



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

Οπτική Ανάλυση Βίντεο Νοηματικής Γλώσσας : Κατάτμηση,  
Παρακολούθηση και Εξαγωγή Χαρακτηριστικών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Όλγας Ν. Διαμαντή

Επιβλέπων: Πέτρος Α. Μαραγκός  
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2007





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

Οπτική Ανάλυση Βίντεο Νοηματικής Γλώσσας : Κατάτμηση,  
Παρακολούθηση και Εξαγωγή Χαρακτηριστικών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Όλγας Ν. Διαμαντή

Επιβλέπων: Πέτρος Α. Μαραγκός  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 24 Οκτωβρίου 2007.

.....  
Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

.....  
Κώστας Τζαφέστας  
Λέκτορας Ε.Μ.Π.

.....  
Στέφανος Κόλλιας  
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2007.



.....  
Όλγα Ν. Διαμαντή

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

Copyright © Όλγα Ν. Διαμαντή, 2007.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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



*Στους γονείς μου*

*Success is the ability to go from one failure to another with no loss of enthusiasm.*  
*Sir Winston Churchill*





# Περίληψη

Η παρούσα διπλωματική εργασία επικεντρώνεται στην υλοποίηση, επέκταση και ανάπτυξη μεθόδων για την κατασκευή ενός συστήματος ανάλυσης βίντεο της νοηματικής γλώσσας, προς χρήση σε εφαρμογή αναγνώρισης. Η εφαρμογή αυτή εμπίπτει στο ερευνητικό πεδίο της Όρασης Υπολογιστών και ειδικότερα στον κλάδο της αλληλεπίδρασης ανθρώπου - μηχανής. Πιο συγκεκριμένα, στην εργασία εξετάζονται όλα τα στάδια που προηγούνται της διαδικασίας της αναγνώρισης, και συγκεκριμένα τα στάδια της κατάκτησης των εικόνων, της παρακολούθησης της κίνησης των χεριών του νοηματιστή, και της εξαγωγής χαρακτηριστικών. Ιδιαίτερη έμφαση δίνεται στην αντιμετώπιση πρακτικών προβλημάτων που συχνά απαντώνται σε εφαρμογές αναγνώρισης νοηματικής γλώσσας, όπως είναι το πρόβλημα των επικαλύψεων (occlusions). Δοκιμάζεται ένα ευρύ σύνολο μεθόδων, που προέκυψε μετά από εκτενή αναζήτηση στη βιβλιογραφία.

Η εργασία είναι δομημένη σε πέντε κεφάλαια, κάθε ένα από τα οποία ασχολείται με ένα ξεχωριστό υποπρόβλημα. Το συνολικό σύστημα πρόκειται να χρησιμοποιηθεί για την κατασκευή πραγματικού και λειτουργικού συστήματος αναγνώρισης της ελληνικής νοηματικής γλώσσας, σε συνδυασμό βέβαια με ένα στάδιο αναγνώρισης, στα πλαίσια του ερευνητικού προγράμματος «ΔΙΑΝΟΗΜΑ».

## Λέξεις - κλειδιά

όραση υπολογιστών, αναγνώριση νοηματικής γλώσσας, αλληλεπίδραση ανθρώπου - μηχανής, παρακολούθηση χειρονομιών, ανάλυση βίντεο, γεωδαιτικά ενεργά περιγράμματα, εξαγωγή χαρακτηριστικών, παρακολούθηση κίνησης.



# Abstract

The goal of this thesis is to test, extend and develop methods for the implementation of a sign language recognition (SLR) system based on the visual analysis of sign language videos. This application is part of the more general research area referred to as «human - machine interaction», which is strongly related to the science and technology of Computer Vision. More specifically, this thesis focuses on all the stages that are required prior to the recognition phase of a SLR system, namely the image segmentation, motion tracking and feature extraction stages. Special effort was made to cope with one of the most commonly met problems in SLR, which is the vagueness introduced in cases of hand - hand or hand - face occlusions. In each stage, a variety of methods and techniques is explored, which were retrieved following an extensive search of the state-of-the-art bibliography.

The thesis is organised in five chapters, each of which focuses on a separate subproblem of the SLR system. The overall system is bound to be used (in combination with the necessary recognition stage of course) in the construction of a real practical SLR system for the recognition of a set of Greek Sign Language signs, as part of the research project «DIANOEMA».

## Keywords

computer vision, sign language recognition, human- machine interaction, gesture tracking, video analysis, geodesic active contours, feature extraction, motion tracking.



# Ευχαριστίες

Θα ήθελα κατ' αρχήν να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Πέτρο Μαραγκό, οι διαλέξεις του οποίου στα πλαίσια του μαθήματος της 'Όρασης Υπολογιστών και η γενικότερη αντιμετώπισή του προς τους φοιτητές του απετέλεσαν το βασικότερο ίσως κίνητρο για να ασχοληθώ με την ερευνητική περιοχή της 'Όρασης. Η καθοδήγηση, η υποστήριξη και η βοήθειά του υπήρξαν καθοριστικές όχι μόνο για την εκπόνηση της παρούσας εργασίας αλλά και για τη διατήρηση του ενδιαφέροντός μου στις περιπτώσεις εκείνες που τα αποτελέσματα των προσπαθειών μου δεν ήταν τα αναμενόμενα. Τον ευχαριστώ ιδιαίτερα για την εμπιστοσύνη που έδειξε προς εμένα από την αρχή της δραστηριοποίησής μου στο ερευνητικό πεδίο της εργασίας αυτής.

Επίσης θα ήθελα να ευχαριστήσω τους υποψηφίους (και μη) διδάκτορες του εργαστηρίου της 'Όρασης Υπολογιστών για την συμπαράσταση και τη βοήθειά τους σε όσες περιπτώσεις τη χρειάστηκα, και για την ευκαιρία που μου έδωσαν να έρθω σε επαφή όχι μόνο με την επιστημονική, αλλά και με την πιο ομαδική και ευχάριστη πλευρά της έρευνας. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα Γιώργο Παπανδρέου για την πολύτιμη βοήθειά του με την παροχή πολύ σημαντικών τμημάτων του κώδικα που χρησιμοποιήθηκε στην παρούσα εργασία. Χωρίς τη συμβολή του η υλοποίηση μεγάλου τμήματος της διπλωματικής θα ήταν ιδιαίτερα δύσκολη.

Τέλος, δε θα μπορούσα να μην αναφερθώ στην οικογένεια και τους φίλους μου, που με πολλαπλούς τρόπους με ενίσχυσαν και με διευκόλυναν στην πορεία μου τα τελευταία 5 χρόνια που υπήρξα φοιτήτρια στο Ε.Μ.Π. Τους ευχαριστώ που έκαναν την καθημερινότητά μου πιο εύκολη και ευχάριστη.



# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>21</b>
1.1	Όραση Υπολογιστών και Επικοινωνία Ανθρώπου - Μηχανής . . . . .	21
1.1.1	Γενικά για την Όραση Υπολογιστών . . . . .	21
1.1.2	Επικοινωνία Ανθρώπου- Μηχανής . . . . .	25
1.1.3	Επικοινωνία Ανθρώπου- Μηχανής με χρήση Χειρονομιών . . . . .	28
1.2	Αναγνώριση Νοηματικής Γλώσσας . . . . .	29
1.2.1	Κατάτμηση και Παρακολούθηση των Χεριών . . . . .	30
1.2.2	Εξαγωγή Χαρακτηριστικών . . . . .	31
1.3	Σύνοψη της παρούσας Εργασίας . . . . .	32
<b>2</b>	<b>Κατάτμηση της Εικόνας</b>	<b>35</b>
2.1	Γενικές Μέθοδοι Κατάτμησης . . . . .	35
2.2	Μέθοδος Γεωδαιτικών Ενεργών Περιγραμμάτων . . . . .	38
2.2.1	Ενεργά Περιγράμματα . . . . .	38
2.2.2	Γεωδαιτικά Ενεργά Περιγράμματα . . . . .	40
2.2.3	Γεωδαιτικά Ενεργά Περιγράμματα και Επιπεδοσύνολα . . . . .	42
2.2.4	Επεκτάσεις του μοντέλου των ενεργών περιγραμμάτων . . . . .	45
2.3	Τεχνικές και Αποτελέσματα Κατάτμησης σε Βίντεο Νοηματικής Γλώσσας . . . . .	53
2.3.1	Κατάτμηση με χρήση δύναμης χρώματος και GACs . . . . .	53
2.4	Εικόνες Αποτελεσμάτων . . . . .	58
<b>3</b>	<b>Περιγραφητές Σχήματος</b>	<b>61</b>
3.1	Boundary - based Μέθοδοι . . . . .	62
3.1.1	Περιγραφητές Fourier . . . . .	63
3.1.2	Χώρος Κλίμακας Καμπυλότητας . . . . .	66
3.1.3	Το Cepstrum της Καμπυλότητας . . . . .	73

3.2	Region - based Μέθοδοι . . . . .	75
3.2.1	Κεντρικές Ροπές . . . . .	75
3.2.2	Χαρακτηριστικά Σχήματος που βασίζονται στις Κεντρικές Ροπές και άλλα Γεωμετρικά Χαρακτηριστικά . . . . .	77
3.3	Εξαγωγή Στατικών Χαρακτηριστικών Σχήματος απο Χειρομορφές . .	79
3.3.1	Region-based Χαρακτηριστικά . . . . .	79
3.3.2	Boundary-based Χαρακτηριστικά . . . . .	79
3.4	Εικόνες Αποτελεσμάτων . . . . .	82
<b>4</b>	<b>Εκτίμηση Οπτικής Κίνησης</b>	<b>91</b>
4.1	Πεδίο Κίνησης . . . . .	92
4.2	Οπτική Ροή . . . . .	94
4.2.1	Μέθοδος των Lucas - Kanade . . . . .	95
4.2.2	Μέθοδος των Horn - Schunck . . . . .	98
4.3	Block Matching . . . . .	100
4.4	Παρακολούθηση Κίνησης: Οπτική Ροή και GACs . . . . .	103
4.5	Εκτίμηση Κίνησης και Αναγνώριση Νοηματικής Γλώσσας . . . . .	109
4.5.1	Πειραματικά Αποτελέσματα και Εξαγωγή Δυναμικών Χαρακτηριστικών Κίνησης . . . . .	110
4.6	Εικόνες Αποτελεσμάτων . . . . .	111
<b>5</b>	<b>Αντιμετώπιση Επικαλύψεων</b>	<b>115</b>
5.1	Σύμπτωση προτύπων . . . . .	116
5.2	Σύμπτωση προτύπων σε συνδυασμό με πληροφορία κίνησης . . . . .	118
5.2.1	Απλή πρόβλεψη κίνησης . . . . .	119
5.2.2	Απομόνωση της μη κινούμενης περιοχής . . . . .	120
5.3	Χρήση της πληροφορίας πρότερου σχήματος . . . . .	122
5.4	«Πεδία δυνάμεων» και «Πηγάδια δυναμικού» . . . . .	123
5.4.1	Κατασκευή του πεδίου δυνάμεων . . . . .	124
5.4.2	Εξαγωγή των καναλιών και πηγαδιών δυναμικού . . . . .	125
5.4.3	Εφαρμογή στην αντιμετώπιση των επικαλύψεων . . . . .	126
5.5	Εικόνες Αποτελεσμάτων . . . . .	128
<b>6</b>	<b>Ανακεφαλαίωση και Μελλοντικές Κατευθύνσεις Έρευνας</b>	<b>141</b>
6.1	Ανασκόπηση της Διπλωματικής Εργασίας . . . . .	141
6.2	Πιθανές Μελλοντικές Επεκτάσεις . . . . .	142



# Λίστα Εικόνων

2.1	Μέθοδος επιπεδοσυνόλων . . . . .	43
2.2	Το μοντέλο χρώματος . . . . .	56
2.3	Ιστογράμματα τιμών των συνιστωσών a και b . . . . .	56
2.4	Αποτέλεσμα εφαρμογής του μοντέλου χρώματος . . . . .	59
2.5	Αποτελέσματα της κατάτμησης . . . . .	60
3.1	Κατασκευή του κώδικα αλυσίδας. Από [29]. . . . .	67
3.2	Υπολογισμός της Καμπυλότητας για ένα απλό σχήμα . . . . .	69
3.3	Η εξέλιξη της καμπύλης ενός σχήματος . . . . .	71
3.4	Ο χώρος κλίμακας καμπυλότητας για την εικόνα του σχ. 3.3 . . . . .	72
3.5	Αποτελέσματα της ανακατασκευής για το Σχ. 3.2 . . . . .	74
3.6	Τα region-based χαρακτηριστικά σχήματος . . . . .	77
3.7	Τρία αποτελέσματα της εξαγωγής της καμπυλότητας . . . . .	84
3.8	Ανακατασκευή της καμπυλότητας από τους 61 πρώτους συντελεστές cepstrum - Χειρομορφή 1 . . . . .	85
3.9	Ανακατασκευή της καμπυλότητας από τους 61 πρώτους συντελεστές cepstrum - Χειρομορφή 2 . . . . .	86
3.10	Ανακατασκευή της καμπυλότητας από τους 61 πρώτους συντελεστές cepstrum - Χειρομορφή 3 . . . . .	87
3.11	Γραμμικοί Χώροι Κλίμακας για τρεις διαφορετικές Χειρομορφές . . . . .	88
3.12	Περιγραφητές Fourier - Μέρος I . . . . .	89
3.13	Περιγραφητές Fourier - Μέρος II . . . . .	90
4.1	Το μοντέλο στενής οπής για μία κάμερα . . . . .	92
4.2	Αλγόριθμοι ταχείας αναζήτησης για την τεχνική block matching. Από [4] . . . . .	104
4.3	Δυναμικά χαρακτηριστικά κίνησης για μια ακολουθία δύο εικόνων . . . . .	111

4.4	Optical Flow vs. Block Matching - Ακολουθία νοήματος «πόρτα» . . .	112
4.5	Optical Flow vs. Block Matching - Ακολουθία νοήματος «κοιτώ προς τα αριστερά» . . . . .	113
4.6	Δυναμικά Χαρακτηριστικά Κίνησης MBD για την ακολουθία νοήματος «κοιτώ προς τα αριστερά» . . . . .	114
5.1	Παραδείγματα επικαλύψεων . . . . .	116
5.2	Εικόνα διαφοράς μεταξύ δύο διαδοχικών frames της ακολουθίας «Πάνω»	121
5.3	Εξαγωγή του πεδίου δυνάμεων. Από [21]. . . . .	125
5.4	Απλό παράδειγμα εξαγωγή των καναλιών και πηγαδιών δυναμικού, από [42] . . . . .	126
5.5	Η διαδικασία της σύμπτωσης προτύπων . . . . .	129
5.6	Αποτελέσματα της απλής σύμπτωσης προτύπων για 5 διαδοχικά frames	130
5.7	Σφάλμα στην παρακολούθηση των κέντρων μάζας στο νεύμα «Βασιλιάς»	130
5.8	Απλή πρόβλεψη κίνησης στο νεύμα «Βασιλιάς» . . . . .	131
5.9	Απλή πρόβλεψη κίνησης στο νεύμα «Βιβλίο» . . . . .	131
5.10	Η οπτική ροή ως πιθανότητα κίνησης. Ακολουθία «Χτίζω» . . . . .	132
5.11	Απομόνωση του μη κινούμενου μέρους της εικόνας με χρήση οπτικής ροής και GAC . . . . .	132
5.12	Αποτελέσματα της Χρήσης Πρότερου Σχήματος - Μέρος I . . . . .	133
5.13	Αποτελέσματα της Χρήσης Πρότερου Σχήματος - Μέρος II . . . . .	134
5.14	Αποτελέσματα της Χρήσης Πρότερου Σχήματος - Μέρος III . . . . .	135
5.15	Αποτελέσματα της Χρήσης Πρότερου Σχήματος - Μέρος IV . . . . .	136
5.16	Αποτυχία της μεθόδου [42] για τρεις τυπικές περιπτώσεις επικαλύψεων	137
5.17	Χρήση του πεδίου δύναμης και των καναλιών δυναμικού για τον εντοπισμό της θέσης των χεριών σε περιπτώσεις επικάλυψης - Μέρος I . . .	138
5.18	Χρήση του πεδίου δύναμης και των καναλιών δυναμικού για τον εντοπισμό της θέσης των χεριών σε περιπτώσεις επικάλυψης - Μέρος II . .	139
5.19	Χρήση του πεδίου δύναμης και των καναλιών δυναμικού για τον εντοπισμό της θέσης των χεριών σε περιπτώσεις επικάλυψης - Μέρος III . .	140

# Λίστα Πινάκων

3.1 Τα region-based χαρακτηριστικά για 3 διαφορετικές χειρομορφές . . .	83
---	----



---

# ΕΙΣΑΓΩΓΗ

---

Στην παρούσα εργασία θα ασχοληθούμε με τρεις διαφορετικούς κλάδους της Όρασης Υπολογιστών, και συγκεκριμένα με την κατάτμηση εικόνων (image segmentation), την εξαγωγή χαρακτηριστικών (feature extraction) και την παρακολούθηση της κίνησης (motion tracking). Η σχετική έρευνα πραγματοποιήθηκε στα πλαίσια ανάπτυξης ενός συστήματος αναγνώρισης της Ελληνικής Νοηματικής Γλώσσας (ΕΝΓ), βασισμένου στην ανάλυση ακολουθιών βίντεο.

Στο κεφάλαιο αυτό θα πραγματοποιήσουμε μια σύντομη εισαγωγή στον επιστημονικό κλάδο της Όρασης Υπολογιστών, και ειδικότερα στις εφαρμογές που σχετίζονται με την αλληλεπίδραση ανθρώπου-μηχανής. Επίσης θα αναφερθούν οι πιο σύγχρονες προσεγγίσεις στο πεδίο της αναγνώρισης νοηματικής γλώσσας, καθώς και στα επιμέρους προβλήματα αυτής. Τέλος, θα παρουσιαστεί εν συντομία ο βασικός άξονας και το περιεχόμενο της παρούσας εργασίας.

## 1.1 Όραση Υπολογιστών και Επικοινωνία Ανθρώπου - Μηχανής

### 1.1.1 Γενικά για την Όραση Υπολογιστών

Με τον όρο «Όραση Υπολογιστών» αναφερόμαστε σε εκείνο τον τομέα της επιστήμης και της τεχνολογίας που ασχολείται με τη θεωρητική μελέτη και το σχεδιασμό μηχανών και συστημάτων που να μπορούν να «βλέπουν», δηλαδή να αξιοποιούν

οπτική πληροφορία από εικόνες ή ακολουθίες εικόνων (βίντεο) προκειμένου να λάβουν κάποιες αποφάσεις. Παραδείγματα εφαρμογών της Όρασης Υπολογιστών περιλαμβάνουν συστήματα για:

1. Τον έλεγχο διαδικασιών (π.χ. σε ένα βιομηχανικό ρομπότ ή ένα αυτόνομο όχημα).
2. Την ανίχνευση συμβάντων (π.χ. σε ένα σύστημα οπτικής παρακολούθησης).
3. Την οργάνωση πληροφοριών (π.χ. για την ταξινόμηση βάσεων δεδομένων από εικόνες ή ακολουθίες εικόνων).
4. Τη μοντελοποίηση αντικειμένων και περιβάλλοντος εργασίας (π.χ. σε εφαρμογές βιομηχανικού ελέγχου, ανάλυσης ιατρικών εικόνων ή τοπογραφική μοντελοποίηση)
5. Αλληλεπίδραση συστημάτων (π.χ. σε εφαρμογές επικοινωνίας ανθρώπου - μηχανής.)

Επίσης η Όραση Υπολογιστών μπορεί να περιγραφεί ως ένα συμπλήρωμα (όχι ως αντίθετο) της βιολογικής όρασης. Στην τελευταία, μελετώνται η οπτική αντίληψη των ανθρώπων και διάφορων ζώων, με στόχο την ανάπτυξη μοντέλων που περιγράφουν τις φυσιολογικές διεργασίες που ελέγχουν αυτά τα συστήματα όρασης. Από την άλλη πλευρά, η Όραση Υπολογιστών μελετά και περιγράφει τεχνητά συστήματα όρασης που υλοποιούνται σε λογισμικό και/ή υλικό. Η ανταλλαγή εμπειρίας και γνώσης μεταξύ της Όρασης Υπολογιστών και της μελέτης της βιολογικής όρασης έχει αποδειχθεί ιδιαίτερα χρήσιμα και για τα δύο πεδία έρευνας.

Η Όραση Υπολογιστών εμφανίστηκε ως επιστημονικό πεδίο την δεκαετία του 1960 από προσπάθειες ερευνητών της Τεχνητής Νοημοσύνης να δημιουργήσουν υπολογιστές με τη δυνατότητα όρασης. Τα επιστημονικά πεδία που συνέβαλαν στη δημιουργία και ανάπτυξη της Όρασης Υπολογιστών είναι η Τεχνητή Νοημοσύνη, η Επεξεργασία Εικόνας και η Αναγνώριση Προτύπων. Η Τεχνητή Νοημοσύνη υποβοήθησε την ανάπτυξη της Όρασης Υπολογιστών με την παροχή αλγορίθμων, όπως οι αλγόριθμοι αναζήτησης, που τροποποιήθηκαν κατάλληλα ώστε να εφαρμοστούν σε εφαρμογές Όρασης. Η Επεξεργασία και Ανάλυση Εικόνας ασχολείται με τη μελέτη διδιάστατων μετασχηματισμών, δηλαδή τελεστών που εφαρμόζονται σε κάθε εικονοστοιχείο και χρησιμεύουν στην βελτίωση της αντίθεσης της εικόνας, την εξαγωγή ακμών ή την αφαίρεση θορύβου, ή ακόμα και στο γεωμετρικό μετασχηματισμό της εικόνας

(μετατόπιση, περιστροφή). Ωστόσο οι διαδικασίες αυτές πραγματοποιούνται χωρίς να γίνονται υποθέσεις ή ερμηνείες για το περιεχόμενο της εικόνας, και αφορούν σε διδιάστατες εικόνες, σε αντίθεση με την Ύραση Υπολογιστών. Τέλος, η Αναγνώριση Προτύπων χρησιμοποιεί διάφορες μεθόδους προκειμένου να εξάγει πληροφορία από σήματα κάθε μορφής, με βάση κυρίως στατιστικές προσεγγίσεις. Ένα μεγάλο μέρος αυτού του πεδίου αφιερώνεται στην εφαρμογή αυτών των μεθόδων σε δεδομένα από εικόνες. Επίσης εξίσου σημαντική είναι προφανώς η συμβολή των Μαθηματικών και του Αυτομάτου Ελέγχου που παρέχουν την απαραίτητη θεωρία.

Οι βασικότερες κατευθύνσεις της Ύρασης Υπολογιστών μπορούν να συνοψιστούν στα παρακάτω:

1. Τεχνητή (Μηχανική) Ύραση: Εφοδιασμός ρομπότ, μηχανών ή αισθητήρων με συστήματα που να προσομοιάζουν τα συστήματα βιολογικής όρασης
2. Βιολογική Ύραση: Ανάπτυξη μοντέλων για την περιγραφή των συστημάτων βιολογικής όρασης με στόχο την αποτελεσματικότερη υλοποίηση της μηχανικής όρασης
3. Μοντελοποίηση Εγκεφάλου: Κατανόηση της φυσιολογικής λειτουργίας της όρασης, θα οδηγήσει σε καλύτερη κατανόηση της βιολογικής λειτουργίας του εγκεφάλου και της αντίληψης γενικότερα.

### **Προβλήματα με τα οποία ασχολείται η Ύραση Υπολογιστών**

**Αναγνώριση :** Το κλασικό πρόβλημα της Ύρασης Υπολογιστών, της Επεξεργασίας Εικόνας και της μηχανικής όρασης είναι ο προσδιορισμός του κατά πόσον μια εικόνα περιέχει κάποιο συγκεκριμένο αντικείμενο, χαρακτηριστικό ή δραστηριότητα. Το πρόβλημα αυτό μπορεί να επιλυθεί αυτόματα και με ευρωστία σε συγκεκριμένες υποπεριπτώσεις αυτού, και όχι στη γενική περίπτωση αντικειμένων τυχαίου σχήματος σε τυχαίο περιβάλλον. Το πρόβλημα της αναγνώρισης μπορεί να αφορά είτε στην αναγνώριση και εντοπισμό γνωστών αντικειμένων ή κατηγοριών αντικειμένων, είτε στην ταυτοποίηση κάποιου συγκεκριμένου αντικειμένου (όπως για παράδειγμα στην αναγνώριση ενός συγκεκριμένου προσώπου ή δακτυλικού αποτυπώματος), ή ακόμα και στην ανίχνευση κάποια συγκεκριμένης μορφής ενός αντικειμένου (π.χ. ανίχνευση ενός προβληματικού κυττάρου σε δείγμα ιστού ή ενός αυτοκινήτου σε ένα αυτόματο σύστημα διοδίων). Ανάμεσα στις εφαρμογές της αναγνώρισης περιλαμβάνονται η ανάκτηση εικόνων βάσει του περιεχομένου (π.χ. η επιλογή όλων των εικόνων από μια

βάση δεδομένων που έχουν ένα συγκεκριμένο περιεχόμενο), η εκτίμηση της θέσης, σε σχέση με την κάμερα, ενός συγκεκριμένου αντικειμένου (π.χ. η επιλογή αντικειμένων από ένα ρομπότ), ή η οπτική αναγνώριση χαρακτήρων από ένα χειρόγραφο ή τυπωμένο κείμενο.

**Κίνηση :** Αρκετές εφαρμογές της Όρασης Υπολογιστών περιλαμβάνουν την εκτίμηση κίνησης. Στις περιπτώσεις αυτές μια ακολουθία εικόνων υφίσταται κατάλληλη επεξεργασία προκειμένου να παραχθεί μια εκτίμηση της ταχύτητας κίνησης είτε σε κάθε σημείο της εικόνας είτε στην τρισδιάστατη σκηνή. Παραδείγματα τέτοιων εφαρμογών είναι η εκτίμηση της τρισδιάστατης κίνησης της κάμερας (egomotion) ή η παρακολούθηση (tracking) των κίνησης κάποιων αντικειμένων (π.χ. αυτοκινήτων ή ανθρώπων).

**Ανακατασκευή Σκηνής :** Δεδομένων δύο ή περισσότερων εικόνων από μια σκηνή, ή ενός βίντεο, η αναπαράσταση σκηνής στοχεύει στον υπολογισμό ενός τρισδιάστατου μοντέλου της σκηνής. Στην απλούστερη περίπτωση το μοντέλο μπορεί να είναι ένα σύνολο τρισδιάστατων σημείων. Πιο εξελιγμένες μέθοδοι παράγουν ένα ολοκληρωμένο τρισδιάστατο μοντέλο επιφάνειας.

**Ανάκτηση Εικόνας και Αποθορυβοποίηση :** Ο στόχος της ανάκτησης εικόνων είναι η αφαίρεση του θορύβου (λόγω της κίνησης ή σφαλμάτων των αισθητήρων). Η πιο απλή μέθοδος αποθορυβοποίησης είναι η εφαρμογή διαφόρων τύπων φίλτρων, ενώ οι πιο εξελιγμένες μέθοδοι χρησιμοποιούν κάποιο μοντέλο για τα διάφορα αντικείμενα της εικόνας, έτσι ώστε να ξεχωρίσουν από το θόρυβο.

Η οργάνωση ενός συστήματος όρασης υπολογιστών εξαρτάται σημαντικά από την εκάστοτε εφαρμογή. Ωστόσο υπάρχουν κάποιες βασικές λειτουργίες που συναντώνται πρακτικά σε όλα τα συστήματα μηχανικής όρασης, όπως η λήψη της διδιάστατης ή τρισδιάστατης εικόνας, η προεπεξεργασία, όπως για παράδειγμα η αποθορυβοποίηση με χρήση κάποιου φίλτρου, η βελτίωση της αντίθεσης ή η επαναδειγματοληψία, η εξαγωγή χαρακτηριστικών, κατά την οποία λαμβάνονται χαρακτηριστικά στοιχεία από την εικόνα σε διάφορα επίπεδα πολυπλοκότητας (π.χ. ακμές, γωνίες, γραμμές ή και ολόκληρες περιοχές, ή ακόμα και πιο αφηρημένες έννοιες όπως η υφή ή το χρώμα), η ανίχνευση/κατάτμηση, κατά την οποία επιλέγονται εκείνα τα σημεία ή περιοχές της εικόνας που είναι κατάλληλα προς περαιτέρω επεξεργασία, και τέλος η επεξεργασία υψηλότερου επιπέδου, κατά την οποία ένα μικρό σύνολο από δεδομένα, όπως ένα σύνολο σημείων ή μια περιοχή που θεωρείται ότι περικλείει ένα αντικείμενο ενδιαφέροντος χρησιμοποιούνται για την επαλήθευση υποθέσεων ή μοντέλων, για την εκτίμηση παραμέτρων όπως η θέση ή το μέγεθος του αντικειμένου ή η ταξινόμηση



ενός ανιχνευθέντος αντικειμένου σε κάποια κατηγορία.

Η ΎΡαση Υπολογιστών βρίςκει εφαρμογές σε πληθώρα πεδίων. Αναφέρουμε ενδεικτικά ορισμένα παραδείγματα:

- Επεξεργασία πληροφορίας σε εικόνες
- Μοντελοποίηση εγκεφάλου και τεχνητή νοημοσύνη
- Αυτοματισμός και ρομποτική
- Βιοϊατρική έρευνα και τεχνολογία, ιατρική απεικόνιση
- Περιβάλλον και γεωεπιστήμες
- Δορυφορική όραση και τηλεπισκόπηση
- Ευφυή συστήματα ψηφιοποίησης και ανάλυσης βλαβών
- Ευφυείς αισθητήρες και ψηφιακές κάμερες
- Τεχνολογία βίντεο
- Τεχνολογία διαδικτύου και αυτόματη αναζήτηση σε βάσεις δεδομένων εικόνων
- Επικοινωνία ανθρώπου - υπολογιστή
- Ψηφιακές εικαστικές τέχνες

### 1.1.2 Επικοινωνία Ανθρώπου- Μηχανής

Ο όρος «επικοινωνία ανθρώπου - μηχανής» (EAM), που αρκετές φορές χρησιμοποιείται ως ταυτόσημος του όρου «επικοινωνία ανθρώπου - υπολογιστή» αναφέρεται στη μελέτη της αλληλεπίδρασης μεταξύ ανθρώπων (χρηστών) και υπολογιστών. Πρόκειται για συνεργασία πολλών επιστημονικών πεδίων, ένα από τα κυριότερα μπορεί να είναι και η ΎΡαση Υπολογιστών. Η αλληλεπίδραση υπολογιστή - χρήστη πραγματοποιείται στη διεπιφάνεια υπολογιστή - χρήστη (user interface), που περιλαμβάνει τόσο υλικό (hardware) όσο και λογισμικό (software), παραδείγματος χάριν στις περιπτώσεις περιφερειακών υπολογιστή και μηχανικών συστημάτων μεγάλης κλίμακας, όπως αεροσκάφη ή συστήματα ηλεκτρικής ενέργειας.

Ένας ορισμός της αλληλεπίδρασης ανθρώπου - υπολογιστή θα μπορούσε να είναι ο εξής: ένα επιστημονικό πεδίο που ασχολείται με το σχεδιασμό, την αξιολόγηση και

την υλοποίηση διαδραστικών υπολογιστικών συστημάτων για ανθρώπινη χρήση, σε συνδυασμό με τη μελέτη των σημαντικών φαινομένων που σχετίζονται με αυτά.

### Στόχοι

Ένας βασικός στόχος της EAM είναι να βελτιώσει την αλληλεπίδραση μεταξύ χρηστών και υπολογιστών κάνοντας τους υπολογιστές πιο εύχρηστους και ικανότερους να αποκριθούν στις ανάγκες των χρηστών. Συγκεκριμένα, η EAM ασχολείται με

- μεθοδολογίες και διαδικασίες για τη σχεδίαση διεπαφών (π.χ., δεδομένου ενός καθήκοντος που πρέπει να επιτελεστεί, και μιας ομάδας χρηστών, ζητείται η σχεδίαση της βέλτιστης διεπαφής, εντός βέβαια κάποιων περιορισμών, που να εμφανίζει βελτιστότητα ως προς κάποια επιθυμητή ιδιότητα όπως η ικανότητα μάθησης ή η ευκολία και αποδοτικότητα χρήσης)
- μεθόδους για την υλοποίηση των παραπάνω διεπαφών (π.χ. πακέτα λογισμικού και βιβλιοθήκες ή αποδοτικοί αλγόριθμοι)
- τεχνικές για την αξιολόγηση, τον έλεγχο και τη σύγκριση των διεπαφών
- την ανάπτυξη νέων τρόπων αλληλεπίδρασης
- την ανάπτυξη σχετικής θεωρίας και κατάλληλων μοντέλων για την περιγραφή και την πρόβλεψη της αλληλεπίδρασης

Ένας μακροπρόθεσμος στόχος της EAM είναι η σχεδίαση συστημάτων που ελαχιστοποιούν το χάσμα μεταξύ του τρόπου που ένας άνθρωπος αντιλαμβάνεται τους στόχους που θέλει να επιτελέσει, και την «αντίληψη» του υπολογιστή σχετικά με το καθήκον που του αναθέτει ο χρήστης.

### Χαρακτηριστικά Συστήματος EAM και Πραγματοποίηση Επικοινωνίας

Για τη σχεδίαση ενός διαδραστικού συστήματος μπορεί να χρειαστούν γνώσεις από πολλά επιστημονικά πεδία, μεταξύ των οποίων: επιστήμη υπολογιστών, τεχνολογία πληροφοριών, αισθητική, ανθρωπολογία, τεχνητή νοημοσύνη, γνωστική επιστήμη (cognitive science), όραση υπολογιστών, σχέδιο (design), εργονομία, επιστήμη της πληροφορίας, φιλοσοφία, φαινομενολογία, ψυχολογία, κοινωνική ψυχολογία, κοινωνιολογία, λεκτική - γλωσσική παθολογία, ψυχοφυσική. Ωστόσο, ανεξάρτητα από την

εκάςτοτε εφαρμογή, κοινό χαρακτηριστικό όλων των συςτημάτων αυτών είναι ότι χρησιμοποιούνται από κάποιους χρήςτες προκειμένου να επιτελέςουν κάποια εργασία. Έτσι, το σύςτημα είναι απαραίητο να χαρακτηρίζεται από λειτουργικότητα, η οποία συνοψίζεται στα εξής:

- ευκολία στην εκμάθηση της λειτουργίας του
- ευκολία στην απομνημόνευση του τρόπου λειτουργίας του
- αποδοτικότητα
- ασφάλεια κατά τη χρήση
- δυνατότητα να διατηρεί το ενδιαφέρον του χρήςτη

Προφανώς, ένα σύςτημα ΕΑΜ πρέπει να είναι σε θέση να πραγματοποιεί αμφίδρομη επικοινωνία μεταξύ του χρήςτη και του υπολογιςτή. Έτσι, εκτός από το υλικό, που μετατρέπει τα εκάςτοτε σήματα στην κατάλληλη μορφή προς επεξεργασία από το χρήςτη ή το μηχάνημα, και το λογισμικό, που εξασφαλίζει την ορθή λειτουργία του υλικού, πολύ σημαντική παράμετρος για το σύςτημα ΕΑΜ είναι και η μεθοδολογία επικοινωνίας που ακολουθείται. Ο δημοφιλέςτερος τρόπος επικοινωνίας σήμερα βασίζεται στις απλές μηχανικές συςκευές, όπως το πληκτρολόγιο και το ποντίκι. Ωςτόςσο οι συςκευές αυτές αναγκαςτικά περιορίζουν την ταχύτητα και τη φυςικότητα της αλληλεπίδρασης με τον υπολογιςτή, με αποτέλεσμα πολλοί άνθρωποι να βρίςκουν την αλληλεπίδραση αυτή κουραστική ή και δυσάρεςτη. Για το λόγο αυτό τα τελευταία χρόνια η ΕΑΜ εξελίςσεται προς την σχεδίαση συςτημάτων που δεν απαιτούν μηχανική επαφή, αλλά χρησιμοποιούν αντιληπτικές ή/και πολυτροπικές διεπαφές με το χρήςτη, που βασίζονται στα «φυςικά» μέσα με τα οποία οι άνθρωποι επικοινωνούν μεταξύ τους. Η αναγνώριση φωνής, πολλές φορές σε συνδυαςμό με οπτική πληροφορία από εικόνες, η καθοδήγηση ρομποτικών συςτημάτων μέσω λεκτικών εντολών ή χειρονομιών, και η αναγνώριση νοηματικής γλώςσας αποτελούν παραδείγματα τέτοιων προςπαθειών. Η υλοποίηση αυτού του τύπου της επικοινωνίας θα διευκολύνει ιδιαίτερα ακόμα και συγκεκριμένες ομάδες πληθυσμού (όπως οι τυφλοί ή οι κωφάλαλοι) να χρησιμοποιήςουν υπολογιςτές, κάτι που προηγουμένων αποτελούσε μια ιδιαίτερα δύςκολη εργασία. Επίςσης, θα μειώσει σημαντικά τον απαιτούμενο χρόνο εκπαίδευσης των χρηςτών προκειμένου να μπορέςουν να πραγματοποιήςουν συγκεκριμένες εργασίες στον υπολογιςτή (αφού ο έλεγχος της λειτουργίας του θα είναι πλέον πολύ πιο

φυσικός), ενώ θα εξαλειφθεί και η ανάγκη για χρήση πολλών συσκευών επικοινωνίας με τον υπολογιστή.

Στην παρούσα εργασία θα ασχοληθούμε με ένα ζήτημα που αποτελεί σαφώς εφαρμογή της ΕΑΜ, και συγκεκριμένα την αναγνώριση χειρονομιών. Πρόκειται για ένα πεδίο όπου η όραση υπολογιστών μπορεί να παρέξει πολύ σημαντική βοήθεια, καθώς επιτρέπει τη σχεδίαση νέων τρόπων αλληλεπίδρασης διεπαφών χρήστη - υπολογιστή, με αξιοποίηση της κίνησης του ανθρώπινου σώματος.

### 1.1.3 Επικοινωνία Ανθρώπου- Μηχανής με χρήση Χειρονομιών

Η έρευνα στη συγκεκριμένη κατεύθυνση προσανατολίζεται στην χρήση χειρονομιών για ΕΑΜ. Το σήμα που λαμβάνει ο υπολογιστής μπορεί είτε να προέρχεται από κάποια εξειδικευμένη συσκευή, όπως κάποιο ειδικό γάντι (cyber-glove), είτε να είναι οπτικό (με χρήση κάμερας) . Στην πρώτη περίπτωση οι δυναμικοί και/ή στατικοί σχηματισμοί του χεριού ή του βραχίονα του ομιλητή λαμβάνονται άμεσα, μετρώντας κατάλληλες γωνίες και θέσεις των αρθρώσεων και των δαχτύλων. Ωστόσο η χρήση γαντιών αναγκαστικά περιορίζει την ευκολία και τη φυσικότητα με την οποία ο χρήστης μπορεί να αλληλεπιδράσει με το ελεγχόμενο από τον υπολογιστή περιβάλλον του. Οι δυσκολίες αυτές δεν υφίστανται στη δεύτερη περίπτωση, που δεν απαιτεί καμιά φυσική επαφή μεταξύ του χρήστη και της μηχανής. Στην προσέγγιση αυτή χρησιμοποιείται ένα σύνολο από κάμερες σε συνδυασμό με τεχνικές της υπολογιστικής όρασης για την ερμηνεία των χειρονομιών.

Στην οπτική αναπαράσταση των χειρονομιών για ΕΑΜ έχουν πραγματοποιηθεί πολλές προσεγγίσεις, οι οποίες συνήθως εστιάζουν είτε στην παρακολούθηση των χεριών (hand tracking), είτε στην εκτίμηση χειρομορφών(hand posture estimation), ή στην ταξινόμηση της μορφής του χεριού σε μια από γνωστές κατηγορίες (hand posture classification). Τα συστήματα υπολογιστικής όρασης που αναλαμβάνουν την αναγνώριση χειρονομιών μπορούν να ταξινομηθούν σε αυτά που βασίζονται σε τρισδιάστατα μοντέλα, και σε αυτά που βασίζονται απλά στην διδιάστατη εμφάνιση των χεριών. Στην πρώτη κατηγορία ανήκουν συστήματα που προσπαθούν να προσαρμόσουν ένα τρισδιάστατο μοντέλο του χεριού στην εικόνα, παρέχοντας μια πλήρη εκτίμηση της διάταξης του χεριού. Τα συστήματα της δεύτερης κατηγορίας στηρίζονται στην εξαγωγή χαρακτηριστικών που περιγράφουν τη διδιάστατη προβολή του χεριού στο επίπεδο της εικόνας, και εφαρμόζονται εφ' όσον δε μας ενδιαφέρει η ακριβής τοποθέτηση του χεριού, αλλά απλά η εκτίμηση της χειρομορφής.

Ένα τέτοιο σύστημα θα μας απασχολήσει και στην παρούσα εργασία. Τα συστήματα αυτά γενικά αποτελούνται από τα εξής τέσσερα στάδια: απόκτηση της εικόνας από κάποια κάμερα - κατάτμηση - εξαγωγή χαρακτηριστικών - ταξινόμηση/αναγνώριση.

Ενδεικτικά, ορισμένες εφαρμογές όπου η χρήση χειρονομιών μπορεί να αποδειχθεί πιο πρακτική από τις παραδοσιακές συσκευές εισόδου είναι οι εξής:

- Αναγνώριση νοηματικής γλώσσας
- Τηλεχειρισμός (οικιακών κυρίως) ρομπότ με χειρονομίες
- Τηλεχειρισμός κάθε είδους κινητής συσκευής όπου η οικονομία χώρου είναι επιθυμητή ή κρίσιμη, καθώς αποφεύγονται περιττές συσκευές εισόδου (όπως για παράδειγμα σε κάποια ιατρικά περιβάλλοντα)
- Χειρισμός κάμερας κατά τη διάρκεια τηλεδιασκέψεων μέσω κάποιας ειδικής χειρονομίας (όπως π.χ. γίνεται σε μια αίθουσα διδασκαλίας), ή ανάλογος χειρισμός ενός μηχανήματος παρουσιάσεων
- Υποβοήθηση της διαδικασίας αναγνώρισης φωνής για πολυτροπική (multimodal) επικοινωνία
- Εικονική Πραγματικότητα

## 1.2 Αναγνώριση Νοηματικής Γλώσσας

Η Αναγνώριση Νοηματικής Γλώσσας (ΑΝΓ) είναι όπως είδαμε μια περίπτωση εφαρμογής του γενικότερου πεδίου της επικοινωνίας ανθρώπου - μηχανής. Τα τελευταία χρόνια έχει σημειωθεί σημαντική κινητικότητα για την ανάπτυξη αποτελεσματικών εφαρμογών προς αυτήν την κατεύθυνση. Αυτό οφείλεται εν μέρει στο γεγονός ότι η αναγνώριση φωνής χρησιμοποιείται ολοένα και ευρύτερα, και μπορεί στο μέλλον να αποτελέσει το κυρίαρχο στοιχείο της επικοινωνίας ανθρώπου - μηχανής. Αν συμβεί κάτι τέτοιο, οι κωφάλαλοι θα αντιμετωπίσουν μεγάλα προβλήματα στη χρήση των υπολογιστών, εάν δεν υπάρξει και ανάλογη πρόοδος και στις εφαρμογές αναγνώρισης νοηματικής γλώσσας. Επιπλέον, η αναγνώριση νοηματικής γλώσσας θα εξυπηρετούσε στην καλύτερη μοντελοποίηση της γλώσσας αυτής, ενώ θα εξυπηρετούσε και στην ορθότερη «μεταγραφή» (transcription) της ομιλούμενης γλώσσας, σε περιπτώσεις όπου δεν υπάρχουν διαθέσιμοι μεταφραστές.

Ένα σύστημα αναγνώρισης νοηματικής γλώσσας αποτελείται από τα εξής στάδια: την απόκτηση της ακολουθίας εικόνων που αντιστοιχεί σε μια σειρά νοημάτων (signs), την εξαγωγή τρισδιάστατων ή δισδιάστατων χαρακτηριστικών, στατικών ή/και δυναμικών, για τα χέρια, τους βραχίονες, το πρόσωπο και τη γενική διάταξη του σώματος του νοηματιστή/ομιλητή, και κατόπιν την αναγνώριση της ακολουθίας νοημάτων με βάση τα χαρακτηριστικά αυτά, και ενδεχομένως την μεταγραφή αυτής σε κείμενο ή σε κάποια άλλη μορφή που είναι δυνατόν να αναγνωσθεί (π.χ. σύστημα HamNoSys).

Στην παρούσα εργασία θα ασχοληθούμε με τα στάδια που προηγούνται της αναγνώρισης των χαρακτηριστικών, και συγκεκριμένα με την επεξεργασία της ακολουθίας εικόνων και την εξαγωγή χαρακτηριστικών. Μέχρι πρότινος (1994-1996) τα εξαγόμενα χαρακτηριστικά προέκυπταν μετά από χρήση ειδικών γαντιών, και η αναγνώριση βασιζόταν στην υπόθεση ότι κάθε νεύμα ήταν μια διακριτή ακολουθία από συγκεκριμένες τοποθετήσεις των χεριών και των δαχτύλων, οπότε για την αναγνώριση αρκούσε η γνώση της διάταξης των δαχτύλων σε όλη τη διάρκεια του νεύματος και την αναγνώριση όλων των ενδιάμεσων χειρομορφών που χρησιμοποιεί ο νοηματιστής (fingerspelling). Η προσέγγιση αυτή είχε το πλεονέκτημα ότι επέτρεπε την ακριβή αναγνώριση των χειρομορφών, ωστόσο δεν λάμβανε υπόψιν ότι η συνολική κίνηση των μελών του σώματος του ομιλητή, και όχι μια ακολουθία από χειρομορφές, είναι αυτή που συμβάλλει στην αναγνώριση (από κάποιον άλλο άνθρωπο) ενός νεύματος, ενώ επιπλέον απαιτούσε τη χρήση εξειδικευμένου υλικού, κάτι που ερχόταν σε αναντιστοιχία με τις αντίστοιχες εφαρμογές αναγνώρισης φωνής, όπου αρκούσε μόνο ο βασικός εξοπλισμός. Έτσι, τελευταία το ενδιαφέρον των ερευνητών έχει στραφεί στην οπτική αναγνώριση νοηματικής γλώσσας. Τα συστήματα αυτά αποτελούνται βασικά από ένα στάδιο κατάτμησης της εικόνας εισόδου και εντοπισμού/παρακολούθησης των χεριών, και από ένα στάδιο εξαγωγής χαρακτηριστικών.

### 1.2.1 Κατάτμηση και Παρακολούθηση των Χεριών

Ιδιαίτερα σημαντική συνιστώσα για την επιτυχή λειτουργία ενός συστήματος ανίχνευσης και παρακολούθησης του ατόμου που χειρονομεί είναι η ακριβής ανάκτηση της γεωμετρικής διάταξης και θέσης των χεριών ή ακόμα και των δαχτύλων του ομιλητή σε μια ακολουθία εικόνων, με χρήση κάποιου αλγορίθμου κατάτμησης των εικόνων. Σε αυτήν την ακολουθία γεωμετρικών χαρακτηριστικών συμπυκνώνεται σε μεγάλο βαθμό το πληροφοριακό περιεχόμενο των χειρονομιών. Το πρόβλημα είναι αρκετά δύσκολο εάν το περιβάλλον λειτουργίας του συστήματος δεν είναι περιορισμένο. Γενικά

για την κατάτμηση της εικόνας στη βιβλιογραφία συναντούμε διάφορες προσεγγίσεις, όπως μεταβολικές (variational) τεχνικές (γεωδαιτικά ενεργά περιγράμματα, ενεργές περιοχές), στατιστικές μεθόδους (Μαρκοβιανά πεδία, γενετικά μοντέλα) ή τεχνικές γράφων. Περισσότερα για τις διάφορες μεθόδους κατάτμησης μπορούν να βρεθούν στο Κεφάλαιο 2.

Ειδικότερα στις εφαρμογές που αφορούν στην ανίχνευση ατόμου που χειρονομεί τα συστήματα που απαντώνται συνήθως βασίζονται σε κάποια από τις παραπάνω μεθόδους, με χρήση πληροφορίας που προέρχεται είτε από το χρώμα, είτε από την κίνηση, ή από τις ακμές της εικόνας. Ενδεικτικές μέθοδοι αναφέρονται συγκεντρωτικά στο [34]. Τα συστήματα της πρώτης ομάδας ([22], [3], [44]) βασίζονται στην υπόθεση ότι το χρώμα του ανθρώπινου δέρματος μπορεί να χρησιμοποιηθεί για το διαχωρισμό των χεριών από το φόντο. Γενικά προϋποτίθεται ότι δεν υπάρχουν άλλα αντικείμενα με παρόμοιο χρώμα στην εικόνα, διαφορετικά πρέπει να εφαρμοστεί κάποια ειδική διαδικασία για την αφαίρεση του μη χρήσιμου φόντου από την εικόνα. Η δεύτερη κατηγορία μεθόδων ([17], [43]) υποθέτει ότι τα χέρια είναι τα μοναδικά κινούμενα αντικείμενα στην εικόνα, ενώ τα υπόλοιπα μέρη του σώματος του ομιλητή (κεφάλι και κορμός) παραμένουν σχετικά σταθερά. Οι μέθοδοι αυτές απαιτούν τα χέρια να κινούνται διαρκώς. Τέλος, οι μέθοδοι που βασίζονται στον εντοπισμό ακμών ([20]) χρησιμοποιούν χαρακτηριστικά σχήματος και θέσης για να ξεχωρίσουν τα χέρια.

### 1.2.2 Εξαγωγή Χαρακτηριστικών

Στις εφαρμογές αναγνώρισης που δε χρησιμοποιούν fingerspelling, αλλά θεωρούν τη συνολική κίνηση των μελών του σώματος κατά την εκτέλεση του νεύματος (full signing scenario), χρησιμοποιούνται για την ταξινόμηση των νευμάτων χαρακτηριστικά που αντιπροσωπεύουν τη συνολική θέση και την κίνηση των χεριών, καθώς και το σχήμα και τον προσανατολισμό των χειρομορφών. Ένα πολύ κοινό χαρακτηριστικό που εξάγεται και δίνει τη θέση των χεριών είναι το κέντρο μάζας. Οι δύο συντεταγμένες θέσεις μπορεί να λαμβάνονται ως οι απόλυτες συντεταγμένες του στην εικόνα, είτε σε σχέση με το κέντρο μάζας του προσώπου του νοηματιστή.

Στη βιβλιογραφία συναντούμε πλήθος περιγραφητών σχήματος, όπως για παράδειγμα οι περιγραφητές Fourier ([48], [25], ή [19] για εφαρμογή σε αναγνώριση χειρονομιών), τα ιστογράμματα ([33], [15]) και οι πίνακες σχήματος, οι κεντρικές ροπές n τάξης ([44]), ο χώρος κλίμακας καμπυλότητας ([32], [10]) για εφαρμογή σε αναγνώριση χειρονομιών) κ.α. Γενικά μπορούμε να ξεχωρίσουμε δύο διαφορετικούς τύπους

περιγραφητών σχήματος, που βασίζονται οι μεν πρώτοι στις ιδιότητες της καμπύλης που ορίζει τα όρια της περιοχής των χειρών (boundary-based), και οι δεύτεροι στις ιδιότητες της ίδιας της περικλειόμενης περιοχής (region-based).

Τέλος, όσον αφορά στα δυναμικά χαρακτηριστικά, συχνά η μόνη πληροφορία που χρησιμοποιείται είναι η ταχύτητα μετατόπισης του κέντρου μάζας ([22],[46]), ενώ άλλοτε χρησιμοποιείται και η οπτική ροή, όπως στο [11].

### 1.3 Σύνοψη της παρούσας Εργασίας

Η παρούσα εργασία ασχολείται με την κατασκευή ενός ολοκληρωμένου συστήματος ανίχνευσης και παρακολούθησης ανθρώπινων χειρονομιών, με προορισμό να χρησιμοποιηθεί για την εξαγωγή χαρακτηριστικών από βίντεο νοηματικής γλώσσας σε εφαρμογή αναγνώρισης. Για το σκοπό αυτό θα δοκιμαστούν και θα τροποποιηθούν ήδη υπάρχουσες μέθοδοι για την κατάτμηση εικόνων και την εξαγωγή χαρακτηριστικών.

Όπως αναφέρθηκε και στην εισαγωγή, ένα σύστημα παρακολούθησης ομιλητή της νοηματικής γλώσσας πρέπει βασικά να αποτελείται από δύο βασικά συστατικά: μια μεθοδολογία κατάτμησης των καρέ του βίντεο και απομόνωσης των κινούμενων μερών, και ένα σύστημα παρακολούθησης της κίνησης. Επιπλέον των παραπάνω, προκειμένου για χρήση σε εφαρμογή αναγνώρισης, το σύστημα πρέπει να συνοδεύεται από έναν αλγόριθμο εξαγωγής κατάλληλων χαρακτηριστικών. Συνολικά λοιπόν η η έρευνά μας επικεντρώθηκε στα παρακάτω σημεία:

1. Κατάτμηση Εικόνας (Κεφάλαιο 2)
2. Εξαγωγή Χαρακτηριστικών (Κεφάλαιο 3)
3. Παρακολούθηση Κίνησης (Κεφάλαιο 4)
4. Αντιμετώπιση Φαινομένων Επικάλυψης (Κεφάλαιο 5)

Κάθε ένα από τα βασικά συστατικά του συστήματος θα αναπτυχθεί διεξοδικά στο αντίστοιχο κεφάλαιο, τόσο ως προς τη σχετική θεωρία όσο και ως προς τα πειραματικά αποτελέσματα.

#### Κατάτμηση Εικόνας

Στην εργασία χρησιμοποιήθηκε μια μέθοδος που συνδυάζει την πληροφορία του χρώματος με την πληροφορία των ακμών, ενώ όπως θα δούμε παρακάτω, σε ορισμέ-



νες περιπτώσεις χρησιμοποιήθηκε και πληροφορία κίνησης για να αυξηθεί η ευρωστία του συστήματος. Ως μέθοδος κατάτμησης χρησιμοποιήθηκε η μέθοδος των ενεργών περιγραμμάτων/περιοχών και επιπεδοσυνόλων ([8], [37]) που όπως θα δούμε δίνει αρκετά ικανοποιητικά αποτελέσματα στην περίπτωση παρακολούθησης χειρονομιών. Ένα ιδιαίτερα σημαντικό χαρακτηριστικό αυτής της κατηγορίας μεθόδων είναι ότι επιτρέπουν την ενσωμάτωση πρότερης πληροφορίας αναφορικά με το σχήμα του αντικειμένου που παρακολουθείται ([12], [28], [38]). Τούτο είναι πολύ σημαντικό στην περίπτωση μας γιατί προσδίδει στο σύστημα επιπλέον αξιοπιστία και αντοχή έναντι θορύβου ή απόκρυψης.

### Εξαγωγή Χαρακτηριστικών

Όσον αφορά στα στατικά χαρακτηριστικά σχήματος, δοκιμάστηκαν διάφοροι περιγραφητές, όπως οι περιγραφητές Fourier και ο χώρος κλίμακας καμπυλότητας. Επίσης εξετάστηκε ένα νέο σύνολο χαρακτηριστικών σχήματος βασισμένο στη φασματική ανάλυση της καμπυλότητας των σχημάτων και συγκεκριμένα στο cepstrum της καμπυλότητας.

Θα εξαχθούν επίσης και ένα σύνολο δυναμικών χαρακτηριστικών, που βασίζονται στην επιλογή εκείνων των τμημάτων της εικόνας που εμφανίζουν τη μέγιστη κινητικότητα. Η μεθοδολογία ακολουθεί τα όσα αναφέρονται στο [11]. Δοκιμάστηκαν δύο τεχνικές ανίχνευσης της κίνησης, οι τεχνικές block matching ([4]) και η οπτική ροή ([30]).

### Παρακολούθηση Κίνησης

Όποτε αυτό ήταν απαραίτητο, για παράδειγμα για την εξαγωγή των δυναμικών χαρακτηριστικών κίνησης ή για την εκτίμηση της επόμενης θέσης του χεριού, εφαρμόστηκε κάποια μεθοδολογία παρακολούθησης της κίνησης. Η παρακολούθηση έγινε με χρήση γεωδαιτικών ενεργών περιοχών και πάλι, με τρόπο παρόμοιο με αυτόν που παρουσιάζεται στο [13], όπου μοντελοποιούνται στατιστικά οι κινούμενες και ακίνητες περιοχές στην εικόνα και η μοντελοποίηση αυτή χρησιμοποιείται για την οδήγηση ενός γεωδαιτικού ενεργού περιγράμματος το οποίο τελικά απομονώνει μόνο την κινούμενη περιοχή. Για τη στατιστική μοντελοποίηση έγινε χρήση της οπτικής ροής.

Επιπλέον, όπως αναφέραμε, σε αρκετές περιπτώσεις κρίθηκε χρήσιμο να υπάρχει μια εκτίμηση της επόμενης θέσης του χεριού. Η εκτίμηση αυτή βασίζεται σε πρότερη γνώση σχετικά με τη θέση και την ταχύτητα των χεριών και ακολουθεί τα όσα

αναφέρονται στο [43].

### Αντιμετώπιση Επικαλύψεων

Ο εντοπισμός και η παρακολούθηση των χεριών στην περίπτωση όπου εμφανίζονται επικαλύψεις είναι ένα ιδιαίτερα δύσκολο καθήκον. Στη βιβλιογραφία αυτό αντιμετωπίζεται είτε με χρήση σύμπτωσης προτύπων ([46]), είτε με χρήση εκτίμησης της κίνησης (π.χ. με χρήση φίλτρων Kalman, [22]), είτε με χρήση πολλαπλών καμερών, οπότε οι διαφορετικές λήψεις της ίδιας σκηνής ελαχιστοποιούν την ασάφεια. Τελευταία προτάθηκε και μια τελείως διαφορετική προσέγγιση στο πρόβλημα των επικαλύψεων, που βασίζεται στο μετασχηματισμό της εικόνας στο πεδίο της εικόνας δυναμικού ([42]).

Εδώ θα δοκιμαστούν οι περισσότερες από τις παραπάνω μεθόδους, με κάποιες τροποποιήσεις και βελτιώσεις, σε συνδυασμό και με κάποια στοιχεία από την παρακολούθηση της κίνησης. Ο συγκερασμός των παραπάνω μεθοδολογιών αναμένεται να βελτιώσει σημαντικά το συνολικό αποτέλεσμα.

---

## ΚΑΤΑΤΜΗΣΗ ΤΗΣ ΕΙΚΟΝΑΣ

---

Η διαδικασία της κατάτμησης των εικόνων είναι ένα από τα δυσκολότερα προβλήματα της Όρασης Υπολογιστών, και συχνά αποτελεί ένα πολύ σημαντικό πρώτο βήμα για τις περισσότερες από τις πρακτικές εφαρμογές στο χώρο αυτό. Με τον όρο «κατάτμηση» εννοούμε τον εντοπισμό των διαφορετικών περιοχών - αντικειμένων που απαρτίζουν την εικόνα, δηλαδή την εύρεση ενός αριθμού μη επικαλυπτόμενων περιοχών, που είναι εμφανώς διαχωρίσιμες από το φόντο της εικόνας και είναι επίσης σχετικά ομογενείς ως προς κάποιο χαρακτηριστικό (π.χ. υφή, χρώμα, βάθος ή κινητικότητα). Είναι φανερό ότι τόσο η έννοια της διαχωρισιμότητας από το φόντο όσο και της ομοιογένειας είναι πρακτικά δύσκολο να οριστούν με αυστηρό και αντικειμενικό τρόπο (ακόμα και μεταξύ ανθρώπων), με αποτέλεσμα το πρόβλημα της κατάτμησης να εξαρτάται έντονα από τις ανάγκες της εκάστοτε εφαρμογής.

Στο κεφάλαιο αυτό θα ασχοληθούμε με ορισμένες μεθόδους κατάτμησης. Αρχικά θα αναφερθούν συνοπτικά ορισμένες από τις πιο ευρέως χρησιμοποιούμενες μεθόδους κατάτμησης εικόνων, και στη συνέχεια θα επικεντρωθούμε στην αναλυτική θεωρητική περιγραφή της μεθόδου που εφαρμόστηκε στην παρούσα εργασία.

### 2.1 Γενικές Μέθοδοι Κατάτμησης

Γενικά οι προσεγγίσεις στο πρόβλημα της κατάτμησης μπορούν να ομαδοποιηθούν σε δύο μεγάλες (και συχνά επικαλυπτόμενες) κατηγορίες, αυτές που αναζητούν τον εντοπισμό των ορίων των αντικειμένων (boundary-based), και αυτών που αναζητούν

τις ίδιες τις περιοχές της εικόνας απευθείας (region-based). Οι πρώτες στοχεύουν στην κατασκευή μιας εικόνας ακμών, όπου απομονώνονται τα σύνορα μεταξύ των διαφορετικών αντικειμένων στην εικόνα, οδηγώντας έτσι στο διαχωρισμό τους, και οι δεύτερες στην ομαδοποίηση των εικονοστοιχείων της εικόνας σε περιοχές, με βάση την τοπική ομοιογένεια του χαρακτηριστικού που μας ενδιαφέρει. Συχνά, για αύξηση της ευρωστίας της κατάτμησης, χρησιμοποιούνται περισσότερα από ένα χαρακτηριστικά, συνήθως σε συνδυασμό με πληροφορία για τη μεταβολή της φωτεινότητας στην εικόνα, που βοηθά στην ανίχνευση των ακμών.

Οι μέθοδοι κατάτμησης μπορούν να κατηγοριοποιηθούν ανάλογα με τη μαθηματική μεθοδολογία που χρησιμοποιούν σε:

- Μεταβολικές μεθόδους
- Στατιστικές μεθόδους
- Μεθόδους που βασίζονται σε γράφους
- Τοπογραφικές - Μορφολογικές μεθόδους

**Μεταβολικές Μέθοδοι:** Οι μέθοδοι αυτές βασίζονται στην ελαχιστοποίηση ενός συναρτησιακού  $J$  ορισμένου στο σύνολο όλων των δυνατών διαχωρισμών της εικόνας σε περιοχές. Το συναρτησιακό αυτό εκφράζει μια μορφή ενέργειας, η οποία είναι κατάλληλα ορισμένη ώστε να ελαχιστοποιείται για εκείνο το διαχωρισμό της εικόνας που εξασφαλίζει τη βέλτιστη κατάτμησή της. Ενσωματώνει δε πληροφορία που αφορά τόσο το σύνορο μεταξύ των αντικειμένων όσο και την εσωτερική επιφάνεια αυτών:

$$J = \sum_{i=1}^N \int_{R_i} f(x, R_i) dx + \sum_{i=1}^N \oint_{\partial R_i} g(s) ds \quad (2.1)$$

Στην κατηγορία αυτή οι κυριότερες μέθοδοι είναι αυτές των ενεργών περιγραμμάτων («φιδιών»), των γεωδαιτικών ενεργών περιγραμμάτων και ενεργών περιοχών, που θα χρησιμοποιηθούν και στην παρούσα εργασία. Οι μέθοδοι αυτές μπορούν να επεκταθούν ώστε να κάνουν χρήση περισσότερων του ενός χαρακτηριστικού για την κατάτμηση.

**Στατιστικές Μέθοδοι:** Στις μεθόδους αυτές ο διαχωρισμός της εικόνας θεωρείται ότι ανήκει σε ένα τυχαίο πεδίο, από το οποίο προσπαθούμε να παράγουμε μια βέλτιστη εκτίμηση, με βάση κάποιο κριτήριο όπως το κριτήριο της μέγιστης πιθανοφάνειας ή του ελάχιστου μέσου τετραγωνικού σφάλματος. Από τις πιο γνωστές προσεγγίσεις στην κατηγορία αυτή είναι η χρήση Τυχαίων Μαρκοβιανών Πεδίων (MRFs,

[41],[23]). Σε αυτήν η εικόνα μοντελοποιείται ως μια τυχαία μεταβλητή στο διδιάστατο χώρο, και θεωρείται ότι η τιμή κάθε εικονοστοιχείου (και κατά συνέπεια η κατάταξή του σε κάποια από τις περιοχές της εικόνας) εξαρτάται μόνο από τις τιμές των γειτονικών του και όχι από το σύνολο των εικονοστοιχείων της εικόνας. Με βάση την παραδοχή αυτή, και την κατάστροψη ενός τύπου γειτονιάς και ενός κατάλληλου συναρτησιακού ενέργειας, βρίσκουμε εκείνο το διαχωρισμό που επιτυγχάνει την ελάχιστη ενέργεια. Μια άλλη μέθοδος είναι αυτή των παραγωγικών μοντέλων, όπου κάθε περιοχή μοντελοποιείται στατιστικά με βάση ένα σύνολο χαρακτηριστικών της, και οι παρατηρήσεις (εικονοστοιχεία) της εικόνας ταξινομούνται στις περιοχές στις οποίες ταιριάζουν καλύτερα. Και οι δύο αυτές μέθοδοι μπορούν να προσαρμοστούν ώστε σε κάθε περιοχή να λαμβάνεται ένα κυρίαρχο χαρακτηριστικό, αντί να χρησιμοποιείται το σύνολο των χαρακτηριστικών για όλες τις περιοχές, έτσι ώστε να γίνεται επιλογή του πιο κατάλληλου χαρακτηριστικού κάθε φορά.

**Μέθοδοι Γράφων:** Οι γραφοθεωρητικές μέθοδοι μοντελοποιούν την εικόνα ως ένα γράφο με βάρη, όπου οι κόμβοι του γράφου είναι τα εικονοστοιχεία της εικόνας, ενώ στις ακμές που συνδέουν τους κόμβους ενσωματώνεται η πληροφορία σχετικά με το είδος της κατάτμησης που ζητούμε: μπορεί να αντιστοιχούν σε εξαρτήσεις μεταξύ γειτονικών εικονοστοιχείων, όπως στα MRF μοντέλα, ή στο αποτέλεσμα της εφαρμογής ενός φίλτρου ανίχνευσης ακμών, που δείχνει κατά πόσον δύο γειτονικά εικονοστοιχεία ανήκουν σε διαφορετικά τμήματα της εικόνας. Η κατάτμηση πραγματοποιείται ακολουθώντας αρχές από τη θεωρία των γράφων, με τελικό στόχο την μεγιστοποίηση κάποιου κατάλληλου κριτηρίου. Οι μέθοδοι γράφων έχουν χρησιμοποιηθεί σε συνδυασμό με τα MRFs, για επιτάχυνση των υπολογισμών, αλλά και ανεξάρτητα. Ένα παράδειγμα είναι η περίπτωση της εύρεσης του ελάχιστου μονοπατιού μεταξύ δύο σημείων στο γράφο, όπου ως «σημείο» ορίζεται είτε ένα εικονοστοιχείο είτε μια περιοχή της εικόνας, ενώ ως μονοπάτι ορίζεται μια κατάλληλα καταστρωμένη «απόσταση», που σχετίζεται με το είδος της κατάτμησης που πραγματοποιούμε.

**Τοπογραφικές - Μορφολογικές Μέθοδοι:** Η μορφολογία είναι ένα μαθηματικό εργαλείο που βασίζεται σε αρχές της θεωρίας των πλεγμάτων και των συνόλων και έχει στόχο την ποσοτικοποίηση των γεωμετρικών χαρακτηριστικών της εικόνας. Μεταξύ των μορφολογικών μεθόδων κατάτμησης κυρίαρχη είναι εκείνη που βασίζεται στο μετασχηματισμό watershed. Η μέθοδος αυτή βασίζεται στην εξαγωγή ορισμένων βασικών χαρακτηριστικών από μια αποθρομβοποιημένη εκδοχή της εικόνας, τα οποία βοηθούν στον εντοπισμό των κέντρων κάποιων ομογενών περιοχών της εικόνας. Στη συνέχεια ξεκινώντας από τα κέντρα αυτά πραγματοποιείται «πλημμυρισμός» των πε-

ριοχών αυτών, μέχρις ότου τα σύνορα των πλημμυρισμένων περιοχών πλησιάσουν, οπότε στο σημείο αυτό ανιχνεύεται το σύνορο μεταξύ των αντίστοιχων αντικειμένων στην εικόνα. Το βασικό πρόβλημα είναι ο εντοπισμός των κέντρων που προαναφέραμε, ο οποίος εξαρτάται έντονα από την εκάστοτε εφαρμογή. Για τη μέθοδο αυτή έχουν προταθεί ένας αριθμός από βελτιώσεις, όπως η εκτέλεση σε πολλαπλές κλίμακες, η εφαρμογή εναλλακτικών σχεδίων πλημμυρισμού και αποθορυβοποίησης και η μοντελοποίηση της διαδικασίας του μετασχηματισμού με μερικές διαφορικές εξισώσεις.

Στην παρούσα εργασία θα εργαστούμε με μεταβολικές μεθόδους, και συγκεκριμένα τη μέθοδο των γεωδαιτικών ενεργών περιγραμμάτων, με χρήση επιπεδοσυνόλων. Η μέθοδος αυτή είχε δοκιμαστεί και νωρίτερα στο [6] και είχε δώσει αρκετά ικανοποιητικά αρχικά αποτελέσματα. Επίσης, ένα ιδιαίτερα σημαντικό χαρακτηριστικό αυτής της κατηγορίας μεθόδων είναι ότι επιτρέπουν την ενσωμάτωση πρότερης πληροφορίας αναφορικά με το σχήμα του αντικειμένου που παρακολουθείται. Τούτο είναι πολύ σημαντικό στην περίπτωσή μας γιατί προσδίδει στο σύστημα επιπλέον αξιοπιστία και αντοχή έναντι θορύβου ή απόκρυψης.

## 2.2 Μέθοδος Γεωδαιτικών Ενεργών Περιγραμμάτων

### 2.2.1 Ενεργά Περιγράμματα

Τα πρώτα βήματα για την ανάπτυξη της μεθόδου αυτής έγιναν από τους Kass, Witkin και Terzopoulos στο [24], όπου πρότειναν το μοντέλο των ενεργών περιγραμμάτων, ή «φιδιών». Η γενική ιδέα της μεθόδου αυτής είναι ότι τα όρια των περιοχών που αντιστοιχούν στα διαφορετικά αντικείμενα μιας εικόνας μπορούν να μοντελοποιηθούν από μια κλειστή καμπύλη  $\vec{C}(p) = [x(p), y(p)]^T : [0, 1] \rightarrow \mathbb{R}^2$ ,  $\vec{C}(0) = \vec{C}(1)$  στο διδιάστατο χώρο, η οποία, ξεκινώντας από μια αρχική θέση  $C_0$ , εξελίσσεται υπό την επίδραση διάφορων δυνάμεων μέχρις ότου καταλήξει (συγκλίνει) σε μια θέση όπου ελαχιστοποιείται ένα συναρτησιακό ενέργειας. Στη θέση αυτή η καμπύλη ταυτίζεται με το σύνορο μεταξύ των εξεταζόμενων αντικειμένων.

Το συναρτησιακό ενέργειας ποικίλλει ανάλογα με τα αντικείμενα προς αναγνώριση και τη μέθοδο που χρησιμοποιείται στην εκάστοτε εφαρμογή. Οι όροι που εμφανίζονται σε αυτό και καθορίζουν την εξέλιξη της καμπύλης μπορεί να σχετίζονται με:

- την ομαλότητα της καμπύλης, εξασφαλίζοντας ότι η καμπύλη δεν θα είναι «πριονωτή»

- την μεταβολή της φωτεινότητας, οπότε οδηγούν στην ανίχνευση των ακμών
- πρότερη γνώση για το σχήμα ή τη θέση των προς αναζήτηση αντικειμένων
- στατιστικές μοντελοποιήσεις ορισμένων χαρακτηριστικών των προς αναζήτηση περιοχών, όπως η υφή, το χρώμα ή η κινητικότητα

Από την παραπάνω λίστα οι τελευταίες δύο επεκτάσεις έγιναν μεταγενέστερα σε σχέση με το αρχικό paper των Kass et al.

Στα παραπάνω η καμπύλη  $\vec{C}$  παραμετροποιείται ως προς κάποια τυχαία παράμετρο  $p$ . Η παράμετρος αυτή δίνει την ταχύτητα κίνησης κατά μήκος της καμπύλης  $\partial\vec{C}/\partial p$ . Η κατεύθυνση της ταχύτητας αυτής είναι ίδια με την κατεύθυνση του μοναδιαίου εφαπτόμενου στην καμπύλη διάνυσματος  $\vec{T}$ , σε κάθε σημείο της καμπύλης (για περισσότερα σε σχέση με τις καμπύλες βλ. Ενότητα 3.1). Αν συμβολίσουμε με  $\vec{N}$  το μοναδιαίο κάθετο στην καμπύλη διάνυσμα, με κατεύθυνση προς το εσωτερικό της καμπύλης, τότε μπορεί να αποδειχθεί (βλ. [39]) ότι η εξέλιξη (παραμορφωτική κίνηση) μιας καμπύλης με ταχύτητα παραμόρφωσης  $\vec{V}$  που αποτελείται τόσο από μια εφαπτομενική όσο και από μια κάθετη συνιστώσα,

$$\frac{\partial\vec{C}(p, t)}{\partial t} = \alpha(p, t)\vec{T}(p, t) + \beta(p, t)\vec{N}(p, t)$$

(όπου  $\alpha = \langle \vec{V} \cdot \vec{T} \rangle$ ,  $\beta = \langle \vec{V} \cdot \vec{N} \rangle$ ) είναι ίδια με αυτήν που θα εμφανιζόταν αν υπήρχε μόνο η κάθετη συνιστώσα, αν η παράμετρος  $\beta$  δεν εξαρτάται από την παραμετροποίηση της καμπύλης, δηλαδή πρόκειται για κάποιο εσωτερικό γεωμετρικό χαρακτηριστικό της καμπύλης. Με άλλα λόγια, η εφαπτομενική συνιστώσα κίνησης επηρεάζει μόνο την παραμετροποίηση της καμπύλης. Συνεπώς, η καμπύλη θα εξελισσεται με βάση τη σχέση:

$$\frac{\partial\vec{C}(p, t)}{\partial t} = V \cdot \vec{N}_o(p, t) \quad (2.2)$$

με την αρχική συνθήκη  $\vec{C}(p, 0) = C_0, \forall p \in [0, 1]$ , όπου  $V = \vec{C}_t^T \vec{N}_o$  η συνιστώσα της ταχύτητα εξέλιξης της καμπύλης που είναι παράλληλη στο κάθετο διάνυσμα. Η ταχύτητα  $V$  προκύπτει από την ελαχιστοποίηση του συναρτησιακού ενέργειας.

Στο [24] οι συγγραφείς χρησιμοποίησαν το εξής συναρτησιακό ενέργειας:

$$E_{snake}(\vec{C}) = \int_0^1 F_{snake}(\vec{C}) = \int_0^1 [E_{int}(\vec{C}) + E_{edge}(\vec{C}) + E_{con}(\vec{C})] \quad (2.3)$$

όπου ο όρος  $E_{int} = a \|\frac{\partial\vec{C}}{\partial p}\|^2 + b \|\frac{\partial^2\vec{C}}{\partial p^2}\|^2$  εξασφαλίζει ότι θα ικανοποιούνται ορισμένες απαιτήσεις ομαλότητας από την καμπύλη, ο όρος  $E_{edge} = -\lambda \|\nabla I(\vec{C})\|^2$  κατευθύνει

την καμπύλη προς τα σημεία όπου η παράγωγος της εικόνας παίρνει μεγάλες τιμές (που αντιστοιχούν στα όρια μεταξύ των αντικειμένων), και ο όρος  $E_{con}$  αφορά τυχόν άλλες εξωτερικές δυνάμεις που καθορίζονται ανάλογα με τις απαιτήσεις της εκάστοτε εφαρμογής.

### 2.2.2 Γεωδαιτικά Ενεργά Περιγράμματα

Η μέθοδος που περιγράφηκε στο [24] πράγματι ήταν πρωτοποριακή για την εποχή της, είχε όμως ορισμένα βασικά μειονεκτήματα. Κατ'αρχήν, το μοντέλο αυτό δεν μπορούσε να χειριστεί με ορθό τρόπο τυχόν μεταβολές στη συνεκτικότητα και την τοπολογία της καμπύλης, καθώς ήταν αδύνατο η αρχική καμπύλη να «σπάσει» για να ανιχνεύσει περισσότερα του ενός αντικείμενα. Επίσης, το μοντέλο εξαρτάται σημαντικά από την παραμετροποίηση της καμπύλης. Τέλος, το σημείο ελαχίστου εξαρτιόταν σημαντικά από την αρχική θέση της καμπύλης.

Μια σημαντική βελτίωση προτάθηκε στο [8], προκειμένου οι παράμετροι της εξέλιξης της καμπύλης να είναι ανεξάρτητες από την παραμετροποίηση αυτής. Οι συγγραφείς εξέτασαν το εξής συναρτησιακό ενέργειας:

$$E = \int_0^1 [aE_{int}(\vec{C}) + bE_{edge}(\vec{C}) + \lambda E_{con}(\vec{C})] dp \quad (2.4)$$

Στην παραπάνω σχέση ο όρος  $E_{int} = a \|\frac{\partial \vec{C}}{\partial p}\|^2$  εξασφαλίζει ότι θα ικανοποιούνται ορισμένες απαιτήσεις ομαλότητας από την καμπύλη, ενώ ο όρος  $E_{edge} = \lambda g(\|\nabla I(\vec{C})\|)$ . Η συνάρτηση  $g : [0, +\infty] \rightarrow \mathbb{R}^+$  είναι μια συνάρτηση ανίχνευσης ακμών, η οποία εφαρμοζόμενη στην παράγωγο της εικόνας  $\nabla I$  εντοπίζει τις ακμές αυτής. Η  $g$  πρέπει να είναι φθίνουσα έτσι ώστε

$$\lim_{r \rightarrow \infty} g(r) = 0$$

. Παραδείγματα κατάλληλων συναρτήσεων είναι για παράδειγμα:

$$g = g(I) = \frac{1}{1 + (\|\nabla G_\sigma * I\|)^n} \quad (2.5)$$

ή

$$g = g(I) = \exp(-\|\nabla G_\sigma * I\|) \quad (2.6)$$

Στην παραπάνω σχέση ο όρος  $E_{int} = a \|\frac{\partial \vec{C}}{\partial p}\|^2$  εξασφαλίζει ότι θα ικανοποιούνται ορισμένες απαιτήσεις ομαλότητας από την καμπύλη, ενώ ο όρος  $E_{edge} = \lambda g(\|\nabla I(\vec{C})\|)$ . Στα παραπάνω,  $\nabla G_\sigma * I$  είναι μια ομαλοποιημένη εκδοχή της εικόνας  $I$  μετά από



συνέλιξη με μια γκαουσιανή. Με βάση τις ιδιότητες της  $g$  προκύπτει ότι ο όρος  $E_{edge}$  οδηγεί (μέσω της ελαχιστοποίησης της ενέργειας) την καμπύλη προς τις ακμές τις εικόνας  $I$ , όπου μεγιστοποιείται το  $\|\nabla I\|$ . Οι όροι  $a$ ,  $b$ ,  $\lambda$  καθορίζουν το ποσοστό συμβολής του κάθε όρου στον υπολογισμό της ενέργειας.

Στην εξίσωση 2.4 φαίνεται να υπάρχει ακόμα εξάρτηση από την παραμετροποίηση της καμπύλης. Ωστόσο, όπως αποδεικνύεται στο [8], αν αγνοήσουμε προσωρινά τους εξωτερικούς όρους, η ελαχιστοποίηση της 2.4 είναι ισοδύναμη με την ελαχιστοποίηση της παρακάτω συνάρτησης:

$$E = \int_0^1 g(I(\vec{C}(p))) \|\vec{C}_p(p)\| dp = \int_0^{L(\vec{C})} \underbrace{g(I(\vec{C}(s)))}_F ds \quad (2.7)$$

όπου  $ds = \|\vec{C}_p\| dp$  είναι το στοιχειώδες διάστημα κατά μήκος της καμπύλης και  $L(\vec{C})$  το συνολικό μήκος αυτής. Στην τελευταία εξίσωση η καμπύλη  $\vec{C} = \vec{C}(s) : [0, L(\vec{C})] \rightarrow \mathbb{R}^2$  είναι παραμετροποιημένη ως προς τη φυσική παράμετρο του μήκους τόξου  $s$ , που εξαρτάται μόνο από τα γεωμετρικά χαρακτηριστικά της καμπύλης. Συνεπώς όλες οι καμπύλες με το ίδιο σχήμα θα αντιμετωπιστούν με ακριβώς τον ίδιο τρόπο από το μοντέλο.

Προκύπτει με χρήση του λογισμού μεταβολών (από [1]) ότι η ελαχιστοποίηση της 2.7 ισοδυναμεί με το μηδενισμό των παραγώγων Euler της συνάρτησης  $F$ ,  $[F]_x$ ,  $[F]_y$ , και μετά από πράξεις καταλήγουμε στις εξής σχέσεις:

$$[F]_x = g_x - (\nabla g \cdot \vec{C}')x_s + g\kappa y_s = 0 \quad (2.8)$$

$$[F]_y = g_y - (\nabla g \cdot \vec{C}')y_s + g\kappa x_s = 0 \quad (2.9)$$

όπου  $[F]_f = \frac{d}{ds} F_{f_s} - F_f$ , για  $f \in \{x, y\}$  και  $\kappa$  η καμπυλότητα της καμπύλης. Αν χρησιμοποιήσουμε τη μέθοδο κατάβασης κλίσης για την εύρεση του σημείου μηδενισμού των παραπάνω εξισώσεων, θα πάρουμε την εξίσωση για τη χρονική εξέλιξη της καμπύλης,

$$\frac{\partial \vec{C}}{\partial t} = - \begin{pmatrix} [F]_x \\ [F]_y \end{pmatrix} = -[g\kappa + \nabla g \cdot \vec{N}_o] \vec{N}_o \quad (2.10)$$

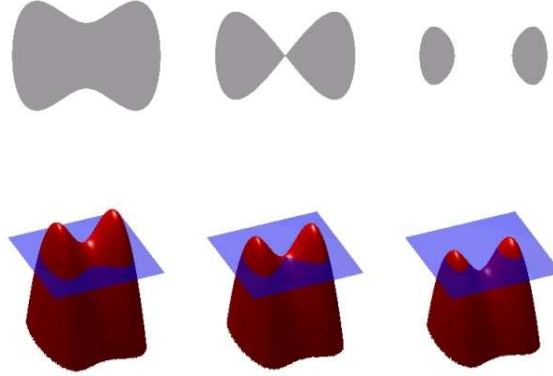
Το μοντέλο αυτό είναι γνωστό ως το μοντέλο γεωδαιτικών ενεργών περιγραμμάτων (geodesic active contours - GAC). Με την παραπάνω μέθοδο εξαλείφεται ουσιαστικά η εξάρτηση της εξέλιξης από την παραμετροποίηση της καμπύλης. Το πρόβλημα της εξάρτησης του σημείου ελαχίστου από την αρχική θέση της καμπύλης εξακολουθεί να υφίσταται, ωστόσο μπορεί να αντιμετωπιστεί με την πραγματοποίηση της εξέλιξης

σε πολλαπλές κλίμακες, όπου αυξανόμενης την κλίμακας η εικόνα γίνεται πιο απλή (coarse) μέσω κατάλληλου φιλτραρίσματος. Το σημείο ισορροπίας της καμπύλης στις μεγαλύτερες κλίμακες χρησιμοποιείται ως αρχικό σημείο για τις μικρότερες. Ωστόσο, παραμένει το μειονέκτημα του μη ορθού χειρισμού των μεταβολών της τοπολογίας της καμπύλης. Συγκεκριμένα, αν έχουμε ένα αντικείμενο που έχει τόσο εξωτερικές όσο και εσωτερικές ακμές, τότε η μέθοδος αδυνατεί να ανιχνεύσει σε ένα πέρασμα τόσο τις εξωτερικές όσο και τις εσωτερικές ακμές. Αντίθετα, ο αλγόριθμος πρέπει να εφαρμόζεται δύο φορές, την πρώτη θεωρώντας ως αρχική συνθήκη μια καμπύλη που να βρίσκεται εξ' ολοκλήρου εκτός του αντικειμένου και τη δεύτερη με αρχική συνθήκη μια καμπύλη εντός του αντικειμένου. Επίσης, πολλαπλά αντικείμενα μπορούν να ανιχνευθούν μόνο με αρχικοποίηση μιας καμπύλης στο εσωτερικό κάθε αντικειμένου. Επιπλέον, προβλήματα προκύπτουν και από το γεγονός ότι η εξέλιξη της καμπύλης πραγματοποιείται αναγκαστικά σε μια διακριτοποιημένη εκδοχή αυτής, με αποτέλεσμα για τη λήψη ορθών αποτελεσμάτων να απαιτείται συχνή επαναδειγματοληψία των σημείων κατά μήκος της καμπύλης, διαδικασία που είναι δύσκολο να πραγματοποιηθεί με αυτόματο τρόπο.

### 2.2.3 Γεωδαιτικά Ενεργά Περιγράμματα και Επιπεδοσύνολα

Για να λυθεί το μειονέκτημα της μεθόδου των GACs ως προς το χειρισμό των αλλαγών στην τοπολογία της καμπύλης, αλλά και προβλήματα που σχετίζονταν με την αριθμητική υλοποίηση της μεθόδου, οι Caselles et al. χρησιμοποίησαν τη μέθοδο των επιπεδοσυνόλων σύμφωνα με τα όσα είχαν προτείνει οι Osher και Sethian στο [35]. Συγκεκριμένα πρότειναν να θεωρηθεί η εξελισσόμενη καμπύλη  $\vec{C}$ , σε κάθε χρονική στιγμή της εξέλιξής της, ως η ισοσταθμική καμπύλη (επιπεδοσύνολο) μιας τρισδιάστατης επιφάνειας  $u : [0, a] \times [0, b] \rightarrow \mathbb{R}$  σε κάποιο σταθερό επίπεδο (στάθμη)  $\lambda$ . Με την επέκταση αυτή η καμπύλη μπορεί να αλλάζει ελεύθερα σχήμα και τοπολογία κατά τη διάρκεια της εξέλιξής της (βλ. Σχ. 2.1), καθώς οι αλλαγές στην τοπολογία προκύπτουν αυτόματα, και εντελώς φυσικά, από την εξέλιξη της επιφάνειας  $u$ . Αρκεί συνεπώς αντί για την καμπύλη να εξελίσσουμε την επιφάνεια  $u$ , και σε κάθε χρονική στιγμή να λαμβάνουμε το επιπεδοσύνολο αυτής στην επιλεγμένη (σταθερή) στάθμη.

Η επιφάνεια  $u$  μπορεί να επιλεγεί τυχαία, αρκεί να ικανοποιεί κάποιους περιορισμούς (να είναι συνεχής κατά Lipschitz). Επίσης συνήθως απαιτούμε η επιφάνεια  $u$  να παίρνει αρνητικές τιμές στο εσωτερικό της καμπύλης ή το αντίστροφο. Με βάση αυτές τις απαιτήσεις, συνήθως η επιφάνεια  $u$  επιλέγεται να αντιστοιχεί στην προσημασμένη



Εικόνα 2.1: Μέθοδος επιπεδοσυνόλων

συνάρτηση απόστασης από την καμπύλη  $\vec{C}$ :

$$u_t(x, y) = \begin{cases} \lambda, & (x, y) \in C(t) \\ \lambda - \bigwedge_{(u,v) \in \vec{C}(t)} \|(x-u, y-v)\|, & (x, y) \in interior(C(t)) \\ \lambda + \bigwedge_{(u,v) \in \vec{C}(t)} \|(x-u, y-v)\|, & (x, y) \in exterior(C(t)) \end{cases} \quad (2.11)$$

Με βάση τα παραπάνω, η εξελισσόμενη καμπύλη μπορεί να περιγραφεί από τη σχέση:

$$C(t) = \{(x, y) : u(x, y, t) = \lambda\}$$

με αρχική συνθήκη:

$$C(0) = \{(x, y) : u(x, y, 0) = u_0(x, y) = \lambda\}$$

ή, ισοδύναμα, η επιφάνεια  $u$  εξελίσσεται με βάση τη σχέση:

$$u(t)(\vec{C}(p, t), t) = \lambda$$

Γενικά, για οποιαδήποτε επιφάνεια  $u$  της οποίας η  $C$  αποτελεί επιπεδοσύνολο, ισχύουν οι σχέσεις:

$$\vec{N}_o = \frac{\nabla u}{\|\nabla u\|} \quad (2.12)$$

και

$$\kappa = -\nabla \vec{N}_o = div \left( \frac{\nabla u}{\|\nabla u\|} \right) \quad (2.13)$$

όπου  $\kappa$  η καμπυλότητα της καμπύλης  $\vec{C}$  και  $\vec{N}_o$  το κάθετο διάνυσμα προς το εξωτερικό της καμπύλης. Ζητούμε η καμπύλη  $C$  να είναι  $\forall t$  επιπεδοσύνολο της  $u$  στο επίπεδο  $\lambda$ , δηλαδή:

$$u(\vec{C}(t), t) = \lambda$$

Με παραγώγιση ως προς το χρόνο παίρνουμε:

$$\begin{aligned} \frac{\partial u}{\partial t} + \nabla u \cdot \frac{\partial \vec{C}}{\partial t} &= 0 \quad (2.12) \\ \frac{\partial u}{\partial t} &= -\vec{N}_o \cdot \frac{\partial \vec{C}}{\partial t} \|\nabla u\| \end{aligned} \quad (2.14)$$

Με χρήση της σχέσης  $V = \vec{N}_o^T \cdot \frac{\partial \vec{C}}{\partial t}$  καταλήγουμε στην εξίσωση εξέλιξης της επιφάνειας  $u$ :

$$\frac{\partial u}{\partial t} = -V \|\nabla u\| \quad (2.15)$$

με αρχική συνθήκη  $u(\vec{C}(0), 0) = C_0$ . Η ταχύτητα  $V$ , όπως αναφέραμε προηγουμένως, επιλέγεται ανάλογα με το συναρτησιακό ενέργειας προς ελαχιστοποίηση. Εξετάσαμε ήδη τη μορφή αυτής στην περίπτωση των γεωδαιτικών ενεργών περιγραμμάτων για την ανίχνευση ακμών. Στην περίπτωση αυτή ο νόμος εξέλιξης της επιφάνειας  $u$  προκύπτει, με αντικατάσταση της  $V$  από την (2.10), και χρήση των (2.13) και (2.12), ως εξής:

$$\frac{\partial u}{\partial t} = \left[ g\kappa + \nabla g \cdot \vec{N}_o \right] \|\nabla u\| = \operatorname{div} \left( g(I) \frac{\nabla u}{\|\nabla u\|} \right) \|\nabla u\| \quad (2.16)$$

Το μοντέλο που παρουσιάσαμε έχει προφανώς τη δυνατότητα άμεσης επέκτασης και σε τρισδιάστατες εικόνες. Το βασικότερο μειονέκτημα τους είναι ο σαφώς μεγαλύτερος χρόνος εκτέλεσης σε σχέση με τις προηγούμενες μεθόδους, καθώς από δισδιάστατο το πρόβλημα μετατρέπεται σε τρισδιάστατο. Για την επιτάχυνση της μεθόδου έχουν προταθεί διάφορες βελτιωμένες αριθμητικές υλοποιήσεις.

Στο [2] προτείνεται η μέθοδος **narrowband**, στην οποία, αντί να πραγματοποιούνται υπολογισμοί για το σύνολο των σημείων του πλέγματος της εικόνας, ελέγχονται μόνο τα σημεία που βρίσκονται στο εσωτερικό μιας στενής ζώνης γύρω από την εξελισσόμενη καμπύλη (μέτωπο). Πράγματι αυτό είναι εφικτό και δίνει σωστά αποτελέσματα, εφ' όσον τελικά μας ενδιαφέρουν μόνο οι θέσεις των σημείων πάνω στο μέτωπο, και εφ' όσον οι τιμές της επιφάνειας  $u$  στα υπόλοιπα σημεία του πλέγματος είναι δυνατόν να προκύψουν από τις τιμές του μετώπου μέσω του μετασχηματισμού απόστασης. Η στενή ζώνη ανανεώνεται οποτεδήποτε το μέτωπο πλησιάσει πάνω από

μια ορισμένη απόσταση στα όρια της ζώνης. Με τον τρόπο αυτό μειώνεται σημαντικά το πλήθος των απαιτούμενων υπολογισμών.

Στο [40] παρουσιάζεται μια ακόμα ταχύτερη μέθοδος, η μέθοδος της **ταχείας προέλασης (fast marching)**<sup>1</sup>, η οποία ωστόσο απαιτεί όλα τα σημεία του μετώπου και σε κάθε χρονική στιγμή να κινούνται προς πάντα την ίδια κατεύθυνση (προς το εσωτερικό ή προς το εξωτερικό της καμπύλης), δηλαδή η ταχύτητα  $V$  εξέλιξης της καμπύλης να διατηρεί πρόσημο. Στην περίπτωση αυτή κάθε σημείο του πλέγματος προσπελάζεται από την καμπύλη μόνο μία φορά, οπότε αν θεωρήσουμε το χρόνο προσπέλασης  $T(x,y)$ , στον οποίο η καμπύλη διέρχεται από το σημείο  $(x,y)$ , θα ισχύει η εξίσωση:  $\|\nabla T(x,y)\| = 1/V$ , με την αρχική συνθήκη  $T(x,y) = 0$ ,  $(x,y) \in \vec{C}_0$ , δηλαδή την χρονική στιγμή της έναρξης η καμπύλη είναι το μηδενικό επιπεδοσύνολο του χρόνο προσπέλασης, ενώ στις επόμενες χρονικές στιγμές η καμπύλη λαμβάνεται ως το επιπεδοσύνολο της  $T$  σε θετικά πάντα επίπεδα. Ο αλγόριθμος αυτός είναι πολύ ταχύς, διότι λαμβάνει υπόψην του ότι η πληροφορία διαδίδεται προς μία μόνο κατεύθυνση. Έτσι, τα σημεία των οποίων οι τιμές έχουν ήδη υπολογιστεί δεν ξαναελέγχονται, αφού οι τιμές τους θα παραμείνουν αμετάβλητες από τους μετέπειτα υπολογισμούς.

Στο [13], προτείνεται μια μέθοδος ενδιάμεση των παραπάνω, η μέθοδος **Hermes**, που εφαρμόζει αρχές της μεθόδου ταχείας προέλασης σε περιπτώσεις που η ταχύτητα εξέλιξης δεν διατηρεί πρόσημο.

Τέλος, στο [16] προτείνεται η πολυπλεγματική μέθοδος υλοποίησης των GACs. Πρόκειται για έναν αριθμητικό αλγόριθμο που χρησιμοποιεί μια ιεραρχική διάταξη από πλέγματα διαφορετικού μεγέθους και λεπτομέρειας για την ταχεία επίλυση αριθμητικών προβλημάτων. Αποτελεί ένα συνδυασμό implicit - explicit αριθμητικών μεθόδων για την επίλυση μερικών διαφορικών εξισώσεων και εμφανίζει πολύ καλή ευστάθεια ακόμα και για πολύ μεγάλα χρονικά βήματα. Συνεπώς μπορούμε να χρησιμοποιούμε ελεύθερα μεγάλο χρονικό βήμα για την εξέλιξη της καμπύλης, με αποτέλεσμα η διαδικασία να επιταχύνεται σημαντικά.

#### 2.2.4 Επεκτάσεις του μοντέλου των ενεργών περιγραμμάτων

Η μέθοδος των γεωδαιτικών ενεργών περιγραμμάτων με χρήση επιπεδοσυνόλων, εκτός από τα πλεονεκτήματα που προαναφέραμε, μπορεί επιπλέον να τροποποιηθεί για να συμπεριλάβει με πολύ απλό τρόπο στην εξέλιξη της καμπύλης και άλλες δυνάμεις,

<sup>1</sup>Ανεξάρτητα από τις εφαρμογές του αλγορίθμου ταχείας προέλασης σε προβλήματα όρασης, στο [45] παρουσιάστηκε μια παρόμοια αποδοτική μέθοδος για τον αλγόριθμο αυτό

που να εκμεταλλεύονται και άλλα χαρακτηριστικά της εικόνας, πλην των ακμών, για την κατάτμηση. Μερικές από τις επεκτάσεις αυτές θα αναφέρουμε στην παρούσα ενότητα.

### Ανταγωνισμός Περιοχών

Στο [49], προτείνεται ένας συνδυασμός των μεταβολικών μεθόδων όπως αυτές που αναφέραμε προηγουμένως με τη στατιστική μοντελοποίηση της εικόνας, και εξάγεται ένα μοντέλο εξέλιξης της καμπύλης που χρησιμοποιεί τόσο γεωμετρικές όσο και στατιστικές ιδιότητες της εικόνας. Ο στόχος είναι να χωριστεί το πεδίο ορισμού  $R$  της εικόνας  $I$  σε περιοχές  $\{R_i, i = 1, 2, 3, \dots, M\}$ , έτσι ώστε  $R = \bigcup_{i=1}^M R_i$ ,  $R_i \cap R_j = \emptyset, i \neq j$  και η εικόνα  $I$  να ικανοποιεί κάποια κριτήρια ομοιογένειας, ως προς κάποιο χαρακτηριστικό, στο εσωτερικό κάθε περιοχής  $R_i$ .

Το πρώτο βήμα στην πρόταση των Zhu και Yuille είναι η στατιστική μοντελοποίηση του εξεταζόμενου χαρακτηριστικού  $F$  (φωτεινότητας, χρώματος, υψής κ.λ.π) εντός κάθε περιοχής με μια απλή παραμετρική κατανομή πιθανότητας  $P(F(x, y) : (x, y) \in R_i | \alpha_i)$ . Οι κατανομές αυτές μπορούν να είναι διαφορετικού τύπου για κάθε περιοχή. Γίνεται η υπόθεση ότι οι κατανομές αυτές είναι ανεξάρτητες μεταξύ τους, οπότε η συνολική κατανομή του χαρακτηριστικού στην εικόνα μπορεί να βρεθεί από τη σχέση:

$$\log \{P(I(x, y) | C, \alpha_1, \alpha_2, \dots, \alpha_M)\} = \sum_{i=1}^M \iint_{R_i} \log \{P(F | \alpha_i)\} \quad (2.17)$$

όπου  $R = \bigcup_{i=1}^M C_i$  είναι το σύνολο των ορίων μεταξύ των περιοχών και  $C_i = \partial R_i$  το (εξελισσόμενο) σύνορο της περιοχής  $i$ .

Ο αλγόριθμος κατόπιν υλοποιείται με δύο εναλλασσόμενα μεταξύ τους βήματα. Το πρώτο βήμα (μετά την αρχικοποίηση των περιοχών), είναι η εξέλιξη των ορίων των περιοχών με στόχο την ελαχιστοποίηση κάποιας ενέργειας, όπως και στα ενεργά περιγράμματα. Η ενέργεια που χρησιμοποιήθηκε εδώ ήταν:

$$E[C, \{\alpha_i\}] = \sum_{i=1}^M \left\{ \frac{\mu}{2} \int_{\partial R_i} ds - \log [P(F(x, y) : (x, y) \in R_i | \alpha_i)] + \lambda \right\} \quad (2.18)$$

Ο πρώτος όρος μέσα στα άγκιστρα είναι το μήκος του συνόρου  $\partial R_i$  της περιοχής  $i$ . Επειδή κάθε περιοχή μοιράζεται ένα τμήμα του συνόρου τους με κάποια άλλη περιοχή,

το συνολικό άθροισμα διαιρείται δια δύο. Ο δεύτερος όρος είναι το «κόστος» της κωδικοποίησης του χαρακτηριστικού εντός κάθε περιοχής με την αντίστοιχη στατιστική κατανομή, και ο τρίτος όρος είναι μια σταθερά κοινή για όλες τις περιοχές. Η εξέλιξη του συνόρου κάθε επιφάνειας προς την ελαχιστοποίηση αυτού του συναρτησιακού προκύπτει ότι διέπεται από την παρακάτω διαφορική εξίσωση:

$$\begin{aligned} \frac{\partial \vec{C}_i}{\partial t} &= -\mu\kappa_i \vec{N}_i + \log\left(\frac{P(F|\alpha_i)}{P(F|\alpha_j)}\right) \vec{N}_i \\ \frac{\partial \vec{C}_j}{\partial t} &= -\frac{\partial \vec{C}_i}{\partial t} \end{aligned} \quad (2.19)$$

όπου  $R_i, R_j$  είναι γειτνιαζουσες περιοχές, και  $N_i$  είναι το μοναδιαίο, κάθετο, και με κατεύθυνση προς το εξωτερικό της καμπύλης  $C_i$  διάνυσμα. Όπως φαίνεται από τις εξισώσεις (2.19), ο πρώτος όρος εξαρτάται από την καμπυλότητα και ευνοεί την ελαχιστοποίηση του μήκους της καμπύλης, με συνέπεια το αποτέλεσμα της κατάτμησης να είναι μια ομαλή καμπύλη, και ο δεύτερος αποτελεί τον όρο που καθορίζει τον «ανταγωνισμό» των γειτονικών περιοχών και οδηγεί στη μεγιστοποίηση της πιθανότητας  $P(I|C, \{\alpha_i\})$ . Αν  $P(F|\alpha_i) > P(F|\alpha_j)$  ο όρος αυτός είναι θετικός και οδηγεί στην επέκταση του συνόρου  $\partial R_i$  και την ενσωμάτωση των εικονοστοιχείων στην περιοχή  $i$ , ενώ αντίθετο αποτέλεσμα έχουμε αν  $P(F|\alpha_i) < P(F|\alpha_j)$ .

Μετά την ολοκλήρωση του βήματος της εξέλιξης και τον επαναπροσδιορισμό των συνόρων πραγματοποιείται εκ νέου μια βέλτιστη εκτίμηση των παραμέτρων  $\{\alpha_i\}$ . Τα δύο βήματα επαναλαμβάνονται μέχρι τη σύγκλιση. Η μέθοδος αυτή έχει δώσει πολύ καλά αποτελέσματα στην κατάτμηση πολύπλοκων εικόνων.

### Γεωδαιτικές Ενεργές Περιοχές

Οι Paragios και Deriche στο [37] εξέλιξαν το μοντέλο του ανταγωνισμού περιοχών για να συμπεριλάβουν και πληροφορία σχετικά με τις ακμές της εικόνας. Για απλότητα μπορούμε να θεωρήσουμε ότι η εικόνα αποτελείται από δύο μόνο περιοχές, και ότι ζητείται ο προσδιορισμός του διαχωρισμού  $\mathcal{P}(R) = \{R_A, R_B\}$ . Πραγματοποιείται η υπόθεση ότι υπάρχει διαθέσιμη πληροφορία για τις αναμενόμενες ιδιότητες των δύο περιοχών, που εκφράζονται μέσω των  $p_{r,A}(\vec{r})$ ,  $p_{r,B}(\vec{r})$  που δίνουν την πιθανότητα το εικονοστοιχείο στη θέση  $\vec{r}$  να διατηρεί τις αναμενόμενες ιδιότητες των περιοχών A και B. Επίσης θεωρούμε ότι υπάρχει και πληροφορία σχετικά με τις αναμενόμενες θέσεις των συνόρων μεταξύ των περιοχών, η οποία κωδικοποιείται μέσω των  $p_{b,A}(\vec{r})$ ,  $p_{b,B}(\vec{r})$  που δίνουν την πιθανότητα το εικονοστοιχείο στη θέση  $\vec{r}$  να ανήκει στο σύνορο της επιφάνειας A και B αντίστοιχα.

Το συναρτησιακό ενέργειας προς ελαχιστοποίηση θα αποτελείται από δύο όρους. Ο πρώτος όρος οδηγεί στην ακριβή εξαγωγή των συνόρων μεταξύ των δύο περιοχών, όπως προβλέπεται από τα γεωδαιτικά ενεργά περιγράμματα:

$$E(\partial R) = \sum_{X \in \{A, B\}} \int_0^1 g(p_{b, X}(I(\partial R_X(c_X)))) \left| \partial \dot{R}_X(c_X) \right| dc_X \quad (2.20)$$

Ο δεύτερος όρος είναι στατιστικής φύσεως και αποσκοπεί στην εύρεση της βέλτιστης διαμέρισης της εικόνας  $\mathcal{P}(R)$  που επιτυγχάνει τη μεγιστοποίηση της a posteriori πιθανοφάνειας, δεδομένης της εικόνας. Συγκεκριμένα, αν υποθέσουμε ότι  $p_S(\mathcal{P}(R)|I)$  είναι η a posteriori κατανομή πυκνότητας πιθανότητας των διαφόρων διαμερίσεων  $\mathcal{P}(R)$  με δεδομένη την εικόνα  $I$ , τότε θα ισχύει η σχέση:

$$p_S(\mathcal{P}(R)|I) = \frac{p(I|\mathcal{P}(R))}{p(I)} p(\mathcal{P}(R)) \quad (2.21)$$

όπου  $p(I|\mathcal{P}(R))$  είναι η a posteriori πιθανότητα κατάτμησης για την εικόνα  $I$ , δεδομένης της διαμέρισης  $\mathcal{P}(R)$ ,  $p(\mathcal{P}(R))$  η πιθανότητα της διαμέρισης  $\mathcal{P}(R)$  ανάμεσα σε όλες τις δυνατές διαμερίσεις του πεδίου της εικόνας, και  $p(I)$  η πιθανότητα εμφάνισης της εικόνας  $I$  ανάμεσα σε όλες τις δυνατές εικόνες. Αν υποθέσουμε ότι όλες οι διαμερίσεις είναι εξίσου πιθανές,  $p(\mathcal{P}(R)) = 1/Z$ , όπου  $Z$  ο συνολικός αριθμός δυνατών διαμερίσεων, τότε μπορούμε να αγνοήσουμε τους σταθερούς όρους  $p_S(\mathcal{P}(R)|I)$ ,  $p(\mathcal{P}(R))$ , και να πάρουμε

$$p_S(\mathcal{P}(R)|I) = p(I|\{R_A, R_B\}) = p\left([I|R_A] \cap [I|R_B]\right) = p(I|R_A) p(I|R_B) \quad (2.22)$$

όπου  $p(I|R_X)$  η a posteriori πιθανότητα για την περιοχή  $X$  δεδομένης της εικόνας. Στην τελευταία εξίσωση έχουμε υποθέσει ότι δεν υπάρχει συσχέτιση στην απόδοση του εικονοστοιχείου στη μία ή την άλλη περιοχή, και οι πιθανότητες για κάθε περιοχή εξαρτώνται μόνο από το σύνολο παρατηρήσεων εντός της περιοχής αυτής.

Επιπλέον, αν υποθέσουμε ότι τα εικονοστοιχεία εντός κάθε περιοχής είναι ανεξάρτητα, τότε μπορούμε να αντικαταστήσουμε την πιθανότητα της περιοχής από την από κοινού πιθανότητα των εικονοστοιχείων της περιοχής:

$$p(I|R_X) = \prod_{s \in R_X} p_X(I(s)) \quad (2.23)$$

οπότε τελικά η πιθανότητα της διαμέρισης γράφεται:

$$p_S(\mathcal{P}(R)|I) = \prod_{s \in R_A} p_A(I(s)) \prod_{s \in R_B} p_B(I(s)) \quad (2.24)$$



και το συναρτησιακό ενέργειας, που θα εξασφαλίζει την μεγιστοποίηση αυτής της πιθανότητας θα ισούται με:

$$E(\partial\mathcal{P}(R)) = - \iint_{R_A} \log [p_A(I(x, y))] dx dy - \iint_{R_B} \log [p_B(I(x, y))] dx dy \quad (2.25)$$

Το συναρτησιακό αυτό μπορεί να ερμηνευθεί ως εξής:

- Έστω ότι το εικονοστοιχείο  $s$  έχει ταξινομηθεί σωστά και η πραγματική του θέση είναι στην περιοχή  $A$ . Τότε η πιθανότητα της πραγματικής θέσης (σωστής επιλογής) είναι  $p_{r,A}(I(s))$ , και θα είναι μεγαλύτερη από την πιθανότητα της λανθασμένης επιλογής  $p_{r,B}(I(s))$ , με συνέπεια η συμβολή του εικονοστοιχείου στο συναρτησιακό να είναι μικρή:  $-\log(p_{r,A}(I(s)))$
- Έστω ότι το εικονοστοιχείο  $s$  έχει ταξινομηθεί λάθος και η πραγματική του θέση είναι στην περιοχή  $B$ , αλλά αυτό έχει ταξινομηθεί στην περιοχή  $A$ . Τότε προφανώς η συμβολή του στο συναρτησιακό θα είναι μεγαλύτερη απ' ό,τι στην περίπτωση που θα είχε γίνει σωστή ταξινόμηση

Τελικά δηλαδή το ολικό συναρτησιακό ενέργειας για το μοντέλο των γεωδαιτικών ενεργών περιοχών (geodesic active regions - GAR) είναι το εξής:

$$E(\partial\mathcal{P}(R)) = \sum_{X \in \{A, B\}} (1 - \alpha) \underbrace{\int_0^1 \overbrace{g[p_{b,X}(I(\partial R_X(c_X)))]}^{\text{έλξη προς το σύνορο}} \overbrace{|\partial \dot{R}_X(c_X)|}^{\text{ομαλότητα συνόρου}} dc_X}_{\text{όρος συνόρου της περιοχής } X} - \sum_{X \in \{A, B\}} \alpha \underbrace{\iint_{R_X} \log [p_X(I(x, y))] dx dy}_{\text{όρος περιοχής } X} \quad (2.26)$$

όπου  $\alpha$  είναι μια θετική σταθερά στο διάστημα  $[0, 1]$  που ελέγχει το ποσοστό επίδρασης κάθε όρου. Με το συναρτησιακό αυτό εξασφαλίζεται ότι:

- Οι καμπύλες που διαχωρίζουν τις περιοχές είναι ομαλές, ελάχιστου μήκους, και αντιστοιχούν στα πραγματικά σύνορα μεταξύ των αντικειμένων.
- Ο διαχωρισμός σε περιοχές είναι αυτός που μεγιστοποιεί την a posteriori πιθανότητα διαμέρισης.

Αν χρησιμοποιήσουμε κατάβαση κλίσης (gradient descent) για τη βελτιστοποίηση, θα πάρουμε της εξής εξισώσεις εξέλιξης GAR για τα σύνορα  $C_A$ ,  $C_B$ :

$$\begin{aligned} \frac{\partial \vec{C}_A}{\partial t} &= \alpha \log \left[ \frac{p_{r,B}(I(\vec{C}_A))}{p_{r,A}(I(\vec{C}_A))} \right] \vec{N}_A(\vec{C}_A) \\ &\quad + (1 - \alpha) \left[ g \left( p_{b,A}(I(\vec{C}_A)) \right) \kappa_A(\vec{C}_A) \right. \\ &\quad \left. - \nabla g \left( p_{b,A}(I(\vec{C}_A)) \right) \cdot \vec{N}_A(\vec{C}_A) \right] \vec{N}_A(\vec{C}_A) \end{aligned} \quad (2.27)$$

$$\begin{aligned} \frac{\partial \vec{C}_B}{\partial t} &= \alpha \log \left[ \frac{p_{r,A}(I(\vec{C}_B))}{p_{r,B}(I(\vec{C}_B))} \right] \vec{N}_B(\vec{C}_B) \\ &\quad + (1 - \alpha) \left[ g \left( p_{b,B}(I(\vec{C}_B)) \right) \kappa_B(\vec{C}_B) \right. \\ &\quad \left. - \nabla g \left( p_{b,B}(I(\vec{C}_B)) \right) \cdot \vec{N}_B(\vec{C}_B) \right] \vec{N}_B(\vec{C}_B) \end{aligned} \quad (2.28)$$

Ο αλγόριθμος GAR μπορεί να γενικευθεί και να υλοποιηθεί και αυτός με χρήση επιπεδοσυνόλων. Στην περίπτωση αυτή η εξίσωση εξέλιξης των τρισδιάστατων επιφανειών  $u_i$ , μια για κάθε περιοχή, δίνεται από τη σχέση:

$$\begin{aligned} \frac{\partial u_i}{\partial t}(\vec{C}_i) &= \underbrace{\alpha (r_i(\vec{C}_i) - r_{o_i}(\vec{C}_i)) \left| \nabla u_i(\vec{C}_i) \right|}_{\text{δυναμη περιοχης } i} \\ &\quad + (1 - \alpha) \underbrace{\left[ b_i(\vec{C}_i) \kappa_i(\vec{C}_i) + \nabla b_i(\vec{C}_i) \cdot \frac{\nabla u_i(\vec{C}_i)}{\left| \nabla u_i(\vec{C}_i) \right|} \right] \left| \nabla u_i(\vec{C}_i) \right|}_{\text{δυναμη συνορου } i} \end{aligned} \quad (2.29)$$

όπου  $r_i$ ,  $r_{o_i}$  είναι συναρτήσεις που ενσωματώνουν τα χαρακτηριστικά των συνόρων για την περιοχή  $i$  και το σύνολο των περιοχών  $\{j\}$ ,  $j \neq i$  αντίστοιχα (εν προκειμένω η συνάρτηση  $g$ ), και  $b_i$ ,  $b_{o_i}$  συναρτήσεις που ενσωματώνουν τα χαρακτηριστικά του εσωτερικού της περιοχής  $i$  και του συνόλου των περιοχών  $\{j\}$ ,  $j \neq i$ .

Οι γεωδαιτικές ενεργές περιοχές μπορούν να επεκταθούν για χρήση σε πλειάδα εφαρμογών, όπως η κατάτμηση με χρήση υφής και η ανίχνευση και παρακολούθηση κινούμενων αντικειμένων. Χρήση αυτών θα γίνει και στην παρούσα εργασία.

### Πρότερα Σχήματα

Το μοντέλο των γεωδαιτικών περιγραμμάτων, όπως αναπτύχθηκε στην προηγούμενη ενότητα, δεν χρησιμοποιεί κάποια πρότερη πληροφορία για το είδος του αντι-

κειμένου που αναζητείται. Σε αρκετές περιπτώσεις είναι αρκετά χρήσιμο να χρησιμοποιούμε κάποια πληροφορία για το σχήμα των αντικειμένων που αναζητούνται, ώστε ο εντοπισμός τους να γίνεται με μεγαλύτερη ασφάλεια, και η διαδικασία της κατάτμησης να γίνει πιο εύρωστη, για παράδειγμα να ανιχνεύει τα αντικείμενα και σε περίπλοκα περιβάλλοντα, ή να αντιμετωπίζει περιπτώσεις όπου το ζητούμενο αντικείμενο επικαλύπτεται μερικώς στην εικόνα από κάποιον άλλο.

Μια θεωρία που μπορεί να χρησιμοποιηθεί προς αυτήν την κατεύθυνση, είναι αυτή των πρότερων σχημάτων, η οποία μπορεί και να συνδυαστεί με το μοντέλο GAC. Συγκεκριμένα, θεωρούμε ότι έχουμε διαθέσιμο ένα πρότυπο σχήμα, το οποίο αναζητούμε στην εικόνα. Το πρότυπο αυτό σχήμα μπορεί, για παράδειγμα, να έχει προκύψει από στατιστική επεξεργασία ενός συνόλου εκπαίδευσης. Συνήθως το σχήμα είναι διαθέσιμο όχι με τη μορφή κάποιας δυαδικής εικόνας, αλλά με την μορφή του προσημασμένου μετασχηματισμού απόστασης  $\Phi$  που προκύπτει από τη δυαδική εικόνα του περιγράμματος του σχήματος, με  $\Phi < 0$  στο εσωτερικό του σχήματος και  $\Phi > 0$  στο εξωτερικό.

Η μέθοδος των πρότερων σχημάτων μπορεί να χρησιμοποιηθεί σε συνδυασμό με τη μέθοδο των γεωδαιτικών ενεργών περιγραμμάτων, με τη δημιουργία μιας εξωτερικής δύναμης που να έλκει την επιφάνεια όπως την κατεύθυνση του προγενέστερου σχήματος.

Στο [28], το πρότερο σχήμα που αναζητείται στην εικόνα σε κάθε χρονική στιγμή όπως εξέλιξης όπως επιφάνειας στη μέθοδο GAC, που έχει προκύψει από ανάλυση πρωτευουσών συνιστωσών (PCA), παριστάνεται από ένα άθροισμα με βάρη, όπως προκύπτει από τη σχέση :

$$\Phi(w, p) = \bar{\Phi}(p) + \sum_{i=1}^k w_i \Phi_i(p) \quad (2.30)$$

Οι παράμετροι  $w$  δίνουν πληροφορία για το σχήμα της τρέχουσας καμπύλης, καθώς αποτελούν τους συντελεστές με τους οποίους κάθε πρωτεύον «σχήμα» της PCA συμμετέχει στο τελικό σχήμα, ενώ οι παράμετροι  $p$  για τη θέση του σχήματος αυτού σε σχέση με το γνωστό πρότερο σχήμα. Οι συγγραφείς πρότειναν σε κάθε βήμα να πραγματοποιείται μια πιθανοτική εκτίμηση της θέσης και του σχήματος της τελικής καμπύλης  $u^*$ , που προκύπτουν από της παραμέτρους  $p_{MAP}$  και  $w_{MAP}$ , με βάση τη μεγιστοποίηση της πιθανότητας:

$$\langle w_{MAP}, p_{MAP} \rangle = \arg \max_{w, p} P(w, p | u, \nabla I) \quad (2.31)$$

και συνεπώς της σχέσης:

$$u_{MAP}^* = \arg \max_{u^*} P(u^* | u, \nabla I) \quad (2.32)$$

όπου  $u$  είναι η εξελισσόμενη καμπύλη την τρέχουσα χρονική στιγμή και  $I$  η φωτεινότητα της εικόνας. Η εκτίμηση γίνεται βάσει ορισμένων λογικών παραδοχών, και χρησιμοποιούνται στη μέθοδο GAC με βάση τη σχέση:

$$u(t+1) = u(t) + \lambda_1 \underbrace{[g(c + \kappa) \|\nabla u(t)\| + \nabla u \cdot \nabla g]}_{\text{κλασική εξέλιξη GAC}} + \lambda_2 \underbrace{[u^*(t) - u(t)]}_{\text{εξέλιξη με βάση το σχήμα}} \quad (2.33)$$

όπου  $\lambda_1, \lambda_2$  παράμετροι που δίνουν κατάλληλα βάρη στις συνιστώσες εξέλιξης, με βάση το πόση εμπιστοσύνη έχουμε στο πρότυπο σχήμα που κατασκευάσαμε.

Η προσέγγιση αυτή είχε το μειονέκτημα ότι δε λάμβανε υπόψιν τυχόν μετασχηματισμούς και διαφοροποιήσεις που μπορεί να έχει υποστεί το πρότερο σχήμα στην εικόνα (οι παράμετροι  $p$  θεωρούνταν ότι ακολουθούν ομοιόμορφη κατανομή). Ωστόσο η ενσωμάτωση του μοντέλου GAC την καθιστούσε ιδιαίτερα εύρωστη, και έδινε τη δυνατότητα εντοπισμού του αντικειμένου ακόμα και αν αυτό παρουσίαζε μικρές αποκλίσεις από την πρότυπή του μορφή.

Αντίθετα, οι Paragios και Rousson στο [38] χρησιμοποιούν μια διαφορετική προσέγγιση. Συγκεκριμένα, έχοντας το πρότερο σχήμα  $\Phi_M$  και την εξελισσόμενη επιφάνεια  $\Phi$  αναζητείται σε κάθε χρονική στιγμή ο μετασχηματισμός  $A$  (περιστροφή κατά  $\theta$  - μετατόπιση κατά  $[T_x, T_y]$  - αλλαγή κλίμακας κατά  $s$ ) που ελαχιστοποιεί το συναρτησιακό ενέργειας:

$$E = \iint_{\Omega} (s\Phi - \Phi_M(A))^2 d\Omega \quad (2.34)$$

Οι παράμετροι του μετασχηματισμού υπολογίζονται με χρήση του αλγορίθμου κατάβασης κλίσης (gradient descent), και η εξέλιξή της προκύπτει ότι καθορίζεται από τις σχέσεις:

$$\frac{\partial s}{\partial t} = 2 \iint_{\Omega} (s\Phi - \Phi_M(A)) \left( \nabla \Phi_M(A) \frac{\partial A}{\partial s} - \Phi \right) d\Omega \quad (2.35)$$

$$\frac{\partial \theta}{\partial t} = 2 \iint_{\Omega} (s\Phi - \Phi_M(A)) \left( \nabla \Phi_M(A) \frac{\partial A}{\partial \theta} \right) d\Omega \quad (2.36)$$

$$\frac{\partial}{\partial t} \begin{bmatrix} T_x \\ T_y \end{bmatrix} = 2 \iint_{\Omega} (s\Phi - \Phi_M(A)) \left( \nabla \Phi_M(A) \frac{\partial A}{\partial \begin{bmatrix} T_x \\ T_y \end{bmatrix}} \right) d\Omega \quad (2.37)$$

ενώ η εξέλιξη της επιφάνειας ακολουθεί το νόμο εξέλιξης:

$$\frac{d\Phi}{dt} = -2s(s\Phi - \Phi_M(A)) \quad (2.38)$$

Ενσωματώνοντας το μοντέλο αυτό στη μέθοδο GAC παίρνουμε τη σχέση:

$$\frac{d\Phi}{dt} = \left[ \operatorname{div} \left( g(I) \frac{\nabla\Phi}{\|\nabla\Phi\|} \right) - \beta g(I) \right] \|\nabla\Phi\| + \gamma s(s\Phi - \Phi_M(A)) \quad (2.39)$$

όπου συμμετέχουν μια δύναμη για την ανίχνευση των ακμών, μια δύναμη μπαλονιού, και μια δύναμη εντοπισμού του πρότερου σχήματος.

## 2.3 Τεχνικές και Αποτελέσματα Κατάτμησης σε Βίντεο Νοηματικής Γλώσσας

Η κατάτμηση στην περίπτωση εικόνων που έχουν εξαχθεί από βίντεο νοηματικής γλώσσας αποσκοπεί στον εντοπισμό και εν τέλει την παρακολούθηση των χειρών του νοηματιστή. Ως μέθοδος κατάτμησης χρησιμοποιήθηκε η μέθοδος των ενεργών περιγραμμάτων/ περιοχών και επιπεδοσυνόλων, που αναπτύχθηκε διεξοδικά στις προηγούμενες ενότητες, και όπως είδαμε έχει μεγάλη ευελιξία και μπορεί να χρησιμοποιηθεί, κατάλληλα προσαρμοσμένη, σε ένα μεγάλο αριθμό από διαφορετικές εφαρμογές. Ένα ιδιαίτερα σημαντικό χαρακτηριστικό αυτής, όπως προαναφέραμε, είναι ότι επιτρέπει την ενσωμάτωση πρότερης πληροφορίας αναφορικά με το σχήμα του αντικειμένου που παρακολουθείται. Τούτο είναι πολύ σημαντικό στην περίπτωσή μας γιατί προσδίδει στο σύστημα επιπλέον αξιοπιστία και αντοχή έναντι θορύβου ή απόκρυψης. Επίσης, η μέθοδος έχει το πλεονέκτημα ότι μπορεί να αξιοποιήσει στατιστική πληροφορία σχετικά με τις ιδιότητες των αντικειμένων που αναζητούμε στην εικόνα.

### 2.3.1 Κατάτμηση με χρήση δύναμης χρώματος και GACs

Στην παρούσα εργασία ήταν επιθυμητό η μέθοδος κατάτμησης να συνδυάζει την πληροφορία του χρώματος με την πληροφορία των ακμών. Η πληροφορία του χρώματος είναι πολύ σημαντική στην περίπτωσή μας, όπως και στις περισσότερες εφαρμογές όπου απαιτείται η ανίχνευση τμημάτων του ανθρώπινου σώματος σε εικόνες ή βίντεο (όπως ανίχνευση και αναγνώριση προσώπου ή κινούμενου ομιλητή), καθώς το ανθρώπινο δέρμα έχει μια αρκετά ομοιόμορφη και συγκεκριμένη χρωματική κατανομή που

στις περισσότερες περιπτώσεις το καθιστά ευδιάκριτα διαφορετικό από τον περιβάλλοντα χώρο. Ειδικότερα δε, τα βίντεο που είχαμε στη διάθεσή μας αποτελούνταν από ένα φόντο σε σκούρο χρώμα, που διέφερε σημαντικά από το χρώμα του δέρματος, οι δε ομιλητές φορούν ως επί το πλείστον σκούρα ρούχα, γεγονός που διευκόλυνε περαιτέρω την απομόνωση των περιοχών με χρώμα δέρματος. Όπως θα δούμε και σε επόμενο κεφάλαιο, στην περίπτωση που ανιχνευθεί επικάλυψη μεταξύ των χεριών και/ή του κεφαλιού, θα χρησιμοποιηθεί και πληροφορία κίνησης για να αυξηθεί η ευρωστία του συστήματος.

Με βάση τα παραπάνω, για τον εντοπισμό του χεριού σε μια εικόνα, πρέπει καταρχήν να δημιουργηθεί ένα μοντέλο χρώματος του ανθρώπινου δέρματος ικανό να περιγράψει τις δεδομένες εικόνες, και κατόπιν η πληροφορία από την εφαρμογή του μοντέλου στις εικόνες να χρησιμοποιηθεί για την κατάτμηση της εικόνας. Πρέπει ωστόσο να παρατηρήσουμε ότι η χρήση της πληροφορίας χρώματος αναγκαστικά περιορίζει κάπως το περιβάλλον στο οποίο χειρονομεί ο νοηματιστής, καθώς πρέπει να αποκλειστούν άλλα αντικείμενα με χρώμα παραπλήσιο προς αυτό του δέρματος.

Γενικά, όταν χρησιμοποιείται το χρώμα, κατά κανόνα αποφεύγεται η χρήση της συνιστώσας της φωτεινότητας, έτσι ώστε αφ' ενός η καλή λειτουργία της εφαρμογής να μην εξαρτάται (κατά το δυνατόν) από τον περιβάλλοντα χώρο, και αφ' ετέρου να μην υπάρχει εξάρτηση από τον ομιλητή, καθώς έχει βρεθεί ότι οι χρωματικές διαφορές στο χρώμα του δέρματος διαφορετικών ανθρώπων εντοπίζονται κυρίως στη συνιστώσα της φωτεινότητας. Ως εκ τούτου οι χώροι χρώματος που ενδείκνυνται για τέτοιου είδους εφαρμογές είναι οι YCbCr, HSV, HSL και Lab.

Η διαδικασία εντοπισμού των περιοχών με χρώμα δέρματος βασίζεται στην κατασκευή ενός μοντέλου, το οποίο θα δίνει για κάθε εικονοστοιχείο οποιασδήποτε εικόνας την πιθανότητα το εικονοστοιχείο αυτό να ανήκει σε περιοχή δέρματος. Το μοντέλο αυτό θα «χτιστεί» επάνω σε ένα σύνολο εκπαίδευσης από εικόνες όπου είναι εκ των προτέρων γνωστό ποιες περιοχές είναι δέρμα. Επειδή όπως είπαμε ο χώρος χρώματος RGB, στον οποίο είναι κωδικοποιημένο το βίντεο, δεν είναι κατάλληλος, πρέπει κατ' αρχήν να μετατρέψουμε όλες τις εικόνες που έχουμε διαθέσιμες, αλλά και τις εικόνες εισόδου, στον κατάλληλο χώρο, π.χ. τον Lab.

Ενδεικτικά αναφέρουμε στο σημείο αυτό ότι ο χώρος Lab είναι ένας χώρος που βασίζεται στην αντιθετικότητα των χρωμάτων. Η συνιστώσα  $L$  ( $L \in [0, 100]$ ) αντιστοιχεί στη φωτεινότητα, και οι συνιστώσες  $a$  και  $b$  περιέχουν την πληροφορία για την αντιθετικότητα, με βάση μια μη γραμμική συμπίεση των RGB συνιστωσών. Η συνιστώσα  $a$  συγκεκριμένα κωδικοποιεί τη θέση του χρώματος μεταξύ του πράσινου και του

κόκκινου (ματζέντα) (αρνητικές τιμές του  $a$  αντιστοιχούν στο πράσινο ενώ θετικές στο ματζέντα) και η  $b$  τη θέση του χρώματος μεταξύ μπλε και κίτρινου (ομοίως). Το χρωματικό μοντέλο Lab δεν ορίζει σε απόλυτες τιμές το κάθε χρώμα, αλλά τις συναρτά ως προς το "λευκό σημείο" του RGB χώρου, δηλαδή εκείνο το διάνυσμα  $(R, G, B)$  το οποίο οπτικά αντιστοιχεί στο λευκό χρώμα. Επακριβής καθορισμός του σημείου αυτού συνεπάγεται και απόδοση των ακριβών αριθμητικών τιμών στα χρώματα με βάση τον Lab. Επίσης, ο χώρος Lab είναι ικανός να περιγράψει το σύνολο των χρωμάτων που είναι ορατά στο ανθρώπινο μάτι, κάτι που δε συμβαίνει π.χ. με τον RGB).

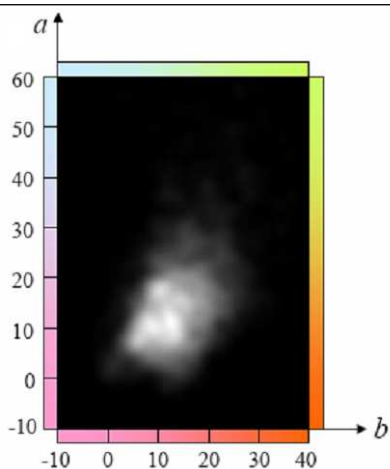
Για την απόκτηση του μοντέλου χρώματος αρχικά απομονώνονται αρκετές περιοχές με χρώμα δέρματος από τις εικόνες που έχουμε διαθέσιμες (μπορεί να χρησιμοποιηθεί και ένα ευρύτερο σύνολο εικόνων εκπαίδευσης ώστε να μπορεί να επεκταθεί η μέθοδος και σε περισσότερους ομιλητές). Για κάθε περιοχή που απομονώθηκε εφαρμόστηκαν τα παρακάτω βήματα:

- Τα εικονοστοιχεία (pixels) της περιοχής μετατρέπονται από τον χρωματικό χώρο RGB στο χώρο Lab.
- Η συνιστώσα φωτεινότητας αγνοείται, ενώ οι συνιστώσες  $a$  και  $b$  αποθηκεύονται σε έναν πίνακα συσσώρευσης, όπου σε κάθε ζεύγος  $(a,b)$  αντιστοιχεί μια τιμή ίση με το συνολικό αριθμό των εμφανίσεων αυτού του ζεύγους σε όλα τα εικονοστοιχεία όλων των εξεταζόμενων περιοχών δέρματος.

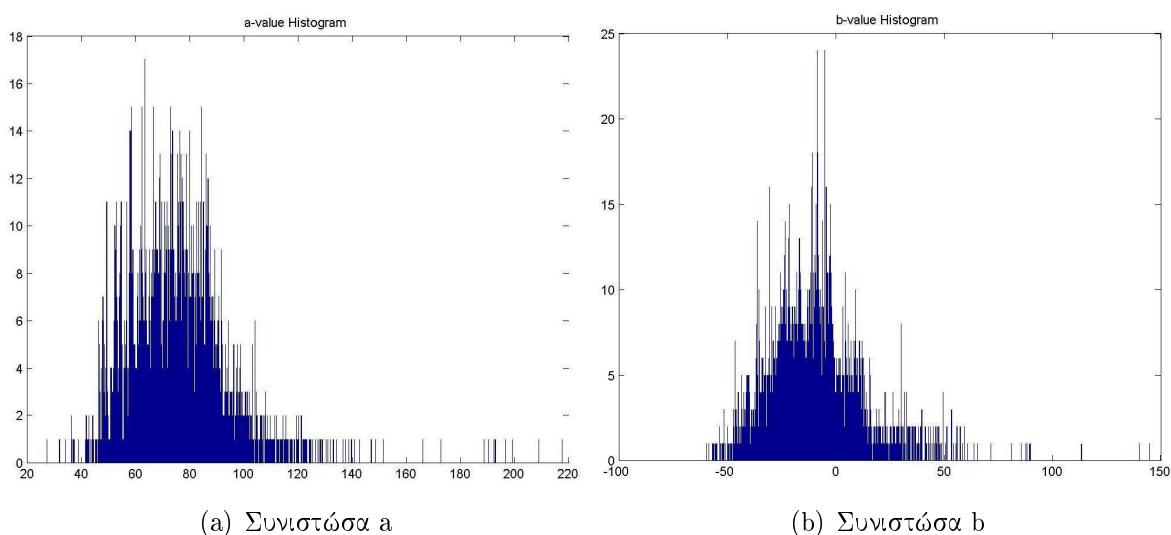
Στη συνέχεια, ο πίνακας συσσώρευσης συνελίσσεται με μια γκαουσιανή συνάρτηση για ομαλοποίηση. Τελικά αυτό που παίρνουμε είναι ένας πίνακας που δίνει κατά προσέγγιση (βάσει του συνόλου εκπαίδευσης που χρησιμοποιήσαμε) την πιθανότητα ένα συγκεκριμένο ζεύγος συνιστωσών  $(a,b)$  να αντιστοιχεί σε εικονοστοιχείο δέρματος. Ένας τέτοιος πίνακας παρουσιάζεται στο Σχ. 2.2

Έχει βρεθεί και επιβεβαιώθηκε και εδώ ότι οι τιμές των  $a$  και  $b$  για το ανθρώπινο δέρμα κυμαίνονται εντός μιας μικρής περιοχής, όπως φαίνεται και από τα ιστογράμματα του Σχ. 2.3.

Με βάση το μοντέλο που κατασκευάσαμε μπορούμε για οποιαδήποτε εικόνα εισόδου να λάβουμε την πιθανότητα ένα εικονοστοιχείο αυτής να ανήκει σε περιοχή δέρματος, αρκεί πρώτα να μετασχηματίσουμε το εικονοστοιχείο αυτό στο χώρο Lab. Η πιθανότητα για κάθε pixel λαμβάνεται από τον παραπάνω πίνακα συσσώρευσης (διδιάστατη κατανομή πιθανότητας) για τις τιμές  $(a,b)$  του pixel. Η εικόνα πιθανότητας που προκύπτει θα είναι καθαρότερη αν σε αυτήν εφαρμοστεί φιλτράρισμα με



Εικόνα 2.2: Το μοντέλο χρώματος



Εικόνα 2.3: Ιστογράμματα τιμών των συνιστωσών a και b

median φίλτρο για απαλοιφή τυχόν τοπικού θορύβου. Κάποια ενδεικτικά αποτελέσματα παρουσιάζονται στο παρακάτω Σχ. 2.4. Βλέπουμε ότι το μοντέλο κατορθώνει να εντοπίσει τις περιοχές δέρματος και σε διαφορετικούς ομιλητές.

Η τελική μέθοδος για την κατάτμηση της εικόνας αποτελεί ουσιαστικά εφαρμογή της μεθόδου των γεωδαιτικών ενεργών περιοχών (η γενική μορφή της οποίας αναφέρθηκε στο τέλος της ενότητας 2.2.4), όπου A είναι η περιοχή που αντιστοιχεί στα χέρια και το κεφάλι του ομιλητή, και B το υπόλοιπο μέρος της εικόνας. Οι δύο περιοχές διαχωρίζονται με χρήση πληροφορίας χρώματος. Συγκεκριμένα, η πιθανότητα



$P_A(\vec{x}) = P_{skin}(\vec{x})$  το εικονοστοιχείο  $\vec{x}$  να ανήκει σε περιοχή δέρματος προκύπτει με βάση ένα μοντέλο χρώματος, η κατασκευή του οποίου εξηγείται στην αμέσως επόμενη ενότητα. Η πιθανότητα  $P_B(\vec{x}) = P_{background}(\vec{x})$  προσδιορίζεται από τη σχέση  $P_{background}(\vec{x}) = 1 - P_{skin}(\vec{x})$ . Εφ'όσον λοιπόν έχουμε για κάθε εικονοστοιχείο την πιθανότητα να ανήκει σε περιοχή δέρματος, μπορούμε να εφαρμόσουμε την μέθοδο των γεωδαιτικών ενεργών περιοχών (GAR). Όπως είδαμε η γενική εξίσωση των γεωμετρικών ενεργών περιγραμμάτων είναι η εξής:

$$\frac{\partial u}{\partial t} = \left[ \text{div} \left( g(I) \frac{\nabla u}{\|\nabla u\|} \right) + F(u) \right] \|\nabla u\| \quad (2.40)$$

Εδώ θα χρησιμοποιήσουμε

$$F(u) = \log \left( \frac{P_{skin}(\vec{x})}{1 - P_{skin}(\vec{x})} \right) + cg(I)$$

δηλαδή μια region-based στατιστική δύναμη χρώματος και μια δύναμη μπαλονιού. Οι δυνάμεις αυτές αθροίζονται με τα κατάλληλα βάρη για να δώσουν την τελική εξίσωση της επιφάνειας  $u$ :

$$\frac{\partial u}{\partial t} = \left[ c_{gac} \text{div} \left( g(I) \frac{\nabla u}{\|\nabla u\|} \right) + c_{color} \log \left( \frac{P_{skin}(\vec{x})}{1 - P_{skin}(\vec{x})} \right) + c_{balloon} g(I) \right] \|\nabla u\| \quad (2.41)$$

Η πιθανότητα ένα εικονοστοιχείο να βρίσκεται σε περιοχή δέρματος προσδιορίζεται από το μοντέλο χρώματος, όπως εξηγήθηκε στην προηγούμενη ενότητα. Η παραπάνω εξίσωση διασφαλίζει ότι η καμπύλη διαχωρισμού θα «σταθεί» σε περιοχές όπου εμφανίζονται οξείες ακμές, και στις διαχωριστικές γραμμές μεταξύ περιοχών δέρματος και φόντου. Όλες οι δυνάμεις (επιμέρους όροι της παραπάνω σχέσης) είναι κανονικοποιημένες στο διάστημα  $[0,1]$  και το ποσοστό συμμετοχής τους στην συνολική δύναμη ελέγχεται από τους συντελεστές  $c_{gac}$ ,  $c_{balloon}$  και  $c_{color}$ .

Σε ότι αφορά την αριθμητική υλοποίηση, για το μοντέλο GAR έχει χρησιμοποιηθεί γρήγορη multigrid υλοποίηση. Πρόκειται για έναν αριθμητικό αλγόριθμο που χρησιμοποιεί μια ιεραρχική διάταξη από πλέγματα διαφορετικού μεγέθους και λεπτομέρειας για την ταχεία επίλυση αριθμητικών προβλημάτων (για λεπτομέρειες βλ. [16]). Αποτελεί ένα συνδυασμό implicit - explicit αριθμητικών μεθόδων για την επίλυση μερικών διαφορικών εξισώσεων και εμφανίζει πολύ καλή ευστάθεια ακόμα και για πολύ μεγάλα χρονικά βήματα. Συνεπώς μπορούμε να χρησιμοποιούμε ελεύθερα μεγάλο χρονικό βήμα για την εξέλιξη της καμπύλης, με αποτέλεσμα η διαδικασία της τμηματοποίησης επιτυγχάνεται σε σημαντικά μικρότερο χρονικό διάστημα, της τάξεως των

0.2-0.4 sec. Ο κώδικας παραχωρήθηκε από τους συγγραφείς του [16]. Το αποτέλεσμα της εφαρμογής της εν λόγω μεθόδου είναι η τελική μορφή της επιφάνειας  $u$ , με το εσωτερικό των περιοχών δέρματος να αντιστοιχεί στην περίπτωση  $u > 0$ .

Στην τμηματοποιημένη εικόνα γενικά αναμένουμε να εμφανίζονται το πολύ 3 περιοχές δέρματος, που αντιστοιχούν στο κεφάλι και στα 2 χέρια (στην περίπτωση που υπάρχει επικάλυψη, π.χ. μεταξύ χεριών και κεφαλιού, οι περιοχές θα είναι λιγότερες). Ωστόσο η τμηματοποιημένη εικόνα συνήθως περιέχει έναν αριθμό από περιοχές που έχουν εσφαλμένα ανιχνευθεί ως περιοχές δέρματος, λόγω διακυμάνσεων στο φωτισμό ή λόγω εμφάνισης μικρών ανοιχτόχρωμων περιοχών στην αρχική εικόνα. Για το λόγο αυτό γίνεται απαλοιφή εκείνων των περιοχών (από τις ευρεθείσες) που έχουν εμβαδό μικρότερο από το  $1/10$  του εμβαδού της μεγαλύτερης περιοχής (βλ. Σχ. 2.5(b)), με χρήση area opening. Επίσης απαλείφονται οι περιοχές που έχουν υπερβολικά μακρόστενο σχήμα, και εμφανίζονται κάποιες φορές στα όρια της εικόνας (βλ. Σχ. 2.5(h)). Τα αποτελέσματα της κατάτμησης πριν και μετά την επεξεργασία αυτή παρουσιάζονται στις εικόνες του Σχ. 2.5.

Γενικά το μέρος της κατάτμησης δίνει πολύ καλά αποτελέσματα, και κατορθώνει να ανιχνεύσει τα χέρια και του κεφάλι του ομιλητή πρακτικά σε κάθε περίπτωση, για τα δεδομένα βίντεο.

Εφ' όσον έχει ολοκληρωθεί το στάδιο της τμηματοποίησης, τα χέρια ξεχωρίζονται από το κεφάλι με βάση το εμβαδό τους, αφού το κεφάλι έχει πάντα μεγαλύτερη επιφάνεια (εξαιρούνται οι περιπτώσεις όπου υπάρχει επικάλυψη μεταξύ κάποιου από τα χέρια και του κεφαλιού, για τις οποίες λαμβάνεται ειδική μέριμνα, όπως θα εξηγηθεί παρακάτω). Έτσι τελικά έχουμε διαθέσιμες τις περιοχές της εικόνας όπου βρίσκονται τα χέρια, καθώς και τα αντίστοιχα όρια αυτών, που θα χρησιμοποιηθούν για την εξαγωγή των χαρακτηριστικών σχήματος.

## 2.4 Εικόνες Αποτελεσμάτων

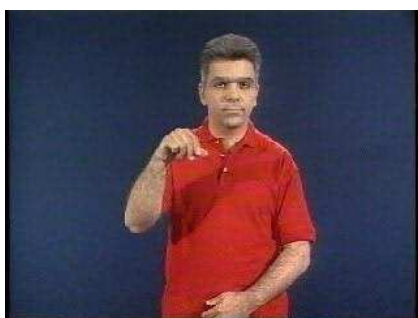
Στις επόμενες σελίδες παρατίθενται τα αποτελέσματα των πειραμάτων του κεφαλαίου 2, με κατάλληλα σχόλια στις λεζάντες όπου χρειάζεται.



(a) Αρχική Εικόνα



(b) Περιοχές δέρματος με την αντίστοιχη πιθανότητα



(c) Αρχική Εικόνα



(d) Περιοχές δέρματος με την αντίστοιχη πιθανότητα



(e) Αρχική Εικόνα



(f) Περιοχές δέρματος με την αντίστοιχη πιθανότητα

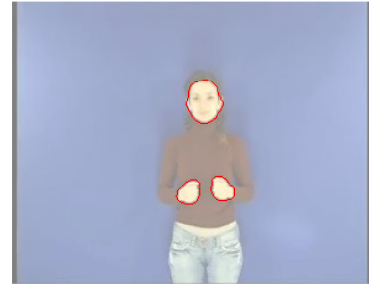
Εικόνα 2.4: Αποτέλεσμα εφαρμογής του μοντέλου χρώματος



(a) Αρχική Εικόνα



(b) Αρχικό αποτέλεσμα τμηματοποίησης



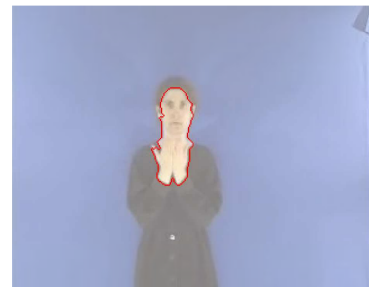
(c) Διορθωμένο αποτέλεσμα



(d) Αρχική Εικόνα



(e) Αρχικό αποτέλεσμα τμηματοποίησης



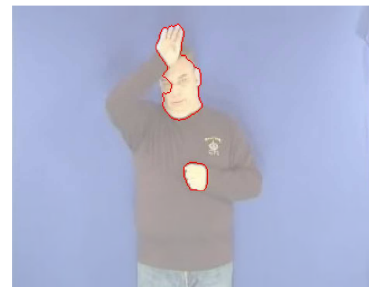
(f) Διορθωμένο αποτέλεσμα



(g) Αρχική Εικόνα



(h) Αρχικό αποτέλεσμα τμηματοποίησης



(i) Διορθωμένο αποτέλεσμα

**Εικόνα 2.5:** Αποτελέσματα της τμηματοποίησης πριν και μετά την επεξεργασία (area opening και αφαίρεση περιοχών μη αποδεκτού σχήματος)

---

## ΠΕΡΙΓΡΑΦΗΤΕΣ ΣΧΗΜΑΤΟΣ

---

Το σχήμα είναι από τα πλέον σημαντικά χαρακτηριστικά για την αναγνώριση ενός αντικειμένου. Ο όρος «σχήμα» μπορεί να αναφέρεται είτε στο εξωτερικό διαστάτο (ή τρισδιάστατο) περίγραμμα του αντικειμένου, ή στη γενική του όψη και διάταξη. Γεωμετρικά, δύο αντικείμενα έχουν το ίδιο σχήμα αν το ένα μπορεί να μετατραπεί στο άλλο με χρήση κάποιου συνδυασμού από μετατοπίσεις, περιστροφές και ομοιόμορφες αλλαγές κλίμακας, με άλλα λόγια το σχήμα είναι όλη εκείνη η γεωμετρική πληροφορία για το αντικείμενο που παραμένει αμετάβλητη από την θέση, την κλίμακα και την περιστροφή του αντικειμένου. Επιπλέον οι άνθρωποι αναγνωρίζουν δύο ίδια σχήματα ακόμα και αν το ένα αποτελεί κατοπτρισμό του άλλου.

Ένα από τα προβλήματα της Όρασης Υπολογιστών είναι και η εύρεση μιας κατάλληλης και ορθής μαθηματικής περιγραφής και μοντελοποίησης του σχήματος. Αυτό γίνεται με εύρεση ενός συνόλου κατάλληλων περιγραφητών ή χαρακτηριστικών σχήματος. Τα χαρακτηριστικά αυτά πρέπει να διατηρούν τις ιδιότητες σε σχέση με το αμετάβλητο που αναφέραμε, και επιπλέον πρέπει να είναι ικανά να κωδικοποιήσουν όλη την πληροφορία του σχήματος σε όσο το δυνατόν μικρότερο χώρο. Στο κεφάλαιο αυτό θα εξετάσουμε αναλυτικά ορισμένους από τους περιγραφητές σχήματος που χρησιμοποιούνται στη βιβλιογραφία, κάποιους από τους οποίους θα χρησιμοποιηθούν και στην παρούσα εργασία.

Γενικά οι περιγραφητές σχήματος μπορούν να χωριστούν σε δύο κατηγορίες: αυτούς που επικεντρώνονται στο περίγραμμα των αντικειμένων (boundary-based) και αυτούς που μελετούν το εσωτερικό της περιοχής που περικλείεται από το περίγραμμα

(region-based). Στην πρώτη κατηγορία ανήκουν π.χ. οι περιγραφητές που βασίζονται στην καμπυλότητα και οι περιγραφητές Fourier, ενώ στη δεύτερη οι κεντρικές ροπές, οι μετασχηματισμοί απόστασης και οι περιγραφητές σκελετού.

### 3.1 Boundary - based Μέθοδοι

Οι μέθοδοι αυτές αποσκοπούν στην ανάλυση του περιγράμματος του σχήματος, το οποίο μοντελοποιείται ως μια επίπεδη καμπύλη  $\Gamma$ , που περιγράφεται από μια συνεχή διανυσματική συνάρτηση  $\vec{C} : [a, b] \rightarrow \mathbb{R}^2$ , όπου  $J = [a, b]$  είναι ένα πεπερασμένο κλειστό διάστημα πραγματικών αριθμών:

$$\Gamma = \{\vec{C}(p) \in \mathbb{R}^2 : p \in [a, b]\}$$

Η συνάρτηση  $\vec{C}$  καλείται «μονοπάτι» της καμπύλης, και αποτελεί μία από τις δυνατές παραμετροποιήσεις αυτής, με παράμετρο  $p$ . Μια μεταβολή στην παραμετροποίηση  $h(p)$  οδηγεί σε ένα διαφορετικό «μονοπάτι»  $\vec{Y}(p) = \vec{C}(h(p))$  το οποίο ωστόσο παριστάνει γεωμετρικά την ίδια καμπύλη  $\Gamma$ . Άρα μπορούμε να θεωρήσουμε ότι κάθε καμπύλη παραμετροποιείται στο διάστημα  $[0, 1]$ . Μια καμπύλη είναι κλειστή αν  $\vec{C}(0) = \vec{C}(1)$ .

Έστω τώρα ότι η συνάρτηση  $\vec{C}(p)$  είναι δύο φορές συνεχώς διαφορίσιμη, και συνεπώς η καμπύλη  $\Gamma$  είναι ομαλή. Τότε η ταχύτητα της κίνησης κατά μήκος της τροχιάς που ορίζει η καμπύλη δίνεται από τη διανυσματική παράγωγο  $d\vec{C}(p)/dp$ , με μέτρο :

$$v(p) \triangleq \left\| \frac{d\vec{C}}{dp} \right\| = \sqrt{\left(\frac{dx_1}{dp}\right)^2 + \left(\frac{dx_2}{dp}\right)^2}$$

και γωνία

$$\theta(p) \triangleq \arctan\left(\frac{y'(p)}{x'(p)}\right)$$

όπου ο τονισμός σημαίνει παραγωγή ως προς  $p$ .

Το μοναδιαίο εφαπτόμενο διάνυσμα στην καμπύλη (παράλληλο στο διάνυσμα της ταχύτητας) δίνεται από τη σχέση

$$\vec{T}(p) \triangleq \frac{\vec{C}'(p)}{\|\vec{C}'(p)\|} = \frac{1}{v(p)} \frac{d\vec{C}(p)}{dp}$$

με  $\vec{T} = \vec{0}$  στις περιοχές όπου  $v=0$ . Αντίστοιχα το μοναδιαίο κάθετο στην καμπύλη διάνυσμα ορίζεται ως:

$$\vec{N}(p) \triangleq \frac{\vec{T}'(p)}{\|\vec{T}'(p)\|}$$

και είναι κάθετο στην καμπύλη διότι  $\vec{T}' \perp \vec{T}$  καθώς  $\|\vec{T}(p)\| = 1$ . Το διάνυσμα αυτό μπορεί να στρέφεται προς το εξωτερικό ή το εσωτερικό της καμπύλης, ανάλογα με την «κατεύθυνση» μεταβολής της γωνίας  $\theta$  της καμπύλης: αν  $\theta'(p) > 0$ , το διάνυσμα  $\vec{N}$  έχει κατεύθυνση προς το εσωτερικό της καμπύλης. Το διάνυσμα  $\vec{U}(p) = \sin(\theta(p))\vec{e}_1 - \cos(\theta(p))\vec{e}_2$ , όπου  $\{\vec{e}_1, \vec{e}_2\}$  το σταθερό σύστημα αναφοράς, είναι το μοναδιαίο κάθετο διάνυσμα με κατεύθυνση προς το εξωτερικό της καμπύλης.

Από όλες τις δυνατές παραμετροποιήσεις μιας (ομαλής) καμπύλης, η πιο σημαντική είναι η παραμετροποίηση βάσει του μήκους τόξου. Το μήκος τόξου ορίζεται για κάθε σημείο της καμπύλης  $(x(p), y(p))$  ως το μήκος του τόξου κατά μήκος της καμπύλης που συνδέει το εν λόγω σημείο με το αρχικό σημείο  $\vec{C}(0)$ . Η σχέση που δίνει το μήκος τόξου σε κάθε σημείο είναι:

$$s(p) = \int_0^p v(q) dq \Rightarrow s'(p) = v(p)$$

Το ολικό μήκος της καμπύλης  $L$  ισούται με  $s(1)$ , δηλαδή  $\int_0^L ds = \int_0^1 v(p) dp$ . Από την τελευταία σχέση προκύπτει ότι το μήκος τόξου παραμένει αμετάβλητο από τυχόν αλλαγές της παραμέτρου. Τέλος ισχύει:

$$\frac{d\vec{C}}{ds} = \frac{d\vec{C}}{dp} \frac{dp}{ds} = \frac{1}{v(p)} \frac{d\vec{C}}{dp} \Rightarrow \left\| \frac{d\vec{C}}{ds} \right\| = \frac{\|d\vec{C}/dp\|}{v(p)} = 1$$

δηλαδή  $\frac{d\vec{C}}{ds} = \vec{T}$ .

Σε κάθε σημείο της καμπύλης, η καμπυλότητα  $\kappa$  ορίζεται ως ο ρυθμός μεταβολής της γωνίας της εφαπτόμενης  $\theta$  σε σχέση με το μήκος τόξου της καμπύλης:

$$\kappa(p) \triangleq \frac{d\theta}{ds} = \frac{\theta'(p)}{v(p)} = \frac{y'x' - x''y'}{[(x')^2 + (y')^2]^{\frac{3}{2}}} \quad (3.1)$$

### 3.1.1 Περιγραφητές Fourier

Οι περιγραφητές Fourier χρησιμοποιούνται αρκετά συχνά για την περιγραφή σχημάτων, καθώς διαθέτουν αρκετές από τις ιδιότητες που είναι επιθυμητό να έχουν οι περιγραφητές σχήματος: ανεξαρτησία από μετασχηματισμούς μετατόπισης, περιστροφής και αλλαγής κλίμακας του υπό περιγραφή αντικειμένου. Οι ιδιότητες αυτές εξασφαλίζουν ότι η περιγραφή του αντικειμένου δεν θα εξαρτάται από την ακριβή θέση και το μέγεθος του αντικειμένου στην εικόνα, και συνεπώς παράμετροι όπως η απόσταση του αντικειμένου από το σημείο λήψης της εικόνας και η θέση του αντικειμένου

ως προς τον οπτικό άξονα της κάμερας λήψης δεν θα επηρεάζουν τα χαρακτηριστικά σχήματος.

Προτού υπολογιστούν οι διακριτοί περιγραφητές Fourier μιας εικόνας, πρέπει πρώτα προφανώς να πραγματοποιηθεί κατάτμηση αυτής και εξαγωγή του συνόρου των αντικειμένων. Το σύνορο αυτό θεωρούμε εδώ ότι αντιπροσωπεύεται από ένα σύνολο  $\{z_0, z_1, \dots, z_N\}$  μιγαδικών αριθμών, των οποίων το πραγματικό και το φανταστικό μέρος αντιστοιχούν στις δύο συνιστώσες της θέσης κάθε σημείου του συνόρου. Οι περιγραφητές Fourier προκύπτουν συνδυάζοντας τους μετασχηματισμούς Fourier (DFT) των συνιστωσών αυτών:

$$Z_k = \sum_{n=0}^{N-1} z_n e^{-2\pi jkn/N}, \quad k = 0, 1, 2, \dots, N-1 \quad (3.2)$$

Συνήθως για τη λήψη των περιγραφητών Fourier οι συντελεστές  $Z_0, Z_1$  απορρίπτονται, και κατόπιν λαμβάνονται οι απόλυτες τιμές των υπόλοιπων συντελεστών διαιρεμένες με την απόλυτη τιμή του  $Z_1$ . Τελικά δηλαδή το σύνολο των περιγραφητών Fourier λαμβάνεται ως εξής:

$$C_{k-2} = \frac{\|Z_k\|}{\|Z_1\|}, \quad k = 2, 3, \dots, N-1 \quad (3.3)$$

### Ιδιότητες των Περιγραφητών Fourier

Οι περιγραφητές Fourier που λαμβάνονται από τη σχέση 3.3 παραμένουν αμετάβλητοι από μετατοπίσεις, περιστροφές και αλλαγές στην κλίμακα των αντικειμένων, όπως θα αποδειχτεί αμέσως παρακάτω.

**Μετατόπιση:** Μια μετατόπιση μπορεί να περιγραφεί από την πρόσθεση ενός μιγαδικού σταθερού αριθμού σε κάθε σημείο του περιγράμματος  $z_n$ . Τότε ο μετασχηματισμός Fourier του νέου συνόλου σημείων θα είναι:

$$\begin{aligned} \sum_{n=0}^{N-1} (z_n - z) e^{-2\pi jkn/N} &= \sum_{n=0}^{N-1} z_n e^{-2\pi jkn/N} + z \sum_{n=0}^{N-1} e^{-2\pi jkn/N} \\ &= \begin{cases} Z_k + z \frac{1 - e^{-2\pi jkn}}{1 - e^{-2\pi jkn/N}} = Z_k, & k \neq 0 \\ Z_k + z(N-1) & k = 0 \end{cases} \end{aligned}$$

Όπως βλέπουμε η μετατόπιση επηρεάζει την τιμή μόνο του πρώτου συντελεστή του μετασχηματισμού Fourier. Η απόρριψη αυτού καθιστά τους περιγραφητές αμετάβλητους από μετατοπίσεις.



**Περιστροφή:** Η περιστροφή κατά γωνία  $\phi$  αντιστοιχεί σε πολλαπλασιασμό κάθε στοιχείου του συνόλου σημείων με  $e^{-j\phi}$ . Ο μετασχηματισμός Fourier του νέου συνόλου θα είναι:

$$\sum_{n=0}^{N-1} (z_n e^{j\phi}) e^{-2\pi jkn/N} = e^{j\phi} \sum_{n=0}^{N-1} z_n e^{-2\pi jkn/N}$$

οπότε επειδή  $|e^{j\phi}| = 1$ , η λήψη των απολύτων τιμών των συντελεστών επιτυγχάνει την ανεξαρτησία από περιστροφές.

**Αλλαγή Κλίμακας:** Η αλλαγή κλίμακας αντιστοιχεί σε πολλαπλασιασμό κάθε σημείου με μια πραγματική σταθερά  $c$ . Ο μετασχηματισμός Fourier είναι τότε:

$$\sum_{n=0}^{N-1} cz_n e^{-2\pi jkn/N} = c \sum_{n=0}^{N-1} z_n e^{-2\pi jkn/N}$$

και η σταθερά  $c$  απαλείφεται με την επιλογή ενός συντελεστή (προφανώς διαφορετικού του  $Z_0$ , για να εξασφαλιστεί η ανεξαρτησία από μετατοπίσεις) με τον οποίο διαιρούμε όλους τους υπόλοιπους.

**Επιλογή Σημείου Έναρξης:** Ένα άλλο πρόβλημα με τους περιγραφητές σχήματος είναι ότι θα πρέπει να μην εξαρτώνται από την επιλογή του σημείου από το οποίο κανείς επιλέγει να αρχίσει να διατρέχει το περίγραμμα. Η επιλογή ενός διαφορετικού σημείου έναρξης ισοδυναμεί με πολλαπλασιασμό κάθε σημείου με  $e^{2\pi jkm/N}$ , όπου  $m$  ο αριθμός των σημείων που μεσολαβούν μεταξύ του νέου αρχικού σημείου και του προηγούμενου. Η περίπτωση αυτή ανάγεται άμεσα στην ανάλυση που πραγματοποιήσαμε για την απόδειξη της ανεξαρτησίας από περιστροφές. Συνεπώς οι συντελεστές Fourier δεν εξαρτώνται ούτε από την επιλογή του αρχικού σημείου της καμπύλης.

### Εναλλακτικές Προσεγγίσεις των Περιγραφητών Fourier

Περιγραφή του σχήματος με χρήση μετασχηματισμών Fourier μπορεί να γίνει και με εφαρμογή του μετασχηματισμού όχι μόνο στη θέση των σημείων του περιγράμματος, αλλά και στη γωνία  $\theta$  της καμπύλης (που είναι περιοδική και συνεπώς μπορεί να παρασταθεί από μια σειρά Fourier), ή και στην καμπυλότητα  $\kappa$ . Οι περιπτώσεις αυτές εξερευνώνται στα [48],[25].

Επίσης, στο [9], περιγράφεται η ανάπτυξη ενός συνόλου περιγραφητών Fourier που να παρουσιάζουν επιπλέον και ανεξαρτησία σε αφινικούς μετασχηματισμούς, ενώ στο [26] χρησιμοποιούνται περιγραφητές Fourier σε πολλαπλές κλίμακες με χρήση του μετασχηματισμού wavelet.

### 3.1.2 Χώρος Κλίμακας Καμπυλότητας

Ο χώρος κλίμακας καμπυλότητας είναι ένας τρόπος αναπαράστασης του σχήματος επίπεδων καμπυλών, που αναπτύχθηκε από τους Mochtarian και Mackworth στο [32]. Πρόκειται για μια μέθοδο που υλοποιείται σε πολλαπλές κλίμακες ανάλυσης του σχήματος, και βασίζεται στον υπολογισμό της καμπυλότητας σε κάθε κλίμακα. Λόγω της πολυκλιμακωτής μορφής του είναι κατάλληλος για την αναγνώριση σχημάτων σε περιπτώσεις που οι καμπύλες σχήματος είναι θορυβώδεις ή έχουν διαφορετικό μέγεθος από το σχήμα αναφοράς. Βρίσκει δε αρκετές εφαρμογές σε περιπτώσεις που απαιτείται κατάτμηση μιας καμπύλης ή εύρεση χαρακτηριστικών σημείων σε αυτή, στην ευθυγράμμιση και ταυτοποίηση επίπεδων σχημάτων, ακόμα και παρουσία επικάλυψων ([31]), και στην αποτελεσματική και συνοπτική αναπαράσταση σχημάτων σε βίντεο.

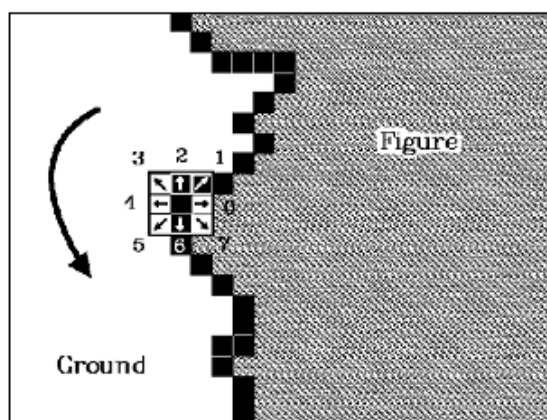
#### Υπολογισμός της Καμπυλότητας

Ο ορθός υπολογισμός της καμπυλότητας ενός σχήματος είναι μια αρκετά περίπλοκη διαδικασία, λόγω της ψηφιακής μορφής των εικόνων και της εμφάνισης δευτέρων παραγώγων στον τύπο υπολογισμού της καμπυλότητας. Οι δύο βασικές μέθοδοι που χρησιμοποιούνται προς αυτήν την κατεύθυνση είναι η χρήση του κώδικα αλυσίδας του Freeman από τη μία ([14]), και η εφαρμογή παρεμβολής (συνήθως με B-splines, [27]) μεταξύ των δειγμάτων-σημείων της καμπύλης οπότε οι παραγωγίσεις εφαρμόζονται άμεσα στις καμπύλες παρεμβολής.

Στα πλαίσια της παρούσας διπλωματικής θα χρησιμοποιηθεί η πρώτη μέθοδος, που είναι και λιγότερο απαιτητική υπολογιστικά. Τα βήματα που ακολουθούνται είναι τα εξής:

1. Δημιουργείται ένα  $3 \times 3$  παράθυρο με κέντρο το τρέχον pixel. Από την θέση (2,3) μέσα στο παράθυρο, και κινούμενοι αριστερόστροφα, τα κουτιά του παραθύρου βαθμονομούνται από το 0 έως το 7 με αύξουσα σειρά διακριτοποιώντας έτσι το διάστημα από  $0^\circ$  σε  $360^\circ$  σε διαστήματα των  $45^\circ$  (βλ. Σχ. 3.1). Καθώς προχωρούμε από το pixel-στόχο στο επόμενο αποθηκεύεται σε έναν πίνακα ο αντίστοιχος αριθμός του επόμενου pixel από 0-7 που υποδηλώνει την σχετική μεταβολή της γωνίας από το ένα στο άλλο. Έτσι λαμβάνουμε μια όχι και τόσο ομαλή συνάρτηση της μεταβολής γωνίας  $C(i)$  καθώς κινούμαστε κατά μήκος του περιγράμματος.

2. Διορθώνουμε το αποτέλεσμα έτσι ώστε να ληφθεί υπόψιν η διαφορά στην απόσταση που διανύουμε όταν κινούμαστε οριζόντια, κατακόρυφα ή διαγώνια. Στις δύο πρώτες περιπτώσεις η ευκλείδεια απόσταση είναι 1 ενώ στις άλλες δύο είναι  $\sqrt{2}$ . Για να έχουμε μια καλύτερη προσέγγιση στην πράξη, θεωρούμε 2 ίδια βήματα για τις οριζόντιες-κατακόρυφες μετατοπίσεις και 3 βήματα για τις διαγώνιες (προσεγγίζοντας το  $\sqrt{2}$  με την τιμή 1,5).
3. Διορθώνουμε το αποτέλεσμα έτσι ώστε η μέγιστη επιτρεπτή μεταβολή γωνίας από το ένα pixel στο επόμενο να είναι  $180^\circ$  (που αντιστοιχεί σε διαφορά τουλάχιστον 4 μονάδων στις διακριτοποιημένες τιμές της αλυσίδας από το ένα σημείο στο άλλο). Σε περίπτωση που η διαφορά είναι μεγαλύτερη από 4 τότε βρίσκουμε έναν ακέραιο  $k_i$  τέτοιο ώστε η διαφορά  $|C_m(i-1) - C(i) - 8k_i|$  να είναι μικρότερη ή ίση με 4 και αντικαθιστούμε το  $C(i)$  με την τιμή  $C(i) + 8k_i$ . Όταν λοιπόν μετά από ένα 7 ακολουθεί 0 αντικαθίσταται το 0 από 8 ( $360^\circ$ ) ενώ όταν από 0 ακολουθεί 7 αντικαθίσταται το 7 από -1 ( $-45^\circ$ ).



Εικόνα 3.1: Κατασκευή του κώδικα αλυσίδας. Από [29].

Ο κώδικας αλυσίδας για τον υπολογισμό της καμπυλότητας βασίζεται στο γεγονός ότι η καμπυλότητα εξ' ορισμού είναι ίση με την παράγωγο της γωνίας της εφαπτόμενης σε κάθε σημείο της καμπύλης,  $\kappa = \partial\theta/\partial s$ . Ο κώδικας αλυσίδας μας δίνει αυτήν ακριβώς τη γωνία, οπότε μπορούμε να βρούμε την καμπυλότητα βρίσκοντας τη (διακριτή) παράγωγο αυτής. Ωστόσο, επειδή η διακριτή παραγωγή αλλά και η διακριτοποίηση της καμπύλης εισάγουν σφάλματα, τα οποία εμφανίζονται ως απότομες μεταβολές στην καμπυλότητα, πριν από την παραγωγή πραγματοποιούμε

ομαλοποίηση της αλυσίδας με συνέλιξη με μια γκαουσιανή, η διασπορά της οποίας τίθεται ίση με τα  $2/35$  του συνολικού μήκους  $L$  της αλυσίδας (η τιμή αυτή βρέθηκε ευριστικά). Το φιλτράρισμα αυτό μπορεί να πραγματοποιηθεί σε 2 βήματα, λαμβάνοντας υπόψιν και τις ιδιότητες της συνέλιξης με γκαουσιανές συναρτήσεις: πρώτα ομαλοποιείται η αλυσίδα με συνέλιξη με μια γκαουσιανή τυπικής απόκλισης  $\sigma_1 = L/35$  και κατόπιν η ομαλοποιημένη αλυσίδα συνελίσσεται με την παράγωγο μιας δεύτερης γκαουσιανής με τυπική απόκλιση  $\sigma_2 = \sigma_1 = L/35$ . Η τελευταία αυτή συνέλιξη μας δίνει την καμπυλότητα.

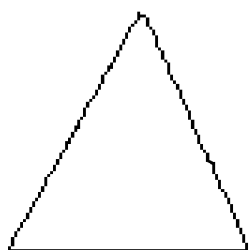
Να σημειώσουμε εδώ ότι η καμπυλότητα απαιτεί η γωνία  $\theta$  να είναι περιοδική συνάρτηση, δηλαδή να «επαναλαμβάνεται» μετά το πέρας της καμπύλης, έτσι ώστε με την παραγωγή να μην προκύψουν σφάλματα στα όρια της καμπύλης. Ωστόσο στη διακριτή περίπτωση που εξετάσαμε παραπάνω, αφ' ενός η γωνία που προκύπτει από τον κώδικα αλυσίδας δεν εμφανίζει περιοδικότητα, καθώς αντιστοιχεί σε μία μόνο διάσχιση της καμπύλης, και αφ' ετέρου η τιμή της στο αρχικό σημείο μπορεί να διαφέρει από την τιμή της στο τελικό σημείο κατά  $\pm 2\pi$ , με αποτέλεσμα να προκύπτουν κορυφές στην αρχή και στο τέλος. Για να εξαλειφθεί αυτό το φαινόμενο, πριν τις δύο διορθώσεις στο βήμα 3 του κώδικα αλυσίδας τοποθετήσαμε έναν ορισμένο αριθμό από δείγματα που αντιστοιχούν στο τέλος της καμπύλης στην αρχή της αλυσίδας και το αντίστροφο, έτσι ώστε μετά τη διόρθωση της γωνίας να υπάρχουν μόνο ομαλές μεταβολές της γωνίας στην αλυσίδα, αλλά και μετά τη συνέλιξη να παίρνουμε σωστά αποτελέσματα.

Ένα παράδειγμα υπολογισμού της καμπυλότητας για ένα απλό σχήμα (τρίγωνο) παρουσιάζεται στο Σχ. 3.2. Η σκιασμένη περιοχή αντιστοιχεί στη γωνία της εφαπτομένης (πολλαπλασιασμένη επί μια σταθερά για λόγους καλύτερης οπτικοποίησης), ενώ η καμπυλότητα παρουσιάζεται με κόκκινο χρώμα. Το σημείο έναρξης της καμπύλης είναι το αριστερότερο σημείο του σχήματος, που αντιστοιχεί σε μία από τις κορυφές του τριγώνου. Όπως ανεμένετο η καμπυλότητα είναι παντού σχεδόν μηδενική με εξαίρεση τις ακμές του τριγώνου.

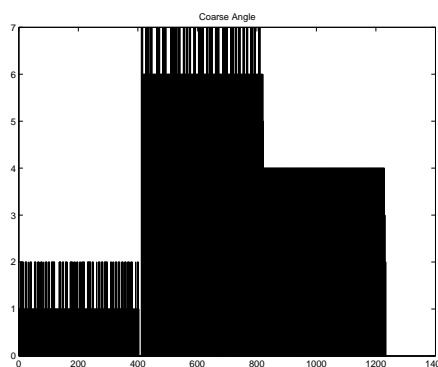
### Κατασκευή του Γραμμικού Χώρου Κλίμακας Καμπυλότητας

Όπως είδαμε στην αρχή της ενότητας, η καμπυλότητα μιας επίπεδης ομαλής καμπύλης μπορεί να υπολογιστεί από την παραμετρική αναπαράσταση αυτής, από τη σχέση (3.1). Αν  $s$  είναι η παράμετρος μήκους τόξου, τότε η καμπυλότητα θα είναι:

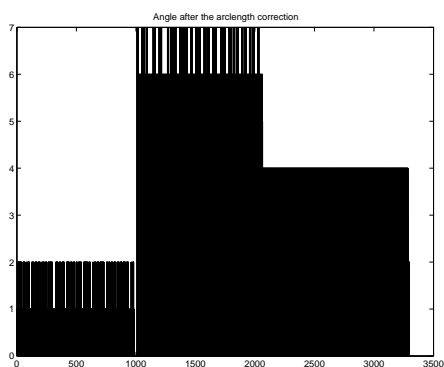
$$\kappa(s) = y(s)''x(s)' - x(s)''y(s)'$$



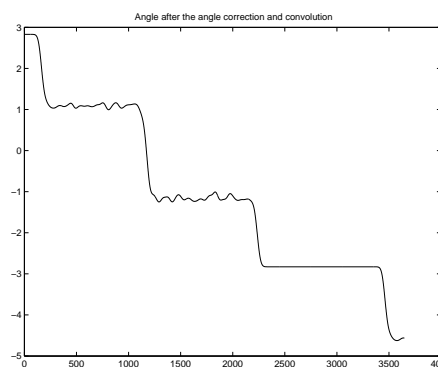
(a) Αρχικό Σχήμα: Τρίγωνο



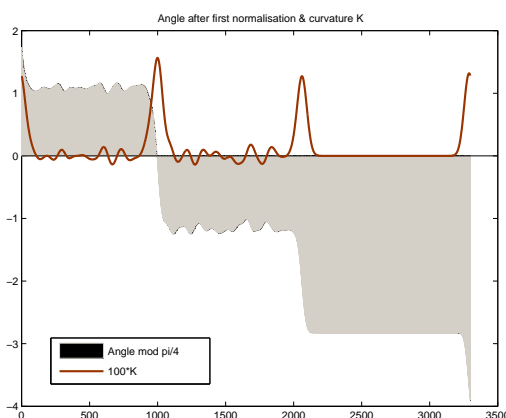
(b) Γωνία εφαπτομένης μετά την εφαρμογή του κώδικα αλυσίδας



(c) Γωνία εφαπτομένης μετά τη διόρθωση μήκους



(d) Γωνία εφαπτομένης μετά τη διόρθωση γωνίας και την ομαλοποίηση



(e) Τελική Γωνία εφαπτομένης και Καμπυλότητα

Εικόνα 3.2: Υπολογισμός της Καμπυλότητας για ένα απλό σχήμα

Δοσμένης μιας καμπύλης  $\Gamma = \{(x(s), y(s)) : |s \in [0, 1]\}$ , μια «εξελιγμένη» εκδοχή αυτής ορίζεται ως:

$$\Gamma_\sigma = \{(X(u, \sigma), Y(u, \sigma)) : u \in [0, 1]\}$$

όπου

$$X(u, \sigma) = x(u) * g(u, \sigma), \quad Y(u, \sigma) = y(u) * g(u, \sigma)$$

είναι ομαλοποιημένες εκδοχές των συνιστωσών της θέσης κάθε σημείου της καμπύλης μετά από συνέλιξη αυτών με μια γκαουσιανή εύρους  $\sigma$ :

$$g(u, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}}$$

Η αναλυτική μορφή των συναρτήσεων  $X$  και  $Y$  είναι η εξής:

$$X(u, \sigma) = \int_{-\infty}^{\infty} x(v) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(u-v)^2}{2\sigma^2}} dv \quad (3.4)$$

$$Y(u, \sigma) = \int_{-\infty}^{\infty} y(v) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(u-v)^2}{2\sigma^2}} dv \quad (3.5)$$

και η καμπυλότητα της ομαλοποιημένης καμπύλης:

$$\kappa(u, \sigma) = \frac{Y_{uu}(u, \sigma)X_u(u, \sigma) - Y_u(u, \sigma)X_{uu}(u, \sigma)}{[(X_u(u, \sigma))^2 + (Y_u(u, \sigma))^2]^{\frac{3}{2}}} \quad (3.6)$$

Ισχύει, βάσει ιδιοτήτων της γκαουσιανής, ότι:

$$\begin{aligned} X_u(u, \sigma) &= \frac{\partial}{\partial u} (x(u) * g(u, \sigma)) = x(u) * g_u(u, \sigma) \\ X_{uu}(u, \sigma) &= \frac{\partial^2}{\partial u^2} (x(u) * g(u, \sigma)) = x(u) * g_{uu}(u, \sigma) \end{aligned}$$

και προφανώς αντίστοιχα για την  $Y$ , δηλαδή η διαδικασία της παραγωγίσης μεταφέρεται στη γκαουσιανή συνάρτηση αντί για τις συναρτήσεις θέσης. Η διαδικασία λήψης των καμπυλών  $\{\Gamma_\sigma | \sigma \geq 0\}$  θα αναφέρεται από εδώ και στο εξής ως «εξέλιξη» της  $\Gamma$ . Καθώς η καμπύλη εξελίσσεται, το συνολικό μήκος της ελαττώνεται, και οι οξείες «ακμές» κατά μήκος αυτής εξομαλύνονται, μέχρις ότου η καμπύλη γίνει τελείως κυρτή.

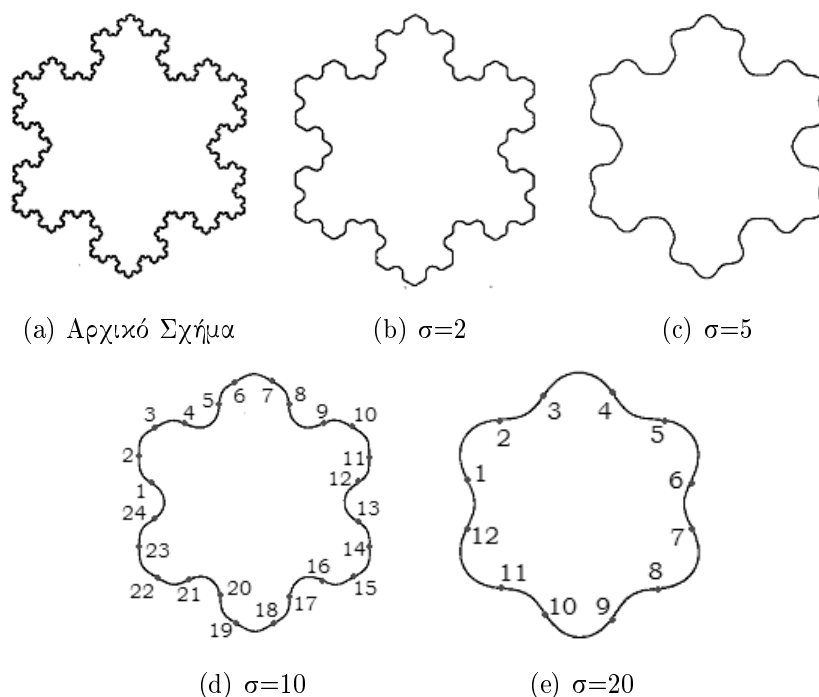
Ο χώρος κλίμακας καμπυλότητας προκύπτει από τη λήψη των zero-crossings της καμπυλότητας σε κάθε κλίμακα, δηλαδή των σημείων  $(u, \sigma)$  που ικανοποιούν τη σχέση:

$$\kappa(u, \sigma) = 0$$

Τα σημεία αυτά αντιστοιχούν σε εκείνες της περιοχές της καμπύλης που βρίσκονται μεταξύ δύο τοπικών ακροτάτων (μεγίστου ή ελαχίστου) της καμπυλότητας, δηλαδή στα σχετικά «ευθεία» τμήματα της καμπύλης. Συνεπώς σε κάθε κοίλο ή κυρτό τμήμα της καμπύλης υπάρχουν δύο zerocrossings. Καθώς η καμπύλη εξελίσσεται, τα σημεία

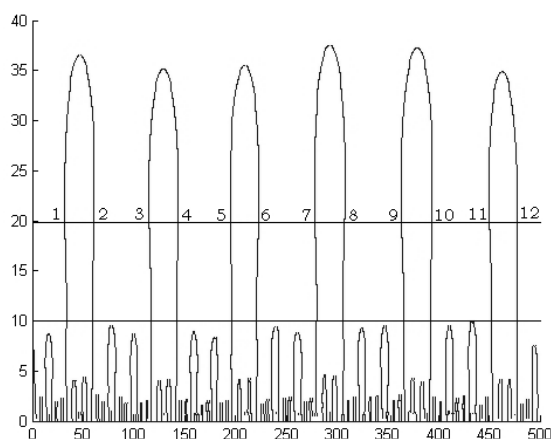
μηδενισμού της καμπυλότητας πλησιάζουν το ένα στο άλλο, και σταδιακά ταυτίζονται και ο αριθμός τους μειώνεται, μέχρις ότου η καμπύλη γίνει τελείως κυρτή οπότε δεν θα εμφανίζονται zero-crossings. Τοποθετώντας τα σημεία αυτά σε ένα γράφημα του οποίου ο οριζόντιος άξονας αντιστοιχεί στη θέση των σημείων κατά μήκος της καμπύλης και ο κατακόρυφος στην τυπική απόκλιση της γκαουσιανής που χρησιμοποιείται σε κάθε κλίμακα, λαμβάνεται η εικόνα CSS (curvature scale-space, χώρος κλίμακας καμπυλότητας). Η κατεύθυνση μετακίνησης των zerocrossings κατά την εξέλιξη της καμπύλης είναι προς τα ελάχιστα ή μέγιστα της καμπυλότητας, με αποτέλεσμα τα σημεία στα οποία δύο zerocrossings ταυτίζονται αντιστοιχούν στα πιο κοίλα ή κυρτά σημεία της καμπύλης. Το σημείο στο οποίο δύο zerocrossings ενώνονται εμφανίζεται σαν κορυφή στο διάγραμμα CSS, με το ύψος της κορυφής να δίνει ένα μέτρο του πόσο κυρτή ή κοίλη είναι η καμπύλη στο αντίστοιχο σημείο.

Ένα παράδειγμα εξέλιξης καμπύλης και κατασκευής του χώρου CSS παρουσιάζεται στα παρακάτω Σχ. 3.3, Σχ. 3.4, με σημειωμένα τα zero-crossings.



Εικόνα 3.3: Η εξέλιξη της καμπύλης ενός σχήματος

Το διάνυσμα χαρακτηριστικών που παράγεται από το χώρο CSS έχει ως στοιχεία τις θέσεις των μεγίστων στο διάγραμμα CSS. Η εξαγωγή των μεγίστων από το διάγραμμα γίνεται με βάση την παρατήρηση ότι οι γραμμές στο σχήμα CSS είναι παντού



Εικόνα 3.4: Ο χώρος κλίμακας καμπυλότητας για την εικόνα του σχ. 3.3

συνεχείς (και συνεπώς η εικόνα είναι συνδεδεμένη) με εξαίρεση συνήθως τα σημεία όπου εμφανίζονται μέγιστα. Εναλλακτικά μπορεί να χρησιμοποιηθεί ένα φίλτρο τύπου hit-miss, με βάση την αναμενόμενη μορφή που θα έχει το διάγραμμα CSS στην περιοχή του μεγίστου. Για τη λήψη του διάνυσματος χαρακτηριστικών, αρχικοποιείται ένα μηδενικό διάνυσμα μήκους ίσου με τον αριθμό των σημείων στα οποία έχει δειγματοληπτηθεί η καμπύλη, και στη συνέχεια τοποθετούνται '1' στις θέσεις όπου ανιχνεύεται μέγιστο. Επειδή προφανώς υπάρχουν περισσότερα του ενός τοπικά μέγιστα στο διάγραμμα CSS, θεωρούνται μόνο εκείνα τα μέγιστα που ξεπερνούν ένα κατώφλι, συνήθως συναρτήσει του μεγέθους του μεγαλύτερου μεγίστου.

### Ιδιότητες Χώρου Κλίμακας Καμπυλότητας

Προφανώς από τον ορισμό της καμπυλότητας προκύπτει ότι το διάνυσμα χαρακτηριστικών θα είναι ανεπηρέαστο από μετατοπίσεις και περιστροφές. Αλλαγή κλίμακας επίσης δεν αναμένεται να επηρεάζει σημαντικά το αποτέλεσμα, εφ' όσον διατηρούμε σταθερό αριθμό δειγμάτων της καμπύλης. Το μόνο μέγεθος που μπορεί να επηρεάσει το διάνυσμα χαρακτηριστικών είναι η επιλογή του σημείου έναρξης της καμπύλης. Αλλαγή του σημείου αυτού συνεπάγεται μια οριζόντια «τοροειδή» μετατόπιση του διαγράμματος CSS, χωρίς να επιφέρει κάποια δομική αλλαγή αυτού. Για το λόγο αυτό συνήθως πριν από την αναγνώριση ενός σχήματος με βάση το χώρο CSS πραγματοποιείται κάποιο matching των CSS εικόνων, ώστε να ευθυγραμμιστούν τα σχήματα. Ωστόσο αυτό δεν αποτελεί ανασταλτικό παράγοντα για τη χρήση της εν λόγω μεθόδου.



δου, καθώς η διαδικασία του matching δίνει παράλληλα και το βαθμό ομοιότητας των δύο σχημάτων με βάση το CSS.

### Επεκτάσεις

Στο [32] αναπτύσσονται και ορισμένες επεκτάσεις στον τρόπο λήψης της εικόνας CSS, που αφορούν κυρίως στην σωστότερη αναγνώριση σχημάτων παρουσία διαφόρων τύπων θορύβου.

Επίσης, στο [29], παρουσιάζεται ο μορφολογικός χώρος κλίμακας καμπυλότητας, όπου στη θέση του γκαουσιανού φιλτραρίσματος πραγματοποιείται μη γραμμικό μορφολογικό φιλτράρισμα.

### 3.1.3 Το Cepstrum της Καμπυλότητας

Το cepstrum είναι μια μέθοδος ανάλυσης σήματος που χρησιμοποιείται ευρύτατα στην ανάλυση και αναγνώριση φωνής. Είναι ιδιαίτερα χρήσιμο για την περιγραφή σημάτων που μπορούν να θεωρηθούν ως συνελίξεις ενός σήματος από μια πηγή με ένα φίλτρο, όπως είναι η φωνή. Το cepstrum εξυπηρετεί στο διαχωρισμό του φάσματος της πηγής από το φάσμα του φίλτρου. Επίσης είναι ικανό να εξάγει τυχόν περιοδικότητες και αρμονικές ζώνες που υπάρχουν στο φάσμα (αν αυτό θεωρηθεί ως ένα νέο σήμα), και οι οποίες μπορεί να μην είναι εμφανείς στο ίδιο το φάσμα λόγω υπέρθεσης. Προκύπτει δε από αντίστροφο Fourier μετασχηματισμό του λογαρίθμου του Fourier μετασχηματισμού του σήματος (το μιγαδικό cepstrum ή του λογαρίθμου του πλάτους του Fourier μετασχηματισμού του σήματος. Με τον τρόπο αυτό το cepstrum του συνολικού σήματος θα προκύπτει από το άθροισμα των cepstra του σήματος πηγής και της απόκρισης του φίλτρου, βάσει των ιδιοτήτων:

$$\begin{aligned} x(t) = h(t) * s(t) &\Rightarrow \mathcal{F}[x(t)] = \mathcal{F}[h(t)] \cdot \mathcal{F}[s(t)] \Rightarrow \\ \log \{ \mathcal{F}[x(t)] \} &= \log \{ \mathcal{F}[h(t)] \} + \log \{ \mathcal{F}[s(t)] \} \Rightarrow \\ \mathcal{F}^{-1} \{ \log \{ \mathcal{F}[x(t)] \} \} &= \mathcal{F}^{-1} \{ \log \{ \mathcal{F}[h(t)] \} \} + \mathcal{F}^{-1} \{ \log \{ \mathcal{F}[s(t)] \} \} \end{aligned} \quad (3.7)$$

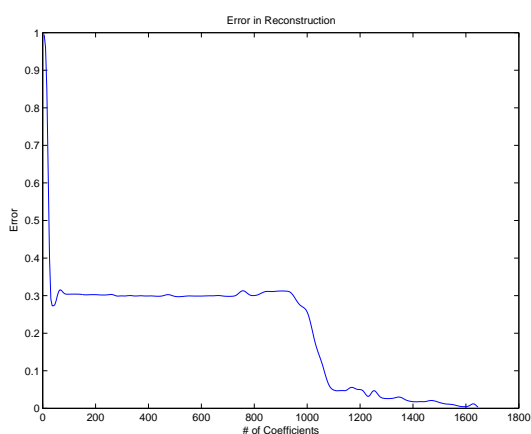
Στη φωνή, όπου μπορούμε να υποθέσουμε ότι η πηγή παράγει μόνο ταχείες ταλαντώσεις, η συνεισφορά της στο cepstrum θα συγκεντρώνεται μόνο στις υψηλές quefrequencies, ενώ η συνεισφορά του φίλτρου (που αντιστοιχεί στο σύστημα παραγωγής της φωνής) θα βρίσκεται στα χαμηλά quefrequencies. Συνεπώς, λόγω της άθροισης, τα δύο φάσματα μπορούν να διαχωριστούν απομονώνοντας τους πρώτους συντελεστές

του cepstrum που βρίσκονται στις χαμηλές frequencies. Για το λόγο αυτό μπορούμε να πούμε ότι το cepstrum πραγματοποιεί ένα βαθυπερατό φιλτράρισμα στο φάσμα του σήματος, και δίνει ένα πιο ομαλά μεταβαλλόμενο φάσμα.

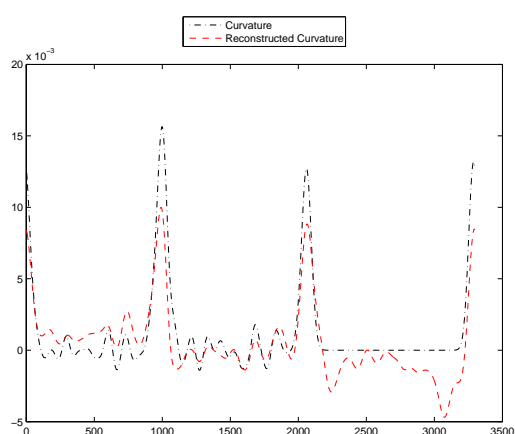
Στην ενότητα αυτή μας ενδιαφέρει να διαπιστώσουμε κατά πόσον μια ίδια τεχνική μπορεί να εφαρμοστεί και στην περίπτωση της εικόνας, και συγκεκριμένα στην καμπυλότητα ενός σχήματος. Συγκεκριμένα, ως χαρακτηριστικά σχήματος θεωρούμε ένα σύνολο από τους  $N_c$  πρώτους συντελεστές του πραγματικού cepstrum της συνάρτησης καμπυλότητας. Επειδή βέβαια το cepstrum είναι συμμετρικό, τελικά στο διάγραμμα χαρακτηριστικών αποθηκεύουμε μόνο τους  $N_c$  συντελεστές στο δεξί ημιπέδο, μαζί με το συντελεστή στο  $t=0$ . Μετά από πειραματισμούς με διάφορα σχήματα διαπιστώθηκε ότι το σφάλμα ανακατασκευής της καμπυλότητας για μικρό  $N_c$  παίρνει πολύ μεγάλες τιμές, ενδεχομένως και μεγαλύτερες από 100%, ενώ καθώς αυξάνεται το  $N_c$  πέφτει βηματικά προς το μηδέν, περνώντας από μερικές περιοχές όπου παραμένει σταθερό.

Η μέθοδος αυτή είναι αντίστοιχη με τη μέθοδο εξαγωγής χαρακτηριστικών cepstrum από σήματα φωνής. Ο αριθμός  $N_c$  πρέπει να είναι αρκετά μικρός, ώστε να μην καθυστερεί σημαντικά η διαδικασία της αναγνώρισης, αλλά και αρκετά μεγάλος ώστε η ανακατασκευή του σήματος της καμπυλότητας από τους  $N_c$  συντελεστές να μην απέχει πολύ από το αρχικό σήμα της καμπυλότητας. Επιλέγεται δε μετά από δοκιμές.

Για το τρίγωνο του Σχ. 3.2, το σφάλμα της ανακατασκευής και η ανακατασκευή της καμπυλότητας από  $N_c = 45$  συντελεστές cepstrum παρουσιάζονται στο Σχ. 3.5.



(a) Σφάλμα Ανακατασκευής συναρτήσεως του αριθμού  $N_c$



(b) Ανακατασκευή της καμπυλότητας από 45 συντελεστές cepstrum

Εικόνα 3.5: Αποτελέσματα της ανακατασκευής για το Σχ. 3.2

Να αναφέρουμε στο σημείο αυτό, ότι το κίνητρο για τη δοκιμή του cepstrum ως περιγραφητή σχήματος ήταν το γεγονός ότι η διαδικασία εξαγωγής του μπορεί να ιδωθεί ως μια μορφή ανάλυσης πρωτευουσών συνιστωσών της καμπυλότητας. Πράγματι, ο διακριτός μετασχηματισμός συνημιτόνου (DCT) που χρησιμοποιείται (συνήθως αντί του FFT) για την εξαγωγή του cepstrum αποτελεί μια προσέγγιση του μετασχηματισμού Karhunen - Lowe (KL), ο οποίος έχει την ιδιότητα να απομονώνει εκείνα τα «κανάλια» στα οποία το σήμα (εδώ η καμπυλότητα) εμφανίζει τη μέγιστη ενέργεια, αντίστοιχα κάπως με την PCA. Συνεπώς τα χαρακτηριστικά του cepstrum μπορούν ενδεχομένως να θεωρηθούν ότι εξάγουν τη μέγιστη πληροφορία από το σήμα της καμπυλότητας.

## 3.2 Region - based Μέθοδοι

Στις μεθόδους αυτές το σχήμα συνήθως παριστάνεται είτε ως ένα σύνολο  $S \subseteq \mathbb{R}^2$  είτε ως μια συνάρτηση

$$s(x, y) = \begin{cases} 1, & (x, y) \in S \\ 0, & (x, y) \in S^c \end{cases}$$

και γίνεται προσπάθεια να εξαχθεί ένα σύνολο χαρακτηριστικών που να περιγράφουν συνολικά ένα σχήμα, με βάση την περιοχή που περικλείεται από αυτό.

### 3.2.1 Κεντρικές Ροπές

Θεωρούμε τη δυαδική εικόνα  $I(x, y)$ , διαστάσεων  $m \times n$ , που ισούται με 1 στην περιοχή εντός του εξεταζόμενου σχήματος, και 0 οπουδήποτε αλλού, και τα σημεία της εικόνας αριθμημένα από το 1 στο  $N = mn$ . Τότε οι ροπές του σχήματος ως προς το σημείο  $(0,0)$  δίνονται από τη σχέση:

$$m_{ij} = \sum_{x=1}^N \sum_{y=1}^N x^i y^j I(x, y) \quad (3.8)$$

όπου προφανώς  $m_{00} = 1$ . Οι συντεταγμένες του κέντρου μάζας ισούνται με

$$(\mu_x, \mu_y) = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right)$$

ενώ προφανώς η ροπή  $m_{00}$  αντιστοιχεί στο εμβαδόν του σχήματος.

Οι κεντρικές ροπές του σχήματος είναι οι ροπές υπολογισμένες γύρω από το κέντρο μάζας του σχήματος, δηλαδή:

$$\mu_{ij} = \sum_{x=1}^N \sum_{y=1}^N (x - \mu_x)^i (y - \mu_y)^j I(x, y) \quad (3.9)$$

και οι κανονικοποιημένες κεντρικές ροπές:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^\gamma}, \quad \gamma = \frac{i+j}{2} + 1 \quad (3.10)$$

Έχοντας κατασκευάσει το σύνολο ροπών, μπορούμε να εξετάσουμε σε ποιο γνώρισμα του σχήματος αντιστοιχεί κάθε μία από αυτές ή και αντίστροφα. Αυτό μπορεί να γίνει μέσω της ανακατασκευής της αρχικής εικόνας από τις υπολογισμένες ροπές. Υποθέτοντας ότι όλες οι κεντρικές ροπές  $\mu_{ij}$  μιας εικόνας  $I(x, y)$ , τάξης  $N = i + j$  είναι γνωστές μέχρι την τάξη  $N_{max}$ , μπορούμε να παράγουμε μια εικόνα της οποίας οι ροπές μέχρι και την τάξη  $N_{max}$  είναι ίσες με τις αντίστοιχες ροπές της  $I$ . Αυτό γίνεται θεωρώντας την εξής μορφή για τη συνάρτηση  $g$ :

$$\begin{aligned} g(x, y) &= g_{00} + g_{10}(x - \mu_x) + g_{01}(y - \mu_y) + g_{20}(x - \mu_x)^2 + \\ &+ g_{11}(x - \mu_x)(y - \mu_y) + g_{02}(y - \mu_y)^2 + \dots + g_{pq}(x - \mu_x)^p (y - \mu_y)^q = \\ &= \sum_{i=0}^{i+j=N_{max}} \sum_{j=0} g_{ij} x^i y^j \end{aligned} \quad (3.11)$$

οπότε οι συντελεστές  $g_{ij}$  υπολογίζονται από την απαίτηση:

$$\sum_x \sum_y g(x, y) (x - \mu_x)^i (y - \mu_y)^j = m_{ij}$$

Η παραπάνω εξίσωση αντιστοιχεί σε ένα σύστημα γραμμικών εξισώσεων, και ο συνολικός αριθμός των εξισώσεων καθορίζεται από την τάξη  $N_{max}$  της ανακατασκευής. Με τον τρόπο αυτό παίρνουμε μια θορυβώδη εκδοχή της  $I$ .

Αποδεικνύεται ότι οι συναρτήσεις βάρη στο άθροισμα 3.9 οδηγούν σε περιγραφητές που περιέχουν πληροφορία για την ύπαρξη ή μη συμμετρίας στην κατανομή της φωτεινότητας στην εικόνα (ως προς το κέντρο μάζας). Σχήματα που είναι συμμετρικά είτε ως προς τον οριζόντιο είτε ως προς τον κατακόρυφο άξονα δίνουν  $\eta_{11} = 0$ . Για σχήματα συμμετρικά ως προς τον  $y$ -άξονα  $\eta_{12} = 0$  και  $\eta_{30} = 0$ , ενώ για σχήματα συμμετρικά ως προς τον  $x$ -άξονα  $\eta_{03} = 0$  και  $\eta_{12} > 0$ . Επιπλέον, για σχήματα συμμετρικά ως προς τον  $x$ -άξονα,  $\eta_{ij} = 0, \forall i = 0, 2, 4, \dots; j = 1, 3, 5, \dots$ , ενώ όταν υπάρχει ασυμμετρία  $\eta_{ij} < 0, \forall i = 0, 2, 4, \dots; j = 1, 3, 5, \dots$ , και  $\eta_{k0} = 0, \eta_{0k} = 0, \forall k = 0, 2, 4, \dots$

Οι κανονικοποιημένες κεντρικές ροπές είναι ανεπηρέαστες από μετατοπίσεις και αλλαγές κλίμακας. Για να εξαλειφθεί και η εξάρτηση από τον προσανατολισμό κατασκευάζουμε το εξής σύνολο ροπών, χρησιμοποιώντας τις κανονικοποιημένες κεντρικές ροπές μέχρι την τάξη 3. Από αυτά, τα μεγέθη  $\phi_1 - \phi_6$  είναι αμετάβλητα σε συμμετρικούς μετασχηματισμούς, ενώ το  $\phi_7$  αλλάζει πρόσημο, για την περίπτωση κατοπτρικών σχημάτων.

$$\phi_1 = \eta_{20} + \eta_{02} \quad (3.12)$$

$$\phi_2 = (\eta_{02} - \eta_{20})^2 + 4\eta_{11}^2 \quad (3.13)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} + \eta_{03})^2 \quad (3.14)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (3.15)$$

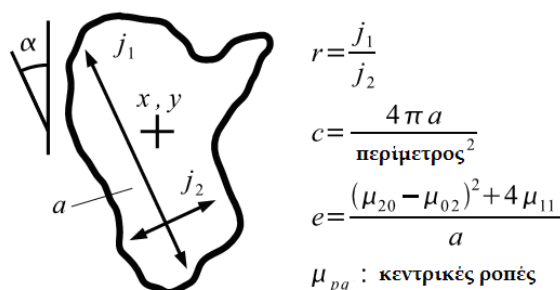
$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] + (3\eta_{21} - \eta_{03})(\eta_{03} + \eta_{21}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] \quad (3.16)$$

$$\phi_6 = (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] + (\eta_{20} - \eta_{02}) [4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21})] \quad (3.17)$$

$$\phi_7 = (3\eta_{21} + \eta_{03})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12}) - 3(\eta_{03} + \eta_{21})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (3.18)$$

### 3.2.2 Χαρακτηριστικά Σχήματος που βασίζονται στις Κεντρικές Ροπές και άλλα Γεωμετρικά Χαρακτηριστικά

Με βάση τις κεντρικές ροπές είναι δυνατόν να παραχθούν και άλλες χρήσιμες ποσότητες, που αποδεικνύονται αποτελεσματικές για τη διάκριση μεταξύ σχημάτων. Οι ποσότητες αυτές παρουσιάζονται στο Σχ. 3.6.



Εικόνα 3.6: Τα region-based χαρακτηριστικά σχήματος

- **Κέντρο μάζας**  $(x, y)$  : Δίνει μια εκτίμηση για τη θέση του σχήματος στην εικόνα.
- **Εμβαδόν**  $a$  : Προσεγγιστικά, ο αριθμός των εικονοστοιχείων που ανήκουν στην περιοχή.
- **Εκκεντρότητα**  $e$  : Ο λόγος του μήκους της μακρύτερης χορδής της περιοχής προς το μήκος της μακρύτερης χορδής μεταξύ των κάθετων στην πρώτη χορδή. Πρόκειται για την εκκεντρότητα της έλλειψης που προσεγγίζει καλύτερα το σχήμα, και δίνει ένα μέτρο του πόσο «επίμηκες» είναι το σχήμα.
- **Βαθμός «Συμπύκνωσης» (Compactness)**  $c$  : Προσεγγιστικά, ο λόγος της επιφάνειας δια το τετράγωνο της περιμέτρου του σχήματος. Δίνει ένα μέτρο του πόσο «κυκλικό» είναι το σχήμα.
- **Κατεύθυνση του κύριου άξονα**  $\alpha$  : Προκύπτει άμεσα από το σχήμα. Πρακτικά στο διάγραμμα χαρακτηριστικών η κατεύθυνση αυτή αντιπροσωπεύεται από το ημίτονο και το συνημίτονο της γωνίας  $2\alpha$ , έτσι ώστε οι γωνίες  $\alpha$  και  $\pi+\alpha$  να δίνουν τα ίδια χαρακτηριστικά κατεύθυνσης.
- **Λόγος αδράνειας**  $r$  : Το πηλίκο του μεγαλύτερου άξονα της έλλειψης που προσεγγίζει καλύτερα το σχήμα προς το μήκος του μικρότερου άξονα αυτής.

Ήδη είδαμε ότι το εμβαδό ισούται με τη μηδενική ροπή του σχήματος. Για τα υπόλοιπα χαρακτηριστικά, χρειάζεται ο πίνακας συμμεταβλητότητας του σχήματος:

$$C = \begin{bmatrix} \eta_{20} & \eta_{11} \\ \eta_{11} & \eta_{02} \end{bmatrix} \quad (3.19)$$

Από τον παραπάνω πίνακα λαμβάνουμε την εκκεντρότητα από τη σχέση:

$$e = 1 - \sqrt{\frac{s_2}{s_1}} \quad (3.20)$$

Μια άλλη εξίσου ορθή αναπαράσταση της εκκεντρότητας είναι:

$$e = \frac{(\eta_{20} - \eta_{02})^2 + 4\eta_{11}}{a} \quad (3.21)$$

Τα μήκη του κύριου και του δευτερεύοντος άξονα προκύπτουν από τις σχέσεις:

$$l_1 = 4\sqrt{s_1} \quad (3.22)$$

$$l_2 = 4\sqrt{s_2} \quad (3.23)$$

όπου  $s_1$  η μεγαλύτερη ιδιοτιμή του πίνακα C και  $s_2$  η μικρότερη. Τέλος, ο βαθμός έσυμπύκνωσης υπολογίζεται από τον τύπο:

$$c = \frac{4\pi a}{P^2} \quad (3.24)$$

όπου η περίμετρος P:

$$P = \sum_{i=1}^M \|\vec{x}_{i+1} - \vec{x}_i\| + \|\vec{x}_M - \vec{x}_1\| \quad (3.25)$$

και  $\vec{x}_i$ ,  $i = 1, 2, \dots, M$  είναι τα σημεία του περιγράμματος του σχήματος.

### 3.3 Εξαγωγή Στατικών Χαρακτηριστικών Σχήματος απο Χειρομορφές

Όπως έχουμε δει, σε κάθε εφαρμογή αναγνώρισης μας ενδιαφέρει η κατασκευή ενός συνόλου χαρακτηριστικών. Εδώ το διάνυσμα χαρακτηριστικών θέλουμε να περιέχει πληροφορία για το σχήμα που έχει το κάθε χέρι του νοηματιστή, καθώς έχει βρεθεί ότι τα περισσότερα νοήματα της ENΓ αποτελούνται από έναν περιορισμένο αριθμό βασικών χειρομορφών. Στην ενότητα αυτή θα δούμε πως εφαρμόζονται τα όσο αναφέραμε στις προηγούμενες ενότητες στην περίπτωση των χειρομορφών.

#### 3.3.1 Region-based Χαρακτηριστικά

Τα region-based χαρακτηριστικά που θα χρησιμοποιήσουμε βασίζονται στις κεντρικές ροπές (βλ. ενότητα 3.2.2), που όπως αναφέραμε ήδη χρησιμοποιούνται συχνά για την ταξινόμηση σχημάτων, γιατί επιτυγχάνουν πολύ μεγάλη συμπίεση, αφού ένας πολύ μικρός αριθμός αυτών (2-3) αρκούν για να περιγράψουν επαρκώς το σχήμα.

Παραθέτουμε τα αποτελέσματα που λάβαμε για 3 διαφορετικές χειρομορφές στον Πίνακα 3.1. Οι εικόνες είναι διαστάσεων  $300 \times 400$ .

#### 3.3.2 Boundary-based Χαρακτηριστικά

Τα boundary-based χαρακτηριστικά που θα χρησιμοποιηθούν περιγράφηκαν στην ενότητα 3.1. Εδώ θα παραθέσουμε ενδεικτικά αποτελέσματα για ορισμένες χειρομορφές.

### Υπολογισμός της Καμπυλότητας

Η καμπυλότητα, όπου αυτή χρειάζεται, υπολογίζεται με βάση τον κώδικα αλυσίδας (βλ. ενότητα 3.1.2). Μερικά αποτελέσματα που πήραμε από τον κώδικα αλυσίδας για κάποιες χειρομορφές, με σημειωμένο το αρχικό σημείο, παρουσιάζονται παρακάτω στο Σχ. 3.7. Στα διαγράμματα αυτά η καμπυλότητα παρουσιάζεται ενισχυμένη για καλύτερη οπτικοποίηση. Μπορούμε να ελέγξουμε την ορθότητα των αποτελεσμάτων μετρώντας τις κορυφές και κοιλάδες της καμπυλότητας, όπως φαίνεται στην εικόνα 3.7(a).

### Ανακατασκευή της Καμπυλότητας μέσω του Cepstrum

Εφ' όσον έχει βρεθεί η καμπυλότητα, μπορούμε να προχωρήσουμε στη φασματική ανάλυση αυτής για την εξαγωγή των χαρακτηριστικών του cepstrum. Συγκεκριμένα, λαμβάνουμε το μετασχηματισμό Fourier της καμπυλότητας, στη συνέχεια το λογάριθμο του πλάτους αυτού και κατόπιν μετασχηματίζουμε αντίστροφα το αποτέλεσμα πίσω στο χώρο του χρόνου. Το αποτέλεσμα είναι το πραγματικό cepstrum της συνάρτησης καμπυλότητας, το οποίο είναι συμμετρικό γύρω από το μηδέν και λαμβάνει μεγάλες τιμές κοντά στο μηδέν αλλά μειώνεται απότομα καθώς προχωράμε προς απολύτως μεγαλύτερες τιμές του χρόνου. Ακολουθώντας μεθοδολογία πανομοιότυπη με αυτή που εφαρμόζεται στην αναγνώριση φωνής, επιχειρούμε να ανακατασκευάσουμε την καμπυλότητα κρατώντας μόνο το πρώτο κομμάτι του cepstrum, αποτελούμενο από  $N_c$  συντελεστές εκατέρωθεν του μηδενικού quefrency. Τα αποτελέσματα της ανακατασκευής για μερικές χειρομορφές παρουσιάζονται αμέσως παρακάτω στα Σχ. 3.8 - 3.10.

Όπως μπορούμε να δούμε, παρότι γενικά η ανακατασκευή του φάσματος δεν είναι τόσο καλή, η μορφή του αρχικού σήματος καμπυλότητας ανακατασκευάζεται αρκετά ικανοποιητικά, με τα σφάλματα να εμφανίζονται μόνο στα ύψη των κορυφών. Γενικά πάντως το αποτέλεσμα είναι αρκετά ελπιδοφόρο.

Να σημειώσουμε εδώ ο αριθμός των συντελεστών που απομονώνονται εξαρτάται από το συνολικό μήκος της καμπύλης που εξετάζεται. Οι εικόνες των χειρομορφών που εξετάσαμε μέχρι τώρα είχαν διάσταση  $300 \times 400$  και οι καμπύλες μήκος γύρω στα 2000-4000 δείγματα. Συνεπώς οι 60 συντελεστές cepstrum επιτυγχάνουν συμπίεση του σήματος κατά 97%. Για τις εικόνες των ομιλητών που είδαμε στο πρώτο μέρος της παρούσας αναφοράς, που έχουν διάσταση  $320 \times 240$  και τα χέρια αποτελούν ένα πολύ μικρό μέρος της εικόνας, οι συντελεστές που θα παραμείνουν στο διάλυσμα



χαρακτηριστικών θα είναι πολύ λιγότεροι. Μια τιμή που βρέθηκε κατάλληλη ήταν  $N_c = 21$  συντελεστές.

#### Γραμμικός Χώρος Κλίμακας Καμπυλότητας

Στο Σχ. 3.11 παρουσιάζονται οι γραμμικοί χώροι κλίμακας καμπυλότητας που προέκυψαν από τρεις διαφορετικές χειρομορφές. Οι κορυφές που τοποθετούνται τελικά στο διάνυσμα κατάστασης έχουν σημειωθεί με μαύρο χρώμα. Οι καμπύλες αυτές έχουν προκύψει από εξέλιξη των περιγραμμάτων των τριών σχημάτων, τα οποία έχουν δειγματοληπτηθεί σε 500 σημεία.

Βλέπουμε ότι τόσο οι θέσεις όσο και τα ύψη των κορυφών αποτελούν σημαντικά διακριτικά στοιχεία για τις χειρομορφές. Ενδεικτικό είναι για παράδειγμα ότι για τη χειρομορφή του Σχ. 3.11(a), που είναι σαφώς πιο «στρογγυλή» από αυτές των Σχ. 3.11(c) και Σχ. 3.11(e), οι κορυφές έχουν πολύ μικρότερο ύψος.

Το βασικότερο πρόβλημα που προκύπτει από τη χρήση του γραμμικού χώρου καμπυλότητας για την εξαγωγή χαρακτηριστικών είναι η εξάρτηση από το αρχικό σημείο της καμπύλης. Αλλαγή του αρχικού σημείου θα έχει ως αποτέλεσμα μετατόπιση στις πιο πάνω εικόνες, πράγμα που μπορεί να οδηγήσει σε εσφαλμένη αναγνώριση μιας χειρομορφής. Για να αντιμετωπιστεί αυτό το ζήτημα συνήθως στις εφαρμογές προηγείται ένα στάδιο ταύτισης των χώρων CSS μεταξύ της χειρομορφής που υφίσταται την επεξεργασία και ενός συνόλου αποθηκευμένων χειρομορφών. Πρακτικά ωστόσο αυτό είναι αρκετά δύσκολο να γίνει σε μια μεγάλη ακολουθία εικόνων, ενώ επίσης θα καθυστερεί σημαντικά την όλη διαδικασία λόγω του αρκετά μεγάλου πλήθους πιθανών χειρομορφών.

#### Περιγραφητές Fourier

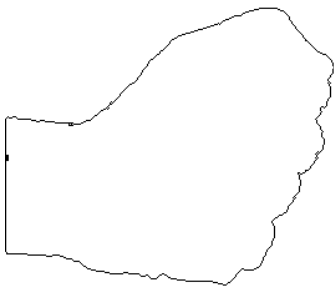
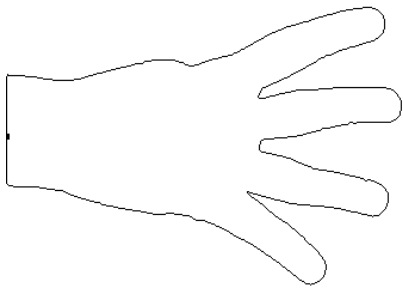
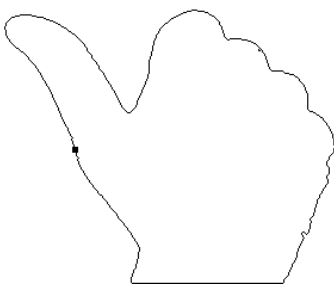
Οι περιγραφητές Fourier είναι ακόμα ένα ευρέως χρησιμοποιούμενο σύνολο χαρακτηριστικών σχήματος, που έχει χρησιμοποιηθεί και για εφαρμογές που σχετίζονται με τη νοηματική γλώσσα. Οι ιδιότητές τους περιγράφηκαν στην ενότητα 3.1.1. Στα Σχ. 3.12 και Σχ. 3.13 παρατίθενται τα αποτελέσματα που πήραμε για τρεις τύπους χειρομορφών.

Όπως μπορούμε να δούμε, χειρομορφές πολύ διαφορετικές μεταξύ τους, όπως αυτές των Σχ. 3.12(c) και Σχ. 3.12(a) εμφανίζουν σημαντικές διαφορές και στους περιγραφητές Fourier, τόσο ως προς το σχήμα της περιβάλλουσας όσο και ως προς τα ύψη των κορυφών τους. Αντίθετα, χειρομορφές που είναι πιο «όμοιες», όπως αυτές των

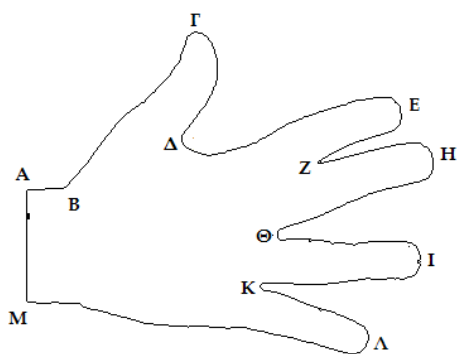
Σχ. 3.12(a) και Σχ. 3.13(a), που είναι και οι δύο πιο «στρογγυλεμένες», έχουν αρκετά παρόμοια φάσματα, που διαφέρουν ωστόσο ως προς τα μεγέθη των κορυφών. Ενδιαφέρον ακόμα είναι το γεγονός ότι οι χειρομορφές των Σχ. 3.13(a) και Σχ. 3.13(c), που έχουν προκύψει η μία από την άλλη μετά από περιστροφή και μετατόπιση, έχουν σχεδόν ταυτόσημους περιγραφητές Fourier. Το γεγονός αυτό αποδεικνύει και έμπρακτα την ανεξαρτησία των περιγραφητών Fourier από γεωμετρικούς μετασχηματισμούς αυτής της μορφής, γεγονός που τους καθιστά ιδιαίτερα εύχρηστους.

### 3.4 Εικόνες Αποτελεσμάτων

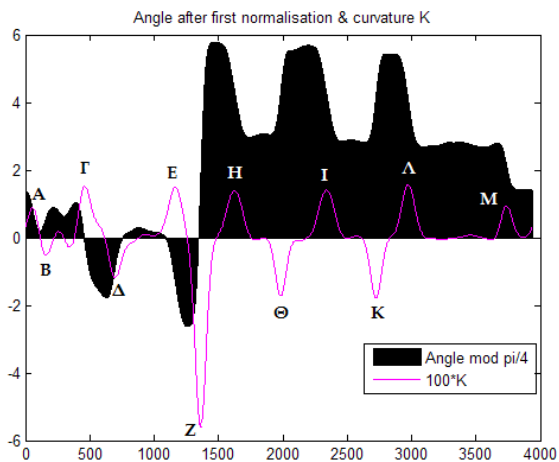
Στις επόμενες σελίδες παρατίθενται τα αποτελέσματα των πειραμάτων του κεφαλαίου 3, με κατάλληλα σχόλια στις λεζάντες όπου χρειάζεται.

 <p style="text-align: center;"><b>Χειρομορφή 1</b></p>	<p>Κέντρο Μάζας : <math>(x_c, y_c) = (138.48, 144.14)</math>  Εμβαδόν : 39684 εικονοστοιχεία  Εκκεντρότητα : 0.7774  Κατεύθυνση κύριου άξονα : <math>29.0968^\circ</math>  Μήκος Κύριου Άξονα : 291.7475  Μήκος Δευτερεύοντος Άξονα : 183.5120  Περιφέρεια : 783 εικονοστοιχεία  Λόγος Αδράνειας : 1.5898  Βαθμός Συμπύκνωσης : 0.8134</p> <p style="text-align: center;"><b>Χαρακτηριστικά Περιοχής</b></p>
 <p style="text-align: center;"><b>Χειρομορφή 2</b></p>	<p>Κέντρο Μάζας : <math>(x_c, y_c) = (166.36, 168.75)</math>  Εμβαδόν : 38745 εικονοστοιχεία  Εκκεντρότητα : 0.8366  Κατεύθυνση κύριου άξονα : <math>2.1803^\circ</math>  Μήκος Κύριου Άξονα : 1187.8114  Μήκος Δευτερεύοντος Άξονα : 342.8127  Περιφέρεια : 1373 εικονοστοιχεία  Λόγος Αδράνειας : 1.8253  Βαθμός Συμπύκνωσης : 0.2583</p> <p style="text-align: center;"><b>Χαρακτηριστικά Περιοχής</b></p>
 <p style="text-align: center;"><b>Χειρομορφή 3</b></p>	<p>Κέντρο Μάζας : <math>(x_c, y_c) = (206.71, 187.90)</math>  Εμβαδόν : 38390 εικονοστοιχεία  Εκκεντρότητα : 0.6917  Κατεύθυνση κύριου άξονα : <math>-39.9985^\circ</math>  Μήκος Κύριου Άξονα : 276.8258  Μήκος Δευτερεύοντος Άξονα : 199.9303  Περιφέρεια : 860 εικονοστοιχεία  Λόγος Αδράνειας : 1.3846  Βαθμός Συμπύκνωσης : 0.6523</p> <p style="text-align: center;"><b>Χαρακτηριστικά Περιοχής</b></p>

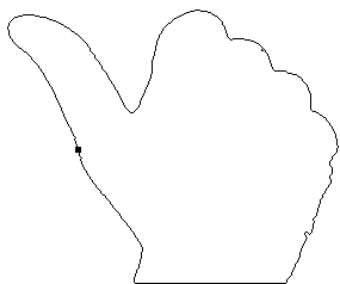
Πίνακας 3.1: Τα region-based χαρακτηριστικά για 3 διαφορετικές χειρομορφές



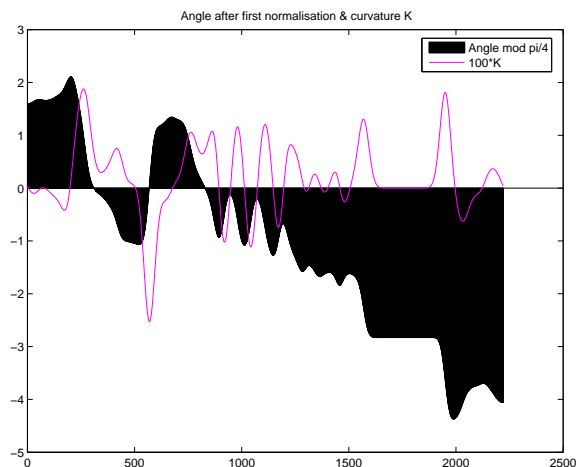
(a) Χειρομορφή 1



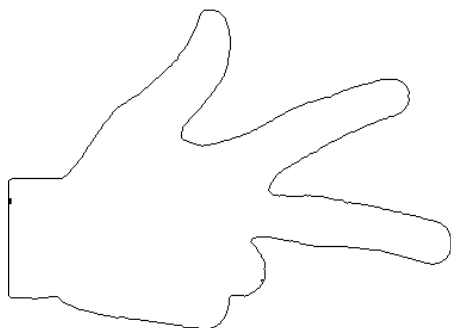
(b) Γωνία Εφαπτόμενης και Καμυλότητα



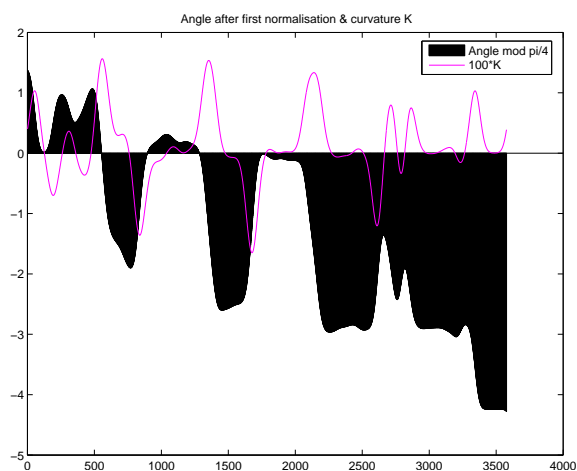
(c) Χειρομορφή 2



(d) Γωνία Εφαπτόμενης και Καμυλότητα

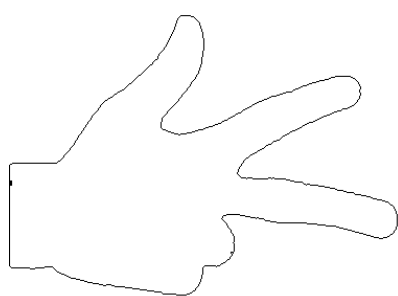


(e) Χειρομορφή 3

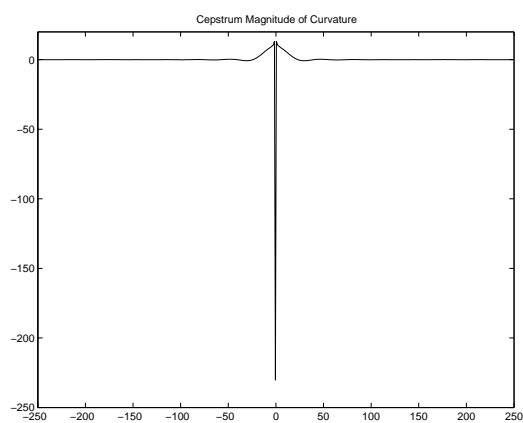


(f) Γωνία Εφαπτόμενης και Καμυλότητα

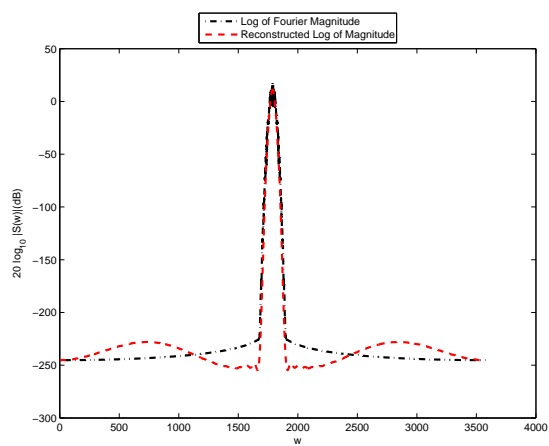
Εικόνα 3.7: Τρία αποτελέσματα της εξαγωγής της καμυλότητας



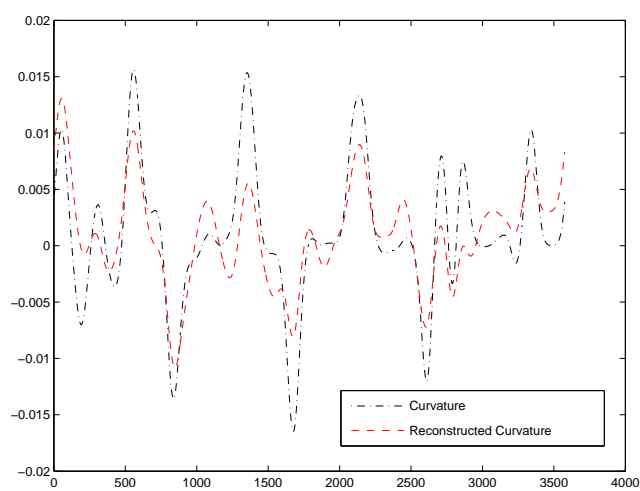
(a) Χειρομορφή



(b) Cepstrum της Καμπυλότητας

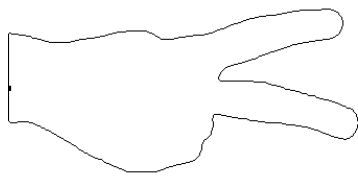


(c) Ανακατασκευασμένο Φάσμα της Καμπυλότητας

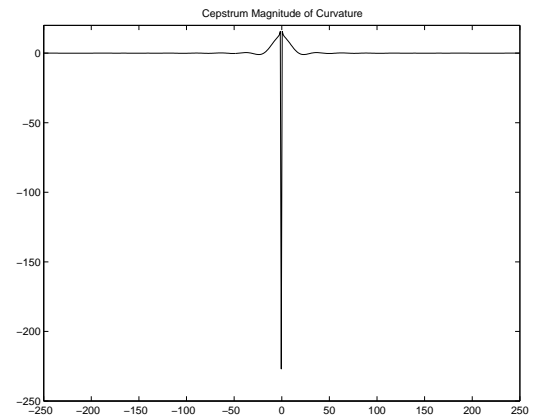


(d) Ανακατασκευή της Καμπυλότητας

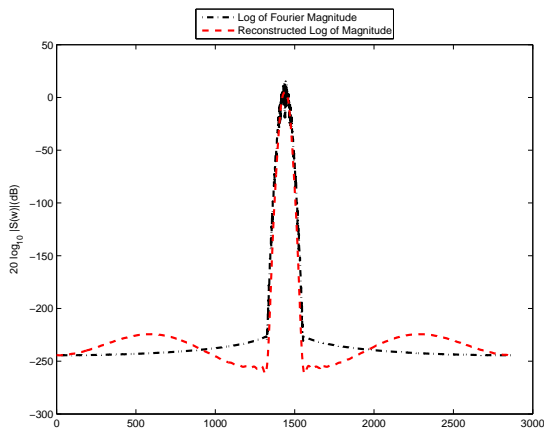
Εικόνα 3.8: Ανακατασκευή της καμπυλότητας από τους 61 πρώτους συντελεστές cepstrum - Χειρομορφή 1



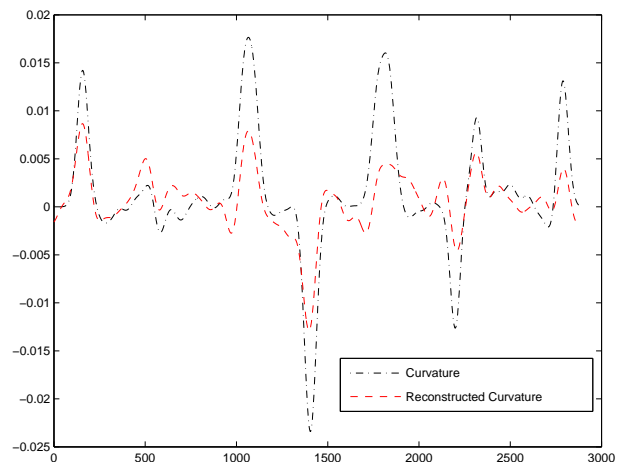
(a) Χειρομορφή



(b) Cepstrum της Καμπυλότητας

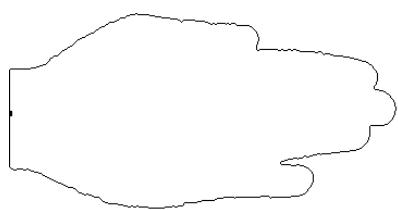


(c) Ανακατασκευασμένο Φάσμα της Καμπυλότητας

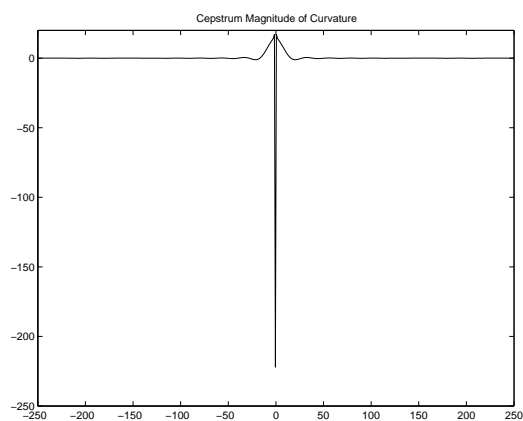


(d) Ανακατασκευή της Καμπυλότητας

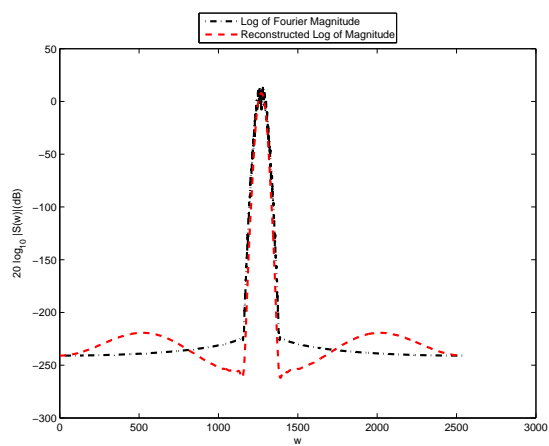
Εικόνα 3.9: Ανακατασκευή της καμπυλότητας από τους 61 πρώτους συντελεστές cepstrum - Χειρομορφή 2



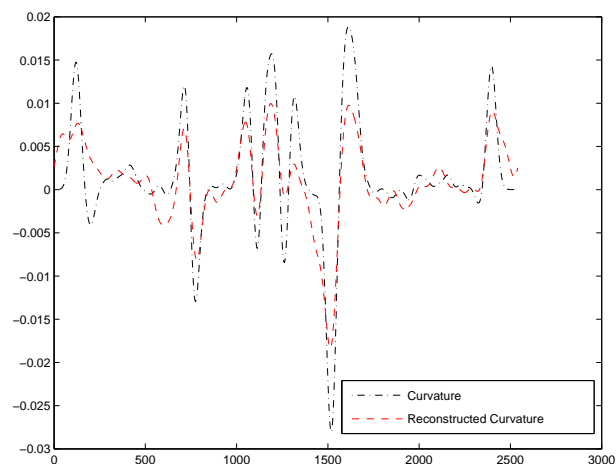
(a) Χειρομορφή



(b) Cepstrum της Καμπυλότητας

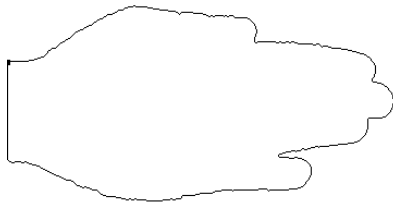


(c) Ανακατασκευασμένο Φάσμα της Καμπυλότητας

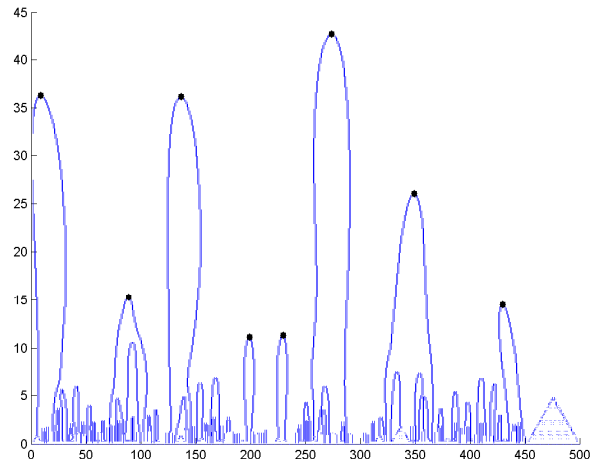


(d) Ανακατασκευή της Καμπυλότητας

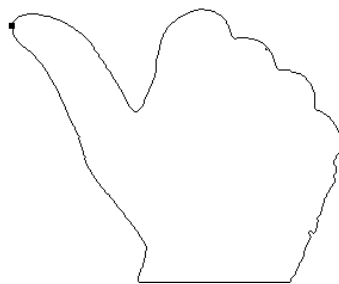
Εικόνα 3.10: Ανακατασκευή της καμπυλότητας από τους 61 πρώτους συντελεστές cepstrum - Χειρομορφή 3



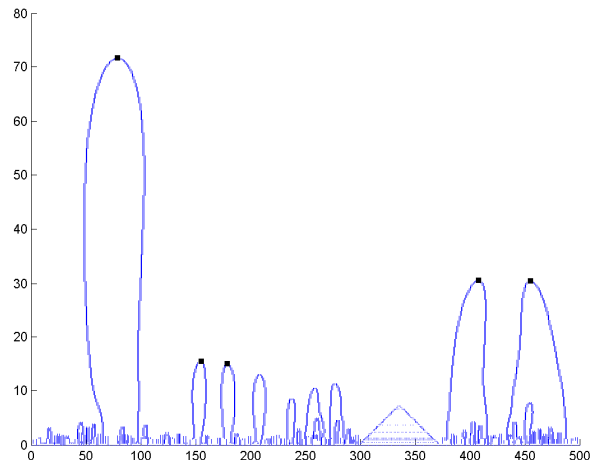
(a) Χειρομορφή



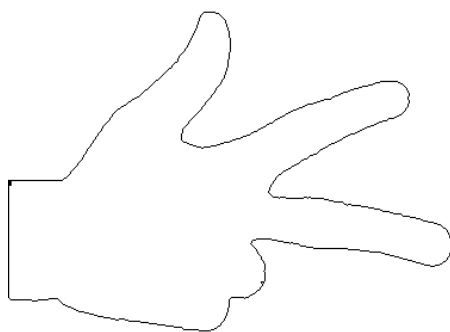
(b) Χώρος Κλίμακας Καμπυλότητας



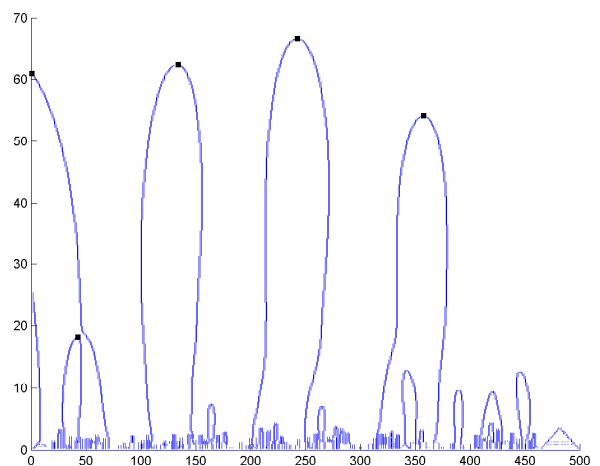
(c) Χειρομορφή



(d) Χώρος Κλίμακας Καμπυλότητας



(e) Χειρομορφή



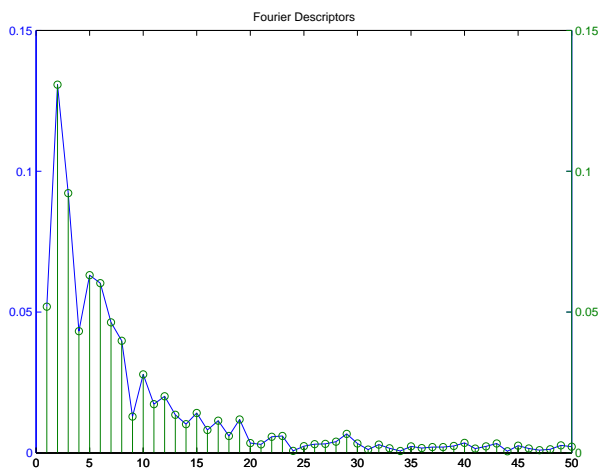
(f) Χώρος Κλίμακας Καμπυλότητας

Εικόνα 3.11: Γραμμικοί Χώροι Κλίμακας για τρεις διαφορετικές Χειρομορφές

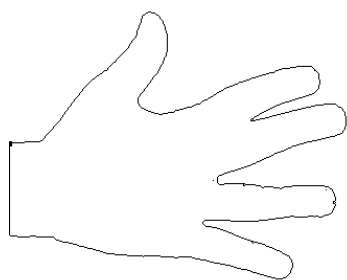




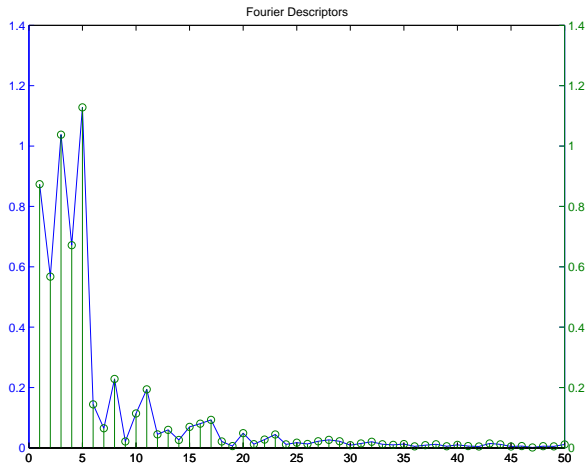
(a) Χειρομορφή



(b) Περιγραφητές Fourier

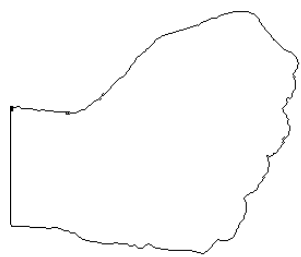


(c) Χειρομορφή

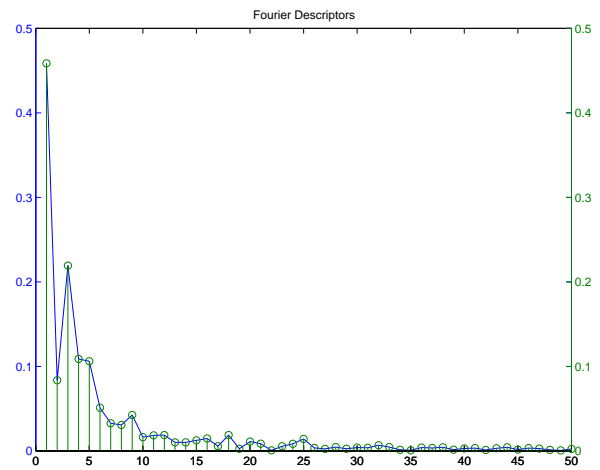


(d) Περιγραφητές Fourier

Εικόνα 3.12: Περιγραφητές Fourier - Μέρος I



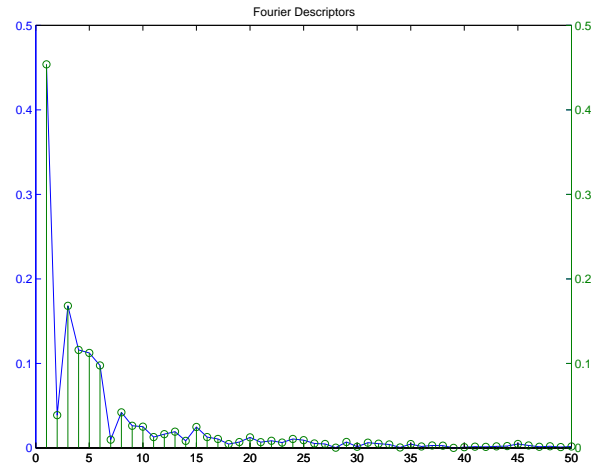
(a) Χειρομορφή



(b) Περιγραφητές Fourier



(c) Χειρομορφή



(d) Περιγραφητές Fourier

Εικόνα 3.13: Περιγραφητές Fourier - Μέρος II

---

## ΕΚΤΙΜΗΣΗ ΟΠΤΙΚΗΣ ΚΙΝΗΣΗΣ

---

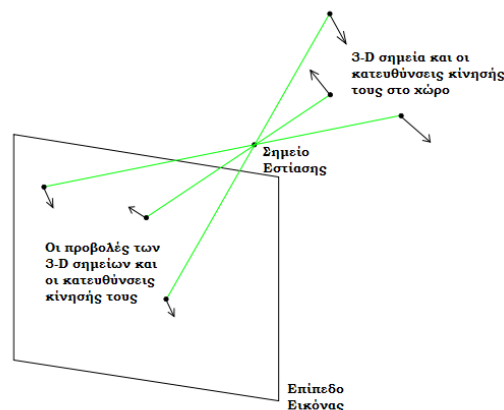
Σε πολλές εφαρμογές της όρασης υπολογιστών μας ενδιαφέρει η ανίχνευση ή και η εκτίμηση της κίνησης κάποιου αντικειμένου. Το κατεξοχήν παράδειγμα είναι η περίπτωση ενός ρομπότ, το οποίο βασίζεται στην εκτίμηση της κίνησης (όπου εδώ ως «κίνηση» ανιχνεύεται η σχετική μετακίνηση των αντικειμένων στο οπτικό πεδίο του ρομπότ λόγω της κίνησης του ίδιου του ρομπότ) για να αποφύγει τυχόν εμπόδια, να επιτελέσει κάποια συγκεκριμένη λειτουργία ή να προχωρήσει προς κάποιο προκαθορισμένο σημείο. Άλλες εφαρμογές που χρησιμοποιούν την πληροφορία της κίνησης, εκτός από το πεδίο των ρομποτικών αυτοματισμών, είναι η συμπίεση βίντεο, η ιατρική παρακολούθηση της δραστηριότητας ζωτικών οργάνων με χρήση μικροκαμερών, και οι δορυφορικές επικοινωνίες. Ειδικότερα στο πεδίο της ανάλυσης βίντεο νοηματικής γλώσσας, που μας απασχολεί στην παρούσα εργασία, η επιτυχής ανίχνευση της κίνησης αποτελεί πολύ σημαντικό συστατικό ενός αποτελεσματικού συστήματος αναγνώρισης, καθώς έχει βρεθεί ότι η κατεύθυνση κίνησης των χεριών και/ή του προσώπου του ομιλητή συμβάλλουν κατά πολύ στην κατανόηση του νοήματος που εκείνος εκτελεί, ίσως και περισσότερο από το σχήμα της χειρομορφής.

Προφανώς η κίνηση των αντικειμένων πραγματοποιείται στον τρισδιάστατο χώρο, ενώ μια κάμερα που παρακολουθεί μια σκηνή μπορεί να ανιχνεύσει μόνο διδιάστατη κίνηση. Η κίνηση αυτή αντιπροσωπεύεται στην εικόνα από κάποια μεταβολή της φωτεινότητας σε ορισμένες περιοχές. Στο παρόν κεφάλαιο μας απασχολεί, αφ' ενός η κατά το δυνατόν ορθή εκτίμηση της διδιάστατης κίνησης από τη μεταβολή αυτή, και αφ' ετέρου η χρήση της πληροφορίας αυτής για τον εντοπισμό κινούμενων αντικειμέ-

νων.

## 4.1 Πεδίο Κίνησης

Στην Όραση Υπολογιστών το πεδίο κίνησης είναι μια ιδανική αναπαράσταση της τρισδιάστατης κίνησης όπως αυτή προβάλλεται σε μια εικόνα που λαμβάνεται από μια κάμερα. Δοσμένου ενός απλοποιημένου μοντέλου για την κάμερα, κάθε σημείο  $(y_1, y_2)$  στην εικόνα είναι η προβολή ενός άλλου σημείου στην τρισδιάστατη σκηνή. Η θέση της προβολής ενός οποιουδήποτε σημείου στην εικόνα μπορεί να μεταβάλλεται με το χρόνο. Το πεδίο κίνησης μπορεί να οριστεί ως η παράγωγος της θέσης στην εικόνα όλων των σημείων αυτής που αποτελούν προβολές των ίδιων πάντα θέσεων στις 3 διαστάσεις. Αυτό σημαίνει ότι το πεδίο κίνησης μπορεί να περιγραφεί από μια συνάρτηση που απεικονίζει τις συνιστώσες θέσης ενός σημείου στην εικόνα σε ένα διδιάστατο διάνυσμα κίνησης. Το πεδίο κίνησης είναι ιδανική αναπαράσταση της προβαλλόμενης τρισδιάστατης κίνησης υπό την έννοια ότι μπορεί να οριστεί τυπικά. Ωστόσο στην πράξη μπορούμε να καθορίσουμε μια προσέγγιση μόνο του πεδίου κίνησης από τα δεδομένα της εικόνας.



Εικόνα 4.1: Το μοντέλο στενής οπής για μία κάμερα

Ένα συχνά χρησιμοποιούμενο μοντέλο για μια κάμερα είναι αυτό της στενής οπής. Αυτό παρουσιάζεται στο ακόλουθο Σχ. 4.1. Κάθε σημείο  $(x_1, x_2, x_3)$  στον τρισδιάστατο χώρο προβάλλεται σε ένα σημείο  $(y_1, y_2)$  της εικόνας σύμφωνα με δύο

συναρτήσεις απεικόνισης  $m_1, m_2$ :

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} m_1(x_1, x_2, x_3) \\ m_2(x_1, x_2, x_3) \end{bmatrix}$$

Υποθέτοντας ότι η σκηνή δεν είναι στατική, θα αποτελείται ενδεχομένως από αντικείμενα που μετακινούνται το ένα σε σχέση με το άλλο, ή που παραμορφώνονται ενώ επίσης και η κάμερα μπορεί να μετακινείται σε σχέση με τη σκηνή. Σε όλες αυτές τις περιπτώσεις, ένα δοθέν σημείο στον τρισδιάστατο χώρο θα απεικονίζεται σε μεταβαλλόμενα σημεία στην εικόνα. Παραγωγίζοντας την παραπάνω έκφραση παίρνουμε:

$$\begin{bmatrix} \frac{dy_1}{dt} \\ \frac{dy_2}{dt} \end{bmatrix} = \begin{bmatrix} \frac{dm_1(x_1, x_2, x_3)}{dt} \\ \frac{dm_2(x_1, x_2, x_3)}{dt} \end{bmatrix} = \begin{bmatrix} \frac{dm_1}{dx_1} & \frac{dm_1}{dx_2} & \frac{dm_1}{dx_3} \\ \frac{dm_2}{dx_1} & \frac{dm_2}{dx_2} & \frac{dm_2}{dx_3} \end{bmatrix} \begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \frac{dx_3}{dt} \end{bmatrix}$$

Εδώ

$$\vec{u} = \begin{bmatrix} \frac{dy_1}{dt} \\ \frac{dy_2}{dt} \end{bmatrix}$$

είναι το πεδίο κίνησης το οποίο εξαρτάται τόσο από τη θέση  $(y_1, y_2)$  όσο και από το χρόνο  $t$ . Επίσης,

$$\vec{x}' = \begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \frac{dx_3}{dt} \end{bmatrix}$$

είναι η κίνηση του αντίστοιχού 3-D σημείου. Τα δύο παραπάνω μεγέθη σχετίζονται ως εξής:

$$\vec{u} = \mathbf{M} \vec{x}'$$

όπου  $\mathbf{M}$  είναι ο εξής  $2 \times 3$  πίνακας, που εξαρτάται από τη σχέση του εξεταζόμενου σημείου στην εικόνα:

$$\mathbf{M} = \begin{bmatrix} \frac{dm_1}{dx_1} & \frac{dm_1}{dx_2} & \frac{dm_1}{dx_3} \\ \frac{dm_2}{dx_1} & \frac{dm_2}{dx_2} & \frac{dm_2}{dx_3} \end{bmatrix}$$

Η παραπάνω σχέση υποδηλώνει ότι το πεδίο κίνησης σε ένα δοσμένο σημείο της εικόνας δεν επηρεάζεται από εκείνα τα είδη τρισδιάστατης κίνησης που εμπίπτουν στο μηδενικό χώρο (null space) του  $\mathbf{M}$ . Για παράδειγμα, στην περίπτωση μιας κάμερας στενής οπής οι συνιστώσες τρισδιάστατης κίνησης που κατευθύνονται προς ή από το εστιακό σημείο της κάμερας δεν είναι δυνατόν να ανιχνευθούν με χρήση του πεδίου κίνησης.

Μπορεί να βρεθεί (βλ.[1]), ότι για το πεδίο κίνησης ισχύει η σχέση:

$$\vec{u} = f \frac{Z\vec{V} - V_z\vec{P}}{Z^2}$$

όπου

$$\vec{V} = -\vec{T} - \vec{\omega} \times \vec{P}.$$

,  $\vec{P}$  είναι ένα σημείο στη σκηνή που απέχει απόσταση  $Z$  από το εξεταζόμενο σημείο. Η  $\vec{V}$  είναι η ταχύτητα της σχετικής κίνησης της κάμερας ως προς τη σκηνή,  $\vec{T}$  είναι η συνιστώσα μετατόπισης της κίνησης και  $\vec{\omega}$  η γωνιακή ταχύτητα αυτής.

Το πεδίο κίνησης είναι μια ιδανική κατασκευή, που βασίζεται στην ιδέα ότι είναι δυνατόν να καθορίσουμε την κίνηση κάθε σημείου της εικόνας, η οποία με τη σειρά της σχετίζεται με την τρισδιάστατη κίνηση ενός σημείου στη σκηνή. Στην πράξη, ωστόσο, το πεδίο κίνησης μπορεί μόνο να προσεγγιστεί βάσει μετρήσεων στα σημεία της εικόνας. Το πρόβλημα είναι ότι στις περισσότερες περιπτώσεις το κάθε σημείο της εικόνας κινείται με την ατομική του ταχύτητα η οποία πρέπει να μετρηθεί τοπικά με χρήση κάποιου τοπικού τελεστή στην εικόνας. Κατά συνέπεια, το ακριβές συνολικό πεδίο κίνησης δε μπορεί να καθοριστεί για κάποιες τοπικές γειτονιές και αντ' αυτού χρησιμοποιείται μια προσέγγιση, γνωστή ως πεδίο οπτικής ροής. Για παράδειγμα, μια γειτονιά όπου η φωτεινότητα παραμένει σταθερή (όπως η επιφάνεια μιας περιστρεφόμενης σφαίρας σε ματ χρωματισμό, υπό σταθερό φωτισμό) μπορεί να αντιστοιχεί σε ένα μη μηδενικό πεδίο κίνησης, ωστόσο η οπτική ροή προκύπτει μηδενική καθώς δεν ανιχνεύεται τοπική κίνηση μέσα στην εικόνα. Ομοίως, μια γειτονιά που είναι εκ φύσεως μονοδιάστατη, όπως για παράδειγμα μια γραμμή ή μια ακμή, μπορεί να αντιστοιχεί σε πεδίο κίνησης οποιασδήποτε κατεύθυνσης, ωστόσο το πεδίο οπτικής ροής μπορεί μόνο να ανιχνεύσει τη συνιστώσα κίνησης που είναι κάθετη στη γραμμή. Ακόμα υπάρχουν περιπτώσεις όπου υπάρχει οπτική ροή αλλά όχι τρισδιάστατη κίνηση (όπως για παράδειγμα μια στατική σκηνή όπου μεταβάλλεται ο φωτισμός). Τέλος, διάφορα άλλα φαινόμενα, όπως επικαλύψεις αντικειμένων, που εμφανίζονται αρκετά συχνά στην τρισδιάστατη κίνηση, προκαλούν αποκλίσεις στο πεδίο οπτικής ροής σε σχέση με το πραγματικό πεδίο κίνησης.

## 4.2 Οπτική Ροή

Με τον όρο "οπτική ροή" εννοούμε την προβολή στο διδιάστατο χώρο των διανυσμάτων κίνησης των τρισδιάστατων αντικειμένων, όπως αυτά απεικονίζονται σε μια

εικόνα. Η οπτική ροή σε κάθε σημείο μιας εικόνας παριστάνεται από ένα διάνυσμα  $(u,v)$ , όπου οι  $u$  και  $v$  συνιστώσες αντιστοιχούν στην ταχύτητα κατά τον οριζόντιο και τον κατακόρυφο άξονα αντίστοιχα. Γενικά οι μέθοδοι εύρεσης της οπτικής ροής συνοψίζονται στο [5], και ακολουθούν τα παρακάτω στάδια:

1. Φιλτράρισμα και ομαλοποίηση της εικόνας με κάποιο βαθυπερατό ή ζωνοπερατό φίλτρο, για την εξάλειψη του υψίσυχνου θορύβου και αύξηση του σηματοθορυβικού λόγου.
2. Διεξαγωγή μετρήσεων στην εικόνα, όπως χρονική και χωρική παραγωγή, ή εύρεση συσχετίσεων μεταξύ περιοχών στην εικόνα, ανάλογα με τη χρησιμοποιούμενη μέθοδο.
3. Χρήση των μετρήσεων και κάποιων παραδοχών για την εύρεση της οπτικής ροής.

Γενικά οι μέθοδοι οπτικής ροής μπορούν να ομαδοποιηθούν σε κατηγορίες όπως διαφορικές τεχνικές (Horn - Schunck, Lucas - Kanade), τεχνικές σύμπτωσης περιοχών (block - matching), τεχνικές που βασίζονται στον υπολογισμό της "ενέργειας" της εικόνας και στην ελαχιστοποίηση του αντίστοιχου συναρτησιακού (Heeger) και τεχνικές φάσης (Fleet - Jepsen). Γενικά πάντως πραγματοποιείται η υπόθεση της διατήρησης της φωτεινότητας, που θεωρεί ότι η φωτεινότητα ή το χρώμα των αντικειμένων δεν αλλάζει σημαντικά μεταξύ δύο διαδοχικών καρέ μιας ακολουθίας εικόνων. Εδώ θα ασχοληθούμε μόνο με τις πρώτες δύο μεθόδους.

### 4.2.1 Μέθοδος των Lucas - Kanade

Η μέθοδος των Lucas και Kanade ([30]) εξακολουθεί να είναι μια από τις πιο δημοφιλείς μεθόδους για την εκτίμηση της κίνησης μεταξύ δύο καρέ (frames) ενός βίντεο. Ανήκει στην κατηγορία των διαφορικών μεθόδων διότι βασίζεται σε προσέγγιση Taylor του φωτεινού σήματος της εικόνας, και συνεπώς χρησιμοποιούν μερικές παραγώγους στο χώρο και στο χρόνο.

Έστω μια ακολουθία εικόνων  $I$ , και ένα pixel στη θέση  $(x, y, z, t)$  με φωτεινότητα  $I(x, y, z, t)$  στην εικόνα αναφοράς. Στο επόμενο καρέ (εικόνα ελέγχου) το pixel αυτό θα έχει μετακινηθεί κατά  $\delta x, \delta y, \delta z, \delta t$ , οπότε η διατήρηση της φωτεινότητας δίνει:

$$I(x, y, z, t) = I(x + \delta x, y + \delta y, z + \delta z, t + \delta t)$$

Υποθέτοντας ότι η κίνηση είναι επαρκώς μικρή, μπορούμε να αναπτύξουμε την παραπάνω εξίσωση κατά Taylor και να πάρουμε:

$$I(x + \delta x, y + \delta y, z + \delta z, t + \delta t) = I(x, y, z, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial z} \delta z + \frac{\partial I}{\partial t} \delta t + H.O.T.$$

όπου H.O.T σημαίνει όροι υψηλότερης τάξεως, που είναι δυνατόν να αγνοηθούν. Από τις εξισώσεις αυτές προκύπτει ότι

$$\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial z} \delta z + \frac{\partial I}{\partial t} \delta t = 0$$

ή

$$\frac{\partial I}{\partial x} \frac{\delta x}{\delta t} + \frac{\partial I}{\partial y} \frac{\delta y}{\delta t} + \frac{\partial I}{\partial z} \frac{\delta z}{\delta t} + \frac{\partial I}{\partial t} \frac{\delta t}{\delta t} = 0$$

που συνεπάγεται ότι

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial z} V_z + \frac{\partial I}{\partial t} = 0$$

όπου  $V_x, V_y, V_z$  είναι οι x,y και z συνιστώσες της ταχύτητας, ή αλλιώς η οπτική ροή για την εικόνα  $I(x, y, z, t)$ , και  $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial z}$  and  $\frac{\partial I}{\partial t}$  είναι οι παράγωγοι της εικόνας στο σημείο  $(x, y, z, t)$  στις αντίστοιχες διευθύνσεις. Οι παραπάνω παράγωγοι μπορούν επίσης να γραφούν ως  $I_x, I_y, I_z$  και  $I_t$ , οπότε η παραπάνω σχέση γράφεται

$$I_x V_x + I_y V_y + I_z V_z = -I_t \Rightarrow \nabla I \cdot \vec{V} = -I_t \quad (4.1)$$

Η παραπάνω εξίσωση έχει τρεις αγνώστους και ως εκ τούτου δε μπορεί να λυθεί άμεσα. Το γεγονός αυτό είναι γνωστό ως το πρόβλημα του διαφράγματος (aperture problem). Για να βρεθεί το πεδίο οπτικής ροής χρειάζεται ακόμα ένα σύνολο εξισώσεων, το οποίο θα προκύψει από κάποιον επιπρόσθετο περιορισμό. Η λύση που έδωσαν οι Lucas και Kanade είναι μια μη επαναληπτική μέθοδος που υποθέτει ότι η οπτική ροή είναι τοπικά σταθερή.

Πράγματι, υποθέτοντας ότι η ροή  $(V_x, V_y, V_z)$  είναι σταθερή σε ένα μικρό παραθύρο γύρω από το εξεταζόμενο pixel, διαστάσεων  $m \times m \times m$  με  $m > 1$ , και αριθμώντας τα pixels εντός του παραθύρου από  $1 \dots n$ , μπορούμε να διατυπώσουμε το εξής σύνολο εξισώσεων:

$$\begin{aligned} I_{x_1} V_x + I_{y_1} V_y + I_{z_1} V_z &= -I_{t_1} \\ I_{x_2} V_x + I_{y_2} V_y + I_{z_2} V_z &= -I_{t_2} \\ &\vdots \\ I_{x_n} V_x + I_{y_n} V_y + I_{z_n} V_z &= -I_{t_n} \end{aligned}$$



Με τις εξισώσεις αυτές έχουμε παραπάνω από 3 εξισώσεις για τους 3 αγνώστους, και συνεπώς το σύστημα εξισώσεων

$$\begin{bmatrix} I_{x_1} & I_{y_1} & I_{z_1} \\ I_{x_2} & I_{y_2} & I_{z_2} \\ \vdots & \vdots & \vdots \\ I_{x_n} & I_{y_n} & I_{z_n} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix} = \begin{bmatrix} -I_{t_1} \\ -I_{t_2} \\ \vdots \\ -I_{t_n} \end{bmatrix}$$

ή

$$A\vec{v} = -b$$

είναι υπερορισμένο. Έτσι, εφαρμόζουμε για την επίλυσή του τη μέθοδο των ελαχίστων τετραγώνων:

$$A^T A\vec{v} = A^T(-b)$$

ή

$$\vec{v} = (A^T A)^{-1} A^T(-b)$$

ή

$$\begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix} = \begin{bmatrix} \sum I_{x_i}^2 & \sum I_{x_i} I_{y_i} & \sum I_{x_i} I_{z_i} \\ \sum I_{x_i} I_{y_i} & \sum I_{y_i}^2 & \sum I_{y_i} I_{z_i} \\ \sum I_{x_i} I_{z_i} & \sum I_{y_i} I_{z_i} & \sum I_{z_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_{x_i} I_{t_i} \\ -\sum I_{y_i} I_{t_i} \\ -\sum I_{z_i} I_{t_i} \end{bmatrix}$$

με τα αθροίσματα να εκτείνονται από  $i=1$  ως  $n$ .

Η παραπάνω εξίσωση υποδηλώνει ότι η οπτική ροή μπορεί να βρεθεί υπολογίζοντας τις παραγώγους τις εικόνες και στις 4 διαστάσεις (3 αν πρόκειται για διδιάστατη εικόνα). Συνήθως προστίθεται κάποια συνάρτηση βάρους  $W(i, j, k)$ ,  $i, j, k \in [1, m]$  που ενισχύει τη συνεισφορά του κεντρικού pixel εντός του παραθύρου. Οι γκαουσιανές συναρτήσεις προτιμώνται για το σκοπό αυτό, χωρίς βέβαια να αποκλείονται και άλλου τύπου συναρτήσεις.

Εκτός από τον υπολογισμό τοπικών μετατοπίσεων, το μοντέλο ροής μπορεί να επεκταθεί και σε αφινικές μεταβολές. Χαρακτηριστικό του αλγορίθμου LK είναι ότι δεν είναι ικανός να παράξει ένα ιδιαίτερα πυκνό πεδίο από διανύσματα οπτικής ροής, ιδιότητα κοινή για όλες τις μεθόδους εκτίμησης του πεδίου ροής που βασίζονται σε τοπική πληροφορία γύρω από κάθε pixel. Με άλλα λόγια, η χρήσιμη πληροφορία του πεδίου ροής μειώνεται γρήγορα στις περιοχές των συνόρων μεταξύ αντικειμένων και στο εσωτερικό μεγάλων ομογενών περιοχών χωρίς ιδιαίτερη υφή, όπου δε μπορεί να ανιχνευθεί ισχυρή κίνηση. Το πλεονέκτημα της εν λόγω μεθόδου είναι ότι είναι αρκετά εύρωστη απέναντι στο θόρυβο.

### 4.2.2 Μέθοδος των Horn - Schunck

Η μέθοδος των Horn-Schunck είναι μια μέθοδος καθολικής εκτίμησης του πεδίου ροής που εισάγει ένα καθολικό κριτήριο ομαλότητας προκειμένου να επιλύσει το πρόβλημα του διαφράγματος. Εδώ ζητείται η ελαχιστοποίηση του εξής συναρτησιακού ενέργειας:

$$f = \iiint ((\nabla I \cdot \vec{V} + I_t)^2 + \alpha(|\nabla V_x|^2 + |\nabla V_y|^2 + |\nabla V_z|^2)) dx dy dz \quad (4.2)$$

όπου  $\nabla I = \begin{bmatrix} I_x \\ I_y \\ I_z \end{bmatrix}$  είναι οι παράγωγοι της εικόνας κατά μήκος των x,y, και z διευ-

θύνσεων, και  $I_t$  είναι η παράγωγος στο χρόνο,  $\vec{V}$  είναι το διάνυσμα οπτικής ροής με συνιστώσες  $V_x, V_y, V_z$ . Η παράμετρος  $\alpha$  είναι μια σταθερά κανονικοποίησης, με τις μεγαλύτερες τιμές του  $\alpha$  να οδηγούν σε ομαλότερο πεδίο ροής. Η ελαχιστοποίηση της συνάρτησης αυτής μπορεί να γίνει υπολογίζοντας τις αντίστοιχες εξισώσεις Euler-Lagrange:

$$\Delta V_x - \frac{1}{\alpha} I_x (I_x V_x + I_y V_y + I_z V_z + I_t) = 0 \quad (4.3)$$

$$\Delta V_y - \frac{1}{\alpha} I_y (I_x V_x + I_y V_y + I_z V_z + I_t) = 0 \quad (4.4)$$

$$\Delta V_z - \frac{1}{\alpha} I_z (I_x V_x + I_y V_y + I_z V_z + I_t) = 0 \quad (4.5)$$

όπου  $\Delta$  συμβολίζει τον τελεστή Laplace, έτσι ώστε  $\Delta V_x = \frac{\partial}{\partial x} \frac{\partial V_x}{\partial x}$ ,  $\Delta V_y = \frac{\partial}{\partial y} \frac{\partial V_y}{\partial y}$ ,  $\Delta V_z = \frac{\partial}{\partial z} \frac{\partial V_z}{\partial z}$ . Επιλύοντας τις εξισώσεις αυτές με τη μέθοδο Gauss - Seidel ως προς τις συνιστώσες  $V_x, V_y, V_z$  της οπτικής ροής παίρνουμε το εξής επαναληπτικό σχήμα:

$$V_x^{k+1} = \frac{\Delta V_x^k - \frac{1}{\alpha} I_x (I_y V_y^k + I_z V_z^k + I_t)}{\frac{1}{\alpha} I_x^2} \quad (4.6)$$

$$V_y^{k+1} = \frac{\Delta V_y^k - \frac{1}{\alpha} I_y (I_x V_x^k + I_z V_z^k + I_t)}{\frac{1}{\alpha} I_y^2} \quad (4.7)$$

$$V_z^{k+1} = \frac{\Delta V_z^k - \frac{1}{\alpha} I_z (I_x V_x^k + I_y V_y^k + I_t)}{\frac{1}{\alpha} I_z^2} \quad (4.8)$$

όπου ο εκθέτης  $k+1$  υποδηλώνει την επόμενη επανάληψη και ο  $k$  δείχνει στο πιο πρόσφατα υπολογισμένο αποτέλεσμα. Οι τιμές  $\Delta V_i$  μπορούν να ληφθούν ως εξής:

$$\Delta V_i = \sum_{N(p)} V_i(N(p)) - V_i(p)$$

όπου  $N(p)$  είναι η γειτονιά με τους 6 πλησιέστερους γείτονες του pixel  $p$ .

Μια εναλλακτική υλοποίηση του αλγορίθμου HS με τη μέθοδο Jacobi δίνει τα εξής:

$$V_x^{k+1} = \overline{V_x^k} - \frac{I_x(I_x \overline{V_x^k} + I_y \overline{V_y^k} + I_z \overline{V_z^k} + I_t)}{\alpha^2 + I_x^2 + I_y^2 + I_z^2} \quad (4.9)$$

$$V_y^{k+1} = \overline{V_y^k} - \frac{I_y(I_x \overline{V_x^k} + I_y \overline{V_y^k} + I_z \overline{V_z^k} + I_t)}{\alpha^2 + I_x^2 + I_y^2 + I_z^2} \quad (4.10)$$

$$V_z^{k+1} = \overline{V_z^k} - \frac{I_z(I_x \overline{V_x^k} + I_y \overline{V_y^k} + I_z \overline{V_z^k} + I_t)}{\alpha^2 + I_x^2 + I_y^2 + I_z^2} \quad (4.11)$$

όπου  $\overline{V_i^k}$  αντιστοιχεί στο μέσο όρο του  $V_i^k$  στην περιοχή του τρέχοντος pixel.

Το πλεονέκτημα του αλγορίθμου HS είναι ότι μπορεί να δώσει μεγάλη πυκνότητα διανυσμάτων ροής, δηλαδή η πληροφορία που λείπει από το εσωτερικό ομογενών περιοχών συμπληρώνεται από τα όρια των περιοχών αυτών. Ωστόσο είναι πιο ευαίσθητος στο θόρυβο απ' ό τι οι τοπικές μέθοδοι όπως ο LK.

### Πυραμιδωτή Εκτέλεση των Αλγορίθμων Οπτικής Ροής

Στο [7] προτείνεται η ιεραρχική υλοποίηση των αλγορίθμων Οπτικής Ροής, που εφαρμόζεται και στους δύο παραπάνω αλγορίθμους. Συγκεκριμένα εξετάζεται η κατασκευή μιας Λαπλασιανής πυραμίδας, όπου στο χαμηλότερο επίπεδο βρίσκονται οι εικόνες στην αρχική τους μορφή, ενώ σε ανώτερα επίπεδα κάποιες ομαλοποιημένες και υποδειγματοληπτημένες εκδοχές αυτών. Ξεκινώντας από τα χαμηλότερα επίπεδα (στην κορυφή της πυραμίδας) εκτελούνται οι αλγόριθμοι οπτικής ροής, και τα αποτελέσματά τους χρησιμοποιούνται ως μια αρχική εκτίμηση για την οπτική ροή του επόμενου επιπέδου. Το σύστημα που προτείνεται αποτελείται από 4 στάδια:

1. Κατασκευή της Λαπλασιανής πυραμίδας.
2. Υπολογισμός του πεδίου ροής στο τρέχον επίπεδο, με οποιονδήποτε αλγόριθμο.
3. Χρησιμοποίηση του υπολογισμένου πεδίου ροής για την εκτίμηση της εικόνας ελέγχου, και σύγκριση για εύρεση του σφάλματος εκτίμησης.
4. Μεταφορά των παραμέτρων του μοντέλου εκτίμησης της οπτικής ροής (εάν υπάρχουν) και μεταφορά στο επόμενο επίπεδο. Το αποτέλεσμα του τρέχοντος επιπέδου τροποποιείται κατάλληλα ώστε να αντιστοιχεί στις διαστάσεις του επόμενου επιπέδου.

Η εκτέλεση των αλγορίθμων σε ιεραρχική πυραμίδα εξυπηρετεί τόσο από άποψη χρόνου εκτέλεσης, καθώς το πεδίο ροής υπολογίζεται ταχύτατα στα υψηλότερα επίπεδα και η αρχική εκτίμηση βοηθά στην επιτάχυνση της διαδικασίας στα χαμηλότερα, όσο και από την άποψη της ποιότητας των αποτελεσμάτων. Για παράδειγμα, η χρήση της πυραμίδας επιτρέπει σε τοπικές μεθόδους όπως ο αλγόριθμος LK να ανιχνεύσουν με επιτυχία και μεγαλύτερες μετατοπίσεις απ' ότι ο απλός αλγόριθμος (που μπορεί να ανιχνεύσει πολύ μικρές μόνο κινήσεις), καθώς οι μετατοπίσεις αυτές ανιχνεύονται με επιτυχία στα υψηλότερα επίπεδα όπου το μέγεθός τους έχει μειωθεί σημαντικά, και μεταφέρονται κατόπιν στα χαμηλότερα επίπεδα.

### 4.3 Block Matching

Ο αλγόριθμος σύμπτωσης block (block-matching) είναι μια μέθοδος εντοπισμού αντίστοιχων blocks σε μια ακολουθία καρτέ ενός ψηφιακού βίντεο, και χρησιμοποιείται κυρίως σε εφαρμογές συμπίεσης βίντεο για την πρόβλεψη κίνησης. Στόχος της τεχνικής block-matching είναι να βρει ένα block σε ένα καρτέ  $j$  που να αντιστοιχεί, σύμφωνα με κάποιο κριτήριο ομοιότητας, σε κάποιο προηγούμενο ή και επόμενο καρτέ  $i$ . Η αντιστοίχιση των blocks εξυπηρετεί στη μείωση της περιττής χρονικής πληροφορίας που εμπεριέχεται στην ακολουθία των καρτέ, καθώς κάποια από τα καρτέ μπορούν να παραληφθούν και αντ' αυτών να χρησιμοποιηθούν οι εκτιμήσεις αυτών όπως προκύπτουν από την πρόβλεψη κίνησης.

Συγκεκριμένα, έστω  $I(x, y, t) = I(p, t)$  μια χωροχρονικά ακολουθία εικόνων στην οποία παρακολουθείται ένα κινούμενο αντικείμενο, όπου  $p = (x, y)$  είναι το χωρικό διάνυσμα ταχύτητας. Η απλή μέθοδος block-matching αναλαμβάνει το χωρισμό μιας εικόνας  $I(p, t_1)$  σε blocks  $R$  σταθερού μεγέθους, και την ελαχιστοποίηση της εξής ποσότητας

$$E = \sum_{p \in R} |I(p, t_1) - I(p + d, t_2)|^2 \quad (4.12)$$

για κάποια άλλη εικόνα  $I(p, t_2)$  προκειμένου να βρεθεί το βέλτιστο διάνυσμα μετατόπισης  $d = (d_x, d_y)$ . Η ελαχιστοποίηση της ποσότητας 4.12 προκύπτει άμεσα ότι σχετίζεται με την εύρεση εκείνου του  $d$  που μεγιστοποιεί τη συσχέτιση

$$\sum_{p \in R} I(p, t_1)I(p + d, t_2)$$

Ένα από τα βασικότερα προβλήματα αυτής της μεθόδου είναι οι υψηλές υπολογιστικές της απαιτήσεις. Για το λόγο αυτό αντί να γίνεται καθολική αναζήτηση για το

block σε όλη την εικόνα, συνήθως η αναζήτηση περιορίζεται σε μια περιοχή γύρω από την αρχική θέση του block. Ακόμα και τότε, οι υπολογισμοί που απαιτούνται είναι συχνά υπερβολικά πολλοί, με αποτέλεσμα να αναπτυχθεί ένας αριθμός από ταχύτερους αλγορίθμους αναζήτησης, που απαριθμούνται παρακάτω:

1. **Αναζήτηση τριών βημάτων:** Πρόκειται για μια από τις παλαιότερες προσπάθειες για την επιτάχυνση της αναζήτησης. Βασίζεται στην επαναληπτική αναζήτηση για το block σε 8 κατευθύνσεις (9 μαζί με την περίπτωση το block να μην μετακινήθηκε) γύρω από την τρέχουσα θέση του block. Σε κάθε βήμα, μόλις εντοπιστεί το σημείο που ελαχιστοποιεί το κόστος (4.12), το βήμα (γειτονιά) αναζήτησης υποδιπλασιάζεται και η αναζήτηση επαναλαμβάνεται. Με τον τρόπο αυτό η υπολογιστική πολυπλοκότητα μειώνεται κατά 9 φορές. Ο αλγόριθμος καλείται «τριών βημάτων» γιατί το βήμα αναζήτησης συνήθως αρχικοποιείται με την τιμή  $S=4$ , οπότε μέχρι το  $S=1$  μεσολαβούν 3 βήματα.
2. **«Προχωρημένη» αναζήτηση τριών βημάτων:** Η μέθοδος αυτή βελτιώνει τα αποτελέσματα της προηγούμενης δίνοντας μεγαλύτερη έμφαση στην περίπτωση η νέα θέση να είναι κοντά, αλλά όχι ακριβώς, στην αρχική του θέση, που συνήθως αντιμετωπίζεται προβληματικά από την αναζήτηση τριών βημάτων. Στο αρχικό βήμα, αντί των 8 θέσεων αναζήτησης της προηγούμενης μεθόδου, που βρίσκονται σε απόσταση  $S$  από το block (συνήθως  $S=4$ ) χρησιμοποιούνται επιπλέον άλλες 8 υποψήφιας θέσεις σε κοντινότερη απόσταση (π.χ. για  $S=1$ ). Αν το κόστος ελαχιστοποιείται σε μια από τις νέες αυτές τις θέσεις, τότε αλλάζουμε το κέντρο αναζήτησης στη νέα αυτή θέση και εξετάζουμε τη συνάρτηση κόστους στις γειτονικές θέσεις του νέου κέντρου. Αν πάλι το κόστος στο πρώτο βήμα ελαχιστοποιηθεί σε κάποια από τις θέσεις με  $S=4$ , τότε αλλάζουμε πάλι το κέντρο της αναζήτησης και χρησιμοποιούμε κανονικά τον αλγόριθμο των τριών βημάτων. Η διαδικασία αυτή, αν και μπορεί να χρειαστεί ένα ελάχιστο 17 σημείων για να ελέγξει κάθε block, στη χειρότερη περίπτωση θα αναγκαστεί να ελέγξει 33 θέσεις.
3. **«Απλή και Αποδοτική» αναζήτηση:** Η μέθοδος αυτή μοιάζει με την αναζήτηση τριών βημάτων, μόνο που εδώ το κάθε σήμα αποτελείται από δύο στάδια. Βασίζεται δε στην υπόθεση ότι για ένα block δεν είναι δυνατόν να ελαχιστοποιείται η συνάρτηση κόστους για 2 διαφορετικές θέσεις που βρίσκονται σε αντίθετες κατευθύνσεις μεταξύ τους, και συνεπώς η αναζήτηση και στις 8 θέσεις δεν είναι πλέον απαραίτητη. Στο πρώτο στάδιο κάθε βήματος εξετάζονται

2 από τις 8 θέσεις σε απόσταση  $S$  γύρω από το κέντρο σε κάθετες μεταξύ τους διευθύνσεις, συν η περίπτωση το block να μην μετακινήθηκε. Ανάλογα με τις τιμές που παίρνουμε για την συνάρτηση κόστους και τις σχετικές διαφορές τους επιλέγουμε στο δεύτερο στάδιο ένα από τα τέσσερα τεταρτημόρια γύρω από το αρχικό κέντρο αναζήτησης, και το νέο κέντρο τοποθετείται σε εκείνο το σημείο του επιλεγμένου τεταρτημορίου που εμφανίζει το μικρότερο κόστος. Στη συνέχεια επαναλαμβάνεται η αναζήτηση με μικρότερο βήμα, όπως στη μέθοδο των 3 βημάτων. Η μέθοδος αυτή είναι σημαντικά ταχύτερη, ωστόσο συχνά τα αποτελέσματά της είναι προβληματικά.

4. **Αναζήτηση τεσσάρων βημάτων:** Η μέθοδος αυτή μοιάζει αρκετά με την «προχωρημένη» αναζήτηση τριών βημάτων. Αποτελείται από 4 βήματα. Στο πρώτο βήμα πάντοτε  $S=2$ , και εξετάζονται 9 γειτονικές θέσεις σε ένα παράθυρο  $5 \times 5$  της εικόνας. Αν το ελάχιστο κόστος προκύψει σε μια από τις 8 γειτονικές θέσεις πλην της αρχικής, τότε μετακινούμε το κέντρο στη θέση αυτή και στα βήματα 2 και 3 επαναλαμβάνεται η αναζήτηση σε  $5 \times 5$  παράθυρο. Διαφορετικά μεταβαίνουμε κατευθείαν στο βήμα 4. Στο 4ο βήμα θέτουμε  $S=1$  και εξετάζουμε πλέον ένα  $3 \times 3$  παράθυρο. Η αναζήτηση αυτή ελέγχει στην καλύτερη περίπτωση 17 θέσεις και στη χειρότερη 27.
5. **Αναζήτηση «διαμαντιού»:** Η αναζήτηση αυτή είναι πολύ παρόμοια με την αναζήτηση 4 βημάτων, με τη διαφορά ότι εδώ η περιοχή αναζήτησης έχει σχήμα ρόμβου (σε δύο διαφορετικά μεγέθη) και όχι τετράγωνο, και ότι δεν υπάρχει περιορισμός στον αριθμό των βημάτων που χρησιμοποιούνται. Η μέθοδος αυτή δίνει πολύ καλή εκτίμηση για το ολικό ελάχιστο της συνάρτησης κόστους, με σημαντικά μικρότερο υπολογιστικό κόστος σε σχέση με την καθολική αναζήτηση.
6. **Προσαρμοστική αναζήτηση:** Η αναζήτηση αυτή βασίζεται στην προηγούμενη, ωστόσο αξιοποιεί το γεγονός ότι η γενικότερη κίνηση σε ένα καρέ εμφανίζει συνήθως σημαντικό ποσοστό συνάφειας, δηλαδή αν τα blocks που γειτνιάζουν προς το τρέχον block κινούνται προς μια συγκεκριμένη κατεύθυνση, τότε είναι πολύ πιθανό ότι και το τρέχον block θα μετακινείται προς την ίδια κατεύθυνση. Εκτός από την αναμενόμενη θέση, εξετάζεται και μια σειρά θέσεων σε μια περιοχή γύρω από αυτήν, της οποίας η διάσταση προσαρμόζεται ανάλογα με το μέγεθος της εκτιμώμενης μετατόπισης.

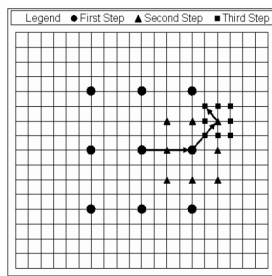
Εκτός των παραπάνω έχουν αναπτυχθεί και μέθοδοι υλοποίησης της τεχνικής block matching και στο πεδίο της (χωρικής) συχνότητας, με χρήση του μετασχηματισμού Fourier, που επιτυγχάνουν επιτάχυνση της διαδικασίας. Τα παραπάνω σχήματα αναζήτησης παρουσιάζονται συνοπτικά στο Σχ. 4.2.

Η μέθοδος του block matching έχει δύο βασικά μειονεκτήματα. Αφ' ενός, είναι πολύ πιθανό μια ορθογωνική περιοχή  $R$  στην εικόνα  $I(p, t_1)$  να αντιστοιχεί σε μια περιοχή  $R'$  στην εικόνα  $I(p, t_2)$  με σχήμα διαφορετικό του ορθογωνίου λόγω της παραμόρφωσης που μπορεί να υφίστανται περιοχές της σκηνής καθώς προβάλλονται στην εικόνα σε διαφορετικές χρονικές στιγμές. Επίσης, οι περιοχές  $R$  και  $R'$  μπορεί να διαφέρουν όχι μόνο ως προς τη θέση τους, αλλά και ως προς τη φωτεινότητά τους, λόγω τυχόν μεταβολών στο φωτισμό ή στη γεωμετρία/θέση της κάμερας. Από την άλλη πλευρά, εφ' όσον αντιμετωπιστούν τα παραπάνω προβλήματα, η μέθοδος block matching εμφανίζει το πλεονέκτημα της ευρωστίας στο θόρυβο και της δυνατότητας εξαγωγής αρκετά πυκνών πεδίων ταχύτητας.

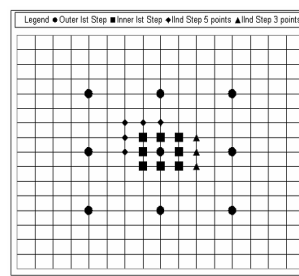
Μια σημαντική προσθήκη στον απλό αλγόριθμο block matching είναι η μέθοδος affine block matching, που επιλύει μερικώς τα παραπάνω ζητήματα επιτρέποντας στα blocks να αλλάζουν σχήμα από τη μία εικόνα στην άλλη, υφιστάμενα αφινικούς μετασχηματισμούς, τόσο ως προς το σχήμα τους, όσο και ως προς τη φωτεινότητα τους. Το affine block matching εμφανίζει θεαματικά καλύτερα αποτελέσματα, με μεγάλη ωστόσο αύξηση της υπολογιστικής πολυπλοκότητας, και δε θα μας απασχολήσει περαιτέρω στην παρούσα εργασία.

## 4.4 Παρακολούθηση Κίνησης: Οπτική Ροή και GACs

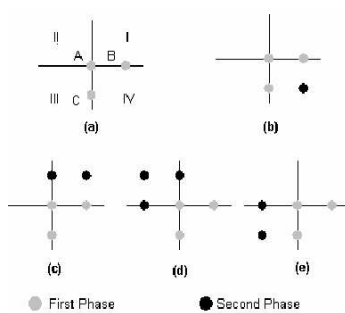
Εκτός από την ανίχνευση και εκτίμηση της κίνησης, στις περισσότερες εφαρμογές, ανάμεσά τους και η αναγνώριση νοηματικής γλώσσας, είναι χρήσιμη και η παρακολούθηση του κινούμενου αντικειμένου στην εικόνα. Γενικά, ένα σύστημα οπτικής παρακολούθησης αντικειμένων αποτελείται από δύο βασικά συστατικά: ένα μοντέλο αναπαράστασης του αντικειμένου-«στόχου» και μια εκτίμηση για την αρχική θέση αυτού, και μια διαδικασία φιλτραρίσματος ή/και συσχέτισης δεδομένων. Η αναπαράσταση του στόχου και ο εντοπισμός της αρχικής του θέσης είναι μια διαδικασία με συχνά μικρότερο υπολογιστικό κόστος, που μπορεί να υλοποιείται μέσω κάποιου αλγορίθμου κατάτμησης και απομόνωσης περιοχών (blob detection), σύμπτωσης προτύπων ή αντιστοίχισης χαρακτηριστικών, μοντελοποίησης του αντικειμένου, ή ανίχνευσης συνόρου όπως ο αλγόριθμος γεωδαιτικών περιγραμμάτων. Στο στάδιο του φιλτραρίσματος και



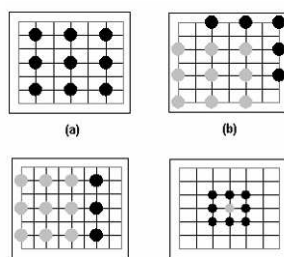
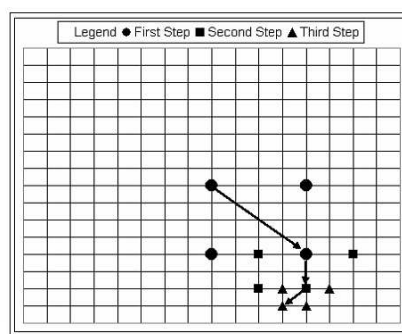
(a) Αναζήτηση 3 βημάτων



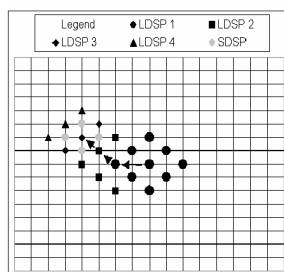
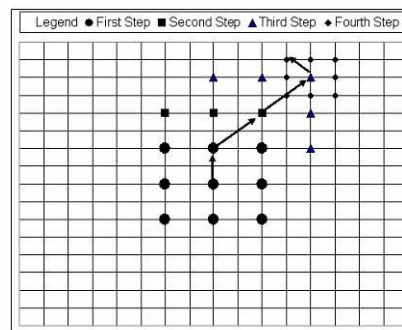
(b) Προχωρημένη αναζήτηση 3 βημάτων



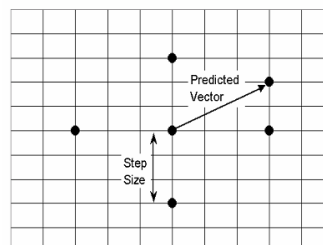
(c) «Απλή και Αποδοτική» αναζήτηση



(d) Αναζήτηση 4 βημάτων



(e) Αναζήτηση «διαμενιού»



(f) Προσαρμοστική αναζήτηση

Εικόνα 4.2: Αλγόριθμοι ταχείας αναζήτησης για την τεχνική block matching. Από [4]



συσχέτισης δεδομένων συνήθως ακολουθείται μια μεθοδολογία «από πάνω προς τα κάτω» (top-down) που περιλαμβάνει τη χρήση πρότερης πληροφορίας για τη σκηνή ή το αντικείμενο, την κίνηση αυτών, και τη διατύπωση και αποτίμηση της ακρίβειας ορισμένων υποθέσεων που οδηγούν τελικά σε μια εκτίμηση για την επόμενη θέση του αντικειμένου. Δύο ευρέως χρησιμοποιούμενες μέθοδοι είναι το φίλτρο Kalman, που υλοποιεί ένα βέλτιστο εκτιμητή για γραμμικές συναρτήσεις και γκαουσιανό θόρυβο, και η τεχνική particle filtering, που είναι χρήσιμη για τη μοντελοποίηση κατανομών μη γραμμικών και μη γκαουσιανών στοχαστικών διαδικασιών.

Στο τμήμα αυτό θα περιγράψουμε μια μέθοδο για την παρακολούθηση κινούμενων αντικειμένων που βασίζεται στα γεωδαιτικά ενεργά περιγράμματα, και ενοποιεί τη διαδικασία ανίχνευσης κίνησης με τη διαδικασία παρακολούθησής της. Περιλαμβάνει δε ένα στάδιο στατιστικής ανάλυσης που αποσκοπεί στο να παρέξει την πληροφορία κίνησης, και ένα στάδιο κατάτμησης, κατά το οποίο ένα γεωδαιτικό περίγραμμα χρησιμοποιεί αυτήν την πληροφορία, μαζί με την πληροφορία των ακμών, για τον εντοπισμό των κινούμενων αντικειμένων. Η μέθοδος αυτή παρουσιάστηκε στο [13].

Θεωρούμε δύο διαδοχικά καρέ ενός βίντεο, τις εικόνες  $I_1(\vec{x}) = I(\vec{x}, t)$  και  $I_2(\vec{x}) = I(\vec{x}, t + 1)$ . Για την ανίχνευση της κίνησης σε μια εικόνα χρησιμοποιούμε τη διαφορά μεταξύ των δύο εικόνων  $D = I_1 - I_2$  για να εξάγουμε την πληροφορία της κίνησης. Συγκεκριμένα, θεωρούμε ότι η εικόνα  $D$  ότι αποτελείται από δύο συνιστώσες, τη στατική  $S$  (που ανήκει στο φόντο και στις 2 εικόνες) και την κινούμενη  $M$ , κάθε μια από τις οποίες μπορεί να μοντελοποιηθεί από μια στατιστική κατανομή. Αν  $p_D$  είναι η κατανομή πυκνότητας πιθανότητας της εικόνας  $D$ , όπου το  $d$  ανήκει στο πεδίο τιμών της  $D$ , τότε η  $p_D$  θα είναι ένα μίγμα αυτών των δύο συνιστωσών. Έστω  $p_S(d)$  η υπό συνθήκη πιθανότητα για την στατική περιοχή ( $p_S(d) = P(D(\vec{x}) = d | \vec{x} \in S)$ ) και  $p_M(d)$  η υπό συνθήκη πιθανότητα για την κινούμενη περιοχή, δεδομένης της  $D$ . Επίσης, έστω  $P_M$  η a-priori πιθανότητα να βρισκόμαστε στο στατικό μέρος της εικόνας, και  $P_S$  αντίστοιχα στο κινούμενο. Θεωρούμε ότι αυτές οι πιθανότητες είναι ανεξάρτητες από τη θέση του pixel  $\vec{x}$  στην εικόνα ενώ επιπλέον οι  $P_M, P_S$  είναι σταθερές. Τότε προφανώς θα ισχύει:

$$p_D(d) = P_S p_S(d) + P_M p_M(d)$$

Στο [13] ως υποψήφιες κατανομές για τις  $p_S(d)$ ,  $p_M(d)$  θεωρούνται οι Gaussian

$$p_X(d) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp -\frac{(d - \mu_X)^2}{2\sigma_X^2}, \quad \Theta_X = \{\sigma_X, \mu_X\}, \quad X \in \{S, M\}$$

ή οι Laplacian

$$p_X(d) = \frac{\lambda_X}{2} \exp -\lambda_X |d - \mu_X|, \quad \Theta_X = \{\lambda_X, \mu_X\}, \quad X \in \{S, M\}$$

Επιπλέον μπορούμε να θεωρήσουμε ότι κάθε μια από τις συνιστώσες της  $D$  αποτελείται από επιμέρους αντικείμενα, οπότε στις παραπάνω εξισώσεις θα εμφανιστεί το αντίστοιχο άθροισμα για όλα τα αντικείμενα κάθε συνιστώσας.

Όσον αφορά στις άγνωστες παραμέτρους  $\Theta$  του μοντέλου, μπορούμε να επιβάλουμε κάποιους περιορισμούς για να μειωθεί ο αριθμός τους. Συγκεκριμένα, γνωρίζοντας ότι οι τιμές της  $D$  στην περιοχή του ακίνητου φόντου θα προκύψουν από τις διαφορές φωτεινότητας πρακτικά των ίδιων σημείων  $\vec{x}$  σε διαφορετικές χρονικές στιγμές, μπορούμε εύλογα να υποθέσουμε ότι οι τιμές αυτές θα αντιστοιχούν τελικά σε λευκό θόρυβο μηδενικής μέσης τιμής. Έτσι  $\mu_S = 0$ . Κάτι αντίστοιχο μπορούμε να υποθέσουμε και για την κινούμενη περιοχή, με την παρατήρηση ότι γενικά τα κινούμενα αντικείμενα αποτελούν ομοιόμορφες περιοχές και ότι οι σημαντικές διαφορές φωτεινότητας που προκύπτουν στο φόντο λόγω της κίνησης (δηλαδή στην περιοχή που αποτελεί φόντο για την πρώτη εικόνα και κινούμενο μέρος για τη δεύτερη) είναι εξίσου έντονες. Άρα έχουμε και  $\mu_M = 0$ .

Η βέλτιστη εκτίμηση των παραμέτρων  $\{(P_L, \lambda_L) : L \in \{S, M\}\}$  γίνεται με τη μέθοδο της κατάβασης κλίσης για τη μεγιστοποίηση της πιθανοφάνειας του μοντέλου. Συγκεκριμένα, το βέλτιστο διάνυσμα  $(\hat{P}, \hat{\Theta}) = (\hat{P}_M, \hat{P}_S, \hat{\sigma}_M, \hat{\sigma}_S)$  για τη Gaussian και  $(\hat{P}, \hat{\Theta}) = (\hat{P}_M, \hat{P}_S, \hat{\lambda}_M, \hat{\lambda}_S)$  για τη Laplacian θα μεγιστοποιεί την από κοινού κατανομή πιθανότητας

$$p(\mathcal{R}|P, \Theta) = \prod_{d \in D} (d|P, \Theta)$$

όπου  $\mathcal{R}$  είναι η διαμέριση της  $D$  σε κινούμενο και στατικό τμήμα.

Μετά τη στατιστική μοντελοποίηση του κινούμενου και του στατικού τμήματος, ακολουθεί η κατάστρωση ενός συναρτησιακού ενέργειας, η ελαχιστοποίηση του οποίου θα οδηγήσει στην ανίχνευση των συνόρων των κινούμενων αντικειμένων. Έστω  $s$  μια θέση στο πλέγμα ορισμού της εικόνας, και  $N$  μια διαμέριση της γειτονιάς αυτού σε δύο τοπικές υποπεριοχές (αριστερή και δεξιά)  $[N_R(s), N_L(s)]$ . Επίσης, έστω  $D(N(s))$  τα δεδομένα που προκύπτουν από την εικόνα διαφοράς  $D$  για την εν λόγω γειτονιά. Αν  $p(B|D(N(s)))$  είναι η πιθανότητα το σημείο  $s$  να βρίσκεται στα όρια ( $B$ : Boundary) μεταξύ ενός κινούμενου αντικειμένου και του στατικού φόντου, τότε σύμφωνα με τον κανόνα του Bayes θα έχουμε :

$$p(B|D(N(s))) = \frac{p_B(D(N(s)))}{p(D(N(s)))} p(B)$$

$$\begin{aligned}
 &= \frac{p_B(D(N(s)))}{p(D(N(s))|B \cup B^c)} p(B) \\
 &= \frac{p_B(D(N(s)))}{p_B(D(N(s))) + p_{B^c}(D(N(s)))} \quad (4.13)
 \end{aligned}$$

όπου  $p_B(d)$  είναι η υπό συνθήκη πιθανότητα για την τιμή  $d$ , όταν βρισκόμαστε εντός του συνόρου,  $p_{B^c}(d)$  είναι η υπό συνθήκη πιθανότητα για την τιμή  $d$ , όταν βρισκόμαστε εκτός του συνόρου, και  $p(B)$  είναι η a-priori πιθανότητα να βρισκόμαστε σε σύνορο, που είναι σταθερά για το συγκεκριμένο πρόβλημα και μπορεί να παραληφθεί. Η παραπάνω πιθανότητα μπορεί να υπολογιστεί βάσει των εξής δύο παρατηρήσεων:

1. Αν το  $s$  είναι στοιχείο του συνόρου, τότε υπάρχει μια διαμέριση  $[N_R(s), N_L(s)]$  της γειτονιάς  $N$ , όπου η πιο πιθανή υπόθεση για την  $N_R(s)$  είναι αυτή της στατικότητας ενώ για την  $N_L(s)$  αυτή της κίνησης, ή το αντίστροφο, αφού πρέπει στην περιοχή του σημείου  $s$  να υπάρχει τόσο η στατική όσο και η κινούμενη συνιστώσα:

$$[(N_R \in S \cap N_L \in M) \cup (N_R \in M \cap N_L \in S)]$$

2. Αν το  $s$  δεν είναι στοιχείο του συνόρου, τότε, για κάθε διαμέριση της γειτονιάς  $N$  η πιο πιθανή υπόθεση και για την  $N_L$  και για την  $N_S$  είναι είτε αυτή της στατικότητας, είτε αυτή της κίνησης, αφού πρέπει το σημείο  $s$  να βρίσκεται εντός μόνο μίας συνιστώσας:

$$[(N_R \in S \cap N_L \in S) \cup (N_R \in M \cap N_L \in M)]$$

Συνεπώς οι δύο πιθανότητες για το σύνορο και για το υπόλοιπο της εικόνας μπορούν πλέον να υπολογιστούν ως εξής:

$$\begin{aligned}
 p_B(D(N(s))) &= p(D(N(s))|B) \\
 &= p(D(N(s))|[N_R \in S \cap N_L \in M]) + p(D(N(s))|[N_R \in M \cap N_L \in S]) \\
 &= \underbrace{p_S(D(N_L(s)))}_{N_L : \text{στατική}} \underbrace{p_M(D(N_R(s)))}_{N_R \text{ κινείται}} + \underbrace{p_M(D(N_L(s)))}_{N_L \text{ κινείται}} \underbrace{p_S(D(N_R(s)))}_{N_R : \text{στατική}}
 \end{aligned}$$

$$\begin{aligned}
 p_{B^c}(D(N(s))) &= p(D(N(s))|B^c) \\
 &= p(D(N(s))|[N_R \in S \cap N_L \in S]) + p(D(N(s))|[N_R \in M \cap N_L \in M]) \\
 &= \underbrace{p_S(D(N_L(s)))}_{N_L : \text{στατική}} \underbrace{p_S(D(N_R(s)))}_{N_R : \text{στατική}} + \underbrace{p_M(D(N_L(s)))}_{N_L \text{ κινείται}} \underbrace{p_M(D(N_R(s)))}_{N_R \text{ κινείται}}
 \end{aligned}$$

Τελικά το μόνο πρόβλημα που απομένει είναι ο ορισμός της διαμέρισης της γειτονιάς  $N$ . Για το σκοπό αυτό στο [13] εξετάζονται 4 τύποι διαμέρισης, που αντιστοιχούν στη διαίρεση της  $N$  (που θεωρείται ως ένα  $3 \times 3$  παράθυρο γύρω από το  $s$ ) κατά μήκος ευθείας που σχηματίζει γωνία  $\theta$  με τον οριζόντιο άξονα, όπου

$$\theta \in \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$$

και η πιθανότητα το  $s$  να βρίσκεται πάνω στο σύνορο υπολογίζεται από τη σχέση (4.13) λαμβάνοντας το μέγιστο αυτής για τις 4 δυνατές διαμερίσεις:

$$\max \left[ \frac{p_B(D(N(s)), \theta)}{p_B(D(N(s)), \theta) + p_{B^c}(D(N(s)), \theta)} \right]$$

Οι τιμές της πιθανότητας αυτής για κάθε  $s$  τοποθετούνται σε μια νέα εικόνα  $I_D$ , η οποία χρησιμοποιείται για την κατασκευή ενός συναρτησιακού με τρόπο ακριβώς όμοιο όπως και στη μέθοδο των γεωδαιτικών ενεργών περιγραμμάτων:

$$E[(C)(p)] = \int_0^1 g(I_D(\vec{C}(p))) \|\vec{C}_p(p)\| dp \quad (4.14)$$

όπου  $g$  γκαουσιανή συνάρτηση. Με τον τρόπο αυτό το γεωδαιτικό ενεργό περίγραμμα επιτρέπει τον εντοπισμό των κινούμενων περιοχών στην εικόνα.

Ωστόσο συχνά η ανίχνευση της κίνησης δεν ταυτίζεται με τον εντοπισμό του αντικειμένου που κινείται, καθώς, επειδή χρησιμοποιούμε την εικόνα διαφοράς  $D$ , στην κινούμενη περιοχή θα περιλαμβάνεται και μέρος του φόντου που αποκαλύπτεται λόγω της κίνησης του αντικειμένου. Για το λόγο αυτό στο παραπάνω συναρτησιακό προστίθεται ένας ακόμη όρος πανομοιότυπος με αυτόν του αρχικού μοντέλου GAC, που εξασφαλίζει ότι το κινούμενο σύνορο θα συμπίσει με το περίγραμμα του αντικειμένου. Έτσι τελικά το ολικό συναρτησιακό ενέργειας γίνεται:

$$E[(C)(p)] = \int_0^1 \left[ \gamma \underbrace{g(I_D(\vec{C}(p)), \sigma_D)}_{\text{ανίχνευση κίνησης}} + (1 - \gamma) \underbrace{g(\|\nabla I(\vec{C}(p), t)\|, \sigma_T)}_{\text{εντοπισμός αντικειμένου}} \right] \|\vec{C}_p(p)\| dp \quad (4.15)$$

Το συναρτησιακό αυτό ελαχιστοποιείται με χρήση του λογισμού μεταβολών και της μεθόδου των επιπεδοσυνόλων, οπότε η εξέλιξη της επιπεδοσυνάρτησης προκύπτει ότι ακολουθεί τη σχέση:

$$\begin{aligned} \phi_t = & \left\{ \gamma \left[ g(I_D, \sigma_D) \kappa + \nabla g(I_D, \sigma_D) \frac{\nabla \phi}{\|\nabla \phi\|} \right] \right. \\ & \left. + (1 - \gamma) \left[ g(\|\nabla I\|, \sigma_T) \kappa + g(\|\nabla I\|, \sigma_T) \frac{\nabla \phi}{\|\nabla \phi\|} \right] \right\} \quad (4.16) \end{aligned}$$

## 4.5 Εκτίμηση Κίνησης και Αναγνώριση Νοηματικής Γλώσσας

Όπως αναφέραμε και στην αρχή του κεφαλαίου, η κίνηση αποτελεί πολύ βασικό συστατικό της νοηματικής γλώσσας. Στις περισσότερες προσεγγίσεις σχετικά με την αναγνώριση ([22],[46]), η κίνηση αντιπροσωπεύεται από τα διανύσματα ταχύτητας των κέντρων των κινούμενων περιοχών (χέρια και πρόσωπο). Πρακτικά, ωστόσο, τα κέντρα μάζας δε μπορούν να θεωρηθούν ως επαρκής πηγή δυναμικών χαρακτηριστικών, καθώς μπορούν μόνο να δείξουν την κατεύθυνση κίνησης των χεριών του ομιλητή και όχι τυχόν παραμορφώσεις και αλλαγές σχήματος που υφίστανται τα χέρια κατά την κίνηση, και που αποτελούν σημαντικό διακριτικό στοιχείο μεταξύ των νευμάτων. Συνεπώς χρειαζόμαστε και κάποια επιπλέον χαρακτηριστικά.

Μια καλύτερη προσέγγιση, που μπορεί να χρησιμοποιηθεί για την περιγραφή οποιασδήποτε ακολουθίας καρτέ, θα ήταν να χρησιμοποιήσουμε το πεδίο οπτικής ροής για να εξάγουμε ένα σύνολο δυναμικών χαρακτηριστικών, προς χρήση στην εφαρμογή αναγνώρισης. Μια τέτοια προσέγγιση παρουσιάζεται στο [11], όπου το πεδίο οπτικής ροής μετά από κατάλληλη επεξεργασία συγκροτεί (χωρίς τη συμμετοχή κανενός άλλου χαρακτηριστικού) ένα διάνυσμα χαρακτηριστικών το οποίο χρησιμοποιείται για την αναγνώριση ενός συνόλου 33 νευμάτων.

Προκειμένου να επιτευχθεί ανεξαρτησία της εφαρμογής από το χρώμα, χρησιμοποιείται μια ενιαία παραμετρική μορφή για τα διανύσματα κίνησης: η απόσταση κινούμενου block (Moving Block Distance - MBD). Συγκεκριμένα, κάθε καρτέ χωρίζεται σε blocks, που αριθμούνται από αριστερά προς τα δεξιά και από πάνω προς τα κάτω, έτσι ώστε οι  $(x, y)$  συνιστώσες του block να αντιστοιχίζονται σε ένα μονοδιάστατο δείκτη  $p$ . Για κάθε block υπολογίζεται το διάνυσμα μετατόπισης, χρησιμοποιώντας είτε τεχνικές εκτίμησης του πεδίου οπτικής ροής ή block - matching. Δημιουργείται έτσι ένα πεδίο διανυσμάτων, με κάθε block της εικόνας να αντιστοιχίζεται σε ένα διάνυσμα εκτίμησης κίνησης. Από το πεδίο αυτό απομονώνονται τα διανύσματα (και συνεπώς τα αντίστοιχα blocks) που έχουν το μέγιστο μήκος, και θεωρούμε ότι τα διανύσματα αυτά είναι αρκετά για να αντιπροσωπεύσουν όλο το πεδίο διανυσμάτων (στο [11] χρησιμοποιήθηκαν τα 10 ισχυρότερα διανύσματα). Τα επιλεγμένα αυτά διανύσματα αντιστοιχούν, μέσω του αντίστοιχου δείκτη, στα κινούμενα blocks. Οι διαφορές μεταξύ των δεικτών  $p$  για κάθε ζεύγρον διαδοχικών (από αριστερά προς τα δεξιά και από πάνω προς τα κάτω) κινούμενων blocks αποθηκεύονται στο διάνυσμα χαρακτηριστικών για το κάθε καρτέ. Οι διαφορές αυτές καλούνται MBDs.

Στην παραπάνω μέθοδο υπάρχουν βασικά δύο παράμετροι που πρέπει να μελετηθούν κατάλληλα. Η πρώτη σχετίζεται με το μέγεθος του block. Αύξηση της διάστασης του block επιτυγχάνει ταχύτερη εκτέλεση και μεγαλύτερη ανεξαρτησία από το νοηματιστή, αλλά και μείωση της λεπτομέρειας που αποτυπώνεται στα διανύσματα κίνησης. Ένα υπερβολικά μεγάλο μέγεθος block είναι δυνατόν να οδηγήσει σε σημαντική απώλεια της χρήσιμης πληροφορίας. Η δεύτερη παράμετρος είναι η μέθοδος εκτίμησης της κίνησης που χρησιμοποιείται.

Η παραπάνω τεχνική παρουσιάζεται στο Σχ. 4.3, για δύο εικόνες από την ακολουθία «Ταξί Αμβούργου». Εδώ κρατάμε τα 30 πιο σημαντικά blocks.

#### 4.5.1 Πειραματικά Αποτελέσματα και Εξαγωγή Δυναμικών Χαρακτηριστικών Κίνησης

Στην ενότητα αυτή θα χρησιμοποιήσουμε τα όσα αναφέραμε πιο πάνω (τεχνικές εκτίμησης της οπτικής κίνησης και αξιοποίησης αυτής) για την εξαγωγή ενός συνόλου χαρακτηριστικών που να αντιπροσωπεύουν την κίνηση των χεριών του νοηματιστή.

Για την εξαγωγή των χαρακτηριστικών πρέπει προηγουμένων για κάθε ζεύγος frames να εξάγουμε το πεδίο κίνησης. Μπορούμε για το σκοπό αυτό να χρησιμοποιήσουμε είτε τεχνικές block matching, είτε optical flow. Στα Σχ. 4.4 και Σχ. 4.5 παρουσιάζονται τα αποτελέσματα που πήραμε και με τις δύο τεχνικές (block-matching/Diamond - Search και Optical Flow/ Lucas - Kanade) για δύο διαφορετικές ακολουθίες βίντεο.

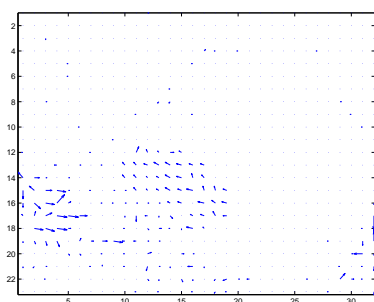
Όπως μπορούμε να δούμε, το block matching αδυνατεί σε πολλές περιπτώσεις, και ειδικά στην πρώτη ακολουθία, να εντοπίσει το σωστό πεδίο κίνησης. Αυτό οφείλεται στο γεγονός ότι το σχήμα του χεριού μεταβάλλεται πολύ έντονα κατά τη διάρκεια της ακολουθίας, ενώ η τεχνική block matching απαιτεί το σχήμα των blocks να διατηρείται. Έτσι η μέθοδος αυτή εντοπίζει κίνηση μόνο στις περιοχές εκείνες όπου η μετατόπιση μπορεί να θεωρηθεί ότι γίνεται κατά blocks, όπως κάποιες περιοχές του φόντου που λόγω της αλλαγής της σκίασης φαίνονται σαν να μετακινούνται ανεπαίσθητα. Προφανώς από τα αποτελέσματα αυτά καταλήγουμε ότι προτιμότερο είναι να χρησιμοποιηθεί ο αλγόριθμος των Lucas και Kanade που εκτιμά την οπτική ροή. Παραθέτουμε ενδεικτικά τα αποτελέσματα για το matching block distance που αντιστοιχεί στην ακολουθία 2 στο Σχ. 4.6.



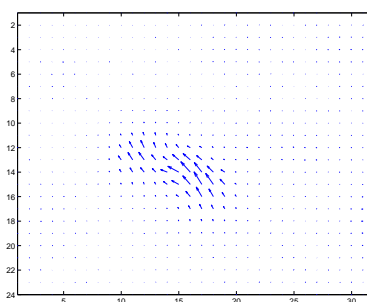
(a) Εικόνα 1



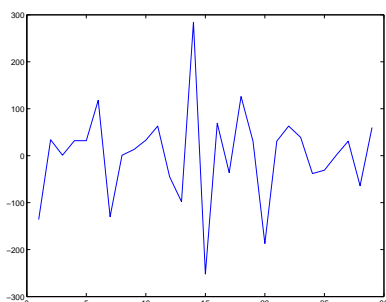
(b) Εικόνα 2



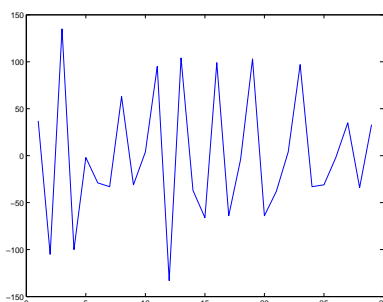
(c) Πεδίο κίνησης ανά blocks με χρήση block matching (Diamond Search)



(d) Πεδίο κίνησης ανά blocks με χρήση optical flow (Lukas - Kanade)



(e) Moving Block distance για το πεδίο κίνησης του Σχ. 4.3(c)



(f) Moving Block distance για το πεδίο κίνησης του Σχ. 4.3(d)

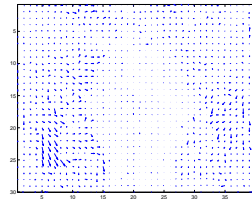
**Εικόνα 4.3:** Δυναμικά χαρακτηριστικά κίνησης για μια ακολουθία δύο εικόνων από την ακολουθία «Ταξί Αμβούργου». Στα Σχ. 4.3(e) και Σχ. 4.3(f) στον οριζόντιο άξονα μεταβάλλονται οι αύξοντες αριθμοί των πιο έντονα κινούμενων blocks ενώ στον κατακόρυφο οι διαφορές των αριθμών αυτών.

## 4.6 Εικόνες Αποτελεσμάτων

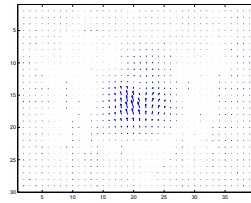
Στις επόμενες σελίδες παρατίθενται τα αποτελέσματα των πειραμάτων του κεφαλαίου 4, με κατάλληλα σχόλια στις λεζάντες όπου χρειάζεται.



(a) Frame 1



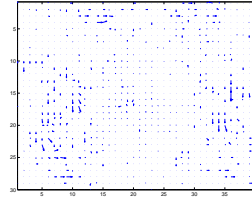
(b) Block Matching



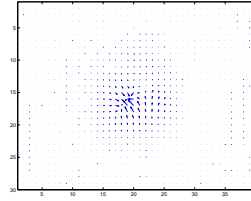
(c) Optical Flow



(d) Frame 2



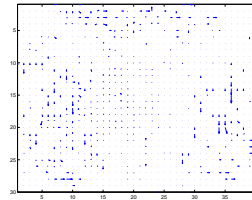
(e) Block Matching



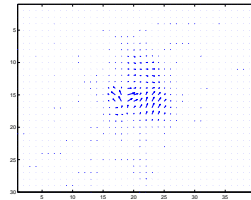
(f) Optical Flow



(g) Frame 3



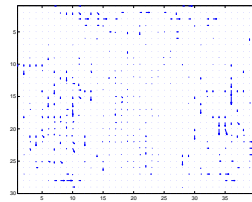
(h) Block Matching



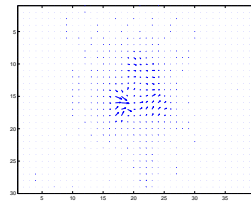
(i) Optical Flow



(j) Frame 4



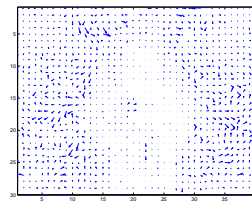
(k) Block Matching



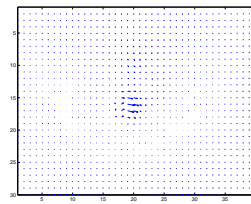
(l) Optical Flow



(m) Frame 5



(n) Block Matching



(o) Optical Flow



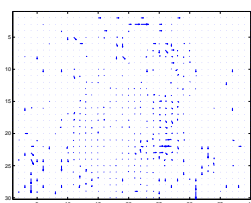
(p) Frame 6

**Εικόνα 4.4:** Optical Flow vs. Block Matching - Ακολουθία νοήματος «πόρτα». Τα πεδία κίνησης αφορούν στο τρέχον frame και το επόμενο.

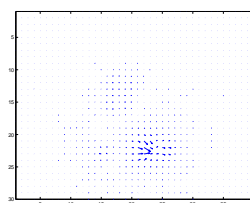




(a) Frame 1



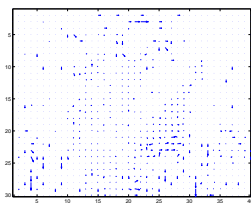
(b) Block Matching



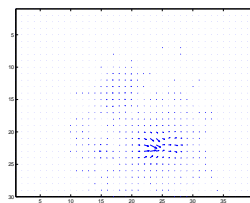
(c) Optical Flow



(d) Frame 2



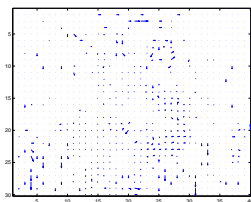
(e) Block Matching



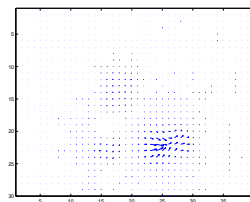
(f) Optical Flow



(g) Frame 3



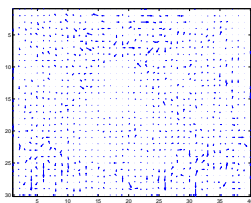
(h) Block Matching



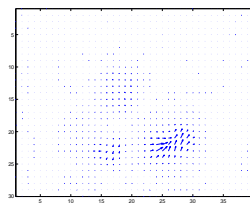
(i) Optical Flow



(j) Frame 4



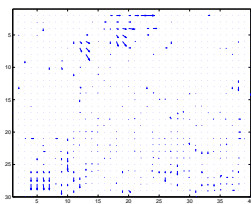
(k) Block Matching



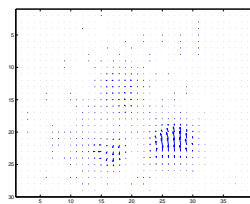
(l) Optical Flow



(m) Frame 5



(n) Block Matching



(o) Optical Flow

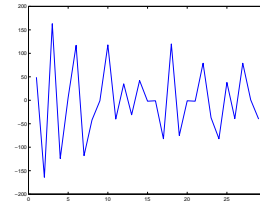


(p) Frame 6

Εικόνα 4.5: Optical Flow vs. Block Matching - Ακολουθία νοήματος «κοιτώ προς τα αριστερά». Τα πεδία κίνησης αφορούν στο τρέχον frame και το επόμενο.



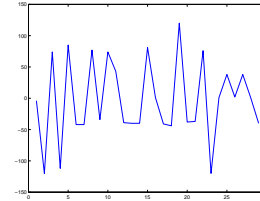
(a) Frame 1



(b) Χαρακ/κά MBD



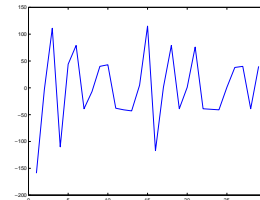
(c) Frame 2



(d) Χαρακ/κά MBD



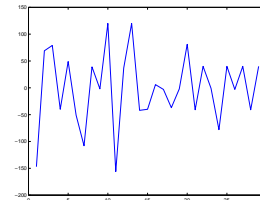
(e) Frame 3



(f) Χαρακ/κά MBD



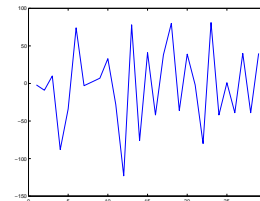
(g) Frame 4



(h) Χαρακ/κά MBD



(i) Frame 5



(j) Χαρακ/κά MBD



(k) Frame 6

**Εικόνα 4.6:** Δυναμικά Χαρακτηριστικά Κίνησης MBD για την ακολουθία νοήματος «κοιτώ προς τα αριστερά». Έχουν απομονωθεί τα 30 πιο έντονα κινούμενα blocks.

---

## ΑΝΤΙΜΕΤΩΠΙΣΗ ΕΠΙΚΑΛΥΨΕΩΝ

---

Το σύστημα που περιγράψαμε στα προηγούμενα κεφάλαια (κατάτμηση εικόνας, απομόνωση χεριών και εξαγωγή στατικών και δυναμικών χαρακτηριστικών) καλύπτει τις περισσότερες περιπτώσεις νευμάτων της νοηματικής γλώσσας. Πρόβλημα εμφανίζεται στις περιπτώσεις όπου εμφανίζονται επικαλύψεις μεταξύ δύο ή περισσότερων μελών του σώματος, όπως για παράδειγμα στην εικόνα του Σχ. 5.1 .

Στις περιπτώσεις αυτές η θέση των δύο χεριών προσδιορίζεται λανθασμένα (όπως φαίνεται και στο Σχ. 5.1 αντιπροσωπεύεται από τη θέση του κέντρου μάζας ολόκληρης της περιοχής της επικάλυψης και λαμβάνεται παίρνοντας το μέσο όρο των θέσεων των εικονοστοιχείων που βρίσκονται εντός αυτής της περιοχής), ενώ δεν είναι δυνατόν και να εξαχθούν στατικά χαρακτηριστικά σχήματος, συνεπώς δεν έχουμε ούτε πληροφορία για τη χειρομορφή που χρησιμοποιείται. Απαιτείται συνεπώς ειδική μέριμνα για την εξάλειψη αυτού του φαινομένου. Στην παρούσα διπλωματική επιχειρήσαμε προσεγγίσεις για την αντιμετώπιση των επικαλύψεων (occlusions), οι οποίες είναι:

1. Χρήση σύμπτωσης προτύπων (template matching)
2. Χρήση σύμπτωσης προτύπων με κινητικό περιορισμό στην περιοχή αναζήτησης
3. Αξιοποίηση της πληροφορίας πρότερου σχήματος
4. Εξαγωγή «πεδίων δυνάμεων» και «πηγαδιών δυναμικού» από την εικόνα



(a) Επικάλυψη χεριού-κεφαλιού

(b) Επικάλυψη χεριών



(c) Ολική επικάλυψη

Εικόνα 5.1: Παραδείγματα επικαλύψεων

## 5.1 Σύμπτωση προτύπων

Πρόκειται για μια τεχνική για την εύρεση των αντιστοιχιών μεταξύ δύο εικόνων, συγκεκριμένα για την εύρεση κάποιας περιοχής σε μια εικόνα που «ταυτίζεται» με κάποια άλλη πρότυπη εικόνα. Για την εξακρίβωση της ταύτισης, η πρότυπη εικόνα «κεντράρεται» πάνω σε όλα τα pixels της εικόνας διαδοχικά και σε κάθε θέση λαμβάνεται η τιμή μιας μετρικής που δίνει ένα μέτρο της διαφορετικότητας μεταξύ της πρότυπης εικόνας και την αντίστοιχης περιοχής στην εικόνα αναφοράς. Ως μετρικές μπορούν να χρησιμοποιηθούν διάφορες συναρτήσεις, όπως:

$$\int_G (f - g)^2, \quad \int_G |f - g|, \quad \max_G |f - g|$$

όπου  $f$  είναι η εικόνα αναφοράς, όπου ζητείται να βρεθεί η θέση του προτύπου  $g$ .

Η θέση ελαχιστοποίησης αυτών των μετρικών λαμβάνεται ως η θέση καλύτερης

σύμπτωσης. Στην περίπτωση της Ευκλείδιας απόστασης

$$\int_G (f - g)^2 = \int_G f^2 + \int_G g^2 - 2 \int_G fg$$

η ελαχιστοποίηση της μετρικής ανάγεται σε μεγιστοποίηση της ποσότητας  $\int_G fg$ , καθώς οι υπόλοιποι 2 όροι είναι σταθερές. Η ποσότητα αυτή αντιστοιχεί στη συσχέτιση μεταξύ της πρότυπης εικόνας και της αντίστοιχης περιοχής της εικόνας αναφοράς. Στην περίπτωση περίπτωση διακριτής εικόνας η συσχέτιση παίρνει τη μορφή

$$R(i, j) = \sum_m \sum_n f(i + m, j + n) \cdot g(m, n) \quad (5.1)$$

Στην περίπτωση που η ενέργεια της εικόνας αναφοράς δεν παραμένει πρακτικά σταθερή εντός της εξεταζόμενης περιοχής, χρησιμοποιούμε την κανονικοποιημένη συσχέτιση

$$R(i, j) = \frac{\sum_m \sum_n f(i + m, j + n) \cdot g(m, n)}{\sqrt{\sum_m \sum_n f(i + m, j + n)^2}} \quad (5.2)$$

Αυτή τη μορφή θα χρησιμοποιήσουμε και εδώ, με πρότυπο την περιοχή του χεριού στο τελευταίο frame πριν την έναρξη της επικάλυψης. Η περιοχή αυτή προσδιορίζεται από το μικρότερο περιγεγραμμένο ορθογώνιο της περιοχής του χεριού που προέκυψε από την κατάτμηση.

Προτού ξεκινήσει η διαδικασία του template-matching, προσδιορίζονται καταρχήν οι περιοχές που επικαλύπτονται. Αυτό γίνεται για να μην γίνει το TM για όλη την εικόνα, πράγμα που θα οδηγούσε σε περαιτέρω καθυστέρηση της όλης διαδικασίας, και ενδεχομένως σε εσφαλμένη ταύτιση. Οι περιοχές που επικαλύπτονται θα είναι εκείνες των οποίων τα κέντρα μάζας στην πρώτη εικόνα (χωρίς επικάλυψη) θα απέχουν περισσότερο από τις αντίστοιχες θέσεις τους στην επόμενη εικόνα (με την επικάλυψη). Αυτό ισχύει στις περισσότερες των περιπτώσεων, εφ' όσον η κίνηση μεταξύ δύο διαδοχικών καρέ είναι σχετικά μικρή, οπότε τα κέντρα μάζας δεν αναμένουμε να μετακινούνται πολύ. Όταν εμφανίζεται επικάλυψη, το κέντρο μάζας του συσσωματώματος θα βρίσκεται σε μια ενδιάμεση θέση μεταξύ των κέντρων μάζας των δύο επικαλυπτόμενων περιοχών και συνεπώς θα εμφανιστεί ασυνήθιστα μεγάλη μετατόπιση στα κέντρα μάζας. Με βάση το κριτήριο αυτό μπορούμε να βρούμε ποιες περιοχές της πρώτης εικόνας είναι εκείνες που επικαλύπτονται στη δεύτερη, και σε ποιο τμήμα της δεύτερης εικόνας αντιστοιχεί το συσσωμάτωμα αυτών. Το τμήμα

αυτό της δεύτερης εικόνας αποτελεί την περιοχή αναζήτησης (συνάρτηση  $f$  στις παραπάνω εξισώσεις), όπου θα γίνει η αναζήτηση για το πρότυπο. Το σημείο στο επόμενο frame όπου εμφανίζεται η καλύτερη σύμπτωση λαμβάνεται ως η νέα θέση του χεριού. Η διαδικασία απεικονίζεται στο παρακάτω Σχ. 5.5. Στη μεσαία γραμμή φαίνονται οι τεμπλέτες και η περιοχή αναζήτησης.

Για καλύτερα αποτελέσματα υλοποιήθηκε και η αντίστροφη διαδικασία, όπου ως πρότυπο χρησιμοποιείται η πρώτη μη επικαλυπτόμενη όψη των χεριών στο πρώτο frame αμέσως μετά το πέρας της επικάλυψης. Οι θέσεις που προκύπτουν από τα δύο TMs αθροίζονται με κατάλληλα βάρη, που προκύπτουν από το βαθμό σύμπτωσης, για να δώσουν την τελική εκτίμηση της θέσης. Ένα ενδεικτικό αποτέλεσμα βλέπουμε στο Σχ. 5.6. Παρατηρούμε ότι τα κέντρα των χεριών ανιχνεύονται με επιτυχία ακόμα και στην περίπτωση που έχουμε ολική επικάλυψη.

Σημειώνουμε ότι, επειδή έχουμε να κάνουμε με RGB εικόνες, το TM πραγματοποιείται ξεχωριστά για κάθε χρωματική συνιστώσα των εικόνων και κατόπιν το σημείο σύμπτωσης λαμβάνεται ως η θέση μεγιστοποίησης της ελάχιστης από τις συσχετίσεις προτύπου-εικόνας των 3 συνιστωσών. Η διαδικασία αυτή προτιμήθηκε της μετατροπής των εικόνων στην κλίμακα του γκρι, διότι έδινε καλύτερα αποτελέσματα στις περισσότερες περιπτώσεις. Επίσης να τονίσουμε ότι, εφ' όσον η διαδικασία της σύμπτωσης προτύπων μπορεί να συμβάλλει στον ακριβή περιορισμό μόνο της θέσης των χεριών, τα χαρακτηριστικά σχήματα είτε «κληρονομούνται» αυτούσια από το τελευταίο frame χωρίς επικάλυψη, είτε παρεμβάλλονται γραμμικά μεταξύ των δύο διανυσμάτων χαρακτηριστικών που προκύπτουν από το τελευταίο frame πριν την επικάλυψη και το πρώτου μετά από αυτήν.

## 5.2 Σύμπτωση προτύπων σε συνδυασμό με πληροφορία κίνησης

Όπως προαναφέραμε, το template matching αποδίδει στον προσδιορισμό μόνο της θέσης των χεριών του ομιλητή σε περιπτώσεις επικάλυψων. Κάποιες φορές, ωστόσο, λόγω της ομοιογένειας του χρώματος και της υφής του δέρματος, είναι πιθανό το σημείο μεγιστοποίησης του ποσοστού ταύτισης να μην αντιστοιχεί στην πραγματική θέση του χεριού, οπότε η μέθοδος TM αποτυγχάνει. Ένα παράδειγμα φαίνεται στο Σχ. 5.7. Το σημείο καλύτερης σύμπτωσης ανιχνεύεται εσφαλμένα εντός της περιοχής του κεφαλιού.

Όπως βλέπουμε, στις περιπτώσεις αυτές είναι απαραίτητο να περιοριστεί η περιοχή της εικόνας στην οποία αναζητούμε το σημείο καλύτερης ταύτισης για την τεμπλέτα. Μια ιδέα θα ήταν να χρησιμοποιήσουμε πληροφορία σχετικά με την κίνηση των αντικειμένων στην εικόνα για να αποκλείσουμε την περιοχή του κεφαλιού από το template-matching. Αυτό μπορεί να γίνει με δύο τρόπους, καθέννας από τους οποίους θα περιγραφεί ξεχωριστά.

### 5.2.1 Απλή πρόβλεψη κίνησης

Ένας τρόπος να ενσωματώσουμε την πληροφορία της κίνησης είναι να προσπαθήσουμε να προβλέψουμε πού θα βρίσκονται τα χέρια του ομιλητή στο επόμενο καρέ. Εφ' όσον έχουμε στη διάθεσή μας μια εκτίμηση για τη θέση του χεριού, μπορούμε να ορίσουμε μια περιοχή αναζήτησης γύρω από τη θέση αυτή, η οποία θα είναι σημαντικά πιο περιορισμένη σε σχέση με την ολική περιοχή αναζήτησης.

Η προσέγγιση που θα ακολουθήσουμε εδώ απαιτεί να έχουμε στη διάθεσή μας τουλάχιστον 3 καρέ πριν από το τρέχον για να πραγματοποιήσουμε την ακριβέστερη δυνατή πρόβλεψη. Επιπλέον, εκμεταλλεύεται την ιδιότητα των βίντεο να διατηρούν κάποια χρονική συνέχεια, που σημαίνει ότι τα κινούμενα αντικείμενα διατηρούν περίπου την ίδια κατεύθυνση και ταχύτητα κίνησης για έναν αριθμό από συνεχόμενα καρέ, με την αλλαγή κατεύθυνσης και ταχύτητας να σηματοδοτούνται από μεταβολές στην ταχύτητα και την επιτάχυνση της κίνησης αντίστοιχα. Πράγματι η ιδιότητα αυτή επιβεβαιώνεται και πειραματικά, τουλάχιστον για το χρονικό διάστημα που αφορά στον «ενεργό» χρόνο του νεύματος (δηλαδή όχι στο χρόνο που χρειάζεται ο ομιλητής να προετοιμαστεί τοποθετώντας τα χέρια του στις κατάλληλες θέσεις προκειμένου να εκτελέσει ένα νεύμα).

Έτσι, υποθέτοντας ότι πράγματι διαθέτουμε πληροφορία για το «παρελθόν» ενός καρέ, ο νόμος που εφαρμόζουμε για την πρόβλεψη είναι:

$$\vec{p}(i) = \vec{p}(i-1) + \frac{1}{2} [\vec{v}(i) + \vec{v}(i-1) + \vec{a}(i)] \quad (5.3)$$

όπου  $\vec{p}(i) = [x(i) \ y(i)]^T$  είναι η θέση του κέντρου του χεριού στο frame  $i$ ,  $\vec{v}(i) = \vec{p}(i) - \vec{p}(i-1)$  και  $\vec{a}(i) = \vec{v}(i) - \vec{v}(i-1)$  είναι η ταχύτητα και επιτάχυνση αυτού στο frame  $i$ . Η περιοχή όπου προβλέπεται ότι βρίσκεται το χέρι θεωρείται σχήματος ορθογωνίου, μεγέθους λίγο μεγαλύτερου από το ελάχιστο περιγεγραμμένο ορθογώνιο του χεριού στο καρέ πριν την επικάλυψη, περαιτέρω μεγενθυμένη ίσως στην περίπτωση που η επιτάχυνση είναι πολύ μεγάλη, προκειμένου να αποφευχθούν τα σφάλματα πρό-

βλεψης. Προφανώς, εφ' όσον διαθέτουμε την πρόβλεψη, η περιοχή αναζήτησης για το TM θα είναι η τομή της περιοχής που προέκυψε από την πρόβλεψη με την αρχική περιοχή αναζήτησης, καθώς δεν έχει νόημα να ψάξουμε αλλού. Τα αποτελέσματα για τα frames του νεύματος «Βασιλιάς» με την επικάλυψη παρουσιάζονται στο Σχ. 5.8, ενώ μια πιο δύσκολη περίπτωση, του νεύματος «Βιβλίο» στο Σχ. 5.9.

Βλέπουμε ότι ακόμα και αν η πρόβλεψη του κεντρικού σημείου δεν είναι απόλυτα ακριβής, η ευρύτερη περιοχή προβλέπεται σωστά, πράγμα που θα βοηθήσει πολύ στην ορθή εύρεση της πραγματικής θέσης του χεριού από το TM.

### 5.2.2 Απομόνωση της μη κινούμενης περιοχής

Εδώ, αντί να προσπαθήσουμε να προβλέψουμε τη θέση του χεριού με βάση την κίνηση του κέντρου μάζας, θα προσπαθήσουμε να εντοπίσουμε στην εικόνα τις μη κινούμενες περιοχές, ώστε να τις αποκλείσουμε από το template matching. Η διαδικασία που θα ακολουθήσουμε είναι πολύ παρόμοια με αυτήν που περιγράψαμε στην ενότητα 4.4, στην οποία η στατιστική μοντελοποίηση της κινούμενης περιοχής καθοδηγούσε μια γεωδαιτική ενεργή καμπύλη η οποία κατέληγε να συμπίσει με το περίγραμμα της κινούμενης περιοχής της εικόνας.

Στην ενότητα 4.4, η στατιστική πληροφορία για την κίνηση προερχόταν από μια στατιστική μοντελοποίηση της εικόνας που προκύπτει από τη διαφορά του τρέχοντος καρέ και του προηγούμενου. Ωστόσο κάτι τέτοιο δεν ενδείκνυται για την περίπτωση του κινούμενου νοηματιστή, καθώς οι κινούμενες περιοχές (χέρια και ενδεχομένως πρόσωπο) έχουν όλες το ίδιο ομοιόμορφο χρώμα, και πρακτικά την ίδια υφή, με αποτέλεσμα στα σημεία της επικάλυψης η διαφορά να προκύπτει σχεδόν μηδενική, με αποτέλεσμα τα σημεία εκείνα να θεωρηθούν ακίνητα, και να απομονωθεί μόνο το τμήμα του χεριού που δεν επικαλύπτεται. Το ίδιο φαινόμενο εμφανίζεται και στα τμήματα εκείνα της εικόνας που τόσο στο τρέχον καρέ όσο και στο προηγούμενο αντιστοιχούσαν σε περιοχή δέρματος. Το πρόβλημα αυτό φαίνεται στο Σχ. 5.2.

Το πρόβλημα αυτό επιλύεται αν χρησιμοποιηθεί το πεδίο οπτικής ροής. Ας θυμηθούμε ότι το πλάτος του πεδίου ροής δίνει την απαιτούμενη πληροφορία για το κατά πόσον κινείται ένα δεδομένο εικονοστοιχείο στην εικόνα. Κανονικοποιώντας το πλάτος (διαιρώντας με τη μέγιστη τιμή του) μπορούμε να πάρουμε κατά κάποιο τρόπο την πιθανότητα ένα εικονοστοιχείο να κινείται, ενώ προφανώς η αντίθετη εικόνα δίνει την πιθανότητα της στατικότητας για το δοθέν pixel. Για παράδειγμα βλ. Σχ. 5.11.

Με βάση τα προηγούμενα, αν θεωρήσουμε το κανονικοποιημένο πλάτος της οπτι-





(a) Frame 9

(b) Frame 10



(c) Επεξεργασμένη εικόνα διαφοράς

**Εικόνα 5.2:** Εικόνα διαφοράς μεταξύ δύο διαδοχικών frames της ακολουθίας «Πάνω». Είναι εμφανής ο μηδενισμός στην τομή των περιοχών που καταλαμβάνει το δεξί χέρι στις δύο εικόνες.

κής ροής ως την πιθανότητα  $P_{mov}(\vec{x})$  και την αντίθετη αυτής ως την πιθανότητα  $P_{notmov}(\vec{x})$ , τότε κατασκευάζουμε την ακόλουθη region-based δύναμη για το μοντέλο GAC:

$$I_R(\vec{x}) = \log \left[ \frac{P_{mov}(\vec{x})}{P_{notmov}(\vec{x})} \right] \quad (5.4)$$

και πραγματοποιούμε κατάτμηση με το μοντέλο GAC. Η επιπεδοσυνάρτηση  $u$  θα εξελίσσεται με βάση τη σχέση:

$$\frac{\partial u}{\partial t} = \left[ c_{gac} \operatorname{div} \left( g(I) \frac{\nabla u}{\|\nabla u\|} \right) + c_{motion} I_R \right] \|\nabla u\| \quad (5.5)$$

Βλέπουμε ότι το μη κινούμενο μέρος της εικόνας (που περικλείεται από τη λευκή καμπύλη) εντοπίζεται σωστά σε κάθε περίπτωση. Βέβαια όπως προκύπτει και από τα σχήματα, επειδή ο αλγόριθμος εκτίμησης της οπτικής ροής είναι πολύ ευαίσθητος

και εντοπίζει και μικρές ακόμα κινήσεις, ο εντοπισμός του κινούμενου μέρους (που είναι και αυτή που μας ενδιαφέρει) δεν είναι ιδιαίτερα ακριβής, και το αποτέλεσμα της κατάτμησης δε μπορεί να χρησιμοποιηθεί αυτούσιο για την απομόνωση του κινούμενου χεριού, αλλά πρέπει να πραγματοποιηθεί και πάλι template matching (ή κάποια από τις μεθόδους που θα αναφέρουμε πιο κάτω). Ωστόσο η περιοχή αναζήτησης έχει πλέον περιοριστεί σημαντικά, ελαχιστοποιώντας τα περιθώρια λάθους του TM. Να σημειώσουμε εδώ ότι πρέπει να ληφθεί κάποια ειδική μέριμνα για τις περιπτώσεις όπου η κίνηση στην περιοχή ενδιαφέροντος είναι πρακτικά μηδαμινή, όπως στα Σχ. 5.11(j), Σχ. 5.11(o), διότι αν χρησιμοποιήσουμε απευθείας την κατατετμημένη εικόνα για να το template matching, η περιοχή αναζήτησης θα έχει εξαλειφθεί τελείως. Η χρήση κάποιου κατωφλίου για τη μέγιστη κινητικότητα, έτσι ώστε αν το μεγαλύτερο ποσοστό της περιοχής αναζήτησης ταξινομηθεί ως μη κινούμενο, το χέρι να θεωρείται στατικό, λύνει και αυτό το πρόβλημα.

### 5.3 Χρήση της πληροφορίας πρότερου σχήματος

Μέχρι στιγμής οι λύσεις που δώσαμε στο πρόβλημα των επικαλύψεων βασιζόνταν στη σύμπτωση προτύπων, και επικεντρώνονταν στη βελτίωση των αποτελεσμάτων αυτής. Όπως όμως αναφέραμε ήδη, οι μέθοδοι αυτές δεν έχουν τη δυνατότητα να εξάγουν την ακριβή πληροφορία του σχήματος της χειρομορφής, και απλά πραγματοποιούν ταίριασμα (matching). Πράγματι αυτή είναι η πιο συχνά εφαρμοζόμενη πρακτική, καθώς στην περίπτωση που αναζητηθεί η επίλυση του ζητήματος των occlusions με ταυτόχρονη εξαγωγή του ακριβούς σχήματος, το πρόβλημα γίνεται σημαντικά πιο δύσκολο. Μια τέτοια προσέγγιση είναι η ενσωμάτωση στο σύστημα της πληροφορίας του σχήματος της χειρομορφής που αναζητείται. Αυτό θα γίνει με χρήση των όσων αναφέρθηκαν στην ενότητα 2.2.4.

Η ιδέα είναι αρκετά απλή: Θεωρούμε ότι έχουμε διαθέσιμο ένα τουλάχιστον καρέ πριν την επικάλυψη, όπου το ακριβές σχήματα του χεριού που θα επικαλυφθεί στο επόμενο καρέ είναι ακριβώς καθορισμένο. Στο επόμενο καρέ, όπου εμφανίζεται η επικάλυψη, αναζητούμε το σχήμα αυτό στην περιοχή αναζήτησης, με τη δυνατότητα η καμπύλη που θα προκύψει να παρουσιάζει διαφορές σε σχέση με το πρότυπο. Χρησιμοποιούμε προς τούτο και πάλι τη μέθοδο GAC, και συγκεκριμένα τη σχέση 2.39. Το πρότυπο σχήμα παριστάνεται μέσω του μετασχηματισμού απόστασης αυτού, και σε κάθε βήμα του αλγορίθμου αναζητείται ο βέλτιστος μετασχηματισμός μετατόπισης, περιστροφής και αλλαγής κλίμακας που ελαχιστοποιεί τη διαφορά μεταξύ του

πρότυπου μετασχηματισμού απόστασης και της τρέχουσας επιπεδοσυνάρτησης, υπολογισμένη πάνω στο εξελισσόμενο σύνορο.

Για να ελέγξουμε την αποτελεσματικότητα της παραπάνω ιδέας, απομονώσαμε τρεις διαφορετικές χειρομορφές, από δύο διαφορετικούς νοηματιστές, και επιχειρήσαμε να τις χρησιμοποιήσουμε για να αντιμετωπίσουμε τις επικαλύψεις σε έναν αριθμό από καρέ από διάφορα νέυματα. Τα αποτελέσματα παρουσιάζονται στα Σχ. 5.12 - Σχ. 5.15, με το σχήμα που απομονώνεται να παρουσιάζεται με λευκό χρώμα.

Βλέπουμε ότι γενικά η μέθοδος επιτυγχάνει σε μεγάλο βαθμό. Η ίδια χειρομορφή ανιχνεύεται με επιτυχία σε διαδοχικά καρέ του ίδιου νοηματιστή, ενώ όπως βλ. στο Σχ. 5.12 εξίσου καλά αποτελέσματα παίρνουμε ακόμα και αν το πρότυπο σχήμα έχει προκύψει από άλλο νοηματιστή. Μια καλύτερη αριθμητική υλοποίηση των GACs ενδεχομένως να βελτιώσει ακόμα περισσότερο τα αποτελέσματα. Ωστόσο πρέπει να παρατηρήσουμε ότι λόγω του γεγονότος ότι ο αλγόριθμος βελτιστοποίησης που αναλαμβάνει την εύρεση του καλύτερου γεωμετρικού μετασχηματισμού αναζητά μόνον τοπικά ελάχιστα, υπάρχει σημαντική εξάρτηση από το αρχικό σημείο που τοποθετείται το πρότυπο σχήμα στην εικόνα. Ως μια καλή προσέγγιση για τη θέση στην οποία αναζητούμε το σχήμα μπορούμε να χρησιμοποιήσουμε το αποτέλεσμα είτε του βελτιωμένου template matching που περιγράψαμε στις προηγούμενες ενότητες, είτε του απλού προβλέπτη κίνησης. Μια μελλοντική βελτίωση θα ήταν να αναζητήσουμε περιγραφές του πρότερου σχήματος που να ελαχιστοποιούν αυτήν την εξάρτηση, και να καθιστούν πιο εύρωστη τη διαδικασία εντοπισμού του σχήματος.

## 5.4 «Πεδία δυνάμεων» και «Πηγάδια δυναμικού»

Η έννοια του πεδίου δυνάμεων για μια εικόνα εισήχθη από τους Hurley, Nixon και Carter στο [21], για την εξαγωγή χαρακτηριστικών από εικόνες ανθρώπινων αυτιών, προς χρήση σε εφαρμογή βιομετρικής αναγνώρισης. Πρόκειται για μια προσέγγιση που αποτελείται από δύο στάδια: μετασχηματισμό της εικόνας στο πεδίο δυνάμεων, και εξαγωγή των καναλιών και πηγαδιών δυναμικού. Κατά το πρώτο στάδιο, η εικόνα μετασχηματίζεται σύμφωνα με ένα γραμμικό μετασχηματισμό στο λεγόμενο «πεδίο δυνάμεων» (force field), υποθέτοντας ότι το κάθε εικονοστοιχείο ασκεί μια ιστροπική δύναμη σε όλα τα υπόλοιπα εικονοστοιχεία, ανάλογα με την ένταση της εικόνας σε αυτό το σημείο. Το φυσικό ανάλογο της διαδικασίας αυτής είναι η εύρεση των αλληλεπιδράσεων μεταξύ ενός συνόλου σημειακών φορτίων, τοποθετημένων πάνω στο πλέγμα ορισμού της εικόνας, και μεγέθους ανάλογου προς τη φωτεινότητα της εικό-

νας στο κάθε σημείο. Κατά το δεύτερο στάδιο, αξιοποιείται η κατευθυντικότητα του πεδίου δυνάμεων για να εντοπιστεί ένας μικρός αριθμός από πηγάδια δυναμικού, που αντιστοιχούν στα ελάχιστα της δυναμικής ενέργειας της εικόνας, όπως αυτή προκύπτει από το πεδίο δυνάμεων. Το στάδιο αυτό είναι αντίστοιχο με την παρακολούθηση των δυναμικών γραμμών του ηλεκτρικού πεδίου που παράγεται από τα σημειακά φορτία που αναφέραμε προηγουμένως, και την εξαγωγή των σημείων ισορροπίας.

#### 5.4.1 Κατασκευή του πεδίου δυνάμεων

Η εικόνα μετασχηματίζεται θεωρώντας ότι αποτελείται από ένα πλέγμα φορτίων, που λειτουργούν ως πηγές έντασης. Κάθε εικονοστοιχείο στη θέση  $\vec{r}_i$  θεωρείται ότι παράγει μια σφαιρικά συμμετρική δύναμη σε όλα τα υπόλοιπα εικονοστοιχεία στις θέσεις  $\vec{r}$  η οποία δίνεται από τη σχέση:

$$\vec{F}_i(\vec{r}) = I(\vec{r}_i) \frac{\vec{r}_i - \vec{r}}{|\vec{r}_i - \vec{r}|} \quad (5.6)$$

Το συνολικό πεδίο δυνάμεων σε κάθε σημείο υπολογίζεται από τη δύναμη που θα ασκούσαν σε ένα μοναδιαίο φορτίο, υπό την επίδραση όλων των υπολοίπων, αν τοποθετούνταν στην αντίστοιχη θέση στην εικόνα. Η δύναμη αυτή προκύπτει από το άθροισμα όλων των επιμέρους δυνάμεων:

$$\vec{F} = \sum_i \vec{F}_i(\vec{r}) = \sum_i I(\vec{r}_i) \frac{\vec{r}_i - \vec{r}}{|\vec{r}_i - \vec{r}|} \quad (5.7)$$

Η γεωμετρία εξαγωγής του πεδίου δυνάμεων παρουσιάζεται στο Σχ. 5.3. Ταυτόχρονα με το πεδίο δυνάμεων μπορεί να υπολογιστεί και η τιμή της δυναμικής ενέργειας σε κάθε σημείο  $\vec{r}$  που προκύπτει λόγω της ύπαρξης φορτίου μεγέθους  $I\vec{r}_i$  στη θέση  $\vec{r}_i$ :

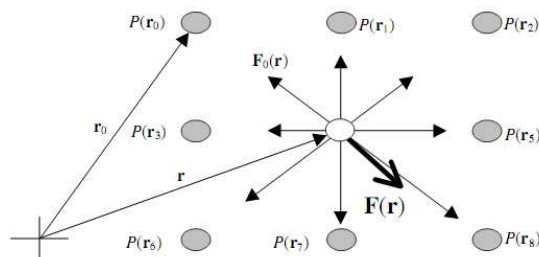
$$\vec{E}_i(\vec{r}) = \frac{I(\vec{r}_i)}{|\vec{r}_i - \vec{r}|} \quad (5.8)$$

καθώς και η συνολική συνάρτηση δυναμικής ενέργειας:

$$\vec{E} = \sum_i \vec{E}_i(\vec{r}) = \sum_i \frac{I(\vec{r}_i)}{|\vec{r}_i - \vec{r}|} \quad (5.9)$$

Το πεδίο δυνάμεων και η συνάρτηση δυναμικού σχετίζονται, και συγκεκριμένα το πεδίο δυνάμεων σε κάθε σημείο είναι η κλίση της συνάρτησης δυναμικού στο ίδιο σημείο:

$$\vec{F}(\vec{r}) = -grad(E(\vec{r})) = -\nabla E(\vec{r})$$

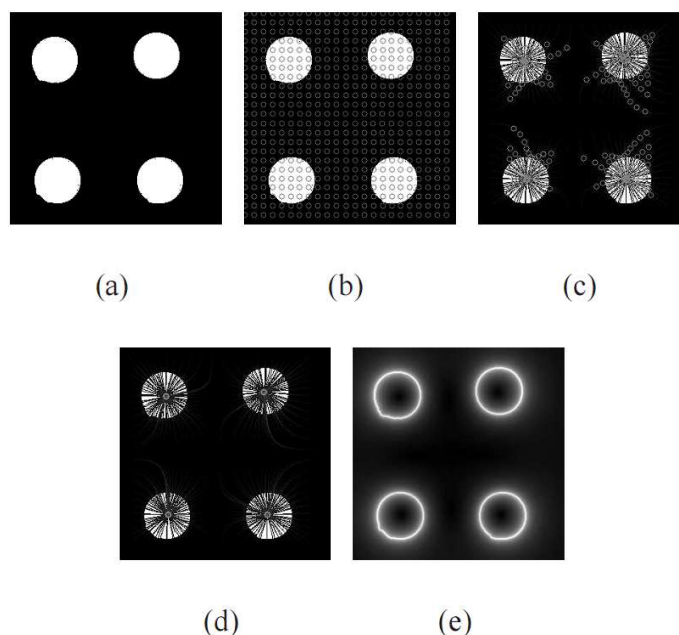


Εικόνα 5.3: Εξαγωγή του πεδίου δυνάμεων. Από [21].

### 5.4.2 Εξαγωγή των καναλιών και πηγαδιών δυναμικού

Θεωρώντας υποθετικά μοναδιαία φορτία να τοποθετούνται σε διάφορα σημεία ελέγχου ομοιόμορφα κατανεμημένα στην εικόνα, και παρατηρώντας την κίνηση που θα πραγματοποιήσουν υπό την επίδραση των δυνάμεων του πεδίου (δηλαδή μετακινώντας επαναληπτικά κάθε σημείο ελέγχου  $t_j$  κατά διάνυσμα  $\vec{F}(r_j)/|\vec{F}(r_j)|$ ), λαμβάνουμε τις δυναμικές γραμμές του πεδίου. Οι δυναμικές αυτές γραμμές ενσωματώνουν την πληροφορία για τη γενικότερη κατεύθυνση του πεδίου. Ένα βασικό χαρακτηριστικό των δυναμικών γραμμών είναι ότι δεν τέμνονται, καθώς το διάνυσμα της δύναμης σε κάθε σημείο είναι μοναδικό. Έτσι, αν δύο τροχιές φτάσουν στο ίδιο σημείο, από το σημείο αυτό και μετά θα ακολουθήσουν την ίδια δυναμική γραμμή. Αν και άλλες τροχιές φτάσουν σε αυτό το σημείο, τότε και αυτές θα ακολουθήσουν το ίδιο μονοπάτι, σχηματίζοντας «κανάλια», που ονομάζονται **κανάλια δυναμικού**. Τα σημεία μηδενισμού (τοπικά ελάχιστα) της δυναμικής ενέργειας αποτελούν τα **πηγάδια δυναμικού**. Καλούνται έτσι διότι τελικά όλα τα κανάλια καταλήγουν εκεί και ισορροπούν, διότι σε αυτά τα σημεία το πεδίο δυνάμεων μηδενίζεται.

Όπως εξηγείται στο [21], ο μετασχηματισμός της εικόνας στο πεδίο δυνάμεων είναι αντιστρέψιμος, εφ' όσον ο συνολικός αριθμός εικονοστοιχείων της εικόνας είναι άρτιος. Ο μετασχηματισμός από την εικόνα στη συνάρτηση δυναμικής ενέργειας είναι πάντοτε αντιστρέψιμος. Η εικόνα δυναμικής ενέργειας δίνει ένα μέτρο της φωτεινότητας στην περιοχή γύρω από κάθε εικονοστοιχείο, ενώ το πεδίο δυνάμεων περιγράφει την τοπική ισχύ των ακμών (ως κλίση της δυναμικής ενέργειας), μοιάζει δηλαδή αρκετά με μετασχηματισμό ανίχνευσης ακμών. Τα πηγάδια δυναμικού είναι εκείνα τα σημεία στην εικόνα όπου η δύναμη μηδενίζεται, δηλαδή οι δυνάμεις από όλα τα υπόλοιπα εικονοστοιχεία εξισορροπούνται, και συνεπώς τείνουν να βρεθούν σε περιοχές μεταξύ των ακροτάτων της δύναμης, συνεπώς στα εσωτερικά φωτεινών περιοχών. Ένα απλό παράδειγμα φαίνεται στο Σχ. 5.4.



**Εικόνα 5.4:** Απλό παράδειγμα εξαγωγή των καναλιών και πηγαδιών δυναμικού, από [42]. (a) : Αρχική Εικόνα, (b) : Αρχικοποίηση των σημείων ελέγχου, (c) : Ενδιάμεσο στάδιο λήψης των καναλιών (d) : Τα κανάλια δυναμικού (e) : Το μέτρο του πεδίου δυνάμεων

Στο [21] επίσης αναφέρεται ότι η διαδικασία εξαγωγής των καναλιών και πηγαδιών δυναμικού είναι ανεξάρτητη από την αρχική επιλογή των σημείων ελέγχου, πράγμα αναμενόμενο καθώς ο παραπάνω μετασχηματισμός χρησιμοποιεί μόνο την εγγενή πληροφορία της εικόνας. Συνεπώς η δομή του πεδίου δυνάμεων και τελικά και η θέση των πηγαδιών είναι ανεξάρτητη από μετασχηματισμούς περιστροφής και μετατόπισης. Για το λόγο αυτό τα πηγάδια/κανάλια δυναμικού μπορούν να χρησιμοποιηθούν ως περιγραφητές σχήματος, για εφαρμογές όπου οι εικόνες εμφανίζουν κάποια σαφή εσωτερική δομή. Επιπλέον, η παραπάνω αναπαράσταση έχει βρεθεί ότι είναι αρκετά εύρωστη απέναντι στο θόρυβο.

### 5.4.3 Εφαρμογή στην αντιμετώπιση των επικαλύψεων

Στο [42], παρουσιάζεται μια μέθοδος αντιμετώπισης των occlusions μεταξύ χεριών και κεφαλιού με χρήση των καναλιών δυναμικού. Βασίζεται στην παρατήρηση ότι τα πεδία δυνάμεων πριν την εμφάνιση της επικάλυψης (για παράδειγμα, τα πεδία που παράγονται από μια εικόνα όπου υπάρχει μόνο το πρόσωπο του ομιλητή) διαταράσσονται

με την είσοδο του χεριού στην εικόνα, καθώς το χέρι έχει το δικό του πεδίο δυνάμεων το οποίο αλληλεπιδρά με το προϋπάρχον πεδίο. Έτσι, αν διαθέτουμε μια εικόνα του προσώπου του ομιλητή μόνο, και το πεδίο δυνάμεων που προκύπτει από αυτήν, και το συγκρίνουμε με το πεδίο δυνάμεων μετά την επικάλυψη, μπορούμε ενδεχομένως να εξάγουμε πληροφορία σχετικά με τη θέση της επικάλυψης και ενδεχομένως και το ακριβές σχήμα των επικαλυπτόμενων περιοχών.

Οι συγγραφείς στο [42] πρότειναν για την ανίχνευση και επίλυση των επικαλύψεων τη στατιστική μοντελοποίηση (με μίγμα γκαουσιανών) της απόστασης που διανύουν τα σημεία ελέγχου μέχρι να ισορροπήσουν στα πηγάδια δυναμικού. Συγκεκριμένα, έστω ότι έχουμε επικάλυψη ενός μόνο χεριού και του προσώπου. Τότε κατασκευάζεται ένα μοντέλο μίγματος γκαουσιανών, με βάση τις αποστάσεις αυτές, για το πρόσωπο του ομιλητή πριν την επικάλυψη. Κατόπιν, επιλέγεται εκείνο το σημείο ελέγχου για το οποίο η διαφορά μεταξύ της απόστασης που διανύθηκε πριν την επικάλυψη από αυτή που διανύθηκε μετά την επικάλυψη είναι μέγιστη. Το πηγάδι στο οποίο κατέληξε το εν λόγω σημείο ελέγχου, μαζί με όσα πηγάδια βρίσκονται σε μια μικρή γειτονιά αυτού, θεωρούνται ότι ανήκουν στην εισερχόμενη διαταραχή (το χέρι του ομιλητή). Το ίδιο το χέρι εξάγεται λαμβάνοντας τις αρχικές θέσεις όλων των σημείων ελέγχου που κατέληξαν στα πηγάδια του χεριού.

Ωστόσο η διαδικασία αυτή αποτυγχάνει σε αρκετές περιπτώσεις, όπως φαίνεται και στο Σχ. 5.16. Βλέπουμε χαρακτηριστικά ότι λόγω του έντονου φωτισμού στην εικόνα (που αποτελεί κοινό χαρακτηριστικό για τις περισσότερες εικόνες που έχουμε στη διάθεσή μας), ολόκληρη η περιοχή επικάλυψης φαίνεται να έχει ομοιόμορφη φωτεινότητα (τα Σχ. 5.16(f)-Σχ. 5.16(l) χρησιμοποιούν τις συνιστώσες φωτεινότητας των αντίστοιχων εικόνων). Ως εκ τούτου, το πεδίο δυνάμεων δεν ανιχνεύει καμιά ακμή, και τελικά εντοπίζεται μόνο ένα πηγάδι δυναμικού, με αποτέλεσμα όλα τα σημεία ελέγχου να ταξινομούνται εσφαλμένα. Αντίστοιχα και στο Σχ. 5.16(m) παρουσιάζεται πρόβλημα, καθώς η μορφή των πεδίων δύναμης είναι τέτοια που ένας μεγάλος αριθμός από εικονοστοιχεία που ανήκουν στο πρόσωπο έχουν ταξινομηθεί στην περιοχή του χεριού. Πρέπει συνεπώς να αναζητήσουμε μια εναλλακτική προσέγγιση.

Η προσέγγιση που θα επιχειρήσουμε στην παρούσα εργασία είναι λίγο διαφορετική. Συγκεκριμένα, σύμφωνα με τα όσα έχουμε αναφέρει, δεν υπάρχει λόγος να ασχοληθούμε με την ανίχνευση των επικαλύψεων, αφού η ύπαρξή τους είναι ανιχνεύεται ήδη από το στάδιο της κατάτμησης. Μας απασχολεί μόνο η επίλυση των επικαλύψεων, δηλαδή η εύρεση των εικονοστοιχείων που ανήκουν στο επικαλυπτόμενο χέρι. Συνεπώς, μια ιδέα θα ήταν να επιχειρήσουμε να ομαδοποιήσουμε τα πηγάδια δυναμικού σε

τόσες κατηγορίες, όσες και τα αντικείμενα που επικαλύπτονται. Η ταξινόμηση αυτή είναι, κατ' αρχήν, εύλογη, εφόσον όπως είδαμε τα κανάλια δυναμικού έχουν την τάση να ακολουθούν τις φωτεινές περιοχές της εικόνας, να τοποθετούνται δηλαδή εντός των περιοχών του προσκηνίου. Ωστόσο επειδή όπως είδαμε τα πηγάδια δυναμικού αρκετές φορές δεν επαρκούν για το διαχωρισμό των επικαλυπτόμενων περιοχών, θα χρησιμοποιήσουμε για την ομαδοποίηση όχι τα πηγάδια δυναμικού, αλλά κάποιες ενδιάμεσες θέσεις των καναλιών δυναμικού. Αντί δηλαδή να εξελίξουμε τα κανάλια μέχρι την ισορροπία, διακόπτουμε νωρίτερα τη διαδικασία της εξαγωγής τους, με την ελπίδα ότι τα κανάλια θα έχουν ήδη τοποθετηθεί εντός των διαφορετικών αντικειμένων, και ότι η ομαδοποίηση τους θα μας δώσει την κατάτμηση της επικαλυπτόμενης περιοχής. Η ομαδοποίηση μπορεί να γίνει με κάποιον μη επιβλεπόμενο αλγόριθμο ταξινόμησης, όπως ο k-means.

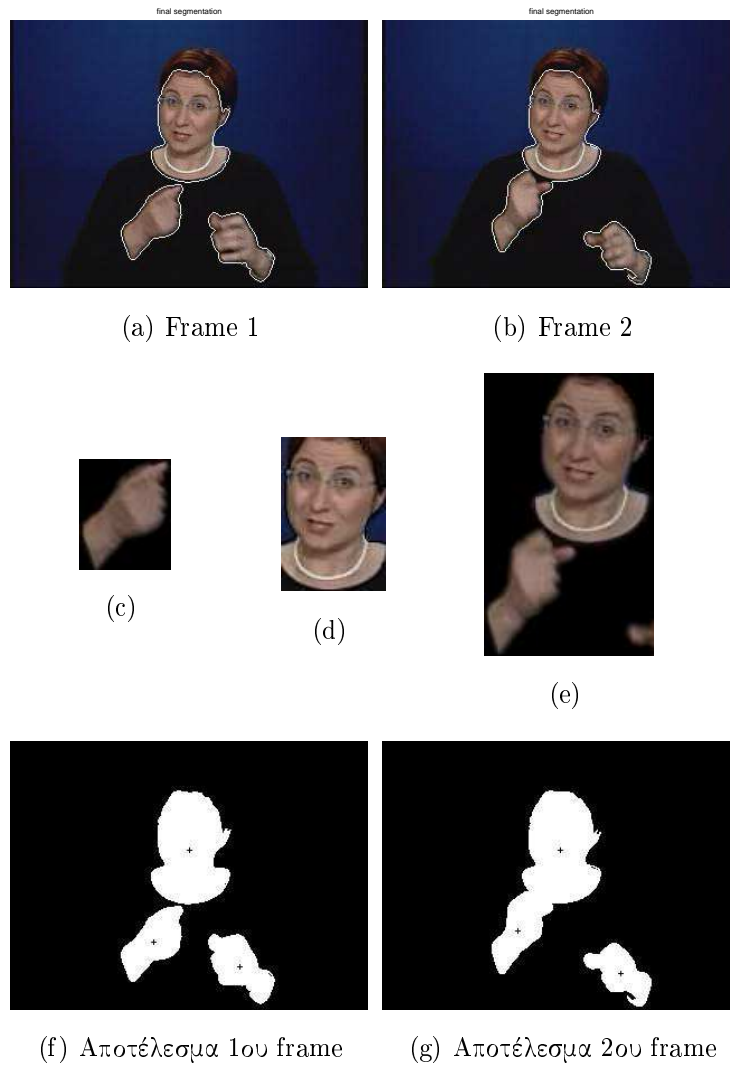
Εναλλακτικά, θα μπορούσαμε να δώσουμε στον ταξινομητή και περαιτέρω πληροφορία για τον τρόπο με τον οποίο πρέπει να πραγματοποιηθεί η ομαδοποίηση. Επειδή, όπως ήδη αναφέραμε, είναι δυνατόν να έχουμε εκ των προτέρων κάποια πολύ καλή γνώση για το πού βρίσκεται το κάθε αντικείμενο (συγκεκριμένα το κινούμενο χέρι), μπορούμε να πραγματοποιήσουμε ομαδοποίηση των σημείων ελέγχου γύρω από την αναμενόμενη αυτές θέσεις. Ειδικότερα, μπορούμε να χρησιμοποιήσουμε τον προβλεπτή κίνησης ή το αποτέλεσμα της σύμπτωσης προτύπων για να τοποθετήσουμε τα κέντρα των χεριών, και για το πρόσωπο (αν υπάρχει στην εικόνα) να χρησιμοποιήσουμε ως κέντρο κάποιο ενδιάμεσο σημείο μεταξύ των δύο.

Τα αποτελέσματα που πήραμε με τις δύο εναλλακτικές προσεγγίσεις παρουσιάζονται στα Σχ. 5.17 - Σχ. 5.19. Βλέπουμε σαφή υπεροχή της 2<sup>ης</sup> μεθόδου, η οποία κατορθώνει και σε δύσκολες περιπτώσεις όπως αυτές των Σχ. 5.17(x), Σχ. 5.18(t), Σχ. 5.19(t) να εντοπίσει τα χέρια στις εικόνες. Βέβαια υπάρχουν και περιπτώσεις όπου αποτυγχάνει, όπως για παράδειγμα στο Σχ. 5.19(p). Γενικότερα πάντως τα αποτελέσματα δείχνουν αρκετά ελπιδοφόρα.

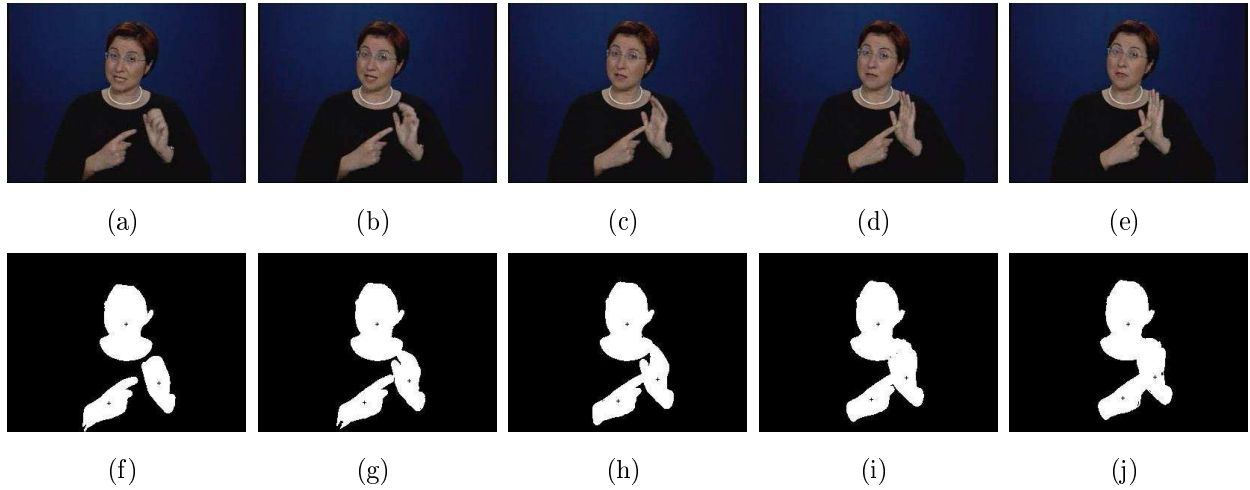
## 5.5 Εικόνες Αποτελεσμάτων

Στις επόμενες σελίδες παρατίθενται τα αποτελέσματα των πειραμάτων του κεφαλαίου 5, με κατάλληλα σχόλια στις λεζάντες όπου χρειάζεται.

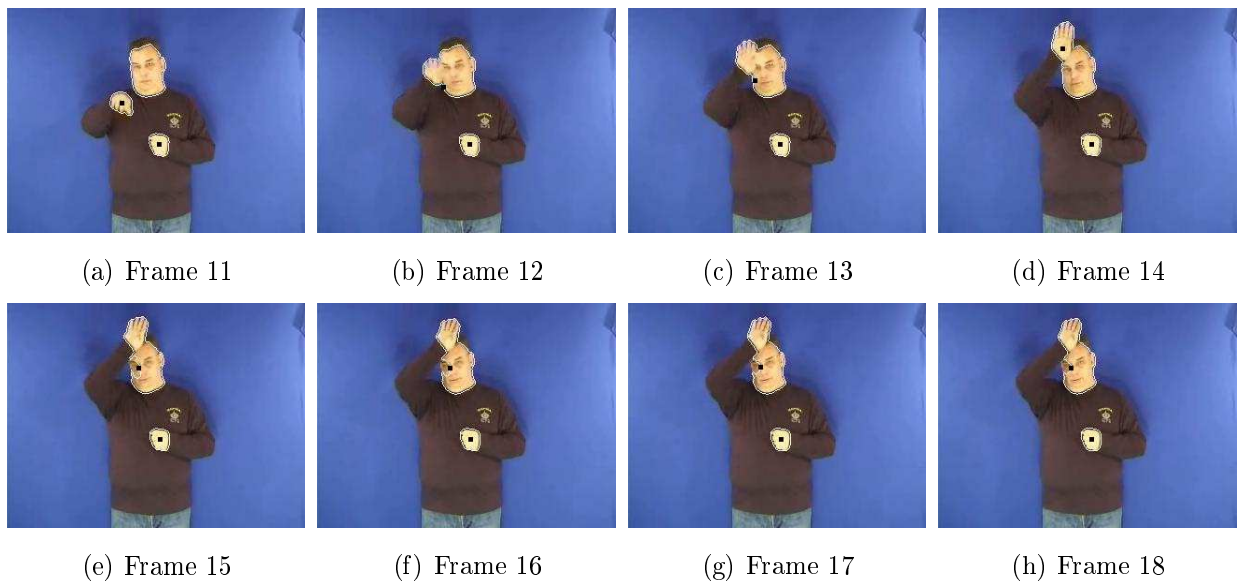




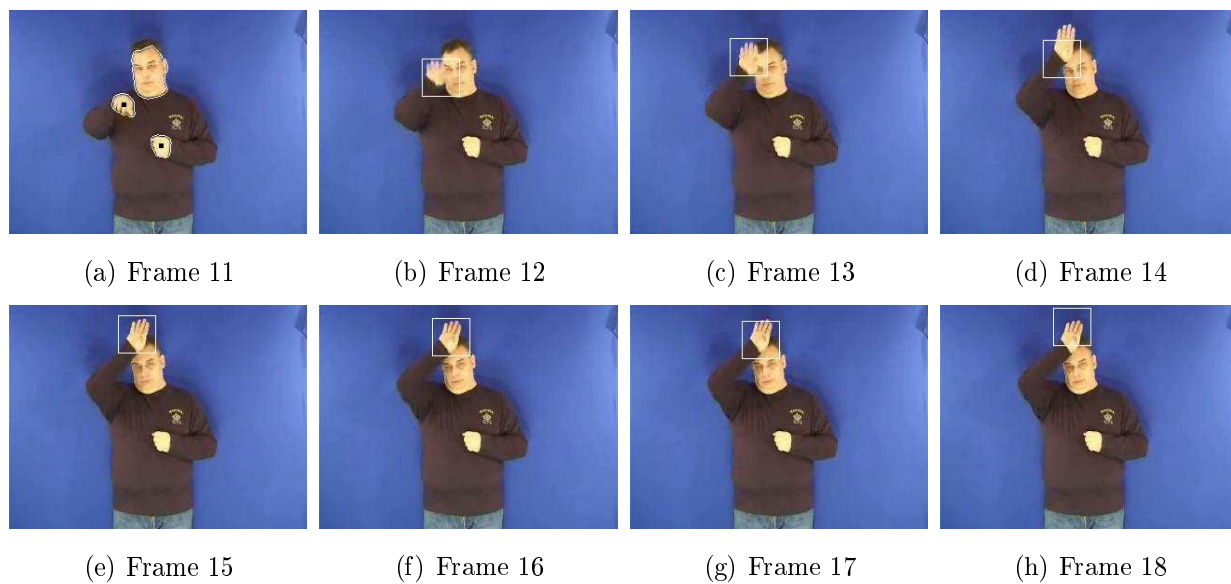
Εικόνα 5.5: Η διαδικασία της σύμπτωσης προτύπων



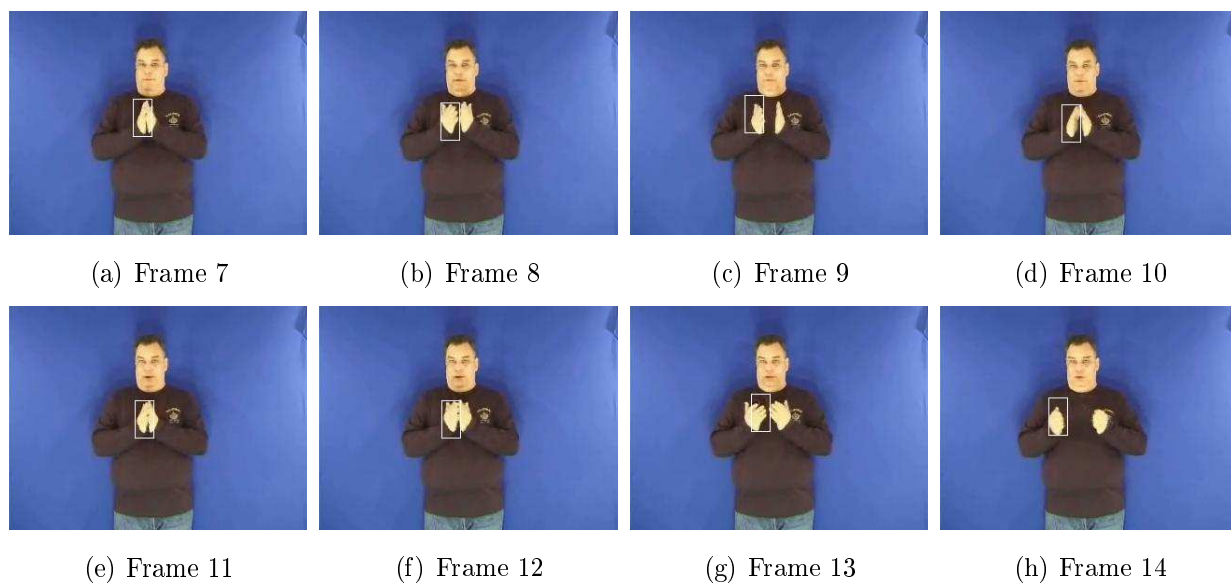
Εικόνα 5.6: Αποτελέσματα της απλής σύμπτωσης προτύπων για 5 διαδοχικά frames



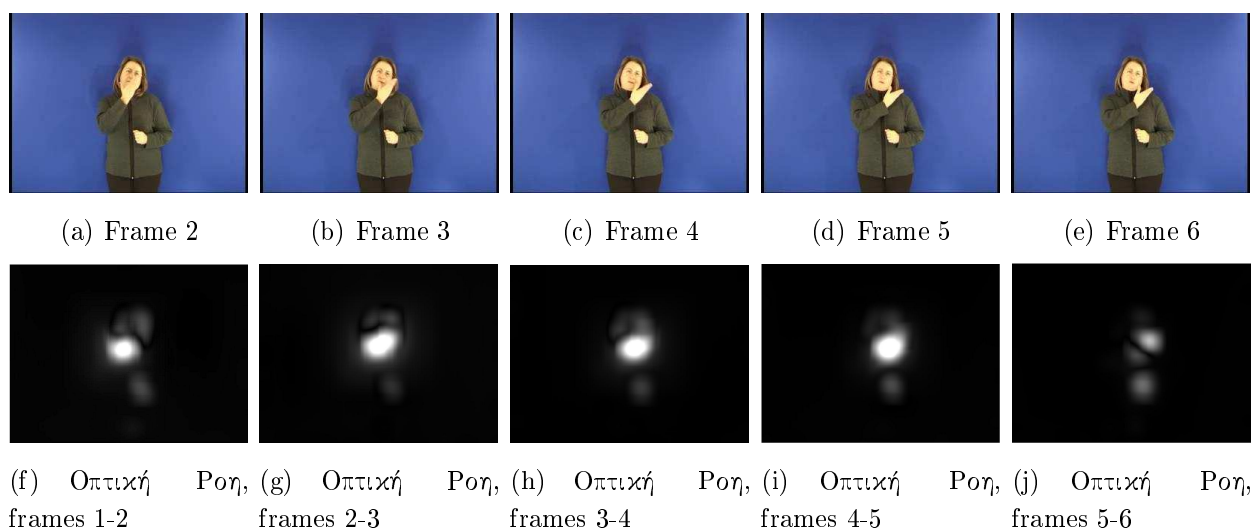
Εικόνα 5.7: Σφάλμα στην παρακολούθηση των κέντρων μάζας στο νεύμα «Βασιλιάς»



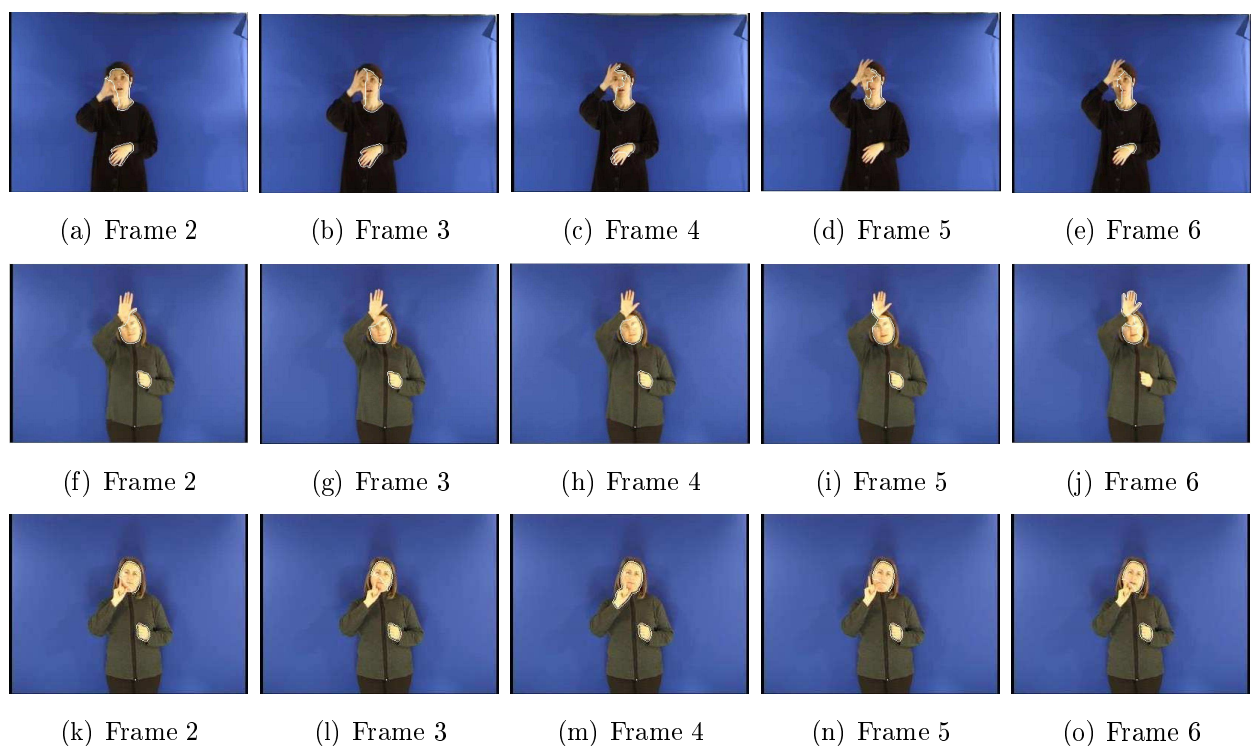
Εικόνα 5.8: Απλή πρόβλεψη κίνησης στο νεύμα «Βασιλιάς»



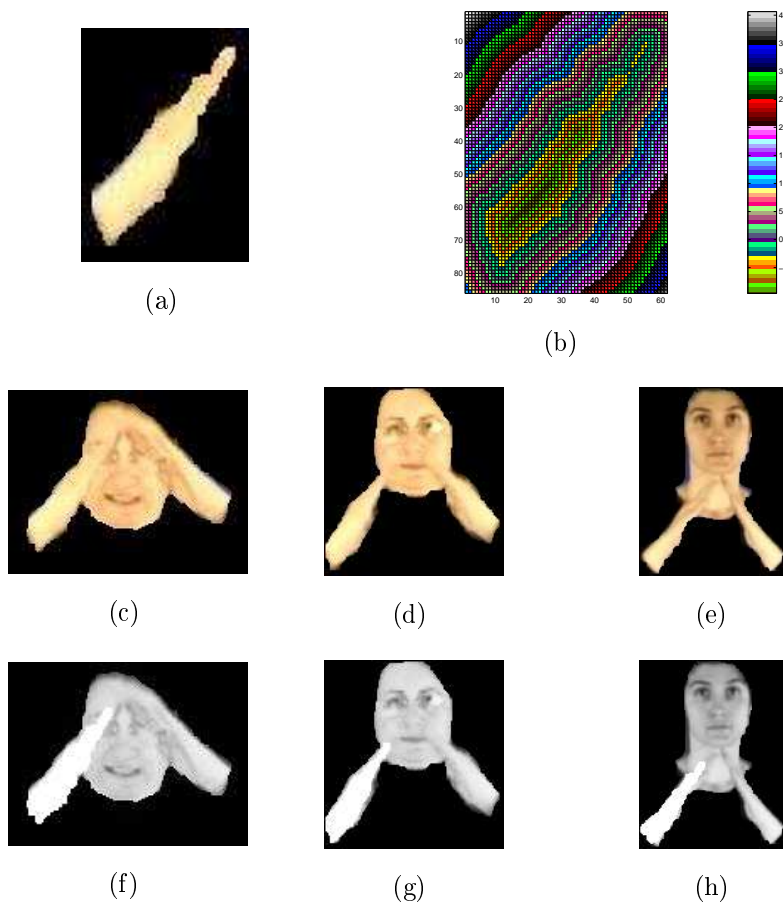
Εικόνα 5.9: Απλή πρόβλεψη κίνησης στο νεύμα «Βιβλίο»



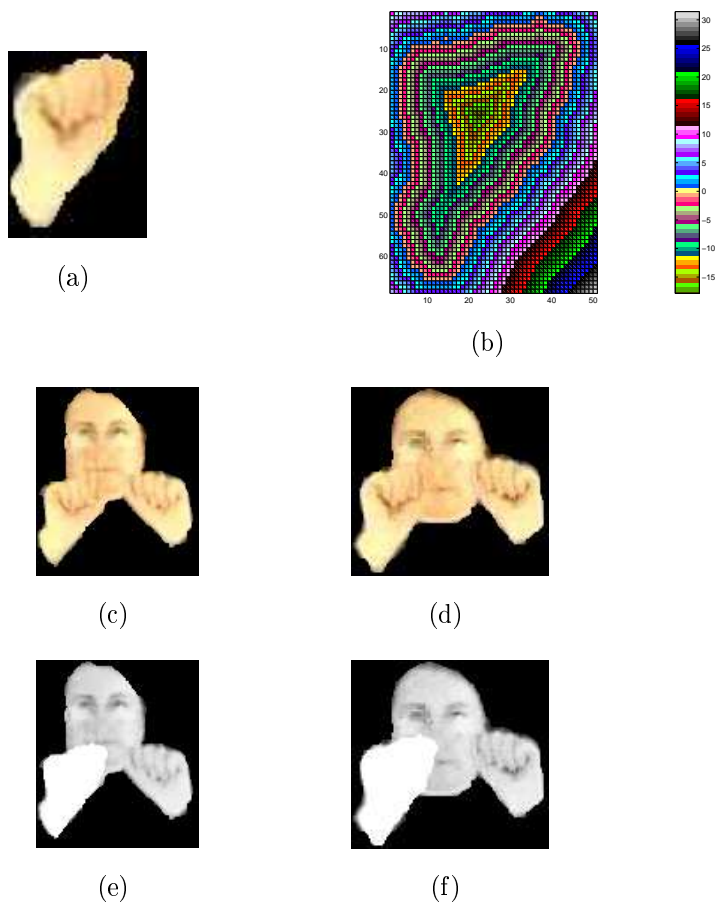
Εικόνα 5.10: Η οπτική ροή ως πιθανότητα κίνησης. Ακολουθία «Χτίζω»



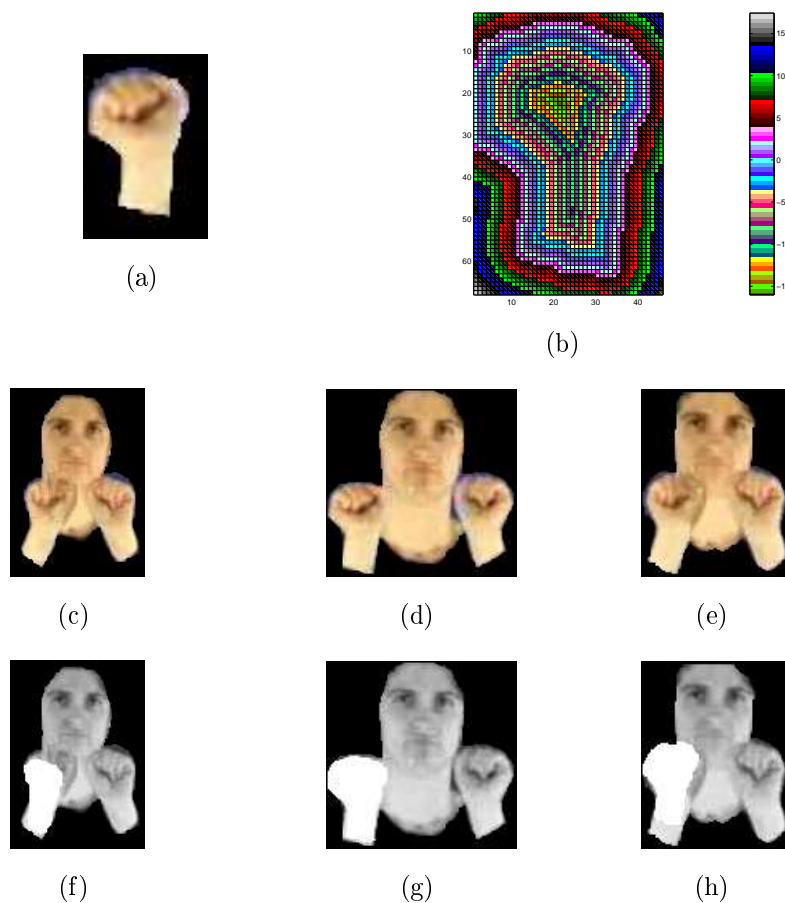
Εικόνα 5.11: Απομόνωση του μη κινούμενου μέρους της εικόνας με χρήση οπτικής ροής και GAC. Σχ. 5.11(a)-Σχ. 5.11(e) : Νεύμα «Αστυνομία», Σχ. 5.11(f)-Σχ. 5.11(j) : Νεύμα «Βασιλιάς», Σχ. 5.11(k)-Σχ. 5.11(o) : Νεύμα «Μύτη».



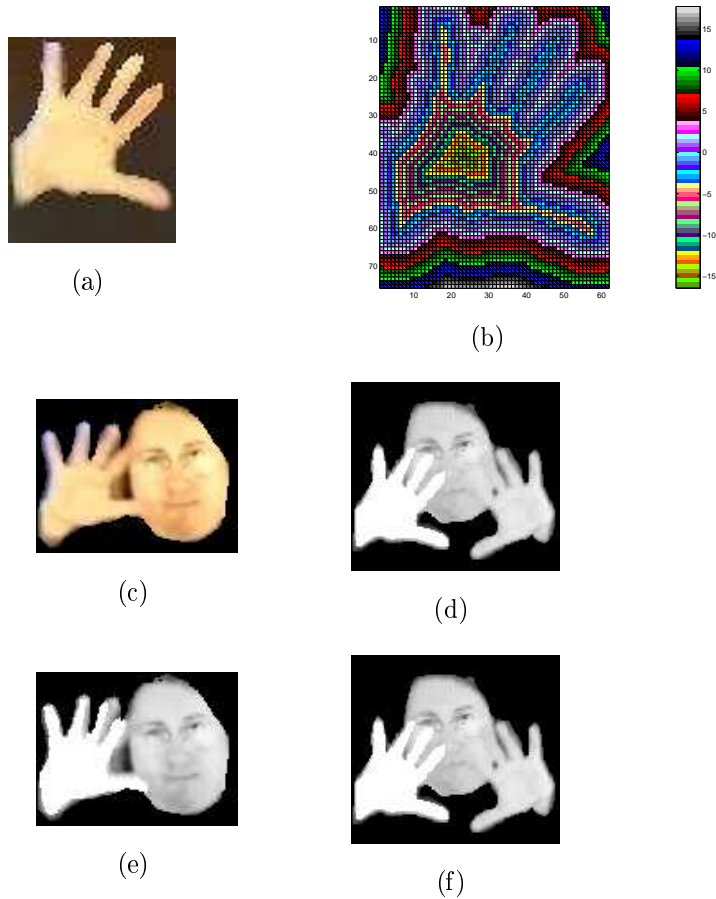
**Εικόνα 5.12:** Αποτελέσματα της Χρήσης Πρότερου Σχήματος - Μέρος Ι. Σχ. 5.12(a): Πρότυπο Σχήμα, Σχ. 5.12(b): Μετασχηματισμός Απόστασης του πρότυπου σχήματος, Σχ. 5.12(c) - Σχ. 5.12(e) : Καρέ όπου εμφανίζεται το πρότυπο σχήμα (Σχ. 5.12(c): Νεύμα «Τέντα», Σχ. 5.12(d)-Σχ. 5.12(e): Νεύμα «Σπίτι» ), Σχ. 5.12(f)-Σχ. 5.12(h):Αποτελέσματα



**Εικόνα 5.13:** Αποτελέσματα της Χρήσης Πρότερου Σχήματος - Μέρος II. Σχ. 5.13(a): Πρότυπο Σχήμα, Σχ. 5.13(b): Μετασχηματισμός Απόστασης του πρότυπου σχήματος, Σχ. 5.13(c) - Σχ. 5.12(d) : Καρέ όπου εμφανίζεται το πρότυπο σχήμα (Δύο διαφορετικά στιγμιότυπα από το νεύμα «Κλείνω τη ντουλάπα»), Σχ. 5.12(f)-Σχ. 5.12(g):Αποτελέσματα

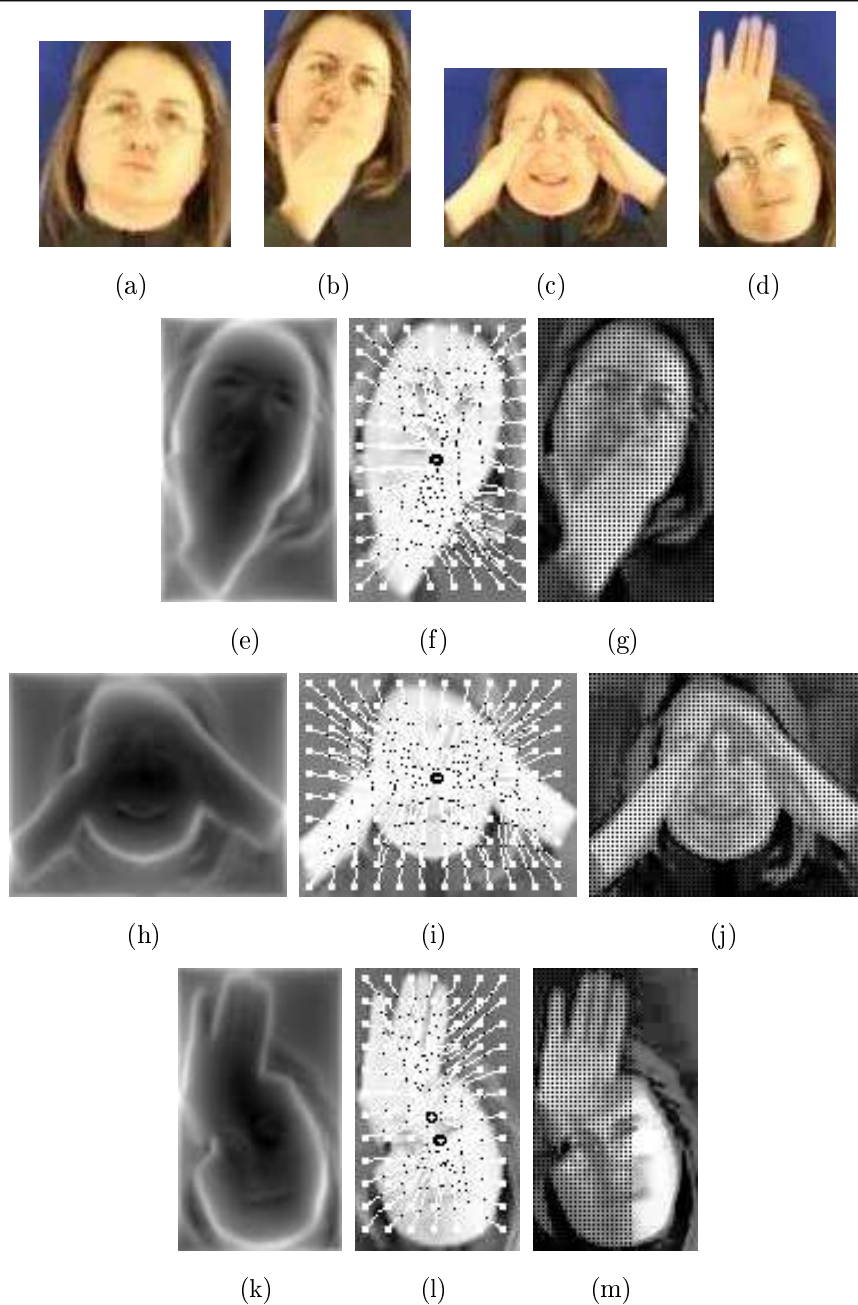


**Εικόνα 5.14:** Αποτελέσματα της Χρήσης Πρότερου Σχήματος - Μέρος III. Σχ. 5.14(a): Πρότυπο Σχήμα, Σχ. 5.14(b): Μετασχηματισμός Απόστασης του πρότυπου σχήματος, Σχ. 5.14(c) - Σχ. 5.14(e) : Καρέ όπου εμφανίζεται το πρότυπο σχήμα (Τρία διαφορετικά στιγμιότυπα από το νεύμα «Κλείνω τη ντουλάπα»), Σχ. 5.14(f)-Σχ. 5.14(h):Αποτελέσματα

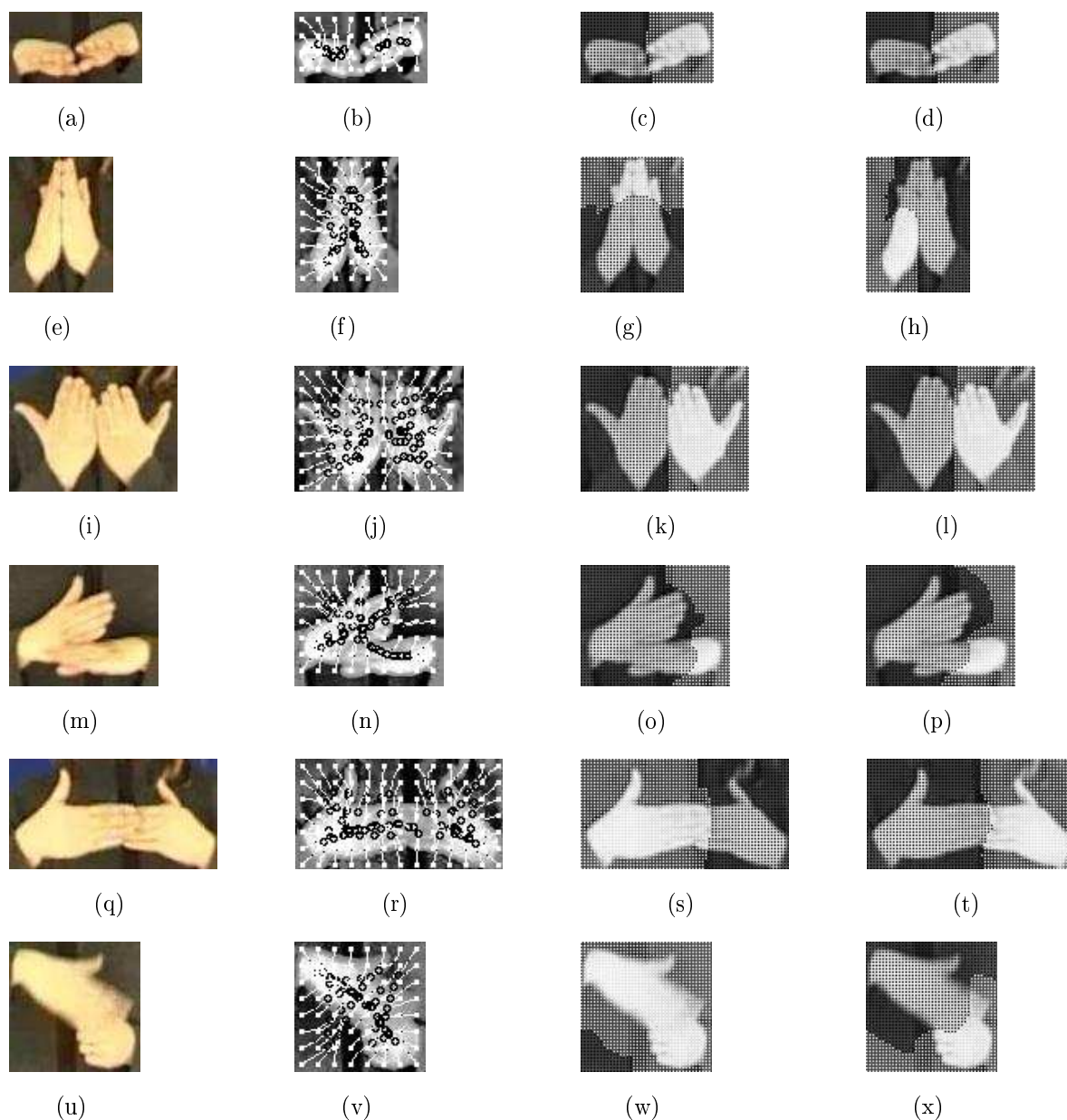


**Εικόνα 5.15:** Αποτελέσματα της Χρήσης Πρότερου Σχήματος - Μέρος IV. Σχ. 5.15(a): Πρότυπο Σχήμα, Σχ. 5.15(b): Μετασχηματισμός Απόστασης του πρότυπου σχήματος, Σχ. 5.15(c) - Σχ. 5.15(d) : Καρέ όπου εμφανίζεται το πρότυπο σχήμα (Δύο διαφορετικά στιγμιότυπα από το νεύμα «Σφαίρα»), Σχ. 5.15(e)-Σχ. 5.15(f):Αποτελέσματα

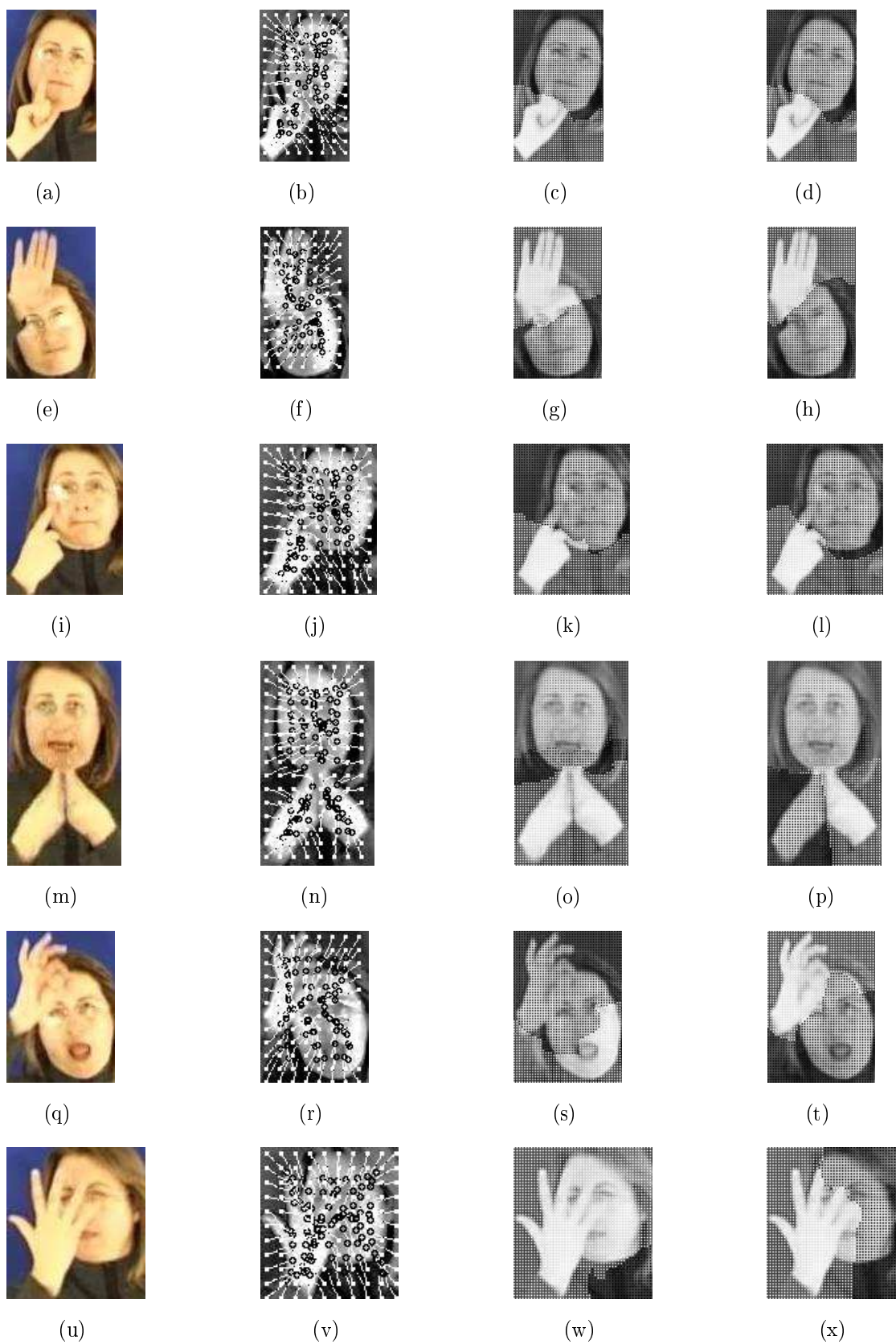




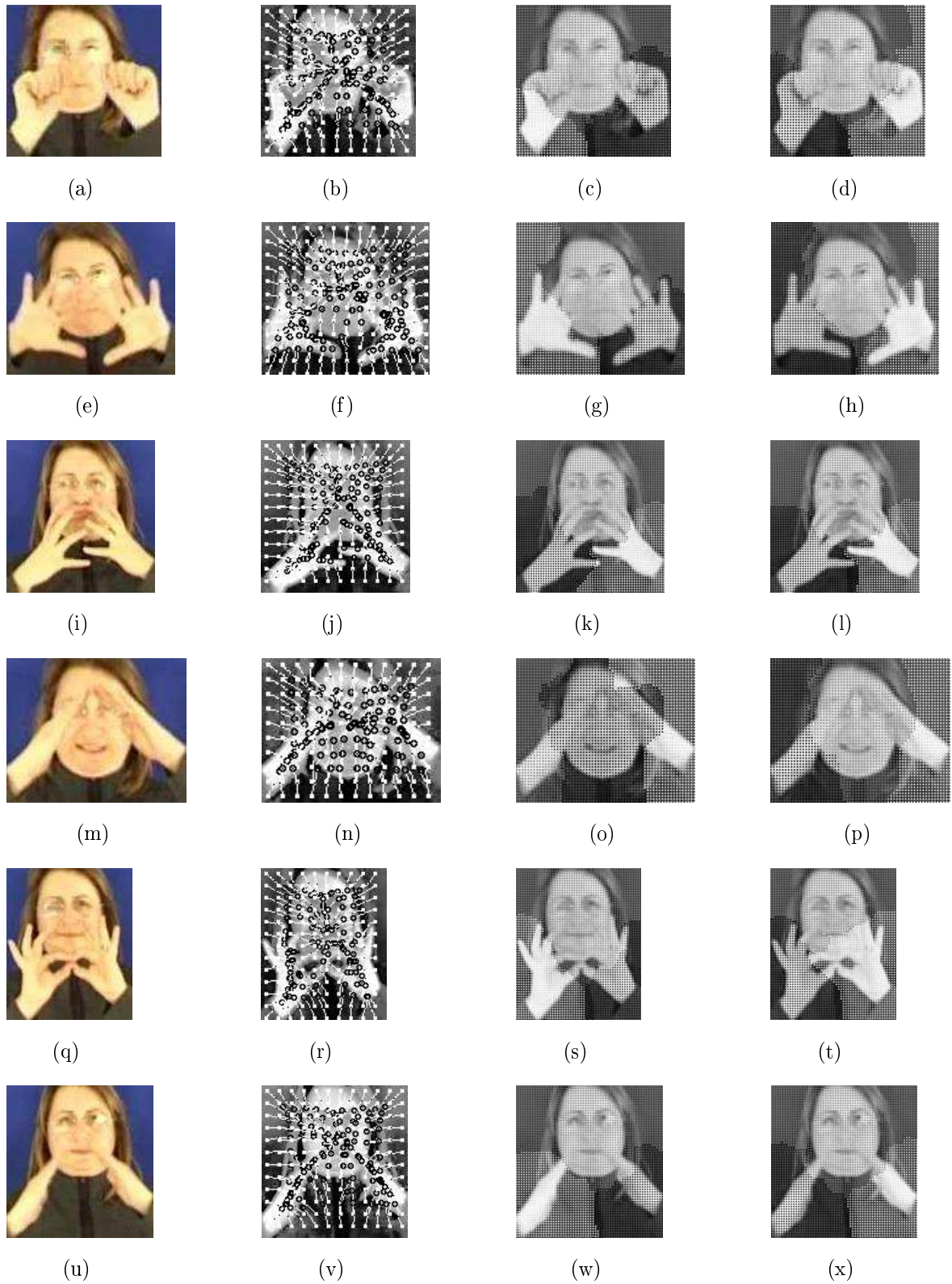
**Εικόνα 5.16:** Αποτυχία της μεθόδου [42] για τρεις τυπικές περιπτώσεις επικαλύψεων (Σχ. 5.16(b):Νεύμα:«Χτίζω», Σχ. 5.16(c):Νεύμα:«Τέντα», Σχ. 5.16(d):Νεύμα:«Βασιλιάς»), με βάση την εικόνα προσώπου του Σχ. 5.16(a). Παρουσιάζονται για κάθε περίπτωση το μέτρο του πεδίου δυνάμεων, τα κανάλια δυναμικού και το αποτέλεσμα της απομόνωσης του χεριού. Στα Σχ. 5.16(g), Σχ. 5.16(j), Σχ. 5.16(m) σημειώνονται με μαύρες κουκκίδες τα σημεία ελέγχου που ταξινομούνται στην περιοχή του/ των χεριού/-ών.



**Εικόνα 5.17:** Χρήση του πεδίου δύναμης και των καναλιών δυναμικού για τον εντοπισμό της θέσης των χεριών σε περιπτώσεις επικάλυψης - Μέρος I. Στην πρώτη στήλη φαίνεται η περιοχή επικάλυψης, στη δεύτερη τα κανάλια δυναμικού μετά από 20 επαναλήψεις (πριν τη σύγκλιση στα πηγάδια δυναμικού), ενώ στην τρίτη και την τέταρτη τα αποτελέσματα της εφαρμογής των δύο μεθόδων που αναφέρθηκαν στην ενότητα 5.4.3. Στις εικόνες των αποτελεσμάτων, έχουν σημειωθεί με κουκκίδες σε διάφορες αποχρώσεις του γκρι (ανάλογα με τον αριθμό των περιοχών που επικαλύπτονται) τα σημεία ελέγχου που ανήκουν στην κάθε περιοχή.



Εικόνα 5.18: Χρήση του πεδίου δύναμης και των καναλιών δυναμικού για τον εντοπισμό της θέσης των χεριών σε περιπτώσεις επικάλυψης - Μέρος II



Εικόνα 5.19: Χρήση του πεδίου δύναμης και των καναλιών δυναμικού για τον εντοπισμό της θέσης των χεριών σε περιπτώσεις επικάλυψης - Μέρος III

---

## ΑΝΑΚΕΦΑΛΑΙΩΣΗ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ ΕΡΕΥΝΑΣ

---

### 6.1 Ανασκόπηση της Διπλωματικής Εργασίας

Στη διπλωματική αυτή έγινε μια προσέγγιση για την ανάπτυξη ενός συστήματος ανάλυσης βίντεο της νοηματικής γλώσσας, με στόχο την εξαγωγή χαρακτηριστικών που θα μπορούσαν να χρησιμοποιηθούν για την αναγνώριση των νοημάτων. Το σύστημα αυτό αποτελείται από τρία στάδια, το στάδιο της κατάτμησης, το στάδιο της παρακολούθησης και το στάδιο της εξαγωγής χαρακτηριστικών.

Για την κατάτμηση χρησιμοποιήθηκε η μέθοδος των γεωδαιτικών ενεργών περιγραμμάτων, η οποία προτιμήθηκε λόγω της επεκτασιμότητάς της, δηλαδή της ικανότητας που διαθέτει να ενσωματώνει και να συνδυάζει πολλαπλή πληροφορία για την κατάτμηση, ανάλογα με την εκάστοτε εφαρμογή. Εδώ χρησιμοποιήθηκε μια νέα στατιστική δύναμη χρώματος, ακολουθώντας το μοντέλο των γεωδαιτικών ενεργών περιοχών (GAR). Η δύναμη αυτή προέκυψε από στατιστική μοντελοποίηση του χρώματος του ανθρώπινου δέρματος, χρησιμοποιώντας δείγματα από πολλές εικόνες.

Όσον αφορά στο κομμάτι της παρακολούθησης των χεριών των νοηματιστών, δοκιμάστηκαν τεχνικές εκτίμησης της οπτικής ροής και της σύμπτωσης blocks, και βρέθηκε ότι οι πρώτες έχουν καλύτερα αποτελέσματα στην περίπτωση της νοηματικής γλώσσας. Τα διανύσματα κίνησης που προκύπτουν μπορούν να χρησιμοποιηθούν για την εκτίμηση των νέων θέσεων των χεριών στα επόμενα καρέ του βίντεο, διαδικασία

που βοηθάει στην εξάλειψη τυχόν ασάφειας που προκύπτει σε περίπτωση επικαλύψεων.

Τελικός στόχος ενός συστήματος ανάλυσης βίντεο της νοηματικής γλώσσας είναι η εξαγωγή χαρακτηριστικών. Τα χαρακτηριστικά που μπορούν να χρησιμοποιηθούν ταξινομούνται σε 2 κατηγορίες : τα **στατικά**, που αποσκοπούν στην περιγραφή του σχήματος των χειρομορφών, και τα **δυναμικά**, που ενσωματώνουν την πληροφορία της κίνησης. Στην παρούσα εργασία δοκιμάσαμε ένα αρκετά μεγάλο σύνολο από χαρακτηριστικά σχήματος, και συγκεκριμένα τους περιγραφητές Fourier, τα χαρακτηριστικά που προκύπτουν από το γραμμικό χώρο καμπυλότητας και τους συντελεστές του cepstrum της καμπυλότητας. Όσον αφορά στα δυναμικά χαρακτηριστικά, χρησιμοποιήθηκε ένα σύνολο χαρακτηριστικών που εξάγεται άμεσα από το πεδίο οπτικής ροής, που είχε εξαχθεί κατά το στάδιο της παρακολούθησης. Το σύνολο αυτό βασίζεται στην απομόνωση εκείνων των τμημάτων της εικόνας που εμφανίζουν τη μεγαλύτερη κινητικότητα.

Μεγάλο μέρος της παρούσας εργασίας αφιερώθηκε και στην επίλυση των προβλημάτων που προκύπτουν από τις επικαλύψεις που πολύ συχνά εμφανίζονται μεταξύ των χειρών του ομιλητή ή μεταξύ των χειρών και του προσώπου. Στις περιπτώσεις οι περιοχές που εξάγονται από την κατάτμηση δεν μπορούν να χρησιμοποιηθούν άμεσα, και πρέπει να ληφθεί ειδική μέριμνα. Δοκιμάστηκαν τέσσερις διαφορετικές μέθοδοι: η μέθοδος της σύμπτωσης προτύπων, η χρήση της πληροφορίας κίνησης (με χρήση του μοντέλου GAR και πάλι) για την απομόνωση μόνο των κινούμενων περιοχών, η εξαγωγή και αξιοποίηση των καναλιών και πηγαδιών δυναμικού από την εικόνα, και η χρήση πρότερου σχήματος. Ο συνδυασμός και η ενοποίηση ορισμένων από αυτές με το στάδιο της κατάτμησης ενδέχεται να αποδώσει σημαντικά στην εξάλειψη του προβλήματος της κατάτμησης.

## 6.2 Πιθανές Μελλοντικές Επεκτάσεις

Με βάση τα όσα αναφέρθηκαν στην παρούσα διπλωματική, η έρευνα θα μπορούσε να συνεχιστεί προς τις παρακάτω κατευθύνσεις:

- Προσπάθεια για την ενσωμάτωση της πληροφορίας της κίνησης ή και κάποιων από τις μεθόδους αντιμετώπισης των επικαλύψεων στο αρχικό στάδιο της κατάτμησης, προκειμένου να ανιχνεύονται και να επιλύονται άμεσα οι επικαλύψεις χωρίς την διαμεσολάβηση ξεχωριστού βήματος

- Ενσωμάτωση στο μοντέλο της κατάτμησης και κάποιας άλλης δύναμης, όπως για παράδειγμα κάποιας δύναμης υψής. Κάτι τέτοιο ενδέχεται να μειώσει σημαντικά την εξάρτηση της διαδικασίας κατάτμησης από τις συνθήκες φωτισμού ή το χρώμα δέρματος του χρήστη, προβλήματα που είναι παρόντα σε όλες τις περιπτώσεις όπου χρησιμοποιείται η πληροφορία του χρώματος.
- Εφαρμογή ενός πιο σύνθετου μοντέλου πρόβλεψης της κίνησης, όπως το φίλτρο Kalman, για τη βελτίωση της εκτίμησης της κίνησης.
- Επέκταση της χρήσης του πρότερου σχήματος στη διαδικασία της παρακολούθησης.
- Έλεγχος της αποδοτικότητας των στατικών και δυναμικών χαρακτηριστικών, σε διάφορους συνδυασμούς, και εύρεση του βέλτιστου συνόλου χαρακτηριστικών που θα επιτρέπει την καλύτερη δυνατή αναγνώριση. Η διαδικασία αυτή βρίσκεται ήδη σε εξέλιξη.





# Βιβλιογραφία

- [1] Π. Μαραγκός. *Ανάλυση Εικόνων και Όραση Υπολογιστών*. Εκδόσεις ΕΜΠ, Αθήνα, 2005.
- [2] D. Adalsteinsson and J. Sethian. A fast level set method for propagating interfaces, 1995.
- [3] A.A. Argyros and M.I.A. Lourakis. Real-time tracking of multiple skin-colored objects with a possibly moving camera. In *Proc. Eur. Conf. Computer Vision*, pages Vol III: 368–379, 2004.
- [4] A. Barjatya. Block matching algorithms for motion estimation, DIP 6620 Spring 2004 Final Project Paper, Utah State University, 2004.
- [5] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *Int'l J. Computer Vision*, 12(1):43–77, February 1994.
- [6] Christos G. Bergeles. Tracking moving objects. *Diploma Thesis*, NTUA 2006.
- [7] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. Eur. Conf. Computer Vision*, pages 237–252, 1992.
- [8] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *Int'l J. Computer Vision*, 22(1):61–79, 1997.
- [9] F. Chaker, M.T. Bannour, and F. Ghorbel. A complete and stable set of affine-invariant fourier descriptors. pages 578–581, 2003.
- [10] Chin-Chen Chang and Chung-Mou Pengwu. Gesture recognition approach for sign language using curvature scale space and hidden markov model. In *ICME*, pages 1187–1190, 2004.

- [11] G. Cortes, L. Garcia, C. Benitez, and J.C. Segura. HMM-based continuous sign language recognition using a fast optical flow parameterization of visual information. In *Proceedings of Interspeech 2006 ICSLP*, Pittsburgh PA, USA, 2006. Interspeech 2006 ICSLP.
- [12] D. Cremers, F. Tischhauser, J. Weickert, and C. Schnorr. Diffusion snakes: Introducing statistical shape knowledge into the mumford-shah functional. *Int'l J. Computer Vision*, 50(3):295–313, December 2002.
- [13] R. Deriche and N. Paragios. Geodesic active contours and level sets for the detection and tracking of moving objects. 22(4):415, April 2000.
- [14] H. Freeman and A. Saghri. Generalized chain codes for planar curves. In *Proceedings of the Joint International Conference on Pattern Recognition, IEEE Computer Society.*, page 701–703, 1978.
- [15] W. Freeman. Orientation histogram for hand gesture recognition, 1995.
- [16] Papandreou G. and Maragos P. Multigrid geometric active contour models. *IEEE Transactions on Image Processing*, 16(1):229–240, January 2007.
- [17] N. Habili, C.C. Lim, and A. Moini. Segmentation of the face and hands in sign language video sequences using color and motion cues. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(8):1086–1097, August 2004.
- [18] B. K.P. Horn and B. G. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17:185–203, August 1981.
- [19] Chung-Lin Huang and Wen-Yi Huang. Sign language recognition using model-based tracking and a 3d hopfield neural network. *Machine Vision and Applications*, 10(5-6):292–307, 1998.
- [20] Y. Huang, T.S. Huang, and H. Niemann. Two-handed gesture tracking incorporating template warping with static segmentation. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 260–265, 2002.
- [21] David J. Hurley, Mark S. Nixon, and John N. Carter. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, 98(3):491–512, 2005.

- 
- [22] K. Imagawa, S. Lu, and S. Igi. Color-based hands tracking system for sign language recognition. In *AFGR*, pages 462–467, 1998.
- [23] J.Goutsias. Markov Random Fields: Interacting Particle Systems for Statistical Image modeling and Analysis. Technical Report 96-01, John Hopkins University, ECE, 1996.
- [24] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. In *Int'l J. Computer Vision*, volume 1, pages 321–333, 1987.
- [25] A. Krzyzak, S.Y. Leung, and C.Y. Suen. Reconstruction of two-dimensional patterns from fourier descriptors. *Machine Vision and Applications*, 2:123–140, 1989.
- [26] I. Kunttu, L. Lepisto, J. Rauhamaa, and A. Visa. Multiscale fourier descriptor for shape-based image retrieval. In *Proc. Int'l Conf. on Pattern Recognition*, pages II: 765–768, 2004.
- [27] Mistsuru Kuroda, Masatake Higashi, Tsuyoshi Saitoh, Yumiko Watanabe, and Tetsuzo Kuragano. Interpolating curve with b-spline curvature function. In *Proceedings of the international conference on Mathematical methods for curves and surfaces II Lillehammer, 1997*, pages 303–310, Nashville, TN, USA, 1998. Vanderbilt University.
- [28] M.E. Leventon, W.E.L. Grimson, and O.D. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. Int'l Conf. on Comp. Vis. & Pat. Rec.*, pages I: 316–323, 2000.
- [29] F. Leymarie and M. D. Levine. Shape Features Using Curvature Morphology. In *Visual Communications and Image Processing IV*, number 1199 in Proc. SPIE, pages 390–401, 1989.
- [30] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of Imaging understanding workshop*, pages 121–130, 1981.
- [31] F. Mokhtarian. Silhouette-based isolated object recognition through curvature scale-space. 17(5):539–544, May 1995.

- [32] F. Mokhtarian and A. Mackworth. A Theory of Multiscale Curvature-based Shape Representation for Planar Curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:789–805, August 1992.
- [33] E.J. Ong and R. Bowden. A boosted classifier tree for hand shape detection. pages 889–894, 2004.
- [34] Sylvie C. W. Ong and Surendra Ranganath. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):873–891, 2005.
- [35] S. Osher and J. Sethian. Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *J. Comp. Phys.*, 79:12–49, 1988.
- [36] N. Paragios and R. Deriche. Geodesic active regions: A new framework to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1/2):249–268, March 2002.
- [37] N. Paragios and R. Deriche. Geodesic Active Regions: A new Paradigm to Deal with Frame Partition Problems in Computer Vision. *Journal of Visual Communication and Image Representation*, 2002.
- [38] M. Rousson and N. Paragios. Shape Priors for Level Set Representations. In *Proc. Eur. Conf. Computer Vision*, 2002.
- [39] G. Sapiro. *Geometric partial differential equations and image analysis*. Cambridge University Press, New York, NY, USA, 2001.
- [40] J. A. Sethian. *Level Set Methods*. Cambridge University Press, 1996.
- [41] K. Sivakumar and J. Goutsias. Morphologically constrained grfs: Applications to texture synthesis and analysis. 21(2), February 1999.
- [42] Paul Smith, Niels da Vitoria Lobo, and Mubarak Shah. Resolving hand over face occlusion. *Image and Vision Computing*, 25(9):1432–1448, 2007.
- [43] N. Soontranon, S. Aramvith, and T.H. Chalidabhongse. Improved face and hand tracking for sign language recognition. *Proceedings of the International Conference on Information Technology: Coding and Computing(ITCC05)*, 2005.

- 
- [44] T.E. Starner, J. Weaver, and A.P. Pentland. Real-time american sign language recognition using desk and wearable computer based video. 20(12):1371–1375, December 1998.
- [45] J. N. Tsitsiklis. Efficient Algorithms for Globally Optimal Trajectories. *IEEE Transactions on Automatic Control*, 49(9):1528–1538, 1995.
- [46] U. von Agris, J. Zieren, U. Canzler, B. Bauer, and K.-F. Kraiss. Recent developments in visual sign language recognition. 2007.
- [47] U. von Agris, J. Zieren, U. Canzler, B. Bauer, and K.-F. Kraiss. Recent developments in visual sign language recognition. 2007.
- [48] C. T. Zahn and R. Z. Roskies. Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers*, 21:269–281, March 1972.
- [49] S.C. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing and Bayes/MDL for Multiband Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.