγιστη διαφοροποίηση και η χαμηλότερη ένταση φωτεινότητας ορίζει την περιοχή ενδιαφέροντος. Έπειτα, χρησιμοποιείται η μέθοδος των Lankton-Tannenbaum [4] για την καθοδήγηση του ενεργού περιγράμματος το οποίο δίνει τη τελική τμηματοποίηση της γλωττίδας. Οι Palm et al. [5] εισάγουν ένα μοντέλο μπαλονιού (balloon model) το οποίο τμηματοποιεί τις φωνητικές χορδές και τη περιοχή της γλωττίδα. Η μέθοδος βασίζεται σε τρεις δυνάμεις που αξιοποιούν πληροφορίες της εικόνας για να αρχικοποιήσουν χαρακτηριστικά του μοντέλου όπως η τάση και η ομαλότητα. Ο αλγόριθμος αναζητεί πρώτα τη σκοτεινότερη περιοχή της εικόνας για να εντοπίσει το χώρο των φωνητικών χορδών και της γλωττίδας. Με τη βοήθεια ενός προγενέστερου σχήματος το μοντέλο μπαλονιού υπολογίζει τα εξωτερικά όρια του περιγράμματος για τις φωνητικές χορδές και μετά οριοθετεί τη γλωττίδα η οποία βρίσκεται στο εσωτερικό του περιγράμματος. Στο τελικό στάδιο γίνεται ο διαχωρισμός των δύο τμημάτων του ζεύγους των φωνητικών χορδών. Η μέθοδος παρουσιάζει ορισμένα μειονεκτήματα με σημαντικότερο το γεγονός πως επικεντρώνεται κυρίως στη τμηματοποίηση των φωνητικών χορδών με αποτέλεσμα όταν το σχήμα της γλωττίδας είναι λεπτό να αδυνατεί να την τμηματοποιήσει επιτυχώς. Ακόμα η μέθοδος είναι αρκετά ευαίσθητη στη παραμετροποίηση και επίσης δεν είναι πλήρως αυτοματοποιημένη οπότε απαιτεί από τον χρήστη να ορίσει τη περιοχή ενδιαφέροντος. Στο [6] οι συγγραφείς χρησιμοποιούν ένα ενεργό περίγραμμα (φίδι, snake) για να παρακολουθήσουν τις δονούμενες φωνητικές χορδές. Όπως αναφέρουν, βελτιώνουν τον αλγόριθμο για το κλασικό φίδι του Kass et al. [7] τροποποιώντας τη συνάρτηση εξωτερικής ενέργειας και προσθέτοντας δύο νέους όρους στη συνάρτηση εσωτερικής ενέργειας. Στο [73] η μέθοδος που υλοποιείται κάνει κατάτμηση γλωττίδας συνδυάζοντας τα ενεργά περιγράμματα με τα προγενέστερα σχήματα. Το ενεργό περίγραμμα βασίζεται στο φίδι του Kass et al. ενώ για τη δημιουργία των προγενέστερων σχημάτων χρησιμοποιείται η τεχνική της PCA.

Οι Cerrolazza et al. [8] παρουσιάζουν μία αυτοματοποιημένη μέθοδο για την κατάτμηση γλωττίδας με τη χρήση των Μοντέλων Ενεργών Σχημάτων (Active Shape Models, ASMs). Η τεχνική ακολουθεί πέντε βήματα. Καταρχήν, με τη μέθοδο της επέκτασης περιοχής (region-growing method) πραγματοποιεί μία κατά προσέγγιση

τμηματοποίηση που οδηγεί σε μία δυαδική εικόνα. Στη συνέχεια, η δυαδική εικόνα φιλτράρεται με σκοπό να οριστούν πιθανές περιοχές όπου θα μπορούσε να βρεθεί η γλωττίδα και άρα οι περιοχές αυτές λειτουργούν ως υποψήφιοι. Έπειτα εφαρμόζεται το μοντέλο ASM για κάθε περιοχή παράγοντας σε κάθε περίπτωση από ένα περίγραμμα το οποίο παρέχει σημαντικές στατιστικές πληροφορίες για το σχήμα τους και την εμφάνισή τους. Στο τέλος, εκτιμώνται όλα τα περιγράμματα και εκείνο που αξιολογείται ως καλύτερο οριοθετεί τη περιοχή της γλωττίδας.

Στο [9] προτείνεται μία άλλη μέθοδος αυτοματοποιημένης κατάτμησης γλωττίδας η οποία βασίζεται στο μετασχηματισμό watershed. Ένα συχνό πρόβλημα της τεχνικής watershed είναι η υπερτμηματοποίηση η οποία οδηγεί σε υπερβολικά μεγάλο πλήθος αντικειμένων. Για να υπερκεραστεί το πρόβλημα, η μέθοδος συνδυάζει το watershed με επιπλέον τεχνικές επεξεργασίας εικόνας. Στη φάση της προεπεξεργασίας, ο αλγόριθμος κατωφλιώνει την κλίση της εικόνα (gradient image) έτσι ώστε ασθενείς ακμές που οφείλονται σε θόρυβο να εξαλειφθούν. Στη φάση της μετεπεξεργασίας εφαρμόζεται η τεχνική Just Noticeable Difference (JND) με αποτέλεσμα περιοχές που εμφανίζουν παρόμοια ένταση φωτεινότητας να συγχωνεύονται. Σε επόμενο βήμα γίνονται επιπλέον συγχωνεύσεις γειτονικών περιοχών σύμφωνα με την ιδέα ότι η γλωττίδα σχηματίζει μία σκούρα περιοχή η οποία περιβάλλεται από φωτεινότερες περιοχές. Στο τέλος, ένας εκπαιδευμένος γραμμικός μάντης (linear predictor) αποφασίζει ποιά από τα εναπομείναντα αντικείμενα είναι πιθανότερο να αναπαριστά το χώρο της γλωττίδας. Παρατηρούμε όμως ότι δεν μπορούμε να βγάλουμε ασφαλή συμπεράσματα για την αποτελεσματικότητα της μεθόδου καθώς δοκιμάστηκε σε μικρό αριθμό εικόνων και επίσης χρησιμοποιήθηκαν μόλις δύο βίντεο στροβοσκόπησης. Συνεπώς είναι πιθανό να υπάρχουν διαφορετικές περιπτώσεις εικόνων όπου η μέθοδος δεν λειτουργεί το ίδιο αποδοτικά.

Η μέθοδος στο [10] επικεντρώνεται στην αυτοματοποιημένη παρακολούθηση της κίνησης των φωνητικών χορδών εκτελώντας παράλληλα τμηματοποίηση της γλωττίδας. Ο αλγόριθμος υποθέτει ότι η εικόνα ακολουθεί κατανομή Rayleigh και κατωφλιώνει την εικόνα παράγοντας μία δυαδική εικόνα. Έπειτα, με τη τεχνική erosion απομακρύνονται ψεύτικα σημεία που βρέθηκαν και είναι εκτός γλωττίδας. Συνεπώς μένει μία δυαδική εικόνα στην οποία έχει τμηματοποιηθεί ένα μέρος της γλωττίδας. Τέλος, με τη βοήθεια της μεθόδου επέκτασης περιοχής γίνεται η πλήρης κατάτμηση και ορίζεται η περιοχή γλωττίδας. Ωστόσο για να καταφέρει η τεχνική να κάνει επιτυχώς την κατάτμηση, θα πρέπει η περιοχή ενδιαφέροντος να είναι σημαντικά πιο σκούρα από τις υπόλοιπες περιοχές της εικόνας ώστε να μπορέσει να εντοπίσει τη τοποθεσία της γλωττίδας.

Συγγραφείς	Παραπομπές	Μέθοδος	Χρονολογία
Palm et al.	[5]	Μοντέλο μπαλονιού, προγενέστερα σχήματα	2001
Marendic et al.	[6]	Ενεργά περιγράμματα, μοντέλο φιδιού	2001
Yan et al.	[10]	Κατωφλίωση, μέθοδος επέκτασης περιοχής	2006
Osma-Ruiz et al.	[9]	Μετασχηματισμός watershed, τεχνική JND	2008
Cerrolazza et al.	[8]	Μοντέλο ενεργών σχημάτων, μέθοδος επέκτασης περιοχής	2011
Karakozoglou et al.	[3]	Τοπικά ενεργά περιγράμματα	2011
Yan et al.	[73]	Ενεργά περιγράμματα, προγενέστερα σχήματα	2012

Πίνακας 1: Πίνακας σχετικών εργασιών

Τεχνικές Στατιστικής Ανάλυσης Δεδομένων Εικόνων

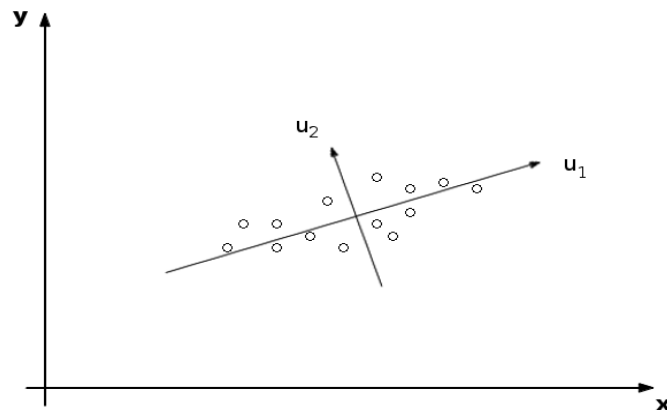
Πολλές φορές τα δεδομένα που συλλέγονται είναι περίπλοκα, γεγονός που τα καθιστά δύσκολα στην ερμηνεία τους. Σε αυτές τις περιπτώσεις είναι αναγκαία η χρήση τεχνικών στατιστικής ανάλυσης οι οποίες μπορούν να αποκαλύψουν τα κρυμμένα μοτίβα πίσω από τα δεδομένα και να υποδείξουν ποιό είναι οι πιο σημαντικοί άξονες στους οποίους πρέπει να επικεντρωθεί κάποιος για την κατανόηση του προβλήματος. Στο κεφάλαιο αυτό εξετάζουμε δύο τέτοιες τεχνικές, την Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis, PCA) και την Ανάλυση Κύριων Συνιστωσών Πυρήνα (Kernel Principal Component Analysis, Kernel PCA).

2.1 Ανάλυση κύριων συνιστωσών

Μία σημαντική τεχνική στατιστικής ανάλυσης αποτελεί η PCA, μία μέθοδος που επιτρέπει την εύρεση των βασικών δομών από ένα σύνολο δεδομένων υψηλών διαστάσεων, με τελικό στόχο την εξαγωγή χαρακτηριστικών (feature extraction) και τη μείωση διαστάσεων (dimensionality reduction). Η PCA δημοσιεύτηκε από τον Karl Pearson [11] ως ένα ανάλογο του θεωρήματος των βασικών αξόνων που χρησιμοποιείται στη μηχανική. Οι δυνατότητες που προσφέρει την έχουν κάνει ιδιαίτερα δημοφιλή σε ένα ευρύ φάσμα επιστημονικών πεδίων από τη νευροφυσιολογία και την επεξεργασία σημάτων μέχρι την όραση υπολογιστών.

Μία εικόνα με διαστάσεις $A \times B$ μπορεί να αναπαρασταθεί από ένα διάνυσμα διαστάσεων $d=AB$, οπότε ένα σύνολο εικόνων απεικονίζεται γραφικά από ένα σύνολο σημείων στο χώρο R^d . Σκοπός της PCA είναι να βρει τους άξονες εκείνους που περιγράφουν καλύτερα το σύνολο των εικόνων και αποτυπώνουν τις μεταβολές που συμβαίνουν μεταξύ τους. Στατιστικά αυτό σημαίνει ότι κάθε άξονας δίνεται από την ευθεία εκείνη για την οποία το σύνολο των προβολών των σημείων προς αυτή μεγιστοποιείται. Οι ευθείες αυτές ορίζουν τα διανύσματα των κύριων συνιστωσών (principal components) και ισχύει ότι κάθε εικόνα μπορεί να εκφραστεί ως γραμμικός συνδυασμός των κύριων συνιστωσών. Συνολικά υπάρχουν d κύριες συνιστώσες αλλά δεν συμμετέχουν όλες ισοδύναμα στην ανακατασκευή μίας εικόνας. Μάλιστα παρατηρείται ότι ένας μικρός αριθμός αυτών συμμετέχει πάντα σε πολύ μεγάλο βαθμό ενώ οι υπόλοιπες κύριες συνιστώσες εμφανίζουν πολύ μικρό βαθμό συμμετοχής.

Επομένως μία εικόνα μπορεί να εκφρασθεί μόνο από ένα μικρό αριθμό κύριων συνιστωσών ενώ η απώλεια πληροφορίας που συμβαίνει θεωρείται αμελητέα. Σε γενικές γραμμές μπορούμε να πούμε ότι χρησιμοποιώντας μόνο τις 3 πρώτες (σημαντικότερες) κύριες συνιστώσες ανακτάται συνήθως το 98-99% της αρχικής εικόνας. Με αυτό τον τρόπο επιτυγχάνεται η μείωση διαστάσεων με αποτέλεσμα την εξοικονόμηση αποθηκευτικού χώρου και τη μείωση του υπολογιστικού κόστους. Συνάμα εντοπίζονται τα στοιχεία που παρουσιάζουν μεγαλύτερη διαφοροποίηση καταλήγοντας έτσι στην εξαγωγή των χαρακτηριστικών του συνόλου δεδομένων.



Εικόνα 2.1: Στο διάγραμμα φαίνονται 14 διανύσματα που ανήκουν στον \mathbb{R}^2 και οι κύριες συνιστώσες που υπολογίζονται με τη PCA. Η u_1 είναι η πρώτη κύρια συνιστώσα και η u_2 είναι η δεύτερη κύρια συνιστώσα.

2.1.1 Υπολογισμός ανάλυσης κύριων συνιστωσών

Ας υποθέσουμε ότι έχουμε ένα σύνολο από M εικόνες x_1, x_2, \dots, x_M σε μορφή διανυσμάτων-στηλών και κάθε διάνυσμα έχει διαστάσεις d . Φτιάχνουμε το πίνακα X με διαστάσεις $d \times M$ και στις στήλες του τοποθετούμε τα διανύσματα x_i . Αρχικά χρησιμοποιώντας τη σχέση

$$X_\mu = \frac{1}{M} \sum_{i=1}^M X_i \quad (2.1)$$

βρίσκουμε το διάνυσμα της μέσης εικόνας (mean image) x_μ . Αφαιρώντας από κάθε εικόνα την μέση εικόνα κεντράρουμε τον X , οπότε στο νέο πίνακα κάθε στήλη i προκύπτει από τη σχέση $x_i - x_\mu$ και αφού ο X κεντραριστεί θα ισχύει για τη μέση εικόνα ότι $x_\mu = 0$.

Στη συνέχεια βρίσκουμε το πίνακα συνδιακύμανσης C για τον X σύμφωνα με τη σχέση

$$C = \frac{1}{M} X X^T \quad (2.2)$$

Τέλος από τη σχέση

$$Cu = \lambda u \quad (2.3)$$

υπολογίζουμε τις λ_i ιδιοτιμές και τα u_i ιδιοδιανύσματα του C και κανονικοποιούμε τα ιδιοδιανύσματα.

Τα ιδιοδιανύσματα u_i αποτελούν τις κύριες συνιστώσες και δείχνουν τους βασικούς άξονες που περιγράφουν καλύτερα τα δεδομένα μας. Οι ιδιοτιμές δείχνουν το βαθμό σημαντικότητας της αντίστοιχης κύριας συνιστώσας και άρα όσο μεγαλύτερο είναι το λ_i τόσο πιο σημαντική είναι η κύρια συνιστώσα. Η επόμενη κίνηση λοιπόν είναι να ιεραρχήσουμε τις κύριες συνιστώσες βάσει των λ_i . Από τις ιδιοτιμές εξετάζονται μόνο όσες έχουν πραγματικές τιμές ενώ οι μιγαδικές δεν λαμβάνονται υπ'όψιν. Τα διανύσματα των κύριων συνιστωσών είναι πάντα κάθετα μεταξύ τους.

2.1.2 Ανάλυση κύριων συνιστωσών μέσω της αποσύνθεσης μοναδικής τιμής

Ένα πρόβλημα που μπορεί να παρατηρήσει κάποιος στη παραπάνω διαδικασία είναι η τάξη μεγέθους του πίνακα συνδιακύμανσης. Αν για παράδειγμα έχουμε ένα σύνολο από 200 εικόνες με κάθε εικόνα να έχει διαστάσεις 100×100 τότε τα διανύσματα x_i θα έχουν διαστάσεις 10^4 , ο πίνακας X διαστάσεις $10^4 \times 200$ και τελικά ο πίνακας συνδιακύμανσης C θα έχει διαστάσεις $10^4 \times 10^4$. Οι διαστάσεις του C εξαρτώνται αποκλειστικά από τις διαστάσεις των εικόνων και για τις τυπικές εικόνες το υπολογιστικό κόστος είναι τεράστιο. Η μέθοδος της Αποσύνθεσης Μοναδικής Τιμής (Singular Value Decomposition, SVD) δίνει λύση σε αυτό το πρόβλημα χρησιμοποιώντας άλλους πίνακες χαμηλότερης τάξης αντί του πίνακα συνδιακύμανσης.

Υπολογισμός

Υποθέτουμε ότι υπάρχουν 3 πίνακες U , Σ , V τέτοιοι ώστε

$$X = U\Sigma V^T \quad (2.4)$$

όπου ο U έχει διαστάσεις $d \times d$, ο Σ είναι διαγώνιος πίνακας με διαστάσεις $d \times M$ και ο V έχει διαστάσεις $M \times M$. Επίσης ορίζουμε ότι οι U και V είναι ορθοκανονικοί οπότε $UU^T = I$ και $VV^T = I$.

Από την (2.4) προκύπτουν οι σχέσεις

$$XX^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma \Sigma U^T = U\Sigma^2 U^T \Rightarrow XX^T = U\Sigma^2 U^T \quad (2.5)$$

$$X^T X = (U\Sigma V^T)^T U\Sigma V^T = V \Sigma^T U^T U \Sigma V^T = V \Sigma \Sigma V^T = V \Sigma^2 V^T \Rightarrow X^T X = V \Sigma^2 V^T \quad (2.6)$$

Οι (2.5) και (2.6) μπορούν να γραφούν επίσης ως

$$\begin{aligned}(\mathbf{X}\mathbf{X}^T)\mathbf{U} &= \Sigma^2\mathbf{U} \\ (\mathbf{X}^T\mathbf{X})\mathbf{V} &= \Sigma^2\mathbf{V}\end{aligned}$$

Παρατηρούμε ότι οι δύο πίνακες $\mathbf{X}\mathbf{X}^T$ και $\mathbf{X}^T\mathbf{X}$ μοιράζονται τις ίδιες ιδιοτιμές και είναι τα τετράγωνα των στοιχείων της διαγωνίου του Σ . Συγκεκριμένα έχουν M κοινές ιδιοτιμές.

Εφαρμόζοντας τη PCA με τις εξισώσεις του SVD έχουμε

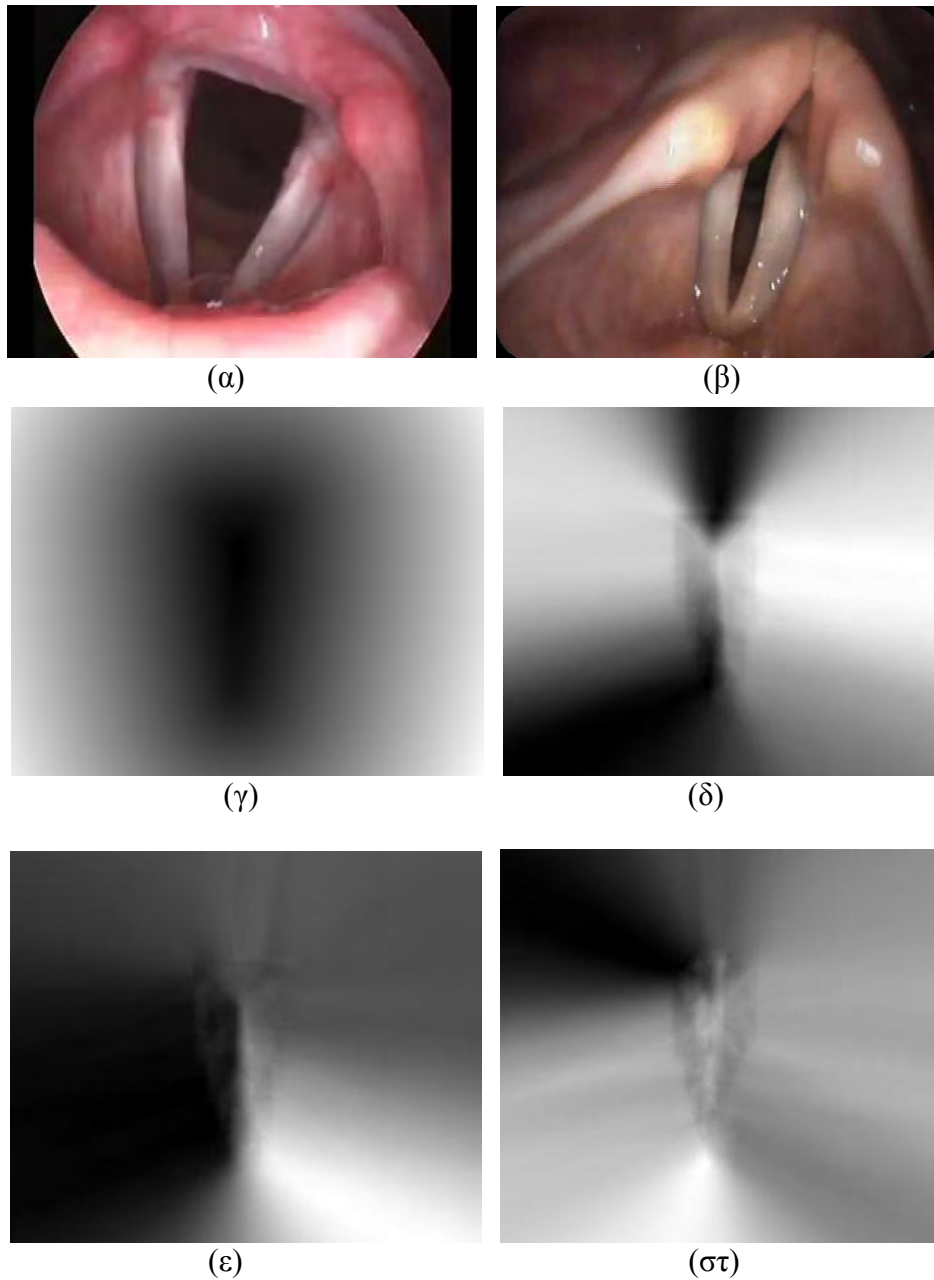
$$\mathbf{C} = \frac{1}{M} \mathbf{X}\mathbf{X}^T \Leftrightarrow \mathbf{C} = \frac{1}{M} \mathbf{U}\Sigma^2\mathbf{U}^T \Leftrightarrow \mathbf{C}\mathbf{U} = \left(\frac{1}{M} \Sigma^2\right)\mathbf{U} \quad (2.7)$$

Επομένως ο \mathbf{U} περιέχει τα ιδιοδιανύσματα του \mathbf{C} και άρα αν βρούμε τον \mathbf{U} μπορούμε να υπολογίσουμε τις κύριες συνιστώσες.

Από τη σχέση (2.6) υπολογίζοντας ιδιοτιμές-ιδιοδιανύσματα βρίσκουμε τους Σ και \mathbf{V} . Στη συνέχεια, λύνοντας την (2.4) ως προς \mathbf{U} βρίσκουμε τις κύριες συνιστώσες.

$$\begin{aligned}\mathbf{X} &= \mathbf{U}\Sigma\mathbf{V}^T \\ \mathbf{X}(\Sigma\mathbf{V}^T)^{-1} &= \mathbf{U}\Sigma\mathbf{V}^T(\Sigma\mathbf{V}^T)^{-1} \\ \mathbf{U} &= \mathbf{X}(\mathbf{V}^T)^{-1}\Sigma^{-1} \\ \mathbf{U} &= \mathbf{X}(\mathbf{V}^{-1})^{-1}\Sigma^{-1} \\ \mathbf{U} &= \mathbf{X}\mathbf{V}\Sigma^{-1}\end{aligned}$$

Τέλος, γίνεται η ιεράρχηση των κύριων συνιστωσών. Όπως μπορούμε να δούμε στη (2.7) οι ιδιοτιμές βρίσκονται στη διαγώνιο του πίνακα $\frac{1}{M}\Sigma^2$. Δεν είναι απαραίτητο όμως να βρούμε επακριβώς τις ιδιοτιμές καθώς αυτό που μας ενδιαφέρει είναι η ταξινόμησή τους και άρα είτε χρησιμοποιήσουμε το πίνακα $\frac{1}{M}\Sigma^2$ είτε τον Σ η ταξινόμηση που προκύπτει είναι η ίδια. Συνεπώς ιεραρχούμε τις κύριες συνιστώσες ανάλογα με τη ταξινόμηση των διαγώνιων στοιχείων του Σ .

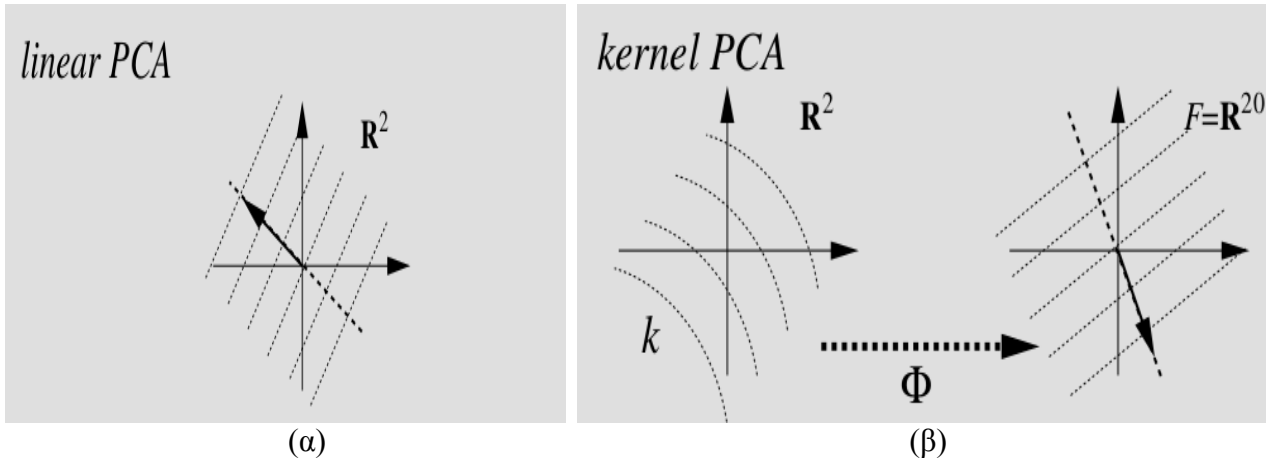


Εικόνα 2.2: Παραπάνω φαίνονται δύο αντιπροσωπευτικές εικόνες γλωττίδας που ελήφθησαν με στροβοσκόπηση και οι τέσσερις εικόνες που συνθέτουν ένα στατιστικό μοντέλο για τη γλωττίδα με τη τεχνική της PCA. Η (α) δείχνει τη γλωττίδα στην ανοιχτή φάση και η (β) στη φάση κλεισίματος. Στην (γ) φαίνεται η μέση εικόνα και στις (δ),(ε),(στ) η πρώτη, δεύτερη και τρίτη κύρια συνιστώσα αντίστοιχα. Στη μέση εικόνα όσο μικρότερη είναι η ένταση φωτεινότητας τόσο μεγαλύτερη είναι η πιθανότητα το συγκεκριμένο εικονοστοιχείο να ανήκει στο σχήμα της γλωττίδας. Στις κύριες συνιστώσες οι περιοχές της εικόνας με τη μεγαλύτερη ένταση φωτεινότητας υποδεικνύουν τις περιοχές του σχήματος που είναι πιθανότερο να παρουσιάζουν μεγαλύτερη διαφοροποίηση.

2.2 Ανάλυση κύριων συνιστωσών πυρήνα

Η PCA έχει χρησιμοποιηθεί ως βάση για την ανάπτυξη βελτιωμένων αλγορίθμων όπως για παράδειγμα την Ανάλυση Ανεξάρτητων Συνιστωσών (Independent Component Analysis, ICA) [12] και την Ανάλυση Γραμμικής Διακρίνουσας (Linear Discriminant Analysis, LDA) [13]. Και οι τρεις μέθοδοι στηρίζονται στις στατιστικές χαμηλής τάξεως καθώς εντοπίζουν τις εξαρτήσεις που ενυπάρχουν μεταξύ δυο εικονοστοιχείων από το δείγμα των εικόνων. Οι εξαρτήσεις αυτές είναι επίσης γνωστές ως γραμμικές συσχετίσεις. Πιο σύνθετα μοτίβα όμως, όπως για παράδειγμα αυτά που εμφανίζονται στα γεωμετρικά σχήματα μίας εικόνας, δεν μπορούν να εντοπιστούν από αυτό το είδος των μεθόδων. Σε αυτή τη περίπτωση είναι αναγκαία η χρήση στατιστικών υψηλότερης τάξεως (Higher Order Statistics, HOS). Σε αυτή τη προσέγγιση ιδιαίτερο ενδιαφέρον παρουσιάζει ο αλγόριθμος της Kernel PCA ή συντομογραφικά KPCA, ο οποίος έχει τη δυνατότητα να εντοπίζει μη γραμμικές συσχετίσεις μεταξύ των εικονοστοιχείων, προσφέροντας έτσι αποτελέσματα παρόμοια και σε ορισμένες περιπτώσεις καλύτερα από αυτά που προκύπτουν με τους αλγορίθμους που εξετάζουν μόνο τις γραμμικές συσχετίσεις.

Η Kernel PCA προτάθηκε από τους Scholkopf, Smola και Muller [14] και λειτουργεί ως η μη γραμμική μορφή της PCA. Η PCA πραγματοποιεί γραμμικό διαχωρισμό των δεδομένων στον χώρο R^d ο οποίος αποκαλείται χώρος εισόδου (Input space) και συμβολίζεται με I . Η KPCA έχει ως στόχο να υλοποιήσει ένα μη γραμμικό διαχωρισμό στον I . Για να το επιτύχει αυτό μεταφέρει τα δεδομένα μέσω μίας συνάρτησης χαρτογράφησης Φ , σε ένα χώρο υψηλότερων διαστάσεων ο οποίος αποκαλείται χώρος χαρακτηριστικών (Feature space) και συμβολίζεται με F . Ο F είναι ένας χώρος Hilbert που παράγεται από τα εσωτερικά γινόμενα των διανυσμάτων των εικόνων και μπορεί να έχει πεπερασμένες ή και άπειρες διαστάσεις. Στον χώρο F εφαρμόζεται η PCA και υπολογίζονται οι κύριες συνιστώσες με αποτέλεσμα τον γραμμικό διαχωρισμό των δεδομένων. Όταν οι κύριες συνιστώσες μεταφερθούν πίσω στον χώρο I τότε εμφανίζονται ως καμπύλες. Αυτά τα διανύσματα αποκαλούνται κύριες συνιστώσες πυρήνα (kernel principal components) και χάρη στη καμπυλόγραμμη μορφή τους έχουν τη δυνατότητα να εντοπίζουν τόσο τις γραμμικές όσο και τις μη γραμμικές συσχετίσεις. Η εικόνα (2.3) παρουσιάζει αυτήν την ιδέα.



Εικόνα 2.3: Στις δύο εικόνες εφαρμόζουμε PCA και Kernel PCA σε ένα δείγμα διανυσμάτων που ανήκουν στον R^2 . Στην αριστερή εικόνα εφαρμόζεται η γραμμική PCA στον χώρο I και προκύπτουν οι κύριες συνιστώσες οι οποίες έχουν γραμμική μορφή. Στη δεξιά εικόνα εφαρμόζεται η KPCA. Στο δεξιό σχήμα τα διανύσματα έχουν μεταφερθεί από τον I χώρο (R^2) στον F χώρο (στη προκειμένη περίπτωση R^{20}) και εκεί εκτελείται η γραμμική PCA. Στο αριστερό διάγραμμα οι κύριες συνιστώσες μεταφέρονται από τον F πίσω στον I οπότε μετατρέπονται σε κύριες συνιστώσες πυρήνα και αποκτούν μη γραμμική μορφή [14].

2.2.1 Υπολογισμός ανάλυσης κύριων συνιστωσών πυρήνα

Δεχόμαστε πάλι ότι έχουμε ένα σύνολο από M διανύσματα εικόνων $x_k \in R^d$ με $k=1, \dots, M$, τα οποία φτιάχνουν τον πίνακα X . Ο χώρος R^d λειτουργεί ως χώρος εισόδου I . Για να εκτελέσουμε τη PCA βρίσκουμε πρώτα το πίνακα συνδιακύμανσης C και ύστερα τις κύριες συνιστώσες. Στη Kernel PCA αρχικά μεταφέρονται τα δεδομένα μέσω μίας συνάρτησης Φ από τον R^d στον F

$$\begin{aligned} \Phi : R^d &\rightarrow F \\ x &\rightarrow \Phi(x) . \end{aligned}$$

Ας υποθέσουμε για αρχή ότι τα δεδομένα είναι κεντραρισμένα στον χώρο χαρακτηριστικών (η διαδικασία του κεντραρίσματος αναλύεται αργότερα), άρα $\sum_{k=1}^M \Phi(x_k) = 0$. Επομένως ο πίνακας συνδιακύμανσης στον F είναι

$$C = \frac{1}{M} \sum_{j=1}^M \Phi(x_j)\Phi(x_j^T) \tag{2.8}$$

Οπότε για να εκτελέσουμε την PCA στον F αρκεί να βρούμε τις πραγματικές ιδιοτιμές $\lambda \geq 0$ και τα ιδιοδιανύσματα $V \in F$ που ικανοποιούν τη σχέση

$$CV = \lambda V \tag{2.9}$$

Παρατηρούμε ότι οι λύσεις του V παράγονται από το σύνολο $\{ \Phi(x_1), \dots, \Phi(x_M) \}$ άρα το σύνολο των $\Phi(x_i)$ λειτουργεί ως η γραμμική θήκη του V . Αυτό σημαίνει ότι

υπάρχουν συντελεστές α_i με $i=1, \dots, M$ τέτοιοι ώστε

$$\mathbf{V} = \sum_{i=1}^M \alpha_i \Phi(\mathbf{x}_i) \quad (2.10)$$

Πολλαπλασιάζοντας και τα δύο μέλη της (2.9) με το $\Phi(\mathbf{x}_k)$ προκύπτει η εξίσωση

$$\lambda(\Phi(\mathbf{x}_k) \cdot \mathbf{V}) = (\Phi(\mathbf{x}_k) \cdot \mathbf{C}\mathbf{V}) \quad (2.11)$$

και συνδυάζοντας τις (2.8), (2.10), (2.11) παίρνουμε την εξίσωση

$$\lambda \sum_{i=1}^M \alpha_i (\Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i)) = \frac{1}{M} \sum_{i=1}^M \alpha_i (\Phi(\mathbf{x}_k) \cdot \sum_{j=1}^M \Phi(\mathbf{x}_j)) (\Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i)), \quad (2.12)$$

για $k=1, \dots, M$.

Ορίζουμε ως πίνακα πυρήνα (kernel matrix) τον πίνακα \mathbf{K} ο οποίος παράγεται από τα εσωτερικά γινόμενα των $\Phi(\mathbf{x}_i)$, $\Phi(\mathbf{x}_j)$ και είναι συμμετρικός με διαστάσεις $M \times M$

$$\mathbf{K}_{ij} = (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) \quad (2.13)$$

επομένως η (2.12) γίνεται

$$\mathbf{M}\lambda\mathbf{K}\mathbf{a} = \mathbf{K}^2\mathbf{a} \quad (2.14)$$

όπου το \mathbf{a} αναπαριστά τα διανύσματα $\alpha_1, \dots, \alpha_M$. Η (2.14) λοιπόν καταλήγει στη τελική της μορφή

$$\mathbf{M}\lambda\mathbf{a} = \mathbf{K}\mathbf{a} \quad (2.15)$$

Από την (2.15) παρατηρούμε ότι η τιμή $M\lambda$ αναπαριστά τις ιδιοτιμές του \mathbf{K} ενώ η \mathbf{a} τα ιδιοδιανύσματα του \mathbf{K} . Διαγωνοποιώντας λοιπόν τον \mathbf{K} βρίσκουμε τις ιδιοτιμές $M\lambda$ του \mathbf{K} , βρίσκουμε τις ιδιοτιμές λ του \mathbf{C} , τις ταξινομούμε σε $\lambda_1 \geq \dots \geq \lambda_p \geq \dots \geq \lambda_M$ και τέλος βρίσκουμε τα αντίστοιχα ιδιοδιανύσματα $\mathbf{a}_1, \dots, \mathbf{a}_p, \dots, \mathbf{a}_M$ του \mathbf{K} , με p τον αριθμό των κύριων συνιστωσών πυρήνα που θα χρησιμοποιήσουμε. Στη συνέχεια κανονικοποιούμε τα ιδιοδιανύσματα \mathbf{V} του \mathbf{C} . Συνδυάζοντας τις σχέσεις (2.10), (2.15) για $k=1, \dots, p$ παίρνουμε

$$\begin{aligned} 1 &= (\mathbf{V}^k \cdot \mathbf{V}^k) \\ &= \sum_{i=1}^M \alpha_i^k \Phi(\mathbf{x}_i) \cdot \sum_{j=1}^M \alpha_j^k \Phi(\mathbf{x}_j) \\ &= \left(\sum_{i,j=1}^M \alpha_i^k \alpha_j^k (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) \right) \\ &= \sum_{i,j=1}^M \alpha_i^k \alpha_j^k \mathbf{K}_{ij} \\ &= (\mathbf{a}^k \cdot \mathbf{K}\mathbf{a}^k) \\ &= \lambda_k (\mathbf{a}^k \cdot \mathbf{a}^k) . \end{aligned} \quad (2.16)$$

Τέλος για την εξαγωγή των κύριων συνιστωσών υπολογίζουμε τις προβολές των ιδιοδιανυσμάτων \mathbf{V}^k στον χώρο F . Αν πάρουμε ως εικόνα δοκιμής μία εικόνα x τότε το διάνυσμά της στον F θα είναι το $\Phi(x)$ και οι κύριες συνιστώσες πυρήνα θα

υπολογιστούν από τη σχέση

$$y^k = (\mathbf{V}^k \cdot \Phi(\mathbf{x})) = \sum_{i=1}^M \alpha_i^k (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})) \tag{2.17}$$

ή ισοδύναμα

$$y^k = (\mathbf{V}^k \cdot \Phi(\mathbf{x})) = \sum_{i=1}^M \alpha_i^k \mathbf{K}^{\text{test}}(\Phi(\mathbf{x}_i) , \Phi(\mathbf{x})) . \tag{2.18}$$

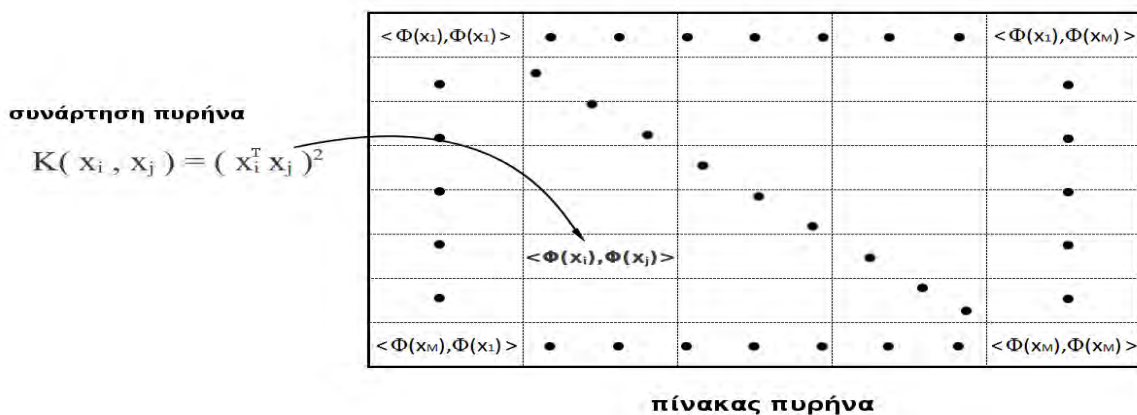
2.2.2 Το τέχνασμα του πυρήνα

Στη Kernel PCA θα υπέθετε κανείς ότι, αφού τα δεδομένα μας μεταφέρονται στον F χώρο, πρέπει να κάνουμε τη πλήρη χαρτογράφηση από τον I στον F . Ο χώρος F όμως έχει πολύ υψηλές πιθανόν και άπειρες διαστάσεις γεγονός που σημαίνει ότι αυξάνεται ραγδαία το υπολογιστικό κόστος. Αν κοιτάξουμε πιο προσεκτικά τις εξισώσεις παρατηρούμε ότι αυτό που χρειαζόμαστε ουσιαστικά για να εφαρμόσουμε τη Kernel PCA είναι τα εσωτερικά γινόμενα των διανυσμάτων. Αν λοιπόν υπήρχε ένας τρόπος για να υπολογίσουμε τα εσωτερικά γινόμενα στον F γνωρίζοντας μόνο τις θέσεις των διανυσμάτων στον I και όχι στον F , θα μπορούσαμε να εφαρμόσουμε την KPCA χωρίς να γνωρίζουμε τη συνάρτηση χαρτογράφησης Φ . Η ιδέα αυτή είναι εφικτή και μπορεί να πραγματοποιηθεί με τη βοήθεια του τεχνάσματος του πυρήνα (kernel trick).

Η βασική εξίσωση στην οποία στηρίζεται το τέχνασμα του πυρήνα είναι η

$$K_{ij} = k(x_i , x_j)$$

Ο πίνακας K_{ij} λέγεται πίνακας πυρήνα (kernel matrix) και κάθε στοιχείο ij αποτελεί το εσωτερικό γινόμενο των διανυσμάτων x_i και x_j στον F , δηλαδή $\langle \Phi(x_i), \Phi(x_j) \rangle$. Για να υπολογίσουμε τον K παίρνουμε ως x_i και x_j όλες τις εικόνες του συνόλου δεδομένων οπότε ο K που προκύπτει είναι συμμετρικός με διαστάσεις το πλήθος M των εικόνων.



Εικόνα 2.4: Κατασκευή πίνακα πυρήνα (kernel matrix) με χρήση της πολυωνυμικής συνάρτησης δευτέρου βαθμού.

Ο πίνακας πυρήνα που θα προκύψει εξαρτάται από τη συνάρτηση πυρήνα (kernel function) $k(x_i, x_j)$ που θα χρησιμοποιήσουμε. Οι πιο γνωστές συναρτήσεις πυρήνα είναι οι εξής:

- 1) πολυωνυμικοί πυρήνες (polynomial kernel functions)

$$k(x_i, x_j) = ((x_i \cdot x_j) + d)^p$$

όπου d σταθερά και p ο βαθμός του πολυωνύμου. Ο πιο απλός μη γραμμικός πολυωνυμικός πυρήνας είναι ο $k(x_i, x_j) = ((x_i \cdot x_j))^2$. Αν ορίσουμε $p=1$ τότε πραγματοποιούμε γραμμική PCA.

- 2) γκαουσιανός πυρήνας (gaussian kernel function)

$$k(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right)$$

Ο συντελεστής σ εξαρτάται από τη διασπορά των διανυσμάτων και συνήθως ισούται με τη διάμεσο των διανυσμάτων $|x_i - x_j|$. Ο γκαουσιανός πυρήνας είναι ο πιο γνωστός από την κατηγορία των συναρτήσεων ακτινωτής βάσης (radial basis functions).

- 3) σιγμοειδείς πυρήνες (hyperbolic tangent (sigmoid) kernel functions)

$$k(x_i, x_j) = \tanh(\alpha(x_i \cdot x_j) + c)$$

όπου c σταθερά. Είναι επίσης γνωστοί ως multilayer perceptron kernels (MLP) και χρησιμοποιούνται κυρίως στο πεδίο των Νευρωνικών Δικτύων.

Μία συνάρτηση για να μπορεί να χρησιμοποιηθεί ως συνάρτηση πυρήνα θα πρέπει να ικανοποιεί το θεώρημα του Mercer. Από το θεώρημα αυτό προκύπτει ότι ο πίνακας πυρήνα θα πρέπει να είναι συμμετρικός και θετικά ημιορισμένος.

2.2.3 Κεντράρισμα πίνακα πυρήνα

Όταν υπολογίσαμε τον kernel matrix κάναμε την υπόθεση ότι ο K που υπολογίστηκε ήταν κεντραρισμένος ως προς τα διανύσματα των εικόνων στον χώρο χαρακτηριστικών, δηλαδή $\sum_{i=1}^M \bar{\Phi}(x_i) = 0$, και βάσει του K υπολογίσαμε τις κύριες συνιστώσες πυρήνα. Για να κεντραριστούν όμως τα δεδομένα του πίνακα πυρήνα ο K πρέπει να περάσει από μία προεπεξεργασία ώστε να προκύψει ο κεντραρισμένος πίνακας που στη παράγραφο αυτή θα συμβολίσουμε με \bar{K} . Ο πίνακας πυρήνα αντιστοιχεί στον F χώρο οπότε δεν μπορούμε να βασιστούμε στη τεχνική που χρησιμοποιήσαμε στη PCA για το κεντράρισμα. Αντιθέτως αντικαθιστούμε τον πίνακα K με τον πίνακα \bar{K} μέσω της εξίσωσης

$$\bar{K} = K - 1_M K - K 1_M + 1_M K 1_M, \quad (2.19)$$

όπου 1_M ένας πίνακας $M \times M$ στον οποίο κάθε στοιχείο ισούται με $\frac{1}{M}$ και M το πλήθος των εικόνων.

Στον F χώρο έχουμε M διανύσματα εικόνων $\Phi(x_i)$. Συμβολίζουμε με $\bar{\Phi}(x_i)$ τα αντίστοιχα κεντραρισμένα διανύσματα οπότε ισχύει ότι

$$\bar{\Phi}(x_i) = \Phi(x_i) - \frac{1}{M} \sum_{i=1}^M \Phi(x_i) \quad (2.20)$$

$$\sum_{i=1}^M \bar{\Phi}(x_i) = 0 \quad (2.21)$$

Ο κεντραρισμένος πίνακας πυρήνα δίνεται από τη σχέση

$$(\bar{K})_{ij} = (\bar{\Phi}(x_i) \cdot \bar{\Phi}(x_j)) \quad (2.22)$$

Αντικαθιστώντας στην (2.22) με την (2.20) παίρνουμε

$$\begin{aligned} (\bar{K})_{ij} &= \left(\left(\Phi(x_i) - \frac{1}{M} \sum_{m=1}^M \Phi(x_m) \right) \cdot \left(\Phi(x_j) - \frac{1}{M} \sum_{n=1}^M \Phi(x_n) \right) \right) \\ &= \left(\left(\Phi(x_i) - \frac{1}{M} \sum_{m=1}^M \Phi(x_m) \right)^T \left(\Phi(x_j) - \frac{1}{M} \sum_{n=1}^M \Phi(x_n) \right) \right) \\ &= \Phi(x_i)^T \Phi(x_j) - \frac{1}{M} \sum_{m=1}^M \Phi(x_m)^T \Phi(x_j) - \frac{1}{M} \sum_{n=1}^M \Phi(x_i)^T \Phi(x_n) + \frac{1}{M} \frac{1}{M} \sum_{m,n=1}^M \Phi(x_m)^T \Phi(x_n) \\ &= K_{ij} - \frac{1}{M} \sum_{m=1}^M 1_{im} K_{mj} - \frac{1}{M} \sum_{n=1}^M 1_{in} K_{nj} + \frac{1}{M} \frac{1}{M} \sum_{m,n=1}^M 1_{im} K_{mn} 1_{nj}. \end{aligned}$$

Χρησιμοποιώντας τον πίνακα 1_M που περιγράψαμε παραπάνω, φθάνουμε στη τελική μορφή της εξίσωσης

$$\bar{K} = K - 1_M K - K 1_M + 1_M K 1_M \quad (2.23)$$

Ακόμη, κατά την εξαγωγή των κύριων συνιστωσών πυρήνα χρησιμοποιήσαμε τον πίνακα πυρήνα K^{test} . Ο πίνακας αυτός θα χρειαστεί επίσης να κεντραριστεί. Αν $t = t_1, \dots, t_L$ το σύνολο των test images τότε σύμφωνα με την (2.18) η εξίσωση υπολογισμού των κύριων συνιστωσών πυρήνα είναι

$$(\mathbf{V}^k \cdot \Phi(t)) = \sum_{i=1}^M \alpha_i^k \bar{K}^{\text{test}}(\Phi(x_i), \Phi(t)).$$

Ισχύει ότι

$$(K^{\text{test}})_{il} = (\Phi(x_i) \cdot \Phi(t_l))$$

άρα

$$\begin{aligned} (\bar{K}^{\text{test}})_{il} &= (\bar{\Phi}(x_i) \cdot \bar{\Phi}(t_l)) \\ &= \left(\left(\Phi(x_i) - \frac{1}{M} \sum_{m=1}^M \Phi(x_m) \right) \cdot \left(\Phi(t_l) - \frac{1}{M} \sum_{n=1}^M \Phi(t_n) \right) \right) \end{aligned}$$

και ακολουθώντας το παραπάνω σκεπτικό καταλήγουμε στην εξίσωση

$$\bar{K}^{\text{test}} = K^{\text{test}} - 1'_M K - K^{\text{test}} 1_M + 1'_M K 1_M, \quad (2.24)$$

όπου $1'_M$ είναι ο πίνακας $L \times M$ στον οποίο κάθε στοιχείο ισούται με $\frac{1}{M}$ και L το πλήθος των test images.

2.2.4 Ανακατασκευή προ-εικόνας

Μέχρι στιγμής έχουν βρεθεί οι κύριες συνιστώσες πυρήνα οι οποίες περιγράφουν τα βασικά χαρακτηριστικά των εικόνων του συνόλου δεδομένων. Ωστόσο οι συνιστώσες αυτές βρίσκονται στον υψηλών διαστάσεων χώρο χαρακτηριστικών (Feature space) με αποτέλεσμα να μην είναι δυνατή η πραγματική αναπαράσταση των κύριων συνιστωσών πυρήνα ως εικόνες καθώς αυτό απαιτεί τη μεταφορά τους στον χαμηλότερων διαστάσεων χώρο εισόδου (Input space) [15]. Η διαδικασία αυτή ονομάζεται ανακατασκευή της προ-εικόνας.

Όπως αναφέρθηκε παραπάνω, στη μέθοδο της Kernel PCA το σύνολο των εικόνων μεταφέρεται από τον χώρο εισόδου, τον χώρο όπου βρίσκονται αρχικά οι εικόνες, στον χώρο των χαρακτηριστικών ο οποίος είναι πολύ υψηλότερων διαστάσεων και εκεί οι εικόνες υποβάλλονται στην PCA με αποτέλεσμα την εξαγωγή των ιδιοδιανυσμάτων που λειτουργούν ως κύριες συνιστώσες. Οι κύριες συνιστώσες πυρήνα που υπολογίζονται στη συνέχεια είναι οι προβολές των κύριων συνιστωσών σε έναν υποχώρο του χώρου χαρακτηριστικών. Για την ανακατασκευή της προ-εικόνας θα πρέπει να ακολουθηθεί η αντίστροφη διαδικασία, δηλαδή να μεταφερθούν οι κύριες συνιστώσες πυρήνα από τον χώρο χαρακτηριστικών πίσω στον χώρο εισόδου. Σε αυτή τη προσπάθεια όμως υπάρχουν δύο δυσκολίες οι οποίες συνθέτουν το αποκαλούμενο 'πρόβλημα της προ-εικόνας'. Πρώτον, η συνάρτηση χαρτογράφησης που χρησιμοποιήθηκε αρχικά είναι άγνωστη καθώς βασιζόμαστε στο τέχνασμα του πυρήνα για το σκοπό της μεταφοράς. Δεύτερον, ο χώρος χαρακτηριστικών έχει πολύ υψηλότερες διαστάσεις από τον χώρο εισόδου, πιθανόν μάλιστα να είναι άπειρων διαστάσεων. Αυτό σημαίνει ότι στο ίδιο στοιχείο από το χώρο εισόδου μπορούν να αντιστοιχούν δύο ή περισσότερα στοιχεία από τον χώρο χαρακτηριστικών. Μερικοί αλγόριθμοι που χειρίζονται το πρόβλημα της προ-εικόνας φθάνοντας στην επιτυχή ανακατασκευή παρουσιάζονται στα [16-17]. Για τους σκοπούς της εργασίας χρησιμοποιήθηκε η μέθοδος του Mika et al.

Υπολογισμός προ-εικόνας

Ορίζουμε το τελεστή προβολής P_M ο οποίος προβάλλει την εικόνα $\Phi(x)$ μέσω της σχέσης

$$P_M \Phi(x) = \sum_{i=1}^M y_i(x) v_i \quad (2.25)$$

όπου y_i η κύρια συνιστώσα πυρήνα και v_i το ιδιοδιάνυσμα του πίνακα συνδιακύμανσης στο χώρο χαρακτηριστικών. Η εύρεση επακριβώς του $\Phi(x)$ είναι ιδιαίτερα δύσκολη, σχεδόν αδύνατη, οπότε υπολογίζεται κατά προσέγγιση ένα z τέτοιο ώστε

$$\Phi(z) \approx P_M \Phi(x). \quad (2.26)$$

Η ακρίβεια του υπολογισμού μπορεί να βελτιωθεί μέσω της ελαχιστοποίησης του (2.27)

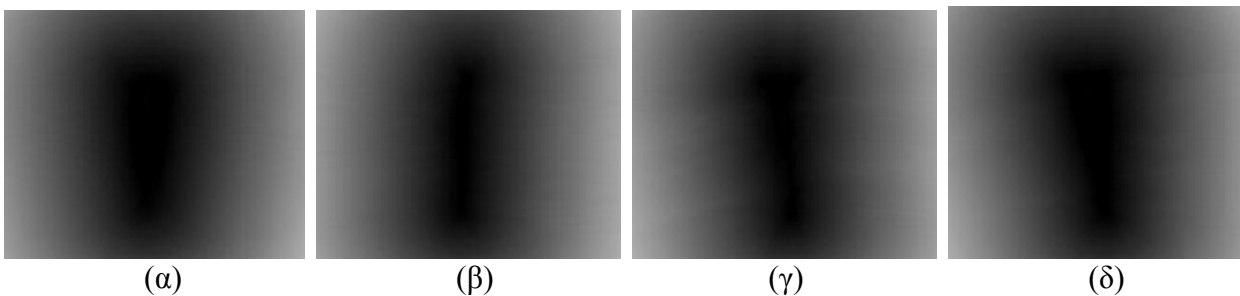
$$\rho(z) = |\Phi(z) - P_M \Phi(x)|^2. \quad (2.27)$$

Εάν η Kernel PCA εφαρμοστεί χρησιμοποιώντας ένα γκαουσιανό πυρήνα τότε σύμφωνα με το [16] η προ-εικόνα μπορεί να βρεθεί με την εξίσωση

$$z_{t+1} = \frac{\sum_{i=1}^M \gamma_i \exp\left(\frac{|z_t - x_i|^2}{2\sigma^2}\right) x_i}{\sum_{i=1}^M \gamma_i \exp\left(\frac{|z_t - x_i|^2}{2\sigma^2}\right)} \quad (2.28)$$

όπου $\gamma_i = \sum_{k=1}^M y_k \mathbf{a}_{ik}$ και t η επανάληψη.

Η προ-εικόνα z υπολογίζεται επαναληπτικά έως ότου η διαφορά $|z_{t+1} - z_t|$ φθάσει σε ένα ικανοποιητικό σημείο. Ως αρχική προ-εικόνα z_0 θέτουμε τη μέση εικόνα του συνόλου των εικόνων στον χώρο I .



Εικόνα 2.5: Τέσσερις εικόνες που συνθέτουν ένα στατιστικό μοντέλο για τη γλωττίδα με τη τεχνική της Ανάλυσης Κύριων Συνιστωσών Πυρήνα. Κατά σειρά εμφανίζονται η μέση εικόνα και η πρώτη, δεύτερη και τρίτη κύρια συνιστώσα πυρήνα. Οι εικόνες των κύριων συνιστωσών πυρήνα προκύπτουν μετά από την ανακατασκευή των προ-εικόνων τους.

Μέθοδοι Τμηματοποίησης

Τα ενεργά περιγράμματα και τα γεωδαιτικά ενεργά περιγράμματα αποτελούν προηγμένες τεχνικές της όρασης υπολογιστών τα οποία μας δίνουν τη δυνατότητα να αναγνωρίζουμε αντικείμενα μέσα σε εικόνες. Οι μέθοδοι συνήθως χρησιμοποιούνται από εφαρμογές με σκοπό την κατάτμηση εικόνας και την παρακολούθηση ενός κινούμενου αντικειμένου. Η ίδια τεχνική μπορεί να αξιοποιηθεί στον τρισδιάστατο χώρο με τη μέθοδο των ενεργών επιφανειών. Επίσης μπορεί να συνδυαστεί με τα προγενέστερα σχήματα τα οποία παρέχουν μία πρότερη πληροφορία όσον αφορά τη μορφή του αντικειμένου που μας ενδιαφέρει. Στη συνέχεια του κεφαλαίου αναλύονται τα μοντέλα των ενεργών περιγραμμάτων και των γεωδαιτικών ενεργών περιγραμμάτων καθώς επίσης και ο τρόπος που συνδυάζονται με τα προγενέστερα σχήματα για να επιτελέσουν τη διεργασία της τμηματοποίησης.

3.1 Ενεργά περιγράμματα

Το μοντέλο των ενεργών περιγραμμάτων ή αλλιώς ‘φιδιών’ (snakes) παρουσιάστηκε από τους Kass et al. [7] και στηρίζεται στη θεωρία των παραμορφώσιμων μοντέλων (deformable models). Η βασική ιδέα της μεθόδου είναι πως αρχικά ορίζεται μία κλειστή καμπύλη κοντά στα όρια του αντικειμένου ενδιαφέροντος και στη συνέχεια επιδρούν πάνω σε αυτήν δυνάμεις που προέρχονται από τα χαρακτηριστικά της καμπύλης και τα στοιχεία της εικόνας. Οι δυνάμεις εξελίσσουν τη καμπύλη με στόχο την ελαχιστοποίηση ενός ενεργειακού συναρτησιακού με αποτέλεσμα στο τέλος της διαδικασίας το περίγραμμα της καμπύλης να συμπίπτει με τα όρια του αντικειμένου. Βελτιωμένες εκδοχές του κλασικού φιδιού είναι το φίδι μπαλονιού (balloon snake) [18], το GVF snake [19], το T-snake [20] και το WVF snake [21].

3.1.1 Μαθηματικό μοντέλο ‘φιδιού’

Στη παράγραφο αυτή εξετάζουμε το φίδι του Kass et al. [7]. Αρχικά ορίζουμε στο επίπεδο την καμπύλη $C(q) = (x(q), y(q)) : [0,1] \rightarrow \mathbb{R}^2$ με παράμετρο q και την εικόνα $I : [0,a] \times [0,b] \rightarrow \mathbb{R}^+$ στην οποία βρίσκεται το αντικείμενο. Θεωρούμε ότι η

καμπύλη έχει μία ενέργεια η οποία δίνεται από το συναρτησιακό

$$\begin{aligned} E_{\text{snake}}^* &= \int_0^1 E_{\text{snake}}(C(q)) dq \\ &= \int_0^1 (E_{\text{int}}(C(q)) + E_{\text{ext}}(C(q))) dq \\ &= \int_0^1 (E_{\text{int}}(C(q)) + E_{\text{image}}(C(q)) + E_{\text{con}}(C(q))) dq . \end{aligned} \quad (3.1)$$

Η συνολική ενέργεια του φιδιού συντίθεται από δύο ειδών ενέργειες, την εσωτερική ενέργεια (E_{int}) και την εξωτερική ενέργεια (E_{ext}) που αποτελείται από την ενέργεια της εικόνας (E_{image}) και την ενέργεια που επιβάλλουν οι εξωτερικοί περιορισμοί (E_{con}).

Η εσωτερική ενέργεια δίνεται από τη σχέση

$$E_{\text{int}} = \frac{1}{2} (\alpha |C'(q)|^2 + \beta |C''(q)|^2) \quad (3.2)$$

όπου τα α και β παίρνουν συνήθως θετικές τιμές. Η εσωτερική ενέργεια τείνει να εξομαλύνει το περίγραμμα της καμπύλης. Ο συντελεστής α ελέγχει την ελαστικότητα της καμπύλης ενώ ο συντελεστής β την ακαμψία.

Η ενέργεια της εικόνας εξαρτάται από τα στοιχεία της εικόνας και έλκει την καμπύλη προς τα προεξέχοντα (salient) χαρακτηριστικά αυτής. Το πιο απλό χαρακτηριστικό στο οποίο μπορούμε να βασιστούμε είναι η ένταση των εικονοστοιχείων. Με την συνάρτηση $E_{\text{line}} = I(x,y)$ το φίδι κινείται προς τις πιο σκούρες ή τις πιο φωτεινές γραμμές τις εικόνας. Ένα άλλο χαρακτηριστικό που μπορούμε να χρησιμοποιήσουμε είναι οι ακμές τις εικόνας. Με τη συνάρτηση $E_{\text{edge}} = -|\nabla I(x,y)|^2$ το φίδι κινείται προς τις περιοχές της εικόνας που εμφανίζουν μεγαλύτερη κλίση. Για να μειώσουμε το θόρυβο και να αυξήσουμε την εμβέλεια του φιδιού μπορούμε να χρησιμοποιήσουμε τη συνάρτηση $E_{\text{edge}2} = -|\nabla(G_\sigma(x,y) * I(x,y))|^2$ όπου $G_\sigma(x,y)$ μία δισδιάστατη γκαουσιανή συνάρτηση με τυπική απόκλιση σ .

Ο όρος E_{con} αναφέρεται στην ενέργεια που προκύπτει από τους εξωτερικούς περιορισμούς που επιβάλλει ο χρήστης. Χαρακτηριστικό παράδειγμα όπου γίνεται χρήση αυτής της ενέργειας είναι το φίδι μπαλονιού του Cohen.

Συνοψίζοντας λοιπόν τα παραπάνω και μη λαμβάνοντας υπ'όψιν εξωτερικούς περιορισμούς, το ενεργειακό συναρτησιακό της καμπύλης περιγράφεται από τη σχέση

$$E(C(q)) = \frac{1}{2}\alpha \int_0^1 |C'(q)|^2 dq + \frac{1}{2}\beta \int_0^1 |C''(q)|^2 dq - \lambda \int_0^1 |\nabla I(C(q))|^2 dq \quad (3.3)$$

Απώτερος στόχος του φιδιού είναι η εύρεση του περιγράμματος του αντικειμένου μέσω της ελαχιστοποίησης της ενέργειάς του (E_{snake}^*). Επομένως εφαρμόζοντας την εξίσωση Euler στο ενεργειακό συναρτησιακό προκύπτει η σχέση

$$\alpha C''(q) - \beta C''''(q) - \nabla E_{\text{ext}} = 0 \quad . \quad (3.4)$$

Τη (3.4) μπορούμε να τη δούμε ως μία εξίσωση εξισορρόπησης δυνάμεων οπότε παίρνει τη μορφή

$$F_{\text{int}} + F_{\text{ext}} = 0. \quad (3.5)$$

με την $F_{\text{int}} = \alpha C''(q) - \beta C'''(q)$ να δρα ως εσωτερική δύναμη και την $F_{\text{ext}} = -\nabla E_{\text{ext}}$ να δρα ως εξωτερική. Αυτές είναι οι δύο δυνάμεις που επιδρούν στο φίδι και ορίζουν το περίγραμμα του. Επιπλέον εισάγουμε στην εξίσωση τον παράγοντα χρόνο t οπότε κάθε χρονική στιγμή το φίδι κινείται ελαττώνοντας το ενεργειακό συναρτησιακό του έως ότου φθάσει στο σημείο να σταθεροποιηθεί σε ένα τοπικό ελάχιστο. Συνεπώς η εξίσωση της καμπύλης του φιδιού στη τελική της μορφή είναι η

$$C(q,t) = \alpha C''(q,t) - \beta C'''(q,t) - \nabla E_{\text{ext}}. \quad (3.6)$$

Το κλασικό φίδι όπως και το balloon snake λειτουργούν ικανοποιητικά αλλά παρουσιάζουν ορισμένα μειονεκτήματα. Είναι πολύ ευαίσθητα στην παραμετροποίηση, έχουν μικρή εμβέλεια (capture range) άρα η αρχική καμπύλη πρέπει να τοποθετηθεί κοντά στο αντικείμενο και επίσης αδυνατούν να τμηματοποιήσουν περιοχές με μεγάλες εσοχές. Τα δύο τελευταία προβλήματα αντιμετωπίζονται στο GVF-snake.

3.1.2 Φίδι ροής κλιμακωτού διανύσματος

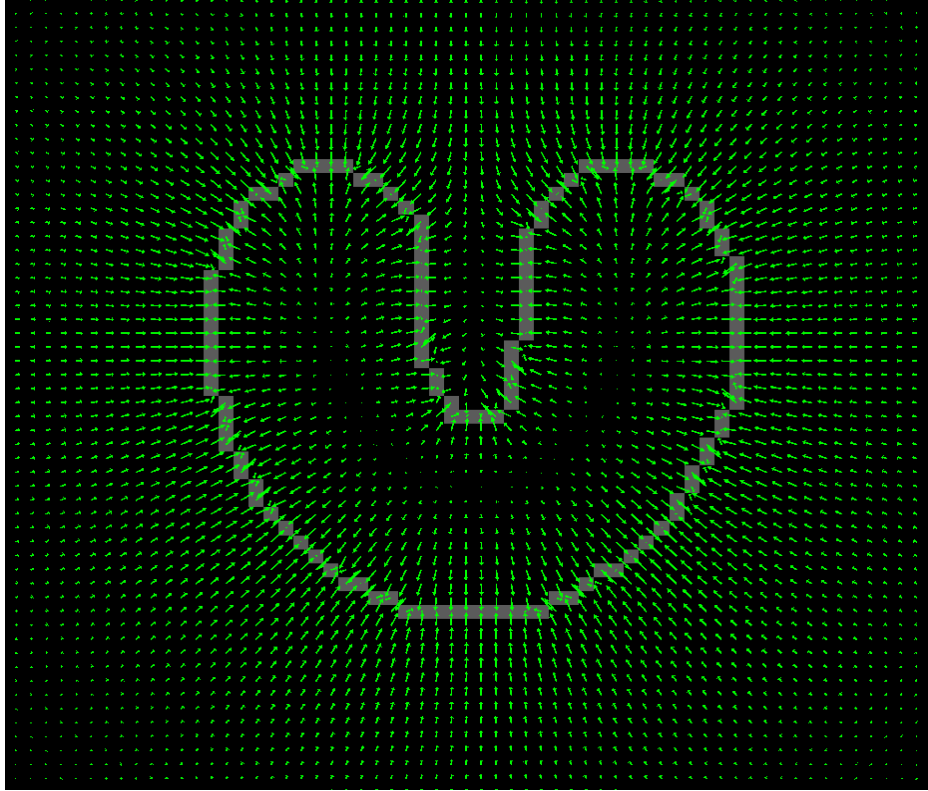
Το φίδι Ροής Κλιμακωτού Διανύσματος (Gradient Vector Flow snake) ή GVF-snake έχει προταθεί από τους Xu και Prince [19]. Το χαρακτηριστικό που κάνει αυτό το φίδι να ξεχωρίζει είναι το πεδίο Ροής Κλιμακωτού Διανύσματος (Gradient Vector Flow field). Το GVF-πεδίο ορίζεται σε ολόκληρη την εικόνα και δημιουργεί διανύσματα που κατευθύνονται ομοιόμορφα προς όλες τις ακμές. Αυτό προσδίδει στο φίδι δύο πολύ σημαντικά πλεονεκτήματα. Το πρώτο είναι ότι μεγιστοποιεί την εμβέλεια του φιδιού με αποτέλεσμα όσο μακριά και αν τοποθετηθεί η αρχική καμπύλη από τα όρια του αντικειμένου, το φίδι με τη βοήθεια του πεδίου να οδηγείται πάντα προς το αντικείμενο. Το δεύτερο πλεονέκτημα είναι ότι το φίδι έχει τη δυνατότητα να συνεχίσει να επεκτείνεται ακόμα και σε δυσπρόσιτες περιοχές του αντικειμένου.

Το GVF-πεδίο ορίζεται από το διανυσματικό πεδίο $g(x,y) = (u(x,y), v(x,y))$ και επιχειρεί να ελαχιστοποιήσει το ενεργειακό συναρτησιακό

$$E = \iint \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |g - \nabla f|^2 dx dy \quad (3.7)$$

όπου f η εικόνα. Από τη (3.7) παρατηρεί κανείς ότι όταν το ∇f είναι μικρό τότε υπερισχύουν τα μερικά παράγωγα του διανυσματικού πεδίου. Αυτό συμβαίνει στις περιοχές της εικόνας που βρίσκονται μακριά από το αντικείμενο και άρα είναι ομοιογενείς. Στη δεύτερη περίπτωση όπου βρισκόμαστε κοντά στο αντικείμενο, το ∇f είναι μεγάλο οπότε το $g \approx \nabla f$ ελαττώνει τη τιμή του δεύτερου όρου. Η μεταβλητή μ

ελέγχει τη σχέση ανάμεσα στο πρώτο και το δεύτερο όρο και η τιμή της ορίζεται ανάλογα με το θόρυβο της εικόνας (όσο μεγαλύτερος ο θόρυβος τόσο μεγαλύτερη η μ).



Εικόνα 3.1: Αναπαράσταση του πεδίου Gradient Vector Flow για ένα αντικείμενο σχήματος U. Οι παραγόμενες ροές κατευθύνουν το φίδι προς το εσωτερικό της εσοχής [57].

Το g μπορεί να βρεθεί από τις εξισώσεις Euler

$$\mu \nabla^2 u - (u - f_x)(f_x^2 + f_y^2) = 0 \quad (3.8\alpha)$$

$$\mu \nabla^2 v - (v - f_y)(f_x^2 + f_y^2) = 0 \quad (3.8\beta)$$

όπου ∇^2 ο λαπλασιανός τελεστής. Για να λύσουμε τις δύο εξισώσεις δουλεύουμε με τα u και v ως συναρτήσεις χρόνου και τις υπολογίζουμε επαναληπτικά καθώς το t τείνει στο άπειρο

$$u_t(x,y,t) = \mu \nabla^2 u(x,y,t) - (u(x,y,t) - f_x(x,y))(f_x^2(x,y) + f_y^2(x,y)) = 0$$

$$v_t(x,y,t) = \mu \nabla^2 v(x,y,t) - (v(x,y,t) - f_y(x,y))(f_x^2(x,y) + f_y^2(x,y)) = 0$$

Στη συνέχεια αντικαθιστούμε στη (3.6) τον όρο $-\nabla E_{\text{ext}}$ με το g οπότε η εξίσωση παίρνει τη μορφή

$$C(q,t) = \alpha C''(q,t) - \beta C''''(q,t) + g. \quad (3.9)$$

Αυτή είναι η εξίσωση της καμπύλης του GVF-snake.

3.2 Γεωδαιτικά ενεργά περιγράμματα

Το μοντέλο των Γεωδαιτικών Ενεργών Περιγραμμάτων (Geodesic Active Contours, GACs) προτάθηκε από τους Caselles, Kimmel και Sapiro [22] και δίνει λύση σε δύο αδυναμίες του μοντέλου των ενεργών περιγραμμάτων. Πιο συγκεκριμένα το μοντέλο απαλλάσσεται από την αυστηρή παραμετροποίηση της καμπύλης και στηρίζεται περισσότερο στα γεωμετρικά χαρακτηριστικά της εικόνας. Επίσης με τη βοήθεια της μεθόδου των επιπεδοσυνόλων (level-sets) [23] αποκτά τη δυνατότητα να διαχειρίζεται αυτόματα τις τοπολογικές αλλαγές με αποτέλεσμα να διευκολύνεται η διαδικασία της τμηματοποίησης.

3.2.1 Υπολογισμός γεωδαιτικού ενεργού περιγράμματος

Σύμφωνα με τη σχέση (3.3) το ενεργειακό συναρτησιακό που χρησιμοποιείται στα ενεργά περιγράμματα είναι το

$$E(C(q)) = \alpha \int_0^1 |C'(q)|^2 dq + \beta \int_0^1 |C''(q)|^2 dq - \lambda \int_0^1 |\nabla I(C(q))|^2 dq$$

το οποίο συνοψίζεται ως ένα άθροισμα εσωτερικών και εξωτερικών ενεργειών

$$E(C(q)) = \int_0^1 (E_{\text{int}}(C(q)) + E_{\text{ext}}(C(q))) dq .$$

Το μοντέλο των GACs θεωρεί ότι η κανονικοποίηση πρώτης τάξης (ο όρος που στηρίζεται στο α) επαρκεί για την κανονικοποίηση της καμπύλης, οπότε θέτοντας το $\beta=0$ παίρνουμε την εξίσωση

$$E(C(q)) = \alpha \int_0^1 |C'(q)|^2 dq - \lambda \int_0^1 |\nabla I(C(q))|^2 dq . \quad (3.10)$$

Επίσης ως ανιχνευτή ακμών (edge detector) ορίζουμε μία γνησίως φθίνουσα συνάρτηση $g:[0,+\infty) \rightarrow \mathbb{R}^+$ όπου $\lim_{r \rightarrow \infty} g(r) = 0$. Παραδείγματα συναρτήσεων που μπορούν να χρησιμοποιηθούν για την g είναι οι

$$g = g(|\nabla I|) = \frac{1}{1 + (|\nabla G_{\sigma} * I|)^n}$$

$$g = g(|\nabla I|) = \exp\left(- \frac{|\nabla G_{\sigma} * I|}{s} \right)$$

Επομένως μπορούμε να αντικαταστήσουμε τον όρο $-\lambda \int_0^1 |\nabla I(C(q))|^2 dq$ με τον όρο $\lambda \int_0^1 g(|\nabla I(C(q))|)^2 dq$ και η (3.10) παίρνει τη μορφή

$$E(C(q)) = \alpha \int_0^1 |C'(q)|^2 dq + \lambda \int_0^1 g(|\nabla I(C(q))|)^2 dq . \quad (3.11)$$

Αποδεικνύεται ότι η ελαχιστοποίηση του συναρτησιακού (3.11) είναι ισοδύναμη με την ελαχιστοποίηση του

$$E(C(q)) = \int_0^1 g(|\nabla I(C(q))|) |C'(q)| dq . \quad (3.12)$$

Θέτοντας $|C'(q)|dq = ds$ με s παράμετρο τόξου παίρνουμε τη συνάρτηση

$$E(C(q)) = \int_0^{L(C)} g(|\nabla I(C(s))|) ds \quad (3.13)$$

όπου $L(C)$ το συνολικό μήκος της καμπύλης. Σχηματίζουμε δηλαδή μία εξίσωση η οποία εξαρτάται αποκλειστικά από τα γεωμετρικά χαρακτηριστικά της καμπύλης.

Η ελαχιστοποίηση του παραπάνω συναρτησιακού επιτυγχάνεται με την εξίσωση Euler-Lagrange και τη μέθοδο κατάβασης κλίσης (gradient descent). Συνεπώς η εξίσωση που περιγράφει την χρονική εξέλιξη της καμπύλης είναι η

$$\frac{\partial C(t)}{\partial t} = g\kappa \vec{N} - (\nabla g \cdot \vec{N}) \vec{N} \quad (3.14)$$

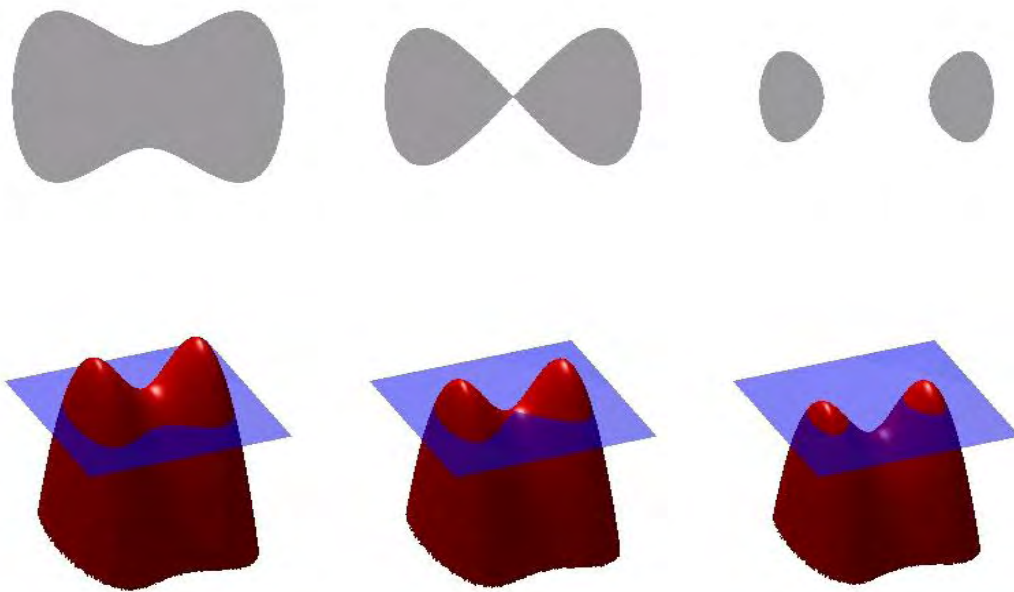
όπου κ η καμπυλότητα της C και \vec{N} το μοναδιαίο κάθετο διάνυσμα προς το εσωτερικό της καμπύλης. Για να αποφευχθεί η σύγκλιση σε τοπικά ελάχιστα χρησιμοποιούμε χώρους πολλαπλής κλίμακας (multiscale) μεταβαίνοντας από μία γενικότερη εικόνα χαμηλότερης ανάλυσης σε μία ειδικότερη εικόνα υψηλότερης ανάλυσης (τεχνική coarse to fine).

3.2.2 Μέθοδος επιπεδοσυνόλων

Το μοντέλο όπως έχει δομηθεί μέχρι στιγμής μπορεί να τμηματοποιήσει τα αντικείμενα μίας εικόνας εντοπίζοντας τις ακμές τους και ορίζοντας την καμπυλότητα της καμπύλης με βάση τα γεωμετρικά χαρακτηριστικά. Ωστόσο φτιάχνει μόνο ένα περίγραμμα που σημαίνει ότι δεν μπορεί να διαχωρίσει τα αντικείμενα ανάλογα με το πλήθος τους. Το πρόβλημα μπορεί να ξεπεραστεί με τη χρήση ενός αλγορίθμου ο οποίος θα έχει τη δυνατότητα να ‘καταλαβαίνει’ αν η καμπύλη περικλείει πολλαπλά αντικείμενα και ανάλογα να προχωράει σε διασπάσεις και επανενώσεις σχηματίζοντας περισσότερες καμπύλες για τις κατατμήσεις των αντικειμένων. Ένα άλλο πρόβλημα που προκύπτει είναι ότι οι κόμβοι της καμπύλης τείνουν να συγκεντρώνεται σε συγκεκριμένες περιοχές της καμπύλης εμποδίζοντάς την να επεκταθεί για να τμηματοποιήσει ολόκληρο το αντικείμενο. Αυτό μπορεί να επιλυθεί κάνοντας ανά τακτά χωρικά διαστήματα δειγματοληψία και επαναπαραμετροποίηση του μοντέλου.

Μία άλλη πολύ πιο αποτελεσματική μέθοδος που μπορούμε να εφαρμόσουμε προκειμένου το μοντέλο να ‘αντιλαμβάνεται’ τις τοπολογικές μεταβολές είναι η μέθοδος των επιπεδοσυνόλων (level sets). Οι βάσεις της θεωρίας των επιπεδοσυνόλων

τέθηκαν από τους Dervieux και Thomasset [27-28], ενώ οι Osher και Sethian στην εργασία [23] παρουσίασαν το πλαίσιο λειτουργίας των επιπεδοσυνόλων και συνέβαλαν καθοριστικά ώστε η μέθοδος να γίνει ευρύτερα γνωστή. Στη τεχνική αυτή η τμηματοποίηση δεν γίνεται μέσω της εξέλιξης της καμπύλης αλλά μέσω της εξέλιξης μίας επιφάνειας u και η καμπύλη που αναπαριστά το περίγραμμα του αντικειμένου προκύπτει από την τομή της επιφάνειας u με την εικόνα. Όπως φαίνεται και στην εικόνα (3.2) με αυτή τη μέθοδο η καμπύλη έχει τη δυνατότητα να διαμορφώνει ελεύθερα το σχήμα της με βάση τις τοπολογικές αλλαγές που παρατηρεί η επιφάνεια u στην εικόνα.



Εικόνα 3.2: Καθώς η επιφάνεια u (μπλε χρώμα) κινείται κατά μήκος των επιπέδων της εικόνας (κόκκινο χρώμα) η καμπύλη διαμορφώνεται αυτόματα και με φυσικό τρόπο. Τα σημεία της εικόνας που βρίσκονται στο ίδιο επίπεδο με την u ορίζουν το περίγραμμα της καμπύλης [58].

Θεωρούμε μία καμπύλη $C(q,t): [0,1] \times [0,+\infty) \rightarrow \mathbb{R}^2$ και μία συνάρτηση $u: [0,a] \times [0,b] \rightarrow \mathbb{R}$ που ορίζεται για ολόκληρη την εικόνα. Η u λειτουργεί όπως η συνάρτηση προσημασμένης απόστασης και κατατάσσει κάθε στοιχείο της εικόνας σύμφωνα με τη συνάρτηση $u(C(t),t)=\lambda$ σε ένα επίπεδο (στάθμη) τιμής λ . Η καμπύλη σχηματίζεται από το σύνολο των σημείων που βρίσκονται στο μηδενικό επίπεδο (zero level set) και ικανοποιούν την σχέση $u(C(t),t)=0$ (επιφάνεια u). Οι περιοχές της εικόνας για τις οποίες ισχύει ότι $u(C(t),t)<0$ θα βρίσκονται εντός της καμπύλης ενώ αν $u(C(t),t)>0$ τότε σημαίνει ότι βρίσκονται εκτός της καμπύλης (ανάλογα με τη σύμβαση που κάνουμε μπορεί να ισχύει το αντίστροφο). Καθώς η επιφάνεια u κινείται από τη μία στάθμη στην άλλη (άρα αλλάζει και το μηδενικό επίπεδο), μεταβάλλεται και η τμηματοποίηση που προκύπτει για την εικόνα. Παρατηρούμε ότι με αυτή τη

μέθοδο η τμηματοποίηση γίνεται με φυσικό τρόπο βάσει της τοπολογίας της εικόνας και χωρίς να χρειάζεται καμία παραμετροποίηση. Ο μόνος περιορισμός που υπάρχει είναι η u να είναι συνεχής κατά Lipschitz.

Αν ενσωματωθεί η θεωρία των επιπεδοσυνόλων στο μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων τότε τα γεωμετρικά στοιχεία της C μπορούν να προσδιοριστούν μέσα από την u , οπότε ισχύει ότι

$$\vec{N} = -\frac{\nabla u}{|\nabla u|} \quad (3.15)$$

$$k = -\nabla \vec{N} = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right). \quad (3.16)$$

Παραγωγίζοντας τη σχέση $u(C(t),t)=0$ ως προς τον χρόνο και χρησιμοποιώντας την (3.16) παίρνουμε

$$\frac{\partial u}{\partial t} + \nabla u C(t) = 0$$

$$\frac{\partial u}{\partial t} = C(t) \vec{N} |\nabla u|$$

και θέτοντας $V=C(t) \vec{N}$ καταλήγουμε στην εξίσωση

$$\frac{\partial u}{\partial t} = V |\nabla u|. \quad (3.17)$$

Η (3.17) είναι η εξίσωση που περιγράφει την εξέλιξη της επιφάνειας u και το V αντιστοιχεί στην ταχύτητα εξέλιξης. Χρησιμοποιώντας τις σχέσεις (3.14),(3.15),(3.16) και προσθέτοντας τη δύναμη μπαλονιού c , καταλήγουμε στη τελική συνάρτηση εξέλιξης για την καμπύλη του GAC

$$\frac{\partial u}{\partial t} = g(c + k)|u| + \nabla u \cdot \nabla g. \quad (3.18)$$

Οι μέθοδοι που χρησιμοποιούνται συνήθως για την εξέλιξη των επιφανειών είναι αυτός της ταχείας προέλασης (fast marching) και αυτός της στενής ζώνης (narrow-band). Η πρώτη μέθοδος είναι ιδιαίτερα γρήγορα χάρη στο γεγονός ότι το μέτωπο της καμπύλης κινείται πάντα προς την ίδια κατεύθυνση (είτε προς στο εσωτερικό είτε προς το εξωτερικό της καμπύλης) με αποτέλεσμα να μειώνεται κατά πολύ η πολυπλοκότητα. Στη μέθοδος στενής ζώνης ο αλγόριθμος καθώς εξελίσσει τη καμπύλη περιορίζεται σε μία λεπτή λωρίδα γύρω από το μέτωπο της καμπύλης για την εύρεση νέων σημείων αντί να ψάχνει σε ολόκληρη την εικόνα, οπότε το υπολογιστικό κόστος ελαττώνεται σημαντικά.

3.3 Τμηματοποίηση με γεωδαιτικά ενεργά περιγράμματα και χρήση προγενέστερου σχήματος

Τα μοντέλα των ενεργών περιγραμμάτων επιχειρούν να τμηματοποιήσουν ένα αντικείμενο στηριζόμενα αποκλειστικά στις πληροφορίες της εικόνας παρέχοντας έτσι μία γενικευμένη μέθοδο. Στη πραγματικότητα όμως, ενδιαφερόμαστε συνήθως για τον εντοπισμό ενός συγκεκριμένου αντικειμένου το σχήμα του οποίου είναι ως επί το πλείστον γνωστό. Εάν λοιπόν μπορούσε να χρησιμοποιηθεί η πληροφορία του εκ των προτέρων γνωστού σχήματος παράλληλα με την τμηματοποίηση, τότε κάτι τέτοιο θα επέτρεπε την ορθότερη καθοδήγηση της εξελισσόμενης καμπύλης και τον ακριβέστερο υπολογισμό του τελικού περιγράμματος. Επίσης μπορεί να βοηθήσει να ξεπεραστούν προβλήματα λόγω θορύβου ή άλλων αιτιών που οφείλονται στη λήψη της εικόνας και να αντιμετωπίσει περιπτώσεις όπου περιοχές του αντικειμένου που μας ενδιαφέρει επικαλύπτονται με περιοχές άλλων αντικειμένων. Εργασίες που έχουν ασχοληθεί με αυτό το θέμα είναι οι εργασίες των Leventon et al. [24], Tsai et al. [25] και Paragios-Rousson [26]. Στην παρούσα εργασία επικεντρωνόμαστε στο μοντέλο του Leventon et al.

Η εισαγωγή της γνώσης γίνεται με τη χρήση του προγενέστερου σχήματος το οποίο παράγεται από το στατιστικό μοντέλο που δίνουν τεχνικές όπως η PCA. Ως προγενέστερο σχήμα ορίζεται αρχικά το μέσο σχήμα του στατιστικού μοντέλου ενώ στη συνέχεια το σχήμα προσαρμόζεται στο αντικείμενο της εικόνας. Για να γίνει αυτό χρησιμοποιούνται δύο είδη παραμέτρων και καθώς εξελίσσεται το περίγραμμα οι παράμετροι επανεκτιμώνται έτσι ώστε η μορφή του προγενέστερου σχήματος να βρίσκεται όσο το δυνατόν πιο κοντά στο αντικείμενο της εικόνας. Για αυτόν το σκοπό χρησιμοποιείται σε κάθε επανάληψη η σχέση

$$\langle \mathbf{a}_{\text{map}}, \mathbf{p}_{\text{map}} \rangle = \operatorname{argmax}_{\mathbf{a}, \mathbf{p}} P(\mathbf{a}, \mathbf{p} \mid \mathbf{u}, \nabla I) \quad (3.19)$$

όπου οι παράμετροι \mathbf{a} είναι υπεύθυνοι για τον υπολογισμό του σχήματος, οι παράμετροι \mathbf{p} είναι υπεύθυνοι για την θέση, και τον προσανατολισμό του σχήματος, το \mathbf{u} αναφέρεται στην εξελισσόμενη καμπύλη και ο τελευταίος όρος στη κλίση της εικόνας. Συμβολίζοντας με \mathbf{u}^* την εκτιμώμενη τελική καμπύλη, τότε στο τέλος κάθε επανάληψης η τελική καμπύλη εκτιμάται από τη σχέση

$$\mathbf{u}_{\text{MAP}}^* = \operatorname{argmax}_{\mathbf{u}^*} P(\mathbf{u}^* \mid \mathbf{u}, \nabla I) \quad (3.20)$$

Μέσω του κανόνα Bayes η (3.19) παίρνει τη μορφή

$$\begin{aligned} P(\mathbf{a}, \mathbf{p} \mid \mathbf{u}, \nabla I) &= \frac{P(\mathbf{u}, \nabla I \mid \mathbf{a}, \mathbf{p}) P(\mathbf{a}, \mathbf{p})}{P(\mathbf{u}, \nabla I)} \\ &= \frac{P(\mathbf{u} \mid \mathbf{a}, \mathbf{p}) P(\nabla I \mid \mathbf{a}, \mathbf{p}, \mathbf{u}) P(\mathbf{a}) P(\mathbf{p})}{P(\mathbf{u}, \nabla I)} \end{aligned} \quad (3.21)$$

Ο όρος $P(u | a, p)$, που αποκαλείται επίσης εσωτερικός όρος, υπολογίζει τη πιθανότητα εμφάνισης μίας καμπύλης u λαμβάνοντας υπ' όψιν τις παραμέτρους a, p της τελικής καμπύλης u^* και ο συγκεκριμένος όρος δεν αξιοποιεί κάποια πληροφορία από την εικόνα. Η λογική που ακολουθείται είναι ότι αν η καμπύλη αρχικοποιηθεί εντός του αντικειμένου στην εικόνα, τότε κατά την εξέλιξη της καμπύλης είναι πιθανότερο να παραμείνει εντός του αντικειμένου παρά να προκύψει μία καμπύλη που θα βρίσκεται ολόκληρη ή τμήμα αυτής εκτός του αντικειμένου. Συνεπώς ο όρος μοντελοποιείται ως μία λαπλασιανή συνάρτηση πυκνότητας πιθανότητας η οποία στηρίζεται στον όγκο της u που βρίσκεται εκτός της u^*

$$P(u | a, p) = \exp(-V_{\text{outside}}) \quad (3.22)$$

Ο εσωτερικός όρος λοιπόν βοηθάει να βρεθούν οι καμπύλες που βρίσκονται εντός του αντικειμένου χωρίς όμως να ευνοεί κάποια συγκεκριμένη από αυτές.

Ο όρος κλίσης $P(|\nabla I| | a, p, u)$, υπολογίζει τη πιθανότητα να βρεθούν κλίσεις της εικόνας με βάση τη τελική καμπύλη και τη νέα καμπύλη κάθε επανάληψης. Παρατηρώντας τα u^* και $|\nabla I|$ θεωρούμε ότι η μεταξύ τους σχέση είναι γκαουσιανή. Αν υποθέσουμε ότι $h(u^*)$ είναι η συνάρτηση που λειτουργεί καλύτερα για τα δεδομένα $(u^*, |\nabla I|)$, τότε ο όρος κλίσης μπορεί να βρεθεί από την εξίσωση

$$P(|\nabla I| | u^*, u) = \exp(-|h(u^*) - |\nabla I||^2) . \quad (3.23)$$

Αναφορικά με τις δύο τελευταίες παραμέτρους, οι παράμετροι a ακολουθούν γκαουσιανή κατανομή οπότε η πιθανότητα δίνεται από την εξίσωση

$$P(a) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_k|}} \exp(-\frac{1}{2} a^T \Sigma_k^{-1} a) \quad (3.24)$$

ενώ οι παράμετροι p θεωρείται ότι ακολουθούν την ομοιόμορφη κατανομή οπότε για το $P(p)$ ισχύει ότι

$$P(p) = U(-\infty, \infty) . \quad (3.25)$$

Στο νέο μοντέλο η καμπύλη ορίζεται από τον συνδυασμό δύο δυνάμεων, αυτή του GAC που στηρίζεται σε χαρακτηριστικά τοπικού επιπέδου της εικόνας και αυτή του προγενέστερου σχήματος που στηρίζεται σε χαρακτηριστικά ενός ευρύτερου επιπέδου της εικόνας. Όπως φαίνεται και στην εικόνα (3.3) σε κάθε επανάληψη η συνιστώσα από το GAC προτείνει την κατεύθυνση προς την οποία πρέπει να κινηθεί το περίγραμμα. Στη συνέχεια ο αλγόριθμος εκτιμά ποιά θα είναι τελική καμπύλη και ανάλογα η δεύτερη συνιστώσα προτείνει το περίγραμμα να κινηθεί προς εκείνη τη κατεύθυνση. Στο τέλος της επανάληψης η συνισταμένη των δύο δυνάμεων ορίζει τη τελική κατεύθυνση προς την οποία θα εξελιχθεί το περίγραμμα. Συνοψίζοντας τα

παραπάνω καταλήγουμε στην εξίσωση υπολογισμού της καμπύλης

$$u(t+1) = u(t) + \lambda_1 (g(c+k)|\nabla u(t)| + \nabla u(t) \nabla g) + \lambda_2 (u^* - u) \quad (3.26)$$

όπου το t συμβολίζει την επανάληψη, το $u(t)$ συμβολίζει την εξελισσόμενη επιφάνεια ενώ ο δεύτερος όρος προκύπτει από το GAC και ο τρίτος όρος από το προγενέστερο σχήμα. Μέσω των συντελεστών λ_1 , λ_2 ο χρήστης καθορίζει το βαθμό συμμετοχής του GAC και του προγενέστερου σχήματος αντίστοιχα.



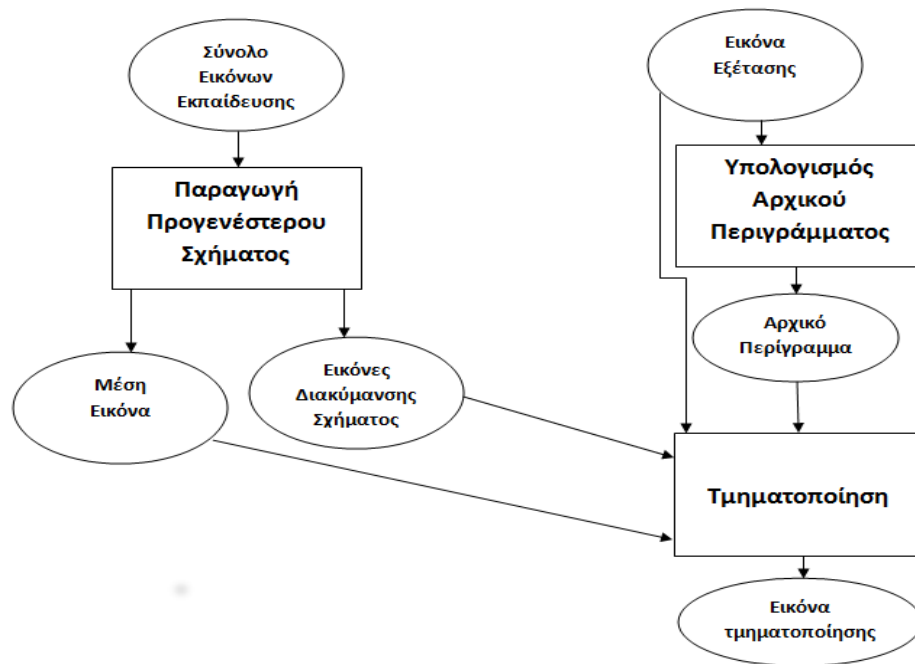
Εικόνα 3.3: Οι δυνάμεις που επιδρούν σε ένα σημείο της καμπύλης και καθορίζουν την εξέλιξή της σε κάθε επανάληψη [24].

Μεθοδολογία

Το κεφάλαιο αυτό εστιάζει στην υλοποίηση των μεθόδων που μελετήθηκαν στη παρούσα πτυχιακή εργασία. Περιγράφεται η διαδικασία της εκπαίδευσης με την οποία δημιουργείται το στατιστικό μοντέλο σχήματος από το οποίο εξάγεται αργότερα το προγενέστερο σχήμα παρέχοντας έτσι μία πρότερη πληροφορία για τη μορφή της γλωττίδας. Ακόμα, αναλύεται η λειτουργία της τμηματοποίησης, παρουσιάζεται η μέθοδος που εντοπίζει τη γλωττίδα μέσα στην εικόνα και διαμορφώνει το αρχικό περίγραμμα όπως επίσης και η μέθοδος που συνδυάζει το μοντέλο GAC με το προγενέστερο σχήμα που οδηγεί στη τελική τμηματοποίηση. Επίσης μελετάμε τη σουίτα λογισμικού Insight Toolkit (ITK) και των αλγορίθμων που παρέχει για την επεξεργασία και κατάτμηση εικόνων. Στο τέλος του κεφαλαίου γίνεται μία ανασκόπηση των εργαλείων που χρησιμοποιούνται για την ανάπτυξη των αλγορίθμων.

4.1 Αλγόριθμοι τμηματοποίησης εικόνων

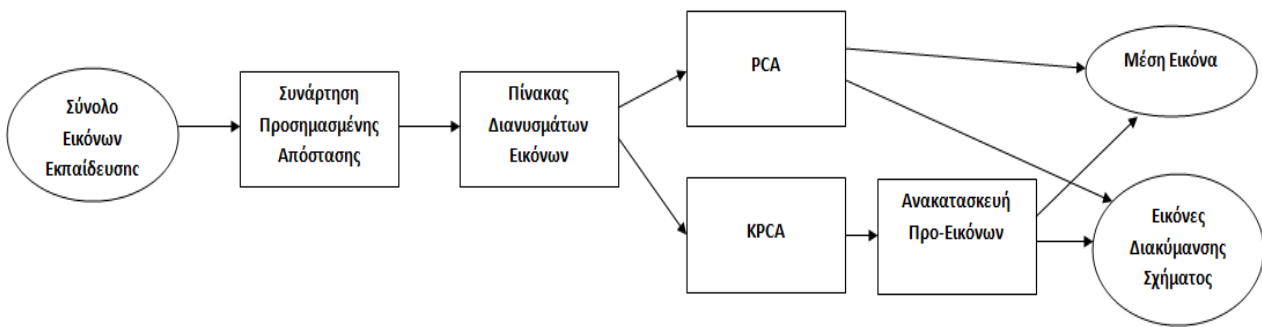
Η εφαρμογή εκτελεί δύο διαδικασίες, την παραγωγή του προγενέστερου σχήματος και την τμηματοποίηση. Στη πρώτη διαδικασία κατασκευάζεται το στατιστικό μοντέλο σχήματος βάσει του οποίου εξάγεται στη συνέχεια το προγενέστερο σχήμα παρέχοντας με αυτόν τον τρόπο μία πρότερη γνώση αναφορικά με το σχήμα της γλωττίδας στη διαδικασία τμηματοποίησης. Η δεύτερη διαδικασία πραγματοποιείται σε δύο στάδια. Στο πρώτο στάδιο εντοπίζεται η τοποθεσία της γλωττίδας μέσα στην εικόνα και υπολογίζεται ένα αρχικό περίγραμμα. Στο δεύτερο στάδιο, ο αλγόριθμος χρησιμοποιώντας ως σημείο εκκίνησης το αρχικό περίγραμμα ολοκληρώνει την τμηματοποίηση με τη βοήθεια των GACs και του προγενέστερου σχήματος. Το μεγαλύτερο μέρος του κώδικα είναι γραμμένο στη γλώσσα προγραμματισμού C++ ενώ για την εξαγωγή των προ-εικόνων χρησιμοποιείται αλγόριθμος υλοποιημένος στο MATLAB.



Εικόνα 4.1: Στο διάγραμμα απεικονίζονται οι διαδικασίες που εκτελεί η εφαρμογή και πως συνδέονται μεταξύ τους.

4.1.1 Παραγωγή προγενέστερου σχήματος

Στο πρώτο βήμα επιλέγεται το σύνολο των εικόνων που θα χρησιμοποιηθεί για την παραγωγή του προγενέστερου σχήματος. Οι εικόνες αυτές είναι δυαδικές καθώς είναι αποτέλεσμα τμηματοποίησης και αναπαριστούν τη μορφή της γλωττίδας σε διάφορες φάσεις του κύκλου ταλάντωσης. Επίσης έχουν υποστεί προεπεξεργασία έτσι ώστε όλες οι γλωττίδες να βρίσκονται στην ίδια θέση και να έχουν το ίδιο μέγεθος και τον ίδιο προσανατολισμό. Στη συνέχεια κάθε εικόνα μετασχηματίζεται με τη συνάρτηση προσημασμένης απόστασης, μετατρέπεται σε διάνυσμα και όλα τα διανύσματα αποθηκεύονται σε ένα αρχείο με τη μορφή πίνακα στον οποίο οι στήλες αντιστοιχούν στις διαστάσεις των διανυσμάτων. Στο επόμενο βήμα υπολογίζεται το στατιστικό μοντέλο του σχήματος. Υπάρχει η δυνατότητα να εφαρμοστεί η μέθοδος της PCA και η μέθοδος της Kernel PCA με χρήση του πυρήνα γκάους (gauss kernel function). Αν επιλεγεί η πρώτη τεχνική τότε υπολογίζονται οι κύριες συνιστώσες και η μέση εικόνα οπότε δεν χρειάζεται γίνει περαιτέρω επεξεργασία. Αν επιλεγεί η δεύτερη τεχνική τότε πρέπει επιπλέον να βρεθούν οι προ-εικόνες. Για την υλοποίηση της PCA βασίζομαστε στον αλγόριθμο [65], για την Kernel PCA χρησιμοποιούμε αλγόριθμο από το MLPACK [66] κάνοντας πρώτα κάποιες βελτιώσεις και για την ανακατασκευή των προ-εικόνων χρησιμοποιούμε τον αλγόριθμο [8] μετά από ορισμένες τροποποιήσεις. Έτσι λοιπόν το στατιστικό μοντέλο σχήματος κατασκευάζεται από 4 εικόνες, τη μέση εικόνα και τις 3 εικόνες που δείχνουν τη διαφοροποίηση της γλωττίδας (κύριες συνιστώσες ή κύριες συνιστώσες πυρήνα) και με βάση το μοντέλο αυτό παράγεται το προγενέστερο σχήμα.



Εικόνα 4.2: Το διάγραμμα ροής για τη διαδικασία της παραγωγής προγενέστερου σχήματος.

4.1.2 Τμηματοποίηση - Διαμόρφωση αρχικού περιγράμματος

Η βασική τεχνική που χρησιμοποιείται συνήθως σε αλγορίθμους κατάτμησης εικόνας με GACs περιλαμβάνει τον ορισμό κάποιων αρχικών σημείων εκκίνησης εντός της περιοχής του αντικειμένου ενδιαφέροντος και έπειτα τη σταδιακή επέκταση του ενεργού περιγράμματος που σχηματίζουν. Ωστόσο στη παράγραφο των σχετικών εργασιών παρατηρούμε ότι υπάρχουν εναλλακτικές μέθοδοι στις οποίες η κατάτμηση βασίζεται σε ένα αρχικό περίγραμμα αντί των σημείων εκκίνησης. Στη παρούσα εργασία προτείνεται μία νέα προσέγγιση για τον εντοπισμό της γλωττίδας και τη δημιουργία ενός αρχικού περιγράμματος κατά τη διαδικασία τμηματοποίησης της γλωττίδας. Ο αλγόριθμος υλοποιείται σε επτά βήματα και σκοπός του είναι να υπολογίσει το αρχικό περίγραμμα το οποίο θα λειτουργήσει αργότερα ως το μηδενικό επιπεδοσύνολο με το οποίο θα ξεκινήσει η τμηματοποίηση και επιπλέον θα βοηθήσει στην τοποθέτηση του προγενέστερου σχήματος. Αρχικά εξάγεται η εικόνα από την ερυθρά συνιστώσα του μοντέλου RGB και γίνεται κατωφλίωση με τη μέθοδο Otsu ώστε να βρεθούν τα αντικείμενα της εικόνας. Στη συνέχεια με τη βοήθεια των ακμών διαμορφώνεται το σχήμα που αντιπροσωπεύει τη γλωττίδα και ορίζεται ως το αρχικό περίγραμμα. Ακολουθεί αναλυτική περιγραφή του αλγορίθμου.

Στο πρώτο βήμα χρησιμοποιείται το χρωματικό μοντέλο RGB για να αναδειχθούν οι διαφορές ανάμεσα στη περιοχή της γλωττίδας και τις υπόλοιπες περιοχές της εικόνας όπως οι φωνητικές χορδές και οι ιστοί του λάρυγγα. Εξετάζοντας τις τρεις συνιστώσες του μοντέλου φαίνεται να βοηθάει περισσότερο ιεραρχικά η ερυθρά συνιστώσα, η πράσινη και τέλος η μπλε. Ως προεπιλογή λοιπόν χρησιμοποιείται η ερυθρά συνιστώσα αλλά ο χρήστης μπορεί να επιλέξει κάποια από τις άλλες δύο αν συναντήσει προβλήματα κατά τον υπολογισμό του αρχικού περιγράμματος.

Στη συνέχεια με τη μέθοδο Otsu γίνεται μία πρώτη τμηματοποίηση στην εικόνα. Η μέθοδος προσπαθεί να διαχωρίσει τα εικονοστοιχεία σε δύο κατηγορίες, το παρασκήνιο ή φόντο (background) και το προσκήνιο (foreground). Για να το επιτύχει αυτό στηρίζεται στο ιστόγραμμα της εικόνας και υπολογίζει το κατώφλι εκείνο για το

οποίο το άθροισμα των διακυμάνσεων των δύο κατηγοριών θα είναι το μικρότερο δυνατό. Από τη διαδικασία αυτή προκύπτει μία δυαδική εικόνα όπου το παρασκήνιο χρωματίζεται με άσπρο χρώμα ενώ το προσκήνιο με μαύρο και σχηματίζει έναν αριθμό αντικειμένων μεταξύ των οποίων θα βρίσκεται και η περιοχή της γλωττίδας.

Η κατωφλίωση κατά Otsu όμως συνήθως δεν αρκεί για να προσδιοριστούν τα όρια της γλωττίδα καθώς το αντίστοιχο αντικείμενο περιλαμβάνει μία ευρύτερη έκταση και επίσης δεν έχει κάποιο χαρακτηριστικό γνώρισμα για να ξεχωρίσει από τα υπόλοιπα αντικείμενα ώστε να οριστεί ως αρχικό περίγραμμα. Γι'αυτούς τους λόγους χρησιμοποιούνται επιπρόσθετα οι ακμές της αρχικής εικόνας. Για τον εντοπισμό των ακμών βασίζομαστε στον τελεστή sobel ο οποίος κάνει συνέλιξη σε οριζόντια και κάθετη διεύθυνση με δύο ισοτροπικές μάσκες διαστάσεων 3x3

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A \quad , \quad G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A$$

όπου A η αρχική εικόνα και * ο τελεστής συνέλιξης. Η τελική εικόνα των ακμών δίνεται από τη σχέση

$$G = \sqrt{G_x^2 + G_y^2}$$

Έπειτα αποτυπώνονται οι ακμές πάνω στη κατωφλιωμένη εικόνα με αποτέλεσμα να υπολογίζονται με μεγαλύτερη ακρίβεια τα όρια της γλωττίδας. Μέσω της παραμέτρου `edge_sensitivity` καθορίζεται η ευαισθησία ως προς στις ακμές που χρησιμοποιούνται επηρεάζοντας με αυτόν το τρόπο τον σχηματισμό των αντικειμένων. Τα αντικείμενα που δημιουργούνται σε αυτή τη φάση λειτουργούν ως υποψήφια αρχικά περιγράμματα για τη γλωττίδα.

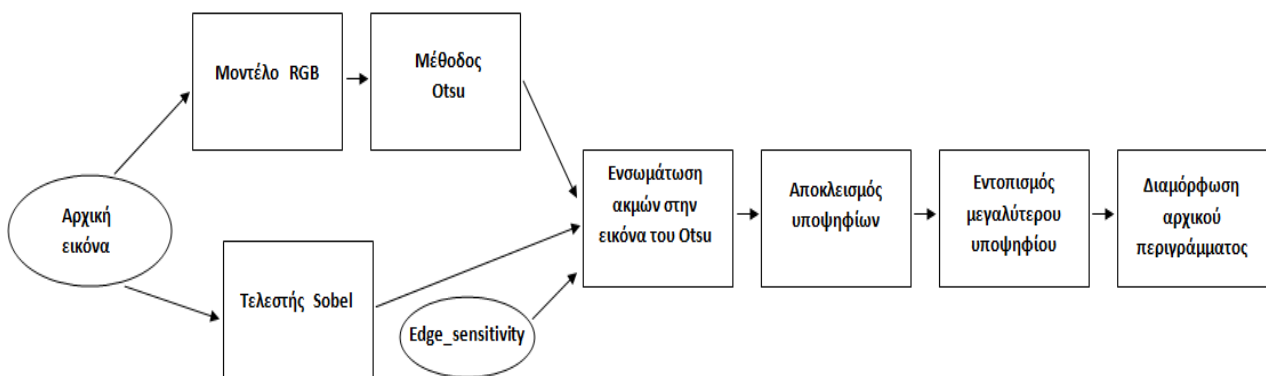
Στο τελικό στάδιο διακρίνεται το αντικείμενο που περιέχει τη γλωττίδα. Το χαρακτηριστικό πάνω στο οποίο βασίζεται ο αλγόριθμος είναι το εμβαδόν των αντικειμένων και θεωρεί ότι η γλωττίδα βρίσκεται στο μεγαλύτερο από αυτά. Ωστόσο είναι πιθανό άλλα αντικείμενα που αντιστοιχούν σε ιστούς του λάρυγγα να έχουν μεγαλύτερο μέγεθος. Σε αυτές τις περιπτώσεις όμως τα αντικείμενα εκτείνονται μέχρι τα άκρα της εικόνας ενώ το αντικείμενο της γλωττίδας, χάρις στις ακμές, διαχωρίζεται πλήρως. Επομένως τα αντικείμενα που εφάπτονται με τα άκρα μπορούν με ασφάλεια να αποκλειστούν από τα υποψήφια αρχικά περιγράμματα και προσαρτώνται από το παρασκήνιο. Σημαντικό ρόλο στη διαδικασία παίζει η παράμετρος `edge_sensitivity` και πρέπει με τη τιμή που λαμβάνει να πετυχαίνεται η κατάλληλη ισορροπία έτσι ώστε το αντικείμενο της γλωττίδας να διαχωρίζεται από άλλα αντικείμενα ή τα άκρα της εικόνας και συνάμα να αποκτά το μεγαλύτερο μέγεθος. Το σχήμα λοιπόν που ικανοποιεί τα παραπάνω κριτήρια προκρίνεται και ορίζει το αρχικό περίγραμμα ενώ τα υπόλοιπα αντικείμενα περνάνε στο παρασκήνιο. Τέλος, εξαλείφεται ο θόρυβος που υπάρχει εντός του αρχικού περιγράμματος, του δίνεται το άσπρο χρώμα και το παρασκήνιο χρωματίζεται με μαύρο.

Κατά την διεξαγωγή των δοκιμών χρησιμοποιήθηκαν εικόνες από βίντεο στροβοσκόπησης και ένα κοινό πρόβλημα που παρουσιάστηκε σε αρκετές εικόνες ήταν ότι περικλείονταν από ένα μαύρο πλαίσιο. Αυτό είχε σαν αποτέλεσμα να εμφανίζονται ιδιαίτερα έντονες ακμές στα όρια του πλαισίου γεγονός που δυσκόλευε τον σωστό προσδιορισμό του αρχικού περιγράμματος αφού προέκυπταν πολλά αντικείμενα τα οποία ήταν μεγαλύτερα από αυτό της γλωττίδας και δεν μπορούσαν να αποκλειστούν καθώς οι ακμές του πλαισίου τα απέκοπταν από τα άκρα της εικόνας. Για να ξεπεραστεί το πρόβλημα, ο αλγόριθμος ελέγχει αν υπάρχει πλαίσιο γύρω από την εικόνα, υπολογίζει χονδρικά το πλάτος του σε κάθε πλευρά και στη συνέχεια στο στάδιο που ενσωματώνει τις ακμές, περιλαμβάνει μόνο τον αντίστοιχο χώρο των ακμών που βρίσκονται εντός του πλαισίου που έχει υπολογίσει.

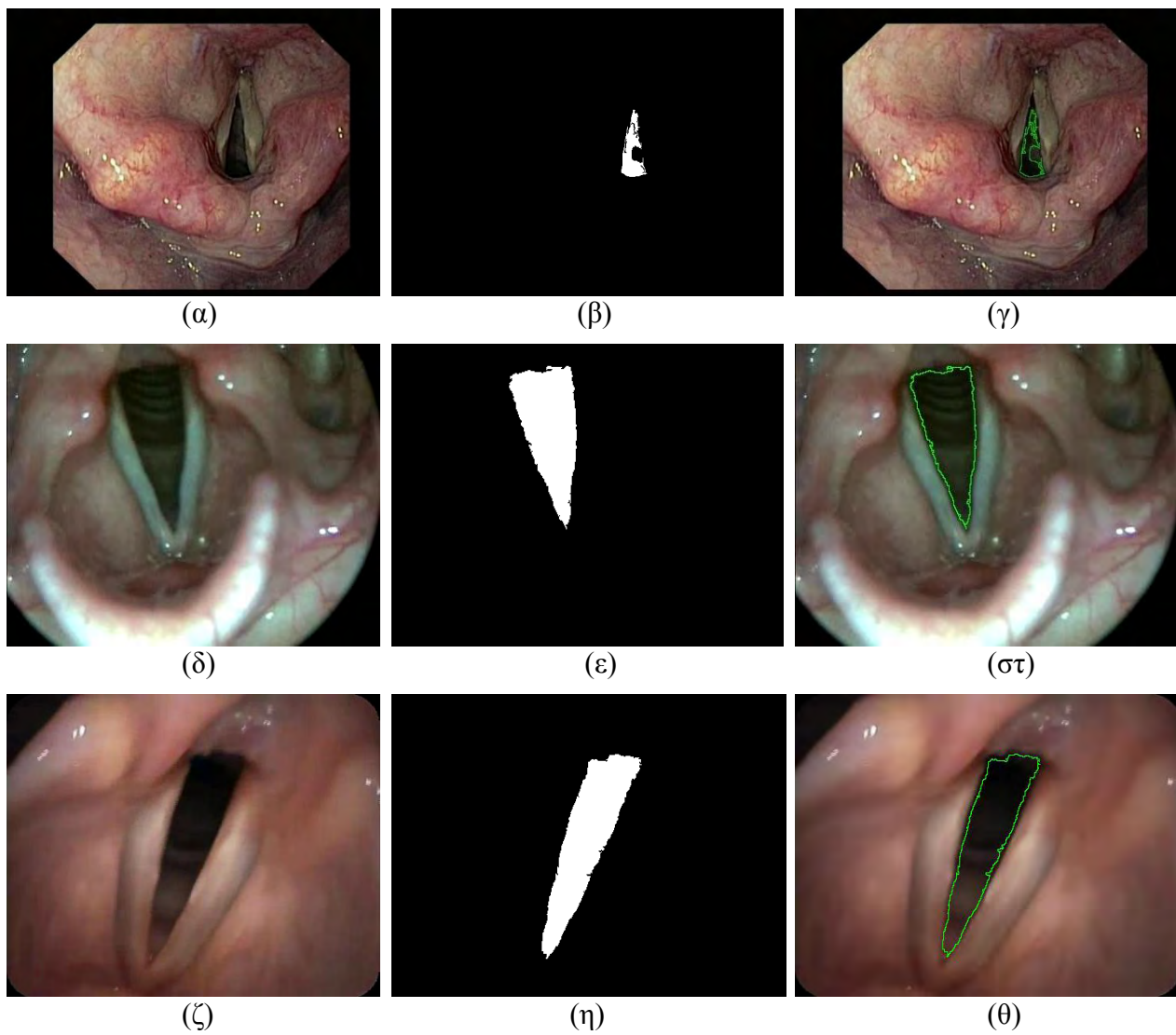
ΑΛΓΟΡΙΘΜΟΣ 1

ΥΠΟΛΟΓΙΣΜΟΣ ΑΡΧΙΚΟΥ ΠΕΡΙΓΡΑΜΜΑΤΟΣ

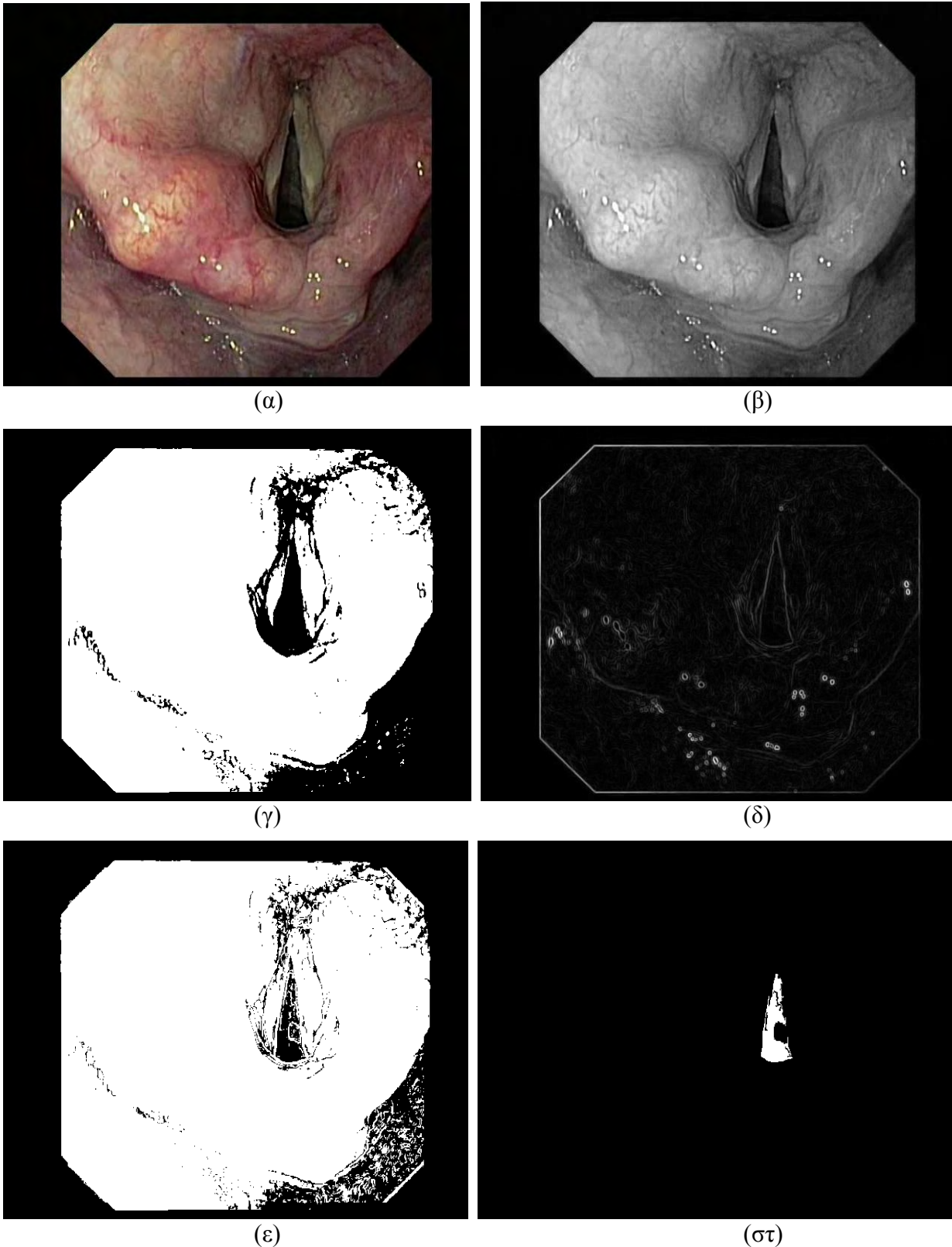
- 1) Εξαγωγή εικόνας από την ερυθρά συνιστώσα του μοντέλου RGB
- 2) Κατωφλίωση Otsu
- 3) Υπολογισμός ακμών αρχικής εικόνας
- 4) Ενσωμάτωση ακμών στη κατωφλιωμένη εικόνα
- 5) Αποκλεισμός υποψηφίων αντικειμένων
- 6) Εντοπισμός του υποψηφίου αντικειμένου με το μεγαλύτερο μέγεθος
- 7) Διαμόρφωση αρχικού περιγράμματος



Εικόνα 4.3: Διάγραμμα ροής αλγορίθμου 1 για τον υπολογισμό του αρχικού περιγράμματος.



Εικόνα 4.4: Στις εικόνες φαίνονται τρία παραδείγματα υπολογισμού αρχικού περιγράμματος. Οι (α),(δ),(ζ) δείχνουν τις αρχικές εικόνες, οι (β),(ε),(η) τα αρχικά περιγράμματα και στις (γ),(στ),(θ) προβάλλεται το αρχικό περίγραμμα πάνω στην αρχική εικόνα.



Εικόνα 4.5: Οι εικόνες που παράγονται σε κάθε βήμα του αλγορίθμου 1. Κατά σειρά φαίνονται η αρχική εικόνα, η εικόνα από την ερυθρά συνιστώσα του μοντέλου RGB, η κατωφλίωση με τη μέθοδο Otsu, οι ακμές που υπολογίζονται με φίλτρο sobel, η αποτύπωση των ακμών πάνω στη κατωφλιωμένη εικόνα και τέλος το αρχικό περίγραμμα.

4.1.3 Τμηματοποίηση - Υπολογισμός τελικής καμπύλης

Για τον υπολογισμό του τελικού περιγράμματος βασιζόμαστε στο φίλτρο `GeodesicActiveContourShapePriorLevelSetImageFilter` του `Insight Toolkit` και η βασική συνάρτηση που χρησιμοποιείται είναι η `GeodesicActiveContour`. Η συνάρτηση δέχεται στις εισόδους της την εικόνα που περιέχει το αρχικό επιπεδοσύνολο με το οποίο θα ξεκινήσει η τμηματοποίηση και την εικόνα των χαρακτηριστικών που παρουσιάζει τις ακμές της αρχικής εικόνας. Επιπρόσθετα δέχεται τρεις εισόδους που σχετίζονται με το στατιστικό μοντέλο, τρεις εισόδους με μεταβλητές που ορίζουν την ταχύτητα επέκτασης, τη μετατόπιση και τη καμπυλότητα της εξελισσόμενης καμπύλης καθώς επίσης και μία είσοδο με τη μεταβλητή που δείχνει το βαθμό εμπιστοσύνης στο προγενέστερο σχήμα. Ο αλγόριθμος υλοποιείται σε δύο μέρη. Στο πρώτο μέρος γίνονται οι αρχικοποιήσεις, βρίσκει το σημείο όπου πρέπει να τοποθετηθεί το προγενέστερο σχήμα, υπολογίζει την εικόνα χαρακτηριστικών και ορίζει το μηδενικό επιπεδοσύνολο. Στο δεύτερο μέρος λειτουργεί επαναληπτικά και επεκτείνει κάθε φορά τη καμπύλη στηριζόμενος στην εξίσωση (3.26) με τη μεταβλητή λ_1 να καθορίζεται από την παράμετρο `PropagationScaling` και τη μεταβλητή λ_2 από τη παράμετρο `ShapePriorScaling`. Οι παράμετροι `Sigma`, `PropagationScaling` και `ShapePriorScaling` αποτελούν τις τρεις παραμέτρους τμηματοποίησης με τις οποίες ο χρήστης ελέγχει τη διαδικασία κατάτμησης και ανήκουν στη συνάρτηση `GeodesicActiveContour`. Στη συνέχεια εξετάζουμε αναλυτικά τη λειτουργία του αλγορίθμου.

Αρχικά υπολογίζεται η τοποθέτηση του προγενέστερου σχήματος μέσα στην εικόνα. Για τον σκοπό αυτό χρησιμοποιούνται η εικόνα που περιέχει το αρχικό περίγραμμα καθώς δείχνει τη περιοχή της γλωττίδας και η εικόνα του μέσου σχήματος για το προγενέστερο σχήμα. Κάθε εικόνα καταφλιώνεται σε μία δυαδική εικόνα και μετασχηματίζεται με τη συνάρτηση προσημασμένης απόστασης. Επομένως βρίσκονται τα σημεία του σχήματος που είναι πιο κοντά στο εσωτερικό του και με βάση τα σημεία αυτά υπολογίζεται το κέντρο κάθε σχήματος. Αφού έχουν βρεθεί τα κέντρα των δύο σχημάτων και είναι γνώστες οι αρχικές τους θέσεις μέσα στην εικόνα, εύκολα μπορεί να υπολογιστεί η μετατόπιση που πρέπει να γίνει έτσι ώστε το προγενέστερο σχήμα να συμπέσει με τη γλωττίδα.

Στο επόμενο βήμα γίνεται η προεπεξεργασία της εικόνας. Στη διαδικασία χρησιμοποιείται ένα φίλτρο εξομάλυνσης για την απαλοιφή του θορύβου, ένα φίλτρο για τον υπολογισμό της κλίσης της εικόνας και ένα φίλτρο για την εύρεση των ακμών. Τα φίλτρα χρησιμοποιούνται διαδοχικά με τη τεχνική της διασωλήνωσης (pipeline) και στο τέλος της διαδικασίας προκύπτει η εικόνα των χαρακτηριστικών που αναπαριστά το μέγεθος των ακμών της αρχικής εικόνας. Η εικόνα χαρακτηριστικών ρυθμίζεται μέσω της παραμέτρου `Sigma`, επηρεάζει την εξίσωση (3.18) και καθορίζει πόσο εύκολα ή δύσκολα επεκτείνεται η καμπύλη στις περιοχές της αρχικής εικόνας.

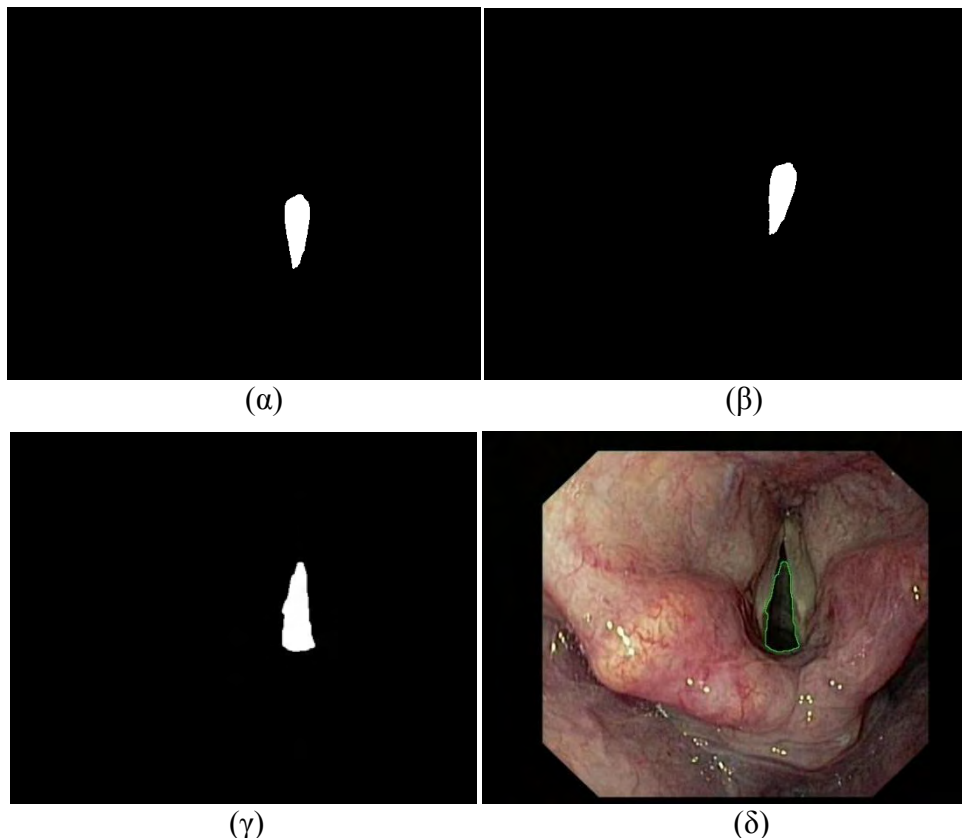
Έπειτα ορίζεται το μοντέλο GAC. Το μηδενικό επιπεδοσύνολο ορίζεται με βάση την εικόνα του αρχικού περιγράμματος που έχει υπολογιστεί προηγουμένως από

τον αλγόριθμο 1. Η τεχνική που χρησιμοποιείται για την εξέλιξη της καμπύλης είναι αυτή της ταχείας προέλασης (fast marching) και έχει σταθερή διεύθυνση. Η ταχύτητα επέκτασης καθορίζεται με τη παράμετρο *PropagationScaling* και από προεπιλογή ορίζεται ότι έχει θετική κατεύθυνση, επομένως το μέτωπο της ταχείας προέλασης κινείται πάντα προς το εξωτερικό της καμπύλης.

Στη συνέχεια υλοποιείται το μέρος που σχετίζεται με το προγενέστερο σχήμα. Αρχικά εισάγονται στο μοντέλο το μέσο σχήμα και οι κύριες συνιστώσες ή κύριες συνιστώσες πυρήνες ανάλογα με τη τεχνική που χρησιμοποιούμε. Σε κάθε επανάληψη του αλγορίθμου, γίνεται μία εκτίμηση για την τελική καμπύλη του αντικειμένου και ανάλογα επανυπολογίζεται το καλύτερα εφαρμοζόμενο σχήμα (best-fit shape) το οποίο βοηθάει στην εξέλιξη της καμπύλης. Ο χρήστης καθορίζει τον βαθμό επιρροής του προγενέστερου σχήματος με τη παράμετρο *ShapePriorScaling* και ο αλγόριθμος υπολογίζει το καλύτερα εφαρμοζόμενο σχήμα από την εξίσωση

$$u^*(x) = \mu(x) + \sum_{k=1}^p \alpha_k \psi_k(x)$$

όπου $\mu(x)$ το μέσο σχήμα, $\psi_k(x)$ οι K πρώτες κύριες συνιστώσες και α_k είναι οι παράμετροι που διαμορφώνουν το σχήμα.



Εικόνα 4.6: Η (α) αντιστοιχεί στο προγενέστερο σχήμα της γλωττίδας κατά τη μηδενική επανάληψη και η (β) δείχνει το προγενέστερο σχήμα που προκύπτει στη τελευταία επανάληψη μετά τις τροποποιήσεις που δέχεται κατά την εκτέλεση του αλγορίθμου 2. Στην (γ) φαίνεται το τελικό περίγραμμα που υπολογίζεται και στην (δ) αποτυπώνεται το τελικό περίγραμμα στην εικόνα όπου έγινε η κατάτμηση.

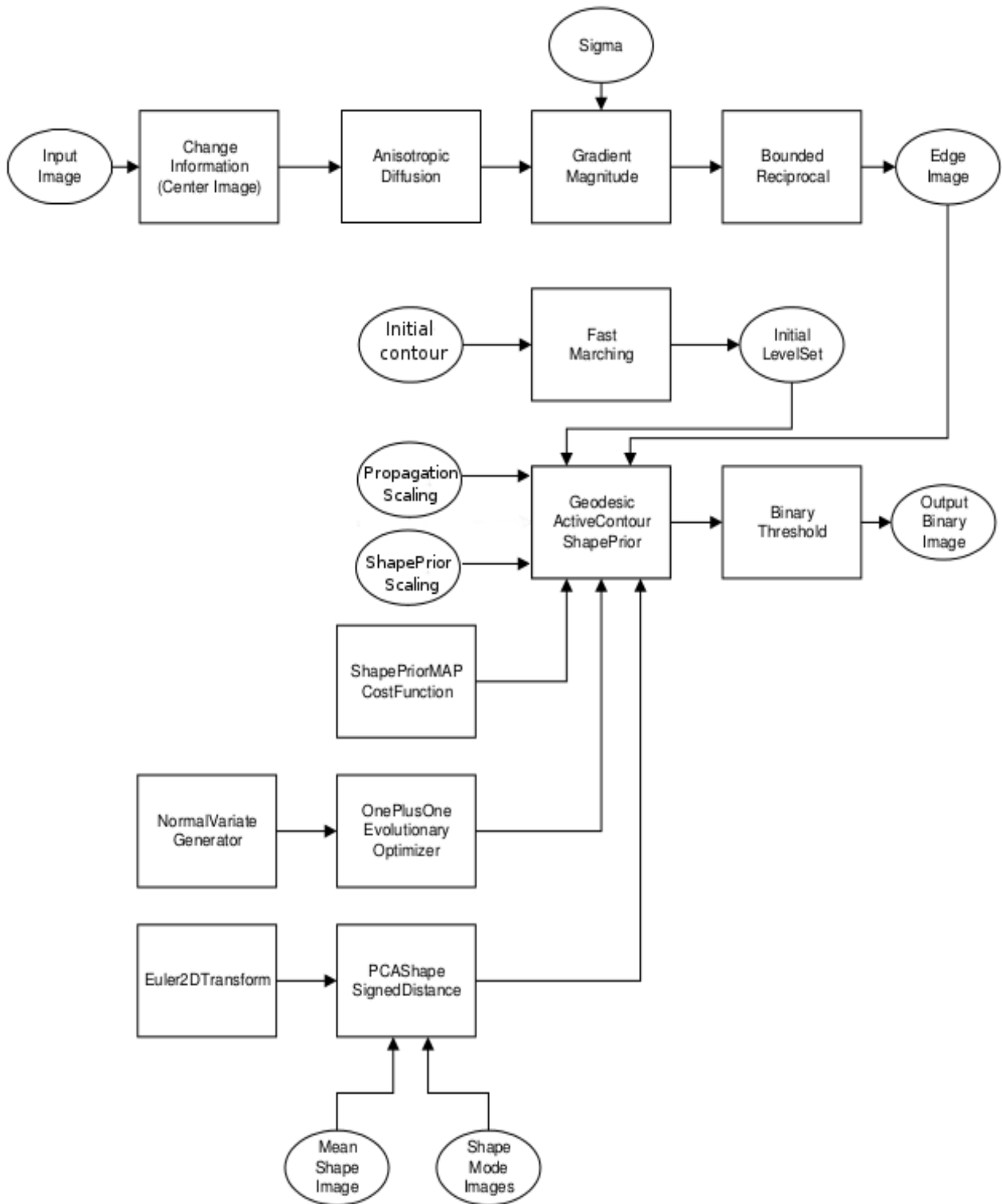
Το καλύτερα εφαρμοζόμενο σχήμα διαμορφώνεται από τις συναρτήσεις Euler2DTransform και ShapePriorMAPCostFunction. Η πρώτη ευθύνεται για τις p παραμέτρους της εξίσωσης (3.19) που ελέγχουν τη θέση και τον προσανατολισμό και η δεύτερη αφορά τις a παραμέτρους της (3.19) που ορίζουν τη μορφή του σχήματος. Η συνάρτηση OnePlusOneEvolutionaryOptimizer λειτουργεί συμπληρωματικά με την ShapePriorMAPCostFunction και έχει στόχο να βελτιστοποιήσει το καλύτερα εφαρμοζόμενο σχήμα ξεκινώντας δοκιμές με τυχαίες τιμές για τις παραμέτρους a , p μέχρι να καταλήξει τελικά στις πιο αποδοτικές. Επίσης χρησιμοποιούνται τέσσερα βάρη τα οποία αντιστοιχούν στις τέσσερις πιθανότητες της εξίσωσης (3.21) και με αυτόν το τρόπο ελέγχεται η επιρροή του κάθε όρου στη μορφοποίηση του τελικού σχήματος.

Ως καλύτερα εφαρμοζόμενο σχήμα ορίζεται καταρχήν το μέσο σχήμα και στη συνέχεια διαμορφώνεται έτσι ώστε να προσαρμοστεί καλύτερα στο αντικείμενο της εικόνας. Η εφαρμογή εκτελεί τον αλγόριθμο δύο φορές. Τη πρώτη φορά ψάχνει μέσα σε ένα εύρος επαναλήψεων που καθορίζεται από τον χρήστη, για ποιόν αριθμό επαναλήψεων πετυχαίνει την ιδανικότερη τμηματοποίηση. Τη δεύτερη φορά εκτελεί πάλι τον αλγόριθμο χρησιμοποιώντας αυτή τη φορά τον αριθμό επαναλήψεων που ξεχώρισε υπολογίζοντας έτσι το τελικό περίγραμμα της γλωττίδας. Για την εύρεση του συνολικού αριθμού επαναλήψεων βασίζεται στη συνάρτηση RMSChange η οποία σε κάθε επανάληψη υπολογίζει το ποσοστό διαφοράς για τις εξελισσόμενες καμπύλες των δύο τελευταίων επαναλήψεων. Η επανάληψη που δίνει τη μικρότερη διαφορά θεωρείται ότι είναι η καλύτερη και ορίζει τον συνολικό αριθμό επαναλήψεων.

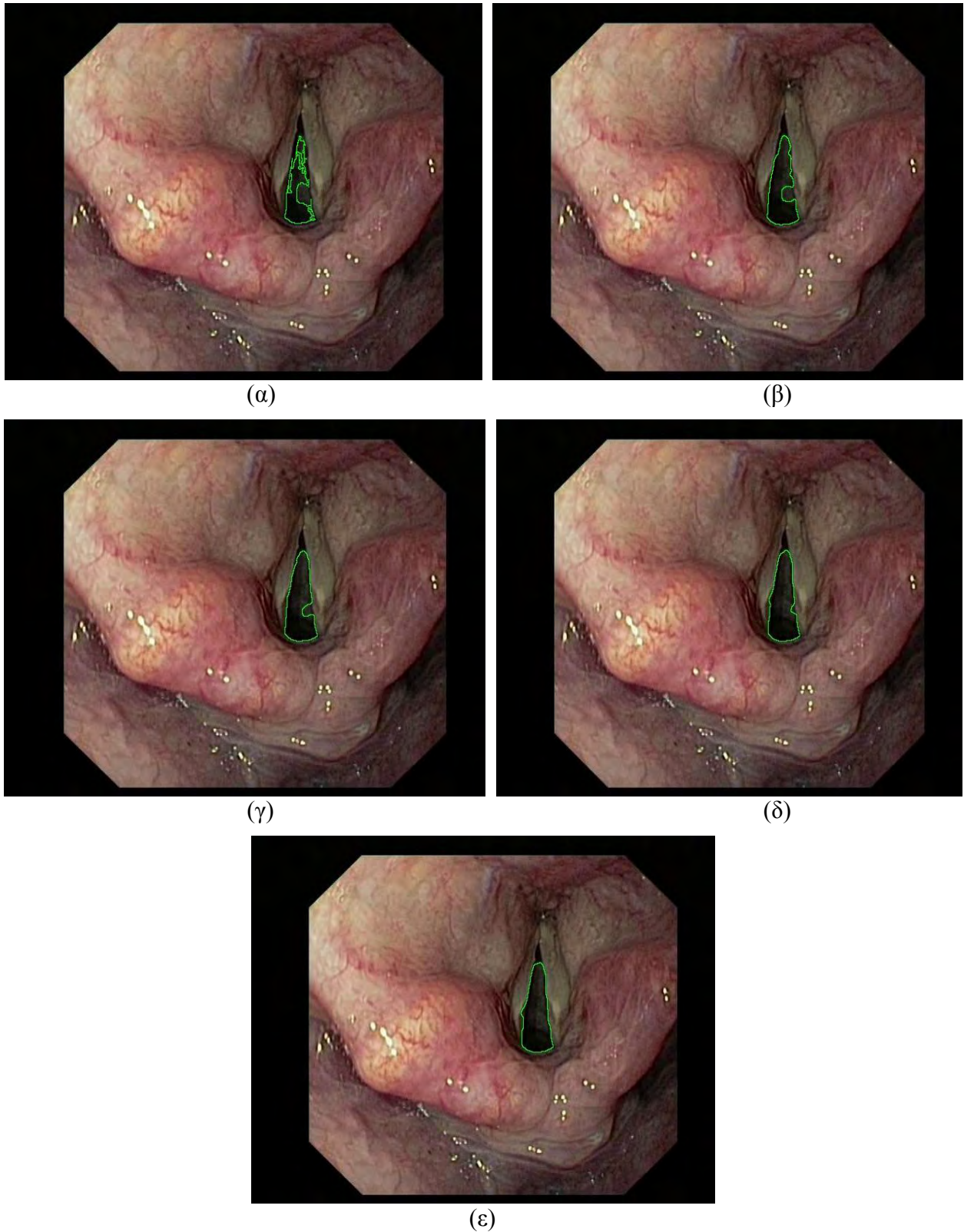
ΑΛΓΟΡΙΘΜΟΣ 2

ΥΠΟΛΟΓΙΣΜΟΣ ΤΕΛΙΚΟΥ ΠΕΡΙΓΡΑΜΜΑΤΟΣ

- 1) Εύρεση θέσης προγενέστερου σχήματος
 - 2) Εξαγωγή εικόνας χαρακτηριστικών
 - 3) Αρχικοποίηση μοντέλου GAC
 - 4) Εισαγωγή προγενέστερου σχήματος
 - 5) Για N επαναλήψεις
 1. Υπολογισμός a, p παραμέτρων προγενέστερου σχήματος
 2. Εξέλιξη καμπύλης μέσα από τον συνδυασμό του γεωδαιτικού ενεργού περιγράμματος και του προγενέστερου σχήματος
- Τέλος επαναλήψεων
-



Εικόνα 4.7: Διάγραμμα ροής αλγορίθμου 2 για την διαδικασία υπολογισμού του τελικού περιγράμματος. Το ‘Mean Shape Image’ συμβολίζει τη μέση εικόνα από το στατιστικό μοντέλο σχήματος και το ‘Shape Mode Images’ τις εικόνες διακύμανσης σχήματος (κύριες συνιστώσες ή κύριες συνιστώσες πυρήνα) [59].



Εικόνα 4.8: Στις εικόνες φαίνεται η εξέλιξη της καμπύλης με τον αλγόριθμο 2 χρησιμοποιώντας τη KPCA για τη παραγωγή του προγενέστερου σχήματος. Η εικόνα (α) προκύπτει από το αρχικό περίγραμμα και αντιστοιχεί στη μηδενική επανάληψη. Στη συνέχεια εμφανίζεται η καμπύλη στη 50ή επανάληψη, στην 100ή, στη 150ή και τέλος στη 200ή επανάληψη προκύπτει η τελική καμπύλη που τμηματοποιεί τη γλωττίδα.

4.2 Εργαλεία υλοποίησης

Για την κατασκευή της εφαρμογής χρησιμοποιούνται εργαλεία που εξασφαλίζουν την υποστήριξη που απαιτείται για την ανάπτυξη των αλγόριθμων. Πιο συγκεκριμένα το Insight toolkit παρέχει τις βιβλιοθήκες που επιτρέπουν την επεξεργασία της εικόνας και το MLPACK περιέχει τις συναρτήσεις με τις οποίες πραγματοποιείται η τεχνική της Kernel PCA. Η ανάπτυξη της εφαρμογής γίνεται στο Eclipse το οποίο προσφέρει ένα κατάλληλο περιβάλλον σχεδίασης λογισμικού που διευκολύνει τη συγγραφή του κώδικα. Οι γλώσσες προγραμματισμού που χρησιμοποιούνται είναι κατά κύριο λόγο η C++ καθώς σε αυτή τη γλώσσα είναι υλοποιημένα τα φίλτρα του ITK και το MATLAB. Ακόμα, το CMake έχει ενεργό ρόλο στη διαδικασία κατασκευής (build process) των επί μέρους τμημάτων της εφαρμογής.

4.2.1 Insight Segmentation and Registration Toolkit

Το Insight Segmentation and Registration Toolkit (ITK) [59] είναι μία βιβλιοθήκη ανάπτυξης λογισμικού που επικεντρώνεται στη επεξεργασία βιοϊατρικών εικόνων και ειδικότερα στην αντιστοίχιση εικόνων (image registration) και την τμηματοποίηση εικόνων. Είναι ανοιχτού κώδικα, ελεύθερα διαθέσιμο και έχει υλοποιηθεί στη γλώσσα προγραμματισμού C++. Επίσης είναι διαπλατφορμικό (cross-platform) παρέχοντας, μέσω της στενής συνεργασίας με το πρόγραμμα CMake, διαλειτουργικότητα μεταξύ των λειτουργικών συστημάτων (Unix, Windows, MacOSX).

Το Insight Toolkit έχει υλοποιήσει μία μεγάλη γκάμα από μεθόδους αιχμής που σχετίζονται με την επεξεργασία εικόνων και μπορεί να χειριστεί εικόνες δισδιάστατης αλλά και τρισδιάστατης απεικόνισης. Ένα επιπλέον χαρακτηριστικό είναι η αυτοματοποιημένη μέθοδος περιτυλίγματος (wrapping process) η οποία επιτρέπει την δημιουργία διεπαφών μεταξύ της C++ και άλλων γλωσσών προγραμματισμού όπως η Java και η Python.

Η ανάπτυξη του ITK ξεκίνησε το 1999 από το Insight Software Consortium με τη χρηματοδότηση του National Library of Medicine των ΗΠΑ. Στη σύμπραξη συμμετείχαν αρχικά τρεις εμπορικοί συνεργάτες, GE Corporate R&D, Kitware, Inc., MathSoft (insightfull) και τρεις ακαδημαϊκοί συνεργάτες, University of North Carolina, University of Tennessee, and University of Pennsylvania. Η άδεια χρήσης που χρησιμοποιεί είναι η Apache 2. Η έκδοση που επιλέχθηκε είναι η InsightToolkit - 4.2.1 καθώς ήταν η πιο πρόσφατη και σταθερή.

4.2.2 MLPACK

Το MLPACK [66] είναι μία βιβλιοθήκη λογισμικού που περιέχει υλοποιημένες μεθόδους από το πεδίο της Μηχανικής Μάθησης. Είναι γραμμένο στη γλώσσα προγραμματισμού C++ και στηρίζεται στη βιβλιοθήκη *armadillo* η οποία παρέχει συναρτήσεις γραμμικής άλγεβρας. Είναι ανοιχτού κώδικα, διαπλατοφορμικό και είναι ελεύθερα διαθέσιμο. Μερικοί από τους αλγορίθμους που προσφέρει είναι η Ανάλυση Κύριων Συνιστωσών (PCA), η Ανάλυση Κύριων Συνιστωσών Πυρήνα (Kernel PCA), η Ανάλυση Ανεξάρτητων Συνιστωσών (ICA), η Συσταδοποίηση K-Means (K-Means Clustering) και τα Κρυπτο-Μαρκοβιανά Μοντέλα (HMMs).

Το MLPACK αναπτύχθηκε από το Fundamental Algorithmic and Statistical Tools Laboratory στο Georgia Institute of Technology και υπόκειται στην άδεια χρήσης LGPL v3+. Παρουσιάστηκε πρώτη φορά στο BigLearning workshop του NIPS το 2011 και έχει δημοσιευτεί στο περιοδικό *Journal of Machine Learning Research*. Η έκδοση που επιλέχθηκε είναι η MLPACK - 1.0.8, ενώ από την πλευρά του *armadillo* χρησιμοποιείται η έκδοση *armadillo* - 4.000.2.

Πειράματα και Αποτελέσματα

Στο πέμπτο κεφάλαιο περιγράφεται η μεθοδολογία που ακολουθήθηκε κατά τη διεξαγωγή των πειραμάτων και παρουσιάζονται τα αποτελέσματα που αποκτήθηκαν. Πραγματοποιήθηκαν δύο πειράματα κατάτμησης γλωττίδας σε εικόνες λάρυγγα επιλέγοντας κάθε φορά διαφορετική τεχνική για την εξαγωγή του προγενέστερου σχήματος. Στο πρώτο πείραμα το στατιστικό μοντέλο σχήματος παράγεται με βάση την PCA ενώ στο δεύτερο πείραμα χρησιμοποιείται η Kernel PCA. Η παραγωγή του προγενέστερου σχήματος και η τμηματοποίηση γίνονται χρησιμοποιώντας τις εικόνες από το σύνολο δεδομένων διαδοχικά με βάση τη τεχνική της διασταυρωμένης επικύρωσης και στη συνέχεια στηρίζομαστε σε διεθνώς αποδεκτούς δείκτες ομοιότητας για την αξιολόγηση των αποτελεσμάτων. Σκοπός των πειραμάτων είναι να εξεταστεί η αποτελεσματικότητα της κάθε τεχνικής στη παραγωγή του προγενέστερου σχήματος για τη διαδικασία της κατάτμησης.

5.1 Σύνολα δεδομένων

Για τα πειράματα βασιζόμαστε σε 5 βίντεο στροβοσκόπησης τα οποία είναι δημόσια διαθέσιμα [60-64]. Από κάθε βίντεο επιλέγονται 10 αντιπροσωπευτικά στιγμιότυπα σχηματίζοντας το αρχικό σύνολο εικόνων το οποίο περιέχει 50 εικόνες και αναπαριστά τη γλωττίδα σε όλες τις πιθανές φάσεις του κύκλου ταλάντωσης όπου μπορεί να εμφανιστεί. Επίσης οι εικόνες του αρχικού συνόλου επισημειώθηκαν από ειδικό δημιουργώντας το σύνολο των εικόνων επαλήθευσης το οποίο χρησιμοποιείται κατά την αξιολόγηση των αποτελεσμάτων.

5.2 Μεθοδολογία εκπαίδευσης και τμηματοποίησης

Η δημιουργία του προγενέστερου σχήματος και η κατάτμηση γίνονται με τη τεχνική της διασταυρωμένης επικύρωσης (cross-validation). Στη μέθοδο αυτή οι εικόνες διαχωρίζονται σε δύο σύνολα, το σύνολο εκπαίδευσης και το σύνολο εξέτασης. Οι εικόνες από το πρώτο σύνολο χρησιμοποιούνται στη διαδικασία της παραγωγής προγενέστερου σχήματος ενώ οι εικόνες από το δεύτερο σύνολο χρησιμοποιούνται στη διαδικασία της τμηματοποίησης. Οι εικόνες του αρχικού

συνόλου διαμοιράζονται κυκλικά στα δύο σύνολα οπότε σε κάθε κύκλο ‘κρατάμε’ έναν αριθμό εικόνων για την τμηματοποίηση και με τις υπόλοιπες εξάγουμε το προγενέστερο σχήμα. Στην εργασία επιλέγουμε την εκδοχή της διασταυρωμένης επικύρωσης πέντε αναδιπλώσεων (five-fold cross-validation) που σημαίνει ότι σε κάθε πείραμα η διεργασία εκπαίδευσης-εξέτασης γίνεται σε 5 κύκλους και σε κάθε κύκλο από τις 50 εικόνες του αρχικού συνόλου χρησιμοποιούμε τις 40 για να δημιουργήσουμε το προγενέστερο σχήμα και στις υπόλοιπες 10 εικόνες κάνουμε τις τμηματοποιήσεις με βάση το προγενέστερο σχήμα που υπολογίσαμε. Η τεχνική της διασταυρωμένης επικύρωσης αποδεικνύεται εξαιρετικά χρήσιμη ιδιαίτερα στις περιπτώσεις όπου το σύνολο δεδομένων είναι μικρό.

Στην εργασία κάνουμε δύο πειράματα όπου στο πρώτο πείραμα η εξαγωγή του προγενέστερου σχήματος γίνεται με την τεχνική της PCA και στο δεύτερο με την τεχνική της Kernel PCA. Τα πειράματα γίνονται σε 3 φάσεις, στη πρώτη φάση δημιουργούνται τα προγενέστερα σχήματα, στη δεύτερη φάση υπολογίζονται τα αρχικά περιγράμματα και στη τρίτη φάση γίνονται οι τμηματοποιήσεις των εικόνων. Σε κάθε αναδίπλωση του k-fold κατασκευάζονται τα στατιστικά μοντέλα σχήματος με PCA και Kernel PCA. Τα μοντέλα αυτά περιλαμβάνουν 4 εικόνες, τη μέση εικόνα και τις 3 συνιστώσες. Όπως είδαμε στη παράγραφο 4.1.3 η διαδικασία της τμηματοποίησης ελέγχεται από τον χρήστη μέσω τριών παραμέτρων, τη Sigma που διαμορφώνει την εικόνα χαρακτηριστικών με βάση τις ακμές, τη PropagationScaling που επηρεάζει την ταχύτητα εξέλιξης της καμπύλης και ορίζει τη μεταβλητή λ_1 της εξίσωσης (3.26) και τη ShapePriorScaling που δείχνει το βαθμό εμπιστοσύνης στο προγενέστερο σχήμα και ορίζει τη μεταβλητή λ_2 της (3.26). Στα πειράματα δίνονται από 3 τιμές σε κάθε παράμετρο οπότε δοκιμάζονται συνολικά 27 συνδυασμοί και με αυτούς τους ελέγχους προσπαθούμε να βρούμε τους συνδυασμούς των παραμέτρων για τους οποίους η τμηματοποίηση αποδίδει καλύτερα.

5.3 Μέτρα αξιολόγησης της τμηματοποίησης

Η αξιολόγηση των αποτελεσμάτων πραγματοποιείται με δύο δείκτες ομοιότητας, τον δείκτη Jaccard και τον δείκτη Dice. Κάθε μέθοδος για να εκτιμήσει το ποσοστό επιτυχίας της τμηματοποίησης σε μία εικόνα συγκρίνει την εικόνα της τμηματοποίησης με την αντίστοιχη εικόνα επαλήθευσης. Οι δύο αυτές εικόνες είναι δυαδικές και αναπαριστούν με άσπρο χρώμα τη περιοχή της γλωττίδας και μαύρο τη περιοχή του παρασκηνίου.

Ο πρώτος δείκτης εισήχθηκε από τον Paul Jaccard [29] και στηρίζεται στον λόγο

$$J = \frac{|A \cap B|}{|A \cup B|}$$

όπου A,B οι δύο συγκρινόμενες εικόνες και $J \in [0,1]$ ο δείκτης Jaccard.

Ο δεύτερος δείκτης προτάθηκε ξεχωριστά από τους Dice το 1945 [30] και Sorensen το 1948 [31] και προκύπτει από τον λόγο

$$D = \frac{2|A \cap B|}{|A| + |B|}$$

όπου A,B είναι οι εικόνες που συγκρίνονται και $D \in [0,1]$ ο δείκτης Dice.

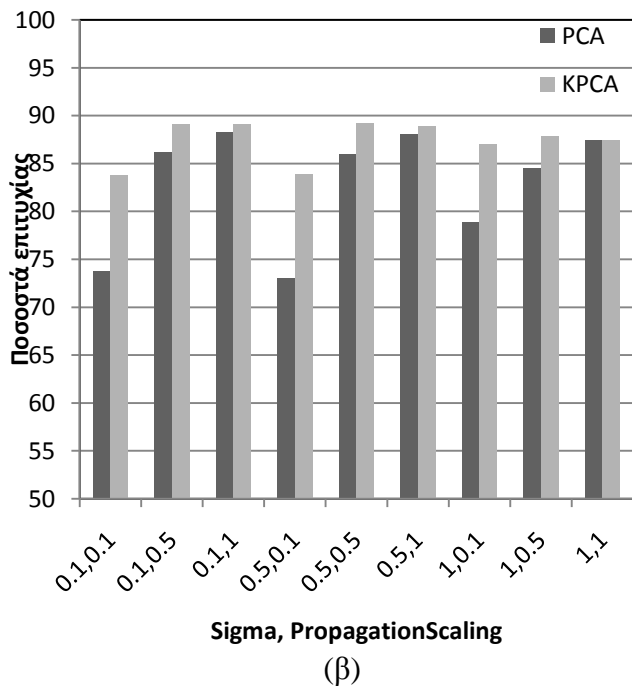
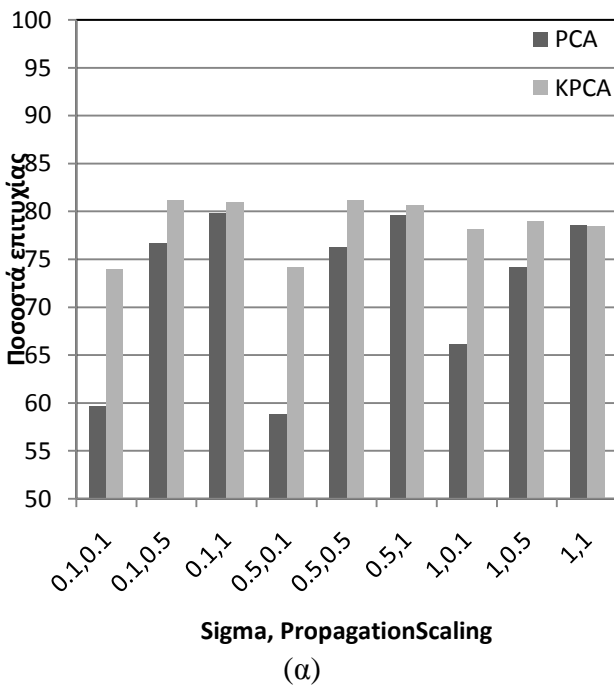
5.4 Αποτελέσματα

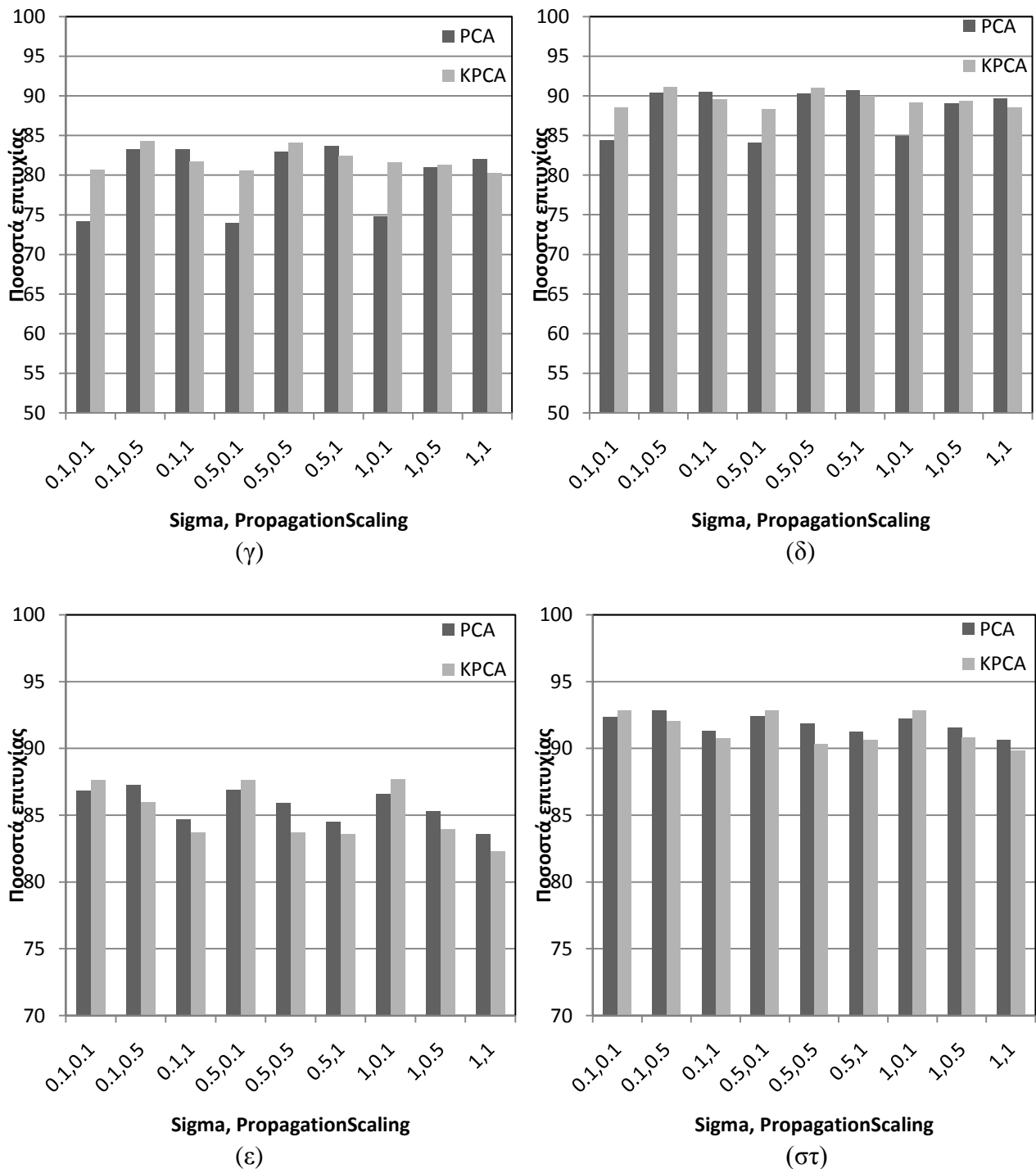
Το σύστημα στο οποίο εκτελέστηκαν τα πειράματα έχει τα ακόλουθα χαρακτηριστικά

- Λειτουργικό σύστημα: Linux, Ubuntu 14.04 (Trusty Tahr)
- Επεξεργαστής: AMD Quad Core A6-5200
- Ταχύτητα επεξεργαστή: 2.00 GHz
- Μνήμη RAM: 4 GB DDR3
- Σκληρός δίσκος: 250 GB

Στη συνέχεια παρουσιάζονται τα διαγράμματα αξιολόγησης για τις τμηματοποιήσεις που έγιναν σε κάθε πείραμα χρησιμοποιώντας τις τεχνικές της PCA και της Kernel PCA για την εξαγωγή του προγενέστερου σχήματος. Στις κατατμήσεις δοκιμάζονται συνολικά 27 συνδυασμοί για τις παραμέτρους τμηματοποίησης με κάθε παράμετρο να παίρνει τις παρακάτω τιμές:

- Sigma: 0.1, 0.5, 1
- PropagationScaling: 0.1, 0.5, 1
- ShapePriorScaling: 0.001, 0.005, 0.01





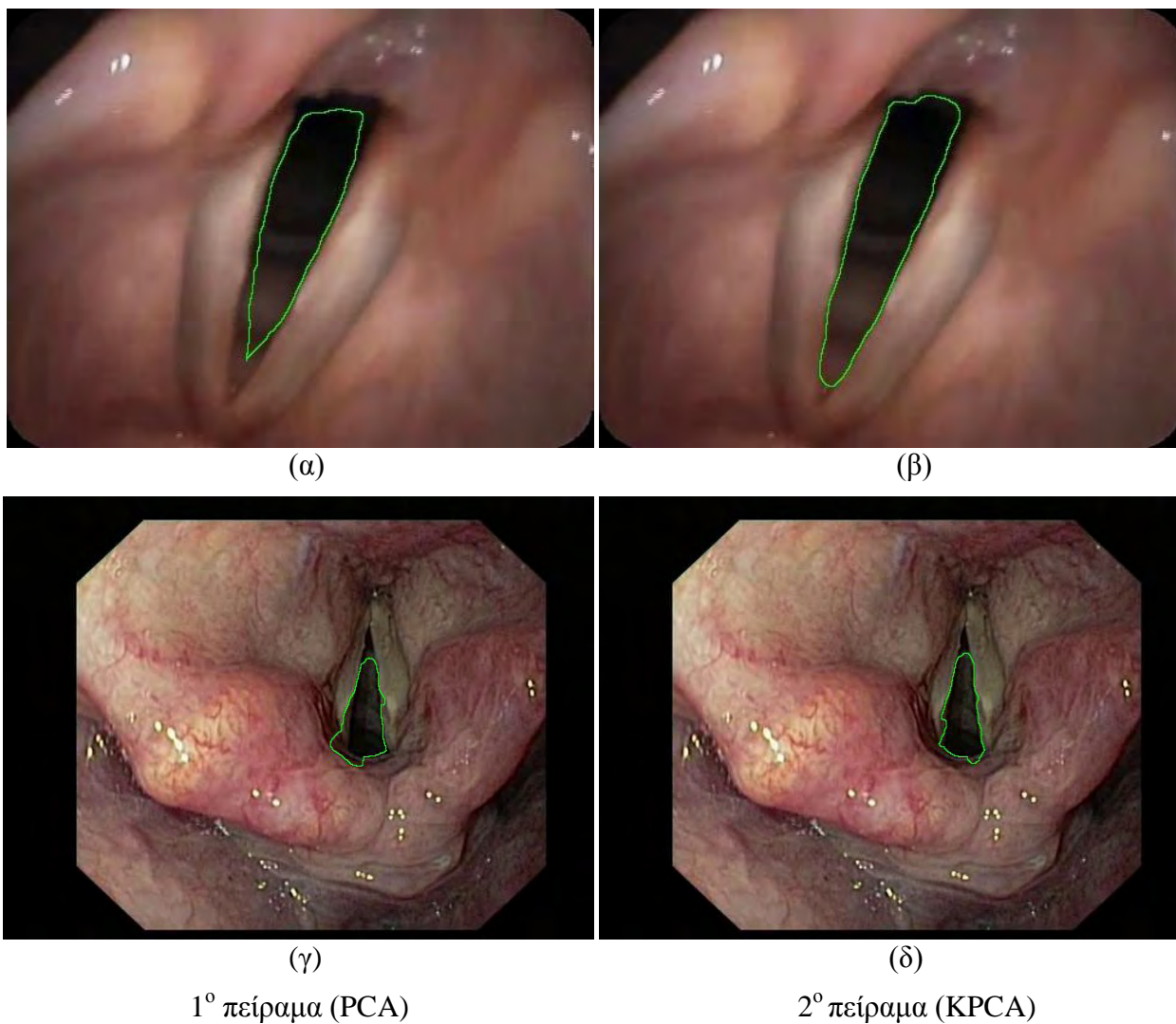
Εικόνα 5.1: Διαγράμματα αξιολόγησης τμηματοποίησης. Τα διαγράμματα (α),(γ),(ε) δείχνουν τα ποσοστά επιτυχίας με τον δείκτη Jaccard και τα (β),(δ),(στ) με τον δείκτη Dice ενώ στον άξονα x φαίνονται κατά σειρά οι τιμές των παραμέτρων Sigma και PropagationScaling. Στα δύο πρώτα διαγράμματα δίνεται η υψηλότερη τιμή στη παράμετρο για την εμπιστοσύνη προς το προγενέστερο σχήμα (ShapePriorScaling=0.01), στα δύο επόμενα η ενδιάμεση τιμή (ShapePriorScaling=0.005) και στα δύο τελευταία δίνεται η χαμηλότερη τιμή (ShapePriorScaling=0.001).

Δείκτης	Sigma	Propagation Scaling	ShapePrior Scaling	Shape prior	Μέσος όρος αξιολόγησης	Τυπική απόκλιση
jaccard	0.1	0.5	0.001	PCA	0.872632	0.1027
jaccard	1	0.1	0.001	KPCA	0.876701	0.1181
dice	0.1	0.5	0.001	PCA	0.928222	0.06900
dice	1	0.1	0.001	KPCA	0.928725	0.08960

Πίνακας 2: Οι συνδυασμοί των παραμέτρων τμηματοποίησης για τους οποίους επιτυγχάνονται οι καλύτερες καταταμήσεις.

Για κάθε συνδυασμό γίνεται τμηματοποίηση στις 50 εικόνες και οι δείκτες αξιολόγησης καταγράφουν το ποσοστό επιτυχίας για κάθε εικόνα. Έπειτα, σε κάθε συνδυασμό υπολογίζεται ο μέσος όρος και η τυπική απόκλιση για τα αποτελέσματα της αξιολόγησης. Στον Πίνακα 2 παρουσιάζονται οι συνδυασμοί εκείνοι που δίνουν τους υψηλότερους μέσους όρους και άρα θεωρούνται οι καλύτεροι για την διαδικασία της κατάτμησης. Στο [73] η τμηματοποίηση της γλωττίδας πραγματοποιείται χρησιμοποιώντας την PCA για την εξαγωγή των προγενέστερων σχημάτων. Στη παρούσα εργασία επιχειρούμε να βελτιώσουμε τη μέθοδο κατάτμησης στηριζόμενοι στη τεχνική της Kernel PCA για τη δημιουργία των προγενέστερων σχημάτων και το μοντέλο GAC για την εξέλιξη του ενεργού περιγράμματος.

Τα αποτελέσματα δείχνουν πως η τμηματοποίηση που εφαρμόζεται παρουσιάζει υψηλά ποσοστά επιτυχίας καθώς ο καλύτερος συνδυασμός δίνει για τον δείκτη Jaccard ένα μέσο όρο επιτυχίας 87,9% και για τον δείκτη Dice 92,9%. Επίσης βλέπουμε από τα διαγράμματα ότι τα αποτελέσματα που παίρνουμε με τις μεθόδους PCA και KPCA είναι παρόμοια και άρα καμία από τις δύο μεθόδους δεν υπερέρχει ξεκάθαρα έναντι της άλλης. Παρατηρούμε όμως ότι στις καταταμήσεις που γίνονται με προγενέστερα σχήματα που παράγει η KPCA τα αποτελέσματα παρουσιάζουν μικρότερη διακύμανση. Συνεπώς καταλήγουμε στο συμπέρασμα ότι η Kernel PCA παρουσιάζει μικρότερη ευαισθησία στις μεταβολές των παραμέτρων τμηματοποίησης συγκριτικά με την PCA.



Εικόνα 5.2: Παραδείγματα από τα πειράματα όπου φαίνεται η συμβολή της κάθε τεχνικής στη διαδικασία της τμηματοποίησης.

Συμπεράσματα και Μελλοντικές Κατευθύνσεις

6.1 Συμπεράσματα

Στη παρούσα πτυχιακή εργασία εξετάσαμε το θέμα της κατάτμησης γλωττίδας και με τη βοήθεια προηγμένων μεθόδων τμηματοποίησης, αναπτύξαμε μία εφαρμογή η οποία υπολογίζει το περίγραμμα της γλωττίδας σε εικόνες λάρυγγα. Στόχος της εργασίας είναι η εφαρμογή που υλοποιείται να λειτουργήσει υποβοηθητικά στην εξέταση του συστήματος φώνησης συμβάλλοντας στον εντοπισμό μη φυσιολογικών περιοχών στη περιοχή της γλωττίδας. Η εφαρμογή έχει τη δυνατότητα να εκτελέσει δύο διαδικασίες, την παραγωγή προγενέστερων σχημάτων για την γλωττίδα και την τμηματοποίηση εικόνων.

Η δημιουργία του προγενέστερου σχήματος βασίζεται σε τεχνικές στατιστικής ανάλυσης δεδομένων και για την εργασία επιλέγονται οι μέθοδοι της PCA και της Kernel PCA. Οι τεχνικές αυτές μπορούν να εντοπίζουν τις γραμμικές και μη γραμμικές συσχετίσεις μεταξύ των στοιχείων ενός συνόλου με αποτέλεσμα την εξαγωγή των σημαντικότερων χαρακτηριστικών και τη μείωση των διαστάσεων. Συνεπώς, κατασκευάζεται ένα στατιστικό μοντέλο σχήματος για την γλωττίδα το οποίο χρησιμοποιείται στη συνέχεια για την παραγωγή του προγενέστερου σχήματος.

Η τμηματοποίηση αποτελείται από δύο στάδια. Στο πρώτο στάδιο ο αλγόριθμος χρησιμοποιεί τη μέθοδο Otsu και σε συνδυασμό με τις ακμές της εικόνας δημιουργεί ένα αρχικό περίγραμμα για το αντικείμενο της γλωττίδας εντοπίζοντας παράλληλα τη θέση της μέσα στην εικόνα. Στο δεύτερο στάδιο ολοκληρώνεται η διαδικασία της τμηματοποίησης μέσα από τη χρήση του μοντέλου GAC και του προγενέστερου σχήματος. Πιο συγκεκριμένα η εξελισσόμενη καμπύλη αρχικοποιείται με το αρχικό περίγραμμα και στη συνέχεια επεκτείνεται με βάση δύο δυνάμεις. Τη δύναμη του GAC που επηρεάζεται από τις ακμές της εικόνας και τη δύναμη του προγενέστερου σχήματος που διαμορφώνει τη καμπύλη έχοντας σαν γνώμονα το στατιστικό μοντέλο σχήματος της γλωττίδας.

Η εφαρμογή διεκπεραιώνει ικανοποιητικά τη διαδικασία τμηματοποίησης σημειώνοντας ποσοστά επιτυχίας κοντά στο 90%. Τα αρχικά περιγράμματα που υπολογίζονται καλύπτουν μεγάλο τμήμα της γλωττίδας παρέχοντας με αυτόν τον τρόπο μία επιπλέον βοήθεια στη μέθοδο κατάτμησης. Επίσης με τη χρήση της τεχνικής Kernel PCA προσδίδουμε στην μέθοδο μεγαλύτερη ανθεκτικότητα όσον αφορά τη μεταβολή των παραμέτρων τμηματοποίησης. Ένα ακόμα συμπέρασμα που

μπορεί να εξαχθεί μελετώντας τη σχετική βιβλιογραφία είναι ότι οι εργασίες στη περιοχή της τμηματοποίησης εικόνων λάρυγγα είναι περιορισμένες και υπάρχει προοπτική για πρωτότυπες συνεισφορές.

6.2 Μελλοντική εργασία και προοπτικές

Η πτυχιακή εργασία μπορεί να χρησιμοποιηθεί ως βάση για την ανάπτυξη αλγορίθμων οι οποίοι θα επεκτείνουν τις δυνατότητες της παρούσας εφαρμογής και θα συμβάλουν στην αποδοτικότερη λειτουργία της. Σε αυτή τη λογική διακρίνουμε τις ακόλουθες μελλοντικές κατευθύνσεις έρευνας:

- ❖ Η μέθοδος που χρησιμοποιείται για να συνδυάσει το μοντέλο GAC με το προγενέστερο σχήμα είναι αυτό του Leventon et al. Η συγκεκριμένη μέθοδος έχει τη δυνατότητα να ρυθμίζει χαρακτηριστικά του προγενέστερου σχήματος όπως η εμφάνιση, ο προσανατολισμός και η θέση αλλά δεν μπορεί να ελέγξει το μέγεθος. Σε άλλες δοκιμές που έγιναν και χρησιμοποιήθηκαν αντικείμενα ενδιαφέροντος και προγενέστερα σχήματα με παρόμοιο μέγεθος, φάνηκε ότι το προγενέστερο σχήμα, έως ένα βαθμό, μπορεί να βοηθήσει στην διαδικασία της κατάτμησης καθώς συγκρατούσε την επέκταση του εξελισσόμενου περιγράμματος και δεν επέτρεπε να προχωρήσει εκτός του αντικειμένου. Συνεπώς μία άλλη μέθοδος όπως αυτή του Tsai et al. ή μία πιο σύγχρονη μέθοδος όπως αυτή των Paragios-Rousson, όπου θα υπάρχει η δυνατότητα ρύθμισης του μεγέθους του προγενέστερου σχήματος, ίσως να έδινε καλύτερα αποτελέσματα.
- ❖ Η μελέτη που έγινε στη παρούσα εργασία είναι προκαταρκτική και σε μελλοντικό χρόνο προβλέπεται η εξέλιξη της μεθόδου χρησιμοποιώντας μεγαλύτερα σύνολα δεδομένων έτσι ώστε να καλυφθούν ακόμα περισσότερες περιπτώσεις.
- ❖ Στη πλειοψηφία των περιπτώσεων η γλωττίδα εμφανίζεται με ένα ελλειψοειδές σχήμα. Αυτό το χαρακτηριστικό μπορούμε να το εκμεταλλευτούμε αξιοποιώντας εξισώσεις έλλειψης και με αυτόν τον τρόπο να εισάγουμε ακόμα μεγαλύτερους περιορισμούς στην εξελισσόμενη καμπύλη.
- ❖ Αντικείμενο μελλοντικής έρευνας θα μπορούσε να είναι η δυνατότητα επέκτασης της εργασίας στη παρακολούθηση της κίνησης της γλωττίδας χρησιμοποιώντας την εφαρμογή σε ένα βίντεο στροβοσκόπησης. Ένα υποθετικό σενάριο για αυτή τη περίπτωση είναι ο αλγόριθμος να αρχικοποιεί ένα περίγραμμα τμηματοποιώντας τη περιοχή της γλωττίδας στα πρώτα

στιγμιότυπα του βίντεο και στα επόμενα στιγμιότυπα με τη βοήθεια επιπλέον μεθόδων να προσαρμόζει το περίγραμμα ανάλογα με τη κίνηση της γλωττίδας.

- ❖ Θα είχε ενδιαφέρον να εξεταστεί αν ο συνδυασμός με κάποιο ακόμα χαρακτηριστικό της εικόνας θα μπορούσε να οδηγήσει στη πλήρη αυτοματοποίηση του υπολογισμού του αρχικού περιγράμματος ή της διαδικασίας τμηματοποίησης.

Παράρτημα Α

Εδώ παρουσιάζονται τα σημαντικότερες συναρτήσεις και αρχεία κώδικα της εφαρμογής. Όπως είδαμε στο 4^ο κεφάλαιο η εφαρμογή εκτελεί τη διαδικασία της παραγωγής προγενέστερου σχήματος και τη διαδικασία της τμηματοποίησης η οποία αποτελείται από δυο διεργασίες, τον υπολογισμό αρχικού περιγράμματος και τον υπολογισμό της τελικής καμπύλης τμηματοποίησης. Η πρώτη διαδικασία εκτελείται από το εκτελέσιμο αρχείο `train.exe` και η δεύτερη από το εκτελέσιμο `test.exe`.

Αλγόριθμος παραγωγής προγενέστερου σχήματος

Η κύρια συνάρτηση που ελέγχει τη διαδικασία για τη δημιουργία του προγενέστερου σχήματος είναι η **train**. Αρχικά με την **image_catalogue** καταγράφονται οι εικόνες που θα χρησιμοποιηθούν για τη κατασκευή του στατιστικού μοντέλου σχήματος. Η εφαρμογή υποθέτει εξ αρχής ότι υπάρχει ένας φάκελος ο οποίος περιέχει όλες τις εικόνες του συνόλου εκπαίδευσης και μέσα σε αυτόν οι εικόνες οργανώνονται σε ομάδες με υποφακέλους. Την πρώτη φορά που εκτελείται η εκπαίδευση ζητείται από τον χρήστη ο φάκελος που περιέχει τις εικόνες εκπαίδευσης και οι υποφάκελοι που θα χρησιμοποιηθούν για την διαδικασία. Το όνομα του φακέλου αποθηκεύεται και άρα την επομένη φορά ζητούνται μόνο οι υποφάκελοι με τους οποίους θα δημιουργηθεί το στατιστικό μοντέλο σχήματος. Στη συνέχεια εφαρμόζονται οι τεχνικές PCA και Kernel PCA.

Για την εκτέλεση της PCA καλείται το φίλτρο **the_pca_2**. Στο φίλτρο αυτό η συνάρτηση **itk::BinaryThresholdImageFilter** κατωφλιώνει κάθε εικόνα μετατρέποντάς την σε δυαδική και η **itk::SignedDanielssonDistanceMapImageFilter** μετασχηματίζει τη δυαδική εικόνα με τη συνάρτηση προσημασμένης απόστασης. Η **itk::ImagePCAShapeModelEstimator** είναι η βασική συνάρτηση του φίλτρου καθώς εκτελεί την PCA παράγοντας τη μέση εικόνα και τις εικόνες των κύριων συνιστωσών.

```
46  if(krnl=="pca"){
47      class4 cl4;
48      cl4.the_pca_2(array_trainingImage_names,array_trainingImage_names_length,3);
49  }
```

Η Kernel PCA εκτελείται μέσω του φίλτρου **the_kernel_pca_2**. Προτού όμως

εφαρμόσουμε την τεχνική, καλούμε το **DMImageRowWrite2CSV** με σκοπό να σχηματίσουμε τον πίνακα που θα περιέχει τα διανύσματα των εικόνων που θα χρησιμοποιήσουμε. Η συνάρτηση αυτή επεξεργάζεται τις εικόνες με τις **itk::BinaryThresholdImageFilter**, **itk::SignedDanielssonDistanceMapImageFilter** και στη συνέχεια τις μετατρέπει σε διανύσματα. Τέλος, αποθηκεύει τα διανύσματα με τη μορφή πίνακα στο αρχείο CSVFile.csv. Επίσης υπολογίζει και επιστρέφει τη διάμεσο από τις ευκλείδειες αποστάσεις μεταξύ των εικονοδιανυσμάτων καθώς με τη τιμή αυτή ορίζεται η μεταβλητή σ της συνάρτησης για τον γκαουσιανό πυρήνα.

```

50 else if(krnl=="kpc") {
51     class2 cl2;
52     double median=cl2.DMImageRowWrite2CSV(array_trainingImage_names,
53                                         size[1], size[0], array_trainingImage_names_length, CSVFileName);
54     cout << "\n\nmedian=" << median << endl;
55     class3 cl3;
56     cl3.the_kernel_pca_2();

```

Αφού υπολογίζουμε τις κύριες συνιστώσες πυρήνα, συνεχίζουμε με την εξαγωγή των εικόνων τους βασιζόμενοι σε κώδικα που έχει γραφεί στη γλώσσα προγραμματισμού MATLAB. Καλούμε λοιπόν την συνάρτηση **prepare_kpca** η οποία συλλέγει τα δεδομένα που απαιτούνται και μετά καλείται η **preimage** που κάνει τις ανακατασκευές των εικόνων ενώ η **v2im** αναλαμβάνει να μετατρέψει τα εικονοδιανύσματα σε εικόνες.

```

10 function out=prepare_kpca(CSVFile,eigenvectors,output,median,num_kpc,
11                          img_height, img_width)
12 CSVFile=load(CSVFile);
13 Eigenvectors=load(eigenvectors);
14 Output=load(output);
15
16 for i=1:num_kpc,
17     display(sprintf('\n\nkpc: %d', i-1));
18     z=preimage(Output(i,:), Eigenvectors(I,:), CSVFile', median, img_height,
19               img_width);
19     img_name1=strcat('kpc', num2str(i-1, '%d'), '.jpg');
20     img_name2=strcat('kpc', num2str(i-1, '%d'), '-bin.jpg');
21     r=v2im(z, 1, img_height, img_width, img_name1, img_name2);
22 end
23 end

```

Αλγόριθμος υπολογισμού αρχικού περιγράμματος

Η διεργασία για τον υπολογισμό του αρχικού περιγράμματος ελέγχεται από την **initial_contour** και οι σημαντικότερες μεταβλητές είναι οι παρακάτω:

half_imageName: το όνομα της εικόνας που επεξεργαζόμαστε
full_imageName: το πλήρες όνομα της εικόνας μαζί με την κατάληξη και τον φάκελο στον οποίο βρίσκεται (πχ “imageFolder1/imageName.png”)
image_color_selection: ορίζει τη συνιστώσα που θα εξάγει η **rgb_image** και η προεπιλεγμένη τιμή αντιστοιχεί στην ερυθρά συνιστώσα
outside_value: η τιμή που θα πάρουν οι περιοχές του παρασκηνίου μετά τη κατωφλίωση Otsu
inside_value: η τιμή που θα πάρουν οι περιοχές των αντικειμένων μετά τη κατωφλίωση Otsu
extra_edge_gap: αυξάνει κατά ένα μικρό αριθμό εικονοστοιχείων το πλάτος του πλαισίου που βρίσκεται στην **edges**
edge_sensitivity: χρησιμοποιείται από την **edges** και λειτουργεί σαν κατώφλι. Οι ακμές που υπολογίζονται και ξεπερνούν την **edge_sensitivity** είναι και αυτές που σχηματίζουν τα υποψήφια αντικείμενα για το αρχικό περίγραμμα

Πρώτα το φίλτρο **rgb_image** εξάγει την ερυθρά συνιστώσα από το μοντέλο RGB για την αρχική εικόνα.

```
61 //call rgb_image
62 class5 cl5;
63 tmp_name2=full_imageName;
64 if(image_color_selection==0)
65     tmp_name1=test_folder+half_imageName+string("_red.png");
66 else if(image_color_selection==1)
67     tmp_name1=test_folder+half_imageName+string("_green.png");
68 else if(image_color_selection==2)
69     tmp_name1=test_folder+half_imageName+string("_blue.png");
70 cl5.rgb_image(tmp_name2, tmp_name1, image_color_selection);
```

Το φίλτρο **Otsu** είναι το σημαντικότερο φίλτρο της διαδικασίας καθώς περιέχει τις περισσότερες συναρτήσεις για τον υπολογισμό του αρχικού περιγράμματος.

```
72 //call Otsu
73 class2 cl2;
74 tmp_name2=test_folder+string("otsu.png");
75 tmp_name3=test_folder+string("initial_contour.png");
```

```
76 cl2.Otsu(tmp_name1, tmp_name2, tmp_name3, outside_value, inside_value,
           test_folder, extra_edge_gap, edge_sensitivity);
```

Αρχικά με τη συνάρτηση **itk::OtsuThresholdImageFilter** κατωφλιώνεται η εικόνα με τη μέθοδο Otsu και εξάγονται τα αντικείμενα αυτής. Στο επόμενο βήμα καλείται το φίλτρο **edges** για να βρεθούν οι ακμές της αρχικής εικόνας και να δημιουργηθούν οι υποψήφιοι για το αρχικό περίγραμμα.

```
66 //call edges
67 class6 cl6;
68 cl6.edges(input_image_name, otsu_name, test_folder, extra_edge_gap,
           edge_sensitivity);
```

Το φίλτρο **Otsu** συνεχίζει με την καταγραφή των υποψήφιων αντικειμένων που έχουν δημιουργηθεί. Έπειτα με τη συνάρτηση **label_union** συγχωνεύει γειτονικά αντικείμενα και με τη συνάρτηση **metrisi_label** υπολογίζει το εμβαδόν κάθε αντικειμένου. Η συνάρτηση **label_detection** μεταθέτει στο παρασκήνιο τους υποψηφίους που εφάπτονται με τα άκρα της εικόνας, εντοπίζει τον υποψήφιο με το μεγαλύτερο εμβαδόν και τον επιλέγει για το αρχικό περίγραμμα. Η συνάρτηση **object_label_swap** χρωματίζει με μαύρο το παρασκήνιο και με άσπρο τη περιοχή του επιτυχόντος υποψηφίου. Η **clear_noise** ελέγχει για την εμφάνιση κενών εντός του αρχικού περιγράμματος και τα καλύπτει. Στο τελευταίο κομμάτι του κώδικα μεταφέρονται στο παρασκήνιο οι υποψήφιοι που δεν επιλέχθηκαν και αποθηκεύεται η εικόνα που δείχνει το αρχικό περίγραμμα.

Το φίλτρο **edges** διαβάζει την αρχική εικόνα, την εικόνα που δημιουργήθηκε με τη μέθοδο Otsu και υπολογίζει τις ακμές της αρχικής εικόνας χρησιμοποιώντας τον τελεστή sobel μέσω της συνάρτησης **itk::SobelEdgeDetectionImageFilter**. Ακολούθως ελέγχει αν υπάρχει πλαίσιο στην αρχική εικόνα και βρίσκει το πλάτος του πλαισίου σε κάθε άκρο. Τέλος, παράγει την εικόνα στην οποία αποτυπώνονται οι ακμές πάνω στην κατά Otsu κατωφλιωμένη εικόνα. Οι ακμές που χρησιμοποιούνται είναι αυτές που βρίσκονται στη περιοχή εντός του πλαισίου και θα πρέπει η ένταση φωτεινότητας να ξεπερνά το κατώφλι του `edge_sensitivity`. Στην εικόνα λοιπόν που επιστρέφει το **edges** σχηματίζονται τα υποψήφια αντικείμενα για το αρχικό περίγραμμα.

Το φίλτρο **perigramma** είναι υπεύθυνο για την προβολή μίας τμηματοποιημένης περιοχής πάνω σε μία άλλη εικόνα. Χρησιμοποιείται από τις δύο διεργασίες της κατάτμησης για την προβολή του αρχικού περιγράμματος και της τμηματοποιημένης γλωττίδας πάνω στην αρχική εικόνα.

```
83 //call perigramma
84 class4 cl4;
85 tmp_name2=test_folder+half_imageName+
           string("_initial_contour_on_image.png");
```

Θεωρούμε ότι η περιοχή ενδιαφέροντος της δυαδικής εικόνας αντιστοιχεί στην άσπρη περιοχή και το παρασκήνιο στο μαύρη περιοχή. Το φίλτρο εφαρμόζει τη συνάρτηση προσημασμένης απόστασης **itk::SignedDanielssonDistanceMapImageFilter** οπότε εντοπίζει τα σημεία που βρίσκονται στο περίγραμμα καθώς όσα σημεία βρίσκονται στο εσωτερικό του περιγράμματος αποκτούν αρνητική τιμή ενώ όσα βρίσκονται στο εξωτερικό αποκτούν τιμή μεγαλύτερη του μηδενός. Στο τελευταίο βήμα αποτυπώνει με πράσινο χρώμα το περίγραμμα πάνω στην αρχική εικόνα.

Αλγόριθμος υπολογισμού τελικής καμπύλης

Η διεργασία που υπολογίζει τη καμπύλη τμηματοποίησης για την γλωττίδα εκτελείται από τη συνάρτηση **segmentation** που βρίσκεται στο φίλτρο **GeodesicActiveContourShapePriorLevelSetImageFilter**. Οι παράμετροι του φίλτρου καθορίζονται στο **test** με τις παρακάτω μεταβλητές:

input_image: η αρχική εικόνα
output_image: το όνομα της εικόνας που περιέχει το αποτέλεσμα της κατάτμησης
sigma_value: ορίζει τη παράμετρο Sigma της **itk::GradientMagnitudeRecursiveGaussianImageFilter**
propagation_Scaling_value: ορίζει τη παράμετρο **propagationScaling** της **geodesicActiveContour** και δείχνει τη ταχύτητα εξέλιξης της καμπύλης
shapePriorScaling_value: ορίζει τη παράμετρο **ShapePriorScaling** της **geodesicActiveContour** και δείχνει το βαθμό εμπιστοσύνης στο προγενέστερο σχήμα
mean_image: η μέση εικόνα
num_kpca: το πλήθος των κύριων συνιστωσών που χρησιμοποιούνται
kpca_image: η εικόνα της πρώτης κύριας συνιστώσας ή της πρώτης κύριας συνιστώσας πυρήνα
iteration_value: το εύρος των επαναλήψεων που εκτελεί η **geodesicActiveContour** για να βρει την επανάληψη όπου σημειώνεται η καλύτερη τμηματοποίηση
initial_contour_image: η εικόνα του αρχικού περιγράμματος
min_Iter: δείχνει τον αριθμό της επανάληψης όπου σημειώθηκε η καλύτερη τμηματοποίηση
test_folder: ο φάκελος ο οποίος περιέχει την εικόνα που επεξεργαζόμαστε

Η σημαντικότερη συνάρτηση εδώ είναι η **itk::GeodesicActiveContourShapePriorLevelSetImageFilter** καθώς υλοποιεί το μοντέλο GAC και συνδέει όλα τα επί μέρους τμήματα για να διεκπεραιώσει τη

κατάτμηση. Στη συνέχεια αναφερόμαστε στη συνάρτηση με το όνομα **geodesicActiveContour**.

```

100 segmentation_class seg;

161 //call segmentation
162 input_image=test_folder+input_image;
163 output_image=test_folder+output_image;
164 initial_contour_image=test_folder+initial_contour_image;
165
166 min_Iter=seg.segmentation(input_image, output_image, sigma_value,
    propagationScaling_value, shapePriorScaling_value, mean_image,
    num_kpca, kpca_image, iteration_value, initial_contour_image,
    test_folder);
167 seg.segmentation(input_image, output_image, sigma_value,
    propagationScaling_value, shapePriorScaling_value, mean_image,
    num_kpca, kpca_image, min_Iter, initial_contour_image, test_folder);
168 cout << "min_Iter: " << min_Iter << endl;

```

Αρχικά το φίλτρο προσδιορίζει το σημείο όπου θα τοποθετηθεί το προγενέστερο σχήμα. Η **center_shape** χρησιμοποιείται για να βρεθεί το κέντρο του σχήματος για το αρχικό περίγραμμα και το κέντρο του σχήματος της μέσης εικόνας οπότε στη πρώτη περίπτωση το κέντρο χρησιμεύει ως σημείο αναφοράς στην εικόνα και στη δεύτερη περίπτωση το κέντρο λειτουργεί ως σημείο αναφοράς για το προγενέστερο σχήμα. Στη συνέχεια με την **DM_mean_shape** εφαρμόζεται η συνάρτηση προσημασμένης απόστασης με σκοπό το σημείο που παίρνει τη χαμηλότερη τιμή να οριστεί ως το κέντρο του σχήματος. Αφού λοιπόν βρεθούν οι συντεταγμένες των κέντρων, γίνονται οι κατάλληλοι υπολογισμοί ώστε να αποφασιστεί η μετατόπιση που πρέπει να υποστεί το προγενέστερο σχήμα (Κατά την προβολή του προγενέστερου σχήματος πάνω στην αρχική εικόνα, το σημείο που βρίσκεται στην πάνω αριστερή γωνία του προγενέστερου σχήματος τοποθετείται στο κέντρο της αρχικής εικόνας).

```

115 //topothesis prior shape
116 string initial_contour=initial_contour_image;
117 int center_meanShape_array[4], center_object_array[4];
118
119 center_shape(neab_image,0);
120 for(int i=0; i<4; i++)
121     center_meanShape_array[i]=center_array[i];
122
123 center_shape(initial_contour, 255);
124 for(int i=0; i<4; i++)

```

```

125   center_object_array[i]=center_array[i];
126
127   int startX = (center_object_array[2]/2) + center_meanShape_array[0] -
           center_object_array[0];
128   int startY = (center_object_array[3]/2) + center_meanShape_array[1] -
           center_object_array[1];

```

Ακολουθούν οι συναρτήσεις που παράγουν την εικόνα χαρακτηριστικών. Η **itk::CurvatureAnisotropicDiffusionImageFilter** εξομαλύνει την εικόνα και η **itk::GradientMagnitudeRecursiveGaussianImageFilter** με την **itk::BoundedReciprocalImageFilter** υπολογίζουν τις ακμές. Με την τεχνική της διασωλήνωσης λοιπόν συνδυάζονται οι παραπάνω συναρτήσεις παράγοντας την εικόνα χαρακτηριστικών η οποία μέσω της **SetImageFeature** εισάγεται στο μοντέλο της **geodesicActiveContour**. Η εικόνα χαρακτηριστικών δείχνει το μέγεθος των ακμών της αρχικής εικόνας ενώ η παράμετρος **Sigma** της **itk::GradientMagnitudeRecursiveGaussianImageFilter** επηρεάζει την ένταση με την οποία αποτυπώνονται οι ακμές.

Η συνάρτηση **itk::FastMarchingImageFilter** ενσωματώνει στο μοντέλο τη θεωρία των επιπεδοσυνόλων χρησιμοποιώντας τη τεχνική της ταχείας προέλασης για την επέκταση του μετώπου της καμπύλης και εισάγεται στο **geodesicActiveContour** μέσω της **SetInput**. Το μηδενικό επιπεδοσύνολο με το οποίο ξεκινάει η κατάτμηση ορίζεται βάσει της εικόνας του αρχικού περιγράμματος ενώ η ταχύτητα εξέλιξης καθορίζεται από την παράμετρο **propagationScaling** που εισάγεται στη **geodesicActiveContour** με τη συνάρτηση **SetPropagationScaling**.

Έπειτα υλοποιούνται οι συναρτήσεις που σχετίζονται με το στατιστικό μοντέλο σχήματος και η βασική συνάρτηση που διαχειρίζεται το στατιστικό μοντέλο είναι η **itk::PCAShapeSignedDistanceFunction**. Με τη συνάρτηση **SetMeanImage** εισάγεται στο μοντέλο η μέση εικόνα και με την **SetPrincipalComponentImages** εισάγονται οι εικόνες των κύριων συνιστωσών και των κύριων συνιστωσών πυρήνα. Η **itk::Euler2DTransform** ρυθμίζει τη μετατόπιση και τον προσανατολισμό του προγενέστερου σχήματος και με την **SetTransform** εισάγεται στην **itk::PCAShapeSignedDistanceFunction**. Οι συναρτήσεις **itk::ShapePriorMAPCostFunction** και **itk::OnePlusOneEvolutionaryOptimizer** συμβάλλουν στην βελτιστοποίηση της τροποποίησης του προγενέστερου σχήματος και παρέχονται ως είσοδοι στη **geodesicActiveContour**. Η **SetWeights** εισάγει στη πρώτη συνάρτηση τα βάρη με τα οποία συμμετέχουν οι τέσσερις όροι πιθανοτήτων της εξίσωσης (3.21) στη διαδικασία βελτιστοποίησης του προγενέστερου σχήματος. Οι δύο συναρτήσεις εισάγονται στη **geodesicActiveContour** μέσω των **SetCostFunction** και **SetOptimizer** αντίστοιχα, για την εισαγωγή του στατιστικού μοντέλου χρησιμοποιείται η **SetShapeFunction** και για την παράμετρο **shapePriorScaling** η **SetShapePriorScaling**.

Τέλος, η **GetRMSChange** που ανήκει στη **geodesicActiveContour** συγκρίνει σε κάθε επανάληψη τις δύο τελευταίες καμπύλες τμηματοποίησης που έχουν υπολογιστεί επιτρέποντας με αυτόν τον τρόπο να προσδιορίσουμε την επανάληψη στην οποία επιτυγχάνεται η καλύτερη τμηματοποίηση. Οι αξιολογήσεις των αποτελεσμάτων γίνονται στο φίλτρο **evaluation** χρησιμοποιώντας την συνάρτηση **itk::LabelOverlapMeasuresImageFilter**.

BIBΛΙΟΓΡΑΦΙΑ

- [1] J. Van Den Berg. Myoelastic-aerodynamic theory of voice production. *J. Speech Hear Res*, 1: 227-44, 1958.
- [2] I. R. Titze. *The Myoelastic Aerodynamic Theory of Phonation*. National Center for Voice and Speech, 2006
- [3] S. Z. Karakozoglou, N. Henrich, C. d'Alessandro, Y. Stylianou. Automatic glottal segmentation using local-based active contours and application to glottovibrography. *Speech Communication*, 2011.
- [4] S. Lankton, A. Tannenbaum. Localizing Region – Based Active Contours. *IEEE Trans Image Process*, 17(11): 2029–2039, 2008.
- [5] C. Palm, T. M. Lehmann, J. Bredno, C. Neuschaefer-Rube, S. Klajman, K. Spitzer. Automated Analysis of Stroboscopic Image Sequences by Vibration Profiles. *Proceedings of the 5th International Workshop on Advances in Quantitative Laryngology, Voice and Speech Research*. Groningen, Netherlands, 2001.
- [6] B. Marendic, N. Galatsanos, D. Bless. A New Active Contour Algorithm for Tracking Vibrating Vocal Folds. *Image Processing, International Conference on Vol. 1 IEEE*, 2001.
- [7] M. Kass, A. Witkin, D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 321-331, 1988.
- [8] J. J. Cerrolaza, V. Osma, N. Sáenz, A. Villanueva, J. M. Gutiérrez, J. I. Godino, R. Cabeza. Full-Automatic Glottis Segmentation with Active Shape Models. *Models and analysis of vocal emissions for biomedical applications : 7th international workshop*, 2011.
- [9] V. Osma-Ruiz, J. I. Godino-Llorente, N. Saenz-Lechon, R. Fraile. Segmentation of the glottal space from laryngeal images using the watershed transform. *Computerized Medical Imaging and Graphics* 32 193–201, 2008.
- [10] Y. Yan , X. Chen, D. Bless. Automatic Tracing of Vocal-Fold Motion From High-Speed Digital Images. *IEEE Transactions on Biomedical Engineering*, VOL. 53, NO. 7, 2006.

-
- [11] K. Pearson. On lines and Planes of Closest fit to Systems of Points in Space. University College, London, 1901.
- [12] P. Comon. Independent component analysis, A new concept?. *Signal Processing* 36, 287-314, 1994.
- [13] R.A. Fisher. *The Use of Multiple Measurements in Taxonomic Problems*, 1936.
- [14] B. Scholkopf, A. Smola, K.-R. Muller. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, vol 10, 1998.
- [15] G. H. Bakir, J. Weston, B. Scholkopf. Learning to Find Pre-Images. *Advances in Neural Information Processing Systems*, 2004.
- [16] S. Mika, B. Scholkopf, A. Smola, K.R. Muller, M. Scholz, G. Ratsch. Kernel PCA and De-Noising in Feature Spaces. In *Advances in Neural Information Processing Systems* 11, 1999.
- [17] J.T. Kwok, I.W. Tsang. *The Pre-Image Problem in Kernel Methods*. Hong Kong University of Science and Technology, 2004.
- [18] L.D. Cohen. On Active Contour Models and Balloons. *CVGIP Image Understanding* Vol. 53, No. 2, March pp. 211-218, 1991.
- [19] C. Xu, J.L. Prince. Gradient Vector Flow: A New External Force for Snakes. *IEEE Proc. Conf. on Comp. Vis. Patt. Recog*, 1997.
- [20] T. McInerney, D. Terzopoulos. T-snakes: Topology adaptive snakes. *Medical Image Analysis*, 2000.
- [21] J. Cheng, Y. Liu, R. Jia, W. Guo. A New Active Contour Model for Medical Image Analysis – Wavelet Vector Flow, 2006
- [22] V. Caselles, R. Kimmel, G. Sapiro. Geodesic Active Contours. *International Journal of Computer Vision* 22(1),61-79, 1997.
- [23] S.J. Osher, J.A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton- Jacobi formulations. *Journal of Computational Physics*, 79:12– 49, 1988.
- [24] M.E. Leventon, W.E.L. Grimson, O. Faugeras. Statistical Shape Influence in Geodesic Active Contours. *Computer Vision and Pattern Recognition* vol. 1, 316-323, 2000.

- [25] A. Tsai, A. Yezzi Jr., W. Wells, C. Tempany, D. Tucker, A. Fan, W.E. Grimson, A. Willsky. A Shape-Based Approach to the Segmentation of Medical Imagery Using Level Sets. *IEEE Transactions on Medical Imaging*, Vol 22, No2, 2003.
- [26] M. Rousson, N. Paragios. *Shape Priors for Level Set Representations*, 2002.
- [27] A. Dervieux and F. Thomasset. A finite element method for the simulation of Rayleigh-Taylor instability. *Lecture Notes Math.*, vol. 771, pp. 145–158, 1980.
- [28] A. Dervieux and F. Thomasset. Multifluid incompressible flows by a finite element method. *Lecture Notes Phys.*, vol. 141, pp. 158–163, 1980.
- [29] P. Jaccard. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37, 241-272, 1901.
- [30] L. R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology* 26 (3): 297–302, 1945.
- [31] T. Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* 5 (4): 1–34, 1948.
- [32] N. Otsu. A Threshold Selection Method from Gray-Level Histogram. *IEEE Trans. Systems Man, and Cybernetics*, Vol. 9, pp. 62-66, 1979.
- [33] P.K. Sahoo, S. Soltani, K.C. Wong, Y.C. Chen. A Survey of Thresholding Techniques. *Computer Vision, Graphics, and Image Processing*, Vol. 41, pp.233-260, 1988.
- [34] R. Adams, L. Bischof. Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intel.*, 16, 641-647, 1994.
- [35] A. Bleau, L.J. Leon. Watershed-based segmentation and region merging. *Computer Vision and Image Understanding*, 77(3):317-370, 2000.
- [36] J. Shi, J. Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, pp 888-905, Vol. 22, No. 8, 2000.
- [37] L. Grady. Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1768–1783, Vol. 28, No. 1, 2006.

-
- [38] Z. Wu, R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1101–1113, Vol. 15, No. 1, 1993.
- [39] L. Grady, Eric L. Schwartz. Isoperimetric Graph Partitioning for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 469–475, Vol. 28, No. 3, 2006.
- [40] C. T. Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers*, pp. 68–86, Vol. 20, No. 1, 1971.
- [41] J. R. Quinlan. Simplifying decision trees. *International journal of man-machine studies*, 27 (3): 221-234, 1987.
- [42] W. McCulloch, W. Pitts. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics* 5 (4): 115–133, 1943.
- [43] C. Cortes, V. Vapnik. Support-vector networks. *Machine Learning* 20 (3): 273, 1995.
- [44] J. Pearl. *Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning*, 1985.
- [45] Ballard, Dana H., and Christopher M. Brown. *Computer Vision* Prentice-Hall. Englewood Cliffs, NJ: 70-72, 1982.
- [46] J. McCarthy. *Programs with common sense*. Defense Technical Information Center, 1963.
- [47] R. Jain, R. Kasturi, B. G. Schunck. *Machine vision*. Vol. 5. New York: McGraw-Hill, 1995.
- [48] J.R. Anderson, R.S. Michalski, J.G. Carbonell, T.M. Mitchell. *Machine learning: An artificial intelligence approach*. Morgan Kaufmann, 1986.
- [49] S. Wold. Pattern recognition by means of disjoint principal components models. *Pattern recognition*, 8.3 (1976): 127-139, 1976.
- [50] R. Malladi, J.A. Sethian , B.C. Vemuri. Shape modeling with front propagation: a level set approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 0162-8828 17, 158–175, 1995.

- [51] T. Chan, L. Vese. Active contours without edges. IEEE Trans. Image Process. 1057-7149 10, 266–277, 2001.
- [52] T.M. Mitchell. Machine learning. Burr Ridge, IL: McGraw Hill 45, 1997.
- [53] <http://www.iatronet.gr/iatriko-lexiko/laryggas.html>
- [54] www.edoctoronline.com
- [55] Ιωάννα Ψύλλα. Μοντελοποίηση των φωνητικών χορδών. Διπλωματική Εργασία, 2013.
- [56] www.iatrikionline.gr
- [57] www.iacl.ece.jhu.edu
- [58] en.wikipedia.org
- [59] www.itk.org
- [60] <http://www.youtube.com/watch?v=v9Wdf-RwLcs>
- [61] <http://www.kolumbus.fi/hans/gastrolab/vid128.htm>
- [62] <http://www.youtube.com/watch?v=bowRggzZ2q4>
- [63] <http://www.youtube.com/watch?v=ljNyf8O32HU>
- [64] <http://www.youtube.com/watch?v=iYpDwhpILkQ>
- [65] <http://www.kitware.com/source/home/post/52>
- [66] <http://www.mlpack.org/>
- [67] <http://www.cmake.org/>
- [68] <http://www.eclipse.org/>
- [69] <http://arma.sourceforge.net/>
- [70] <http://www.mathworks.com/>
- [71] <http://www.ecse.rpi.edu/sites/sampl/quant/>

- [72] <http://www.hygeia.gr>
- [73] Y. Yan, G. Du, G. Marriott. Snake based automatic tracing of vocal-fold motion from high-speed digital images. Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on IEEE, 2012.

