



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ & ΡΟΜΠΟΤΙΚΗΣ

Παρακολούθηση Κινούμενων Αντικειμένων με Έμφαση στις
Ανθρώπινες Χειρονομίες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Χρήστου Γ. Μπεργελέ

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

Αθήνα, Ιούλιος 2006



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

Παρακολούθηση Κινούμενων Αντικειμένων με Έμφαση στις
Ανθρώπινες Χειρονομίες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Χρήστου Γ. Μπεργελέ

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

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 17η Ιουλίου 2006.

.....
Μαραγκός Πέτρος Καθηγητής
Ε.Μ.Π

.....
Τζαφέστας Κώστας Λέκτορας
Ε.Μ.Π

.....
Μαράτος Νίκος Καθηγητής
Ε.Μ.Π

Αθήνα, Ιούλιος 2006

.....
Χρήστος Γ. Μπεργελές

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών

© (2006) Εθνικό Μετσόβιο Πολυτεχνείο. All rights reserved. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Στόχος της διπλωματικής αυτής είναι η μελέτη και η κατανόηση του προβλήματος της “Παρακολούθησης Κινούμενων Αντικειμένων”, με έμφαση στην “Παρακολούθηση Ανθρώπινων Χειρονομιών”. Γίνεται προσπάθεια διεξοδικής ανάλυσης όλων των παραμέτρων και των προσεγγίσεων του προβλήματος, αντλώντας υλικό από μια ευρεία βιβλιογραφία. Η διπλωματική αυτή είναι διαρθρωμένη σε πέντε κεφάλαια. Στο πρώτο κεφάλαιο γίνεται η εισαγωγή στο πεδίο της Όρασης Υπολογιστών και η παρουσίαση του ειδικού προβλήματος της “παρακολούθησης κινούμενων αντικειμένων” και των εφαρμογών του. Γίνεται μια εκτενής αναφορά σε εφαρμογές προκειμένου να γίνει αντιληπτή η σηματικότητα εύρεσης καλών λύσεων για το πρόβλημα που μελετούμε. Στο δεύτερο κεφάλαιο γίνεται μια συνοπτική παρουσίαση στις τεχνικές κατάτμησης εικόνας με χρήση εξελισσόμενων καμπυλών. Παρουσιάζονται μέθοδοι εξαγωγή και χρήσης προγενέστερων σχημάτων. Στο τρίτο κεφάλαιο παρουσιάζονται αναλυτικά μέθοδοι εύρεσης της κίνησης σε μια εικόνα, και μέθοδοι παρακολούθησης κινούμενων αντικειμένων. Οι μέθοδοι καλύπτουν μια πληθώρα προσεγγίσεων, προκειμένου να γίνει αντιληπτή η ευρύτητα των λύσεων που μπορούν να διατυπωθούν. Στο τέταρτο κεφάλαιο παρουσιάζεται ο αλγόριθμος που υλοποιήθηκε, καθώς και τα αποτελέσματα των προσομοιώσεων. Ο αλγόριθμος συνδυάζει πολλές μεθόδους, προκειμένου να εξελίξει γεωδαιτικές ενεργές καμπύλες με βάση πληροφορία κίνησης και πρότερη γνώση, και τελικά να παρακολουθεί κινούμενα αντικείμενα και ειδικά, ανθρώπινες χειρονομίες. Το τελευταίο κεφάλαιο περιλαμβάνει μια ανακεφαλαίωση της διπλωματικής αυτής και παρατίθενται και μελλοντικές κατευθύνσεις έρευνας.

Λέξεις-Κλειδιά

Παρακολούθηση Κινούμενων Αντικειμένων, Παρακολούθηση Κίνησης, Παρακολούθηση Ανθρώπινων Χειρονομιών, Οπτική Ροή, Προγενέστερα Σχήματα, Γεωδαιτικές Ενεργές Καμπύλες

Abstract

The goal of this thesis is to study and understand the problem of “Tracking Moving Object”, giving emphasis in “Tracking Human Gestures”. Effort has been made in order to fully analyze all the parameters and the approaches of the problem, by utilizing material from an extensive bibliography. This thesis is divided in five chapters. In the first chapter, an introduction in the field of Computer Vision is made, and the problem of “Tracking Moving Objects” is presented. Moreover, a wide variety of applications is discussed, in order to fully appreciate the importance of finding fast and reliable solutions to our problem. In the second chapter, an introduction to the segmentation techniques that use propagating fronts is made. It is made clear that standard segmentation techniques fail when they evolve “blindly”, without any prior information on the object they are seeking. Thus, different approaches to extract and use a prior shape are discussed. In the fourth chapter, different algorithms for the tracking of moving objects are presented. These solutions cover different approaches, in order to clarify the diversity of solutions available. In the fourth chapter the implemented algorithm is presented, together with the results of the simulations. The algorithm incorporates a lot of different methods in order to evolve contours based on motion and prior information. The goal is to track moving objects, and especially human gestures. In the fifth chapter some conclusions are drawn, and the directions of possible future research on the topic are given.

Keywords

Motion Tracking, Tracking of Moving Objects, Gesture Tracking, Optical Flow, Prior Shapes, Geodesic Active Contours.

Ευχαριστίες

Για την εκπόνηση της διπλωματική αυτής σημαντική ήταν η συμβολή του καθηγητή Μαραγκού Πέτρου, που δίνοντάς μου τις βασικές κατευθύνσεις, με βοήθησε να προσεγγίσω και να μελετήσω ένα μεγάλο φάσμα των μεθόδων που χρησιμοποιούνται στην Όραση Υπολογιστών. Για το λόγο αυτό, τον ευχαριστώ θερμά.

Θα ήθελα να επίσης να ευχαριστήσω όσους με στήριξαν και πίστεψαν σε εμένα κατά τη διάρκεια των σπουδών μου και όλα τα σημαντικά πρόσωπα που μου έδιναν την ώθηση να συνεχίσω να προσπαθώ. Ιδίως κατά το τελευταίο έτος των σπουδών μου, ευχαριστώ όσους συνέβαλλαν στις αποφάσεις μου για τις διδακτορικές μου σπουδές, και όσους είχαν την ικανότητα να ομορφαίνουν την καθημερινότητά μου.

“Αυτά τα λόγια ανάρμοστα και βλασφημίες είναι.
Κι αν είσαι στο σκαλί το πρώτο, πρέπει νάσαι υπερήφανος κ' ευτυχισμένος.
Εδώ που έφθασες, λίγο δεν είναι· τόσο που έκαμες, μεγάλη δόξα.
Κι αυτό ακόμη το σκαλί το πρώτο πολύ από τον κοινό τον κόσμο απέχει.
Εις το σκαλί για να πατήσεις τούτο πρέπει με το δικαίωμά σου νάσαι
πολίτης εις των ιδεών την πόλι.
Και δύσκολο στην πόλι εκείνην είναι και σπάνιο να σε πολιτογραφήσουν.
Στην αγορά της βρίσκεις Νομοθέτας που δεν γελά κανένας τυχοδιώκτης.
Εδώ που έφθασες, λίγο δεν είναι· τόσο που έκαμες, μεγάλη δόξα”.

Το Πρώτο Σκαλί, Κ. Π. Καβάφης

Acknowledgments

Towards carrying out this thesis, valuable have been the advices of Professor Maragos Petros. Professor Maragos Petros gave me basic and meaningful directions, and so, I have been able to study a handfull of methods applicable in the field of Computer Vision. For this reason, I express my thanks.

I would also like to thank those who supported me and believed in me during the whole period of my academic studies, and all the important persons that gave me the “push up” to keep trying. Especially during the final year of my studies, I am grateful to those that contributed to my decisions for my graduate studies, and the person who had the ability of making my everyday life special.

“Words like that are improper, blasphemous.
Just to be on the first step should make you happy and proud.
To have reached this point is no small achievement:
what you’ve done already is a wonderful thing.
Even this first step is a long way above the ordinary world.
To stand on this step you must be in your own right
a member of the city of ideas.
And it’s a hard, unusual thing to be enrolled as a citizen of that city.
Its councils are full of Legislators no charlatan can fool.
To have reached this point is no small achievement:
what you’ve done already is a wonderful thing.”
The First Step, by C. P. Cavafis

Πίνακας Περιεχομένων

1	Η Όραση Υπολογιστών και η Παρακολούθηση Κινούμενων Αντικειμένων	1
1.1	Εισαγωγή	1
1.2	Το Πεδίο της Όρασης Υπολογιστών	1
1.3	Το Πρόβλημα της Παρακολούθησης Κινούμενων Αντικειμένων .	5
1.3.1	Εύρεση του Αντικειμένου	5
1.3.2	Παρακολούθηση του Αντικειμένου	6
1.3.3	Εφαρμογές της Παρακολούθησης Κινούμενων Αντικειμένων	7
1.4	Εικόνες Κεφαλαίου 1	12
2	Εύρεση Αντικειμένων σε Εικόνες	17
2.1	Εισαγωγή	17
2.2	Εύρεση Αντικειμένων σε Εικόνα	17
2.2.1	Μέθοδος των Ενεργών Περιγραμμάτων	17
2.2.2	Μέθοδος των Γεωδαιτικών Ενεργών Περιγραμμάτων . .	19
2.2.3	Γεωδαιτικά Ενεργά Περιγράμματα και Επιπεδοσύνολα .	22
2.3	Προγενέστερα Σχήματα και Επιπεδοσύνολα	25
2.3.1	Επεξεργασία του Συνόλου Εκπαίδευσης	26
2.3.2	Εξαγωγή του Προγενέστερου Σχήματος	31
2.3.3	Χρήση του Προγενέστερου Σχήματος	38
2.4	Εναλλακτικές Προσεγγίσεις στο Πρόβλημα των Προγενέστερων Σχημάτων	47
3	Παρακολούθηση Κίνησης	49
3.1	Εισαγωγή	49
3.2	Οπτική Ροή	49
3.2.1	Εισαγωγή στην Οπτική Ροή	49
3.2.2	Μέθοδοι Υπολογισμού της Οπτικής Ροής	50
3.3	Παρακολούθηση Κινούμενων Αντικειμένων	57
3.3.1	Χρήση Οπτικής Ροής και Ενεργών Περιγραμμάτων . . .	58

3.3.2	Εξελίσσοντας το Μοντέλο των Γεωδαιτικών Ενεργών Περιγραμμάτων	59
3.3.3	Παρακολούθηση Κινούμενων Περιοχών χωρίς Υπολογισμό της Κίνησης	60
3.3.4	Παρακολούθηση Αντικειμένων με Μοντελοποίηση Χρώματος	62
3.3.5	Παρακολούθηση Περιοχών με Εύρεση Αφινικών Μετασχηματισμών	65
4	Ανάπτυξη Αλγορίθμου για την Παρακολούθηση Κινούμενων Αντικειμένων	69
4.1	Εισαγωγή	69
4.2	Αλγόριθμοι Εξέλιξης των Γεωδαιτικών Ενεργών Περιγραμμάτων	69
4.2.1	Μέθοδος Στενής Ζώνης	70
4.2.2	Μέθοδος Ταχείας Προέλασης	71
4.2.3	Μέθοδος HERMES	73
4.2.4	Υλοποιώντας το Μοντέλο Εξέλιξης των Γεωδαιτικών Ενεργών Περιγραμμάτων	74
4.2.5	Αποτελέσματα	76
4.3	Αλγόριθμοι Υπολογισμού της Οπτικής Ροής	77
4.3.1	Υπολογισμός της Οπτικής Ροής με τη Μέθοδο Horn-Schunck	77
4.3.2	Αποτελέσματα	79
4.4	Αλγόριθμοι Εύρεσης και Χρήσης Προγενέστερου Σχήματος	80
4.4.1	Επεξεργασία του Συνόλου Εκπαίδευσης	80
4.4.2	Εξαγωγή του Προγενέστερου Σχήματος	81
4.4.3	Χρήση του Προγενέστερου Σχήματος	81
4.4.4	Αποτελέσματα	82
4.5	Αλγόριθμοι Παρακολούθησης Κίνησης	82
4.5.1	Παρακολουθώντας Κινούμενα Αντικείμενα	83
4.5.2	Παρακολουθώντας Ανθρώπινες Χειρονομίες	83
4.6	Εικόνες Κεφαλαίου 4	85
5	Ανακεφαλαίωση και Μελλοντικές Κατευθύνσεις Έρευνας	111
5.1	Ανακεφαλαίωση της Διπλωματικής	111
5.2	Μελλοντικές Κατευθύνσεις	112

Κατάλογος Εικόνων

- 1.1 Το ελικόπτερο του [33] (a) Εν πτήσει, (b) Κατά την προσεδάφιση. Από το [33]. 12
- 1.2 (a) Το ελικοδρόμιο για το ελικόπτερο του [33], (b) Το τελικό αποτέλεσμα της κατάτμησης. Από το [33]. 12
- 1.3 Η παρακολούθηση του αυτοκινητόδρομου στο [8]. (a) Τα χαρακτηριστικά που εξάγονται από τα αυτοκίνητα, (b) Η παρακολούθηση των χαρακτηριστικών των αυτοκινήτων, (c) Ο συσχετισμός των χαρακτηριστικών για τον προσδιορισμό του οχήματος. Από το [8]. 13
- 1.4 Αναγνώριση ενός άρματος M60 από τα κομμάτια του. (a) Κατακόρυφη προβολή του κυρίου σώματος του M60. Τα σημεία του σχηματικού είναι χρωματικά κωδικοποιημένα με βάση το ύψος (πράσινο-κόκκινο-κίτρινο), ενώ τα σημεία του M60 είναι κωδικοποιημένα σε κλίμακα μαύρο-άσπρο. (b) Μια άλλη όψη του M60, η οποία δείχνει ότι η αναγνώριση του κυρίως σώματος έγινε επιτυχημένα και για τους 6 βαθμούς ελευθερίας. (c) Κατακόρυφη προβολή του κανονικού του M60. Η χρωματική κωδικοποίηση είναι παρόμοια με το (a). (d) Μια άλλη όψη του M60, η οποία δείχνει ότι η αναγνώριση του κανονικού έγινε επιτυχημένα και για του 6 βαθμούς ελευθερίας. Από το [39]. . . 14
- 1.5 Οι δύο δυνατές τοποθετήσεις της κάμερας σύμφωνα με το [16] Από το [16]. 14
- 1.6 Το σύστημα παρακολούθησης χειρονομιών για αναγνώριση της νοηματικής γλώσσας, που αναπτύχθηκε στο Georgia Tech. (a) Εικόνα από την κάμερα που βρίσκεται στο καπέλο, (b) Εικόνα από την κάμερα που βρίσκεται στον πάγκο εργασίας. Από το [37]. 15

1.7	(a), (b) Το interface ζωγραφικής του συστήματος που αναπτύχθηκε στο [23], (c) Το Enhanced Desk, ένα σύστημα διαπροσωπείας που εκτελεί παρακολούθηση και αναγνώριση ανθρώπινων χειρονομιών, επιτρέπει στους χρήστες να χειριστούν αληθινά και εικονικά αντικείμενα. Από το [23].	16
1.8	Το σύστημα παρακολούθησης χειρονομιών που αναπτύχθηκε στο [25].	16
4.1	Εξέλιξη Γεωδαιτικού Ενεργού Περιγράμματος με τη μέθοδο της Στενής Ζώνης. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις. (Εικόνα: Γεωμετρικά Σχήματα (128x128)).	85
4.2	Η Στενή Ζώνη που υπολογίζεται. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις. (Εικόνα: Γεωμετρικά Σχήματα (128x128)).	86
4.3	Εξέλιξη Γεωδαιτικού Ενεργού Περιγράμματος με τη μέθοδο της Στενής Ζώνης. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις, (d) Η Στενή Ζώνη μετά από 5 επαναλήψεις, (e) Η Στενή Ζώνη μετά από 10 επαναλήψεις, (f) Η Στενή Ζώνη μετά από 15 επαναλήψεις. (Εικόνα: Χέρι και Mickey (144x144)).	87
4.4	Εξέλιξη Γεωδαιτικού Ενεργού Περιγράμματος με τη μέθοδο της Στενής Ζώνης. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις, (d) Μετά από 20 επαναλήψεις, (e) Μετά από 25 επαναλήψεις. (Εικόνα: Ταξί Αμβούργου (191x191)).	88
4.5	Η Στενή Ζώνη που υπολογίζεται. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις, (d) Μετά από 20 επαναλήψεις, (e) Μετά από 25 επαναλήψεις. (Εικόνα: Ταξί Αμβούργου (191x191)).	89
4.6	Εύρεση Οπτικής Ροής (a) Frame 1, (b) Frame 2: Κινήθηκε ο δίσκος και το μικρό ορθογώνιο, (c) Με χρήση του αλγορίθμου Horn-Schunck, (d) Μέτρο της οπτικής ροής Horn-Schunck, (e) Με χρήση του αλγορίθμου Lukas-Kanade, (e) Μέτρος της οπτικής ροής Lukas-Kanade. (Εικόνα: Γεωμετρικά Σχήματα (128x128)).	90
4.7	Εύρεση Οπτικής Ροής (a) Frame 1, (b) Frame 2: Κινήθηκε ο δίσκος και ο Mickey, (c) Με χρήση του αλγορίθμου Horn-Schunck, (d) Μέτρο της οπτικής ροής Horn-Schunck, (e) Με χρήση του αλγορίθμου Lukas-Kanade, (e) Μέτρος της οπτικής ροής Lukas-Kanade. (Εικόνα: Διάφορα Αντικείμενα (120x120)).	91

- 4.8 Εύρεση Οπτικής Ροής (a) Frame 1, (b) Frame 2: Η παλάμη μετακινήθηκε και άνοιξε, (c) Με χρήση του αλγορίθμου Horn-Schunck, (d) Μέτρο της οπτικής ροής Horn-Schunck, (e) Με χρήση του αλγορίθμου Lukas-Kanade, (e) Μέτρος της οπτικής ροής Lukas-Kanade. (Εικόνα: Χέρι (144x144)). 92
- 4.9 Εύρεση Οπτικής Ροής (a) Frame 1, (b) Frame 2: Μικρή μετακίνηση του πεζού (επάνω αριστερά), του ταξί (κάτω δεξιά), και του εισερχόμενου αυτοκινήτου (αριστερά), (c) Με χρήση του αλγορίθμου Horn-Schunck, (d) Μέτρο της οπτικής ροής Horn-Schunck, (e) Με χρήση του αλγορίθμου Lukas-Kanade, (e) Μέτρος της οπτικής ροής Lukas-Kanade. (Εικόνα: Ταξί Αμβούργου (191x191)). 93
- 4.10 Ευθυγράμμιση των αεροπλάνων (δυναμικές εικόνες 75x75), σύμφωνα με τη μέθοδο του [38]. (a) Το αεροπλάνο περιστρέφεται και διαστέλλεται, (b) Το αεροπλάνο περιστρέφεται και διαστέλλεται, (c) Το αεροπλάνο περιστρέφεται και συστέλλεται, (d) Το αεροπλάνο περιστρέφεται και συστέλλεται. (Εικόνες: Μαχητικά Αεροπλάνα (75x75) (από το [38])). 94
- 4.11 Εικόνες χεριών που ευθυγραμίζονται. (a) Απουσιάζει το τέταρτο δάχτυλο, (b) Απουσιάζει το τέταρτο δάχτυλο και το σχήμα είναι περιστραμμένο, (c) Μετασχηματισμένο σχήμα, (d) Μετασχηματισμένο Σχήμα. Οι εικόνες τραβήχθηκαν με Web-Cam, και στη συνέχεια τροποποιήθηκαν. (Εικόνες: Χέρια (144x144)). 95
- 4.12 Το σύνολο εκπαίδευσης για την εξαγωγή του προγενέστερου σχήματος του χεριού. Τα χέρια φωτογραφήθηκαν με Web-Cam, και οι δυναμικές εικόνες δημιουργήθηκαν χειροκίνητα με χρήση του MS Paint. (Εικόνες: Χέρια (144x144)). 96
- 4.13 Ευθυγράμμιση των χεριών (μετασχηματισμοί προσημασμένης απόστασης 144x144) με βάση τη μέθοδο του [30]. (Εικόνες: Χέρια (144x144)). 97
- 4.14 Το προγενέστερο σχήμα ως μετασχηματισμός προσημασμένης απόστασης, όπως προέκυψε από τη μέθοδο του [32], μετά την ευθυγράμμιση και επεξεργασία των σχημάτων της εικόνας (4.13). (Εικόνα: Προγενέστερο Σχήμα Χεριού (144x144)). 98
- 4.15 Η εξέλιξη του γεωδαιτικού ενεργού περιγράμματος κατευθύνεται και από το προγενέστερο σχήμα. (a) Το τελικό γεωδαιτικό περίγραμμα, στο οποίο έχει ανγοηθεί η σφαίρα, (b) Ο μετασχηματισμός προσημασμένης απόστασης για το γεωδαιτικό περίγραμμα του (a), (c) Το προγενέστερο σχήμα περιστρέφεται κατάλληλα προκειμένου να ευθυγραμμίζονται οι μετασχηματισμοί προσημασμένης απόστασης. (Εικόνα 144x144). 99

- 4.16 Η εξέλιξη του γεωδαιτικού ενεργού περιγράμματος κατευθύνεται και από το προγενέστερο σχήμα. (a) Το τελικό γεωδαιτικό περίγραμμα στο οποίο έχει αγνοηθεί ο Mickey, και έχει εντοπιστεί μόνο το χέρι. Το γάντι εξασφαλίζει έντονες ακμές. Είναι εμφανής η διαφορά στην κατάτμηση, σε σχέση με την εικόνα (4.3), (b) Ο μετασχηματισμός προσημασμένης απόστασης για το γεωδαιτικό περίγραμμα του (a), (c) Το περιστραμμένο προγενέστερο σχήμα που κατευθύνει την κατάτμηση. (Εικόνα: 144x144). 100
- 4.17 Εξέλιξη γεωδαιτικού ενεργού περιγράμματος. (a), (c), (e) Χωρίς χρήση προγενέστερου σχήματος για 2, 20 και 40 επαναλήψεις αντίστοιχα. (b), (d), (f) Με χρήση προγενέστερου σχήματος για 2, 20 και 40 επαναλήψεις αντίστοιχα. (Εικόνα: Άνθρωπος και Χέρι (352x288)). 101
- 4.18 Το γεωδαιτικό ενεργό περίγραμμα αρχικοποιείται στα όρια της εικόνας. Με τη χρήση του προγενέστερου σχήματος, αγνοεί τις ακμές του σώματος και εντοπίζει το χέρι και στις τρεις περιπτώσεις. Οι εικόνες έχουν ληφθεί από Web-Cam. (Εικόνα: Άνθρωπος και Χέρι(144x176)). 102
- 4.19 Το γεωδαιτικό ενεργό περίγραμμα αρχικοποιείται στα όρια της εικόνα. Με τη χρήση του προγενέστερου σχήματος, αγνοεί τις ακμές του σώματος και εντοπίζει το χέρι και στις δύο περιπτώσεις. Οι εικόνες έχουν ληφθεί από Web-Cam. (Εικόνα: Άνθρωπος και Χέρι (352x288)). 103
- 4.20 Εξέλιξη γεωδαιτικού ενεργού περιγράμματος στα όρια του αλγορίθμου. (a), (b) Εξέλιξη της καμπύλης με μεγάλο συντελεστή δύναμης μπαλονιού, για 10 και 40 επαναλήψεις αντίστοιχα, (c), (d) Εξέλιξη της καμπύλης με μεγάλο συντελεστή δύναμης προγενέστερου σχήματος, για 10 και 40 επαναλήψεις αντίστοιχα. (Εικόνα: Άνθρωπος και Χέρι (352x288)). 104
- 4.21 Η γεωδαιτική ενεργή καμπύλη αρχικοποιείται περικλύοντας μόνο τη γροθιά. Η καμπύλη σε κάθε frame εξελίσσεται πρώτα με χρήση των διανυσμάτων οπτικής ροής, και στη συνέχεια ως γεωδαιτική ενεργή καμπύλη. (a) Frame 10, (b) Frame 40, (c) Frame 80, (d) Frame 120. Παρατηρούμε ότι παρακαλουθείται αρκετά περίπλοκη κίνηση. (Video: Χέρι (352x288), 350 frames). 105
- 4.22 Η γεωδαιτική καμπύλη εξελίσσεται, όπως και στην εικόνα (4.21). (a) Frame 160, (b) Frame 200, (c) Frame 280, (d) Frame 350. (Video: Χέρι (352x288), 350 frames). 106

- 4.23 Παρακολούθηση κίνησης. (a) Frame 1: Το γεωδαιτικό ενεργό περίγραμμα εξελίσσεται ώστε να βρει τα αντικείμενα, (b) Frame 2: Χρησιμοποιώντας πληροφορία οπτικής ροής, η καμπύλη εξελίσσεται ώστε να βρει το κινούμενο αντικείμενο, (c) Η καμπύλη μετά από 5 επαναλήψεις, (d) Η καμπύλη μετά από 9 επαναλήψεις. Παρατηρούμε ότι τα όρια του ταξί παραμένουν, ενώ τα άλλα αντικείμενα χάνονται. (Video: Ταξί Αμβούργου (258x192), 41 frames). 107
- 4.24 Παρακολούθηση κίνησης. Το γεωδαιτικό ενεργό περίγραμμα εξελίσσεται, με βάση την πληροφορία της οπτικής ροής, τόσο ως διάνυσμα όσο και ως μέτρο. (a) Frame 10, (b) Frame 20, (c) Frame 30, (d) Frame 40. Παρατηρούμε ότι τελικά η καμπύλη περικλύει μόνο το κινούμενο αντικείμενο της αρχικής εικόνας, δηλαδή το ταξί. (Video: Ταξί Αμβούργου (258x192), 41 frames).108
- 4.25 Παρακολούθηση κίνησης. Η καμπύλη αρχικοποιείται στα όρια της εικόνας, και εξελίσσεται με βάση το προγενέστερο σχήμα, προκειμένου να βρει το επιθυμητό αντικείμενο. Στη συνέχεια, η θέση της ανανεώνεται με χρήση των διανυσμάτων οπτικής ροής, και εξελίσσεται σύμφωνα με το μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων. (a) Frame 10, (b) Frame 40, (c) Frame 70, (d) Frame 100. (Video: Χέρι (352x288), 240 frames).109
- 4.26 Παρακολούθηση της κίνησης του χεριού, όπως και στην εικόνα (4.25). (a) Frame 150, (b) Frame 200, (c) Frame 230. (Video: Χέρι (352x288), 240 frames). 110

Κεφάλαιο 1

Η Όραση Υπολογιστών και η Παρακολούθηση Κινούμενων Αντικειμένων

1.1 Εισαγωγή

Η συγκεκριμένη διπλωματική είναι πάνω στο πρόβλημα της παρακολούθησης κινούμενων αντικειμένων, του επιστημονικού κλάδου της Όρασης Υπολογιστών. Έτσι, κρίνεται απαραίτητο να γίνει μια σύντομη εισαγωγή τόσο στον τομέα της Όρασης Υπολογιστών, όσο και στο πρόβλημα της παρακολούθησης κινούμενων αντικειμένων. Αρχικά γίνεται μια σύντομη αλλά περιεκτική περιγραφή του κλάδου της Όρασης Υπολογιστών, των επιρροών που έχει δεχθεί αυτός και των προβλημάτων που καλείται να αντιμετωπίσει. Επίσης, αναφέρονται οι πιθανές εφαρμογές των αλγορίθμων που αναπτύσσονται από τους επιστήμονες της Όρασης Υπολογιστών. Στη συνέχεια, γίνεται μια πρώτη προσέγγιση του προβλήματος παρακολούθησης κινούμενων αντικειμένων, και προσπαθούμε να διασπάσουμε το πρόβλημα σε επιμέρους υποπροβλήματα έτσι ώστε να είναι ευκολότερη η κατανόησή του. Ιδιαίτερη έμφαση δίνεται στις εφαρμογές που μπορούν να έχουν οι αλγόριθμοι παρακολούθησης κινούμενων αντικειμένων. Έτσι, καθίσταται εμφανής η σημαντικότητα καλών και γρήγορων λύσεων για το πρόβλημα της παρακολούθησης κινούμενων αντικειμένων.

1.2 Το Πεδίο της Όρασης Υπολογιστών

Με τον όρο Όραση Υπολογιστών ή Υπολογιστική Όραση, εννοούμε το επιστημονικό πεδίο που ασχολείται με την εξαγωγή χρήσιμων πληροφοριών από εικόνες ή από αλληλουχίες εικόνων (*video*). Οι αλγόριθμοι οι οποίοι αναπτύσ-

σονται στο πεδίο της Όρασης Υπολογιστών, καλούνται να προσεγγίσουν όσο το δυνατόν πληρέστερα την ανθρώπινη δυνατότητα αντίληψης των εικόνων. Τα συστήματα τα οποία αναπτύσσονται με βάση αυτούς τους αλγορίθμους, προσπαθούν να δώσουν στον υπολογιστή δυνατότητες τις οποίες θεωρούμε *de facto* στον άνθρωπο. Γενικά, τρεις είναι οι κύριες κατευθύνσεις της Όρασης Υπολογιστών, όπως αναφέρεται στο [1]:

- Τεχνητή (Μηχανική) Όραση: εφοδιασμός μηχανών, ρομπότ και αισθητήρων, με δυνατότητες και στοιχεία βιολογικής όρασης.
- Βιολογική Όραση: ανάλυση και κατανόηση της βιολογικής όρασης με υπολογιστικά μοντέλα και χρησιμοποίηση των ενδείξεων για τη δημιουργία μηχανικής όρασης.
- Μοντελοποίηση Εγκεφάλου: η όραση είναι η σημαντικότερη αίσθηση για την αντίληψη του κόσμου. Η κατανόηση του πώς σκεπτόμαστε εξαρτάται από την κατανόηση του πώς βλέπουμε.

Ο κλάδος της Υπολογιστικής Όρασης είναι διεπιστημονικός και, αν και προέκυψε ως παρακλάδι της Τεχνητής Νοημοσύνης, έχει δεχθεί επιρροές από κλάδους όπως η Επεξεργασία Σήματος, η Αναγνώριση Προτύπων, η Φυσική, τα Μαθηματικά, ο Αυτόματος Έλεγχος και η Βιολογική Όραση. Το διεπιστημονικό πεδίο της Όρασης Υπολογιστών άρχισε την ανάπτυξή του στην δεκαετία του 1960.

Στην επιρροή από την Τεχνητή Νοημοσύνη οφείλεται το γεγονός ότι πολλοί από τους αλγόριθμους που χρησιμοποιούνται είναι αλγόριθμοι αναζήτησης, οι οποίοι έχουν αναπτυχθεί για τις εφαρμογές της συγκεκριμένης επιστήμης. Οι τεχνικές της Επεξεργασίας Σήματος, αν και αφορούν μονοδιάστατα σήματα, μπορούν εύκολα να επεκταθούν ώστε να εφαρμοστούν στα δισδιάστατα σήματα τα οποία επεξεργάζεται η Όραση Υπολογιστών. Η Αναγνώριση Προτύπων είναι άμεσα συνδεδεμένη με την Όραση Υπολογιστών, δεδομένου ότι η τελευταία ξεκίνησε μετά από προσπάθειες κατάταξης αντικειμένων μέσω απλής καταωφλιοποίησης εικόνων. Η επιστήμη του Αυτόματου Ελέγχου και η Υπολογιστική Όραση έχουν πολλά κοινά σημεία. Άλλωστε, το πρόβλημα της παρακολούθησης κινούμενων αντικειμένων μπορεί να θεωρηθεί ως πρόβλημα αυτομάτου ελέγχου. Θέλουμε η έξοδος του αλγορίθμου να παρακολουθεί ένα σήμα αναφοράς, ελαχιστοποιώντας το σφάλμα παρακολούθησης. Μέσω της Φυσικής είναι δυνατόν να εξαχθούν πληροφορίες όπως οι ταλαντώσεις των κυμάτων της θάλασσας, από μια απλή εικόνα αυτών. Επίσης, η εξήγηση της συμπεριφοράς του φωτός και της διαδρομής που ακολουθεί από την πηγή μέχρι την κάμερα, είναι απαραίτητη για την κατασκευή τρισδιάστατων αντικειμένων από εικόνες. Τα Μαθηματικά υποστηρίζουν την Όραση Υπολογιστών μέσω

μεθόδων στατιστικής ενώ τέλος, η κατανόηση της Βιολογικής Όρασης βοηθάει στην δημιουργία καλύτερων και πιο ακριβών μοντέλων για το ανθρώπινο σύστημα όρασης.

Ενδεικτικά, κάποια προβλήματα που καλείται να λύσει η Όραση Υπολογιστών είναι η κατάτμηση εικόνας, ο εντοπισμός αντικειμένων, ο διαχωρισμός υφών, η ανίχνευση ακμών, αλλά και προβλήματα τα οποία έχουν να κάνουν με τη σημασιολογία των εικόνων, όπως για παράδειγμα η αναγνώριση μιας ανθρώπινης δραστηριότητας, η αναγνώριση ενός αντικειμένου και η κατάταξη αυτού σε μια ομάδα, η αναγνώριση και κατάταξη ενός τοπίου σε θαλάσσιο, πεδινό κ.τ.λ. Βλέπουμε ότι τα προβλήματα μπορούν να χωριστούν σε αυτά τα οποία αφορούν την εξαγωγή πληροφορίας από μια εικόνα ή από μια αλληλουχία εικόνων, και στα προβλήματα τα οποία αφορούν την εξαγωγή μετα-πληροφορίας από μια εικόνα, ή από μια αλληλουχία εικόνων. Στο [10] αναφέρεται ότι η Όραση Υπολογιστών καλείται να κατανοήσει τις φυσικές διαδικασίες σχηματισμού των εικόνων, να εξάγει συμπεράσματα απλά από τις τιμές των *pixels*, να συνδυάσει τις πληροφορίες από μια πλειάδα εικόνων με ένα συνεκτικό τρόπο, να ταξινομήσει ομάδες από *pixels* έτσι ώστε να τα διαχωρίσει από τα υπόλοιπα, ή να εξάγει πληροφορίες για σχήματα και να αναγνωρίσει αντικείμενα με βάση πληροφορία γεωμετρίας ή με πιθανοτικές τεχνικές. Τα προβλήματα που καλείται να επιλύσει ο κλάδος διαχωρίζονται σε διάφορες υποκατηγορίες, δημιουργώντας μια πυραμιδοειδή ιεραρχία. Για να επιλυθούν τα προβλήματα στην κορυφή της πυραμίδας, πρέπει να χρησιμοποιηθούν οι λύσεις των προβλημάτων χαμηλότερου επιπέδου. Μια σύντομη κατάταξη κάποιων από τα προβλήματα, από τα πιο απλά στα πιο σύνθετα είναι η παρακάτω ([1]):

- Σχηματισμός εικόνων από τον άνθρωπο και τη μηχανή.
- Μοντελοποίηση και ανάλυση χρώματος.
- Ομαλοποίηση εικόνων και ανάλυση σε πολλαπλές κλίμακες.
- Ανίχνευση ακμών.
- Ανάλυση και σύνθεση υφής.
- Στερέοψη.
- Εντοπισμός και εκτίμηση της ταχύτητας των κινούμενων αντικειμένων.
- Κατάτμηση εικόνων σε ομογενείς περιοχές.
- Αναγνώριση και κατηγοριοποίηση αντικειμένων.

- Μοντελοποίηση φυσικών και τεχνητών σχημάτων.

Παρατηρώντας τα προβλήματα τα οποία καλείται να επιλύσει η επιστήμη της Όρασης Υπολογιστών, αντιλαμβανόμαστε και την πληθώρα εφαρμογών που μπορούν να έχουν και οι οποίες καλύπτουν διάφορους τομείς. Στην Ιατρική, η παρακολούθηση αντικειμένων μπορεί να συνδυαστεί με τις υπάρχουσες μεθόδους τηλε-ιατρικής για ακόμα πιο ακριβείς επεμβάσεις. Επίσης, οι μέθοδοι της κατάτμησης εικόνας μπορούν να βοηθήσουν στην επεξεργασία και κατανόηση των εικόνων που προέρχονται από μηχανήματα ακτίνων x, μαγνητικής τομογραφίας κ.τ.λ. Μέθοδοι κατάτμησης εικόνας και παρακολούθησης κινούμενων αντικειμένων μπορούν να βρουν ευρεία χρήση στην παρακολούθηση της κυκλοφορίας στους αυτοκινητόδρομους. Η τρισδιάστατη ανακατασκευή έχει μεγάλη χρησιμότητα σε εφαρμογές εικονικής πραγματικότητας, ενώ η κατάλληλη πληροφορία σκίασης μπορεί να βοηθήσει στην αναγνώριση ανθρώπων και αντικειμένων υπό διάφορες συνθήκες φωτισμού, έχοντας ως αρχική πληροφορία μόνο μερικές εικόνες. Η κατηγοριοποίηση αντικειμένων ανήκει στα προβλήματα που μπορούν να χρησιμοποιηθούν πιο άμεσα από εφαρμογές Τεχνητής Νοημοσύνης, ενώ είναι προφανής η χρήση που μπορούν να έχουν τέτοιοι αλγόριθμοι στον στρατιωτικό τομέα. Γενικά ο στρατιωτικός τομέας έχει μεγάλη ανάγκη από εφαρμογές της Όρασης Υπολογιστών, ωστόσο όπως είναι αναμενόμενο, λίγα αποτελέσματα γίνονται γνωστά στο ευρύ κοινό. Συνοπτικά, οι περιοχές εφαρμογών της Όρασης Υπολογιστών είναι ([1]):

- 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.
- Ευφείς αισθητήρες και ψηφιακές κάμερες.
- Οπτική παρακολούθηση οχημάτων.
- Τεχνολογία *video*.
- Τεχνολογία διαδικτύου και αυτόματη αναζήτηση.

- Επικοινωνία ανθρώπου με υπολογιστή.
- Ψηφιακές εικαστικές εικόνες.

Όπως είναι κατανοητό, οι εφαρμογές προκύπτουν από την επίλυση των προβλημάτων ανωτέρου επιπέδου. Αυτός είναι άλλωστε και ο λόγος που το μεγαλύτερο μέρος της έρευνας γίνεται πλέον πάνω σε αυτά τα προβλήματα, ενώ μειώνεται η προσπάθεια εύρεσης καλύτερων λύσεων για τα προβλήματα χαμηλού επιπέδου. Γενικά πάντως, ο κλάδος αναπτύσσεται με γοργούς ρυθμούς. Ενώ στο παρελθόν η απαίτηση των αλγορίθμων αλλά και των συσκευών εισόδου-εξόδου εικόνας και *video* για μεγάλη υπολογιστική ισχύ καθιστούσε για πολλούς απαγορευτική την διεξαγωγή έρευνας στον τομέα, πλέον η εμφάνιση φθηνού και γρήγορου υλισμικού (*hardware*) και λογισμικού (*software*) έχει κάνει τον κλάδο πιο προσιτό και ολοένα και περισσότερες εφαρμογές κάνουν την εμφάνισή τους.

1.3 Το Πρόβλημα της Παρακολούθησης Κινούμενων Αντικειμένων

Το πρόβλημα της παρακολούθησης κινούμενων αντικειμένων (Tracking of Moving Objects) ή της παρακολούθησης κίνησης (Motion Tracking) μπορεί γενικά να θεωρηθεί διτό. Όπως περιγράφεται στο [28], περιλαμβάνει το στάδιο κατά το οποίο γίνεται εύρεση του κινούμενου αντικειμένου, και το στάδιο στο οποίο γίνεται η παρακολούθηση αυτού. Έτσι, κάθε σκέλος μπορεί να αντιμετωπιστεί είτε ξεχωριστά (όπως π.χ στα [36], [37], [2]), είτε ενιαία (όπως π.χ στα [28], [21]).

1.3.1 Εύρεση του Αντικειμένου

Το υποπρόβλημα της εύρεσης του αντικειμένου (Object Detection), μπορεί να θεωρηθεί μέρος του γενικότερου προβλήματος της κατάτμησης εικόνας (Image Segmentation). Κατάτμηση εικόνας είναι η διαδικασία κατά την οποία τα αντικείμενα μιας εικόνας διαχωρίζονται από το φόντο της εν λόγω εικόνας¹. Είναι διαδικασία απαραίτητη για όλα τα συστήματα αυτόματης αναγνώρισης εικόνας, καθώς οδηγεί στο διαχωρισμό των αντικειμένων ώστε να μπορούν να υποστούν επεξεργασία στα επόμενα στάδια. Πρακτικά, κατά την κατάτμηση εικόνας γίνεται κατάταξη των *pixels* σε διάφορες κατηγορίες-τιμήματα της εικόνας. Ο διαχωρισμός μπορεί να γίνει είτε χειροκίνητα είτε αυτόματα. Στην περίπτωση που γίνεται χειροκίνητα, κάποιος έμπειρος χειριστής με κατάλληλο

¹Από το <http://rkb.home.cern.ch/rkb/AN16pp/node131.html>

χειρισμό μονάδων αλληλεπίδρασης, κατατμει την εικόνα με βάση διάφορα κριτήρια. Ο χειροκίνητος τρόπος κατάτμησης παρουσιάζει δυσκολία στην επανάληψη, και είναι πολύ χρονοβόρος. Επομένως, ιδιαίτερη έμφαση δίνεται στις αυτόματες μεθόδους κατάτμησης, οι απλούστερες από τις οποίες, αφορούν κατάλληλη κατωφλιοποίηση και χρήση της *grey level* τιμής του *pixel*. Ωστόσο, αυτό δεν οδηγεί πάντα σε ικανοποιητικά αποτελέσματα και επομένως γίνεται χρήση και άλλων παραμέτρων όπως η υφή, το χρώμα, οι ακμές, και ο προσανατολισμός. Στην περίπτωση παρακολούθησης κίνησης είναι αρκετά συχνή η χρησιμοποίηση των ακμών ως επιπρόσθετη πληροφορία, δεδομένου ότι καταδεικνύουν ως ένα βαθμό τα όρια των αντικειμένων. Για την επίλυση του ίδιου προβλήματος, συχνά γίνεται χρήση στατιστικών παραμέτρων που έχουν σχέση με την κίνηση ([28], [21]). Στα [37], [2] χρησιμοποιείται ένα κατασκευασμένο μοντέλο υφής του ανθρωπίνου δέρματος, έτσι ώστε να βρεθούν τα ανθρώπινα χέρια και στη συνέχεια να παρακολουθηθεί η κίνησή τους. Σε άλλες περιπτώσεις σκόπιμα χρωματίζεται το επιθυμητό αντικείμενο ([36]), ή εξοπλίζεται με φωτεινές ενδείξεις, έτσι ώστε να είναι πιο εύκολος ο εντοπισμός του. Τέλος, είναι δυνατόν να εντοπιστεί συγκεκριμένο αντικείμενο στην εικόνα, χρησιμοποιώντας αποτελέσματα που προκύπτουν από την κατάλληλη επεξεργασία ενός συνόλου εκπαίδευσης, δηλαδή από μια βάση επιθυμητών σχημάτων ([18], [19], [29], [32], [31], [38]). Έτσι, χρησιμοποιώντας τις εικόνες π.χ χεριών από ένα σύνολο εκπαίδευσης, κατασκευάζονται μοντέλα-αναπαραστάσεις αυτών, και στη συνέχεια γίνεται κατάτμηση της επιθυμητής εικόνας με βάση την πρότερη γνώση, έτσι ώστε στην εικόνα να εντοπιστούν τα χέρια, και όχι οποιαδήποτε αντικείμενα.

Συνοψίζοντας, η διαδικασία της κατάτμησης εικόνας έχει ως είσοδο μια εικόνα και ως έξοδο τον εντοπισμό των αντικειμένων τα οποία φέρουν κάποια επιθυμητά χαρακτηριστικά από πλευράς υφής, χρώματος, σχήματος ή κίνησης.

Οι εφαρμογές της κατάτμησης εικόνας είναι πολύ ευρείες και κυμαίνονται στους τομείς της ιατρικής (*medical image segmentation*), στους τομείς διαχείρισης κυκλοφορίας, ως αρχικό στάδιο σε εφαρμογές παρακολούθησης κινούμενων αντικειμένων κ.τ.λ. Για να δώσουμε κάποιες πιο λεπτομερείς εφαρμογές της αυτόματης κατάτμησης εικόνας, αναφέρουμε ότι στον τομέα της ιατρικής είναι δυνατόν να εξαχθούν πολύ χρήσιμες ανατομικές πληροφορίες, από εικόνες που λαμβάνονται από μηχανήματα ακτίνων x, υπερήχων και άλλων συσκευών απεικόνισης.

1.3.2 Παρακολούθηση του Αντικειμένου

Όπως είπαμε στην προηγούμενη παράγραφο, το πρώτο σκέλος της παρακολούθησης κινούμενων αντικειμένων μπορεί να θεωρηθεί η εύρεση του επιθυμητού αντικειμένου. Το δεύτερο σκέλος αφορά την παρακολούθηση αυτού του

αντικειμένου (Object Tracking), και αποτελεί το κύριο μέρος της παρακολούθησης κινούμενων αντικειμένων. Σε συνδυασμό με την κατάκτηση εικόνας, οδηγεί στην επιθυμητή λύση του προβλήματος. Παρακολούθηση είναι δυνατόν να επιτευχθεί με διάφορους τρόπους, όπως με χρήση μηχανικών, μαγνητικών, ακουστικών και οπτικών μεθόδων, όπως αναφέρεται στο [22]. Ωστόσο, χρησιμοποιείται γενικά η οπτική μέθοδος, δηλαδή αυτή της επεξεργασίας εικόνας, διότι είναι δυνατόν να απλοποιηθεί πολύ. Αυτό επιτυγχάνεται χρησιμοποιώντας ένα προκαθορισμένο περιβάλλον όπως αναφέραμε και προηγουμένως (χρωματισμός αντικειμένων, μοντελοποίηση υψής κ.τ.λ). Η οπτική μέθοδος είναι η μέθοδος η οποία θα μας απασχολήσει στη διπλωματική αυτή.

Κατά καιρούς έχουν προταθεί διάφοροι αλγόριθμοι για την επίλυση του προβλήματος παρακολούθησης κίνησης. Αρχικά, χρησιμοποιούνταν απλές τεχνικές που υπολόγιζαν τις διαφορές μεταξύ των εικόνων και επιχειρούσαν να παρακολουθήσουν τα αντικείμενα με χρήση κατάλληλων κατωφλιών. Σε άλλες περιπτώσεις γίνονταν χρήσεις στατιστικών αποτελεσμάτων, ενώ κάποιοι άλλοι αλγόριθμοι προσπαθούσαν να βρουν τις τοπικές αλλαγές. Άλλοι αλγόριθμοι προσπαθούσαν να βρουν βέλτιστους μετασχηματισμούς μεταξύ των διαδοχικών εικόνων, έτσι ώστε από την μια εικόνα, με εφαρμογή του μετασχηματισμού να προκύπτει η επόμενη (Image Registration). Τέλος, σε πολλές περιπτώσεις γίνονταν παρακολούθηση κάποιων χαρακτηριστικών (features) των αντικειμένων, ή τα τελευταία οριοθετούνταν και παρακολουθούνταν με τη χρήση των ακμών. Καινούριες τεχνικές χρησιμοποιούν ως κεντρική ιδέα την ελαχιστοποίηση κάποιων συναρτησιακών (minimization of energy functionals) που έχουν σχέση με την κίνηση. Με τη χρήση των συναρτησιακών αυτών είναι δυνατόν το πρόβλημα της παρακολούθησης κινούμενων αντικειμένων να έχει μια εννιαία αντιμετώπιση όπως στα [28], [21].

Συνοψίζοντας, η παρακολούθηση ενός αντικειμένου έχει ως είσοδο ένα επιθυμητό αντικείμενο, και ως έξοδο τον εντοπισμό της θέσης του σε διάδοχικές χρονικές στιγμές.

1.3.3 Εφαρμογές της Παρακολούθησης Κινούμενων Αντικειμένων

Οι εφαρμογές των αλγορίθμων που έχουν αναπτυχθεί για την παρακολούθηση κινούμενων αντικειμένων συναντιούνται σε πολλούς τομείς. Για παράδειγμα στο [28], αναφέρεται ότι το πρόβλημα του εντοπισμού και της παρακολούθησης κινούμενων αντικειμένων έχει πολλές εφαρμογές πάνω σε κωδικοποίηση, παρακολούθηση με κάμερες, επαυξημένη πραγματικότητα (augmented reality) και ρομποτική. Επίσης, μπορεί να δημιουργήσει τα δεδομένα για υψηλότερου επιπέδου εφαρμογές Όρασης, όπως 3Δ-ανακατασκευή και

3Δ-αναπαράσταση. Στο [22], οι εφαρμογές ταξινομούνται σε κάποιες γενικές κατηγορίες, οι οποίες είναι:

- **Οδήγηση και Έλεγχος Οχημάτων:** Στο [33] περιγράφεται μια μέθοδος που επιτρέπει σε ένα αυτόνομο ελικόπτερο τον εντοπισμό του ελικοδρομίου μέσω ενός συστήματος Όρασης, και την ομαλή προσγείωση του εκεί. Τέτοιου είδους εφαρμογές μπορούν να επιτρέψουν την πρόσβαση του ανθρώπου σε περιοχές επικίνδυνες και μη προσβάσιμες. Επίσης, αφαιρώντας την αναγκαιότητα ύπαρξης του ανθρώπου-οδηγού είναι δυνατή μεγάλη μείωση των διαστάσεων του οχήματος αλλά και του κόστους κατασκευής, ενώ δεν πρέπει να αμελήσουμε ότι δεν τίθεται σε κίνδυνο η ζωή του πιλότου (βλ. Εικόνα 1.1, Εικόνα 1.2)
- **Παρακολούθηση και Αναγνώριση:** Εφαρμογές σε αυτόν τον τομέα έχουν σχέση με την παρακολούθηση χώρων και την αναγνώριση υπόπτων, και με την αναγνώριση και ταυτοποίηση προσώπων κατά την είσοδο σε εγκαταστάσεις. Επίσης, μπορούν να χρησιμοποιηθούν και για την παρακολούθηση της κίνησης στους αυτοκινητόδρομους. Στο [8], εξάγονται κάποια σημαντικά χαρακτηριστικά των αυτοκινήτων, και παρακολουθούνται στον αυτοκινητόδρομο (βλ. Εικόνα 1.3). Στο [39] περιγράφεται μια στρατιωτική εφαρμογή του προβλήματος. Εικόνες οι οποίες λαμβάνονται μέσω ενός laser radar υπόκεινται σε επεξεργασία και γίνεται αναγνώριση πιθανών στόχων. Το σύστημα αναγνώρισης χρησιμοποιεί πληροφορίες σχετικές με το γεωμετρικό σχήμα και το μέγεθος διαφόρων πιθανών στόχων, προκειμένου να μπορέσει να τους αναγνωρίσει όταν βρίσκονται καμουφλαρισμένοι (βλ. Εικόνα 1.4). Οι εικόνες λαμβάνονται από ένα αεροσκάφος, και το σύστημα είναι ικανό να αναγνωρίσει σωστά τους στόχους, έχοντας χαμηλό ποσοστό λανθασμένων αποφάσεων. Για την κατασκευή του συστήματος χρησιμοποιήθηκαν αλγόριθμοι για την αντίχνευση στόχων, και για την αναγνώριση αυτών.
- **Ρομποτική και Κατασκευές:** Παρόμοια με την πρώτη κατηγορία, είναι δυνατός ο έλεγχος του σερβομηχανισμού των ρομποτικών βραχιόνων με βάση κάποια αντικείμενα που παρακολουθούνται. Επίσης, είναι δυνατόν να εντοπίζονται και να αφαιρούνται οι κατασκευές της γραμμής παραγωγής που είναι ελλατωματικές. Στο [16] γίνεται μια εκτενής αναφορά στις εφαρμογές ελέγχου ρομποτικών βραχιόνων με βάση οπτική πληροφορία. Πιο συγκεκριμένα, γίνεται κατάταξη των συστημάτων που χρησιμοποιούν τέτοιου είδους πληροφορία, και στη συνέχεια γίνεται μια προσπάθεια αξιολόγησης και σύγκρισης αυτών. Συνοπτικά, αναφέρουμε ότι είναι δυνατόν να συλλέγονται πληροφορίες από κάμερες οι οποίες είναι τοποθετημένες είτε πάνω στην αρπάγη του ρομπότ (και επομένως

υπάρχει μια σταθερή σχέση μεταξύ των συντεταγμένων της κάμερας και της ρομποτικής αρπάγης) είτε από κάμερες οι οποίες είναι σταθερές στο περιβάλλον (και επομένως υπάρχει μια σταθερή σχέση μεταξύ των συντεταγμένων της κάμερας και της βάσης του ρομποτικού συστήματος) (βλ. Εικόνα 1.5). Από τις εικόνες που συλλέγουν οι κάμερες είναι δυνατόν να εξαχθούν χαρακτηριστικά των αντικειμένων (όπως ακμές), ή και συγκεκριμένες περιοχές προς παρακολούθηση. Οι πληροφορίες αυτές χρησιμοποιούνται για την καθοδήγηση της ρομποτικής αρπάγης μέσω ενός συστήματος αυτομάτου ελέγχου. Το σύστημα οπτικής καθοδήγησης μπορεί να είναι είτε ανοικτού βρόχου, είτε κλειστού βρόχου. Στην πρώτη περίπτωση υπάρχει μεγαλύτερη ευκολία στην υλοποίηση, ωστόσο τα λάθη είναι πιο σημαντικά, ενώ στη δεύτερη περίπτωση αν και η συμπεριφορά είναι καλύτερη, απαιτείται πιο εξειδικευμένο υλισμικό και λογισμικό για τον έλεγχο. Φυσικά, ανάλογες τεχνικές χρησιμοποιούνται και για την καθοδήγηση κινούμενων ρομπότ μέσα σε ένα περιβάλλον στο οποίο υπάρχουν κάποια αναγνωρίσιμα σημάδια (*landmarks*).

- Συστήματα Διαπροσωπείας: Όπως στα [36], [37], [23] είναι δυνατόν να χρησιμοποιηθεί τεχνολογία παρακολούθησης αντικειμένων ώστε να γίνει πιο εύκολη η αλληλεπίδραση του ανθρώπου με τη μηχανή. Συγκεκριμένα στα [36], [37] γίνεται παρακολούθηση ανθρώπινων χειρονομιών (βλ. Εικόνα 1.6) και ερμηνεία αυτών με βάση ένα υποσύνολο της νοηματικής γλώσσας, το οποίο ωστόσο έχει μια πολύ αυστηρή και συγκεκριμένη δομή. Τέτοιου είδους εφαρμογές μπορούν να χρησιμοποιηθούν μαζί με εφαρμογές σύνθεσης φωνής, αποτελώντας ένα σημαντικό βοήθημα για τα άτομα με αναπηρίες. Στο [23] γίνεται παρακολούθηση των χειρονομιών του χρήστη για χειρισμό ενός προγράμματος. Πιο συγκεκριμένα, το αριστερό χέρι του χρήστη επιλέγει εντολές από ένα οπτικό μενού, και το δεξί του χέρι επιλέγει τα αντικείμενα στα οποία θα εφαρμοστούν οι εντολές (βλ. Εικόνα 1.7). Όπως αναφέρεται στο [11], κινήσεις του κεφαλιού, των χεριών και των ματιών, μπορούν στο μέλλον να χρησιμοποιούνται για την αλληλεπίδραση με κάθε είδους ηλεκτρονικές συσκευές. Σύμφωνα με μια έρευνα Γάλλων επιστημόνων, ο άνθρωπος μπορεί να εκπληρώσει περίπου μισή φορά πιο γρήγορα κάποια εργασία στον υπολογιστή, αν αντί για το ποντίκι και το πληκτρολόγιο κατευθύνει τον κέρσορα με τα μάτια του. Ωστόσο, απαιτείται κατασκευή ειδικών συστημάτων ώστε να επιτυγχάνεται ταχύτητα και ακρίβεια. Η κατασκευή ενός συστήματος παρακολούθησης ανθρώπινων χειρονομιών περιγράφεται στο [25] (βλ. Εικόνα 1.8).
- Ιατρικές εφαρμογές: Τα συστήματα παρακολούθησης κινούμενων αντι-

κειμένων μπορούν να βοηθήσουν τους χειρουργούς, ειδικά σε συνδυασμό με τα ήδη υπάρχοντα μηχανήματα τηλε-ιατρικής. Ειδικά στον τομέα της μικρο-χειρουργικής μπορούν να οδηγήσουν σε σημαντική πρόοδο. Όπως είναι αναμενόμενο, σε αυτόν τον τομέα απαιτούνται κινήσεις μεγάλης ακρίβειας. Για παράδειγμα, σε πολλές επεμβάσεις απαιτείται ακρίβεια 10 μm , ενώ η πλειοψηφία των ικανών μικρο-χειρουργών μπορεί να έχει ακρίβεια κινήσεων γύρω στα 60 μm , και μάλιστα για περιορισμένο χρονικό διάστημα. Έτσι, είναι απαραίτητη η κατασκευή ρομποτικών χειριστών και συστημάτων που να μπορούν να παρακολουθήσουν τις κινήσεις με ακρίβεια 10 μm . Στο [15] των Hotraphinyo και Riviere, περιγράφεται η κατασκευή ενός συστήματος που να παρακολουθεί την άκρη ενός μικρο-χειρουργικού εργαλείου. Δεδομένου ότι τα μικρο-χειρουργικά εργαλεία πρέπει να είναι πολύ ελαφριά, πρέπει και ο συγκεκριμένος αισθητήρας να παρέχει τις πληροφορίες για την ακριβή παρακολούθηση, αλλά χωρίς να προσθέτει βάρος στο εργαλείο. Επίσης, θα πρέπει να παρέχει μετρήσεις μεγάλης ακρίβειας. Το συγκεκριμένο σύστημα μπορεί να χρησιμοποιηθεί για την αξιολόγηση και την εκπαίδευση των μικρο-χειρουργών, ή και σαν μονάδα εισόδου σε προσομοιωτές μεγάλης ακρίβειας. Συνοπτικά, στο [15], γίνεται φωτισμός του χώρου εργασίας με υπέρυθη ακτινοβολία, και μια μικρή μπάλα από ανακλαστικό υλικό στην άκρη του εργαλείου παρακολουθείται μέσω της αντανάκλασης της ακτινοβολίας.

Όπως είναι εμφανές, το πρόβλημα της παρακολούθησης κινούμενων αντικειμένων έχει ευρείες και πολύ χρήσιμες εφαρμογές. Δυστυχώς, δεν είναι δυνατόν να επιλυθεί εύκολα. Συγκεκριμένα, όπως αναφέρεται στο [40], υπάρχουν ήδη πολλά συστήματα παρακολούθησης κίνησης τα οποία λειτουργούν υπό τη βάση διαφορετικών αρχών. Τα συστήματα αυτά είναι ειδικά σχεδιασμένα για διαφορετικές εφαρμογές και παρουσιάζουν διαφορετικές επιδόσεις μεταξύ τους. Δεν υπάρχει δηλαδή κάποιο ιδανικό σύστημα παρακολούθησης κίνησης, το οποίο οι χρήστες να μπορούν να χρησιμοποιούν σε οποιοδήποτε εφαρμογές, αλλά κάθε φορά, το σύστημα κατασκευάζεται έτσι ώστε να ικανοποιεί κάποια συγκεκριμένα κριτήρια. Με τον όρο ιδανικό, αντιλαμβανόμαστε ένα σύστημα το οποίο να είναι:

- Μικρό: Με μέγεθος μιας 8-DIP ψηφίδας.
- Αυτάρκες: Για τη λειτουργία του να μην απαιτείται να τοποθετηθούν άλλα συστήματα στο περιβάλλον ή στο χρήστη.
- Πλήρες: Να μπορεί να παρακολουθήσει και τους έξι βαθμούς ελευθερίας της κίνησης.

- Ακριβές: Να έχει ανάλυση καλύτερη από 1 mm στη θέση, και 0.1 μοίρες στον προσανατολισμό.
- Γρήγορο: Να λειτουργεί σε συχνότητα 1.000 Hz και με καθυστέρηση μικρότερη από 1ms, ανεξάρτητα από το πόσα τέτοια συστήματα χρησιμοποιούνται ταυτόχρονα.
- Ανεπηρέαστο από επικαλύψεις: Να μην απαιτεί απευθείας οπτική επαφή με το αντικείμενο, και να μπορεί να ξεχωρίσει τα επικαλυπτόμενα αντικείμενα μεταξύ τους.
- Σταθερής απόδοσης: Να παρουσιάζει σταθερή απόδοση ανεξάρτητα από μεταβολές στον φωτισμό, στον ήχο, στη θερμοκρασία, στα μαγνητικά πεδία, στα ραδιοκύματα και την παρουσία άλλων παρόμοιων συστημάτων στο περιβάλλον.
- Ανθεκτικό: Να παρουσιάζει σταθερή απόδοση ανεξάρτητα από το πόσο γρήγορα ή πόσο μακριά κινούνται τα αντικείμενα.
- Ασύρματο: Να μην απαιτεί καλώδια για τη σύνδεση με το περιβάλλον, και να λειτουργεί για μεγάλο χρονικό διάστημα με μικρή μπαταρία.
- Φθινό: Να έχει χαμηλό κόστος, έτσι ώστε να είναι δυνατή η μαζική αγορά και χρήση του.

Όπως φαίνεται από τα κριτήρια που διατυπώθηκαν, το πρόβλημα της παρακολούθησης κινούμενων αντικειμένων δεν μπορεί να αντιμετωπίζεται ανεξάρτητα από την κατασκευή του συστήματος, παρά το γεγονός ότι στη διπλωματική αυτή εμείς θα ασχοληθούμε μόνο με την αλγοριθμική προσέγγιση. Πάντως έως τώρα, δεν υπάρχει σύστημα που να έχει την επιθυμητή συμπεριφορά σε όλα τα παραπάνω σημεία. Σύμφωνα με το [40], κάθε σύστημα που έχει κατασκευαστεί, παρουσιάζει ελλειμματική συμπεριφορά στα επτά από τα δέκα παραπάνω σημεία, και δεν αναμένεται αισθητή βελτίωση στο βραχύ μέλλον.

1.4 Εικόνες Κεφαλαίου 1

Στην ενότητα αυτή παρατίθενται οι εικόνες του πρώτου κεφαλαίου, μαζί με τα κατάλληλα συνοδευτικά σχόλια-λεζάντες.



(a)

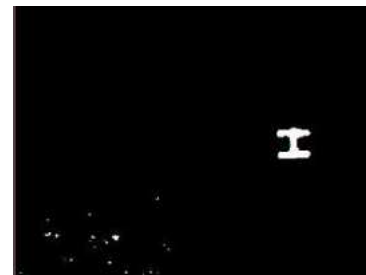


(b)

Εικόνα 1.1: Το ελικόπτερο του [33] (a) Εν πτήσει, (b) Κατά την προσεδάφιση. Από το [33].



(a)



(b)

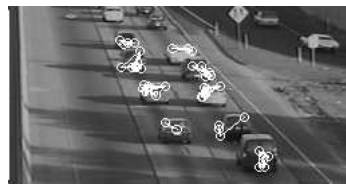
Εικόνα 1.2: (a) Το ελικοδρόμιο για το ελικόπτερο του [33], (b) Το τελικό αποτέλεσμα της κατάτμησης. Από το [33].



(a)

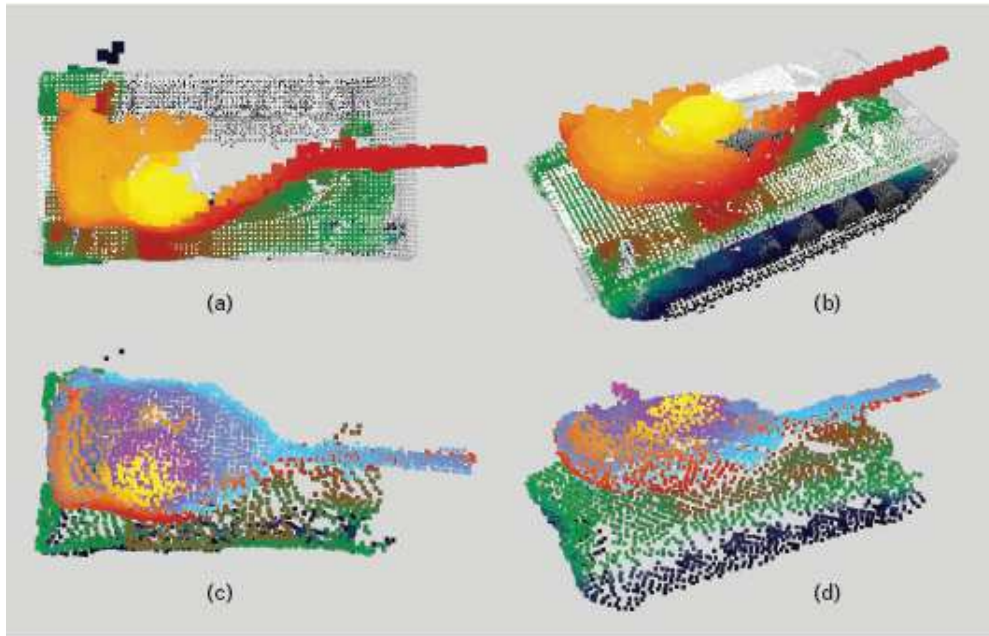


(b)

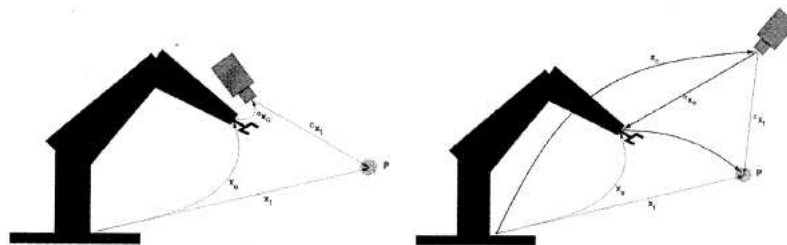


(c)

Εικόνα 1.3: Η παρακολούθηση του αυτοκινητόδρομου στο [8]. (a) Τα χαρακτηριστικά που εξάγονται από τα αυτοκίνητα, (b) Η παρακολούθηση των χαρακτηριστικών των αυτοκινήτων, (c) Ο συσχετισμός των χαρακτηριστικών για τον προσδιορισμό του οχήματος. Από το [8].



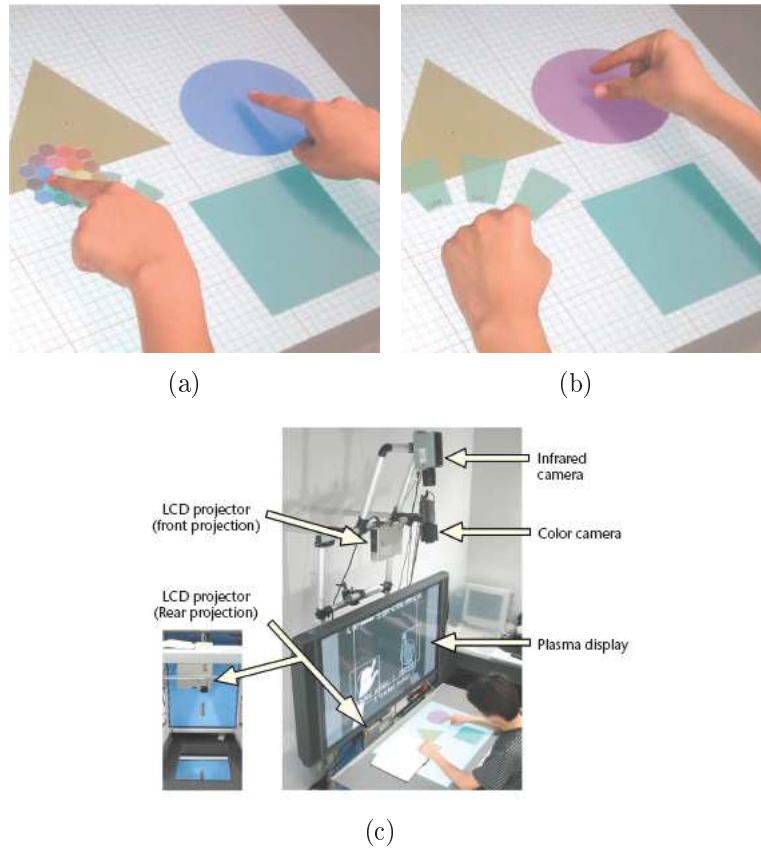
Εικόνα 1.4: Αναγνώριση ενός άρματος M60 από τα κομμάτια του. (a) Κατακόρυφη προβολή του κυρίως σώματος του M60. Τα σημεία του σκηηνικού είναι χρωματικά κωδικοποιημένα με βάση το ύψος (πράσινο-κόκκινο-κίτρινο), ενώ τα σημεία του M60 είναι κωδικοποιημένα σε κλίμακα μαύρο-άσπρο. (b) Μια άλλη όψη του M60, η οποία δείχνει ότι η αναγνώριση του κυρίως σώματος έγινε επιτυχημένα και για τους 6 βαθμούς ελευθερίας. (c) Κατακόρυφη προβολή του κανονικού του M60. Η χρωματική κωδικοποίηση είναι παρόμοια με το (a). (d) Μια άλλη όψη του M60, η οποία δείχνει ότι η αναγνώριση του κανονικού έγινε επιτυχημένα και για του 6 βαθμούς ελευθερίας. Από το [39].



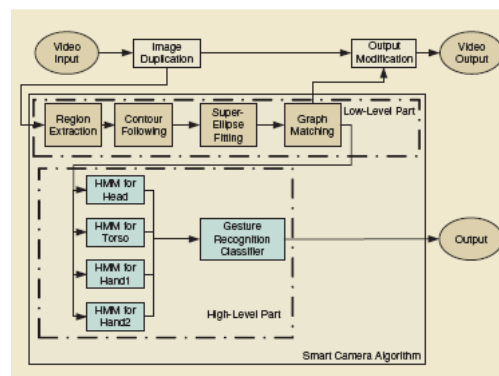
Εικόνα 1.5: Οι δύο δυνατές τοποθετήσεις της κάμερας σύμφωνα με το [16] Από το [16].



Εικόνα 1.6: Το σύστημα παρακολούθησης χειρονομιών για αναγνώριση της νοηματικής γλώσσας, που αναπτύχθηκε στο Georgia Tech. (a) Εικόνα από την κάμερα που βρίσκεται στο καπέλο, (b) Εικόνα από την κάμερα που βρίσκεται στον πάγκο εργασίας. Από το [37].



Εικόνα 1.7: (a), (b) Το interface ζωγραφικής του συστήματος που αναπτύχθηκε στο [23], (c) Το Enhanced Desk, ένα σύστημα διαπροσωπείας που εκτελεί παρακολούθηση και αναγνώριση ανθρώπινων χειρονομιών, επιτρέπει στους χρήστες να χειριστούν αληθινά και εικονικά αντικείμενα. Από το [23].



Εικόνα 1.8: Το σύστημα παρακολούθησης χειρονομιών που αναπτύχθηκε στο [25].

Κεφάλαιο 2

Εύρεση Αντικειμένων σε Εικόνες

2.1 Εισαγωγή

Στο κεφάλαιο αυτό θα παρουσιάσουμε αναλυτικά τους τρόπους με τους οποίους είναι δυνατόν να εντοπίσουμε αντικείμενα σε μια εικόνα. Αρχικά θα αναφερθούμε σε μεθόδους οι οποίες αναζητούν και βρίσκουν οποιαδήποτε αντικείμενα σε μια εικόνα, όπως οι μέθοδοι των Ενεργών Περιγραμμάτων (Active Contours) και Γεωδαιτικών Ενεργών Περιγραμμάτων (Geodesic Active Contours). Στη συνέχεια θα παρουσιάσουμε μεθόδους με τις οποίες είναι δυνατόν να βρούμε ένα επιθυμητό αντικείμενο στην εικόνα με χρήση Προγενέστερων Σχημάτων (Prior Shapes). Οι μέθοδοι στις οποίες βασίζεται ο κώδικας που αναπτύξαμε παρουσιάζονται εκτενώς, και γίνεται σύγκριση μεταξύ τους, παρουσιάζοντας πλεονεκτήματα και μειονεκτήματα. Τέλος, γίνεται μια σύντομη παρουσίαση κάποιων άλλων μεθόδων κατάτμησης με βάση προγενέστερα σχήματα, οι οποίες είναι διαφορετικής φιλοσοφίας αλλά έχουν πολύ ενδιαφέροντα αποτελέσματα. Στο τέταρτο κεφάλαιο παρουσιάζονται προσομοιώσεις των μεθόδων που υλοποιήσαμε.

2.2 Εύρεση Αντικειμένων σε Εικόνα

2.2.1 Μέθοδος των Ενεργών Περιγραμμάτων

Τα ενεργά περιγράμματα (active contours) ή φίδια (snakes), εισήχθησαν από τους Kass, Witkin και Terzopoulos το 1987 ([17]), και αποτέλεσαν μια ανατρεπτική μέθοδο για τον εντοπισμό αντικειμένων σε εικόνες. Το μοντέλο των ενεργών περιγραμμάτων βασίζεται στην ελαχιστοποίηση ενός συναρτησιακού ενέργειας. Αρχικοποιώντας μια καμπύλη επί της εικόνας είναι δυνατόν, με κατάλληλη κατασκευή του συναρτησιακού, να οδηγήσουμε την καμπύλη έτσι

ώστε να εντοπίσει ένα αντικείμενο. Η ιδέα είναι παρόμοια με την ιδέα πίσω από το μοντέλο των παραμορφώσιμων μοντέλων. Σύμφωνα με αυτό, έχοντας την παραμετρική αναπαράσταση ενός σχήματος-μοντέλου, είναι δυνατόν να βρούμε τις κατάλληλες τιμές των παραμέτρων, αν προσπαθήσουμε να τις μεταβάλουμε με κριτήριο την ελαχιστοποίηση κάποιας συνάρτησης ενέργειας. Όπως σε αυτή την περίπτωση, έτσι και στην περίπτωση των ενεργών περιγραμμάτων, αρχικά απαιτείται η παραμετροποίηση της καμπύλης. Στη συνέχεια, κατασκευάζεται το συναρτησιακό ενέργειας, που γενικά μπορεί να περιλαμβάνει όρους που εξαρτώνται από την καμπυλότητα, όρους που εξαρτώνται από τις ακμές των αντικειμένων στην εικόνα, στατιστικούς όρους που έχουν σχέση με υφή, φωτεινότητα, χρώμα ή κίνηση και τέλος, από όρους που σχετίζονται με πρότερη γνώση σχετικά με το αντικείμενο, τη θέση του κ.τ.λ (από το [1]).

Η τυπική παραμετροποίηση του εξελισσόμενου περιγράμματος, είναι σαν μια συνεχή καμπύλη μιας παραμέτρου: $C \equiv [x(q), y(q)] : [0, 1] \rightarrow \mathbb{R}^2$, όπου q είναι η παράμετρος της καμπύλης και παίρνει τιμές στο $[0, 1]$. Η εικόνα θεωρείται ότι είναι μια απεικόνιση από ένα πλέγμα δύο διαστάσεων στο σύνολο των μη αρνητικών αριθμών, δηλαδή $I : [0, a] \times [0, b] \rightarrow \mathbb{R}^+$. Η ελαχιστοποίηση του συναρτησιακού ενέργειας γίνεται μέσω της εξέλιξης της καμπύλης με βάση τη διαφορική εξίσωση εξέλιξης:

$$\frac{\partial \vec{C}(q, t)}{\partial t} = V(q, t) \vec{N}_o(q, t)$$

Παρατηρούμε την εισαγωγή της παραμέτρου του χρόνου, η οποία αντανακλά την έννοια της εξέλιξης της καμπύλης. Ο όρος V είναι η ταχύτητα εξέλιξης, ενώ το διάνυσμα \vec{N}_o είναι το κάθετο μοναδιαίο διάνυσμα προς το εξωτερικό της καμπύλης. Μπορεί να αποδειχθεί ότι η μόνη ενδιαφέρουσα κατεύθυνση εξέλιξης της καμπύλης, είναι κατά μήκος του εξωτερικού κάθετου διανύσματος. Ο όρος της ταχύτητας προκύπτει από την ελαχιστοποίηση του συναρτησιακού ενέργειας. Στην περίπτωση των φιδιών, το συναρτησιακό εξαρτάται από την ομαλότητα, την εγγύτητα ως προς τις ακμές και από κάποιους εξωτερικούς περιορισμούς. Η ενέργεια της καμπύλης σε αυτή την περίπτωση, δίνεται από τη σχέση:

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

Υποθέτοντας ότι δεν υπάρχουν εξωτερικοί περιορισμοί, το συναρτησιακό γράφεται:

$$E_{snake}[\vec{C}] = \int_0^1 \underbrace{(\alpha \|\vec{C}_q\|^2 + \beta \|\vec{C}_{qq}\|^2)}_{\text{όρος ομαλότητας}} dq - \lambda \int_0^1 \underbrace{\|\nabla I(\vec{C})\|^2}_{\text{όρος έλξης προς ακμές}} dq \quad (2.1)$$

Οι δύο πρώτοι όροι είναι η εσωτερική ενέργεια που οδηγεί στην ομαλότητα της καμπύλης. Πιο συγκεκριμένα, σύμφωνα με το [17], ο πρώτος όρος κάνει την καμπύλη να συμπεριφέρεται σαν μεμβράνη, ενώ ο δεύτερος όρος σαν λεπτή πλάκα. Ο τελευταίος όρος έλκει την καμπύλη προς τις ακμές και δίνει την εξωτερική ενέργεια. Οι θετικές παράμετροι α , β , λ , προσδιορίζουν την σημαντικότητα του κάθε όρου στην συνολική ενέργεια της καμπύλης. Η καμπύλη που εξελίσσεται κατ' αυτόν τον τρόπο ονομάστηκε φίδι, λόγω της κίνησης που κάνει πάνω στην εικόνα κατά την ελαχιστοποίηση της ενέργειάς της.

Τελικά, η εξέλιξη της καμπύλης την οδηγεί σε θέση τοπικού ελαχίστου ενέργειας. Έτσι, σε μεγάλο βαθμό, το τελικό αποτέλεσμα εξαρτάται από την αρχική τοποθέτηση του φιδιού. Στο [17] προβλέπονταν η δυνατότητα αρχικής τοποθέτησης της καμπύλης από το χρήστη, και στη συνέχεια η αυτόματη εξέλιξή της, οπότε συνολικά, η εξέλιξη του ενεργού περιγράμματος ήταν μια ημιαυτόματη διαδικασία. Για την υπερπήδηση του προβλήματος της ανεπιθύμητης σύγκλισης σε τοπικό ελάχιστο, η καμπύλη μπορεί να εξελιχθεί σε διάφορους χώρους κλίμακας. Μόλις συγκλίνει σε μια πολύ θολωμένη εικόνα, μπορεί το αποτέλεσμα να χρησιμοποιηθεί ως αρχικοποίηση στην εικόνα με τη μεγαλύτερη λεπτομέρεια (coarse-to-fine approach). Έτσι, τελικά θα συγκλίνει στο επιθυμητό ελάχιστο.

Τα φίδια, όπως αναφέραμε, ήταν πρωτοποριακά για την εποχή τους. Το μεγάλο τους μειονέκτημα ωστόσο, ήταν η εξάρτηση από την παραμετροποίηση της καμπύλης. Επίσης, απαιτούνταν να κρατιέται πληροφορία σχετικά με την τοπολογία της καμπύλης, σχετικά δηλαδή με τη συνεκτικότητά των τμημάτων της. Δεν ήταν δυνατόν το αρχικό μοντέλο να χειριστεί αυτόματα αλλαγές στην τοπολογία της καμπύλης (διασπάσεις και ενώσεις), κάτι που είναι πολύ αρνητικό ιδίως στην περίπτωση που δεν γνωρίζουμε τον αριθμό των αντικειμένων που αναζητούμε στην εικόνα.

Η βασική βελτίωση του μοντέλου των φιδιών έτσι ώστε να επιτρέπονται αλλαγές στην τοπολογία και να υπάρχει ανεξαρτησία από την παραμετροποίηση, έγινε από τους Caselles, Kimmel και Sapiro, στο [6].

2.2.2 Μέθοδος των Γεωδαιτικών Ενεργών Περιγραμμάτων

Προκειμένου να αντιμετωπίσουν τα μειονεκτήματα του μοντέλου των φιδιών, οι Caselles, Kimmel και Sapiro πρότειναν με το [6] το 1997, το μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων.

Για να επιτευχθεί ανεξαρτησία από την παραμετροποίηση ήταν απαραίτητο οι παράμετροι της καμπύλης να εξαρτώνται μόνο από τα γεωμετρικά χαρακτηριστικά αυτής, όπως για παράδειγμα την καμπυλότητα και το κάθετο διάνυσμα. Θεωρώντας ότι ο πρώτος όρος της εξίσωσης (2.1) αρκεί για να εξασφαλιστεί η ομαλότητα της καμπύλης, οι Caselles, Kimmel και Sapiro, απαίτησαν την

ελαχιστοποίηση του συναρτησιακού:

$$E_{snake}[\vec{C}] = \alpha \int_0^1 \|\vec{C}_q\|^2 dq - \lambda \int_0^1 \|\nabla I(\vec{C})\|^2 dq$$

Δεδομένου ότι ο όρος $\|\nabla I(\vec{C})\|$ είναι αυτός που εξαρτάται από τις ακμές του αντικειμένου και οδηγεί την καμπύλη προς τα όρια των αντικειμένων, πρότειναν την αντικατάστασή του από μια γενική συνάρτηση εύρεσης ακμών. Συγκεκριμένα, η γενικευμένη συνάρτηση που πρότειναν ήταν μια $g : [0, +\infty] \rightarrow \mathbb{R}^+$, φθίνουσα ώστε $g(r) \rightarrow 0$ καθώς $r \rightarrow +\infty$. Πιθανές τέτοιες συναρτήσεις είναι οι $g = g(\|\nabla I\|) = \frac{1}{1+\|\nabla G_{\sigma^* I}\|^n}$, $n = 1$ ή 2 , και $g = g(\|\nabla I\|) = \exp\left(-\frac{\|\nabla G_{\sigma^* I}\|}{s}\right)$.

Έτσι, με αντικατάσταση του τελευταίου όρου $-\|\nabla I(C(q))\|$ από τον όρο $g(\|\nabla I(C(q))\|)$ το συναρτησιακό ενέργειας γίνεται:

$$E(C) = \alpha \int_0^1 \|C'(q)\|^2 dq + \lambda \int_0^1 g(\|\nabla I(C(q))\|)^2 dq \quad (2.2)$$

Η τελευταία εξίσωση περιέχει έναν όρο εσωτερικής ενέργειας που διασφαλίζει την ομαλότητα της καμπύλης και έναν όρο εξωτερικής ενέργειας που διασφαλίζει την κίνηση της καμπύλης προς τις ακμές. Μας ενδιαφέρει μόνο η τιμή του λόγου $\frac{\lambda}{\alpha}$. Αν και εξακολουθεί να υπάρχει εξάρτηση από την παραμετροποίηση, στο [6] αποδεικνύεται ότι ελαχιστοποίηση του συναρτησιακού (2.2) ισοδυναμεί με ελαχιστοποίηση του:

$$E(C) = \int_0^1 g(\|\nabla I(C(q))\|) \|C'(q)\| dq \quad (2.3)$$

Δεδομένου ότι $\|C'(q)\|dq = ds$, όπου s είναι παράμετρος τόξου, προκύπτει:

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

Στην τελευταία σχέση παρατηρούμε την φυσική παραμετροποίηση της καμπύλης C , με βάση το μήκος του τόξου της καμπύλης. Επομένως, η καμπύλη ορίστηκε με βάση γεωμετρικά χαρακτηριστικά της, ξεπερνώντας τον πρώτο βασικό περιορισμό των φιδιών.

Συγκρίνοντας την τελευταία εξίσωση με τον κλασικό ορισμό του μήκους μιας καμπύλης, βλέπουμε ότι η ενέργεια προσδιορίζεται από το μήκος της καμπύλης με κατάλληλο βάρος. Το βάρος σε κάθε σημείο δίνεται από την $g(\|\nabla I(C(q))\|)$, εξαρτάται δηλαδή από την θέση των ορίων των αντικειμένων, των ακμών. Όπως καταλήγουν στο [6] "όταν προσπαθούμε να εντοπίσουμε ένα αντικείμενο, δεν μας ενδιαφέρει να ελαχιστοποιήσουμε το μήκος της

καμπύλης με την κλασσική έννοια, αλλά να ελαχιστοποιήσουμε ένα μήκος ορισμένο με τέτοιο τρόπο, ώστε να λαμβάνει υπόψη του τα χαρακτηριστικά της εικόνας”.

Η ελαχιστοποίηση της ενέργειας επιτυγχάνεται με τη μέθοδο κατάβασης κλίσης (gradient descent). Στο [6], βρίσκεται η Euler-Lagrange παράγωγος της (2.3), και η σχέση μεταβολής της καμπύλης είναι η:

$$\frac{\partial C(t)}{\partial t} = g(I) \kappa \vec{N}_i - (\nabla g \cdot \vec{N}_i) \cdot \vec{N}_i \quad (2.4)$$

όπου κ είναι η καμπυλότητα της καμπύλης και \vec{N}_i το κάθετο μοναδιαίο διάνυσμα προς το εσωτερικό της καμπύλης. Η εξίσωση (2.4) δείχνει πώς εξελίσσεται χρονικά το κάθε σημείο της καμπύλης. Για την αποφυγή του προβλήματος σύγκλισης σε τοπικά ελάχιστα είναι δυνατόν, όπως και στην περίπτωση των φιδιών, να γίνει εξέλιξη της καμπύλης σε διάφορους χώρους κλίμακας.

Συχνά στην εξίσωση (2.4) προστίθεται και ένας τρίτος όρος, προκειμένου να επιταχύνει την εξέλιξη της καμπύλης, ενώ βοηθάει την καμπύλη να αποφύγει τα τοπικά ελάχιστα. Ο όρος αυτός ονομάζεται δύναμη μπαλονιού. Ανάλογα με το πρόσημό του και την κατεύθυνση εξέλιξης της καμπύλης, μπορεί να ευνοεί την ελαχιστοποίηση της επιφάνειας της καμπύλης και να οδηγεί στην συρρίκνωσή της, ή να οδηγεί στην αύξηση της επιφάνειας που καλύπτεται από την καμπύλη, δηλαδή στην διαστολή της καμπύλης. Τελικά, η εξέλιξη καθορίζεται από την επίδραση τριών δυνάμεων: της δύναμης μπαλονιού (balloon force), της δύναμης καμπυλότητας (curvature force) και της δύναμης ελατηρίου (spring force), σύμφωνα με τη σχέση:

$$\frac{\partial C(t)}{\partial t} = (-g\beta + g\kappa - \nabla g \cdot \vec{N}_i) \vec{N}_i \quad (2.5)$$

Στον πρώτο όρο οφείλεται η επιτάχυνση της σύγκλισης, στο δεύτερο όρο η διατήρηση της ομαλότητας, και στον τρίτο, η τοπική έλξη προς τις ακμές της εικόνας. Η επίδραση των παραγόντων εξαρτάται από την επιλογή της συνάρτησης τερματισμού g , και η κατάλληλη επιλογή δεν αποτελεί απλό πρόβλημα. Αν λάβουμε υπόψη ότι η g εισάγει στο συναρτησιακό ενέργειας την πληροφορία για τις ακμές της εικόνας, αντιλαμβανόμαστε ότι η επιλογή της αντιμετωπίζει τα προβλήματα που αντιμετωπίζει ο εντοπισμός ακμών με βάση την κλίση (gradient) των εικόνων. Σε ιδανικές εικόνες, οποιαδήποτε συνάρτηση που πληρεί τις ιδιότητες $g : [0, +\infty] \rightarrow \mathbb{R}^+$, με $g(r) \rightarrow 0$ καθώς $r \rightarrow \infty$, θα ήταν κατάλληλη. Για τις φυσικές εικόνες ωστόσο, πρέπει να αντιμετωπιστούν προβλήματα όπως οι διακυμάνσεις στην τιμή της κλίσης, ρωγμές στις ακμές κ.τ.λ. Η εύρεση κατάλληλης g που να ικανοποιεί και αυτές τις περιπτώσεις, αποτέλεσε από μόνο του ένα καινούριο ερευνητικό πρόβλημα.

Το μοντέλο είναι ικανό να εντοπίσει τόσο τις εσωτερικές ακμές των αντικειμένων, όσο και τις εξωτερικές ακμές αυτών. Στην πρώτη περίπτωση η καμπύλη αρχικοποιείται ως μικρός κύκλος στο εσωτερικό κάθε αντικειμένου, και στη δεύτερη περίπτωση ως περικλύουσα το επιθυμητό αντικείμενο. Ωστόσο, δεν μπορεί να χειριστεί αυτόματα τις τοπολογικές αλλαγές της καμπύλης, και για κάθε αντικείμενο που πρέπει να εντοπίσουμε θα πρέπει να αρχικοποιούμε και μια διαφορετική καμπύλη στο εσωτερικό του ή στο εξωτερικό του. Προκειμένου να χρησιμοποιήσουμε αλγοριθμικά τις σχέσεις (2.4) και (2.5), είναι απαραίτητο να γίνεται δειγματοληψία της καμπύλης ανά κάποια τακτά χωρικά διαστήματα, και να εφαρμόζεται η διαφορική εξίσωση εξέλιξης για κάθε σημείο ξεχωριστά. Ωστόσο, αυτή η μέθοδος απαιτεί συχνή επαναδειγματοληψία, δεδομένου ότι τα σημεία έχουν την τάση να διασκορπίζονται ή να συγκεντρώνονται. Έτσι, είναι απαραίτητη η “συνένωση” σημείων μεταξύ τους, ή και η εισαγωγή νέων σημείων, και αυτό δεν είναι δυνατό να γίνει με αυτόματο τρόπο. Σύμπερασματικά, οι γεωδαιτικές ενεργές καμπύλες αντιμετωπίζουν το ίδιο πρόβλημα με τις ενεργές καμπύλες όσον αφορά τη διαχείριση των τοπολογικών αλλαγών.

Στο πρόβλημα της τοπολογίας έδωσαν λύση οι Osher και Sethian στο [24]. Ακολουθώντας τη μέθοδο που περιγράφεται στο [24], το υπάρχον μοντέλο βελτιώνεται και οι καμπύλες μπορούν να εξελίσσονται αδιαφορώντας για την τοπολογία, αρκεί να αντιμετωπίζονται ως το επιπεδοσύνολο μιας κατάλληλα επιλεγμένης επιφάνειας.

2.2.3 Γεωδαιτικά Ενεργά Περιγράμματα και Επιπεδοσύνολα

Οι Osher και Sethian παρουσίασαν το 1988 με τη δημοσίευσή τους [24], ένα μοντέλο το οποίο να μπορεί να χειρίζεται άμεσα αλλαγές στην τοπολογία μιας εξελισσόμενης καμπύλης. Η βασική ιδέα είναι η ενσωμάτωση της εξελισσόμενης καμπύλης σε μια επιφάνεια με τέτοιο τρόπο, ώστε η καμπύλη να αποτελεί ανά πάσα στιγμή μια ισοσταθμική καμπύλη της επιφάνειας. Έτσι, αντί να εξελίσσεται η καμπύλη, εξελίσσεται η επιφάνεια και η καμπύλη προκύπτει κάθε φορά ως τομή της επιφάνειας με το κατάλληλο επίπεδο δυναμικού.

Έστω $C(q,t)$ η καμπύλη που εξελίσσεται, και μια συνάρτηση u η οποία ορίζεται πάνω σε ολόκληρη την εικόνα I . Έστω ότι η u ορίζεται κατά τέτοιο τρόπο ώστε $u : [0, a] \times [0, b] \rightarrow \mathbb{R}$, και τη χρονική στιγμή $t = 0$ η C είναι η ισοσταθμική επιφάνεια αυτής για κάποιο τυχαίο αλλά σταθερό δυναμικό λ . Επομένως ισχύει $u(C(0), 0) = \lambda$ και θέλουμε να βρούμε τη διαφορική εξίσωση εξέλιξης της u , έτσι ώστε για τα σημεία Γ για τα οποία ισχύει $u(\Gamma(t), t) = \lambda$, να είναι $\Gamma(t) \equiv C(t) \forall t$. Παρατηρούμε ότι με αυτή την προσέγγιση, δεν απαιτείται καμία παραμετροποίηση. Επίσης, οι τοπολογικές αλλαγές της καμπύλης-επιπεδοσυνόλου C , δεν συνεπάγονται ασυνέχεια ή αλλαγή στην τοπολογία

της επιπεδοσυνάρτησης u , παρά μόνο απλή εξέλιξή της. Έτσι, στο καινούριο μοντέλο, οι τοπολογικές αλλαγές αντιμετωπίζονται αυτόματα¹. Αναζητούμε επομένως τη διαφορική εξίσωση εξέλιξης της u , ώστε να μπορούμε να χρησιμοποιούμε και να εξελίσσουμε αυτή, αντί να χρησιμοποιούμε την καμπύλη C . Φυσικά, αρχίζει και γίνεται αντιληπτό ότι εφόσον για τον προσδιορισμό της εξέλιξης μιας καμπύλης απαιτείται η εξέλιξη μιας επιφάνειας (μια διάσταση επιπλέον), η πολυπλοκότητα αυξάνει κατακόρυφα.

Λόγω της σχέσης ανάμεσα στην καμπύλη C και την συνάρτηση επιπεδοσυνόλων u , είναι δυνατόν τα γεωμετρικά στοιχεία της C (η καμπυλότητα και το κάθετο διάνυσμα), να προσδιορίζονται απευθείας από την u . Ισχύει ότι:

$$\vec{N}_i = -\frac{\nabla u}{\|\nabla u\|} \quad (2.6)$$

$$\kappa = -\nabla \vec{N}_i = \operatorname{div}\left(\frac{\nabla u}{\|\nabla u\|}\right) \quad (2.7)$$

Η σχέση $u(\Gamma(t), t) = 0$ (θεωρήσαμε ότι $\lambda = 0$, χ.β.τ.γ) ισχύει για κάθε χρονική στιγμή t . Με παραγωγή της ως προς t προκύπτει ότι:

$$\begin{aligned} \frac{\partial u}{\partial t} + \nabla u \Gamma_t &= 0 \Rightarrow^{2.7} \\ \frac{\partial u}{\partial t} - \Gamma_t \vec{N}_i \|\nabla u\| &= 0 \end{aligned}$$

Γνωρίζουμε ότι $\Gamma_t \vec{N}_i = V$, όπου V είναι η ταχύτητα εξέλιξης προς το εσωτερικό της καμπύλης, οπότε προκύπτει η σχέση:

$$u_t = V \|\nabla u\|$$

Επομένως, η επιφάνειά μας εξελίσσεται σαν ένα πρόβλημα συνοριακών τιμών, από τις σχέσεις:

$$\begin{aligned} u_t &= V \|\nabla u\| \\ u(\vec{x}, 0) &= 0 \text{ για } (x, y) \in \Gamma(0) \end{aligned}$$

Μπορεί να αποδειχθεί ότι η εξέλιξη της C_0 είναι ανεξάρτητη της επιλογής της συνάρτησης u , και ως εκ τούτου η εξέλιξη C_0 μπορεί να βρεθεί αποκλειστικά από την εξελιγμένη u , παίρνοντας το κατάλληλο επιπεδοσύνολο αυτής. Ο μόνος περιορισμός είναι η συνάρτηση u να είναι συνεχής κατά Lipschitz. Οι συναρτήσεις λαμβάνονται έτσι ώστε να είναι $u(\vec{x}, t) < 0$ για σημεία στο

¹Η προσέγγιση αυτή γενικεύεται. Αν θέλουμε για παράδειγμα να χειριζόμαστε αυτόματα τις τοπολογικές αλλαγές σε μια επιφάνεια, αρκεί να την θεωρήσουμε ως ισοσταθμική επιφάνεια σε ένα υπερεπίπεδο 3 διαστάσεων, και να προσδιορίσουμε την εξέλιξη του υπερεπιπέδου.

εσωτερικό της καμπύλης $C(t)$ και $u(\vec{x}, t) > 0$ για σημεία στο εξωτερικό της καμπύλης, ή το αντίστροφο. Οι θεωρήσεις αυτές όμως είναι αυθαίρετες και έχουν σχέση μόνο με τη φορά εξέλιξης της καμπύλης. Μια συνάρτηση που ικανοποιεί αυτούς τους περιορισμούς και χρησιμοποιείται συχνά στην πράξη, είναι η συνάρτηση προσημασμένης ευκλείδειας απόστασης από το εξελισσόμενο σύνορο-καμπύλη.

Συνοψίζοντας, αυτό που κερδίζουμε από την προσέγγιση των επιπεδοσυνόλων είναι η αυτόματη αντιμετώπιση των τοπολογικών αλλαγών της C , ξεπερνώντας έτσι και το δεύτερο πρόβλημα των ενεργών περιγραμμάτων που παρουσιάστηκαν στο 2.2.1.

Έχοντας εξάγει το γενικό μοντέλο εξέλιξης, θα μελετήσουμε την ειδική περίπτωση των γεωδαιτικών ενεργών περιγραμμάτων. Σε αυτή την περίπτωση έχουμε ότι $V = g(I)x - \nabla g \vec{N}_i$, όπως προκύπτει από την εξίσωση (2.4). Έτσι, η διαφορική εξίσωση εξέλιξης της συνάρτησης επιπεδοσυνόλων είναι η:

$$u_t = [g(I)x - \nabla g \vec{N}_i] \|u\| = \left[g(I) \operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right) + \nabla g \frac{\nabla u}{\|\nabla u\|} \right] \|\nabla u\|$$

που απλοποιείται στην:

$$u_t = \operatorname{div} \left(g(I) \frac{\nabla u}{\|\nabla u\|} \right) \|\nabla u\|$$

Αν προστεθεί και η δύναμη μπαλονιού, τότε $V = (-g\beta + gx - \nabla g \vec{N}_i)$, και η διαφορική εξίσωση εξέλιξης είναι η:

$$u_t = \left[\operatorname{div} \left(g(I) \frac{\nabla u}{\|\nabla u\|} \right) - g(I)\beta \right] \|\nabla u\|$$

Ένα μεγάλο πλεονέκτημα των γεωδαιτικών ενεργών καμπυλών είναι η δυνατότητα εξέλιξης της καμπύλης και υπό την επιρροή άλλων δυνάμεων. Έτσι, είναι δυνατόν να εισαχθούν δυνάμεις οι οποίες βασίζονται σε κάποια πρότερη γνώση και να εντοπίσουν συγκεκριμένα επιθυμητά αντικείμενα, δυνάμεις που έχουν σχέση με κίνηση ώστε να εντοπίσουν κινούμενα αντικείμενα, ή άλλες δυνάμεις που να εξαρτώνται και από άλλα χαρακτηριστικά της εικόνας πλην των ακμών.

Για την εξέλιξη των επιφανειών έχουν αναπτυχθεί αρκετοί αλγόριθμοι, με τους βασικότερους και γνωστότερους να είναι οι Ταχείας Προέλασης (Fast Marching, στο [34]) και Στενής Ζώνης (Narrow Band, στο [35]). Ο αλγόριθμος ταχείας προέλασης βασίζεται στα χαρακτηριστικά κάποιων ειδικών περιπτώσεων εξέλιξης, ενώ ο αλγόριθμος στενής ζώνης είναι πιο γενικός. Και οι δύο έχουν ως στόχο την μείωση της πολυπλοκότητας του προβλήματος εξέλιξης της επιφάνειας u , και μια λεπτομερής περιγραφή τους θα δοθεί στο 4ο κεφάλαιο που αφορά το σύστημα παρακολούθησης κινούμενων αντικειμένων που αναπτύξαμε.

2.3 Προγενέστερα Σχήματα και Επιπεδοσύνολα

Οι μέθοδοι που μέχρι τώρα περιγράψαμε δεν παρέχουν καμία πληροφορία για τα αντικείμενα που εντοπίζονται στην εικόνα. Επίσης, ο τρόπος εξέλιξης των καμπυλών τις καθιστά ευαίσθητες στις διακυμάνσεις της κλίσης, στο θόρυβο, στη χαμηλή διαφορά στην αντίθεση της φωτεινότητας, καθώς και σε πιθανά σφάλματα-τοπικές καταστροφές της εικόνας. Έτσι, οι μέθοδοι αυτοί δεν αρκούν σε πολλές περιπτώσεις για την εξαγωγή της χρήσιμης πληροφορίας από μια εικόνα. Για να έχουμε την εξαγωγή της επιθυμητής πληροφορίας, είναι απαραίτητο να αναζητήσουμε τα αντικείμενα που μας ενδιαφέρουν και να μπορέσουμε να τα εντοπίσουμε παρά τις διάφορες ατέλειές τους.

Η κατάτμηση εικόνας, στις περισσότερες περιπτώσεις, βασίζεται σε ένα συνδυασμό αυτόματων και ημιαυτόματων τεχνικών. Απαιτείται δηλαδή από τον άνθρωπο αρκετή προεργασία, προκειμένου να καθοδηγηθεί ο εκάστοτε αλγόριθμος. Η προεργασία αυτή είναι αναγκαία, ιδίως όταν τα αντικείμενα που αναζητούνται δεν έχουν ευκρινείς ακμές, ή όταν περιοχές του ενός επικαλύπτονται με περιοχές του άλλου. Δεδομένου ότι οι μέθοδοι εύρεσης αντικειμένων που περιγράψαμε βασίζονται στις διαφορές φωτεινότητας της εικόνας, αντιλαμβανόμαστε ότι δεν είναι δυνατόν να εντοπιστούν σχηματισμοί των οποίων η φωτεινότητα δεν μεταβάλλεται έντονα.

Τα προβλήματα αυτά πηγάζουν από το γεγονός ότι οι αλγόριθμοι κατάτμησης δεν χρησιμοποιούν πληροφορία για το είδος των αντικειμένων που αναζητούν. Επομένως, πρέπει να αναπτυχθούν τεχνικές όπου η κατάτμηση, πέρα από τις πληροφορίες φωτεινότητας, υψής ή και προσανατολισμού που χρησιμοποιεί, θα κάνει και χρήση κάποιας προγενέστερης γνώσης επί των αντικειμένων που αναζητούμε. Είναι λοιπόν φανερό ότι πρέπει να δημιουργηθεί ένα μοντέλο (προγενέστερο σχήμα) του αντικειμένου που αναζητούμε, και αυτό να χρησιμοποιηθεί κατά τη διαδικασία κατάτμησης. Για τη δημιουργία του μοντέλου πρέπει να χρησιμοποιηθεί ένα σύνολο δεδομένων εκπαίδευσης, δηλαδή ένα σύνολο από κατάλληλες απεικονίσεις του αντικειμένου που μας ενδιαφέρει.

Καταλήγουμε επομένως στο συμπέρασμα ότι ένας αλγόριθμος κατάτμησης εικόνας με βάση κάποιο προγενέστερο σχήμα (prior shape), πρέπει να αντιμετωπίζει τα ακόλουθα τρία υποπροβλήματα:

- Επεξεργασία ενός συνόλου εκπαίδευσης, ώστε οι εικόνες να μπορούν να χρησιμοποιηθούν από τα επόμενα στάδια.
- Εξαγωγή ενός μοντέλου του αντικειμένου (προγενέστερο σχήμα), με βάση το σύνολο εκπαίδευσης.
- Χρήση του προγενέστερου σχήματος έτσι ώστε να εντοπιστεί στην εικόνα το επιθυμητό αντικείμενο.

2.3.1 Επεξεργασία του Συνόλου Εκπαίδευσης

Το σύνολο εκπαίδευσης αποτελείται από εικόνες οι οποίες περιέχουν το αντικείμενο που μας ενδιαφέρει. Προκειμένου να μπορέσουμε να χρησιμοποιήσουμε την πληροφορία που περιέχεται στις εικόνες, θα πρέπει πρώτα να βρούμε πώς θα την κωδικοποιήσουμε. Στα [18], [19], [29], [38] όπου οι συγγραφείς ασχολούνται με τις μεθόδους κατάκτησης με βάση πρότερη γνώση, επιλέγεται ως μέθοδος κωδικοποίησης της πληροφορίας η προσημασμένη συνάρτηση απόστασης από το περίγραμμα του επιθυμητού αντικειμένου. Η συγκεκριμένη συνάρτηση έχει τα επιθυμητά χαρακτηριστικά των συναρτήσεων επιπεδοσυνόλων (βλ. παράγραφο 2.2.3), ενώ δεν παρουσιάζει μεγάλη διαφοροποίηση αν τα σχήματα διαφέρουν λίγο μεταξύ τους. Η σημαντικότητα της τελευταίας ιδιότητας εντοπίζεται στο γεγονός ότι τα σχήματα από το σύνολο εκπαίδευσης είναι δυνατόν να διαφοροποιούνται ελαφρώς, αλλά οι συναρτήσεις που τα αναπαριστούν να είναι αρκετά όμοιες. Έτσι, ενώ είναι απαραίτητο να μεριμνήσει κάποιος για την ευθυγράμμιση των σχημάτων προκειμένου να εξάγει κάποιο μέσο σχήμα, δεν είναι αναγκαίο η ευθυγράμμιση αυτή να είναι τέλεια.

Λόγω του ανωτέρω παράγοντα, στο [19], δεν υπάρχει μέριμνα για την ευθυγράμμιση των σχημάτων του συνόλου εκπαίδευσης. Θεωρείται ότι εξ αρχής τα σχήματα έχουν παρόμοιο προσανατολισμό και ότι τυχόν διαφοροποιήσεις σε αυτόν καλύπτονται επαρκώς από την χρήση της προσημασμένης συνάρτησης απόστασης. Αντιθέτως, στα [38], [29] και [32] γίνεται προσπάθεια ευθυγράμμισης των εικόνων εκπαίδευσης. Κατά τη διαδικασία της ευθυγράμμισης των εικόνων αναζητούμε τους αφινικούς μετασχηματισμούς αυτών, έτσι ώστε να είναι όσο το δυνατόν πιο όμοιες. Το κριτήριο ομοιότητας ορίζεται διαφορετικά για κάθε ερευνητή, αν και στη γενική περίπτωση έχει να κάνει με μεγάλη επικάλυψη των αντικειμένων και του φόντου.

Το Πρόβλημα της Ευθυγράμμισης

Στο [30] το πρόβλημα της ευθυγράμμισης δύο εικόνων διατυπώνεται ως ακολούθως: *δεδομένων δύο σχημάτων, ενός σχήματος εισόδου D και ενός σχήματος στόχου S , και ενός μέτρου ανομοιότητας, αναζητείται ο καλύτερος μετασχηματισμός ο οποίος αντιστοιχεί σε κάθε σημείο του D ένα σημείο στο S , ενώ παράλληλα ελαχιστοποιεί το μέτρο ανομοιότητας μεταξύ του μετασχηματισμένου σχήματος \hat{D} και του στόχου S . Το μέτρο ανομοιότητας μπορεί να ορίζεται είτε πάνω στο περίγραμμα του αντικειμένου, είτε πάνω σε όλο το αντικείμενο.*

Το κεντρικό κομμάτι της ευθυγράμμισης των δύο εικόνων, είναι η εύρεση του μετασχηματισμού. Το πρόβλημα επομένως περιπλέκεται, ανάλογα με το είδος του μετασχηματισμού που αναζητούμε. Οι πιο συνηθισμένες περιπτώ-

σεις είναι ο μετασχηματισμός με περιστροφή και μετατόπιση, και ο γενικός αφινικός μετασχηματισμός. Στις περιπτώσεις που εμείς θα εξετάσουμε, αναζητείται η μετατόπιση, περιστροφή και κλίμακα του μετασχηματισμού. Σε κάποιες περιπτώσεις λαμβάνονται υπόψη και πιθανές τοπικές διαταραχές.

Ευθυγράμμιση με Χρήση Δυαδικών Σχημάτων

Στο [38] οι εικόνες του συνόλου εκπαίδευσης είναι δυαδικές. Το αντικείμενο που μας ενδιαφέρει είναι το μόνο το οποίο βρίσκεται στην εικόνα (όπως και σε όλες τις προσεγγίσεις) και είναι άσπρο σε μαύρο φόντο. Θεωρώντας ότι το σύνολο εκπαίδευσης είναι της μορφής $T = \{I^1, I^2, \dots, I^n\}$, όπου I^i με $i = 1 \dots n$ είναι οι δυαδικές εικόνες των αντικειμένων, αναζητούνται οι παράμετροι $\{\bar{p}^1, \bar{p}^2, \dots, \bar{p}^n\}$ των αφινικών μετασχηματισμών, έτσι ώστε οι μετασχηματισμένες εικόνες να είναι ευθυγραμμισμένες. Οι παράμετροι του αφινικού μετασχηματισμού είναι η μετατόπιση (translation) στους άξονες x και y , η περιστροφή (rotation), και η μεγέθυνση (scaling), δεν ασχολούμαστε δηλαδή με τη γενικότερη περίπτωση. Για να υπολογιστούν οι τρεις παράμετροι, ορίζεται πρώτα μια ενέργεια πάνω στο σύνολο T , που εξαρτάται από τις σχετικές θέσεις των αντικειμένων στις εικόνες:

$$E_{align} = \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left\{ \frac{\int_{\Omega} (T^i - T^j)^2 dA}{\int_{\Omega} (T^i + T^j)^2 dA} \right\} \quad (2.8)$$

όπου $T^i = I^i(T(p^i)[xy1]^T)$ η εικόνα που προκύπτει από εφαρμογή του αφινικού μετασχηματισμού με παραμέτρους p .

Η ενέργεια είναι μικρή όταν πολλές εικόνες ταυτίζονται μεταξύ τους και ιδανικά είναι μηδέν, όταν όλες οι εικόνες είναι ίδιες. Ο παράγοντας του αριθμητή δείχνει τον βαθμό επικάλυψης των εικόνων, ενώ ο όρος του παρονομαστή αποτελεί όρο κανονικοποίησης. Προφανώς, για να βρούμε τις παραμέτρους των αφινικών μετασχηματισμών, αρκεί να ελαχιστοποιήσουμε την ενέργεια E_{align} . Αυτό επιτυγχάνεται με τη μέθοδο κατάβασης κλίσης, θεωρώντας αρχικά μια εικόνα ως πρότυπο. Παραγωγίζοντας τη σχέση (2.8) ως προς p^i , προκύπτει η διαφορική εξίσωση εξέλιξης των παραμέτρων:

$$\begin{aligned} \nabla_{p^i} E_{align} = & \sum_{j=1, j \neq i}^n \left(\frac{2 \int_{\Omega} (T^i - T^j) \nabla_{p^i} T^i dA}{\int_{\Omega} (T^i + T^j)^2 dA} \right) - \\ & - \sum_{j=1, j \neq i}^n \left(\frac{2 \int_{\Omega} (T^i - T^j)^2 dA \int_{\Omega} (T^i + T^j) \nabla_{p^i} T^i dA}{(\int_{\Omega} (T^i + T^j)^2 dA)^2} \right) \end{aligned}$$

Μετά την εύρεση των παραμέτρων p^i , οι εικόνες περιστρέφονται, μετατοπίζονται, και μεγενθύνονται κατάλληλα, έτσι ώστε να ευθυγραμμιστούν. Στη

συνέχεια χρησιμοποιούνται προκειμένου να εξαχθεί το προγενέστερο σχήμα του επιθυμητού αντικειμένου.

Ευθυγράμμιση με Χρήση Συναρτήσεων Επιπεδοσυνόλων

Στο [29] ακολουθείται μια παρόμοια προσέγγιση, με μια βασική διαφορά: Ενώ στην περίπτωση του [38] συγκρίνονται δυαδικές εικόνες που περιέχουν το αντικείμενο, στο [29] πρώτα βρίσκονται οι συναρτήσεις προσημασμένης απόστασης και στη συνέχεια γίνεται η σύγκριση και η εύρεση των αφινικών μετασχηματισμών. Οι λεπτομέρειες αυτής της προσέγγισης αναφέρονται στο [30].

Αποδεικνύεται ότι ο μετασχηματισμός προσημασμένης απόστασης παραμένει αναλλοίωτος μετά από περιστροφή και μετατόπιση, και άρα ενδείκνυται η χρήση του ως μέτρου ομοιότητας, για τις περιπτώσεις που εξετάζουμε. Θεωρούμε ένα σχήμα D που προκύπτει μετά από την μετατόπιση κατά T και την περιστροφή κατά θ , ενός σχήματος S . Τότε, θεωρώντας ένα pixel (x, y) , η τιμή του μετασχηματισμού απόστασης στο σημείο (x, y) δίνεται από τη σχέση:

$$d = D((x, y), D) = \min_{(u,v) \in D} \left\{ \sqrt{(u-x)^2 + (v-y)^2} \right\}$$

Αποδεικνύεται τελικά ότι $D((\hat{x}, \hat{y}), S) = D((x, y), D)$, όπου:

$$\begin{cases} \hat{x} = x \cos \theta - y \sin \theta - T_x \\ \hat{y} = x \sin \theta + y \cos \theta - T_y \end{cases}$$

Επομένως, ο μετασχηματισμός απόστασης δεν εξαρτάται από τη μετατόπιση και την περιστροφή. Αποδεικνύεται ότι εξαρτάται από τον παράγοντα κλίμακας, και πιο συγκεκριμένα:

$$D((x, y), D) = sD((\hat{x}, \hat{y}), S) \quad (2.9)$$

Η σχέση (2.9) μπορεί να αποτελέσει επομένως το κριτήριο ομοιότητας μεταξύ δύο εικόνων, εφόσον έχουμε τους μετασχηματισμούς προσημασμένης απόστασης αυτών. Πιο συγκεκριμένα, αναζητούμε τον μετασχηματισμό A ο οποίος δημιουργεί αντιστοίχιση pixel-προς-pixel μεταξύ της δεδομένης αναπαράστασης $s\Phi_D$, και του σχήματος στόχου Φ_S . Δεδομένου ότι οι αναπαραστάσεις $[\Phi_D, \Phi_S]$ είναι αμετάβλητες από την μετατόπιση και περιστροφή, γράφουμε:

$$\left\{ \begin{array}{l} A(x, y) = \begin{pmatrix} A_x \\ A_y \end{pmatrix} = s \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \end{pmatrix} \\ \forall (x, y) \in \Omega : s\Phi_D(x, y) = \Phi_S(A(x, y)) \end{array} \right\}$$

Έχοντας ως βάση τα ανωτέρω, μπορούμε να θεωρήσουμε την ευθυγράμμιση ως πρόβλημα βελτιστοποίησης. Αναζητούμε δηλαδή τον βέλτιστο μετασχηματισμό A , έτσι ώστε να ελαχιστοποιηθεί το παρακάτω συναρτησιακό ενέργειας:

$$E(s, \theta, T) = \int_{\Omega} (s\Phi_D(x, y) - \Phi_S(A(x, y)))^2 dx dy \quad (2.10)$$

Το συναρτησιακό (2.10) ελαχιστοποιείται με βάση τη μέθοδο κατάβασης κλίσης. Παραγωγίζοντας ως προς τις παραμέτρους του συναρτησιακού, προκύπτουν οι παράμετροι εξέλιξης αυτών:

$$\begin{aligned} \frac{d\theta}{dt} &= 2 \int_{\Omega} (\nabla \Phi_S \nabla_{\theta} A^T) [s\Phi_D - \Phi_S(A^T)] \\ \frac{ds}{dt} &= 2 \int_{\Omega} (\Phi_D + \nabla \Phi_S \nabla_s A^T) [s\Phi_D - \Phi_S(A^T)] \\ \frac{dT}{dt} &= 2 \int_{\Omega} \left(\nabla \Phi_S \begin{pmatrix} \nabla_{T_x} A^T \\ \nabla_{T_y} A^T \end{pmatrix} \right) [s\Phi_D - \Phi_S(A^T)] \end{aligned}$$

Οι τελευταίες σχέσεις είναι δυνατόν να υπολογίζονται μέσα σε μια στενή ζώνη κοντά στα όρια του αντικειμένου, εμφανίζοντας αρκετά καλή συμπεριφορά. Αυτό οφείλεται στο γεγονός ότι ο μετασχηματισμός προσημασμένης απόστασης περιέχει πλεονάζουσα πληροφορία, καθώς ένα πλήθος επιπεδοσυνόλων ουσιαστικά δεν χρειάζεται για τον προσδιορισμό του αντικειμένου. Επομένως, είναι δυνατόν να χρησιμοποιούνται μόνο τα επιπεδοσύνολα που βρίσκονται κοντά στην ισοσταθμική επιφάνεια μηδενικού δυναμικού, εξακολουθώντας να έχουμε ικανοποιητικά αποτελέσματα.

Με την παρούσα μοντελοποίηση και τις εξισώσεις εξέλιξης, είναι δυνατόν να ευθυγραμμιστούν τα σχήματα. Ωστόσο, στο [30], οι συγγραφείς προχωρούν περισσότερο, και εξελίσσουν τη μέθοδο ευθυγράμμισης. Ενώ μέχρι τώρα γίνονταν αναζήτηση ενός μετασχηματισμού που εφαρμόζεται σε όλα τα σημεία της εικόνας, τώρα λαμβάνονται υπόψη και πιθανές τοπικές αποκλίσεις για λίγα pixel. Έτσι, στην ιδανική περίπτωση:

$$\left\{ \begin{array}{l} (u(x, y), v(x, y)) : (x, y) \in \Omega \\ \forall (x, y) \in \Omega : s\Phi_D(x, y) = \Phi_S(A^T + (u, v)) \end{array} \right\}$$

Οι ανωτέρω σχέσεις οδηγούν στην κατασκευή του ακόλουθου συναρτησιακού ενέργειας:

$$E(s, \theta, T, (u, v)) = \alpha \int_{\Omega} \underbrace{(s\Phi_D - \Phi_S(A^T))^2}_{\text{ολική ευθυγράμμιση}} + (1 - \alpha) \int_{\Omega} \underbrace{(s\Phi_D - \Phi_S(A^T + (u, v)))^2}_{\text{τοπικές διαφορές}}$$

Η παράμετρος α δίνει μεγαλύτερη ή μικρότερη βαρύτητα στους όρους της τοπικής ταύτισης, ή στον γενικό ολικό μετασχηματισμό. Παραγωγίζοντας την παραπάνω σχέση, χρησιμοποιούμε τον αλγόριθμο κατάβασης κλίσης, προκειμένου να υπολογίσουμε τις παραμέτρους s , θ , T , u , v . Σημειώνουμε ότι είναι δυνατόν στο τελικό μοντέλο να υπάρχει και κάποιος όρος ομαλότητας, ο οποίος εξαρτάται από τις τιμές $u_x^2 u_y^2 v_x^2 v_y^2$. Φυσικά, αλλάζει ανάλογα το συναρτησιακό ενέργειας, όπως και οι διαφορικές εξισώσεις που καθορίζουν την εξέλιξη. Για λόγους πληρότητας, παραθέτουμε το τελευταίο συναρτησιακό όπως προκύπτει στο [30]:

$$\begin{aligned}
 E(s, \theta, T, (u, v)) &= \underbrace{\alpha \int_{\Omega} \int (s\Phi_D - \Phi_S(A^T))^2}_{\text{όρος καθολικού μετασχηματισμού}} \\
 &+ \underbrace{(1 - \alpha)\beta \int_{\Omega} \int (s\Phi_D - \Phi_S(A^T + (u, v)))^2}_{\text{όρος τοπικών αποκλίσεων}} \\
 &+ \underbrace{(1 - \alpha)(1 - \beta) \int_{\Omega} \int (u_x^2 + u_y^2 + v_x^2 + v_y^2)}_{\text{όρος ομαλότητας}}
 \end{aligned}$$

Η μέθοδος κατάβασης κλίσης για τον υπολογισμό των παραμέτρων, χρησιμοποιεί τις διαφορικές εξισώσεις εξέλιξης που προκύπτουν από την εύρεση των Euler-Lagrange παραγώγων του συναρτησιακού ενέργειας:

$$\begin{aligned}
 \frac{ds}{dt} &= 2 \int_{\Omega} \int (\Phi_D + \nabla \Phi_S \nabla_s A^T) \cdot [\alpha(s\Phi_D - \Phi_S(A^T)) + \beta(1 - \alpha)(s\Phi_D - \Phi_S(A^T + (u, v)))] \\
 \frac{dT}{dt} &= 2 \int_{\Omega} \int (\nabla \Phi_S \begin{pmatrix} \nabla_{T_x} A^T \\ \nabla_{T_y} A^T \end{pmatrix}) \cdot [\alpha(s\Phi_D - \Phi_S(A^T)) + \beta(1 - \alpha)(s\Phi_D - \Phi_S(A^T + (u, v)))] \\
 \frac{d\theta}{dt} &= 2 \int_{\Omega} \int (\nabla \Phi_S \nabla_{\theta} A^T) \cdot [\alpha(s\Phi_D - \Phi_S(A^T)) + \beta(1 - \alpha)(s\Phi_D - \Phi_S(A^T + (u, v)))] \\
 \frac{du}{dt} &= 2(1 - \alpha)(1 - \beta)(u_{xx} + u_{yy}) + \\
 &+ 2\beta(1 - \alpha) \nabla_x \Phi_S(A^T + (u, v))(s\Phi_D - \Phi_S(A^T + (u, v)))
 \end{aligned}$$

$$\frac{dv}{dt} = 2(1 - \alpha)(1 - \beta)(v_{xx} + u_{yy}) + \\ + 2\beta(1 - \alpha)\nabla_y \Phi_S(A^T + (u, v))(s\Phi_D - \Phi_S(A^T + (u, v)))$$

Με την χρήση των παραπάνω σχέσεων είναι δυνατόν να υπολογιστούν οι απαιτούμενες παράμετροι έτσι ώστε τα σχήματα D και S να ευθυγραμμιστούν όσο το δυνατόν καλύτερα. Ας σημειωθεί ότι όλοι οι υπολογισμοί των ολοκληρωμάτων είναι δυνατόν να γίνονται σε μια στενή ζώνη, αντί σε ολόκληρη την εικόνα. Έτσι, όπως έχουμε εξηγήσει και παραπάνω, ελαττώνεται πολύ η πολυπλοκότητα του προβλήματος, χωρίς να υπάρχει μεγάλη επίδραση στη επίδοση, λόγω της πλεονάζουσας πληροφορίας που περιέχει η επιπεδοσυνάρτηση.

Συνοψίζοντας, στην παράγραφο αυτή περιγράψαμε δύο διαφορετικές μεθόδους που μπορούν να χρησιμοποιηθούν για την ευθυγράμμιση σχημάτων. Και οι δύο τεχνικές χρησιμοποιούν την μέθοδο κατάβασης κλίσης προκειμένου να ελαχιστοποιήσουν ένα συναρτησιακό ενέργειας. Η ενέργεια του εν λόγω συναρτησιακού πηγάζει από την ανομοιότητα των εικόνων μεταξύ τους. Η διαφορά των δύο τεχνικών είναι ότι στη μια περίπτωση γίνεται χρήση δυαδικών εικόνων, ενώ στη δεύτερη γίνεται χρήση του μετασχηματισμού προσημασμένης ευκλείδειας απόστασης. Επίσης, η δεύτερη περίπτωση επεκτείνεται έτσι ώστε πέρα από την εύρεση ενός καθολικού μετασχηματισμού, να λαμβάνει υπόψη της και πιθανές τοπικές διακυμάνσεις, ή και παράγοντες ομαλότητας. Τέλος, η δεύτερη περίπτωση εξετάζει την κάθε εικόνα ξεχωριστά, η ελαχιστοποίηση δηλαδή της ενέργειας γίνεται ξεχωριστά για κάθε εικόνα, αν θεωρήσουμε μια εικόνα ως αναφορά. Αντίθετα, στην πρώτη περίπτωση, το συναρτησιακό ενέργειας περιλαμβάνει ταυτόχρονα πληροφορία για όλες τις εικόνες. Ας σημειώσουμε ότι μελετήσαμε δύο πλήρως αυτόματες τεχνικές ευθυγράμμισης. Μια χρήσιμη αλλά γενικά απλή ημιαυτόματη τεχνική είναι η εύρεση του αφινικού μετασχηματισμού μέσω της ταύτισης σημείων της εικόνας. Τα σημεία τα επιλέγει ο χρήστης, έτσι ώστε να ανταποκρίνονται στο σχήμα του ενδιαφέροντός μας. Γενικά πάντως, μετά την εφαρμογή κάποιου από τους αλγορίθμους αυτούς στο σύνολο εκπαίδευσης, αποκτούμε ένα σύνολο από ευθυγραμμισμένες εικόνες του αντικειμένου που μας ενδιαφέρει, και μπορούμε πλέον να προχωρήσουμε στην κατασκευή του προγενέστερου σχήματος.

2.3.2 Εξαγωγή του Προγενέστερου Σχήματος

Στην προηγούμενη παράγραφο παρουσιάσαμε την προεπεξεργασία που πρέπει να γίνει στο σύνολο εκπαίδευσης. Κάποιες μέθοδοι απαιτούν την ευθυγράμμιση των εικόνων εκπαίδευσης, ενώ άλλες όχι. Η ευθυγράμμιση μπορεί να γίνει με χρήση είτε δυαδικών εικόνων, είτε με χρήση των επιπεδοσυναρτήσεων και συγκεκριμένα με την χρήση της συνάρτησης προσημασμένης απόστασης

από το μέτωπο. Σε κάθε περίπτωση, το σύνολο εκπαίδευσης υπόκειται σε επεξεργασία με την εφαρμογή κάποιου αλγορίθμου. Πλέον, πρέπει να χρησιμοποιηθεί η πληροφορία του συνόλου εκπαίδευσης έτσι ώστε να δημιουργηθεί ένα μοντέλο (προγενέστερο σχήμα) του επιθυμητού αντικειμένου. Οι μέθοδοι τις οποίες θα εξετάσουμε εισάγονται στα [32], [29], [19] και [38].

Οι μέθοδοι που περιγράφονται στις ανωτέρω δημοσιεύσεις έχουν μια βασική ομοιότητα, και αρκετές διαφορές. Συγκεκριμένα, όλες χρησιμοποιούν ως βάση την αναπαράσταση των σχημάτων του συνόλου εκπαίδευσης ως επιπεδοσυναρτήσεις, με χρήση του μετασχηματισμού προσημασμένης απόστασης. Ωστόσο, ο τρόπος με τον οποίο χρησιμοποιείται αυτή η πληροφορία διαφέρει μεταξύ των δημοσιεύσεων. Θα εξετάσουμε την κοινή μέθοδο που προτείνουν οι ομάδες των Leventon et al. ([19]) και Tsai et al. ([38]), σε αντιπαράθεση με την μέθοδο των Paragios et al. ([32]). Στη συνέχεια θα εξετάσουμε τη δεύτερη μέθοδο που προτείνεται από τους Paragios et al. ([29]), και σε επόμενο κεφάλαιο κάποιες μεθόδους οι οποίες δεν κάνουν χρήση επιπεδοσυνόλων.

Εξαγωγή Προγενέστερου Σχήματος μέσω Ανάλυσης σε Κύριες Συνιστώσες

Η μέθοδος της ανάλυσης σε κύριες συνιστώσες (singular value decomposition), χρησιμοποιείται προκειμένου να βρεθούν οι κύριες συνιστώσες ενός πίνακα που αποτελείται από διανύσματα δεδομένων εισόδου. Αλλιώς, δεδομένου ενός πίνακα που έχει για γραμμές κάποια διανύσματα δεδομένων, εφαρμόζοντας την ανάλυση σε κύριες συνιστώσες, μπορούμε να βρούμε ποιές είναι κύριες συνιστώσες των δεδομένων, δηλαδή τα διανύσματα εκείνα, η γνώση των οποίων αρκεί για να περιγράψουμε επαρκώς τα δεδομένα. Στη συνέχεια, αρκεί να γίνεται χρήση αυτών των συνιστωσών αντί για το σύνολο των δεδομένων, και να έχουμε ικανοποιητικά αποτελέσματα. Είναι δυνατόν να επιλέγεται ο αριθμός των συνιστωσών που χρησιμοποιούνται, έτσι ώστε να ελαχιστοποιείται κάποιο σφάλμα προσέγγισης². Επίσης, με την ανάλυση σε κύριες συνιστώσες είναι δυνατόν να εντοπίζουμε τους χώρους στους οποίους κινούνται τα δεδομένα μας.

Η χρήση της ανάλυσης σε κύριες συνιστώσες είναι η κεντρική ιδέα του αλγορίθμου των [19], και [38], προκειμένου να εξαχθεί το μοντέλο του επιθυμητού αντικειμένου. Προκειμένου να χρησιμοποιηθεί η μέθοδος, είναι απαραίτητο να κατασκευαστούν τα διανύσματα δεδομένων. Η κάθε επιπεδοσυνάρτηση του συνόλου εκπαίδευσης αποτελεί μια επιφάνεια που έχει υποστεί δειγματοληψία, και αποτελείται από $n \times m$ δείγματα, όπου n και m είναι οι διαστάσεις της εικόνας. Το κάθε σημείο έχει τιμή ίση με την προσημασμένη απόσταση

². Θεωρείται ότι το μαθηματικό υπόβαθρο της μεθόδου singular value decomposition είναι γνωστό στον αναγνώστη, και επομένως παραλείπεται.

από το σύνορο του αντικειμένου, και η τιμή αυτή μπορεί να αποτελεί στοιχείο σε ένα διάνυσμα διάστασης $1 \times (nm)$. Επομένως, διατρέχοντας τους πίνακες των επιπεδοσυναρτήσεων με κάποιο συγκεκριμένο τρόπο, δημιουργούμε για κάθε επιφάνεια του συνόλου $T = \{u_1, u_2, \dots, u_n\}$ του συνόλου εκπαίδευσης, ένα διάνυσμα Φ_i , το οποίο έχει στοιχεία ίσο με το πλήθος των στοιχείων της εικόνας.

Το σύνολο των διανυσμάτων-δεδομένων $\{\Phi_1, \Phi_2, \dots, \Phi_n\}$ περιέχει ένα μεγάλο ποσοστό πλεονάζουσας πληροφορίας, και αυτό οφείλεται σε δύο λόγους. Ο πρώτος λόγος είναι ότι ο μετασχηματισμός προσημασμένης απόστασης μπορεί να υπολογιστεί έχοντας μόνο τη γνώση της ισοσταθμικής καμπύλης μηδενικού δυναμικού. Επομένως, όλες οι υπόλοιπες ισοσταθμικές καμπύλες αποτελούν πλεονάζουσα πληροφορία. Ο δεύτερος λόγος έχει αναφερθεί ήδη στην παράγραφο (2.3.1): Για μικρές διαφοροποιήσεις μεταξύ των εικόνων εκπαίδευσης, οι μετασχηματισμοί προσημασμένης απόστασης δεν έχουν σημαντικά διαφοροποιημένη μορφή. Επομένως, είναι δυνατόν τα $\{\Phi_1, \Phi_2, \dots, \Phi_n\}$ να περιέχουν πολλαπλά την ίδια πληροφορία.

Με βάση τα παραπάνω μπορούμε να καταλήξουμε στο συμπέρασμα ότι η μέθοδος της ανάλυσης σε κύριες συνιστώσες μπορεί να εφαρμοστεί. Με την εξαγωγή των κύριων συνιστωσών, θα δημιουργήσουμε ένα μοντέλο το οποίο θα περιέχει ικανοποιητική πληροφορία ώστε να προσεγγίζει όλα τα δεδομένα εισόδου. Επίσης, μπορεί να λαμβάνεται κατάλληλος αριθμός συνιστωσών, έτσι ώστε να έχουμε το δυνατόν μικρότερο σφάλμα. Τέλος, στην περίπτωση που υπάρχουν διακυμάνσεις μεταξύ των εικόνων εισόδου, μπορεί να δωθεί κατάλληλο βάρος σε συγκεκριμένες κύριες συνιστώσες ώστε να προσεγγίζεται το αντικείμενο με την συγκεκριμένη διακύμανση, με μεγαλύτερη ακρίβεια.

Προκειμένου να εφαρμοστεί η μέθοδος ανάλυσης σε κύριες συνιστώσες, πρέπει να βρεθεί ένα μέσο διάνυσμα των δεδομένων εισόδου, που να περιέχει τις τιμές της μέσης επιφάνειας. Έτσι:

$$\mu = \frac{1}{n} \sum_{i=1}^n u_i$$

όπου μ είναι η μέση επιφάνεια, και Φ_i είναι τα διανύσματα που προκύπτουν από τις επιφάνειες του συνόλου εκπαίδευσης. Στη συνέχεια, το μέσο σχήμα αφαιρείται από κάθε διάνυσμα πληροφορίας, δημιουργώντας έτσι τα διανύσματα $\hat{\Phi}_i = \Phi_i - \mu$, $i=1 \dots n$, τα οποία θα χρησιμοποιηθούν κατά τη διαδικασία της ανάλυσης:

$$U \Sigma U^T = \frac{1}{n} M M^T$$

Ο πίνακας M περιέχει ως γραμμές τα διανύσματα Φ_i , και επομένως ο πίνακας $M M^T$ αποτελεί τον πίνακα συνδιακύμανσης των Φ_i . Ο πίνακας U περιέχει τις κύριες συνιστώσες στις οποίες αναλύεται ο πίνακας M , και ο πίνακας Σ

περιέχει τα βάρη των συνιστωσών αυτών, έτσι ώστε συνδυασμένες να μπορούν να δώσουν τον πίνακα συνδιακύμανσης.

Έχοντας πλέον την ανάλυση σε συνιστώσες, μπορεί να επιλεγεί ο αριθμός των συνιστωσών και κατάλληλα βάρη για αυτές, έτσι ώστε να μοντελοποιηθεί το αντικείμενο που αναζητούμε. Επομένως, επιλέγονται παράμετροι α :

$$\bar{\Phi} = \alpha U^k + \mu$$

όπου U^k είναι οι πρώτες k στήλες του πίνακα U . Όλες οι συνιστώσες U^k είναι γνωστές, όπως επίσης και το μέσο διάνυσμα μ . Η επιφάνεια $\bar{\Phi}$ η οποία προκύπτει από την παραπάνω εξίσωση δεν αποτελεί μετασχηματισμό προσημασμένης απόστασης, διατηρεί ωστόσο κάποιες βασικές ιδιότητες των επιπεδοσυναρτήσεων, όπως η ομαλότητα, η τοπικότητα και η ιδιότητα το επιπεδοσύνολο μηδενικού δυναμικού να δίνει το περίγραμμα του αντικειμένου. Σε κάθε περίπτωση, κατά την κατάτμηση εικόνας, αναζητούμε τους συντελεστές α έτσι ώστε να έχουμε όσο το δυνατόν καλύτερη προσέγγιση του επιθυμητού σχήματος. Επομένως, πρέπει να μοντελοποιήσουμε την εξάρτηση των συντελεστών από την δοσμένη εικόνα, και να βρεθεί μια διαφορική εξίσωση εξέλιξης αυτών, ώστε η χρήση του προγενέστερου σχήματος Φ που κατασκευάζεται, να οδηγεί προς την επιθυμητή κατάτμηση.

Γενικά, πρέπει να λαμβάνεται υπόψη και η περίπτωση το επιθυμητό αντικείμενο να βρίσκεται περιστραμένο, διεσταλμένο, ή μετατοπισμένο στην τελική εικόνα, σε σχέση με τις εικόνες του συνόλου εκπαίδευσης. Επομένως, θα πρέπει να εισάγεται και η πληροφορία του μετασχηματισμού. Η γενική περίπτωση επομένως, μαζί με την κατάτμηση εικόνας, επιδιώκει και την εύρεση του βέλτιστου προγενέστερου σχήματος, σύμφωνα με την εξίσωση:

$$\bar{\Phi}(\alpha, p) = \mu(p) + \sum_1^k \alpha_i v_i(p) \quad (2.11)$$

όπου p είναι το διάνυσμα που περιέχει τις παραμέτρους s, θ, \vec{T} του αφινικού μετασχηματισμού, v_i με $i = 1 \dots k$ τα k πρώτα διανύσματα του πίνακα U , και α_i τα αντίστοιχα βάρη.

Το μοντέλο που προκύπτει από την εξίσωση (2.11) θα χρησιμοποιηθεί κατά την κατάτμηση εικόνας σύμφωνα με τις μεθόδους στα [19] και [38], προσπαθώντας να προσδιοριστούν ταυτόχρονα οι βέλτιστες παράμετροι.

Εξαγωγή Προγενέστερου Σχήματος μέσω Βέλτιστου Μέσου Σχήματος

Στο [32] οι εικόνες του συνόλου εκπαίδευσης έχουν ευθυγραμμιστεί, και έχουν βρεθεί οι επιπεδοσυναρτήσεις τους, έστω $\{\Phi_1, \Phi_2, \dots, \Phi_n\}$. Σκοπός είναι

να βρεθεί μια αναπαράσταση με τη μορφή επιπεδοσυνάρτησης, η οποία να προσεγγίζει όσο το δυνατόν καλύτερα τις υπάρχουσες αναπαραστάσεις, η εύρεση δηλαδή ενός βέλτιστου σχήματος. Η προσέγγιση αυτή διαφέρει από την απλή μέθοδο του της ανάλυσης σε κύριες συνιστώσες, στον τρόπο που βρίσκεται το μέσο σχήμα.

Συγκεκριμένα, οι Paragios et al. ([32], [30]), επιθυμούν το βέλτιστο σχήμα να διαθέτει όλες τις ιδιότητες επιπεδοσυναρτήσεων. Το μέσο σχήμα, όπως υπολογίζεται στα [38] και [19], έχει την ιδιότητα της ομαλότητας, και της τοπικότητας, ενώ το επιπεδοσύνολο μηδενικού δυναμικού δίνει το περίγραμμα του αντικειμένου. Ωστόσο, δεν ισχύει $\|\nabla\bar{\varphi}\| = 1$, που είναι η πιο βασική ιδιότητα της επιφάνειας που προκύπτει από τον προσημασμένο μετασχηματισμό ευκλείδειας απόστασης. Στο [32] ακολουθείται μια ελαφρώς διαφορετική προσέγγιση για το μέσο σχήμα, με αποτέλεσμα όλες οι ιδιότητες να ισχύουν.

Η προσέγγιση τους μπορεί να γίνει εύκολα αντιληπτή, αν σκεφτούμε ότι το μέσο σχήμα είναι αυτό το οποίο απέχει το λιγότερο από όλες τις αναπαραστάσεις. Αλλιώς, το μέσο-βέλτιστο σχήμα, δίνεται από την επιπεδοσυνάρτηση εκείνη, για την οποία ελαχιστοποιείται το συναρτησιακό:

$$\left\{ \begin{array}{l} E(\Phi_M) = \sum_{i=1}^n \int_{\Omega} (\Phi_i - \Phi_M)^2 d\Omega \\ \text{δεδομένου ότι } \|\nabla\Phi_M\|^2 = 1 \end{array} \right\} \quad (2.12)$$

Η πρώτη εξίσωση του συνόλου εξισώσεων (2.12) δημιουργεί ένα συναρτησιακό που ισούται με το άθροισμα των τετραγώνων των σφαλμάτων μεταξύ κάθε επιπεδοσυνάρτησης του συνόλου εκπαίδευσης, και του μέσου σχήματος που υπολογίζουμε. Η δεύτερη εξίσωση παρέχει τον απαιτούμενο περιορισμό προκειμένου η τελική επιφάνεια που υπολογίζεται, να μπορεί να αποτελεί μετασχηματισμό προσημασμένης απόστασης.

Η ελαχιστοποίηση του παραπάνω συναρτησιακού πετυχαίνεται με τη μέθοδο κατάβασης κλίσης. Υπολογίζοντας τις παραγώγους της ενέργειας, προκύπτει η διαφορική εξίσωση εξέλιξης:

$$\frac{d}{dt}\Phi_M = 2 \sum_{i=1}^n (\Phi_i - \Phi_M) \quad (2.13)$$

ενώ η προβολή της επιφάνειας που εξελίσσεται σύμφωνα με την παραπάνω διαφορική εξίσωση γίνεται μέσω της διαφορικής εξίσωσης:

$$\frac{d}{dt}\Phi_M = (1 - \text{sign}(\Phi_M^0))(1 - \|\nabla\Phi_M\|) \quad (2.14)$$

όπου Φ_M^0 είναι η επιφάνεια που προκύπτει από την διαφορική εξίσωση (2.13) με την μέθοδο κατάβασης κλίσης. Τα δυο βήματα εναλλάσσονται μέχρι να προκύψει μια λύση τελικής κατάστασης.

Για την εφαρμογή των ανωτέρω διαφορικών εξισώσεων, είναι απαραίτητη κατάλληλη αρχικοποίηση της επιπεδοσυνάρτησης που αντιστοιχεί στο μέσο σχήμα, ώστε να περιοριστεί ο απαιτούμενος αριθμός επαναλήψεων. Ως κατάλληλη αρχικοποίηση θεωρούμε την μέση επιφάνεια υπολογισμένη όπως στο [38] και [19], σύμφωνα με τη σχέση:

$$\Phi_M^0 = \frac{1}{n} \sum_{i=1}^n (\Phi_i)$$

Το μέσο σχήμα το οποίο έχει προκύψει μπορεί στη συνέχεια να χρησιμοποιηθεί όπως στην προηγούμενη παράγραφο, είτε να χρησιμοποιηθεί ως έχει και να καθοδηγήσει την κατάτμηση προς την εύρεση του επιθυμητού αντικειμένου. Τον τρόπο χρήσης του μέσου σχήματος θα τον δούμε πιο αναλυτικά στην επόμενη ενότητα.

Εξαγωγή Προγενέστερου Σχήματος με χρήση Πιθανοτήτων

Στις μέχρι τώρα προσεγγίσεις που περιγράψαμε, η προσπάθεια μοντελοποίησης των τοπικών διαφοροποιήσεων μεταξύ των σχημάτων του συνόλου εκπαίδευσης, γίνεται με χρήση της μεθόδου ανάλυσης σε κύριες συνιστώσες, με εφαρμογή κατάλληλων βαρών στις συνιστώσες οι οποίες εξάγονται. Στο [29] γίνεται προσπάθεια να ενσωματωθούν όλες οι τοπικές διαφοροποιήσεις από το μέσο σχήμα, σε ένα γενικό πιθανοτικό μοντέλο.

Το πιθανοτικό μοντέλο το οποίο αναζητείται έχει δύο παραμέτρους οι οποίες πρέπει να βρεθούν, ένα μέσο σχήμα Φ_M , και μια διακύμανση σ_M , η οποία μοντελοποιεί τις τοπικές διαφοροποιήσεις. Η τιμή σε κάθε σημείο της επιφάνειας του μοντέλου, θεωρείται ότι ακολουθεί ένα Gaussian μοντέλο, με συνάρτηση πυκνότητας πιθανότητας:

$$p_{(x,y)}^M(\Phi) = \frac{1}{\sqrt{2\pi\sigma_M(x,y)}} e^{-\frac{(\Phi(x,y) - \Phi_M(x,y))^2}{2\sigma_M^2(x,y)}}$$

Σύμφωνα με το μοντέλο αυτό, κάθε σημείο μιας επιφάνειας Φ από το σύνολο εκπαίδευσης, έχει τιμή με πιθανότητα εμφάνισης που εξαρτάται από το βέλτιστο-μέσο σχήμα Φ_M και την διακύμανση σ_M . Μέσω του παραπάνω μοντέλου μπορούμε να δημιουργήσουμε ένα συναρτησιακό ενέργεια, και να ανάγουμε το πρόβλημα της εύρεσης των Φ_M και σ_M σε πρόβλημα βελτιστοποίησης, όπως και στις περιπτώσεις που περιγράφηκαν στις προηγούμενες ενότητες. Δεδομένου ότι το σημείο κάθε επιφάνειας αποτελεί μια πιθανότητα, το μέσο σχήμα Φ_M και η τοπική διακύμανση σ_M που αναζητούμε, θα είναι αυτές για τις οποίες οι εν λόγω πιθανότητες γενικά μεγιστοποιούνται. Έτσι, ορίζεται το ακόλουθο συναρτησιακό:

$$E(\Phi_M, \sigma_M) = - \sum_{i=1}^n \int \int_{(x,y)} \log \left[p_{(x,y)}^M \left(\hat{\Phi}_i(x,y) \right) \right] dx dy$$

όπου $\hat{\Phi}_i$ με $i = 1..n$ είναι οι ευθυγραμμισμένοι προσημασμένοι μετασχηματισμοί απόστασης για τα σχήματα του συνόλου εκπαίδευσης.

Προκειμένου να βρούμε τις επιθυμητές επιφάνειες, μπορούμε να βρούμε τις παραγώγους Euler-Lagrange, και να εφαρμόσουμε τη μέθοδο κατάβασης κλίσης. Ωστόσο, η μέθοδος κατάβαση κλίσης θα μας οδηγήσει στην εύρεση της βέλτιστης επιφάνειας όσον αφορά το πιθανοτικό μοντέλο. Η επιφάνεια αυτή δεν θα αποτελεί προσημασμένο μετασχηματισμό απόστασης, και επομένως, πρέπει να εισαχθεί και επιπλέον περιορισμός:

$$\|\nabla \Phi_M(x, y)\|^2 = 1$$

Εισάγοντας και κάποιες επιπλέον παραμέτρους ομαλότητας στο συναρτησιακό ενέργειας, δημιουργείται το σύνολο των εξισώσεων που απαιτείται για το πρόβλημα βελτιστοποίησης ([29]):

$$\left\{ \begin{array}{l} E(\Phi_M, \sigma_M) = \underbrace{(1 - \alpha) \int \int_{\Omega} \left(\left(\frac{d}{dx} \sigma_M(x, y) \right)^2 + \left(\frac{d}{dy} \sigma_M(x, y) \right)^2 \right) dx dy}_{\text{όρος ομαλοποίησης}} \\ \underbrace{\alpha \int \int_{\Omega} \sum_{i=1}^N \left(\log [\sigma_M(x, y)] + \frac{(\hat{\Phi}_i(x, y) - \Phi_M(x, y))^2}{2\sigma_M^2(x, y)} \right) dx dy}_{\text{όρος προσαρμογής στα δεδομένα σχήματα}} \end{array} \right\} \quad \text{Δεδομένου ότι: } \|\nabla \Phi_M(x, y)\|^2 = 1 \quad (2.15)$$

Επομένως, η μέθοδος κατάβασης κλίσης πρέπει να εφαρμοστεί στο σύνολο των δύο εξισώσεων (2.15). Συνεπώς, πρέπει να βρεθεί η παράγωγος κατά Euler-Lagrange, και να δημιουργηθεί η διαφορική εξίσωση εξέλιξης των παραμέτρων του ενεργειακού. Ωστόσο, λόγω της μορφής των εξισώσεων και της εμφάνισης των δεύτερων παραγώγων, δεν είναι δυνατόν να αποδειχθεί η ισχύς του θεωρήματος Lagrange. Τελικά, όπως και στην προηγούμενη παράγραφο, δημιουργούνται δύο διαφορετικά σύνολα διαφορικών εξισώσεων εξέλιξης, ένα από την ελαχιστοποίηση του συναρτησιακού ενέργειας το οποίο οδηγεί στην εύρεση της βέλτιστης επιφάνειας σχετικά με την πληροφορία που υπάρχει στο σύνολο εκπαίδευσης, και ένα που να προβάλλει τις εν λόγω επιφάνειες στον χώρο των επιφανειών που αποτελούν προσημασμένο μετασχηματισμό απόστασης. Έτσι, προκύπτουν οι ακόλουθες τρεις διαφορικές εξισώσεις:

$$\frac{d}{dt} \Phi_M = \alpha \sum_{i=1}^n \frac{(\hat{\Phi}_i - \Phi_M)}{2\sigma_M^2}$$

$$\frac{d}{dt}\sigma_M = \alpha \sum_{i=1}^n \left[-\frac{1}{\sigma_M} + \frac{(\hat{\Phi}_i - \Phi_M)^2}{\sigma_M^3} \right] + (1 - \alpha) \left[\frac{\partial^2}{\partial x^2} \sigma_M + \frac{\partial^2}{\partial y^2} \sigma_M \right]$$

$$\frac{d}{dt}\Phi_M = (1 - \text{sign}(\Phi_M^0))(1 - \|\nabla \Phi_M\|)$$

Αρχικά γίνεται χρήση των δύο πρώτων εξισώσεων για κάποιες επαναλήψεις, και στη συνέχεια η επιφάνεια Φ_M που προκύπτει χρησιμοποιείται ως η Φ_M^0 , στην τρίτη διαφορική εξίσωση. Όσον αφορά τις αρχικές τιμές των επιφανειών, η μέση επιφάνεια αρχικοποιείται σε ένα από τα σχήματα του συνόλου εκπαίδευσης, και κάθε σημείο της επιφάνειας διακυμάνσεων αρχικοποιείται στη μονάδα. Προκειμένου να αποφευχθούν τυχόντα προβλήματα ευστάθειας, η διακύμανση μπορεί να θεωρηθεί ως $\sigma_M = 1 + \hat{\sigma}_M$, και να βρίσκονται οι τιμές της $\hat{\sigma}_M$, υπό τον περιορισμό να είναι αυστηρά θετικές.

Συνοψίζοντας, με εισαγωγή των πιθανοτήτων στα επιπεδοσύνολα, κατασκευάστηκε ένα προγενέστερο σχήμα το οποίο λαμβάνει υπόψη του όχι μόνο το βέλτιστο σχήμα που εξάγεται από το σύνολο εκπαίδευσης, αλλά και τις πιθανές τοπικές διακυμάνσεις. Οι διακυμάνσεις είναι ενσωματωμένες στο προγενέστερο σχήμα, και δεν απαιτείται η εξαγωγή των κύριων συνιστωσών και η εφαρμογή βαρών σε αυτές, όπως στις προηγούμενες μεθόδους που περιγράψαμε.

2.3.3 Χρήση του Προγενέστερου Σχήματος

Στην προηγούμενη ενότητα παρουσιάστηκαν τρόποι με τους οποίους κατασκευάζεται ένα μοντέλο του επιθυμητού σχήματος, από ένα σύνολο σχημάτων εκπαίδευσης. Το προγενέστερο σχήμα είναι σε όλες τις περιπτώσεις στη μορφή επιπεδοσυνάρτησης. Στην ενότητα αυτή θα παρουσιάσουμε τον τρόπο χρήσης του μοντέλου, προκειμένου να εντοπιστεί το επιθυμητό αντικείμενο σε μια εικόνα.

Προγενέστερα Σχήματα και Κατάτμηση Εικόνας

Στο [38] δημιουργήθηκε μια αναπαράσταση του αντικειμένου μέσω της μεθόδου ανάλυσης σε κύριες συνιστώσες. Έτσι, πλέον υπάρχει μια αναπαράσταση του επιθυμητού αντικειμένου με χρήση των κύριων συνιστωσών, και αυτή μπορεί να χρησιμοποιηθεί κατά τη διαδικασία της κατάτμησης.

Οι συγγραφείς Tsai et al. κάνουν χρήση της αναπαράστασης που προέκυψε και του μοντέλου που έχει δημιουργηθεί, προκειμένου να κατασκευαστούν συναρτησιακά ενέργειας, και να δημιουργηθεί ένα πρόβλημα βελτιστοποίησης η επίλυση του οποίου θα οδηγήσει στην επιθυμητή κατάτμηση της εικόνας. Συγκεκριμένα, το μοντέλο που χρησιμοποιούν, όπως είδαμε από την προηγούμενη

παράγραφο, είναι το:

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

όπου p είναι οι παράμετροι που καθορίζουν τον αφινικό μετασχηματισμό, w_i , με $i = 1 \dots k$ το βάρος της i -οστής συνιστώσας, και Φ_i , με $i = 1 \dots k$, η i -οστή κύρια συνιστώσα. Στόχος είναι να βρεθούν τα βάρη w_i και οι παράμετροι p , έτσι ώστε να γίνεται η βέλτιστη κατάτμηση, δηλαδή ο εντοπισμός του επιθυμητού αντικειμένου.

Έστω οι περιοχές R^u και R^v , οι περιοχές εσωτερικά και εξωτερικά της καμπύλης κατάτμησης αντίστοιχα. Τότε, ορίζεται η ενέργεια:

$$\begin{aligned} E_{cv} &= -(\mu^2 A_u + v^2 A_v) \\ &= -\left(\frac{S_u^2}{A_u} + \frac{S_v^2}{A_v}\right) \end{aligned} \quad (2.16)$$

όπου μ : μέση φωτεινότητα της περιοχής R_u

S_u : άθροισμα της φωτεινότητας στην περιοχή R_u

A_u : εμβαδό της φωτεινότητας στην περιοχή R_u

και v , S_v και A_v τα αντίστοιχα μεγέθη στην περιοχή R_v . Η συγκεκριμένη σχέση αποδεικνύεται στο [7] από τους Chan T. και Vese L. Η χρήση της φωτεινότητας κατά την κατάτμηση πηγάζει από το γεγονός ότι ένα συγκεκριμένο αντικείμενο το οποίο αναζητούμε, θα έχει στη γενική περίπτωση φωτεινότητα παρόμοια στο εσωτερικό του, και έντονα διαφορετική φωτεινότητα στο εξωτερικό του. Επομένως, είναι δυνατόν η φωτεινότητα να χρησιμοποιηθεί ως πληροφορία προκειμένου να οδηγήσει την κατάτμηση.

Όλα τα μεγέθη του συναρτησιακού (2.16) μπορούν να υπολογιστούν. Δεδομένου ότι το μοντέλο που έχει κατασκευαστεί πληρεί κάποιες από τις ιδιότητες του προσημασμένου μετασχηματισμού απόστασης, μπορεί να χρησιμοποιηθεί για τον προσδιορισμό των περιοχών στο εσωτερικό και το εξωτερικό της καμπύλης μηδενικού δυναμικού. Αρχικά, ορίζουμε την συνάρτηση Heaviside:

$$H(x) = \begin{cases} 1 & \text{εάν } x < 0 \\ 0 & \text{εάν } x \geq 0 \end{cases}$$

Κάθε επιπεδοσυνάρτηση με ιδιότητες μετασχηματισμού προσημασμένης απόστασης, έτσι όπως έχει οριστεί από τα προηγούμενα κεφάλαια, διατηρεί αρνητικό πρόσημο στο εσωτερικό της καμπύλης μηδενικού δυναμικού, και θετικό πρόσημο στο εξωτερικό της καμπύλης μηδενικού δυναμικού. Επομένως, με βάση τους παραπάνω ορισμούς και τη συνάρτηση Heaviside, προκύπτει ότι:

$$S_u = \int \int_{\Omega} IH(\Phi(w, p)) dA, A_u = \int \int_{\Omega} H(\Phi(w, p)) dA, S_v = \int \int_{\Omega} IH(-\Phi(w, p)) dA$$

και $A_v = \int_{\Omega} H(-\Phi(w, p)) dA$. Με I συμβολίζουμε την φωτεινότητα της εικόνας. Παρατηρούμε ότι η ενέργεια (2.16) έχει άγνωστες μόνο τις παραμέτρους w και p . Επομένως, στο πρόβλημα βελτιστοποίησης που προέκυψε, υπολογίζουμε τις παραγώγους ως προς τις παραμέτρους ενδιαφέροντος:

$$\begin{aligned} \nabla_{(w,p)} E_{cv} = & -2(\mu \nabla_{(w,p)} S_u + v \nabla_{(w,p)} S_v) + \\ & + (\mu^2 \nabla_{(w,p)} A_u + v^2 \nabla_{(w,p)} A_v) \end{aligned}$$

όπου:

$$\begin{aligned} \nabla_{w_i} A_u &= -\nabla_{w_i} A_v = -\int_{\vec{C}} \Phi_i(p) ds \\ \nabla_{w_i} S_u &= -\nabla_{w_i} S_v = -\int_{\vec{C}} I \Phi_i(p) ds \\ \nabla_{p_i} A_u &= -\nabla_{p_i} A_v = -\int_{\vec{C}} \nabla_{p_i} \Phi(w, p) ds \\ \nabla_{p_i} S_u &= -\nabla_{p_i} S_v = -\int_{\vec{C}} \nabla_{p_i} I \Phi(w, p) ds \end{aligned}$$

Η καμπύλη \vec{C} αποτελεί την ισοσταθμική επιφάνεια μηδενικού δυναμικού της επιφάνειας Φ . Έχοντας τις ανωτέρω διαφορικές εξισώσεις, μπορούμε με τη μέθοδο κατάβασης κλίσης να βρούμε τις διαφορικές εξισώσεις εξέλιξης των παραμέτρων.

Συνοψίζοντας, η μέθοδος που περιγράφεται στο [38], αρχικά ευθυγραμμίζει τα σχήματα του συνόλου εκπαίδευσης, και στη συνέχεια δημιουργεί από αυτά ένα μοντέλο του επιθυμητού αντικειμένου. Το μοντέλο αυτό χρησιμοποιείται στη συνέχεια για την κατάτμηση μιας εικόνας, παρουσιάζοντας πολύ ενδιαφέροντα αποτελέσματα τόσο σε συνθήκες υψηλού θορύβου, όσο και σε περιπτώσεις κατεστραμμένων ακμών. Ο αλγόριθμος που παρουσιάζεται, μερική και για τις περιπτώσεις όπου το αντικείμενο βρίσκεται μετασχηματισμένο στην εικόνα του ενδιαφέροντος. Ωστόσο, η μέθοδος εξαρτάται σε πολύ μεγάλο βαθμό από την αναπαράσταση του αντικειμένου, εφόσον δεν υπεισέρχεται κάποιος άλλος παράγοντας ενέργειας στην κατάτμηση. Επομένως, στην περίπτωση που το αντικείμενο υπάρχει στην εικόνα πολύ διαφοροποιημένο, ή δεν υπάρχει καθόλου, θα υπάρχει έντονο πρόβλημα στην κατάτμηση. Το πρόβλημα αυτό αντιμετωπίζεται κάπως, με τις μεθόδους που παρουσιάζονται στην επόμενη παράγραφο.

Προγενέστερα Σχήματα και Γεωδαιτικά Ενεργά Περιγράμματα

Οι μέθοδοι που θα περιγράψουμε σε αυτή την παράγραφο, κάνουν χρήση του προγενέστερου σχήματος σε συνδυασμό με το μοντέλο των ενεργών γεωδαιτικών περιγραμμάτων. Περιληπτικά αναφέρουμε ότι η αναπαράσταση του

αντικειμένου η οποία έχει δημιουργηθεί, χρησιμοποιείται είτε ως εξωτερική δύναμη η οποία κατευθύνει την εξέλιξη του γεωδαιτικού ενεργού περιγράμματος, είτε ενσωματώνεται στο μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων, ως κάποιος παράγοντας ενέργειας.

Παρά το γεγονός ότι η μοντελοποίηση του προς αναζήτηση αντικειμένου στις δημοσιεύσεις [38] και [19] είναι η ίδια, ο τρόπος χρήσης του προγενέστερου σχήματος είναι διαφορετικός. Στο [18] έχοντας το προγενέστερο σχήμα:

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

αναζητούνται μόνο τα βάρη w , και οι παράμετροι p θεωρούνται αδιάφορες, ή καλύτερα, θεωρείται ότι το αντικείμενο θα βρίσκεται στην ίδια μορφή στις εικόνες του συνόλου εκπαίδευσης και στην τελική εικόνα. Θεωρείται ότι η χρήση του μετασχηματισμού προσημασμένης απόστασης αρκεί προκειμένου να αντιμετωπιστούν οι τοπικές διακυμάνσεις μεταξύ του μέσου σχήματος και του αντικειμένου που βρίσκεται στην τελική εικόνα. Η ιδέα της τεχνικής που θα παρουσιαστεί, είναι η δημιουργία μιας δύναμης έτσι ώστε να έλκει την επιφάνεια προς την κατεύθυνση του “εκ των υστέρων” υπολογισμένου προγενέστερου σχήματος. Οι παράμετροι w δίνουν το σχήμα που αναζητείται, και οι παράμετροι p δίνουν τον προσανατολισμό και την θέση του σχήματος που αναζητείται.

Προκειμένου να γίνει χρήση μιας τέτοιας δύναμης, σε συνδυασμό με τις ήδη υπάρχουσες δυνάμεις ομαλότητας, μπαλονιού και ελατηρίου που συναντώνται στις γεωδαιτικές ενεργές καμπύλες, πρέπει σε κάθε στιγμή να υπολογίζονται οι παράμετροι βάρους w , οι οποίοι δίνουν την τελική εκτίμηση του σχήματος που αναζητείται. Για την εύρεση των παραμέτρων, έχουμε:

$$\langle w_{MAP}, p_{MAP} \rangle = \arg \max_{w, p} P(w, p | u, \nabla I) \quad (2.17)$$

Στην εξίσωση αυτή, u είναι η εξελισσόμενη επιφάνεια, της οποίας η ισοσταθμική καμπύλη μηδενικού δυναμικού δίνει το αντικείμενο, σε μια χρονική στιγμή t , και I είναι η φωτεινότητα της εικόνας που περιέχει το αντικείμενο. Σύμφωνα με την προσέγγιση των Leventon et al. ([19]), οι παράμετροι w και p δίνουν ανά πάσα στιγμή την τελική εκτίμηση για το αντικείμενο. Επομένως, η επίλυση της παραπάνω σχέσης σχετίζεται με τη σχέση:

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

Με χρήση του κανόνα του Bayes, η σχέση (2.17) αναπτύσσεται και γίνεται:

$$\begin{aligned} P(w, p | u, \nabla I) &= \frac{P(u, \nabla I | w, p) P(w, p)}{P(u, \nabla I)} \\ &= \frac{P(u | w, p) P(\nabla I | w, p, u) P(w) P(p)}{P(u, \nabla I)} \end{aligned} \quad (2.18)$$

Παρόλου που δεν θα μας απασχολήσουν οι παράμετροι p , σημειώνουμε ότι γενικά θεωρούνται ανεξάρτητες από τις παραμέτρους w , και για το λόγο αυτό είναι δυνατή η διάσπαση κάποιων πιθανοτήτων σε γινόμενο πιθανοτήτων.

Ο όρος $P(u|w, p)$ της σχέσης (2.18) υπολογίζει την πιθανότητα εμφάνισης μιας εξελισσόμενης καμπύλης u , δεδομένης της τελικής εκτίμησης u^* (ή αλλιώς των παραμέτρων (w, p)). Η σχέση αυτή δεν περιλαμβάνει καμία πληροφορία σχετικά με την εικόνα. Αν υποθεθεί ότι η εξελισσόμενη καμπύλη αρχικοποιείται στο εσωτερικό του αντικειμένου, τότε είναι αναμενόμενο η καμπύλη να βρίσκεται συνέχεια στο εσωτερικό του αντικειμένου. Μια τέτοια καμπύλη δηλαδή, είναι πιο πιθανή από κάποια καμπύλη που βρίσκεται στο εξωτερικό του αντικειμένου, ολικά ή μερικά. Επομένως, μπορούμε να μοντελοποιήσουμε αυτή την πληροφορία με κατασκευή μιας συνάρτησης πυκνότητας πιθανότητας τύπου Laplace, χρησιμοποιώντας τον όγκο της u ο οποίος βρίσκεται εκτός της u^* :

$$P(u|w, p) = \exp(-V_{outside})$$

Στο [19] υποστηρίζεται ότι η σχέση μεταξύ $\|\nabla I\|$ και u^* είναι Gaussian φύσεως. Επομένως, θεωρώντας ότι η συνάρτηση $h(u^*)$ είναι η συνάρτηση που προσαρμόζεται βέλτιστα στα δεδομένα $(u^*, \|\nabla I\|)$, ο δεύτερος όρος της σχέσης (2.18) μπορεί να υπολογιστεί ως:

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

Οι παράμετροι w , υποτίθεται ότι ακολουθούν ένα Gaussian μοντέλο, και επομένως η πιθανότητά του μπορεί να βρίσκεται από τη σχέση:

$$P(w) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_k|}} \exp\left(-\frac{1}{2} w^T \Sigma_k^{-1} w\right)$$

Οι παράμετροι p που δίνουν τις παραμέτρους του αφινικού μετασχηματισμού που είναι δυνατόν να έχει υποστεί το αντικείμενο στην τελική εικόνα, δεν μας αφορούν, και επομένως θεωρούμε ότι έχουν ομοιόμορφη κατανομή $P(p) = U(-\infty, \infty)$.

Οι όροι που υπολογίστηκαν δίνουν την “εκ των υστέρων” εκτίμηση του σχήματος και του προσανατολισμού-θέσης, απ’ όπου μπορεί να προκύψει η τελική καμπύλη κατάτμησης του αντικειμένου. Οι ποσότητες είναι δυνατόν να υπολογίζονται μόνο σε μια στενή ζώνη γύρω από την εξελισσόμενη καμπύλη. Οι παράμετροι είναι δυνατόν να υπολογιστούν με κατάβαση κλίσης στην λογαριθμική πιθανότητα της σχέσης (2.18). Μετά την εύρεση των παραμέτρων, και τον υπολογισμό της u^* , γίνεται χρήση αυτής για την καθοδήγηση της τελικής κατάτμησης. Συγκεκριμένα, όπως είναι γνωστό, η γεωδαιτική ενεργή καμπύλη εξελίσσεται υπό την επίδραση τριών δυνάμεων. Οι Leventon et al. προτείνουν η εξέλιξη της καμπύλης να κατευθύνεται και από τη διαφορά μεταξύ

της εξελισσόμενης επιφάνειας, και της υπολογισμένης βέλτιστης επιφάνειας u^* , σύμφωνα με τη σχέση:

$$u(t+1) = u(t) + \underbrace{\lambda_1(g(c+x)\|\nabla u(t)\| + \nabla u(t)\nabla g)}_{\text{κλασσική εξέλιξη γεωδαιτικής καμπύλης}} + \underbrace{\lambda_2(u^*(t) - u(t))}_{\text{εξέλιξη με βάση το βέλτιστο σχήμα}}$$

Οι παράμετροι λ_1 και λ_2 δίνουν κατάλληλα βάρη στις δύο συνιστώσες εξέλιξης, και εξαρτώνται από το πόση πίστη έχει κάποιος στο μοντέλο το οποίο έχει κατασκευάσει.

Γενικά, η προσέγγιση που παρουσιάζεται στο [19], δεν λαμβάνει υπόψη της τοπικές διαφορές, ή την έλλειψη ευθυγράμμισης, κατά την κατασκευή του μοντέλου. Ωστόσο, η χρήση των γεωδαιτικών ενεργών καμπυλών και η ταυτόχρονη εξέλιξή τους από τις κλασσικές δυνάμεις και από τη δύναμη που βασίζεται στο μοντέλο, το καθιστά ιδιαίτερα εύρωστο, αφού με κατάλληλη επιλογή των βαρών μεταξύ των τρόπων εξέλιξης, είναι δυνατόν τόσο να εντοπίζεται το προγενέστερο σχήμα, όσο και οι μικρές διαφοροποιήσεις αυτού.

Στο [32] ακολουθείται μια παρόμοια προσέγγιση. Το αντικείμενο αναπαρίστανται με τη χρήση της μέσης επιφάνειας, η οποία έχει υπολογιστεί όπως περιγράφεται στην προηγούμενη ενότητα, και αποτελεί μια επιφάνεια που είναι μετασχηματισμός προσημασμένης απόστασης. Το μοντέλο είναι δυνατόν να χρησιμοποιηθεί κατά τρόπο παρόμοιο με τη μέθοδο που περιγράφηκε για την ευθυγράμμιση των επιφανείων του [30].

Έτσι, έχοντας το μέσο σχήμα Φ_M , και την εξελισσόμενη επιφάνεια Φ , αναζητείται ο βέλτιστος μετασχηματισμός A , προκειμένου να ελαχιστοποιηθεί η ενέργεια:

$$E(\Phi, A) = \int_{\Omega} \int_{\Omega} (s\Phi - \Phi_M(A))^2 d\Omega$$

Ο αλγόριθμος κατάβασης κλίσης, δίνει τις ακόλουθες διαφορικές εξισώσεις εξέλιξης:

$$\begin{aligned} \frac{d}{dt}\Phi &= -2s(s\Phi - \Phi_M(A)) \\ \frac{d}{dt}s &= 2 \int_{\Omega} \int_{\Omega} (s\Phi - \Phi_M(A))(-\Phi + \nabla\Phi_M(A)\frac{\partial}{\partial s}A) d\Omega \\ \frac{d}{dt}\theta &= 2 \int_{\Omega} \int_{\Omega} (s\Phi - \Phi_M(A))(\nabla\Phi_M(A)\frac{\partial}{\partial\theta}A) d\Omega \end{aligned}$$

$$\frac{d}{dt} \begin{bmatrix} T_x \\ T_y \end{bmatrix} = 2 \int_{\Omega} (s\Phi - \Phi_M(A)) (\nabla \Phi_M(A) \frac{\partial}{\partial \begin{bmatrix} T_x \\ T_y \end{bmatrix}} A) d\Omega$$

Οι τρεις τελευταίοι όροι δίνουν τις παραμέτρους του αφινικού μετασχηματισμού, και ο πρώτος όρος την εξέλιξη της επιπεδοσυνάρτησης, με βάση τον όρο του προγενέστερου σχήματος.

Μια πιθανή εξέλιξη του τρόπου χρήσης είναι να γίνεται πρώτα ανάλυση σε κύριες συνιστώσες με βάση το μέσο σχήμα που έχει υπολογιστεί, και οι συντελεστές των κύριων συνιστωσών να υπόκεινται επίσης σε βελτιστοποίηση, όπως συνέβαινε και στα [19] και [38]. Μια άλλη εξέλιξη, η οποία μπορεί να αποτελέσει και εισαγωγή στην επόμενη μέθοδο που θα περιγράψουμε, είναι η χρήση της πρώτης διαφορικής εξίσωσης, με το μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων. Έτσι, σε συνδυασμό με τους όρους μπαλονιού, ομαλότητας και ελατηρίου, προστίθεται και η δύναμη του προγενέστερου σχήματος. Επομένως, το τελικό μοντέλο εξέλιξης της γεωδαιτικής ενεργής καμπύλης γίνεται:

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

Παρόμοια γίνεται χρήση και του πιθανοτικού μοντέλου που αναπτύχθηκε στο [29]. Αρχικά, ορίζουμε τις συναρτήσεις $H_\alpha(\Phi)$ και $\delta_\alpha(\Phi)$, με:

$$\delta_\alpha(\Phi) = \begin{cases} 0 & |\Phi| > \alpha \\ \frac{1}{2\alpha} (1 + \cos(\frac{\pi\Phi}{\alpha})) & |\Phi| \leq \alpha \end{cases}$$

$$H_\alpha(\Phi) = \begin{cases} 1 & \Phi > \alpha \\ 0 & \Phi < -\alpha \\ \frac{1}{2} (1 + \frac{\Phi}{\alpha} + \frac{1}{\pi} \sin(\frac{\pi\Phi}{\alpha})) & |\Phi| < \alpha \end{cases}$$

Οι συναρτήσεις αυτές αποτελούν προσεγγίσεις των γνωστών συναρτήσεων Dirac (δ) και Heaviside (H). Πράγματι, καθώς $\alpha \rightarrow 0$ τότε $\delta_\alpha \rightarrow \delta$ και $H_\alpha \rightarrow H$. Επίσης, δεδομένων των ιδιοτήτων των συναρτήσεων επιπεδοσυνόλων, εφόσον έχουμε θεωρήσει ότι στο εσωτερικό της ισοσταθμικής καμπύλης μηδενικού δυναμικού τα σημεία έχουν αρνητικές τιμές, μπορούμε να εξάγουμε και τις ακόλουθες δύο σχέσεις:

$$\{(x, y) \in \Omega : \lim_{\alpha \rightarrow 0^+} [H_\alpha(\Phi((x, y); t))] = 1\} \equiv R(t) \quad (2.19)$$

$$\{(x, y) \in \Omega : \lim_{\alpha \rightarrow 0^+} [\delta_\alpha(\Phi((x, y); t))] = 1\} \equiv \partial R(t) \quad (2.20)$$

Παρατηρούμε επομένως, ότι με τις συναρτήσεις δ_α, H_α και χρησιμοποιώντας την επιπεδοσυνάρτηση Φ , μπορούμε να υπολογίζουμε εύκολα σε κάθε χρονική στιγμή, τα σημεία τα οποία ανήκουν στο εσωτερικό της καμπύλης και άρα στο

αντικείμενο που κατατμίζεται, και τα σημεία τα οποία ανήκουν στο σύνορο του αντικειμένου, δηλαδή στην καμπύλη. Όπως έχουμε αναφέρει στις προηγούμενες παραγράφους, οι υπολογισμοί είναι δυνατόν να γίνονται σε μια στενή ζώνη γύρω από την καμπύλη, με αποτέλεσμα να ελαττώνεται πολύ η πολυπλοκότητα. Το ρόλο του εύρους ζώνης της στενής ζώνης, παίζει η παράμετρος α . Επομένως, χρησιμοποιώντας τις συναρτήσεις που εισάγαμε, οι διαφορικές εξισώσεις εξέλιξης που θα προκύψουν, θα είναι ήδη κατά το δυνατόν βελτιστοποιημένες από άποψη πολυπλοκότητας υλοποίησης.

Η αναζήτηση του ιδανικού μετασχηματισμού, έχοντας ως βάση το πιθανοτικό μοντέλο που αναπτύχθηκε στο [29], έχει ως βάση την ιδέα ότι κάθε τιμή της τρέχουσας αναπαράστασης θα αντιστοιχίζεται στην πιο πιθανή τιμή του μοντέλου:

$$\begin{cases} (x, y) \rightarrow A(x, y) \\ \max_{x,y} \{p_A^M(x, y)(s\Phi(x, y))\} \forall (x, y) : H_\alpha(\Phi(x, y)) \geq 0 \end{cases}$$

Ο πιο πιθανός μετασχηματισμός είναι αυτός που προκύπτει από τη μεγιστοποίηση της πιθανοφάνειας όλων των pixels. Θεωρώντας ότι οι πυκνότητες είναι ανεξάρτητες μεταξύ των pixels, η ελαχιστοποίηση της $-\log$ συνάρτησης της μέγιστης πιθανοφάνειας, μπορεί να αποτελέσει ένα κριτήριο για βελτιστοποίηση. Έτσι, με άγνωστες παραμέτρους τις Φ και A , σχηματίζεται το ακόλουθο συναρτησιακό ενέργειας:

$$\begin{aligned} E(\Phi, A) &= - \int \int_{\Omega} H_\alpha(\Phi(x, y)) \log \left[p_{A(x,y)}^M(s\Phi(x, y)) \right] dx dy \\ &= \int \int_{\Omega} H_\alpha(\Phi(x, y)) \left[\log(\sigma_M(A(x, y))) + \frac{(s\Phi(x,y) - \Phi_M(A(x,y)))^2}{2\sigma_M^2(A(x,y))} \right] dx dy \end{aligned}$$

Αναζητούμε ένα μετασχηματισμό A και μια επιπεδοσυνάρτηση Φ , έτσι ώστε να μεγιστοποιείται για κάθε pixel η εκ των υστέρων πιθανότητα, δεδομένου του πιθανοτικού μοντέλου. Έχοντας το συναρτησιακό ενέργειας, και θεωρώντας ότι ο μετασχηματισμός A μπορεί στη γενική περίπτωση να αναπαρασταθεί από N παραμέτρους $\{\alpha_1, \alpha_2, \dots, \alpha_N\}$, μπορούμε να βρούμε τις Euler-Lagrange παραγώγους, και να βρούμε τις διαφορικές εξισώσεις εξέλιξης:

$$\begin{aligned} \frac{d}{dt} \Phi &= -s H_\alpha(\Phi) \left[\frac{(s\Phi - \Phi_M(A))}{\sigma_M^2(A)} \right] - \\ &\quad - \delta_\alpha(\Phi) \left[\log(\sigma_M(A)) + \frac{(s\Phi - \Phi_M(A))^2}{2\sigma_M^2(A)} \right] \end{aligned}$$

$$\begin{aligned} \frac{d}{dt} s &= -2 \int \int_{\Omega} H_\alpha(\Phi) \left[\frac{1}{2\sigma_M(A)} \nabla \sigma_M(A) \frac{\partial}{\partial s} (A_x, A_y) + \frac{(s\Phi - \Phi_M(A))^2 [\nabla \sigma_M(A) \frac{\partial}{\partial s} (A_x, A_y)]}{\sigma_M^3(A)} \right] - \\ &\quad - \frac{(s\Phi - \Phi_M(A)) (\Phi - \nabla \Phi_M(A) \frac{\partial}{\partial s} (A_x, A_y))}{\sigma_M^2(A)} \end{aligned}$$

$$\frac{d}{dt} \alpha_j = -2 \int_{\Omega} H_{\alpha}(\Phi) \left[\frac{1}{2\sigma_M(A)} \nabla \sigma_M(A) \frac{\partial}{\partial \alpha_j} (A_x, A_y) - \frac{(s\Phi - \Phi_M(A)) [\nabla \Phi_M(A) \frac{\partial}{\partial s} (A_x, A_y)]}{\sigma_M^2(A)} - \frac{(s\Phi - \Phi_M(A))^2 \left[\nabla \sigma_M(A) \frac{\partial}{\partial \alpha_j} (A_x, A_y) \right]}{\sigma_M^3(A)} \right]$$

Το συναρτησιακό ενέργειας που προτάθηκε μπορεί να ενσωματωθεί σε ένα κλασσικό μοντέλο κατάτμησης, όπως αυτό των γεωδαιτικών ενεργών καμπυλών, ως μια επιπλέον εξωτερική ενέργεια. Έτσι, είναι δυνατόν να προκύψει ένα ολικό μοντέλο εξέλιξης της καμπύλης, το οποίο εξαρτάται τόσο από τις κλασσικές δυνάμεις, όσο και από το προγενέστερο σχήμα:

$$\frac{d}{dt} \Phi = \alpha \delta_{\alpha}(\Phi) [g(I)x - \nabla g(I) \nabla \Phi] + b H_{\alpha}(\Phi) s \frac{(s\Phi - \Phi_M(A))}{\sigma_M^2(A)} - b \delta_{\alpha}(\Phi) \left[\log(\sigma_M(A) + \frac{(s\Phi - \Phi_M(A))^2}{2\sigma_M^2(A)}) \right]$$

Η επιπεδοσυνάρτηση Φ εξελίσσεται υπό την επίδραση δυνάμεων που δρουν κατά την κατεύθυνση του κάθετου διανύσματος, τοπικά. Η πρώτη δύναμη συρικνώνει το εξελισσόμενο μέτωπο προς τα όρια του αντικειμένου, και η δεύτερη δύναμη εξελίσσει το μέτωπο προς τέτοια κατεύθυνση, έτσι ώστε η κατάτμηση που προκύπτει να πληρεί τους περιορισμούς που εισάγει η χρήση του προγενέστερου σχήματος. Παρατηρούμε ότι η δύναμη του προγενέστερου σχήματος δεν εξαρτάται μόνο από τα σημεία κοντά στην καμπύλη, αλλά από όλα τα σημεία στο εσωτερικό της καμπύλης. Γενικά, η προσέγγιση κατά την οποία η καμπύλη εξελίσσεται με βάση πληροφορία από την επιφάνεια που αυτή περικλύει, αποτελεί την βασική ιδέα πίσω από το μοντέλο που ονομάζεται Γεωδαιτικές Ενεργές Επιφάνειες, και που εισήχθη από τους Paragios N. και Deriche R., προκειμένου να επιλυθούν κάποια από τα προβλήματα που αντιμετωπίζουν οι γεωδαιτικές ενεργές καμπύλες. Το εν λόγω μοντέλο δεν θα μας απασχολήσει στη διπλωματική αυτή, ωστόσο είναι αυτό που προτείνουν οι συγγραφείς στο [29]. Για το λόγο αυτό, στη διαφορική εξίσωση εξέλιξης που προτείνουν οι συγγραφείς στην εν λόγω δημοσίευση, περιλαμβάνονται και κάποιοι επιπλέον όροι που σχετίζονται με πληροφορία των κατατμηθέντων επιφανειών της εικόνας, και οι οποίοι δεν θα μας απασχολήσουν.

Γενικά, το μοντέλο που αναπτύχθηκε είναι ιδιαίτερα ικανοποιητικό διότι έχει την ικανότητα να λαμβάνει υπόψη του τόσο μικρές διαφοροποιήσεις μεταξύ των σχημάτων που παρουσιάζονται, όσο και να μπορεί να βρει με ακρίβεια το επιθυμητό αντικείμενο. Αυτό οφείλεται τόσο στο γεγονός ότι χρησιμοποιείται ένα πιθανοτικό μοντέλο για το προγενέστερο σχήμα, όσο και το γεγονός ότι το μοντέλο αυτό εισάγει μια ενέργεια που μπορεί να χρησιμοποιηθεί απ'ευθείας με το μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων.

Συνοψίζοντας, παρουσιάστηκαν τέσσερις εναλλακτικές λύσεις του προβλήματος κατάτμησης. Όλες κάνουν χρήση προγενέστερων σχημάτων, και

προσπαθούν, χρησιμοποιώντας ένα προγενέστερο σχήμα που κατασκευάζουν, να κατευθύνουν την κατάτμηση έτσι ώστε να βρεθεί το επιθυμητό αντικείμενο. Η κάθε μέθοδος έχει μειονεκτήματα και πλεονεκτήματα απέναντι στις άλλες, και για αυτό έγινε προσπάθεια αντιπαράθεσης μεταξύ τους, στάδιο-στάδιο. Ωστόσο, το βασικό μειονέκτημα που αντιμετωπίζουν όλες οι μέθοδοι, είναι ότι μπορούν να εντοπίσουν μόνο ένα αντικείμενο του συγκεκριμένου είδους στην τελική εικόνα. Έτσι, αν στην τελική εικόνα υπάρχουν πολλά αντικείμενα ίδια με το προγενέστερο σχήμα, οι μέθοδοι που παρουσιάσαμε αντιμετωπίζουν προβλήματα.

2.4 Εναλλακτικές Προσεγγίσεις στο Πρόβλημα των Προγενέστερων Σχημάτων

Στη διπλωματική αυτή θα κάνουμε χρήση των επιπεδοσυνόλων για να περιγράψουμε τα προγενέστερα σχήματα και αυτό διότι μπορούν να χρησιμοποιηθούν σε συνδυασμό με τις γεωδαιτικές ενεργές καμπύλες. Ωστόσο, κρίνεται ενδιαφέρον να γίνει μια σύντομη αναφορά σε κάποιες μεθόδους οι οποίες ακολουθούν τελείως διαφορετική προσέγγιση, και έχουν αρκετά ικανοποιητικά αποτελέσματα.

Στο [31] προτείνεται μια καινούρια μέθοδος κατάτμησης εικόνας με βάση προγενέστερα σχήματα. Με τη μέθοδο αυτή, απαιτείται μόνο μια απεικόνιση του επιθυμητού αντικειμένου, και στη συνέχεια, χωρίς να αναζητούνται pixel-προς-pixel αντιστοιχίες, το αντικείμενο αυτό εντοπίζεται στην εικόνα. Επίσης, τα σύνορα του αντικειμένου εντοπίζονται πολύ σωστά. Η κατάτμηση επιτυγχάνεται με χρήση προβολικών μετασχηματισμών, οι παράμετροι των οποίων επίσης υπολογίζονται από τον προτεινόμενο αλγόριθμο. Η μέθοδος είναι ιδιαίτερα αποτελεσματική, και η καινοτομία της έγκειται στη χρήση ενός μόνο αντικειμένου αναφοράς, καθώς και στην χρήση καινούριων τεχνικών ευθυγράμμισης.

Στο [9] γίνεται προσπάθεια δημιουργίας στατιστικών μοντέλων από παραμορφώσιμα αντικείμενα, όπως για παράδειγμα από έναν κινούμενο άνθρωπο. Αναφέρεται ότι τέτοιου είδους αντικείμενα είναι αυτά τα οποία παρατηρούνται σε πραγματικές συνθήκες, και είναι αυτά τα οποία απαιτούνται από εφαρμογές οι οποίες παρακολουθούν κινούμενα αντικείμενα. Η προσέγγιση βασίζεται στο γεγονός ότι η αλληλουχία των παραμορφωμένων αντικειμένων εμφανίζεται με κάποια συγκεκριμένη σειρά, ή καλύτερα, κάθε στιγμιότυπο του αντικειμένου εμφανίζεται μετά από κάποιο άλλο με κάποια πιθανότητα. Τα αντικείμενα ενσωματώνονται ως προγενέστερα σχήματα σε ένα Bayesian μοντέλο που βασίζεται σε κατάτμηση εικόνων με χρήση επιπεδοσυνόλων. Στη συγκεκριμένη εργασία γίνεται προσπάθεια αντιμετώπισης και του προβλήματος εύρεσης πολ-

λαπλών όμοιων αντικειμένων σε μια εικόνα, με χρήση των προγενέστερων σχημάτων.

Στο [18] γίνεται χρήση προγενέστερων πιθανοτήτων καμπυλότητας και φωτεινότητας. Συγκεκριμένα, θεωρείται ότι οι τιμές τις επιπεδοσυνάρτησης που προκύπτει από τον μετασχηματισμό προσημασμένης απόστασης σχετίζονται μεταξύ τους, ιδίως σε μικρές γειτονίες. Το ίδιο συμβαίνει και για τις τιμές της φωτεινότητας της εικόνας. Οι πιθανότητες αυτές υπολογίζονται για τα σχήματα του συνόλου εκπαίδευσης, και στη συνέχεια χρησιμοποιούνται κατάλληλα κατά τη διαδικασία κατάτμησης, προκειμένου να βρεθεί το αντικείμενο που αναζητείται.

Κεφάλαιο 3

Παρακολούθηση Κίνησης

3.1 Εισαγωγή

Στο κεφάλαιο αυτό θα παρουσιάσουμε τρόπους με τους οποίους είναι δυνατόν να βρεθεί και να παρακολουθηθεί η κίνηση σε μια εικόνα. Καταρχήν, θα αναφερθούμε στο πρόβλημα εύρεσης της Οπτικής Ροής (Optical Flow), η οποία μας επιτρέπει να βρίσκουμε την ταχύτητα κίνησης κάθε pixel μιας εικόνας. Θα δούμε πώς είναι δυνατόν η μέθοδος αυτή να χρησιμοποιηθεί για παρακολούθηση κινούμενων αντικειμένων. Στη συνέχεια του κεφαλαίου θα παρουσιάσουμε μεθόδους παρακολούθησης κίνησης, άλλες εκ των οποίων παρακολουθούν αντικείμενα μοντελοποιώντας το χρώμα τους, και άλλες οι οποίες κάνουν χρήση στατιστικών παραμέτρων.

3.2 Οπτική Ροή

3.2.1 Εισαγωγή στην Οπτική Ροή

Η διαρκής κίνηση είναι ένα από τα βασικά χαρακτηριστικά του κόσμου μας. Επομένως, τα βιολογικά και μηχανικά συστήματα τα οποία αναπτύσσονται, πρέπει να έχουν την ικανότητα να την αντιληφθούν. Με τον όρο κίνηση, στην Όραση Υπολογιστών, εννοούμε την σχετική κίνηση μεταξύ κάμερας και αντικειμένων. Δεδομένου ότι τα 3-Δ αντικείμενα απεικονίζονται σε 2-Δ σχήματα επί των εικόνων, αντιλαμβανόμαστε ότι και τα διανύσματα της 3-Δ κίνησης του ανθρώπινου κόσμου, θα προβάλλονται σε 2-Δ διανύσματα επί των εικόνων. Το πεδίο που προκύπτει από την προβολή αυτή, ονομάζεται οπτική ροή. Στόχος είναι η όσο το δυνατόν καλύτερη προσέγγιση του 2-Δ αυτού πεδίου διανυσμάτων, δεδομένου ότι μια καλή προσέγγιση αποτελεί τη βάση για την εύρεση της πλήρους 3-Δ κίνησης στον πραγματικό κόσμο.

3.2.2 Μέθοδοι Υπολογισμού της Οπτικής Ροής

Η οπτική ροή σε κάθε σημείο της εικόνας αναπαρίσταται από ένα διάνυσμα (u,v) , όπου u είναι η ταχύτητα κατά την κατεύθυνση x , και v είναι η ταχύτητα κατά την κατεύθυνση y . Για τον υπολογισμό των διανυσμάτων αυτών, έχουν αναπτυχθεί αρκετοί διαφορετικοί αλγόριθμοι. Στην επόμενη παράγραφο θα παρουσιάσουμε γενικά τις διαφορετικές προσεγγίσεις, και στη συνέχεια θα επικεντρωθούμε σε ορισμένες από αυτές.

Γενικές Προσεγγίσεις Εύρεσης Οπτικής Ροής

Μια ανασκόπηση και αξιολόγηση πολλών διαφορετικών προσεγγίσεων για την εύρεση της οπτικής ροής από δύο διαδοχικές εικόνες, γίνεται από τους Barron et al. στο [3]¹. Πολλές από τις τεχνικές αυτές, αν και έχουν διαφορές μεταξύ τους, ακολουθούν τα ακόλουθα τρία στάδια:

1. Φιλτράρισμα ή ομαλοποίηση με βαθυπερατό/ζωνοπερατό φίλτρο, έτσι ώστε να αναδειχθεί η δομή ενδιαφέροντος στο σήμα και να αυξηθεί ο σηματοθορυβικός (SNR) λόγος.
2. Διεξαγωγή βασικών μετρήσεων, όπως οι χωρικές και χρονικές παράγωγοι και τοπικές συσχετίσεις επιφανειών.
3. Χρήση των μετρήσεων προκειμένου να εξαχθεί το 2-Δ διανυσματικό πεδίο. Στο στάδιο αυτό γίνονται υποθέσεις περί ομαλότητας του πεδίου.

Γενικά, οι τεχνικές εύρεσης της οπτικής ροής, μπορούν να ταξινομηθούν σε κατηγορίες, ανάλογα με τη φιλοσοφία υπό την οποία προσεγγίζουν το πρόβλημα:

- Διαφορικές τεχνικές: Οι διαφορικές τεχνικές υπολογίζουν την ταχύτητα από χωρο-χρονικές παραγώγους της φωτεινότητας της εικόνας, κάνοντας χρήση βαθυπερατών ή ζωνοπερατών φίλτρων. Βασική υπόθεση αυτών των μεθόδων είναι ότι η φωτεινότητα ενός pixel δεν μεταβάλλεται λόγω μετατόπισης σε μικρά χρονικά διαστήματα, διότι οι ταχύτητες δεν θεωρούνται μεγάλες. Αντιπροσωπευτικά παραδείγματα αυτής της προσέγγισης αποτελούν οι μέθοδοι των Horn και Schunck, των Lukas και Kanade, του Nagel και των Uras, Giroso, Verri και Torre.
- Τεχνικές αντιστοίχισης περιοχών: Η ακριβής εύρεση των παραγώγων με μεθόδους διαφορικής ανάλυσης είναι δυνατόν να μην είναι πρακτική, λόγω παραγόντων όπως ο μικρός αριθμός των frames. Σε αυτές τις

¹Η παράγραφος αυτή έχει βασιστεί κυρίως στο [3].

περιπτώσεις, η μέθοδος αντιστοίχισης περιοχών δίνει καλύτερα αποτελέσματα. Η ταχύτητα v θεωρείται ως ένα διάνυσμα μετατόπισης $\vec{d} = (d_x, d_y)$, το οποίο δίνει την καλύτερη ταύτιση μεταξύ περιοχών της εικόνας, σε διαφορετικές χρονικές στιγμές. Η εύρεση της καλύτερης ταύτισης έγγειται στην μεγιστοποίηση ενός μέτρου ομοιότητας μεταξύ των εικόνων, όπως του τετραγωνικού σφάλματος ή της κανονικοποιημένης ετεροσυσχέτισης. Την προσέγγιση αυτή ακολουθούν για παράδειγμα ο Anandan και ο Singh.

- Τεχνικές που βασίζονται στην ενέργεια: Οι μέθοδοι της κατηγορίας αυτής βασίζονται στην ενέργεια εξόδου φίλτρων τα οποία είναι ρυθμισμένα κατάλληλα ως προς την ταχύτητα. Συγκεκριμένα, κάνουν χρήση του γεγονότος ότι η ενέργεια του Fourier μετασχηματισμού της εικόνας βρίσκεται σε ένα επίπεδο που εξαρτάται από την ταχύτητα. Οι τεχνικές αυτές, σε κάποιες ειδικές περιπτώσεις, είναι ισοδύναμες με διαφορικές τεχνικές ή τεχνικές που χρησιμοποιούν την ετεροσυσχέτιση μεταξύ εικόνων για την εύρεση της οπτικής ροής. Για περισσότερες πληροφορίες μπορεί κανείς να ανατρέξει στην εργασία του Heeger.
- Τεχνικές που βασίζονται στη φάση: Η τέταρτη κατηγορία μεθόδων χαρακτηρίζονται ως βασιζόμενες στη φάση, διότι η ταχύτητα ορίζεται σε σχέση με την συμπεριφορά φάσης κάποιων ζωνοπερατών φίλτρων. Η πρώτη γενικευμένη χρήση αυτής της τεχνική έγινε από τους Fleet και Jepson, ενώ χρησιμοποιήθηκε επίσης από τους Waxman, Wu και Bergholm.

Στο [3], γίνεται μαθηματική ανάλυση των παραπάνων μεθόδων. Στη συνέχεια, προγραμματίζονται και αξιολογούνται αναδεικνύοντας τελικά τα πλεονεκτήματα και τα μειονεκτήματα της κάθε μιας. Γενικά, οι καλύτερες μέθοδοι θεωρούνται οι διαφορικές και οι τεχνικές που βασίζονται στη φάση.

Από τις πρώτες και πιο σημαντικές μεθόδους ήταν οι μέθοδοι που αναπτύχθηκαν από τις ομάδες των Lukas-Kanade, και Horn-Schunck. Οι δύο αυτοί μέθοδοι χρησιμοποιούν διαφορικές τεχνικές, και παρουσιάζουν αρκετά καλά αποτελέσματα. Στη συνέχεια, θα παρουσιάσουμε αναλυτικά τις δύο αυτές μεθόδους, καθώς και μια τρίτη μέθοδο που τις ενσωματώνει και ως εκ τούτου διαθέτει τα πλεονεκτήματα και των δύο.

Μέθοδος των Lukas B. D. και Kanade T.

Οι Lukas Bruce D. και Kanade Takeo παρουσίασαν στο 1981 μια καινούρια μέθοδο προκειμένου να επιτυγχάνεται ευθυγράμμιση σημάτων. Αρχικά ανέπτυξαν τη μέθοδό τους για μονοδιάστατα σήματα, και στη συνέχεια εξέλιξαν τη μέθοδο για να χειρίζεται σήματα πολλών διαστάσεων. Η παρουσίαση των

μεθόδων αυτών γίνεται στο [20], μαζί με μια χρήση της μεθόδου για εύρεση βάθους από εικόνες. Η μέθοδος που ανέπτυξαν είναι δυνατόν να χρησιμοποιηθεί για την εύρεση της οπτικής ροής, δεδομένου ότι επιθυμούμε να βρούμε τις μετατοπίσεις των “όμοιων” pixels, ανάμεσα σε δύο διαδοχικές εικόνες.

Έστω ότι σε κάποια χρονική t , η φωτεινότητα της εικόνας σε κάποιο pixel (x,y) είναι ίση με $I(x,y,t)$. Η φωτεινότητα οφείλεται στην προβολή ενός αντικειμένου, στον 2-Δ χώρο. Υποθέτουμε ότι μετά από χρόνο Δt , η προβολή του αντικειμένου έχει μετατοπιστεί κατά $(\Delta x, \Delta y)$ αλλά η φωτεινότητα δεν έχει αλλάξει. Γίνεται δηλαδή η υπόθεση ότι:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) \approx I(x, y, t) \quad (3.1)$$

Για μικρές μετατοπίσεις $(\Delta x, \Delta y)$ σε μικρό χρονικό διάστημα Δt , μπορούμε να αναπτύξουμε τον όρο $I(x + \Delta x, y + \Delta y, t + \Delta t)$ κατά Taylor, παίρνοντας:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) \approx I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t$$

Εισάγοντας τις ταχύτητες $u \equiv \frac{dx}{dt}$ και $v \equiv \frac{dy}{dt}$ στην παραπάνω εξίσωση, και χρησιμοποιώντας την εξίσωση (3.1) καταλήγουμε στην εξίσωση:

$$\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} = 0$$

Η παραπάνω εξίσωση μπορεί να γραφεί με χρήση του gradient ως:

$$\nabla I \cdot (u, v) = -\frac{\partial I}{\partial t} \quad (3.2)$$

Από την εξίσωση αυτή είναι δυνατόν να υπολογιστεί η συνιστώσα της ταχύτητας η οποία είναι κάθετη στις ισοσταθμικές καμπύλες της εικόνας. Ωστόσο, για την εύρεση και των δύο συνιστωσών, απαιτείται μια επιπλέον εξίσωση.

Οι Lukas και Kanade στο [20], λύνουν το πρόβλημα θεωρώντας ότι η ταχύτητα είναι σταθερή γύρω από ένα κεντρικό pixel \vec{x} . Τότε, η επίλυση της εξίσωσης (3.2) ισοδυναμεί με την ελαχιστοποίηση του συναρτησιακού:

$$\sum_{\vec{x} \in \Omega} W^2(\vec{x}) [\nabla I(\vec{x}, t) \vec{v} + I_t(\vec{x}, t)]^2 \quad (3.3)$$

όπου το σύνολο Ω είναι το σύνολο των σημείων της εικόνας στα οποία η ταχύτητα θεωρείται σταθερή, μια γειτονιά δηλαδή του \vec{x} . Ο πίνακας βαρών $W(\vec{x})$ χρησιμοποιείται για να δώσει περισσότερη έμφαση στα pixels του κέντρου της γειτονιάς απ' ό,τι στα pixels της περιφέρειας.

Η σχέση (3.3) αποτελεί ένα πρόβλημα βέλτιστοποίησης, και η λύση του δίνει:

$$\begin{bmatrix} I_x(\vec{x}_1) & \dots & I_x(\vec{x}_n) \\ I_y(\vec{x}_1) & \dots & I_y(\vec{x}_n) \end{bmatrix} \begin{bmatrix} W^2(\vec{x}_1) & 0 & \dots & 0 \\ 0 & W^2(\vec{x}_2) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & W^2(\vec{x}_n) \end{bmatrix} \begin{bmatrix} I_x(\vec{x}_1) & I_y(\vec{x}_1) \\ \vdots & \vdots \\ I_x(\vec{x}_n) & I_y(\vec{x}_n) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \\ \begin{bmatrix} I_x(\vec{x}_1) & \dots & I_x(\vec{x}_n) \\ I_y(\vec{x}_1) & \dots & I_y(\vec{x}_n) \end{bmatrix} \begin{bmatrix} W^2(\vec{x}_1) & 0 & \dots & 0 \\ 0 & W^2(\vec{x}_2) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & W^2(\vec{x}_n) \end{bmatrix} \begin{bmatrix} I_t(\vec{x}_1) \\ \vdots \\ I_t(\vec{x}_n) \end{bmatrix}$$

ή αλλιώς $A^T W^2 A v = A^T W^2 b$, όπου τα μεγέθη A , W , b προκύπτουν με απλή επισκόπηση της παραπάνω εξίσωσης. Ο πίνακας $A^T W^2 A$ είναι πίνακας 2×2 και συγκεκριμένα:

$$\begin{bmatrix} \sum_{\vec{x} \in \Omega} W^2(\vec{x}) I_x^2(\vec{x}) & \sum_{\vec{x} \in \Omega} W^2(\vec{x}) I_x(\vec{x}) I_y(\vec{x}) \\ \sum_{\vec{x} \in \Omega} W^2(\vec{x}) I_x(\vec{x}) I_y(\vec{x}) & \sum_{\vec{x} \in \Omega} W^2(\vec{x}) I_y^2(\vec{x}) \end{bmatrix}$$

Επομένως, προκειμένου να βρεθεί η ταχύτητα v σε μια γειτονιά ενός κεντρικού pixel x_i , αρκεί ο πίνακας $A^T W^2 A$ να είναι αντιστρέψιμος. Για βελτιστοποίηση τόσο από άποψη ταχύτητας όσο και από άποψη αποτελεσμάτων, είναι δυνατόν να γίνεται εκτέλεση του αλγορίθμου σε χώρους κλίμακας, σύμφωνα και με όσα έχουμε περιγράψει στο 2ο Κεφάλαιο, στην παράγραφο (2.2.1).

Ο αλγόριθμος Lukas-Kanade είναι τοπικός, δηλαδή προσπαθεί να ελαχιστοποιήσει μια ενέργεια η οποία υπολογίζεται τοπικά. Έχει το πλεονέκτημα ότι είναι εύρωστος απέναντι στο θόρυβο. Ωστόσο, δεν είναι δυνατόν να δώσει πυκνά πεδία ταχυτήτων, εφόσον θεωρείται ότι η ταχύτητα είναι σταθερή σε μια γειτονιά ενός κεντρικού σημείου.

Μέθοδος των Horn B. K. P. και Schunck B. G.

Οι Horn B. K. P. και Schunck B. G., παρουσιάζουν στο [14] έναν διαφορετικό τρόπο εύρεσης των συνιστωσών της ταχύτητας. Σε συνδυασμό με την εξίσωση (3.2), απαιτούν η ταχύτητα να μεταβάλλεται ομαλά σε ολόκληρη την εικόνα. Επομένως, η ενέργεια των παραγώγων της ταχύτητας:

$$E_s = \int \int_{\Omega} [(u_x^2 + u_y^2) + (v_x^2 + v_y^2)] dx dy$$

πρέπει να είναι ελάχιστη. Επίσης, ελάχιστο πρέπει να είναι και το σφάλμα από την μη ικανοποίηση της εξίσωσης (3.2), δηλαδή πρέπει να ελαχιστοποιείται το συναρτησιακό:

$$E_c = \int_{\Omega} (I_x u + I_y v + I_t)^2 dx dy$$

Τελικά, συνδυάζοντας τα δύο συναρτησιακά E_c και E_s κατασκευάζεται ένα πρόβλημα βελτιστοποίησης, και η οπτική ροή υπολογίζεται ελαχιστοποιώντας το συναρτησιακό:

$$J(u, v) = E_c + \lambda E_s \quad (3.4)$$

Η παράμετρος λ εκφράζει πόσο σημαντική είναι η ομαλότητα του πεδίου ταχυτήτων, σε σχέση με την ικανοποίηση της εξίσωσης (3.2).

Το συναρτησιακό (3.4) ελαχιστοποιείται όταν οι συνιστώσες u και v της ταχύτητας ικανοποιούν το σύστημα των μερικών διαφορικών εξισώσεων Euler (από το [1]):

$$\begin{aligned} \nabla^2 u &= \lambda (I_x u + I_y v + I_t) I_x \\ \nabla^2 v &= \lambda (I_x u + I_y v + I_t) I_y \end{aligned}$$

Στο [14] προτείνεται και ο αλγόριθμος υπολογισμού των συνιστωσών u και v . Ο αλγόριθμος θα παρουσιαστεί στο 4ο κεφάλαιο, όπου θα αναφερθούμε στην υλοποίηση του συστήματος παρακολούθησης κινούμενων αντικειμένων που κατασκευάστηκε για την διπλωματική αυτή.

Η μέθοδος Horn-Schunck είναι “καθολική” (global), δηλαδή προσπαθεί να ελαχιστοποιήσει ένα συναρτησιακό το οποίο υπολογίζεται σε όλη την εικόνα. Έχει το πλεονέκτημα ότι μέσω της ομαλοποίησης του πεδίου των ταχυτήτων, δίνει πυκνά πεδία οπτικής ροής. Συγκεκριμένα, οι ταχύτητες στο εσωτερικό ενός αντικειμένου (ή οπουδήποτε $\nabla I \approx 0$), μπορούν και υπολογίζονται από τις ταχύτητες του αντικειμένου στα όριά του. Το βασικό μειονέκτημα της μεθόδου είναι ότι είναι αρκετά ευαίσθητη στο θόρυβο, ενώ δεν έχει ικανοποιητικά αποτελέσματα όταν τα αντικείμενα στην εικόνα κινούνται με ταχύτητα μεγαλύτερη από μερικά pixels ανά frame.

Μέθοδος των Bruhn et al.

Οι μέθοδοι των Lukas-Kanade και Horn-Schunck παρουσιάζουν πλεονεκτήματα και μειονεκτήματα τα οποία παρουσιάστηκαν συνοπτικά στις παραπάνω παραγράφους. Μάλιστα, τα μειονεκτήματα της μιας μεθόδου, αποτελούν πλεονεκτήματα της άλλης και το αντίστροφο. Επομένως, αν μπορούσαν οι δύο αυτοί μέθοδοι να συνδυαστούν, τότε θα μπορούσαν να αναπτυχθούν αλγόριθμοι που να έχουν τα πλεονεκτήματα και των δύο μεθόδων, και επομένως πολύ καλύτερα αποτελέσματα, δίνοντας πυκνά πεδία οπτικής ροής, ομαλά, και έχοντας μεγάλη αντοχή στο θόρυβο.

Μια τέτοια καινούρια συνδυαστική μέθοδος παρουσιάζεται από τους Bruhn A., Weickert J. και Schnoerr C., στο [5]. Το μοντέλο που κατασκευάζεται, επεκτείνεται έτσι ώστε να χρησιμοποιεί πληροφορία και από το πεδίο του

χρόνου (spatio-temporal approach) κατά την ομαλοποίηση και τις παραγωγίσεις. Επίσης, το πρόβλημα μετατρέπεται σε μη-γραμμικό, με τη χρήση μιας μη-γραμμικής συνάρτησης κόστους στο συναρτησιακό ενέργειας. Τέλος, προκειμένου να είναι δυνατή η χρήση του μοντελου και για περιπτώσεις όπου οι ταχύτητες u και v δεν είναι μικρές (και άρα η γραμμικοποίηση στην εξίσωση (3.1) δεν ισχύει), γίνεται επίλυση σε χώρους κλίμακας, υιοθετώντας μια coarse-to-fine στρατηγική.

Αν K_ρ είναι ο Gaussian πυρήνας με διακύμανση ρ , τότε εισάγοντας τους συμβολισμούς:

$$\begin{aligned} w &:= (u, v, 1)^T \\ \|\nabla w\|^2 &:= \|\nabla u\|^2 + \|\nabla v\|^2 \\ \nabla_3 I &:= (f_x, f_y, f_t)^T \\ J_\rho(\nabla_3 I) &:= K_\rho * (\nabla_3 I \nabla_3 I^T) \end{aligned}$$

οι εξισώσεις (3.3) και (3.4) που δίνουν την τοπική ενέργεια των Lukas-Kanade και την ολική ενέργεια Horn-Schunck αντίστοιχα, μπορούν να ξαναγραφούν ως:

$$\begin{aligned} E_{LK}(w) &= w^T J_\rho(\nabla_3 I) w \\ E_{HS}(w) &= \int_{\Omega} (w^T J_0(\nabla_3 I) w + \alpha \|\nabla w\|^2) dx dy \end{aligned}$$

Παρατηρείται πλέον ότι οι δύο ενέργειες αυτές, είναι δυνατόν να συνδυαστούν σε ένα ολικό-τοπικό συναρτησιακό ενέργειας:

$$E_{CLG}(w) = \int_{\Omega} (w^T J_\rho(\nabla_3 I) w + \alpha \|\nabla w\|^2) dx dy$$

Το συναρτησιακό που προέκυψε μοιάζει αρκετά με την ενέργεια που εισήγαγαν οι Horn-Schunck, και η ελαχιστοποίησή του βρίσκοντας τις παραγώγους Euler-Lagrange δεν αποτελεί δυσκολότερο πρόβλημα (σύμφωνα με το [5]).

Η συνάρτηση ενέργειας που δημιουργήθηκε παραπάνω, βασίζεται μόνο σε χωρική πληροφορία. Εισάγοντας τον χρόνο ως τρίτη διάσταση, οι αλλαγές είναι ελάχιστες:

$$E_{CLG3}(w) = \int_{\Omega_x[0,T]} (w^T J_\rho(\nabla_3 I) w + \alpha \|\nabla_3 w\|^2) dx dy dt \quad (3.5)$$

Συγκεκριμένα, οι μόνες αλλαγές που απαιτούνται είναι η επέκταση του ∇w στις τρεις διαστάσεις ($\|\nabla_3 w\|^2 := \|\nabla_3 u\|^2 + \|\nabla_3 v\|^2$), και η 3- Δ (2- Δ χώρος και χρόνος) συνέλιξη με τις Gaussians (για τον υπολογισμό του παράγοντα J_ρ). Οι Gaussians μπορούν να έχουν διαφορετική κλίμακα στο χώρο και στο χρόνο, λόγω της διαφορετικής ερμηνείας των δύο περιοχών.

Οι παράγωγοι Euler-Lagrange της ενέργειας (3.5) είναι:

$$\begin{aligned}\nabla_3^2 u - \frac{1}{\alpha}(J_{11}u + J_{12}v + J_{13}) &= 0 \\ \nabla_3^2 v - \frac{1}{\alpha}(J_{21}u + J_{22}v + J_{23}) &= 0\end{aligned}$$

όπου J_{ij} δηλώνει το τμήμα του σύνθετου πίνακα $J_\rho(\nabla_3 I)$. Το παραπάνω σύστημα διαφορικών εξισώσεων έχει παρόμοια μορφή και στην περίπτωση της χωρικής προσέγγισης.

Οι Bruhn et al. προτείνουν την εισαγωγή μιας μη γραμμικής συνάρτησης, προκειμένου το συναρτησιακό της ενέργειας να γίνει πιο εύρωστο απέναντι σε ασυνέχειες των δύο όρων του συναρτησιακού ενέργειας. Η συνάρτηση που εισάγεται είναι η:

$$\psi_i(s^2) = 2\beta_i^2 \sqrt{1 + \frac{s^2}{\beta_i^2}}$$

οπότε το συναρτησιακό ενέργειας γίνεται:

$$E_{CLG3}(w) = \int_{\Omega_x[0,T]} (\psi_1(w^T J_\rho(\nabla_3 I)w) + \alpha \psi_2(\|\nabla_3 w\|^2)) dx dy dt$$

Οι παράγωγοι Euler-Lagrange του παραπάνω συναρτησιακού είναι:

$$\begin{aligned}0 &= \operatorname{div}(\psi_2(\|\nabla_3 w\|^2)\nabla_3 u) - \frac{1}{\alpha}\psi_1(w^T J_\rho(\nabla_3 I)w)(J_{11}u + J_{12}v + J_{13}) \\ 0 &= \operatorname{div}(\psi_2(\|\nabla_3 w\|^2)\nabla_3 v) - \frac{1}{\alpha}\psi_1(w^T J_\rho(\nabla_3 I)w)(J_{11}v + J_{12}u + J_{13})\end{aligned}$$

Η μη γραμμικότητα του παραπάνω συστήματος διαφορικών εξισώσεων εξαρτάται από την επιλογή της παραμέτρου β . Συγκεκριμένα:

$$\lim_{\beta \rightarrow +\infty} (\psi_i(s^2)) = \lim_{\beta \rightarrow +\infty} \left(\frac{1}{\sqrt{1 + \frac{s^2}{\beta_i^2}}} \right) \approx 1$$

δηλαδή για μεγάλες τιμές της παραμέτρου β , το πρόβλημα μετατρέπεται σε γραμμικό.

Οι μέθοδοι εύρεσης της οπτικής ροής των Lukas-Kanade και Horn-Schunk αλλά και πολλών άλλων, χρησιμοποιούν ως βάση τους την υπόθεση (3.1). Η ισχύς της σχέσης αυτής προϋποθέτει μικρές τιμές για τα u και v , και αυτό δεν είναι δυνατόν να εξασφαλιστεί για οποιαδήποτε κίνηση. Το μειονέκτημα αυτό μπορεί να αρθεί αν εφαρμοστούν τεχνικές χώρου κλίμακας. Οι Bruhn et al., δεν χρησιμοποιούν το πεδίο που προκύπτει από τις χαμηλές κλίμακες, ως αρχικοποίηση της μεθόδου στις υψηλές κλίμακες. Τα αποτελέσματα της

κάθε κλίμακας χρησιμοποιούνται διορθωτικά στις ήδη υπάρχοντες προβλέψεις (warping), προκειμένου οι ανώτερες κλίμακες να βρουν το επιθυμητό πεδίο με όσο το δυνατόν μικρότερο σφάλμα. Η μέθοδος αυτή παρουσιάζεται επίσης στο [4], όπου γίνεται υπολογισμός της οπτικής ροής εισάγοντας και τον περιορισμό της σταθερότητας της κλίσης της εικόνας σε συνδυασμό με την (3.1).

Αν δw^m είναι η διόρθωση του πεδίου στην κλίμακα m , με $m = 0$ και $w^0 = (0, 0, 0)^T$ να είναι η αρχική κλίμακα και η αρχικοποίηση αντίστοιχα, τότε το δw^m προκύπτει από την ελαχιστοποίηση της ενέργειας:

$$E_{CLG3}(w) = \int_{\Omega_x[0,T]} (\psi_1((\delta w^m)^T J_\rho(\nabla_3 I(p + w^m)) \delta w^m) + \alpha \psi_2(\|\nabla_3(w^m + \delta w^m)\|^2)) dp$$

όπου $w^{m+1} = w^m + \delta w^m$ και $p = (x, y, t)$. Οι χώροι κλίμακας μπορούν να δημιουργηθούν είτε ομαλοποιώντας με Gaussians διαφορετικής κλίμακας, είτε, όπως στη συγκεκριμένη περίπτωση, με υποδειγματοληψία της εικόνας.

Για να ελαχιστοποιηθεί η ενέργεια, βρίσκονται οι παράγωγοι Euler-Lagrange, η μορφή των οποίων δεν διαφέρει σημαντικά από αυτή που είχαν για τις προηγούμενες προσεγγίσεις. Στο [5] παρατίθεται και ένας αλγόριθμος υπολογισμού των παραμέτρων u και v , με βάση τη μέθοδο που έχει περιγραφεί.

Γενικά, οι Bruhn et al. συνδύασαν δύο κλασσικές μεθόδους υπολογισμού της οπτικής ροής, και, μετατρέποντας το πρόβλημα σε μη-γραμμικό, και με επίλυση του σε διαδοχικούς χώρους κλίμακας, κατάφεραν να δημιουργήσουν μια καινούρια μέθοδο που παρουσιάζει τα προτερήματα και των δύο αρχικών μεθόδων, δίνοντας πολύ καλύτερα αποτελέσματα. Ωστόσο, η μη γραμμικοποίηση του προβλήματος, καθώς και η ενσωμάτωση όλων των περιορισμών, έχει αρνητικό αντίκτυπο στην ταχύτητα σύγκλισης. Επίσης, η επίλυση σε χώρους κλίμακας, αν και βοηθάει την αποφυγή των τοπικών ελαχίστων και οδηγεί σε καλύτερα αποτελέσματα, οδηγεί σε αύξηση της πολυπλοκότητας.

3.3 Παρακολούθηση Κινούμενων Αντικειμένων

Το αντικείμενο της συγκεκριμένης διπλωματικής, είναι η παρακολούθηση κινούμενων αντικειμένων. Έχοντας παρουσιάσει μεθόδους εύρεσης επιθυμητών αντικειμένων σε εικόνες, και μεθόδους εντοπισμού της κίνησης, μπορούμε πλέον να παρουσιάσουμε διάφορες μεθόδους παρακολούθησης κινούμενων αντικειμένων. Άλλες μέθοδοι κάνουν χρήση κάποιων εργαλείων που έχουμε ήδη περιγράψει, όπως επιπεδοσύνολα και γεωδαιτικά ενεργά περιγράμματα, ενώ άλλες ακολουθούν εντελώς διαφορετικές προσεγγίσεις. Κάποιοι συγγραφείς παρακολουθούν αντικείμενα μοντελοποιώντας την υφή τους, ενώ άλλοι χρησιμοποιούν στατιστική και πιθανότητες προκειμένου να εντοπίσουν τα κινούμενα αντικείμενα.

3.3.1 Χρήση Οπτικής Ροής και Ενεργών Περιγραμμάτων

Τα ενεργά περιγράμματα είναι δυνατόν να χρησιμοποιηθούν μαζί με την οπτική ροή, για την παρακολούθηση των κινούμενων αντικειμένων σε μια εικόνα. Η τεχνική αυτή προτείνεται στο [12], αλλά αποτελεί έναν εύλογο συνδυασμό δύο υπάρχοντων μεθόδων για την παρακολούθηση κίνησης.

Όπως είχαμε περιγράψει στο πρώτο κεφάλαιο, το πρόβλημα της παρακολούθησης κινούμενων αντικειμένων μπορεί να θεωρηθεί διτό, αποτελούμενο από το υποπρόβλημα της εύρεσης του επιθυμητού αντικειμένου, και από το πρόβλημα παρακολούθησης της κίνησης αυτού του αντικειμένου. Στην εργασία [12], κάθε πρόβλημα αντιμετωπίζεται ξεχωριστά.

Αρχικά, το εργαλείο των ενεργών περιγραμμάτων χρησιμοποιείται προκειμένου να εντοπιστεί το ιπτάμενο αντικείμενο. Δεν γίνεται αναφορά στον τρόπο με τον οποίο εντοπίζεται το ιπτάμενο αντικείμενο, και αγνοούνται τα υπόλοιπα αντικείμενα στην εικόνα, οπότε υποθέτουμε ότι το ιπτάμενο αντικείμενο είναι το μοναδικό στην εικόνα, είτε η αρχικοποίηση της καμπύλης περικλύει μόνο το αντικείμενο του ενδιαφέροντος μας. Επίσης, στη συγκεκριμένη εργασία, τελικά γίνεται χρήση των γεωδαιτικών ενεργών περιγραμμάτων, και όχι των ενεργών περιγραμμάτων όπως αρχικά αναφέρεται, και αυτό διότι εισάγεται η παραμετροποίηση ως προς το μήκος τόξου της καμπύλης. Σε κάθε περίπτωση, τελικά εντοπίζεται το επιθυμητό αντικείμενο.

Δεδομένων δύο διαδοχικών εικόνων υπολογίζεται η οπτική ροή, που αποτελεί όπως έχουμε αναφέρει την προβολή των 3-Δ ταχυτήτων στη 2-Δ εικόνα. Δεδομένου ότι το σύνορο του αντικειμένου έχει υπολογιστεί στο προηγούμενο στάδιο, είναι δυνατόν να χρησιμοποιηθεί η οπτική ροή προκειμένου να μετατοπίσει κατάλληλα την καμπύλη, κοντά στο σύνορο του αντικειμένου στην επόμενη εικόνα. Επομένως, αντί η εξέλιξη της καμπύλης με στόχο την ελαχιστοποίηση της ενέργειάς της να γίνεται εξαρχής, η καμπύλη τοποθετείται κοντά στο ελάχιστο της ενέργειας και έτσι απαιτούνται λιγότερες επαναλήψεις.

Ουσιαστικά, η καθ'αυτόν τον τρόπο αντιμετώπιση του προβλήματος παρακολούθησης κινούμενων αντικειμένων, έχει πολύ μεγάλη συνάφεια με την προσέγγιση που ακολουθείται στον αυτόματο έλεγχο. Δεδομένου μιας αρχικής καμπύλης C_i που αποτελεί το πραγματικό σύνορο του αντικειμένου, βρίσκεται με την πληροφορία της οπτικής ροής η καμπύλη \hat{C}_{i+1} , η οποία εξελίσσεται έτσι να ελαχιστοποιηθεί το σφάλμα μεταξύ C_{i+1} και \hat{C}_{i+1} . Στα προβλήματα κλειστού βρόχου, γίνεται προσπάθεια να ελαχιστοποιηθεί η διαφορά μεταξύ πρόβλεψης από το σύστημα ανοικτού βρόχου, και του επιθυμητού σήματος που επιβάλλει η ανάδραση.

Στο [12] παρουσιάζεται και ένας τρόπος υπολογισμού της οπτικής ροής σε χώρους κλίμακας. Η προσέγγισή μας απέναντι στο πρόβλημα παρακολούθησης κινούμενων αντικειμένων είναι παρόμοια και θα αναλυθεί στο τέταρτο κεφά-

λαιο. Συνοπτικά, αναφέρουμε ότι πέραν της οπτικής ροής και των γεωδαιτικών ενεργών καμπυλών, γίνεται χρήση και προγενέστερων σχημάτων, ενώ παρουσιάζονται ιδέες με χρήση των οποίων ο αλγόριθμος είναι δυνατόν να αγνοεί τα στάσιμα αντικείμενα κατά την εξέλιξη της καμπύλης.

3.3.2 Εξελίσσοντας το Μοντέλο των Γεωδαιτικών Ενεργών Περιγραμμάτων

Το πρόβλημα παρακολούθησης κινούμενων αντικειμένων από εικόνες που λαμβάνονται από ένα σταθερό παρατηρητή αντιμετωπίζεται στο [28]. Οι Paragios N. και Deriche R., εξελίσσουν το μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων, έτσι ώστε να μπορεί να χρησιμοποιηθεί για την παρακολούθηση κινούμενων αντικειμένων, χωρίς να απαιτείται κάποιου είδους πρότερη γνώση και εξωτερική πληροφορία για την κίνηση. Εξάγοντας πληροφορία για την κίνηση των αντικειμένων από την εικόνα, επεκτείνουν το συναρτησιακό ενέργειας των γεωδαιτικών ενεργών περιγραμμάτων έτσι ώστε να εντοπίζονται τα κινούμενα αντικείμενα. Η μέθοδος που παρουσιάζεται ανήκει στο σύνολο των μεθόδων που κάνουν χρήση των ακμών για τον εντοπισμό αντικειμένων σε εικόνες. Επίσης, παρουσιάζεται ένας τρόπος ώστε ο εντοπισμός του κινούμενου αντικειμένου και η παρακολούθησή του, να γίνεται σε ένα στάδιο.

Αρχικά υπολογίζεται η διαφορά μεταξύ δύο διαδοχικών εικόνων $D(s; t) = I(s; t) - I(s; t - 1)$. Ένα pixel ανήκει στο σταθερό μέρος της εικόνας αν αποτελεί μέρος του φόντου και στις δύο εικόνες, ενώ στο κινητό μέρος, αν ανήκει σε κινούμενο αντικείμενο, σε κάποια από τις δύο εικόνες. Στο κάθε pixel του D μπορεί να αποδοθεί μια πιθανότητα, χρησιμοποιώντας τις δεσμευμένες πιθανότητες να ανήκει στο στατικό ή στο κινητό μέρος. Εισάγεται η συνάρτηση πυκνότητας πιθανότητας (σ.π.π):

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

όπου P_L είναι οι προγενέστερες πιθανότητες της υπόθεσης L και $p_L(d)$ είναι η δεσμευμένη σ.π.π υπό την υπόθεση L , όπου $L \in \{S:\text{στατικό}, M:\text{κινητό}\}$. Οι σ.π.π $p_L(d)$ θεωρείται ότι είναι Laplacian ή Gaussian φύσεως.

Το μοντέλο αυτό χρησιμοποιείται στη συνέχεια προκειμένου να βρεθεί η πιθανότητα ένα pixel να βρίσκεται στο σύνορο της κινούμενης περιοχής. Λαμβάνονται υπόψη πιθανοί διαχωρισμοί των επιφανειών με τέσσερις προσανατολισμούς διεπιφάνειας. Τελικά, σχηματίζεται μια εικόνα I_D που περιέχει πληροφορία σχετικά με τα όρια της κίνησης. Η εικόνα αυτή χρησιμοποιείται σε ένα συναρτησιακό ενέργειας παρόμοιο με αυτό που χρησιμοποιείται στα γεωδαιτικά ενεργά περιγράμματα:

$$E[C(p)] = \int_0^1 g(I_D(C(p))) \|\dot{C}(p)\| dp$$

Η εξέλιξη της καμπύλης C με στόχο την ελαχιστοποίηση της παραπάνω ενέργειας, έχει ως αποτέλεσμα η καμπύλη C να περικλύει όλες τις κινούμενες περιοχές. Η εξέλιξη της καμπύλης γίνεται με χρήση επιπεδοσυνόλων, και έτσι οι τοπολογικές διαφορές είναι αυτόματα διαχειρίσιμες.

Πρέπει να σημειωθεί ότι με τον όρο κινούμενη περιοχή, εννοούμε τόσο το αντικείμενο, όσο και την περιοχή-φόντο της εικόνας, που καλύπτονταν από αντικείμενο στο προηγούμενο frame. Επομένως, είναι απαραίτητο να υπάρχει και ένα επιπλέον βήμα, το οποίο να βρίσκει ακριβώς τα όρια του αντικειμένου. Αυτό επιτυγχάνεται με την ελαχιστοποίηση του γενικευμένου συναρτησιακού ενέργειας:

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

όπου σ_D και σ_T είναι η κλίμακα των Gaussians που χρησιμοποιούνται σε κάθε συνάρτηση g .

Η καμπύλη που προκύπτει σε ένα frame μπορεί να χρησιμοποιηθεί ως αρχικοποίηση στο επόμενο frame, δεδομένου ότι βρίσκεται πάντα στο εσωτερικό της περιοχής που χαρακτηρίζεται ως κινούμενη. Μόνο στη περίπτωση του πρώτου frame αρχικοποιείται στο σύνορο της εικόνας, και εντοπίζει όλες τις κινούμενες περιοχές και στη συνέχεια όλα τα κινούμενα αντικείμενα.

Η μέθοδος των Paragios et al. έχει το πλεονέκτημα ότι αντιμετωπίζει εννιαία το πρόβλημα παρακολούθησης κινούμενων αντικειμένων, ενώ ταυτόχρονα δεν απαιτεί κάποιου είδους προγενέστερης γνώσης. Τα όρια των αντικειμένων βρίσκονται πολύ ικανοποιητικά με τη στατιστική μέθοδο που προτείνεται, αλλά όχι στην περίπτωση που υπάρχει έντονη πληροφορία υφής κοντά στο κινούμενο αντικείμενο. Η μέθοδος που παρουσιάστηκε παρουσιάζει προβλήματα και στην περίπτωση που οι κινήσεις των αντικειμένων είναι πολύ μεγάλες, και επίσης όταν η αρχική καμπύλη βρίσκεται τόσο στο εξωτερικό όσο και το εσωτερικό των κινούμενων αντικειμένων. Τέλος, στο [28] προτείνεται και ένας καινούριος τρόπος εξέλιξης των επιπεδοσυναρτήσεων, ο οποίος συνδυάζει τις μεθόδους Στενής Ζώνης και Ταχείας Προέλασης, και θα παρουσιαστεί στο επόμενο κεφάλαιο.

3.3.3 Παρακολούθηση Κινούμενων Περιοχών χωρίς Υπολογισμό της Κίνησης

Ο Mansouri A. R. προσπαθεί στο [21] να λύσει το πρόβλημα παρακολούθησης κινούμενων αντικειμένων σε εικόνες που έχουν ληφθεί από τυχαίο πα-

ρατηρητή. Ο παρατηρητής είναι δυνατόν να είναι κινούμενος ή στατικός, και ο τρόπος αντιμετώπισης του προβλήματος είναι εννιαίος. Γίνεται χρήση όσο το δυνατόν λιγότερων περιορισμών στην κίνηση της περιοχής που παρακολουθείται, σε αντίθεση με τις προσεγγίσεις άλλων συγγραφέων.

Η πρωτοτυπία και τα βασικά στοιχεία της προσέγγισης του Mansouri A. R. συνίστανται στα ακόλουθα:

- Δεν γίνεται καμία υπόθεση για το σχήμα της περιοχής που παρακολουθείται, εκτός από γενική ομαλότητα.
- Δεν γίνεται καμία υπόθεση για την φωτεινότητα στα όρια του αντικειμένου.
- Παρακολουθούνται ολόκληρες περιοχές και όχι μόνο τα όρια των αντικειμένων.
- Επιτρέπονται αλλαγές στην τοπολογία της περιοχής υπό παρακολούθηση, λόγω της χρήσης εξελισσόμενων επιπεδοσυναρτήσεων.
- Χρησιμοποιούνται πιθανότητες προκειμένου να μοντελοποιηθούν όλες οι παραδοχές που γίνονται, και να κατασκευαστούν οι διαφορικές εξισώσεις εξέλιξης.

Το πρόβλημα επιλύεται κυρίως με χρήση πιθανοτήτων, και συγκεκριμένα υπολογίζεται η πιθανότητα να ανήκει ένα pixel στην περιοχή που παρακολουθείται. Βασική υπόθεση του αλγορίθμου είναι ότι η φωτεινότητα και το χρώμα μιας περιοχής δεν αλλάζει σημαντικά από frame σε frame σε μια αλληλουχία εικόνων. Οι πιθανότητες P_{in} και P_{out} , δηλαδή οι πιθανότητες να ανήκει ένα pixel στην εξελισσόμενη επιφάνεια ή όχι, χρησιμοποιούνται για την δημιουργία δυνάμεων υπό τις οποίες η επιφάνεια που παρακολουθείται, εξελίσσεται. Με χρήση των πιθανοτήτων αυτών κατασκευάζεται ένα συναρτησιακό ενέργειας και το πρόβλημα της παρακολούθησης της κινούμενης περιοχής μοντελοποιείται ως πρόβλημα βελτιστοποίησης. Η εύρεση των Euler-Lagrange παραγώγων της συνάρτησης ενέργειας δίνει τις διαφορικές εξισώσεις εξέλιξης της επιφάνειας του ενδιαφέροντος. Προκειμένου να είναι αυτόματα διαχειρίσιμες οι τοπολογικές αλλαγές, γίνεται χρήση των επιπεδοσυναρτήσεων, και της θεωρίας που παρουσιάστηκε στην παράγραφο (2.2.3).

Πιο ειδικά, γίνεται προσπάθεια να προσδιοριστεί η περιοχή R_{t+1} , έχοντας ως δεδομένη την περιοχή R_t , και την τρέχουσα και προηγούμενη εικόνα I_{t+1} και I_t αντίστοιχα. Οι τρόποι με τους οποίους είναι δυνατόν η επιφάνεια R_t να εξελιχθεί στην επιφάνεια R_{t+1} ενσωματώνονται στο μοντέλο ως πιθανότητες. Για παράδειγμα, μια προγενέστερη πιθανότητα επί της R_{t+1} διασφαλίζει την ομαλότητα του συνόρου ∂R_{t+1} . Ο τελικός στόχος είναι η εύρεση της καμπύλης

γ^* που διαχωρίζει τις περιοχές κατάλληλα και οδηγεί στην ελαχιστοποίηση της ενέργειας. Τα pixels που βρίσκονται στο εξωτερικό της καμπύλης αλλά με μεγάλη πιθανότητα έπρεπε να βρίσκονται στο εσωτερικό της, δημιουργούν δυνάμεις που εξελίσσουν την καμπύλη και άρα την επιφάνεια, προς τα έξω. Η επιφάνεια δηλαδή επεκτείνεται, έτσι ώστε να τα συμπεριλάβει. Το αντίθετο συμβαίνει με τα pixels τα οποία βρίσκονται στο εσωτερικό της καμπύλης αλλά έχουν μεγάλη πιθανότητα να βρίσκονται εκτός επιφάνειας.

Οι υπολογισμοί των δυνάμεων εξέλιξης είναι δυνατόν να είναι πιο ακριβείς, αν λαμβάνονται υπόψη μόνο τα σημεία τα οποία βρίσκονται μακριά από το εξελισσόμενο σύνορο. Ο λόγος είναι, ότι τα σημεία κοντά στα όρια της επιφάνειας κατατάσσονται εύκολα σε λάθος κατηγορίες. Επίσης, αν υπάρχει γνώση σχετικά με το σχήμα της επιφάνειας που εξελίσσεται ή με το είδος της κίνησης που αυτή εκτελεί, είναι δυνατόν να ενσωματωθεί εύκολα στο μοντέλο, ως όρος πιθανότητας. Με τον ίδιο τρόπο είναι δυνατόν να ενσωματωθούν κάθε είδους περιορισμοί, με αποτέλεσμα η μέθοδος να θεωρείται εύκολα επεκτάσιμη. Τέλος, ο αλγόριθμος που προτείνεται είναι πολύ αποδοτικός, ακόμα και για περιπτώσεις κινούμενου φόντου.

Το βασικό μειονέκτημα της μεθόδου που παρουσιάστηκε είναι ότι δεν κάνει εντοπισμό του κινούμενου αντικειμένου αυτόματα, εννοώντας ότι απαιτεί ένα συγκεκριμένο βήμα αρχικοποίησης. Πρέπει δηλαδή, στο πρώτο frame, να προσδιοριστεί χειροκίνητα η περιοχή που θέλουμε να παρακολουθήσουμε. Αυτό καθιστά την μέθοδο ημιαυτόματη. Επίσης, είναι απαραίτητο να είναι από πριν γνωστό το μέγιστο εύρος της κίνησης μεταξύ των εικόνων. Αυτός ο περιορισμός ωστόσο δεν είναι ιδιαίτερα αυστηρός, δεδομένου ότι πάντα το εύρος της κίνησης περιορίζεται από τις διαστάσεις της εικόνας. Τέλος, στη δημοσίευση [21], αν και παρουσιάζεται πληθώρα πολύ ικανοποιητικών αποτελεσμάτων, δεν γίνεται καμία αναφορά στην πολυπλοκότητα του αλγορίθμου, ούτε και σύγκρισή του με υπάρχουσες μεθόδους. Ενδέχεται λόγω της πληθώρας υπολογισμών που απαιτούνται, να είναι αρκετά πιο αργός από πολλές μεθόδους της βιβλιογραφίας.

3.3.4 Παρακολούθηση Αντικειμένων με Μοντελοποίηση Χρώματος

Μια διαφορετική προσέγγιση στο πρόβλημα της παρακολούθησης κινούμενων αντικειμένων αποτελεί η παρακολούθηση αντικειμένων με μοντελοποίηση του χρώματός τους. Οι μέθοδοι που ακολουθούν την προσέγγιση αυτή δημιουργούν μια αναπαράσταση του χρώματος του αντικειμένου που θέλουν να εντοπίσουν. Στη συνέχεια, αναζητούνται στην εικόνα τα pixels που θεωρείται ότι ανήκουν στο επιθυμητό αντικείμενο, υπόκεινται σε κατάλληλη ομαδο-

ποίηση, και παρακολουθούνται στην αλληλουχία των εικόνων.

Αυτή η προσέγγιση ακολουθείται στα [36], [37] και [2]. Στα [36] και [37] των Starner et al., γίνεται προσπάθεια ερμηνείας ενός υποσυνόλου της αμερικάνικης νοηματικής γλώσσας μέσω οπτικής παρακολούθησης των ανθρώπινων χειρονομιών, και αναγνώρισης αυτών. Δεδομένου ότι απαιτείται ένας αλγόριθμος παρακολούθησης χειρονομιών ο οποίος να λειτουργεί κοντά σε πραγματικό χρόνο, επιλέχθηκε η προσέγγιση της μοντελοποίησης του χρώματος. Πιο συγκεκριμένα, στο [36] το πρόβλημα απλοποιείται πάρα πολύ διότι ο χρήστης φοράει κατάλληλα χρωματισμένα γάντια και αναζητείται στην εικόνα το χρώμα των γαντιών. Στο [37] γίνεται μοντελοποίηση του ανθρώπινου δέρματος με χρήση ενός μείγματος από Gaussians.

Η παρακολούθηση των ανθρώπινων χειρονομιών είναι το πρόβλημα που αντιμετωπίζεται στο [2] των Argyros et al., και η μοντελοποίηση του ανθρώπινου δέρματος που υιοθετούν παρουσιάζεται αναλυτικά. Χρησιμοποιείται ο χώρος χρώματος YUV, αγνοώντας την παράμετρο Y. Η παράμετρος Y αναπαριστά την φωτεινότητα ενός pixel, και επομένως αγνοώντας τη, πέραν του γεγονότος ότι μειώνεται η πολυπλοκότητα του προβλήματος, το μοντέλο που θα δημιουργηθεί θα έχει μια σχετική ανεκτικότητα στη φωτεινότητα.

Προκειμένου να δημιουργηθεί το μοντέλο του ανθρώπινου δέρματος, γίνεται χρήση ενός συνόλου εκπαίδευσης. Στις εικόνες του συνόλου εκπαίδευσης επισημαίνονται οι περιοχές που αντιστοιχούν σε ανθρώπινο δέρμα, όπως για παράδειγμα χέρια και πρόσωπο. Από τα δεδομένα αυτά υπολογίζονται οι πιθανότητες $P(c|s)$, $P(c)$, $P(s)$ που αντιστοιχούν στην προγενέστερη πιθανότητα ένα χρώμα c να είναι χρώμα δέρματος, στην προγενέστερη πιθανότητα του χρώματος c , και στην προγενέστερη πιθανότητα του χρώματος του ανθρώπινου δέρματος αντίστοιχα. Από αυτές τις τιμές υπολογίζεται η πιθανότητα ένα χρώμα να είναι χρώμα ανθρώπινου δέρματος, σύμφωνα με τον νόμο του Bayes:

$$P(s|c) = \frac{P(c|s)P(s)}{P(c)}$$

Οι πιθανότητες αυτές χρησιμοποιούνται με κατάλληλα κατώφλια στην αλληλουχία των εικόνων που μας ενδιαφέρουν, έτσι ώστε να εξαχθούν οι περιοχές του ανθρώπινου δέρματος. Ενδιαφέρον είναι ότι αν ένα χρώμα βρίσκεται πλησίον περιοχής που αντιστοιχεί σε ανθρώπινο δέρμα, τότε αντιμετωπίζεται ως ανθρώπινο δέρμα με αρκετά μεγαλύτερη πιθανότητα. Επίσης, δεδομένου ότι είναι δυνατόν κατά τη διάρκεια της παρακολούθησης να αλλάζουν οι συνθήκες φωτισμού, οι συγγραφείς κάνουν χρήση και των τελευταίων w εικόνων, προκειμένου να κάνουν το μοντέλο τους πιο εύρωστο. Έτσι, τελικά:

$$P(s|c) = \gamma P(s|c) + (1 - \gamma)P_w(s|c)$$

όπου γ είναι η παράμετρος εμπιστοσύνης και $P_w(s|c)$ οι πιθανότητες που εξάχθηκαν από τα w τελευταία frames.

Τελικά, με εξέταση κάθε pixel της εικόνας, προκύπτουν περιοχές οι οποίες δυνητικά αντιστοιχούν σε περιοχές ανθρώπινου δέρματος. Αρχικά, κάποιες μικρές περιοχές αγνοούνται διότι αποδίδονται σε θόρυβο, και οι περιοχές που απομένουν τιτλοφορούνται (labeled), έτσι ώστε να είναι δυνατός ο διαχωρισμός τους. Στη συνέχεια, σε κάθε περιοχή αντιστοιχίζεται μια έλλειψη, δεδομένου ότι οι ελλείψεις μπορούν να αποδώσουν αρκετά ικανοποιητικά περιοχές όπως το πρόσωπο και τα χέρια. Η κάθε έλλειψη είναι ένα διάνυσμα $h_i = h_i(c_{x_i}, c_{y_i}, \alpha_i, \beta_i, \theta_i)$, όπου (c_{x_i}, c_{y_i}) είναι το κέντρο της έλλειψης, α_i, β_i το μήκος του πρωτεύοντος και του δευτερεύοντος άξονα αντίστοιχα, και θ_i ο προσανατολισμός της έλλειψης στο 2-Δ επίπεδο της εικόνας. Η απόσταση ενός pixel από μια έλλειψη, χρησιμοποιείται προκειμένου να βρεθεί η περιοχή στην οποία πρέπει να αντιστοιχηθεί το συγκεκριμένο pixel.

Οι περιπτώσεις που πρέπει λαμβάνει υπόψη της η μέθοδος είναι:

- Μια καινούρια περιοχή εμφανίζεται στην εικόνα. Στην περίπτωση αυτή τα pixels της είναι εκτός κάθε ήδη υπάρχουσας περιοχής, και επομένως πρέπει να υπολογιστούν οι παράμετροι μιας καινούριας έλλειψης και να δοθεί ένας καινούριος τίτλος.
- Μια υπάρχουσα περιοχή συνεχίζει να υπάρχει και στο επόμενο frame. Στην περίπτωση αυτή, ο αλγόριθμος πρέπει να διαχειρίζεται αποτελεσματικά τις επικαλύψεις, αλλά και τις περιπτώσεις που μια έλλειψη δεν μπορεί να καλύψει πλήρως όλη την περιοχή του δέρματος.
- Μια υπάρχουσα περιοχή εξαφανίζεται από την εικόνα. Στην περίπτωση αυτή, η περιοχή δε διαγράφεται αυτόματα, αλλά δίνεται ένα χρονικό περιθώριο μισού δευτερολέπτου (περίπου δεκατέσσερα frames για την ανάλυση των πειραμάτων), για να αντιμετωπίζεται και η περίπτωση λανθασμένης καταχώρησης των pixels.

Προκειμένου να υπάρχει κάποια πρόβλεψη για τη θέση των ελλείψεων στα επόμενα frames, χρησιμοποιείται πληροφορία κίνησης από τις προηγούμενες εικόνες με βάση ένα γραμμικό μοντέλο.

Ο αλγόριθμος που αναπτύχθηκε μπορεί να λειτουργεί σε πραγματικό χρόνο και να αντιμετωπίζει αποτελεσματικά τις περιπτώσεις επικάλυψης περιοχών, εμφάνισης νέων, και αλλαγής φωτισμού. Επίσης, μπορεί να χρησιμοποιηθεί ικανοποιητικά ακόμα και στην περίπτωση γρήγορων κινήσεων. Τέλος, η μοντελοποίηση των χεριών και του προσώπου με ελλείψεις, μπορεί να κάνει απλή την επέκταση της μεθόδου έτσι ώστε να γίνεται και αναγνώριση των χειρονομιών. Γενικά, η έλλειψη, αν και δεν είναι η πιο περίτεχνη μοντελοποίηση του ανθρώπινου χεριού και προσώπου, είναι ικανοποιητική ως μοντελοποίηση διότι μπορεί να βρεθεί πολύ εύκολα, ενώ για την περιγραφή της απαιτούνται μόλις πέντε τιμές.

3.3.5 Παρακολούθηση Περιοχών με Εύρεση Αφινικών Μετασχηματισμών

Στην παράγραφο (2.3.1) αναφερθήκαμε σε μεθόδους ευθυγράμμισης ενός συνόλου από εικόνες. Μια παρόμοια τεχνική είναι δυνατόν να χρησιμοποιηθεί στο πρόβλημα παρακολούθησης κινούμενων αντικειμένων, για την παρακολούθηση μιας συγκεκριμένης περιοχής. Η προσέγγιση αυτή ακολουθείται στο [13], όπου θεωρείται ότι κάθε εικόνα της αλληλουχίας προκύπτει από τον μετασχηματισμό της αρχικής εικόνας. Επομένως, αρκεί να βρεθούν οι παράμετροι του μετασχηματισμού και αυτό μπορεί να γίνει ελαχιστοποιώντας μια συνάρτηση σφάλματος. Επίσης, στο μοντέλο που αναπτύσσεται, ενσωματώνεται πληροφορία φωτισμού και έτσι τα αποτελέσματα δεν είναι ευαίσθητα στις μεταβολές φωτισμού. Τέλος, η μέθοδος που προτείνεται μπορεί να παρακολουθήσει σωστά τις περιοχές ενδιαφέροντος, ακόμα και όταν υπάρχουν επικαλύψεις (occlusions).

Στο [13] των Hager et al., ο τυχαίος μετασχηματισμός f μεταξύ της αρχικής εικόνας και της τρέχουσας, έχει ως παραμέτρους τις $\mu(t) = (\mu_1(t), \mu_2(t), \dots, \mu_n(t))^T$. Οι ιδανικές παράμετροι $\mu^*(t) = (\mu_1^*(t), \mu_2^*(t), \dots, \mu_n^*(t))^T$ είναι αυτές για τις οποίες:

$$I(x, t_0) = I(f(x; \mu^*(t)), t) \forall x \in \Omega$$

όπου Ω είναι η περιοχή που παρακολουθούμε. Στόχος είναι η εύρεση των βέλτιστων παραμέτρων μ έτσι ώστε να ελαχιστοποιείται το άθροισμα των τετραγωνικών διαφορών μεταξύ των εικόνων:

$$O(\mu) = \sum_{x \in \Omega} (I(f(x; \mu), t) - I(x, t_0))^2$$

Δεδομένου ότι η κίνηση είναι μια διαδικασία που χαρακτηρίζεται από συνέχεια των μεγεθών που την περιγράφουν, είναι δυνατόν οι παράμετροι μ σε κάθε χρονική στιγμή, να υπολογίζονται από τις προηγούμενες παραμέτρους μ , μαζί με κάποια αλλαγή $\delta\mu$. Έτσι τελικά, το ζητούμενο είναι η εύρεση του $\delta\mu$, έτσι ώστε να ελαχιστοποιείται η συνάρτηση:

$$O(\delta\mu) = \|I(\mu(t) + \delta\mu, t + \tau) - I(0, t_0)\|^2$$

όπου $I(\mu, t) = [I(f(x_1; \mu), t) I(f(x_2; \mu), t) \dots I(f(x_N; \mu), t)]^T$, με N να είναι τα σημεία της περιοχής Ω . Με χρήση κατάλληλων γραμμικοποιήσεων, ανάλογων της υπόθεσης (3.1), προκύπτει ως τελική συνάρτηση προς ελαχιστοποίηση η:

$$O(\delta\mu) \approx \|M\delta\mu + I(\mu, t + \delta\tau) - I(0, t_0)\|^2$$

από όπου προκύπτει αν απαιτηθεί $\nabla O = 0$:

$$\delta\mu = -(M^T M)^{-1} M^T [I(\mu, t + \tau) - I(0, t_0)]$$

υπό την προϋπόθεση ότι ο $M^T M$ είναι πλήρους κλίμακας, όπου $M = M(\mu, t) = [I_{\mu_1}(\mu, t) I_{\mu_2}(\mu, t) \dots I_{\mu_n}(\mu, t)]$. Επομένως, οι παράμετροι μ υπολογίζονται με βάση την σχέση:

$$\mu(t+1) = \mu(t) - (M^T M)^{-1} M^T [(I(\mu, t + \tau) - I(0, t_0))]$$

Στο [13] παρουσιάζεται και ένας αποδοτικός αλγόριθμος υπολογισμού του πίνακα M . Συγκεκριμένα, αποδεικνύεται ότι ο M μπορεί, σε κάθε χρονική στιγμή, να γραφτεί ως το γινόμενο ενός σταθερού πίνακα $N_{xk} M_0$, και ενός χρονομεταβλητού πίνακα $\Sigma(\mu)$:

$$\begin{aligned} M(\mu, t) &= \begin{bmatrix} \nabla_f I(f(x_1; \mu), t)^T f_\mu(x_1; \mu) \\ \nabla_f I(f(x_2; \mu), t)^T f_\mu(x_2; \mu) \\ \vdots \\ \nabla_f I(f(x_N; \mu), t)^T f_\mu(x_N; \mu) \end{bmatrix} \\ &= \begin{bmatrix} \nabla_x I(x_1; t_0)^T f_x(x_1; \mu)^{(-1)} f_\mu(x_1; \mu) \\ \nabla_x I(x_2; t_0)^T f_x(x_2; \mu)^{(-1)} f_\mu(x_2; \mu) \\ \vdots \\ \nabla_x I(x_N; t_0)^T f_x(x_N; \mu)^{(-1)} f_\mu(x_N; \mu) \end{bmatrix} \\ &= \begin{bmatrix} \nabla_x I(x_1, t_0)^T \Gamma(x_1) \\ \nabla_x I(x_2, t_0)^T \Gamma(x_2) \\ \vdots \\ \nabla_x I(x_N, t_0)^T \Gamma(x_1) \end{bmatrix} \\ &= M_0 \Sigma(\mu) \end{aligned}$$

και επομένως, ο πίνακας M_0 μπορεί να υπολογιστεί και να αποθηκευτεί πριν την έναρξη της διαδικασίας παρακολούθησης, συμβάλλοντας κατά πολύ στη μείωση της πολυπλοκότητας.

Η μέθοδος που έχει προταθεί, επεκτείνεται έτσι ώστε να αντιμετωπίσει και αλλαγές στις συνθήκες φωτισμού. Χρησιμοποιείται ένα γραμμικό μοντέλο που χρησιμοποιεί οχτώ ως δεκαπέντε διαφορετικές εικόνες οι οποίες λαμβάνονται υπό την ίδια οπτική γωνία, αλλά με διαφορετικές συνθήκες φωτισμού. Οι εικόνες οι οποίες τελικά θα χρησιμοποιηθούν από αυτό το σύνολο, μπορούν να βρεθούν με ανάλυση σε κύριες συνιστώσες, και έστω ότι συγκροτούν τη βάση B . Τότε, η συνάρτηση σφάλματος που πρέπει να ελαχιστοποιηθεί είναι η:

$$O(\delta\mu, \lambda) = \|M\delta\mu + B\lambda + I(\mu(t), t + \tau) - I(0, t_0)\|^2$$

εφόσον ιδανικά θεωρείται ότι:

$$I(\mu^*(t), t) = I(0, t_0) + B\lambda(t)$$

όπου λ είναι τα βάρη που προκύπτουν από την ανάλυση σε κύριες συνιστώσες. Οι επιθυμητές παράμετροι και βάρη δίνονται από τη σχέση:

$$\begin{bmatrix} \delta\mu \\ \lambda \end{bmatrix} = - \begin{bmatrix} M^T M & M^T B \\ B^T M & B^T B \end{bmatrix}^{-1} \begin{bmatrix} M^T \\ B^T \end{bmatrix} (I(\mu(t), t + \tau) - I(0, t_0))$$

Το μοντέλο μπορεί να επεκταθεί και έτσι ώστε να αντιμετωπίζει τις περιπτώσεις επικαλύψεις της περιοχής που παρακολουθείται, από άλλα αντικείμενα. Η προσέγγιση που ακολουθείται σε αυτή την περίπτωση, είναι η επαναληπτική μέθοδος ελαχίστων τετραγώνων με μεταβαλλόμενα βάρη. Με βάση τη μέθοδο αυτή, δεν έχουν όλα τα σφάλματα εντός της περιοχής ενδιαφέροντος τον ίδιο συντελεστή στο τελικό άθροισμα, και συγκεκριμένα, τα pixels που βρίσκονται υπό επικάλυψη έχουν πολύ χαμηλότερο βάρος. Επομένως, στην περίπτωση αυτή απαιτείται η ελαχιστοποίηση του συναρτησιακού:

$$O_R(\mu) = \sum_{x \in \Omega} \rho(I(f(x; \mu), t) - I(x, t_0))$$

όπου η συνάρτηση ρ δίνει τα κατάλληλα βάρη στα τετραγωνικά σφάλματα, έχοντας ως στόχο την ευρωστία απέναντι σε επικαλύψεις.

Γενικά, η μέθοδος που προτείνεται από τους Hager et al. μπορεί να υλοποιηθεί ώστε να είναι γρήγορη και να μπορεί να κάνει παρακολούθηση σε πραγματικό χρόνο. Επίσης, μπορεί να αντιμετωπίσει αποτελεσματικά τις περιπτώσεις επικαλύψεων και αλλαγών φωτισμού. Ωστόσο, η περιοχή που θα παρακολουθηθεί πρέπει να επισημανθεί χειροκίνητα, και επομένως η μέθοδος δεν μπορεί να χαρακτηριστεί αυτόματη.

Κεφάλαιο 4

Ανάπτυξη Αλγορίθμου για την Παρακολούθηση Κινούμενων Αντικειμένων

4.1 Εισαγωγή

Στα προηγούμενα κεφάλαια παρουσιάσαμε τις μεθόδους που μας επιτρέπουν να εντοπίζουμε στις εικόνες τα αντικείμενα που μας ενδιαφέρουν και να τα παρακολουθούμε. Στο κεφάλαιο αυτό, θα αναφερθούμε στην αλγοριθμική υλοποίηση κάποιων από τις μεθόδους αυτές. Στο τέλος του κεφαλαίου θα παρουσιάσουμε έναν εννιαίο αλγόριθμο, ο οποίος επιτρέπει την παρακολούθηση κινούμενων χειρών, και αναπτύχθηκε από το μηδέν για τις ανάγκες της παρούσας διπλωματικής. Ο προγραμματισμός του αλγορίθμου έγινε στο Matlab, και έγινε προσπάθεια να είναι ταχύς και αποτελεσματικός. Για το λόγο αυτό, κάποια μέρη του τελικού προγράμματος είναι γραμμένα σε γλώσσα C. Ο κώδικας του λογισμικού συστήματος παρακολούθησης κινούμενων αντικειμένων και συγκεκριμένα ανθρώπινων χειρονομιών, θα γίνει σύντομα διαθέσιμος στο δίκτυο.

4.2 Αλγόριθμοι Εξέλιξης των Γεωδαιτικών Ενεργών Περιγραμμάτων

Στο δεύτερο κεφάλαιο αναφερθήκαμε διεξοδικά στη μέθοδο των γεωδαιτικών ενεργών περιγραμμάτων, και διαπιστώσαμε τις δυνατότητές της στην κατάτμηση εικόνων, κυρίως χάρη στην ευκολία επέκτασής της έτσι ώστε να κάνει χρήση στατιστικών δεδομένων και προγενέστερων σχημάτων. Οι το-

πολογικές διαφορές καθίστανται αυτόματα διαχειρίσιμες όταν οι γεωδαιτικές ενεργές καμπύλες χρησιμοποιούνται μαζί με επιπεδοσυναρτήσεις, προσδίδοντας στην προσέγγιση ακόμα μεγαλύτερη δύναμη. Είναι επομένως εύλογο το ότι στον αλγόριθμό μας θα χρησιμοποιήσουμε το μοντέλο των ενεργών γεωδαιτικών περιγραμμάτων. Για το λόγο αυτό, κάνουμε μια αναφορά στους πιθανούς αλγόριθμους εξέλιξης των περιγραμμάτων.

4.2.1 Μέθοδος Στενής Ζώνης

Η προφανής μέθοδος εξέλιξης των επιπεδοσυναρτήσεων, είναι η εξέλιξη όλων των τιμών του πίνακα που τις αναπαριστά. Ωστόσο, όταν κάποιος ενδιαφέρεται μόνο για το εξελισσόμενο σύνορο, δηλαδή μόνο για την εξέλιξη του επιπεδοσυνόλου μηδενικού δυναμικού, τότε υπάρχει μεγάλο κόστος αν εξελίσσονται όλες οι τιμές του πίνακα. Δεδομένου ότι σε κάθε χρονική στιγμή οι τιμές της επιπεδοσυνάρτησης μπορούν να προκύψουν από το κινούμενο σύνορο (υποθέτουμε μετασχηματισμό προσημασμένης απόστασης), αρκεί να περιορίσουμε το ενδιαφέρον μας σε μια στενή ζώνη (narrow band) κοντά στο σύνορο.

Η προσέγγιση αυτή, η οποία παρουσιάζεται στο [34] και περιγράφεται αναλυτικά στο [35] παρουσιάζει πολλά πλεονεκτήματα κυρίως από άποψη ταχύτητας εκτέλεσης. Αν γίνεται ανανέωση όλων των τιμών της επιπεδοσυνάρτησης, τότε απαιτούνται $o(N^2)$ υπολογισμοί, όπου N είναι το πλήθος των σημείων κατά μήκος μιας διάστασης. Αντίθετα, αν οι υπολογισμοί και οι ανανεώσεις γίνουν μόνο σε μια γειτονιά του κινούμενου συνόρου, η πολυπλοκότητα μειώνεται αισθητά. Αν η γειτονιά περιέχει k σημεία, τότε η πολυπλοκότητα μειώνεται στο $o(kN)$, που αποτελεί σημαντική βελτίωση.

Τα σημεία του ενδιαφέροντος, δηλαδή τα σημεία της στενής ζώνης, αποθηκεύονται σε ένα μονοδιάστατο πίνακα και η επιπεδοσυνάρτηση Φ ανανεώνεται μόνο σε αυτά τα σημεία. Οι τιμές στα όρια της στενής ζώνης αποθηκεύονται και όταν το σύνορο πλησιάσει στις τιμές αυτές, τότε οι υπολογισμοί σταματούν. Στο στάδιο αυτό, η επιφάνεια επαναρχικοποιείται ως μετασχηματισμός προσημασμένης απόστασης από το εξελιγμένο πλέον σύνορο. Επομένως, η μέθοδος στενής ζώνης είναι η επανάληψη των ακόλουθων βημάτων:

1. Σημειώνονται τα σημεία τα οποία ανήκουν στη στενή ζώνη.
2. Αποθηκεύονται οι τιμές που βρίσκονται στα όρια της στενής ζώνης.
3. Τα σημεία εκτός της στενής ζώνης και στο εξωτερικό (εσωτερικό) της καμπύλης αρχικοποιούνται σε αρνητικές (θετικές) τιμές.
4. Η επιφάνεια ανανεώνεται μέχρι το κινούμενο σύνορο να πλησιάσει τα όρια της στενής ζώνης.

5. Η επιφάνεια επαναρχικοποιείται και η διαδικασία επαναλαμβάνεται.

Η μέθοδος Στενής Ζώνης έχει ως αποτέλεσμα η μέθοδος εξέλιξης με επιπεδοσύνολα να έχει την ίδια περίπου πολυπλοκότητα με τις παραδοσιακές μεθόδους διακριτοποίησης της εξελισσόμενης καμπύλης, ενώ χειρίζεται αυτόματα αλλαγές στην τοπολογία.

Το βασικό ζήτημα στην μέθοδο της Στενής Ζώνης είναι η κατάλληλη επιλογή του εύρους k . Εάν η ζώνη είναι απείρως μεγάλη, τότε καλύπτεται όλη η επιφάνεια και δεν υπάρχει καμία μείωση της πολυπλοκότητας. Αν μόνο ένας μικρός αριθμός από σημεία περιλαμβάνεται στη στενή ζώνη, για παράδειγμα το k ισούται με 2, τότε η εξέλιξη εντός της ζώνης είναι μεν πολύ γρήγορη, αλλά απαιτούνται συχνές επαναρχικοποιήσεις. Έτσι, η πολυπλοκότητα αυξάνεται λόγω των πολλών επαναρχικοποιήσεων. Μια εμπειρική τιμή του εύρους ζώνης είναι έξι σημεία. Ωστόσο, κάποιες προσεγγίσεις χρησιμοποιούν μεταβλητό εύρος ζώνης και συγκεκριμένα μικρή ζώνη όταν η ταχύτητες εξέλιξης είναι μικρές και μεγάλο εύρος όταν οι ταχύτητες είναι μεγάλες. Κάποιες άλλες τεχνικές χρησιμοποιούν τις τιμές της επιπεδοσυνάρτησης Φ για να καθορίσουν τα σημεία που ανήκουν στη στενή ζώνη. Έτσι, στη στενή ζώνη τοποθετούνται σημεία που αντιστοιχούν σε τιμές $|\Phi|$ κάτω από συγκεκριμένο κατώφλι. Μετά από κάθε εξέλιξη, απαιτείται επαναρχικοποίηση της επιφάνειας και αυτό μπορεί να εισάγει επιπλέον κόστος στην εκτέλεση.

4.2.2 Μέθοδος Ταχείας Προέλασης

Η μέθοδος Ταχείας Προέλασης (Fast Marching Method) παρουσιάστηκε στο [34] και αποτελεί μια ταχύτατη μέθοδο, η οποία ωστόσο μπορεί να εφαρμοστεί μόνο για μια συγκεκριμένη οικογένεια εξισώσεων εξέλιξης.

Συγκεκριμένα, ο αλγόριθμος Ταχείας Προέλασης χρησιμοποιείται όταν όλα τα σημεία του μετώπου εξελίσσονται προς την ίδια κατεύθυνση, καθόλη της διάρκεια της εξέλιξης. Τότε, το σύνορο επισκέπτεται κάθε σημείο της εικόνας μόνο μια φορά και το μέτωπο κάθε φορά μπορεί να καθοριστεί από το σύνολο των σημείων για τα οποία ισχύει $\Gamma(t) = \{(x, y) : T(x, y) = t\}$, δηλαδή τα σημεία στα οποία βρίσκεται η καμπύλη μετά από χρόνο t . Στην περίπτωση αυτή, η μερική διαφορική εξίσωση που πρέπει να λυθεί είναι η:

$$\begin{cases} \|\nabla T\|F(x, y) = 1 \\ T(x, y) = 0 \text{ για } (x, y) \in \Gamma(0) \end{cases}$$

Αν $F > 0$, τότε το πρόβλημα ανάγεται στην εύρεση του D_{ij} με βάση τη σχέση $\nabla^+T = \frac{1}{F_{ij}}$, όπου ∇^+T είναι μια προσέγγιση του $\|\nabla T\|$. Αντικαθιστώντας το ∇^+T , παίρνουμε:

$$\max^2(D_{ij}^{-x}, 0) + \min^2(D_{ij}^{+x}, 0) + \max^2(D_{ij}^{-y}, 0) + \min^2(D_{ij}^{+y}, 0) = \frac{1}{F_{ij}^2} \quad (4.1)$$

όπου $D_{ij}^{\pm x}$ και $D_{ij}^{\pm y}$ είναι οι προσεγγίσεις των παραγώγων T_x και T_y από αριστερά (-) ή από δεξιά (+). Η λύση του προβλήματος (4.1) μπορεί να βρεθεί γρήγορα, κάνοντας χρήση του αλγορίθμου Ταχείας Προέλασης.

Ο αλγόριθμος Ταχείας Προέλασης βασίζεται στο γεγονός ότι η πληροφορία διαδίδεται μόνο προς μια κατεύθυνση, και άρα από τις μικρότερες προς τις μεγαλύτερες τιμές του T . Επομένως, αρκεί τα pixels να σαρώνονται κατά συγκεκριμένο τρόπο, αρχίζοντας από αυτά τα οποία έχουν μικρές τιμές T και είναι κοντά στην καμπύλη και απομακρυνόμενοι από την καμπύλη. Τα σημεία των οποίων τις τιμές ανανεώνουμε, δεν χρειάζεται να τα ξαναεπισκεφτούμε ποτέ. Οι τιμές αυτές θα παραμείνουν αμετάβλητες από τους μετέπειτα υπολογισμούς. Έτσι, σταδιακά, βρίσκονται όλες οι τιμές T_{ij} στο χώρο.

Σύμφωνα με το [27] όπου γίνεται μια παρουσίαση του αλγορίθμου Ταχείας Προέλασης, τα σημεία κατάσσονται σε τρεις κατηγορίες:

1. Alive: Τα σημεία στα οποία έχει ήδη φθάσει το σύνορο, και γνωρίζουμε την τελική τους τιμή.
2. Active: Τα σημεία στα οποία ενδέχεται να φθάσει το σύνορο στην τρέχουσα επανάληψη.
3. Far Away: Τα σημεία που δεν μας απασχολούν ακόμα, και στα οποία δεν γνωρίζουμε πότε θα φθάσει το σύνορο.

Στη συνέχεια, με βάση αυτή την κατηγοριοποίηση, ο αλγόριθμος χωρίζεται σε δύο στάδια, στο στάδιο αρχικοποίησης, και στο στάδιο προέλασης:

- Αρχικοποίηση

1. Τα σημεία που αντιστοιχούν στο σύνορο $\Gamma(0)$ εντάσσονται στην κατηγορία Alive και αντιστοιχούν σε μηδενικό χρόνο διάσχισης από το σύνορο [$T_{ij} = 0$].
2. Για κάθε σημείο στο σύνολο Alive ελέγχονται οι άμεσοι γείτονες και αν δεν ανήκουν στο σύνολο Alive τοποθετούνται στο σύνολο Active με χρόνο διάσχισης αντίστροφο της ταχύτητας διάδοσής τους [$T_{ij} = \frac{1}{F_{ij}}$].
3. Τα υπόλοιπα σημεία ανήκουν στο σύνορο Far Away, και έχουν χρόνο διάσχισης κοντά στο άπειρο [$T_{ij} = \infty$].

- Προέλαση

1. Το σημεία του συνόλου Active, με τον μικρότερο χρόνο T_{ij} μεταφέρεται στο σύνολο Alive.

2. Οι άμεσοι γείτονες του εν λόγω σημείου επιλέγονται, και αν ανήκουν στο σύνολο Far Away, τότε μεταφέρονται στο σύνολο Active με χρόνο διάσχισης που δίνεται κατά τα γνωστά. Στη συνέχεια, οι τιμές χρόνου διάσχισης των σημείων του συνόλου Active, επαναπροσδιορίζονται με τη χρήση της σχέσης (4.1).
3. Αν το σύνολο Active περιέχει στοιχεία, τότε το στάδιο της Προέλασης επανεκινεί.

Ο αλγόριθμος Ταχείας Προέλασης είναι γενικά πολύ ταχύς, ειδικά όταν χρησιμοποιούνται δομές τύπου Heap, οι οποίες βελτιστοποιούν από άποψη χρόνου την ταξινόμηση στοιχείων με βάση κάποιο χαρακτηριστικό όπως ο χρόνος διάσχισης στη συγκεκριμένη περίπτωση. Το βασικό μειονέκτημα του αλγορίθμου Ταχείας Προέλασης είναι ότι απαιτείται η ταχύτητα εξέλιξης να διατηρεί σταθερό πρόσημο, καθόλη τη διάρκεια εξέλιξης.

4.2.3 Μέθοδος HERMES

Ο αλγόριθμος HERMES είναι μια καινούρια μέθοδος εξέλιξης των συναρτήσεων επιπεδοσυνόλων, η οποία παρουσιάζεται στο [28]. Η μέθοδος αυτή αποτελεί ένα συνδυασμό της μεθόδου Στενής Ζώνης, και της μεθόδου Ταχείας Προέλασης. Έτσι, διαθέτει τη γενικότητα της πρώτης αφού η ταχύτητα χρειάζεται να διατηρεί πρόσημο κατά την εξέλιξη, και την ταχύτητα της δεύτερης.

Ο αλγόριθμος βασίζεται στην ιδέα της επιλεκτικής εξέλιξης (Ταχεία Προέλαση) σε ένα στενό παράθυρο (Στενή Ζώνη). Η επιπεδοσυνάρτηση εξελίσσεται κάθε φορά γύρω από το pixel το οποίο έχει τη μεγαλύτερη ταχύτητα. Ουσιαστικά, η μέθοδος μπορεί να κατανοηθεί ως ένας έξυπνος αλγόριθμος Στενής Ζώνης. Αγνοούνται σημεία τα οποία είναι δυνατόν να έχουν μηδενική ταχύτητα εξέλιξης, κάτι που δεν συμβαίνει με τη μέθοδο Στενής Ζώνης. Στην τελευταία, όλα τα σημεία αντιμετωπίζονται εξίσου.

Στο [27] δίνεται μια αναλυτική περιγραφή του αλγορίθμου και επισημαίνεται ότι αποτελείται από δύο στάδια, το στάδιο αρχικοποίησης και το στάδιο προέλασης, όπως άλλωστε και ο αλγόριθμος Ταχείας Προέλασης:

- Αρχικοποίηση: Η επιπεδοσυνάρτηση αρχικοποιείται ως ο μετασχηματισμός προσημασμένης απόστασης από το σύνορο. Τότε, τα σημεία του συνόρου εισάγονται στο σύνολο Active.
- Προέλαση:
 1. Επιλέγεται το σημείο με τη μεγαλύτερη ταχύτητα διάδοσης.
 2. Δημιουργείται ένα παράθυρο γύρω από το σημείο, και η επιπεδοσυνάρτηση εξελίσσεται μόνο για τα σημεία που βρίσκονται εντός

του παραθύρου. Για τη διασφάλιση της ευστάθειας, χρησιμοποιείται διαφορετικό βήμα εξέλιξης για κάθε σημείο του παραθύρου.

3. Όταν η εξέλιξη εντός του παραθύρου σταματήσει, εξάγεται το επιπεδοσύνολο μηδενικού δυναμικού και γίνεται επαναρχικοποίηση της επιπεδοσυνάρτησης εντός του παραθύρου.
4. Τα σημεία που είναι πλέον μέρος του συνόρου εισάγονται στο σύνολο Active, ενώ τα σημεία που δεν αποτελούν πλέον μέρος του συνόρου αφαιρούνται από τη λίστα. Επίσης, ανανεώνεται η ταχύτητα των σημείων που ανήκαν και στις δύο περιπτώσεις στο σύνορο.
5. Ο αλγόριθμος τερματίζει όταν πραγματοποιηθεί ένα προκαθορισμένος αριθμός επαναλήψεων, ή όταν το σύνορο δεν εξελίσσεται πλέον.

Το μειονέκτημα του αλγορίθμου είναι ότι λόγω του τρόπου εξέλιξης, δεν επιλύεται ακριβώς η διαφορική εξίσωση εξέλιξης. Ωστόσο, η τελική καμπύλη ικανοποιεί τους περιορισμούς καμπυλότητας που εισάγονται από το συναρτησιακό ενέργειας. Αυτό, σε συνδυασμό με την ταχεία σύγκλιση με ταχύτητες εφάμιλλες της μεθόδου Ταχείας Προέλασης, καθιστά τη μέθοδο ιδιαίτερα εύχρηστη για εφαρμογές που απαιτούν εξέλιξη καμπυλών.

4.2.4 Υλοποιώντας το Μοντέλο Εξέλιξης των Γεωδαιτικών Ενεργών Περιγραμμάτων

Λόγω της γενικότητας της μεθόδου Στενής Ζώνης, και αφού μελετήθηκε η υλοποίηση της μεθόδου Ταχείας Προέλασης, αποφασίστηκε ότι αυτή θα είναι η μέθοδος που θα υλοποιήσουμε.

Η μέθοδος Στενής Ζώνης απαιτεί συναρτήσεις για την αρχικοποίηση της καμπύλης, για την κατασκευή της Στενής Ζώνης (tube), για την εξέλιξη της επιπεδοσυνάρτησης εντός της Στενής Ζώνης, και για την επαναρχικοποίηση της επιπεδοσυνάρτησης ως μετασχηματισμό προσημασμένης απόστασης από το κινούμενο σύνορο. Έτσι, κατασκευάστηκαν οι ακόλουθες συναρτήσεις:

- **[tube, tube_bounds] = buildTube(bandwidth, phi)**: Η συνάρτηση buildTube παίρνει ως παραμέτρους την επιπεδοσυνάρτηση phi, και το εύρος ζώνης bandwidth, και επιστρέφει ως tube τα σημεία στα οποία $|\Phi| \leq bandwidth$, και ως tube_bounds τα όρια της περιοχής tube. Για επιτάχυνση στο Matlab, τα σημεία δεν αποθηκεύονται ως (x,y), αλλά ως τα indexes των αντίστοιχων σημείων της επιπεδοσυνάρτησης.
- **phi = evolve(lsf, nb, nb_bounds, g, g_x, g_y, opt_mag, prior, coef_balloon, coef_optical, coef_prior, rotate, dt)**: Η συνάρτηση evolve δέχεται ως ορίσματα την επιπεδοσυνάρτηση (lsf), τη στενή ζώνη

(nb), τα όρια της στενής ζώνης (nb_bounds), την κλίση της εικόνας (g), τις παραγώγους της κλίσης (g_x, g_y), το μέτρο της οπτικής ροής (opt_mag), το προγενέστερο σχήμα (prior), το συντελεστή της δύναμης μπαλονιού (coef_balloon), το συντελεστή της δύναμης κίνησης (coef_optical), το συντελεστή της δύναμης του προγενέστερου σχήματος (coef_prior) και το βήμα εξέλιξης (dt). Καλεί ένα εξωτερικό module σε γλώσσα C, και εκτελεί την ανανέωση της επιπεδοσυνάρτησης στα σημεία που ανήκουν στη στενή ζώνη (nb). Οι δυνάμεις που προέρχονται από την κίνηση ή από το προγενέστερο σχήμα, είναι δυνατόν να είναι μηδέν, ανάλογα με το τι επιθυμούμε.

- **[u, c] = initializeLSF(Img)**: Η συνάρτηση initializeLSF δέχεται ως παράμετρο μια εικόνα, αρχικοποιεί την καμπύλη c κοντά στα όρια της εικόνας, και θέτει κάθε σημείο της u ως την προσημασμένη απόσταση από την c.
- **phi = reinitializeLSF(c, phi)**: Η συνάρτηση reinitializeLSF δέχεται ως όρισμα την τρέχουσα επιπεδοσυνάρτηση phi, και το σύνορο c. Επαναρχικοποιεί την phi ως μετασχηματισμό προσημασμένης απόστασης από το σύνορο c, βρίσκοντας τα πρόσημα από τις υπάρχουσες τιμές της phi (βλ. [35]). Χρησιμοποιείται μια υπάρχουσα συνάρτηση του Matlab, για λόγους ταχύτητας. Εναλλακτικά μπορούσε να χρησιμοποιηθεί ο αλγόριθμος Ταχείας Προέλασης.

Έτσι, μπορεί να υλοποιηθεί ο αλγόριθμος Στενής Ζώνης, σύμφωνα και με τα όσα έχουν περιγραφεί στην σχετική παράγραφο. Για την εύρεση του μέτρου της κλίσης χρησιμοποιούνται οι προσεγγίσεις:

$$\begin{aligned}\nabla^+ \Phi &= \max^2(D_{ij}^{-x}, 0) + \min^2(D_{ij}^{+x}, 0) + \max^2(D_{ij}^{-y}, 0) + \min^2(D_{ij}^{+y}, 0) \\ \nabla^- \Phi &= \min^2(D_{ij}^{-x}, 0) + \max^2(D_{ij}^{+x}, 0) + \min^2(D_{ij}^{-y}, 0) + \max^2(D_{ij}^{+y}, 0)\end{aligned}$$

και για τον υπολογισμό της κλίσης χρησιμοποιούνται οι πυρήνες Sobel:

$$\begin{aligned}\Phi_x &= \Phi * \begin{bmatrix} -1 & 0 & 2 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \\ \Phi_y &= \Phi * \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}\end{aligned}$$

Για την εξέλιξη της επιπεδοσυνάρτησης Φ , χρησιμοποιείται η σχέση:

$$\frac{\partial}{\partial t} = \left[g(I)\beta - g(I)\kappa - \nabla g \vec{N}_o \right] \vec{N}_o \quad (4.2)$$

όπου β είναι ο συντελεστής της δύναμης μπαλονιού και η καμπυλότητα κ δίνεται από τη σχέση:

$$\kappa = \frac{\Phi_{xx}\Phi_y^2 - 2\Phi_x\Phi_y\Phi_{xy} + \Phi_{yy}\Phi_x^2}{(\Phi_x^2 + \Phi_y^2)^{\frac{3}{2}}}$$

Επίσης, όταν θέλουμε να συμπεριλάβουμε πληροφορίες κίνησης ή προγενέστερου σχήματος, ενεργοποιούμε και τις αντίστοιχες δυνάμεις, οι οποίες θα περιγραφούν στις επόμενες παραγράφους.

4.2.5 Αποτελέσματα

Τα αποτελέσματα του κώδικά μας φαίνονται στις εικόνες (4.1 – 4.4). Στις εικόνες (4.1, 4.3, 4.4) φαίνεται η εξέλιξη του γεωδαιτικού ενεργού περιγράμματος για διάφορες περιπτώσεις, όπως για συνθετικές εικόνες με γεωμετρικά σχήματα (4.1) και για πραγματικές εικόνες (4.3) και (4.4). Παρατηρείται ότι οι τοπολογικές διαφορές είναι αυτόματα διαχειρίσιμες και ότι προκύπτουν κλειστές και ομαλές καμπύλες. Στις εικόνες (4.2, 4.3, 4.5) φαίνεται και η στενή ζώνη στην οποία γίνονται οι υπολογισμοί, καθώς και τα όρια αυτής. Οι εικόνες (4.4) και (4.5) είναι παρμένες από το video “Hamburg Taxi Sequence”, που χρησιμοποιείται γενικά για την αξιολόγηση αλγορίθμων της Όρασης Υπολογιστών.

Οι βασική παράμετρος που καθορίζει την εξέλιξη των ενεργών γεωδαιτικών περιγραμμάτων, είναι η συνάρτηση εύρεσης ακμών (g). Η εν λόγω συνάρτηση, σύμφωνα με τον ορισμό και τη συνάρτηση που προτάθηκε στο [6], δεν σταματάει εντελώς την εξέλιξη της καμπύλης στα όρια των αντικειμένων, αλλά την επιβραδύνει. Αυτό, δεν κρίνεται ικανοποιητικό στην περίπτωσή μας. Επομένως, επιλέγουμε να ενισχύσουμε την πληροφορία της g με κατάλληλη κατωφλιοποίηση. Συγκεκριμένα, όπως φαίνεται από την εξίσωση (4.2), πρέπει να μηδενίζουμε την συνάρτηση g , έτσι ώστε να μην υπάρχει απλή επιβράδυνση, αλλά πλήρης ακινητοποίηση. Για το λόγο αυτό, ανάλογα με την εικόνα, επιλέγουμε και το κατώφλι. Τυπική τιμή κατωφλιοποίησης είναι η $\text{thres} = 0.6$, και τότε $g(g \geq 0.6) = 0$. Γενικά, προτιμούμε να έχουμε ισχυρή πληροφορία ακμών στην εικόνα και όχι ελλιπή, διότι μετά την εισαγωγή των δυνάμεων προγενέστερου σχήματος και δύναμης κίνησης, οι δυνάμεις αυτές θα μπορούν να οδηγούν την καμπύλη έτσι ώστε να υπερβαίνει τις ακατάλληλες ακμές. Για παράδειγμα, όταν έγινε χρήση των ακμών Canny (που γενικά θεωρείται η καλύτερη μέθοδος εξαγωγής ακμών) για την εξαγωγή της συνάρτησης g , τα αποτελέσματα ήταν απογοητευτικά.

4.3 Αλγόριθμοι Υπολογισμού της Οπτικής Ροής

Η χρήση της οπτικής ροής μεταξύ δύο εικόνων, μπορεί να αποτελέσει οδηγό της διαδικασίας κατάτμησης έτσι ώστε να διαχωριστούν τα κινούμενα αντικείμενα. Επίσης, είναι δυνατόν να χρησιμοποιηθούν σε συνδυασμό με μεθόδους εξέλιξης καμπυλών όπως Γεωδαιτικά Ενεργά Περιγράμματα ή Ενεργά Περιγράμματα, προκειμένου να παρακολουθούν τα κινούμενα αντικείμενα. Έτσι, πρέπει να προγραμματίσουμε αλγόριθμους οι οποίοι βρίσκουν την οπτική ροή. Δεδομένου ότι η μέθοδος Lukas-Kanade προγραμματίζεται άμεσα από τις σχέσεις στις οποίες έχουμε καταλήξει στο προηγούμενο κεφάλαιο, στο σημείο αυτό θα αναφερθούμε μόνο στις σχέσεις που απαιτούνται για τον προγραμματισμό της μεθόδου Horn-Schunck.

4.3.1 Υπολογισμός της Οπτικής Ροής με τη Μέθοδο Horn-Schunck

Η μέθοδος Horn-Schunck είναι μια καθολική μέθοδος, η οποία προκειμένου να λύσει το πρόβλημα της οπτικής ροής, απαιτεί την ελαχιστοποίηση ενός συναρτησιακού που υποθέτει την συνθήκη (3.1) και την ομαλότητα του πεδίου ταχυτήτων. Έτσι, οι ταχύτητες προκύπτουν από:

$$\arg \min_{u,v} \left(\int \int_{\Omega} ((I_x u + I_y v + I_t)^2 + \lambda [(u_x^2 + u_y^2) + (v_x^2 + v_y^2)]) dx dy \right)$$

όπου ο όρος λ επιλέγεται εμπειρικά.

Οι Horn-Schunck προτείνουν στο [14] έναν αλγόριθμο που ελαχιστοποιεί το παραπάνω συναρτησιακό, και δίνει τις ταχύτητες u και v . Συγκεκριμένα, στο διακριτό χώρο των εικόνων, το συναρτησιακό που πρέπει να ελαχιστοποιηθεί είναι το ¹:

$$J(U, V) = \sum_i \sum_j (E_s[i, j, k] + \lambda E_c[i, j, k])$$

όπου τα i και j δηλώνουν τη διακριτοποίηση στο επίπεδο της εικόνας, και το k τη διακριτοποίηση στο πεδίο του χρόνου. Επίσης:

$$\begin{aligned} E_s[i, j, k] &= \frac{1}{4}(U_x^2 + U_y^2 + V_x^2 + V_y^2)[i, j, k] \\ E_c[i, j, k] &= (I_x U + I_y V + I_t)^2[i, j, k] \end{aligned}$$

¹Ο αλγόριθμος είναι παρμένος από το [1].

όπου οι χωρικές παράγωγοι των συνιστωσών της ταχύτητας σε κάθε frame προσεγγίζονται από:

$$U_x[i, j, k] = (U[i + 1, j, k] - U[i, j, k]) / \Delta x, \quad U_y[i, j, k] = (U[i, j + 1, k] - U[i, j, k]) / \Delta y$$

$$V_x[i, j, k] = (V[i + 1, j, k] - V[i, j, k]) / \Delta x, \quad V_y[i, j, k] = (V[i, j + 1, k] - V[i, j, k]) / \Delta y$$

και οι χωρικές και χρονικές παράγωγοι της φωτεινότητας της εικόνας, προσεγγίζονται από:

$$I_x[i, j, k] = \frac{1}{4\Delta x} (I[i + 1, j, k] - I[i, j, k] + I[i + 1, j + 1, k] - I[i, j + 1, k] +$$

$$+ I[i + 1, j, k + 1] - I[i, j, k + 1] + I[i + 1, j + 1, k + 1] - I[i, j + 1, k + 1])$$

$$I_y[i, j, k] = \frac{1}{4\Delta y} (I[i, j + 1, k] - I[i, j, k] + I[i + 1, j + 1, k] - I[i + 1, k, k] +$$

$$+ I[i, j + 1, k + 1] - I[i, j, k + 1] + I[i + 1, j + 1, k + 1] - I[i + 1, j, k + 1])$$

$$I_t[i, j, k] = \frac{1}{4\Delta t} (I[i, j, k + 1] - I[i, j, k] + I[i + 1, j, k + 1] - I[i + 1, j, k] +$$

$$+ I[i, j + 1, k + 1] - I[i, j + 1, k] + I[i + 1, j + 1, k + 1] - I[i + 1, j + 1, k])$$

Υποθέτοντας ότι $\Delta x = \Delta y = \Delta t = 1$, και απαιτώντας τα βέλτιστα U και V των ταχυτήτων να ικανοποιούν τις εξισώσεις $\frac{\partial}{\partial U} J = 0$ και $\frac{\partial}{\partial V} J = 0$ σε κάθε pixel, παίρνουμε τις ακόλουθες δύο εξισώσεις:

$$(1 + \lambda(I_x^2 + I_y^2))U = (1 + \lambda I_y^2)\bar{U} - \lambda I_x I_y \bar{V} - \lambda I_x I_t$$

$$(1 + \lambda(I_x^2 + I_y^2))V = (1 + \lambda I_x^2)\bar{V} - \lambda I_x I_y \bar{U} - \lambda I_y I_t$$

όπου \bar{U} και \bar{V} είναι οι μέσες ταχύτητες σε ένα συμμετρικό χωρικό παράθυρο που περιλαμβάνει τους άμεσους γείτονες για κάθε pixel $[i, j, k]$:

$$\bar{U}[i, j, k] = (U[i + 1, j, k] + U[i, j + 1, k] + U[i - 1, j, k] + U[i, j - 1, k]) \frac{1}{4}$$

$$\bar{V}[i, j, k] = (V[i + 1, j, k] + V[i, j + 1, k] + V[i - 1, j, k] + V[i, j - 1, k]) \frac{1}{4}$$

Με βάση τις παραπάνω εξισώσεις, κατασκευάζεται ένα αριθμητικό σχήμα αναδρομικής εξέλιξης:

$$U_{n+1} = \bar{U}_n - \lambda \frac{I_x \bar{U}_n + I_y \bar{V}_n + I_t}{1 + \lambda(I_x^2 + I_y^2)} I_x$$

$$V_{n+1} = \bar{V}_n - \lambda \frac{I_x \bar{U}_n + I_y \bar{V}_n + I_t}{1 + \lambda(I_x^2 + I_y^2)} I_y$$

Στις σχέσεις αυτές, η αρχικοποίηση μπορεί να είναι οι μηδενικές τιμές. Όταν το επαναληπτικό σύστημα συγκλίνει, έχουν βρεθεί οι επιθυμητές τιμές των ταχυτήτων U και V .

4.3.2 Αποτελέσματα

Ο προγραμματισμός των παραπάνω σχέσεων είναι ευθεία υπόθεση. Ο αλγόριθμος Lukas-Kanade βρέθηκε στο Internet (Sohaib Khan, University of Central Florida), και ο αλγόριθμος Horn-Schunck κατασκευάστηκε. Μάλιστα, εκμεταλλευόμαστε τη δυνατότητα του Matlab να χειρίζεται εύκολα και γρήγορα πίνακες και αποφεύγουμε κάθε είδους επανάληψη κατά μήκος των χωροχρονικών διαστάσεων. Τα πλεονεκτήματα από τη χρήση του αλγορίθμου Horn-Schunck είναι ότι είναι πολύ γρήγορος, και ότι τα πεδία που προκύπτουν είναι πυκνά. Η ευαισθησία του στο θόρυβο δεν αποτέλεσε ανασταλτικό παράγοντα, δεδομένου ότι οι κινήσεις που παρακολουθούσαμε περιορίζονται σε λίγα pixels ανά frame.

Τα αποτελέσματά μας φαίνονται στις εικόνες (4.6 – 4.9). Στην εικόνα (4.6) φαίνονται δύο frames που περιέχουν γεωμετρικά σχήματα. Υπάρχει ελαφριά κίνηση κάποιων σχημάτων, και χρησιμοποιούμε τους αλγορίθμους Lukas-Kanade και Horn-Schunck προκειμένου να υπολογίσουμε την οπτική ροή. Ο αλγόριθμος Horn-Schunck έχει καλά αποτελέσματα, επειδή η κίνηση είναι μικρή (4.6(c)) και επειδή η ομαλότητα του πεδίου ταχυτήτων έχει ως αποτέλεσμα να εμφανίζονται ταχύτητες σε όλα τα σημεία. Αντίθετα, ο αλγόριθμος Lukas-Kanade δεν έχει καλά αποτελέσματα, διότι το “block matching” που επιχειρεί παρουσιάζει αποτελέσματα μόνο στα όρια των κινούμενων αντικειμένων (4.6(e)). Βρίσκοντας το μέτρο της οπτικής ροής σε κάθε σημείο της εικόνας, παρατηρούμε ότι η πληροφορία που εξάγεται είναι δυνατόν να βοηθήσει στην κατάτμηση της εικόνας στην περίπτωση του Horn-Schunck (4.6(d)), αλλά όχι στην περίπτωση του Lukas-Kanade (4.6(f)).

Στην εικόνα (4.7) έχουμε κίνηση του χεριού. Υπολογίζουμε την οπτική ροή με χρήση και των δύο αλγορίθμων (4.7(c)–4.7(e)), και παρατηρούμε ότι ο αλγόριθμος Lukas-Kanade έχει καλύτερα αποτελέσματα από τον Horn-Schunck. Αυτό είναι λογικό δεδομένου ότι μεταξύ των εικόνων (4.7(a)) και (4.7(b)) υπάρχει μεγάλη κίνηση, και αρκετός θόρυβος. Παίρνοντας το μέτρο της οπτικής ροής (4.7(d)–4.7(f)), φαίνεται καθαρά η υπεροχή του Lukas-Kanade στη συγκεκριμένη περίπτωση.

Στις εικόνες (4.8–4.9), ο αλγόριθμος που υλοποιήσαμε υπερέχει. Στις εικόνες (4.8(a)) και (4.8(b)) υπάρχει μια παλάμη η οποία κινείται και κλείνει. Στις εικόνες (4.8(c)) και (4.8(e)) υπολογίζεται η οπτική ροή, και στις (4.8(d)) και (4.8(f)) υπολογίζεται το μέτρο της οπτικής ροής. Το μέτρο της οπτικής ροής που εξάγεται από τον αλγόριθμο Horn-Schunck δίνει περισσότερες πληροφορίες για το κινούμενο αντικείμενο. Το ίδιο γίνεται και στην εικόνα (4.9), όπου οι δύο αλγόριθμοι δεν εμφανίζουν μεγάλες διαφορές. Παρατηρώντας ότι το μέτρο της οπτικής ροής είναι δυνατόν να χρησιμοποιηθεί για κατάτμηση εικόνας, θα εξελίσσουμε τις γεωδαιτικές ενεργές καμπύλες και με έναν και-

νούριο όρο, όπως θα περιγράψουμε στην παράγραφο (4.5).

4.4 Αλγόριθμοι Εύρεσης και Χρήσης Προγενέστερου Σχήματος

Στόχος της διπλωματικής αυτής είναι να προγραμματιστεί αλγόριθμος ο οποίος να μπορεί να παρακολουθεί κινούμενα αντικείμενα, με έμφαση στις ανθρώπινες χειρονομίες. Δεδομένου ότι τα ανθρώπινα χέρια δεν αποτελούν να μοναδικά αντικείμενα στην εικόνα, επιβάλλεται να υπάρχει ένας τρόπος έτσι ώστε να γίνεται η εύρεση μόνο αυτών στην εικόνα. Για το σκοπό αυτό, χρησιμοποιήθηκε η θεωρία των προγενέστερων σχημάτων που παρουσιάστηκε στο δεύτερο κεφάλαιο.

4.4.1 Επεξεργασία του Συνόλου Εκπαίδευσης

Στον αλγόριθμό μας, αποφασίστηκε ότι η κατάλληλη επεξεργασία του συνόλου εκπαίδευσης είναι ιδιαίτερα σημαντική για να βασίζεται μόνο στις ιδιότητες του μετασχηματισμού προσημασμένης απόστασης. Έτσι, επιλέχθηκε να γίνεται ευθυγράμμιση των σχημάτων, ακολουθώντας τόσο τις μεθόδους των Tsai et al. ([38]) όσο και των Paragios et al. ([30]).

Θεωρούμε ότι τα σχήματα του συνόλου εκπαίδευσης διαφέρουν μεταξύ τους εξαιτίας περιστροφής, διαστολής ή/και μετατόπισης. Αναζητούμε δηλαδή τις παραμέτρους s , θ , T_x , T_y , έτσι ώστε οι επιφάνειες των σχημάτων να επικαλύπτονται όσο το δυνατόν περισσότερο. Ο βαθμός της επικάλυψης αντιπροσωπεύεται από μια συνάρτηση ενέργειας της μορφής:

$$E(s, \theta, \vec{T}) = \int_{\Omega} \int (s\Phi_D(x, y) - \Phi_S(A(x, y))) dx dy$$

στην περίπτωση που ταυτίζουμε τις συναρτήσεις προσημασμένης απόστασης (κατά [30]), ή:

$$E(s, \theta, \vec{T}) = \int_{\Omega} \int (I_D(x, y) - I_S(A(x, y))) dx dy$$

στην περίπτωση που ταυτίζουμε τις επιφάνειες των δυαδικών εικόνων (κατά [38]).

Για την ευθυγράμμιση των σχημάτων σύμφωνα με οποιαδήποτε από τις δύο μεθόδους, χρησιμοποιείται η συνάρτηση **fminsearch** του Matlab, η οποία αναζητεί τις παραμέτρους που ελαχιστοποιούν μια συνάρτηση που δέχεται ως όρισμα. Έτσι, δημιουργήσαμε μια συνάρτηση υπολογισμού της ενέργειας, με παραμέτρους τις s , θ , T_x και T_y , και με χρήση της **fminsearch** βρίσκουμε

κάθε φορά τις s^* , θ^* , T_x^* , T_y^* προκειμένου να ελαχιστοποιείται η ενέργεια. Ως αποτέλεσμα, μπορούμε με χρήση των παραμέτρων αυτών, να ευθυγραμμίσουμε τις εικόνες. Στην περίπτωση που χρησιμοποιούμε την ενέργεια που εισήχθη στο [30], λαμβάνουμε υπόψη μας μόνο το μέρος της εικόνας που βρίσκεται εντός μιας στενής ζώνης από το σύνορο, σύμφωνα και με τα όσα έχουμε περιγράψει στο δεύτερο κεφάλαιο. Αυτό είναι αναγκαίο διότι κατά την περιστροφή της εικόνας κατά τυχαία γωνία θ , εμφανίζονται pixel των οποίων η τιμή δεν είναι καθορισμένη. Επομένως, δεδομένου ότι η πρόβλεψη των τιμών των pixels θα είναι με μεγάλο σφάλμα, πρέπει να περιοριστούμε σε περιοχή κοντά στο σύνορο του αντικειμένου.

Τα αποτελέσματα του κώδικά μας φαίνονται στις εικόνες (4.10) και (4.13). Στην εικόνα (4.10) φαίνεται η ευθυγράμμιση των μαχητικών αεροπλάνων (από το [38]), με βάση τη μέθοδο που περιγράφεται στο [38]. Στην εικόνα (4.13) φαίνεται η ευθυγράμμιση κάποιων χεριών του συνόλου εκπαίδευσης (βλ. Εικόνα (4.12)), με βάση τη μέθοδο που περιγράφεται στο [30].

4.4.2 Εξαγωγή του Προγενέστερου Σχήματος

Σύμφωνα με την προσέγγιση των Paragios et al. στο [32], αναζητούμε το βέλτιστο σχήμα, έτσι ώστε να ελαχιστοποιείται το άθροισμα των τετραγωνικών σφαλμάτων μεταξύ των εικόνων. Φυσικά, η επιφάνεια που προκύπτει διατηρεί τις ιδιότητες του μετασχηματισμού προσημασμένης απόστασης. Για την εύρεση της επιφάνειας χρησιμοποιούμε τη μέθοδο κατάβασης κλίσης, προγραμματίζοντας τις σχέσεις (2.13) και (2.14). Η χρήση του Matlab μας επιτρέπει να ανανεώνουμε σε κάθε βήμα ολόκληρο τον πίνακα, κερδίζοντας έτσι σημαντικά σε ταχύτητα υλοποίησης αλλά και εκτέλεσης.

Το μοντέλο που προκύπτει με χρήση των ευθυγραμμισμένων εικόνων (4.12), φαίνεται στην εικόνα (4.14).

4.4.3 Χρήση του Προγενέστερου Σχήματος

Το προγενέστερο σχήμα που εξήχθη από το παραπάνω βήμα χρησιμοποιήθηκε όπως περιγράφουν οι Paragios et al. στο [29]. Πιο συγκεκριμένα, ενσωματώθηκε ως επιπλέον όρος στην διαφορική εξίσωση των γεωδαιτικών ενεργών περιγραμμάτων. Κάθε φορά που ξεκινάει μια καινούρια διαδικασία ανανέωσης υπολογίζεται όπως στην ενότητα (4.4.1), η καλύτερη προσέγγιση αφινικού μετασχηματισμού (με παραμέτρους s , θ , T_x και T_y), έτσι ώστε το προγενέστερο σχήμα να ευθυγραμμίζεται όσο το δυνατόν καλύτερα με την εικόνα στην οποία εξελίσσεται η καμπύλη. Έτσι, αγνοούνται τα υπόλοιπα αντικείμενα στην εικόνα, και εντοπίζεται το αντικείμενο του ενδιαφέροντος. Επίσης, ανάλογα με

το βάρος που δίνουμε στην χρήση του προγενέστερου σχήματος, είναι δυνατόν να βρεθεί το σωστό σχήμα αποκαθιστώντας βλάβες λόγω έλλειψης ακμών. Το μετασχηματισμένο προγενέστερο σχήμα το οποίο υπολογίζεται από την ευθυγράμμιση, χρησιμοποιείται προκειμένου να κατευθύνει την εξελισσόμενη καμπύλη, σύμφωνα με τις μεθόδους που περιγράψαμε στην παράγραφο (2.3.3).

4.4.4 Αποτελέσματα

Τα αποτελέσματα του κώδικά μας φαίνονται στις εικόνες στο τέλος του κεφαλαίου. Στις προσομοιώσεις μας χρησιμοποιήσαμε εικόνες του συνόλου εκπαίδευσης που έχουν υποστεί μικρή επεξεργασία (4.15), ρεαλιστικές εικόνες με εμπόδια (4.16, πρβλ. με 4.3), καθώς και εικόνες που περιλαμβάνουν χέρια και μέρος του ανθρώπινου σώματος (4.18–4.19). Μια σύγκριση μεταξύ των διαφορετικών τρόπων εξέλιξης της καμπύλης, με εφαρμογή και χωρίς εφαρμογή του προγενέστερου σχήματος, φαίνεται στην Εικόνα (4.17).

Σε όλες τις περιπτώσεις τα αποτελέσματα είναι άκρως ικανοποιητικά, ιδίως αν σκεφτούμε την πολυπλοκότητα των ακμών των αντικειμένων που εμφανίζονται στις εικόνες. Με τη χρήση του προγενέστερου σχήματος, είναι δυνατόν να αγνοούνται άλλα αντικείμενα στην εικόνα, αλλά και κάποιες έντονες ακμές που θα σταματούσαν την εξέλιξη της καμπύλης. Ωστόσο, οι παράμετροι εξέλιξης οι οποίοι υπεισέρχονται στο μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων, παίζουν πολύ σημαντικό ρόλο στην ποιότητα των αποτελεσμάτων. Πέραν της συνάρτησης g , η σημασία της οποία αναφέρθηκε στην παράγραφο (4.2.5), σημαντικό ρόλο παίζει και ο λόγος μεταξύ του συντελεστή της δύναμης μπαλονιού (`coef_balloon`) και του συντελεστή της δύναμης προγενέστερου σχήματος (`coef_prior`). Συγκεκριμένα, αν ο συντελεστής προγενέστερου σχήματος είναι πολύ μεγάλος, τότε αγνοούνται οι ακμές και η επιπεδοσυνάρτηση κατευθύνεται χωρίς καμία πληροφορία προς το αποθηκευμένο σχήμα (βλ. Εικόνα (4.20)). Αν η δύναμη μπαλονιού είναι πολύ μεγάλη, τότε κάποιες ακμές εμποδίζουν την εξέλιξη της καμπύλης (βλ. Εικόνα (4.20)). Επομένως, είναι απαραίτητο να γίνονται δοκιμές για την εύρεση των κατάλληλων συντελεστών, και μάλιστα αυτό απαιτείται για κάθε εικόνα που θέλουμε να κατατιμήσουμε.

4.5 Αλγόριθμοι Παρακολούθησης Κίνησης

Ο αλγόριθμος που αναπτύξαμε βασίζεται στην προσέγγιση που περιγράφεται συνοπτικά στο [12]. Ωστόσο, στη δική μας περίπτωση ενσωματώνουμε πληροφορίες από το προγενέστερο σχήμα ή και την οπτική ροή, εντοπίζοντας τελικά το επιθυμητό αντικείμενο στην εικόνα, κάτι το οποίο δεν γίνεται (ή δεν περιγράφεται) στη δημοσίευση [12].

4.5.1 Παρακολουθώντας Κινούμενα Αντικείμενα

Δεδομένης μιας αλληλουχίας εικόνων, βρίσκουμε τα αντικείμενα στο πρώτο frame εξελίσσοντας ένα γεωδαιτικό ενεργό περίγραμμα με τη μέθοδο Στενής Ζώνης. Στη συνέχεια, υπολογίζουμε την οπτική ροή ανάμεσα στο τρέχον frame και το επόμενο με βάση τον αλγόριθμο Horn-Schunck. Κατόπιν, γίνεται χρήση της οπτικής ροής ώστε να μετατοπιστεί το κινούμενο σύνορο, και η καινούρια καμπύλη εξελίσσεται στο επόμενο frame.

Με βάση το μέτρο της οπτικής ροής, κατασκευάζεται και μια δύναμη που έλκει την καμπύλη προς την κινούμενη περιοχή. Η συνάρτηση που χρησιμοποιούμε έχει τη μορφή:

$$g(O) = \frac{1}{1+G_\sigma * O}$$

όπου G_σ ένα πυρήνας Gauss, και $O = \sqrt{u^2 + v^2}$, με u και v τα διανύσματα της οπτικής ροής. Με βάση αυτή τη συνάρτηση δημιουργούμε έναν επιπλέον όρο ταχύτητας για την εξέλιξη των ενεργών γεωδαιτικών καμπυλών. Έτσι, τελικά η επιπεδοσυνάρτηση εξελίσσεται σύμφωνα με τη σχέση:

$$\Phi(t+1) = \Phi(t) + (\alpha speed_{curvature} - \beta speed_{spring} - \gamma speed_{balloon} - \delta speed_{optical\ flow}) dt_{gac}$$

όπου οι συντελεστές α , β , γ , δ των δυνάμεων, βρίσκονται πειραματικά, μετά από δοκιμές. Ο συνδυασμός των δυνάμεων έχει ως αποτέλεσμα, μετά από μερικά frames, τα αντικείμενα τα οποία έχουν εντοπιστεί από το κλασσικό μοντέλο των γεωδαιτικών ενεργών καμπυλών να εξαφανίζονται, και να παρακολουθείται μόνο το αντικείμενο που κινείται. Λόγω της καταρχήν μετατόπισης της καμπύλης με τη χρήση των διανυσμάτων οπτικής ροής, απαιτείται μόλις μια επανάληψη για τον τελικό υπολογισμό της εξελισσόμενης καμπύλης, σε κάθε frame.

Τα αποτελέσματα του κώδικά μας, φαίνονται στις εικόνες (4.23–4.24). Παρατηρούμε ότι τελικά, παρακολουθείται μόνο το κινούμενο αντικείμενο.

4.5.2 Παρακολουθώντας Ανθρώπινες Χειρονομίες

Όταν επιθυμούμε να παρακολουθήσουμε ανθρώπινα χέρια στις αλληλουχίες εικόνων, τότε αρχικά δημιουργούμε ένα σύνολο από δυαδικές εικόνες που απεικονίζουν χέρια. Στη συνέχεια, ευθυγραμμίζουμε τις εικόνες και εξάγουμε το προγενέστερο σχήμα. Το προγενέστερο σχήμα χρησιμοποιείται σε συνδυασμό με τα γεωδαιτικά ενεργά περιγράμματα, προκειμένου να εντοπιστεί το χέρι στην εικόνα, σύμφωνα με το [32]. Στη συνέχεια, ακολουθούνται τα βήματα που περιγράφονται στην προηγούμενη παράγραφο προκειμένου να παρακολουθηθεί το αντικείμενο που έχει ήδη εντοπιστεί. Ας σημειωθεί ότι επειδή

οι χειρονομίες είναι δυνατόν να είναι γενικά περίπλοκες, η χρήση του προγενέστερου σχήματος σταματάει στις επόμενες εικόνες, και παρακολουθείται το χέρι που έχει ήδη εντοπιστεί, ανεξαρτήτως της κίνησης και του σχήματος που παίρνει. Έτσι, αρχικά η επιπεδοσυνάρτηση εξελίσσεται σύμφωνα με τη σχέση:

$$\Phi(t+1) = \Phi(t) + (\alpha speed_{curvature} - \beta speed_{spring} - \gamma speed_{balloon} - \delta speed_{prior}) dt_{gac}$$

όπου α , β , γ , δ είναι οι συντελεστές των δυνάμεων εξέλιξης, όπως βρίσκονται μετά από δοκιμές. Μετά από τον εντοπισμό του επιθυμητού αντικειμένου, η επιπεδοσυνάρτηση εξελίσσεται σύμφωνα με τη σχέση:

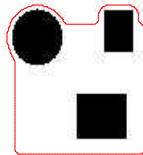
$$\Phi(t+1) = \Phi(t) + (\alpha speed_{curvature} - \beta speed_{spring} - \gamma speed_{balloon})$$

όπου οι συντελεστές α , β , γ , δ βρίσκονται πειραματικά μετά από δοκιμές.

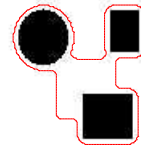
Τα αποτελέσματα του κώδικά μας φαίνονται στις εικόνες (4.21–4.26). Παρατηρούμε ότι ο αλγόριθμός μας εντοπίζει το χέρι ικανοποιητικά, και το παρακολουθεί καθόλη της διάρκεια της κίνησης. Σημαντικό στο σημείο αυτό, είναι να προσέξουμε την παρουσία του “βραχιολιού”. Η ύπαρξή του δημιουργεί έντονη ακμή, η οποία σταματάει την εξέλιξη της καμπύλης στο τμήμα του καρπού. Αν το βραχιόλι απουσίαζε, τότε η δύναμη μπαλονιού θα εξέλιξε την καμπύλη μέχρι το εσωτερικό της παλάμης, και δεν θα παίρναμε τα αποτελέσματα που θέλουμε.

4.6 Εικόνες Κεφαλαίου 4

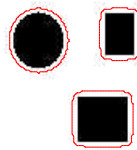
Στην ενότητα αυτή παρατίθενται οι εικόνες του τέταρτου κεφαλαίου, μαζί με τα κατάλληλα συνοδευτικά σχόλια-λεζάντες.



(a)

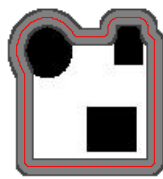


(b)

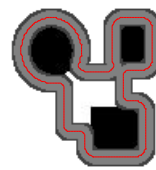


(c)

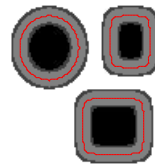
Εικόνα 4.1: Εξέλιξη Γεωδαιτικού Ενεργού Περιγράμματος με τη μέθοδο της Στενής Ζώνης. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις. (Εικόνα: Γεωμετρικά Σχήματα (128x128)).



(a)



(b)



(c)

Εικόνα 4.2: Η Στενή Ζώνη που υπολογίζεται. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις. (Εικόνα: Γεωμετρικά Σχήματα (128x128)).



(a)



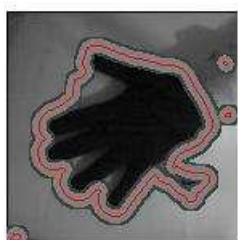
(b)



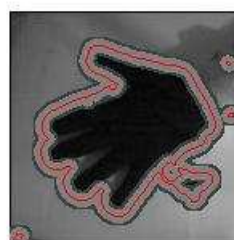
(c)



(d)



(e)



(f)

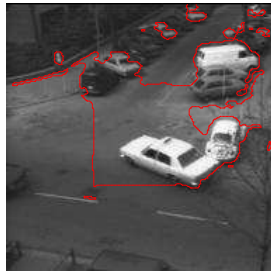
Εικόνα 4.3: Εξέλιξη Γεωδαιτικού Ενεργού Περιγράμματος με τη μέθοδο της Στενής Ζώνης. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις, (d) Η Στενή Ζώνη μετά από 5 επαναλήψεις, (e) Η Στενή Ζώνη μετά από 10 επαναλήψεις, (f) Η Στενή Ζώνη μετά από 15 επαναλήψεις. (Εικόνα: Χέρι και Mickey (144x144)).



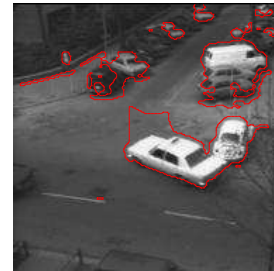
(a)



(b)



(c)



(d)



(e)

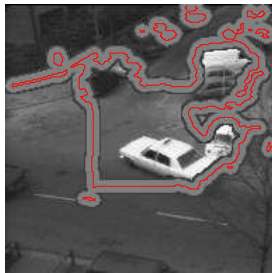
Εικόνα 4.4: Εξέλιξη Γεωδαιτικού Ενεργού Περιγράμματος με τη μέθοδο της Στενής Ζώνης. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις, (d) Μετά από 20 επαναλήψεις, (e) Μετά από 25 επαναλήψεις. (Εικόνα: Ταξί Αμβούργου (191x191)).



(a)



(b)



(c)

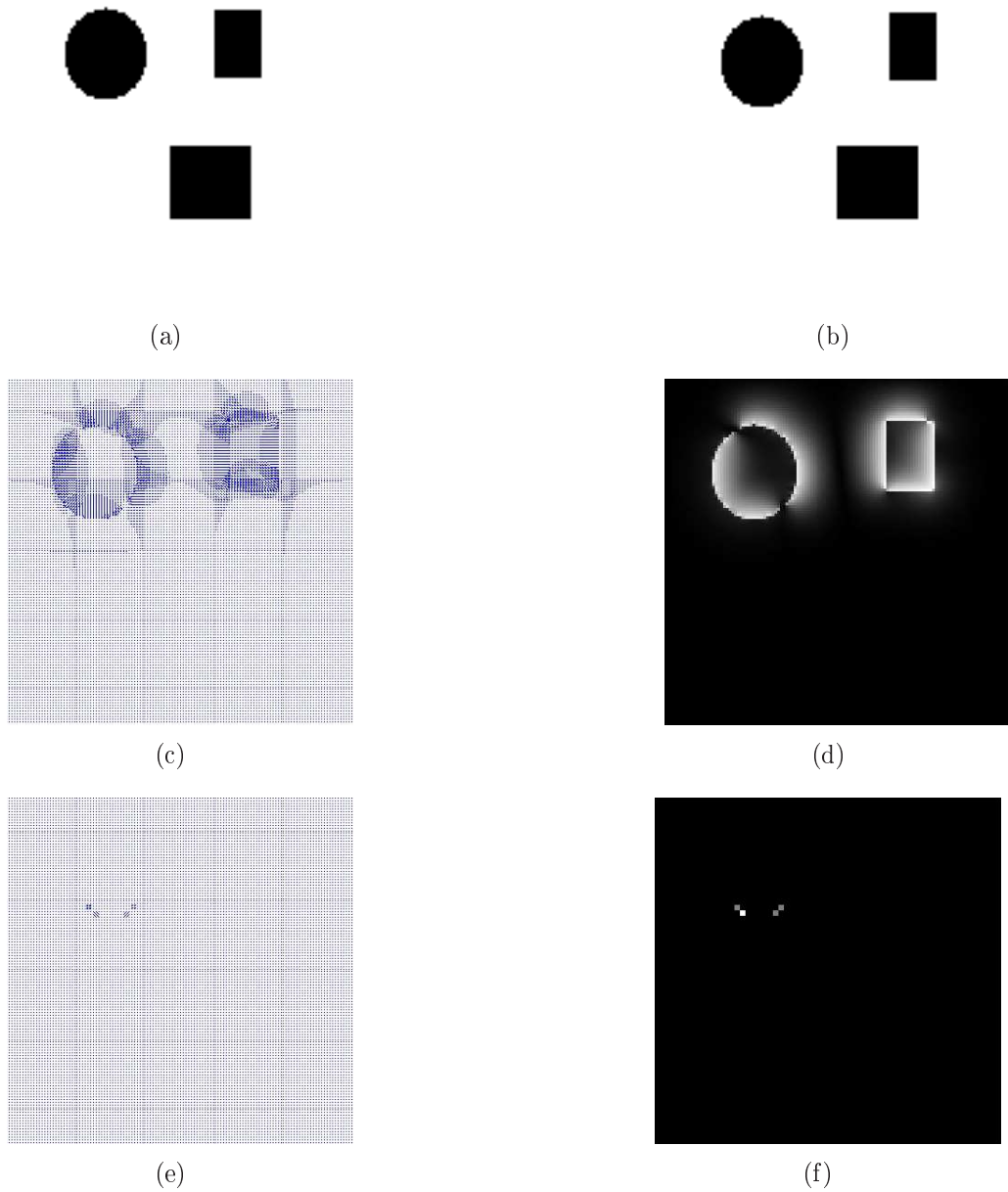


(d)



(e)

Εικόνα 4.5: Η Στενή Ζώνη που υπολογίζεται. (a) Μετά από 5 επαναλήψεις, (b) Μετά από 10 επαναλήψεις, (c) Μετά από 15 επαναλήψεις, (d) Μετά από 20 επαναλήψεις, (e) Μετά από 25 επαναλήψεις. (Εικόνα: Ταξί Αμβούργου (191x191)).



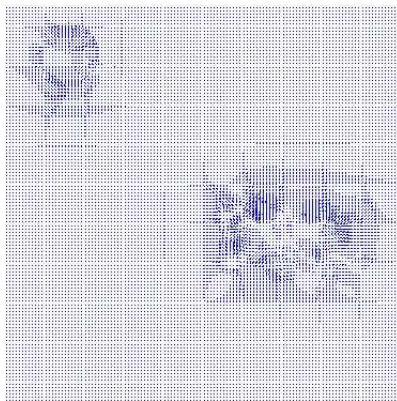
Εικόνα 4.6: Εύρεση Οπτικής Ροής (a) Frame 1, (b) Frame 2: Κινήθηκε ο δίσκος και το μικρό ορθογώνιο, (c) Με χρήση του αλγορίθμου Horn-Schunck, (d) Μέτρο της οπτικής ροής Horn-Schunck, (e) Με χρήση του αλγορίθμου Lukas-Kanade, (e) Μέτρος της οπτικής ροής Lukas-Kanade. (Εικόνα: Γεωμετρικά Σχήματα (128x128)).



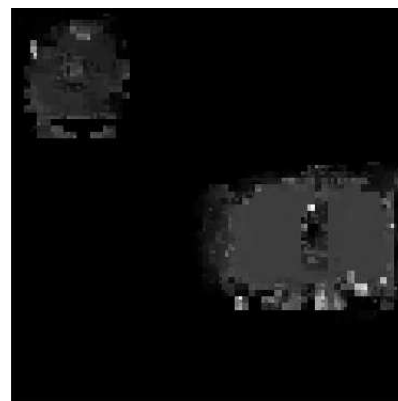
(a)



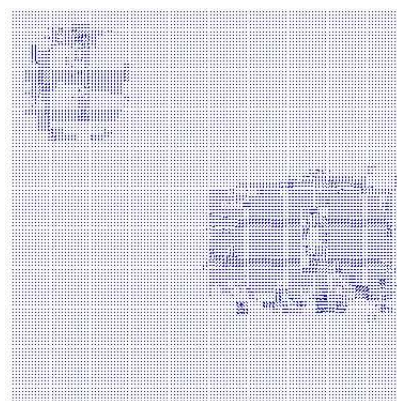
(b)



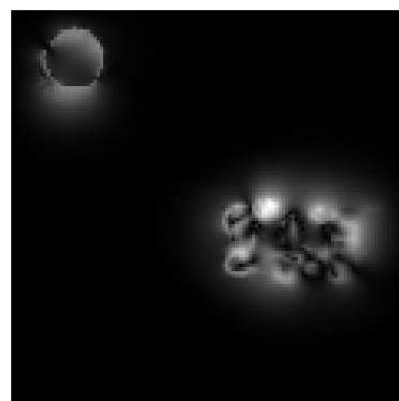
(c)



(d)



(e)



(f)

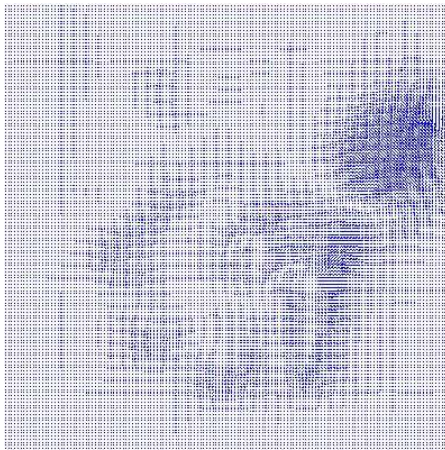
Εικόνα 4.7: Εύρεση Οπτικής Ροής (a) Frame 1, (b) Frame 2: Κινήθηκε ο δίσκος και ο Mickey, (c) Με χρήση του αλγορίθμου Horn-Schunck, (d) Μέτρο της οπτικής ροής Horn-Schunck, (e) Με χρήση του αλγορίθμου Lukas-Kanade, (e) Μέτρος της οπτικής ροής Lukas-Kanade. (Εικόνα: Διάφορα Αντικείμενα (120x120)).



(a)



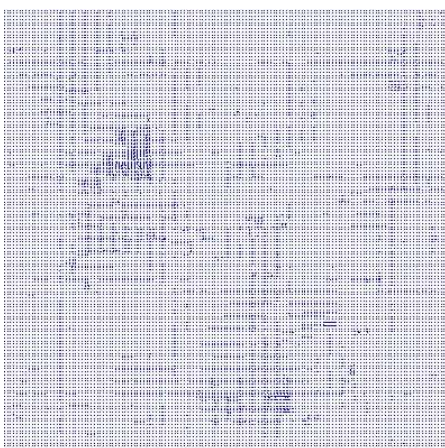
(b)



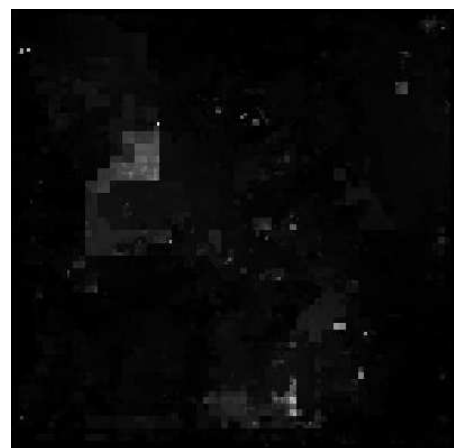
(c)



(d)



(e)



(f)

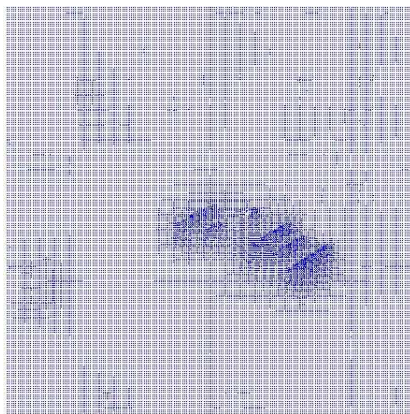
Εικόνα 4.8: Εύρεση Οπτικής Ροής (a) Frame 1, (b) Frame 2: Η παλάμη μετακινήθηκε και άνοιξε, (c) Με χρήση του αλγορίθμου Horn-Schunck, (d) Μέτρο της οπτικής ροής Horn-Schunck, (e) Με χρήση του αλγορίθμου Lukas-Kanade, (f) Μέτρος της οπτικής ροής Lukas-Kanade. (Εικόνα: Χέρι (144x144)).



(a)



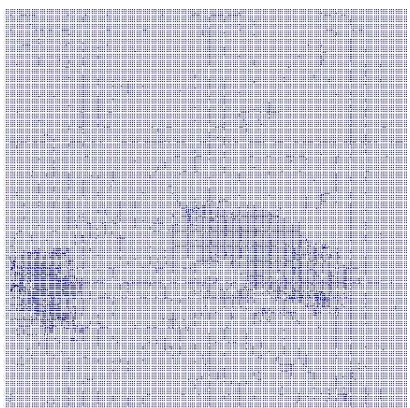
(b)



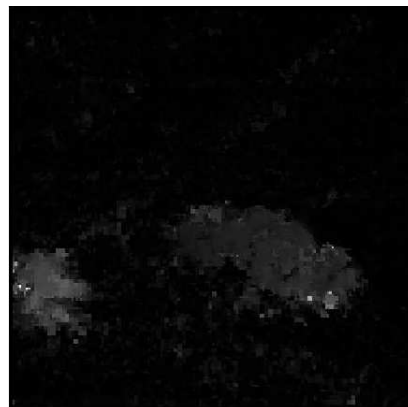
(c)



(d)

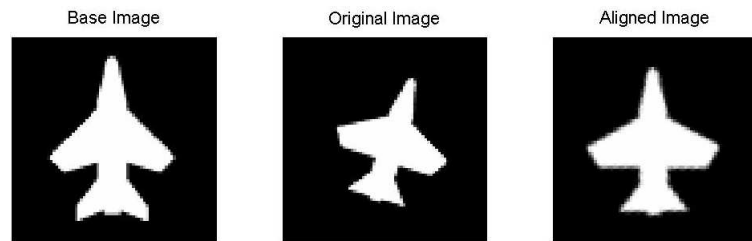


(e)

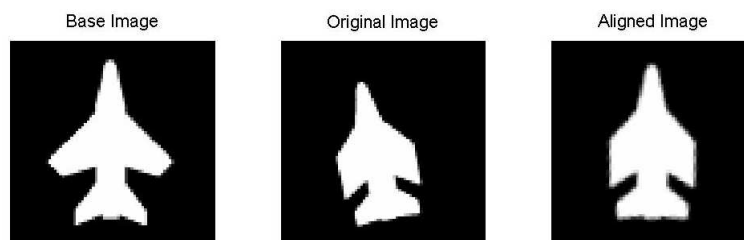


(f)

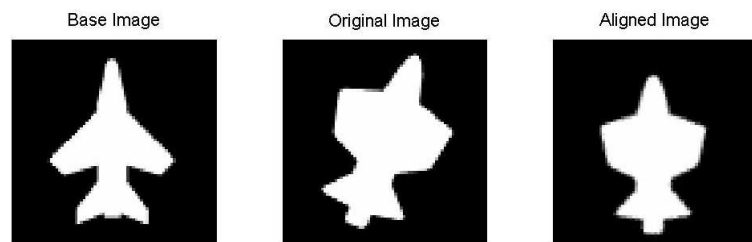
Εικόνα 4.9: Εύρεση Οπτικής Ροής (a) Frame 1, (b) Frame 2: Μικρή μετακίνηση του πεζού (επάνω αριστερά), του ταξί (κάτω δεξιά), και του εισερχόμενου αυτοκινήτου (αριστερά), (c) Με χρήση του αλγορίθμου Horn-Schunck, (d) Μέτρο της οπτικής ροής Horn-Schunck, (e) Με χρήση του αλγορίθμου Lukas-Kanade, (f) Μέτρος της οπτικής ροής Lukas-Kanade. (Εικόνα: Ταξί Αμβούργου (191x191)).



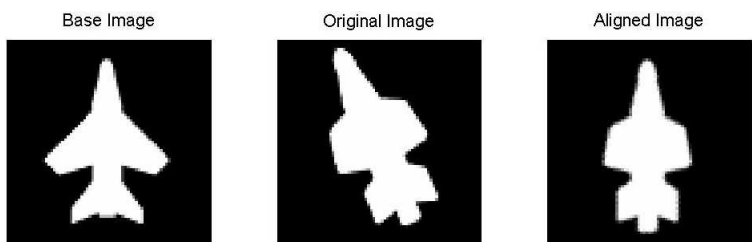
(a) Αεροπλάνο 2



(b) Αεροπλάνο 5

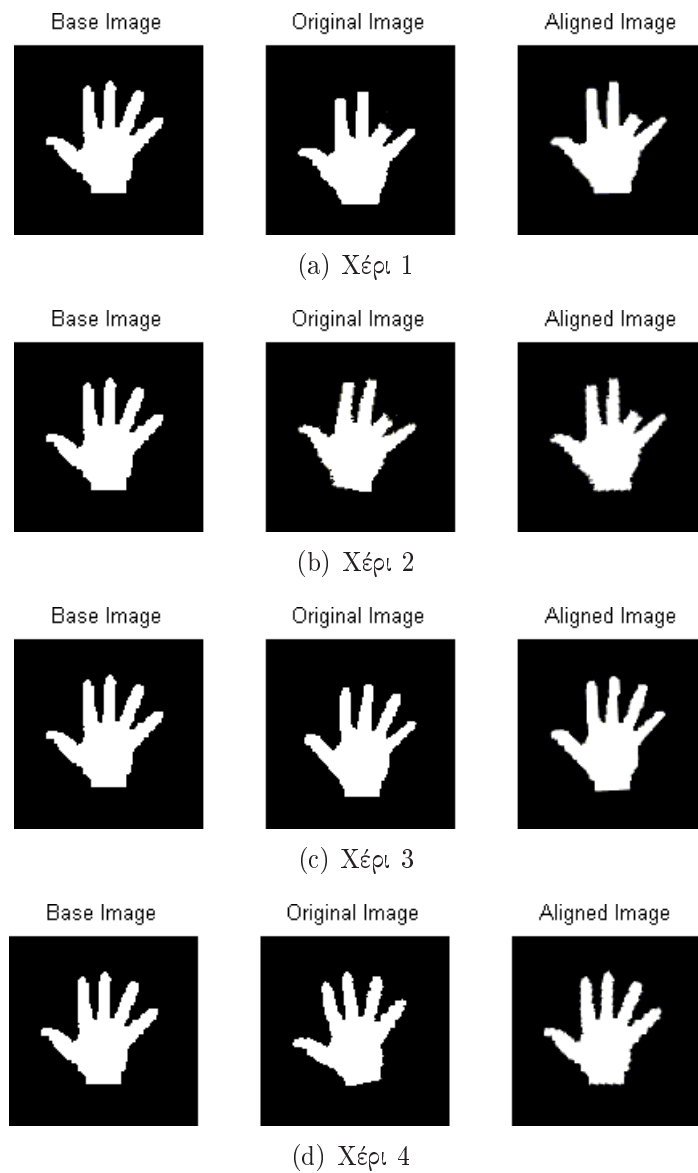


(c) Αεροπλάνο 9

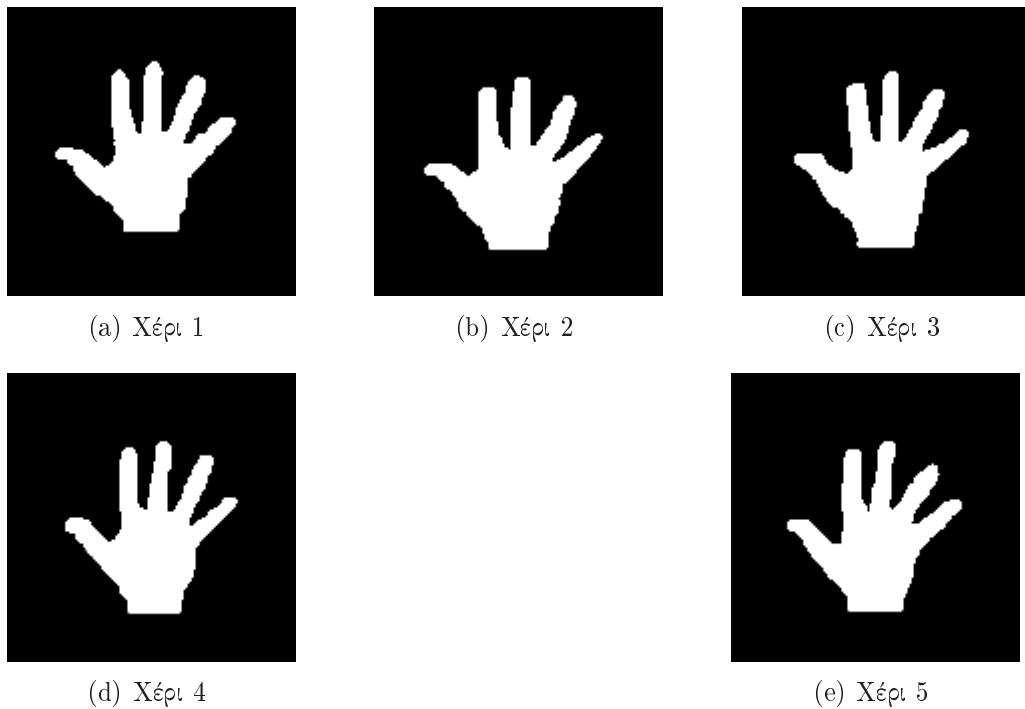


(d) Αεροπλάνο 10

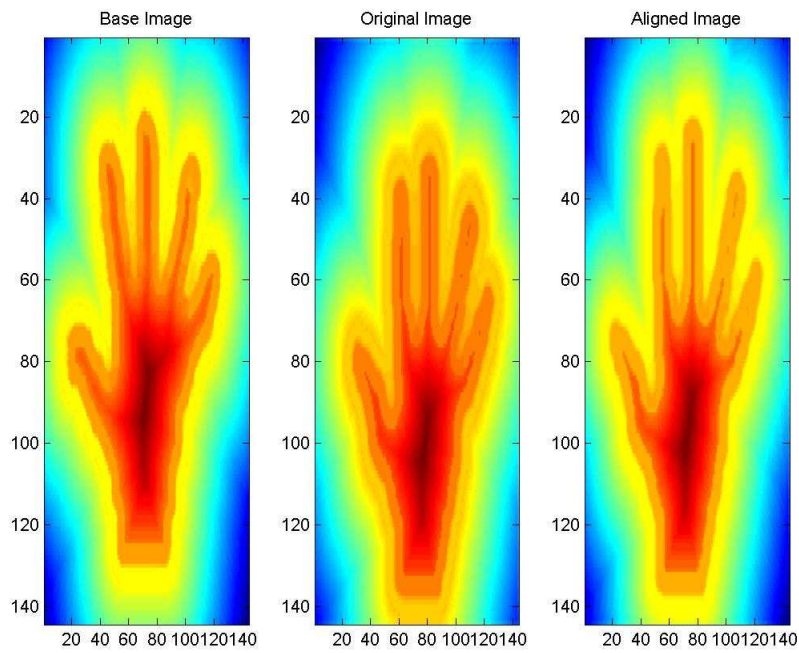
Εικόνα 4.10: Ευθυγράμμιση των αεροπλάνων (δυναδικές εικόνες 75x75), σύμφωνα με τη μέθοδο του [38]. (a) Το αεροπλάνο περιστρέφεται και διαστέλλεται, (b) Το αεροπλάνο περιστρέφεται και διαστέλλεται, (c) Το αεροπλάνο περιστρέφεται και συστέλλεται, (d) Το αεροπλάνο περιστρέφεται και συστέλλεται. (Εικόνες: Μαχητικά Αεροπλάνα (75x75) (από το [38])).



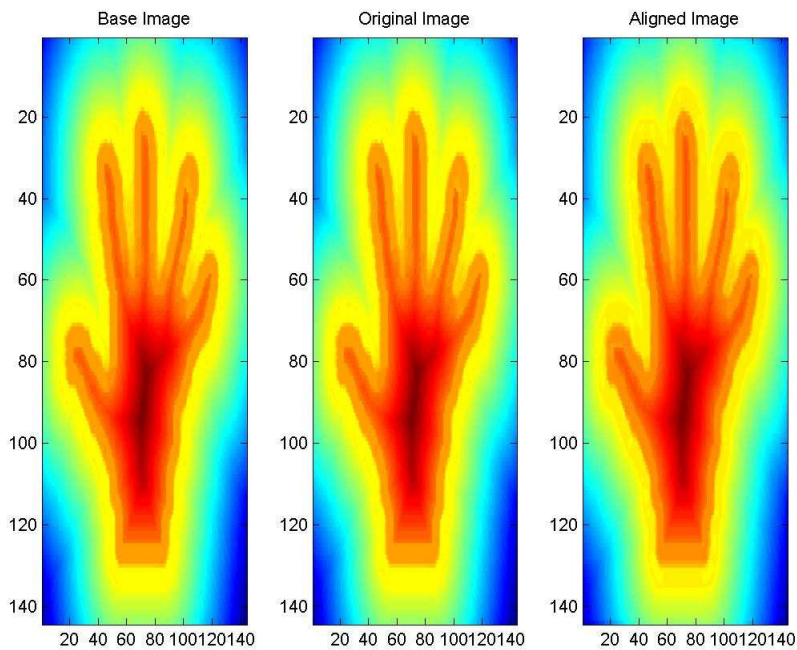
Εικόνα 4.11: Εικόνες χεριών που ευθυγραμίζονται. (a) Απουσιάζει το τέταρτο δάχτυλο, (b) Απουσιάζει το τέταρτο δάχτυλο και το σχήμα είναι περιστραμμένο, (c) Μετασχηματισμένο σχήμα, (d) Μετασχηματισμένο Σχήμα. Οι εικόνες τραβήχθηκαν με Web-Cam, και στη συνέχεια τροποποιήθηκαν. (Εικόνες: Χέρια (144x144)).



Εικόνα 4.12: Το σύνολο εκπαίδευσης για την εξαγωγή του προγενέστερου σχήματος του χεριού. Τα χέρια φωτογραφήθηκαν με Web-Cam, και οι δυαδικές εικόνες δημιουργήθηκαν χειροκίνητα με χρήση του MS Paint. (Εικόνες: Χέρια (144x144)).

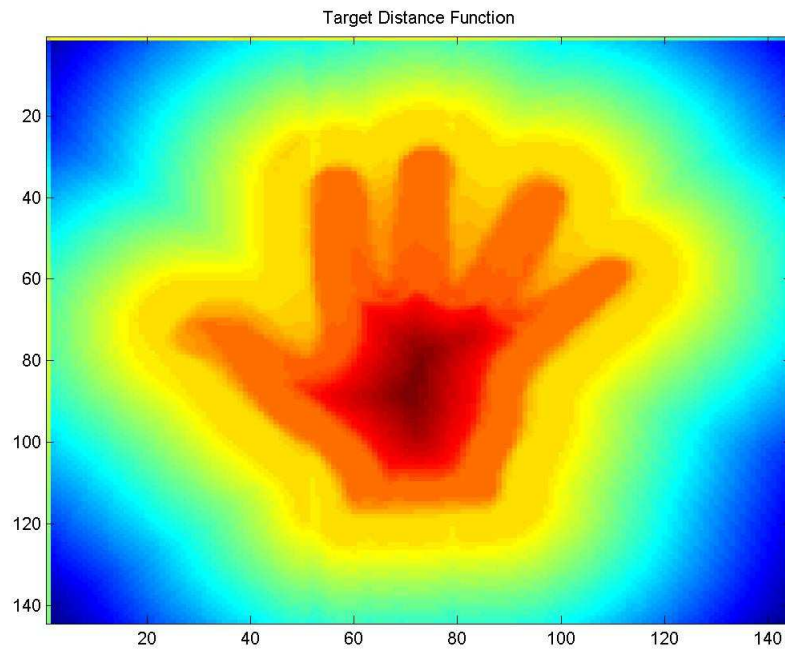


(a) Χέρι 2

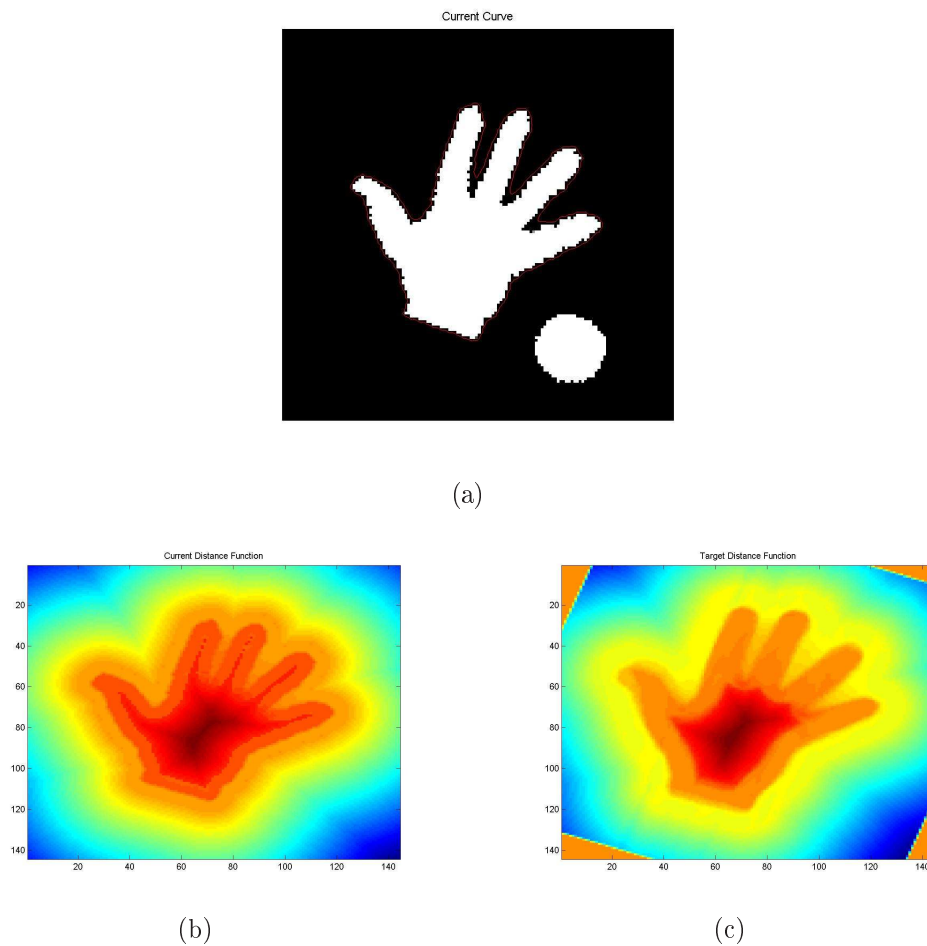


(b) Χέρι 8

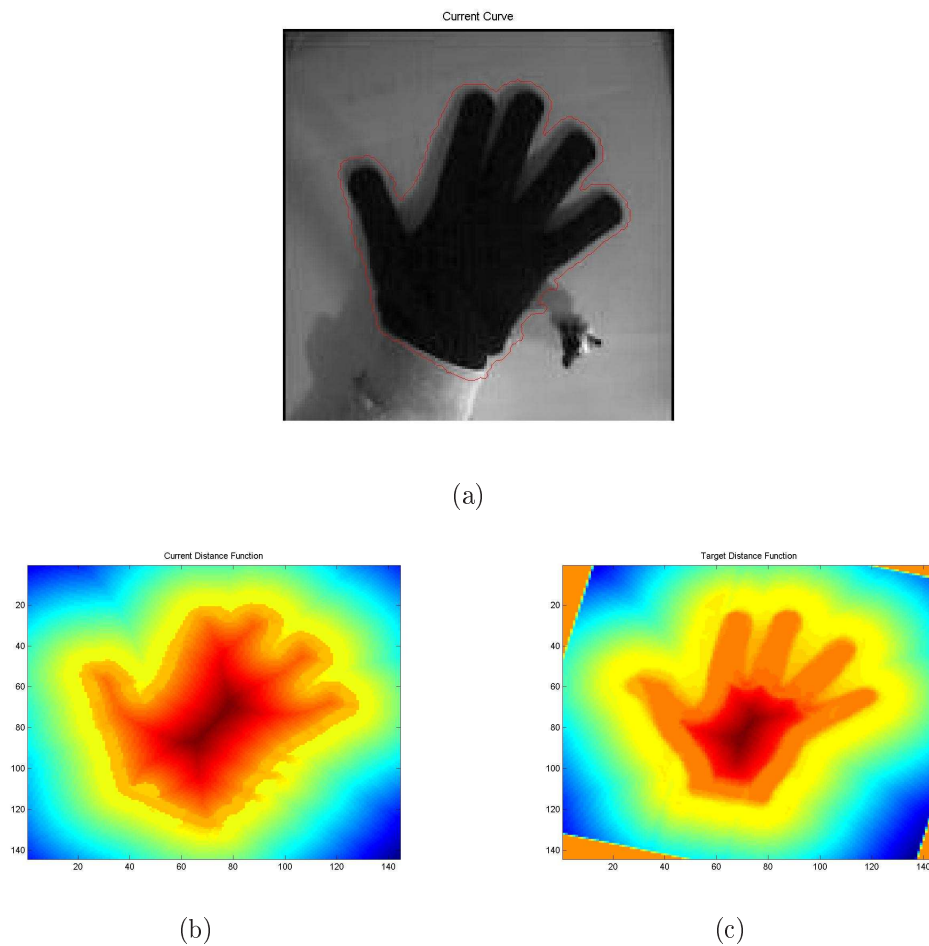
Εικόνα 4.13: Ευθυγράμμιση των χεριών (μετασχηματισμοί προσημασμένης απόστασης 144x144) με βάση τη μέθοδο του [30]. (Εικόνες: Χέρια (144x144)).



Εικόνα 4.14: Το προγενέστερο σχήμα ως μετασχηματισμός προσημασμένης απόστασης, όπως προέκυψε από τη μέθοδο του [32], μετά την ευθυγράμμιση και επεξεργασία των σχημάτων της εικόνας (4.13). (Εικόνα: Προγενέστερο Σχήμα Χεριού (144x144)).



Εικόνα 4.15: Η εξέλιξη του γεωδαιτικού ενεργού περιγράμματος κατευθύνεται και από το προγενέστερο σχήμα. (a) Το τελικό γεωδαιτικό περίγραμμα, στο οποίο έχει αγγισθεί η σφαίρα, (b) Ο μετασχηματισμός προσημασμένης απόστασης για το γεωδαιτικό περίγραμμα του (a), (c) Το προγενέστερο σχήμα περιστρέφεται κατάλληλα προκειμένου να ευθυγραμμίζονται οι μετασχηματισμοί προσημασμένης απόστασης. (Εικόνα 144x144).



Εικόνα 4.16: Η εξέλιξη του γεωδαιτικού ενεργού περιγράμματος κατευθύνεται και από το προγενέστερο σχήμα. (a) Το τελικό γεωδαιτικό περίγραμμα στο οποίο έχει αγνοηθεί ο Mickey, και έχει εντοπιστεί μόνο το χέρι. Το γάντι εξασφαλίζει έντονες ακμές. Είναι εμφανής η διαφορά στην κατάτμηση, σε σχέση με την εικόνα (4.3), (b) Ο μετασχηματισμός προσημασμένης απόστασης για το γεωδαιτικό περίγραμμα του (a), (c) Το περιστραμμένο προγενέστερο σχήμα που κατευθύνει την κατάτμηση. (Εικόνα: 144x144).



(a)



(b)



(c)



(d)

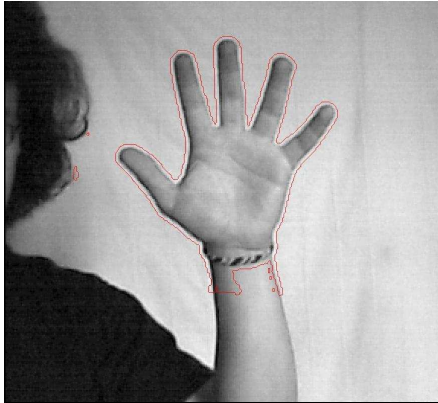


(e)

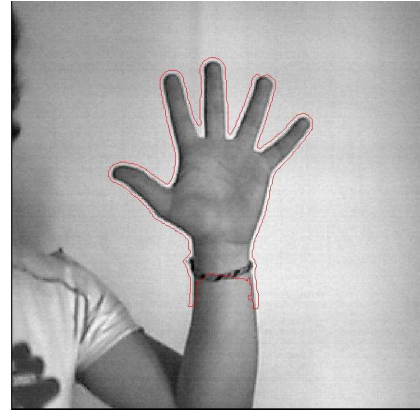


(f)

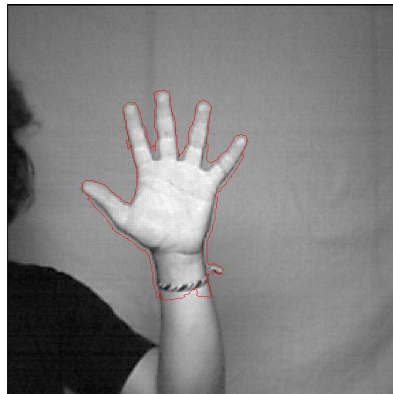
Εικόνα 4.17: Εξέλιξη γεωδαιτικού ενεργού περιγράμματος. (a), (c), (e) Χωρίς χρήση προγενέστερου σχήματος για 2, 20 και 40 επαναλήψεις αντίστοιχα. (b), (d), (f) Με χρήση προγενέστερου σχήματος για 2, 20 και 40 επαναλήψεις αντίστοιχα. (Εικόνα: Άνθρωπος και Χέρι (352x288)).



(a) Περίπτωση 1



(b) Περίπτωση 2



(c) Περίπτωση 3

Εικόνα 4.18: Το γεωδαιτικό ενεργό περίγραμμα αρχικοποιείται στα όρια της εικόνας. Με τη χρήση του προγενέστερου σχήματος, αγνοεί τις ακμές του σώματος και εντοπίζει το χέρι και στις τρεις περιπτώσεις. Οι εικόνες έχουν ληφθεί από Web-Cam. (Εικόνα: 'Ανθρωπος και Χέρι(144x176)).

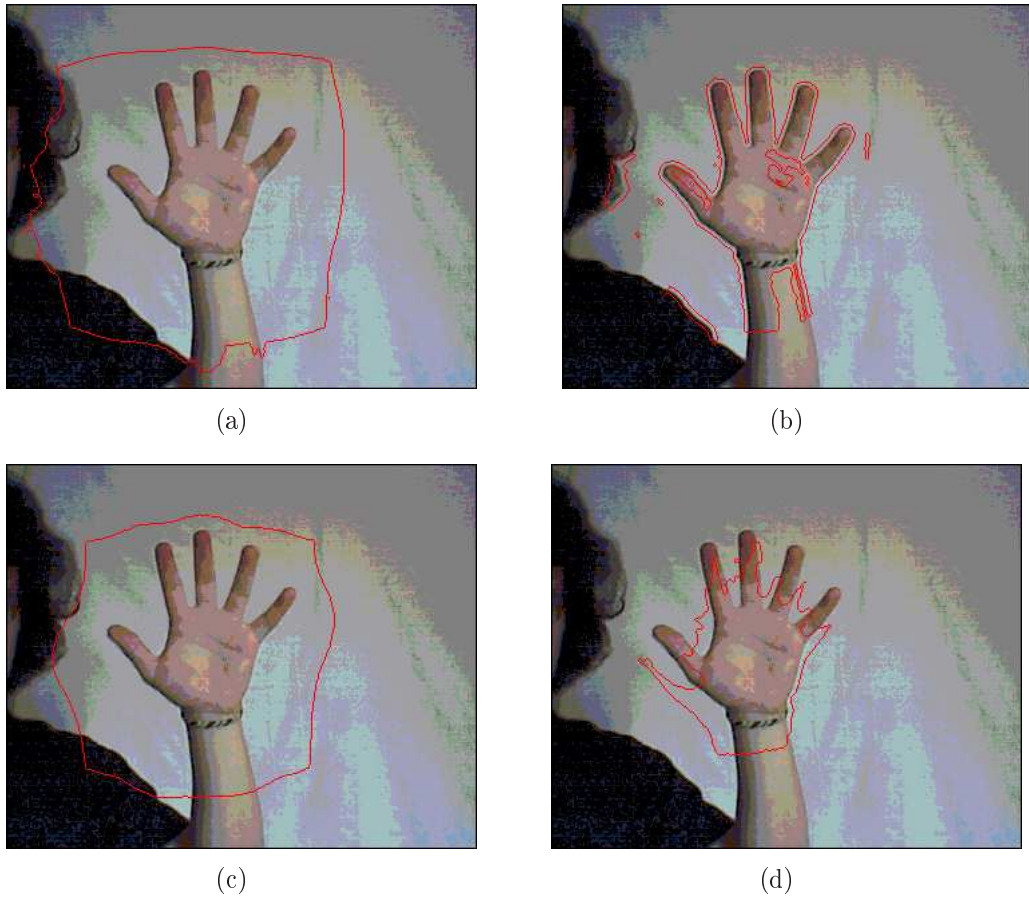


(a) Περίπτωση 1

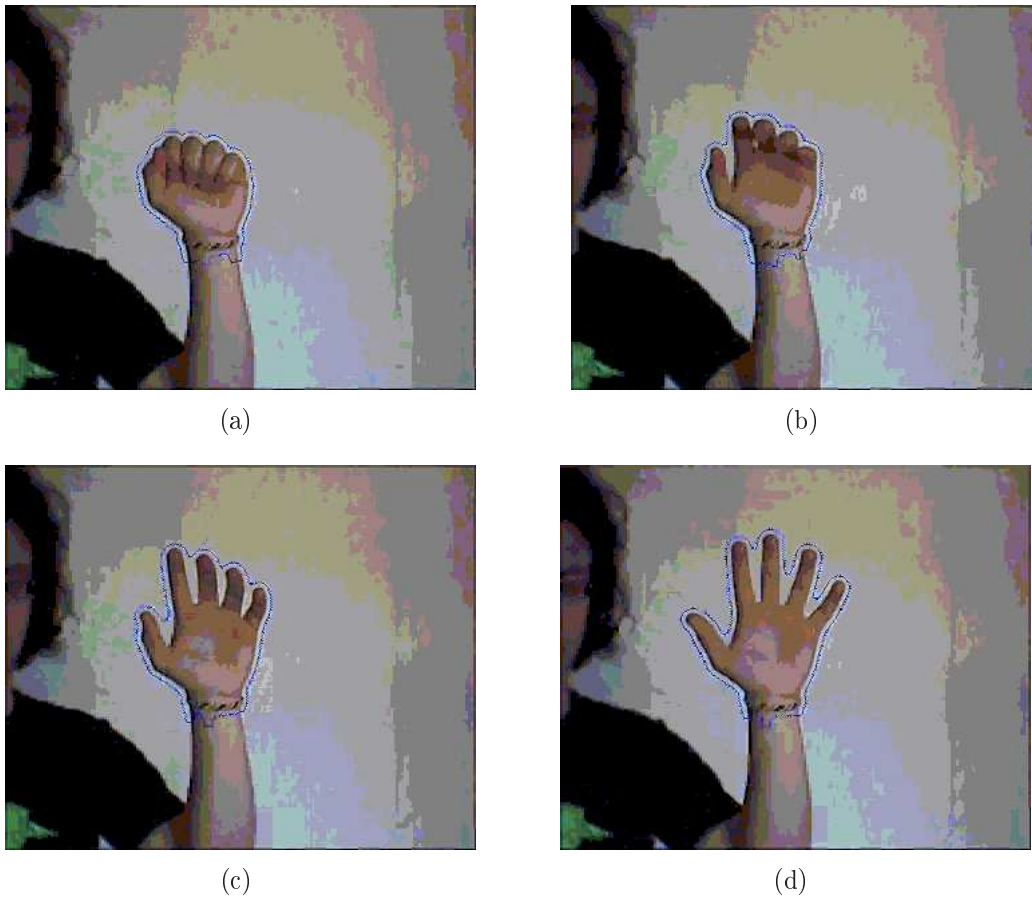


(b) Περίπτωση 2

Εικόνα 4.19: Το γεωδαιτικό ενεργό περίγραμμα αρχικοποιείται στα όρια της εικόνα. Με τη χρήση του προγενέστερου σχήματος, αγνοεί τις ακμές του σώματος και εντοπίζει το χέρι και στις δύο περιπτώσεις. Οι εικόνες έχουν ληφθεί από Web-Cam. (Εικόνα: 'Ανθρωπος και Χέρι (352x288)).



Εικόνα 4.20: Εξέλιξη γεωδαιτικού ενεργού περιγράμματος στα όρια του αλγορίθμου. (a), (b) Εξέλιξη της καμπύλης με μεγάλο συντελεστή δύναμης μπαλονιού, για 10 και 40 επαναλήψεις αντίστοιχα, (c), (d) Εξέλιξη της καμπύλης με μεγάλο συντελεστή δύναμης προγενέστερου σχήματος, για 10 και 40 επαναλήψεις αντίστοιχα. (Εικόνα: Άνθρωπος και Χέρι (352x288)).



Εικόνα 4.21: Η γεωδαιτική ενεργή καμπύλη αρχικοποιείται περικλύοντας μόνο τη γροθιά. Η καμπύλη σε κάθε frame εξελίσσεται πρώτα με χρήση των διανυσμάτων οπτικής ροής, και στη συνέχεια ως γεωδαιτική ενεργή καμπύλη. (a) Frame 10, (b) Frame 40, (c) Frame 80, (d) Frame 120. Παρατηρούμε ότι παρακαλουθείται αρκετά περίπλοκη κίνηση. (Video: Χέρι (352x288), 350 frames).



(a)



(b)

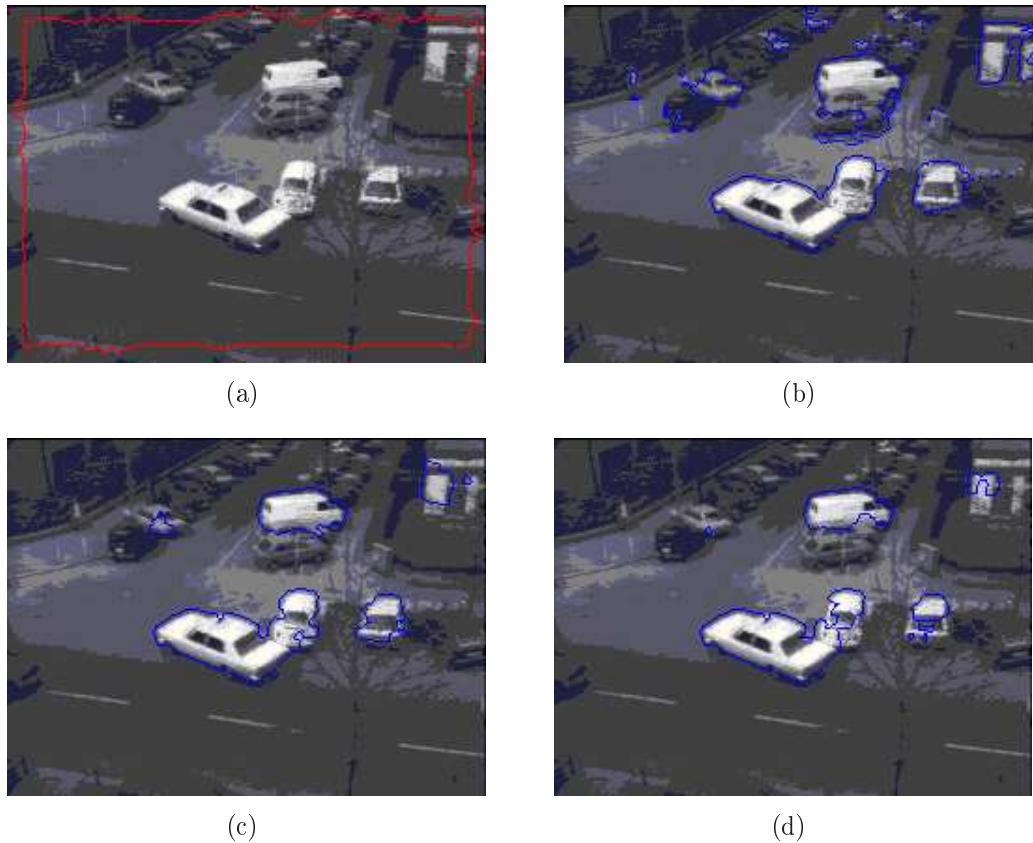


(c)



(d)

Εικόνα 4.22: Η γεωδαιτική καμπύλη εξελίσσεται, όπως και στην εικόνα (4.21).
(a) Frame 160, (b) Frame 200, (c) Frame 280, (d) Frame 350. (Video: Χέρι
(352x288), 350 frames).



Εικόνα 4.23: Παρακολούθηση κίνησης. (a) Frame 1: Το γεωδαιτικό ενεργό περίγραμμα εξελίσσεται ώστε να βρει τα αντικείμενα, (b) Frame 2: Χρησιμοποιώντας πληροφορία οπτικής ροής, η καμπύλη εξελίσσεται ώστε να βρει το κινούμενο αντικείμενο, (c) Η καμπύλη μετά από 5 επαναλήψεις, (d) Η καμπύλη μετά από 9 επαναλήψεις. Παρατηρούμε ότι τα όρια του ταξί παραμένουν, ενώ τα άλλα αντικείμενα χάνονται. (Video: Ταξί Αμβούργου (258x192), 41 frames).



(a)



(b)

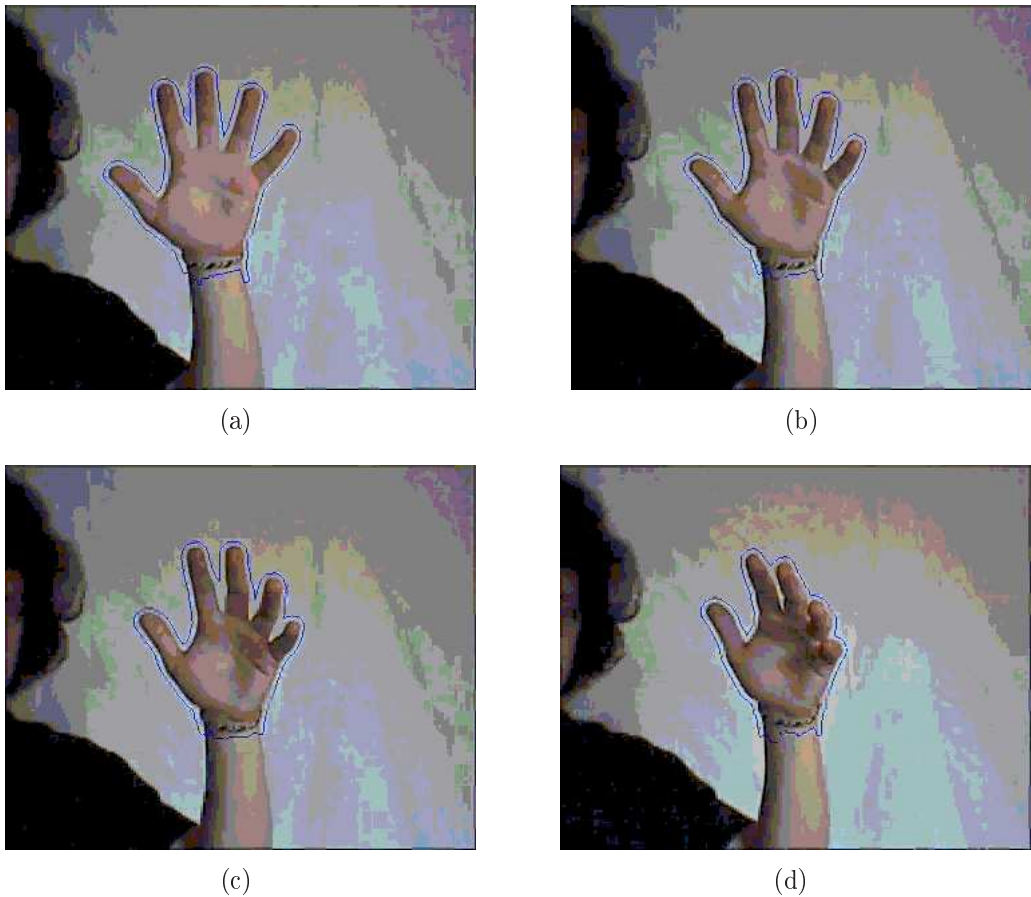


(c)

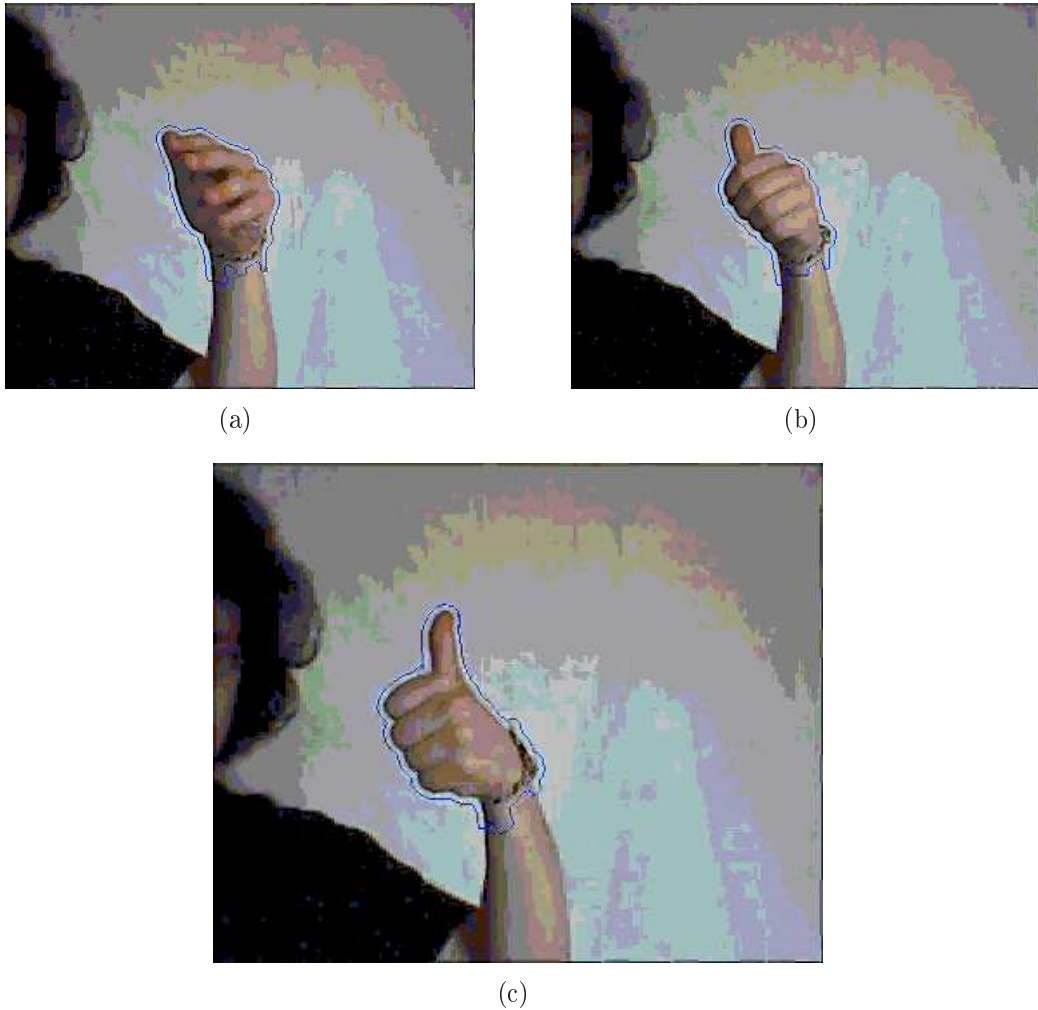


(d)

Εικόνα 4.24: Παρακολούθηση κίνησης. Το γεωδαιτικό ενεργό περίγραμμα εξελίσσεται, με βάση την πληροφορία της οπτικής ροής, τόσο ως διάνυσμα όσο και ως μέτρο. (a) Frame 10, (b) Frame 20, (c) Frame 30, (d) Frame 40. Παρατηρούμε ότι τελικά η καμπύλη περικλύει μόνο το κινούμενο αντικείμενο της αρχικής εικόνας, δηλαδή το ταξί. (Video: Ταξί Αμβούργου (258x192), 41 frames).



Εικόνα 4.25: Παρακολούθηση κίνησης. Η καμπύλη αρχικοποιείται στα όρια της εικόνας, και εξελίσσεται με βάση το προγενέστερο σχήμα, προκειμένου να βρει το επιθυμητό αντικείμενο. Στη συνέχεια, η θέση της ανανεώνεται με χρήση των διανυσμάτων οπτικής ροής, και εξελίσσεται σύμφωνα με το μοντέλο των γεωδαιτικών ενεργών περιγραμμάτων. (a) Frame 10, (b) Frame 40, (c) Frame 70, (d) Frame 100. (Video: Χέρι (352x288), 240 frames).



Εικόνα 4.26: Παρακολούθηση της κίνησης του χεριού, όπως και στην εικόνα (4.25). (a) Frame 150, (b) Frame 200, (c) Frame 230. (Video: Χέρι (352x288), 240 frames).

Κεφάλαιο 5

Ανακεφαλαίωση και Μελλοντικές Κατευθύνσεις Έρευνας

5.1 Ανακεφαλαίωση της Διπλωματικής

Στη διπλωματική αυτή, μελετήσαμε το πρόβλημα παρακολούθησης κινούμενων αντικειμένων. Έγινε εξαρχής σαφές ότι το πρόβλημα αυτό παρουσιάζει άμεση συνάφεια με το πρόβλημα κατάτμησης εικόνας, και έτσι θεωρήθηκε απαραίτητη η μελέτη των τεχνικών κατάτμησης με χρήση εξελισσόμενων καμυλών. Δόθηκε έμφαση σε αυτές τις τεχνικές, διότι αποτελούν τεχνικές που βασίζονται στην ελαχιστοποίηση ενέργειας. Τεχνικές αυτή της φιλοσοφίας είναι εύκολα επεκτάσιμες, μέσω της προσθήκης διαφόρων όρων ενέργειας.

Γρήγορα έγινε αντιληπτό ότι οι κλασσικές μέθοδοι εξέλιξης των καμυλών δεν επαρκούν για την κατάτμηση της εικόνας σε περίπτωση που αυτή παρουσιάζει θόρυβο, ή έντονη πληροφορία ακμών. Έτσι, είναι απαραίτητο να γίνει χρήση πρότερης γνώσης, προκειμένου να εξαχθεί από την εικόνα η επιθυμητή πληροφορία. Η διαπίστωση αυτή μας οδήγησε στη δημιουργία προγενέστερων σχημάτων από ένα σύνολο εκπαίδευσης, και στην χρήση των σχημάτων αυτών ως οδηγών κατά την κατάτμηση. Έμφαση δόθηκε στις τεχνικές οι οποίες ενσωματώνονται άμεσα στο μοντέλο εξέλιξης των γεωδαιτικών ενεργών καμυλών, εξελίσσοντας έτσι την καμπύλη προς την επιθυμητή κατεύθυνση.

Προκειμένου να παρακολουθηθεί το κινούμενο αντικείμενο σε μια αλληλουχία εικόνων, είναι απαραίτητο να βρεθεί πρώτα πληροφορία για την κίνηση. Έτσι, αναφερθήκαμε σε μεθόδους εύρεσης των διανυσμάτων της ταχύτητας στον 2-Δ χώρο (οπτική ροή). Με χρήση των διανυσμάτων αυτών, καταφέραμε να παρακολουθήσουμε ένα αντικείμενο το οποίο έχουμε ήδη εντοπίσει στην εικόνα, ή να παρακολουθήσουμε το αντικείμενο το οποίο κινείται σε μια αλληλουχία εικόνων.

Το βασικό στοιχείο της διπλωματική αυτής είναι η μελέτη πολλών και διαφορετικών μεθόδων κατάτμησης, χρήσης προγενέστερου σχήματος, εύρεσης της οπτικής ροής και παρακολούθησης κίνησης. Τελικά, ενσωματώνουμε στο μοντέλο εξέλιξης των γεωδαιτικών ενεργών περιγραμμάτων, δύο επιπλέον όρους. Ο πρώτος όρος έχει σχέση με την προγενέστερη πληροφορία και οδηγεί την καμπύλη προς την εξαγωγή του επιθυμητού σχήματος. Ο δεύτερος όρος προκύπτει από την εύρεση της οπτικής ροής και οδηγεί την καμπύλη προς των εντοπισμό των κινούμενων περιοχών. Επίσης, η καμπύλη εξελίσσεται και με βάση τα διανύσματα των ταχυτήτων που υπολογίζουμε.

Υλοποιήθηκαν διάφοροι αλγόριθμοι εξέλιξης καμπυλών, εύρεσης και χρήσης προγενέστερων σχημάτων και εύρεσης της οπτικής ροής. Μετά από προσομοιώσεις, διαπιστώθηκαν πλεονεκτήματα και μειονεκτήματα για κάθε αλγόριθμο, και αυτά παρατέθηκαν στις εκάστοτε παραγράφους. Τα τελικά αποτελέσματα παρουσιάστηκαν στο τέταρτο κεφάλαιο, με μια προσπάθεια αντιπροσώπευσης όλων των πεδίων που μελετήθηκαν. Οι κώδικες παρουσιάζουν πολύ καλά αποτελέσματα στην πλεινότητα των περιπτώσεων, ενώ ο συνδυασμός των γεωδαιτικών ενεργών καμπυλών με την πληροφορία της οπτικής ροής, έχει ιδιαίτερα ενδιαφέροντα αποτελέσματα.

5.2 Μελλοντικές Κατευθύνσεις

Έχοντας ως βάση τη διπλωματική αυτή μπορεί κάποιος να προχωρήσει σε βελτιώσεις του αλγορίθμου, και να εισέλθει σε καινούρια πεδία έρευνας:

- Αρχικά, είναι δυνατόν να βρίσκεται το πλήθος και το κέντρο βάρους των αντικειμένων τα οποία βρίσκονται σε μια εικόνα. Η εύρεση του πλήθους των αντικειμένων μπορεί να βρεθεί βρίσκοντας τις συνεκτικές συνιστώσες της επιπεδοσυνάρτησης μηδενικού δυναμικού και το κέντρο βάρους, με απλές μαθηματικές μεθόδους. Η πληροφορία του κέντρου βάρους είναι δυνατόν να χρησιμοποιηθεί για την εύρεση των κινούμενων αντικειμένων. Έτσι, αν σε κάποια διαδοχικά frames το κέντρο βάρους παραμένει περίπου σταθερό, τότε το αντικείμενο αυτό χαρακτηρίζεται ακίνητο, και η καμπύλη που το περικλύει αφαιρείται.
- Στην περίπτωση παρακολούθησης των ανθρώπινων χειρονομιών, έχοντας εξάγει την καμπύλη η οποία παρακολουθεί το χέρι, είναι δυνατόν να δημιουργηθεί μια μοντελοποίηση του χεριού εκείνη τη χρονική στιγμή, και να γίνει τελικά αναγνώριση-ερμηνεία της κίνησης. Ο συνηθέστερος τρόπος είναι η πρόσεγγιση του χεριού με τη χρήση ελλείψεων. Το κέντρο βάρους, οι κύριοι άξονες και ο προσανατολισμός της έλλειψης, χρησιμοποιούνται στη συνέχεια από Μαρκοβιανά Μοντέλα, προκειμένου

η κίνηση να αναγνωριστεί. Ωστόσο, με τη χρήση των προγενέστερων σχημάτων, είναι δυνατόν να ακολουθηθεί μια διαφορετική προσέγγιση. Συγκεκριμένα, μπορούν να δημιουργηθούν διαφορετικά προγενέστερα σχήματα για διάφορες διατάξεις του χεριού (π.χ. Ανοιχτός αντίχειρας, κλειστός δείκτης κ.τ.λ.). Τα προγενέστερα αυτά σχήματα είναι δυνατόν να χρησιμοποιηθούν στη συνέχεια κατά την ελαχιστοποίηση της ενέργειας, έχοντας διαφορετικά βάρη τα οποία υπόκεινται σε βελτιστοποίηση. Έτσι, καθώς εξελίσσεται η καμπύλη, το πιο πιθανό σχήμα θα αποκτά μεγαλύτερο βάρος. Μετά την σύγκλιση, οι συντελεστές μπορούν να χρησιμοποιηθούν ως γνώση της διάταξης του χεριού, και κατ' επέκταση για την αναγνώριση των διαδοχικών χειρονομιών. Αυτό μπορεί και πάλι να γίνει με χρήση Μαρκοβιανών μοντέλων, έχοντας ωστόσο πληροφορία που είναι πιο αντιπροσωπευτική από τις ελλείψεις, για τις πιθανές διατάξεις του χεριού. Γενικά τα Μαρκοβιανά μοντέλα θεωρούνται η εγκυρότερη μέθοδος για την αναγνώριση ανθρώπινων χειρονομιών και η χρήση του είναι αρκετά μελετημένη.

- Όσον αφορά την ταχύτητα εξέλιξης των καμπυλών, είναι δυνατόν να χρησιμοποιηθούν σχήματα ταχείας εξέλιξης. Συγκεκριμένα, η επιβράδυνση στην σύγκλιση οφείλεται στον περιορισμό του βήματος εξέλιξης. Μεγάλο βήμα εξέλιξης οδηγεί σε αστάθεια, ενώ μικρό βήμα εξέλιξης οδηγεί σε πολύ αργή σύγκλιση. Επομένως, πρέπει να χρησιμοποιηθούν κατάλληλοι μέθοδοι επιλογής του βήματος εξέλιξης, όπως για παράδειγμα προτείνεται στο [26].

Όπως είναι εμφανές, η εργασία που έχει γίνει στην παρούσα διπλωματική μπορεί να αποτελέσει βάση για την ανάπτυξη καλών και γρήγορων αλγορίθμων παρακολούθησης κίνησης και ερμηνείας ανθρώπινων χειρονομιών.

Βιβλιογραφία

- [1] Μαραγκός Π. *Ανάλυση Εικόνων και Όραση Υπολογιστών*. Ιούνιος 2005. Αθήνα.
- [2] A. Argyros and M. Lourakis. Tracking skin-colored objects in real-time.
- [3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *Int'l Journal of Computer Vision*, 12(1):43–77, 1994.
- [4] T. Brox, A. Bruhn, N. Papenbergh, and J. Weickert. High accuracy optical flow based on a theory for warping. *Proc. 8th European Conference on Computer Vision*, 4:25–36, May 2004.
- [5] A. Bruhn, J. Weickert, and C. Schnoerr. Lukas/kanade meets horn/schunck: Combining local and global optical flow techniques. *Int'l Journal of Computer Vision*, 61(3):211–231, 2005.
- [6] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *Int'l Journal of Computer Vision*, 22(1):61–79, 1997.
- [7] T. Chan and L. Vese. Active contours without edges.
- [8] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik. A real-time computer vision system for vehicle tracking and traffic surveillance.
- [9] D. Cremers. Dynamical statistical shape priors for level set based tracking.
- [10] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.
- [11] D. Geer. Will gesture recognition technology point the way? *Computer Magazine*, October 2004.

- [12] J. Ha, C. Alvino, G. Pryor, M. Niethammer, E. Johnson, and A. Tannenbaum. Active contours and optical flow for automatic tracking of flying vehicles. *Proc. American Control Conference*, pages 3441–3446, June 2004.
- [13] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, October 1998.
- [14] B. K. P Horn and B. G Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
- [15] L. F. Hotraphinyo and C. N. Riviere. Precision measurement for microsurgical instrument evaluation. *Proc. IEEE 23rd Annual EMBS International Conference*, pages 3454–3457, October 2001.
- [16] S. Hutchinson, G. D. Hager, and Corke P. I. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, October 1996.
- [17] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int'l Journal of Computer Vision*, 1987.
- [18] M. Leventon, O. Faugeras, W. Grimson, and III W. Wells. Level set based segmentation with intensity and curvature priors. In *Workshop on Mathematical Methods in Biomedical Image Analysis Proceedings*, pages 4–11, June 2000.
- [19] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. pages 316–323.
- [20] B. D. Lukas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. Image Understanding Workshop*, pages 121–130, 1981.
- [21] A. R Mansouri. Region tracking via level-set pdes without motion computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):947–961, July 2002.
- [22] M. Niethammer, A. Tannenbaum, and S. Angenent. Dynamical active contours for visual tracking. *IEEE Transactions on Automatic Control*, October 2005. Draft.

-
- [23] K. Oka, Y. Sato, and H. Koike. Real-time fingertip tracking and gesture recognition. *IEEE Computer Graphics and Applications*, pages 64–71, November/December 2002.
- [24] S. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, (79):12–49, 1988.
- [25] I. B Ozer, T. Lu, and W. Wolf. Design of a real-time gesture recognition system. *IEEE Signal Processing Magazine*, May 2005.
- [26] G. Papandreou and P. Maragos. A fast multigrid implicit algorithm for the evolution of geodesic active contours. *Proc. Int’l Conference on Computer Vision and Pattern Recognition (CVPR-2004)*, June 2004.
- [27] N. Paragios. *Geodesic Active Regions and Level Set Methods: Contributions and Applications in Artificial Vision*. PhD thesis, University of Nice Sophia Antipolis, January 2000.
- [28] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3):266–280, 2000.
- [29] N. Paragios and M. Rousson. Shape priors for level set representation.
- [30] N. Paragios, M. Rousson, and V. Ramesh. Matching distance functions: A shape-to-area variational approach for global-to-local registration.
- [31] T. Riklin-Raviv, N. Kiryati, and N. Sochen. Prior-based segmentation by projective registration and level sets.
- [32] M. Rousson, N. Paragios, and R. Deriche. Active shape models from a level set perspective. *INRIA Rapport de Recherche*, 3(4984), October 2003.
- [33] S. Saripalli, J. F. Montgomery, and G. S. Sukhatme. Vision based autonomous landing of an unmanned aerial vehicle. *Proc. IEEE Int’l Conference on Robotics and Automation*, pages 2799–2804, May 2002.
- [34] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision and Materials Science*. Cambridge University Press, 2nd edition, 1999.

-
- [35] J. A. Sethian and D. Adalsteinsson. A fast level set method for propagating interfaces. September 1994.
 - [36] T. Starner and A. Pentland. Real-time american sign language recognition from video using hidden markov models. In *SCV95*, page 5B Systems and Applications, 1995.
 - [37] T. Starner, J. Weaver, and A. Pentland. Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375, 1998.
 - [38] A. Tsai, Jr. A. Yezzi, III W. Wells, and C. Tempny. Model based curve evolution technique for image segmentation.
 - [39] A. Vasile and R. Marino. Pose independent automatic target detection and recognition. *Lincoln Laboratory Journal*, 15(1):61–78, 2005.
 - [40] G. Welch and E. Foxlin. Motion tracking: No sliver bullet but a respectable arsenal. *IEEE Computer Graphics and Applications*, pages 24–38, November/December 2002.