

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

**Μια εναλλακτική προσέγγιση για τη μελέτη της
αντίληψης του χώρου:
Χρήση διατακτικού αντί μετρικού βάθους**

Μανόλης Ι. Α. Λουράκης

Μεταπτυχιακή Εργασία

Ηράκλειο 1995

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

**Μια εναλλακτική προσέγγιση για τη μελέτη της
αντίληψης του χώρου:
Χρήση διατακτικού αντί μετρικού βάθους**

Εργασία που υποβλήθηκε από τον
Μανόλη Ι. Α. Λουράκη
ως μερική απαίτηση για την απόκτηση του
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Ιανουάριος 1995

Συγγραφέας:

Τμήμα Επιστήμης Υπολογιστών

Εισηγητική Επιτροπή:

Στέλιος Ορφανουδάκης, Καθηγητής, Επόπτης

Γιώργος Τζιρίτας, Αναπληρωτής Καθηγητής, Μέλος

Χρήστος Νικολάου, Αναπληρωτής Καθηγητής, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος, Αναπληρωτής Καθηγητής
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Στην οικογένειά μου

Ευχαριστίες

Κατ' αρχήν ευχαριστώ την οικογένειά μου για την υποστήριξή της και την αγάπη με την οποία με περιέβαλλε μέχρι σήμερα.

Η εργασία αυτή πραγματοποιήθηκε με την καθοδήγηση του καθηγητή του Τμήματος Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης Στέλιου Ορφανουδάκη, τον οποίο και ευχαριστώ για την συνεργασία και τις συμβουλές του.

Πολύτιμες υπήρξαν οι συζητήσεις που είχα κατά καιρούς με τον καθηγητή Γιάννη Αλοίμονο, την Dr. Cornelia Fermüller, τον καθηγητή Πάνο Τραχανιά και τους συναδέλφους μου Αντώνη Αργυρό και Γιάννη Καβακλή. Τους ευχαριστώ όλους θερμά.

Ευχαριστώ επίσης τα μέλη της εξεταστικής επιτροπής, καθηγητές Γιώργο Τζιρίτα και Χρήστο Νικολάου για τις παρατηρήσεις και υποδείξεις τους.

Τελειώνοντας, θα ήθελα να ευχαριστήσω το Ινστιτούτο Πληροφορικής του ΙΤΕ, η οικονομική και τεχνική συνδρομή του οποίου υπήρξε πολύ σημαντική για την ολοκλήρωση της εργασίας.

Μια εναλλακτική προσέγγιση για τη μελέτη της αντίληψης του χώρου: Χρήση διατακτικού αντί μετρικού βάθους

Μανόλης Ι. Α. Λουράκης

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Περίληψη

Μια από τις πιο σημαντικές ικανότητες ενός έλλογου συστήματος είναι αυτή της πλοήγησης, η ικανότητα δηλαδή της αυτόνομης κίνησης στο περιβάλλον του με βάση τα δεδομένα που προκύπτουν από τις μετρήσεις διαφόρων αισθητήρων. Για να μπορέσει ένα τέτοιο σύστημα να επιτύχει τους εκάστοτε στόχους του, θα πρέπει να αλληλεπιδρά με έναν δυναμικά μεταβαλλόμενο και κάποιες φορές άγνωστο περιβάλλοντα χώρο. Κατά συνέπεια, το σύστημα θα πρέπει να διαθέτει κάποιας μορφής αντίληψη του χώρου, να διαθέτει δηλαδή ικανότητες που να υποστηρίζουν διάφορες λειτουργίες πλοήγησης. Αναμφίβολα, ένα πολύ σημαντικό στοιχείο για την αντίληψη του χώρου είναι η αντίληψη του βάθους, της τρίτης διάστασης του κόσμου η οποία χάνεται κατά την προβολή των τμημάτων του σε επίπεδες επιφάνειες.

Διάφορες μέθοδοι έχουν χρησιμοποιηθεί μέχρι σήμερα για τον υπολογισμό του βάθους των σημείων που απεικονίζονται σε μια ή περισσότερες εικόνες. Πληροφορίες που προέρχονται από περισσότερες από μια κάμερα, από την κίνηση μιας κάμερας ως προς το περιβάλλον, την ένταση του φωτός που ανακλούν τα αντικείμενα μιας εικόνας και την κατανομή κάποιων φωτεινών προτύπων (brightness patterns) σε μια εικόνα, αποτέλεσαν την βάση για την ανάπτυξη μεθόδων γνωστών με το γενικό όνομα *structure from X*, όπου X μπορεί να είναι stereo, motion, shading, texture. Παραδοσιακά, οι μέθοδοι αυτές στοχεύουν στον υπολογισμό μετρικού (ποσοτικού) βάθους, δηλαδή στην ακριβή εκτίμηση της απόστασης κάθε σημείου από τον παρατηρητή. Η μεγάλη και μακρόχρονη προσπάθεια που έχει γίνει

προς την κατεύθυνση υπολογισμού μετρικού βάθους κατέστησε σαφές ότι ο υπολογισμός αυτός είναι πολύ δύσκολος. Έτσι, παρά την εμφάνιση συστημάτων τα οποία καταφέρνουν να υπολογίζουν μετρικό βάθος σε πολύ ειδικές περιπτώσεις, μέχρι σήμερα δεν έχει γίνει δυνατή η ανάπτυξη τεχνικών ικανών να υπολογίζουν με ακρίβεια μετρικό βάθος σε γενικά, πραγματικά περιβάλλοντα. Είναι όμως πιθανό ότι η κατανόηση κάποιων χαρακτηριστικών του χώρου δεν απαιτεί κατ' ανάγκη τον υπολογισμό μετρικού βάθους.

Τα αποτελέσματα ψυχοφυσικών πειραμάτων που έγιναν χρησιμοποιώντας ανθρώπους ή ζώα ως υποκείμενα, ενισχύουν την άποψη ότι ένας οργανισμός δεν εκτιμά τα χαρακτηριστικά του περιβάλλοντός του με βάση “εξωγενείς” μονάδες μέτρησης (π.χ. μέτρα, δευτερόλεπτα κ.λπ.) αλλά χρησιμοποιεί μονάδες μέτρησης που σχετίζονται με το σώμα του, τις ικανότητες κίνησής του και την θέση του στο περιβάλλον. Επιπλέον, οι πληροφορίες που εξάγονται είναι ποιοτικές και όχι μετρικές. Βασισμένη σε τέτοιες παρατηρήσεις, η εργασία αυτή παρουσιάζει μια πρόταση για τη μελέτη της αντίληψης του χώρου και ασχολείται με τον υπολογισμό διατακτικού βάθους. Με τον όρο “διατακτικό βάθος” εννοούμε την διάταξη των σημείων μιας σκηνής με βάση το βάθος τους, την απόστασή τους δηλαδή από τον παρατηρητή. Το διατακτικό βάθος παρέχει σαφώς λιγότερες πληροφορίες από ότι το μετρικό, πιστεύουμε όμως ότι μπορεί να είναι πολύ χρήσιμο στην κατανόηση του χώρου.

Η προσέγγιση που παρουσιάζουμε, στοχεύει στον υπολογισμό ποιοτικών πληροφοριών, κατάλληλων για την κατανόηση του χώρου και ακολουθεί τις κατευθύνσεις της θεωρίας της τελεολογικής και ποιοτικής ενεργούντος όρασης. Συνοπτικά, η μέθοδος υπολογισμού διατακτικού βάθους που αναπτύχθηκε, χρησιμοποιεί σαν είσοδο ένα fixating ζεύγος εικόνων. Κάθε ζεύγος εικόνων που καταγράφεται μια δεδομένη χρονική στιγμή από τις δύο κάμερες ενός στερεοσκοπικού συστήματος, μπορεί να θεωρηθεί ότι προέρχεται από την κίνηση της μιας κάμερας στην θέση της άλλης. Αυτή η υποθετική κίνηση και η προσέγγιση της διαδικασίας σχηματισμού των εικόνων με το μοντέλο της ορθογραφικής προβολής, οδηγεί σε απλές εξισώσεις κίνησης για τα σημεία των εικόνων, οι οποίες επιτρέπουν τον υπολογισμό της σχέσης δύο σημείων ως προς το βάθος. Μια πιθανή χρήση της αναπαράστασης που προκύπτει, αφορά την ποιοτική περιγραφή του σχήματος των επιφανειών που περιέχονται σε μια σκηνή.

Επόπτης : Στέλιος Ορφανουδάκης, Καθηγητής

An alternative approach for studying space perception: Use of ordinal instead of metric depth

Manolis I. A. Lourakis

Master of Science Thesis

Department of Computer Science
University of Crete

Abstract

One of the most important capabilities of an intelligent system is that of navigation, that is the capability of autonomous motion in the environment, based on the measurements provided by various sensors. In order for such a system to achieve its goals, it should interact with dynamically changing and probably unknown surroundings. Consequently, the system must possess some sort of space perception, in other words possess capabilities that support various navigation functions. Undoubtedly, an important component of space perception is the perception of depth, the third world dimension that is lost during the projection of world parts on planar surfaces.

Up to now, various methods have been used to compute the depth of points imaged in one or more frames. Information derived from more than one cameras, the motion of a camera with respect to the environment, the intensity of light reflected by the objects in an image and the distribution of some brightness patterns in an image, has formed the basis for the development of methods known with the generic name *structure from X*, where X can be one of stereo, motion, shading, texture. Traditionally, these methods aim at computing metric (quantitative) depth, in other words they aim at accurately estimating the distance of every point from the observer. The enormous and long-term effort towards computing metric depth, has made clear that this computation is very difficult. Although systems capable of computing metric depth in restricted cases have appeared, their generalization in general, real-world environments has not been successful. It is possible, however, that the understanding of some aspects of space does not necessarily require the computation of metric depth.

Results of psychophysical experiments performed using human or animal subjects, suggest that an organism does not perceive the characteristics of the environment in extrinsic units (such

as meters, seconds, etc) but employs units related to its body, motion capabilities and position in the environment instead. Moreover, the extracted information is qualitative rather than metric. Relying on such observations, this thesis proposes a qualitative approach for studying space perception and deals with the computation of *ordinal depth*. The term “ordinal depth” denotes the ordering of the points of a scene according to their depth, that is their distance from the observer. Clearly, ordinal depth provides less information compared to that provided by metric depth, nevertheless we believe that it can be very useful in understanding space.

The approach that will be presented, aims at computing qualitative information, appropriate for space understanding, and follows the guidelines of the theory of purposive and qualitative active vision. In short, the method we developed for ordinal depth computation, uses a fixating image pair as input. Each image pair, acquired by the two cameras of a stereo system at a given time instance, can be seen as resulting from the movement of the one camera to the position of the other. This hypothetical motion and the approximation of the image formation process with the orthographic projection model, leads to simple motion equations for the image points, which permit the computation of the relationship of two points with respect to depth. A possible use of the resulting representation regards the qualitative description of the shape of surfaces contained in a scene.

Supervisor : Stelios Orphanoudakis, Professor

Περιεχόμενα

| | |
|--|-------------|
| Περιεχόμενα | v |
| Κατάλογος Σχημάτων | viii |
| Κατάλογος Πινάκων | x |
| 1 Εισαγωγή | 1 |
| 1.1 Υπόβαθρο | 1 |
| 1.2 Οργάνωση της εργασίας | 4 |
| 2 Τελεολογική και ποιοτική ενεργός όραση | 6 |
| 2.1 Εισαγωγή | 6 |
| 2.2 Οι δεσπόζουσες θεωρίες για την όραση | 6 |
| 2.3 Μια νέα προσέγγιση | 9 |
| 2.4 Υπολογιστικά πλεονεκτήματα της μεθοδολογίας της ενεργούς όρασης | 12 |
| 2.5 Ερευνητικά προβλήματα στα πλαίσια της θεωρίας της ενεργού όρασης | 13 |
| 3 Οπτική πλοιήγηση | 19 |
| 3.1 Εισαγωγή | 19 |
| 3.2 Ανασκόπηση της περιοχής | 20 |
| 3.2.1 Structure from motion | 21 |
| 3.2.2 Προγραμματισμός κίνησης | 22 |

| | | |
|----------|--|-----------|
| 3.3 | Μια εναλλακτική προσέγγιση | 25 |
| 4 | Κάθετη ροή | 29 |
| 4.1 | Εισαγωγή | 29 |
| 4.2 | Οπτική ροή και πεδίο κίνησης | 29 |
| 4.3 | Ποσοτική περιγραφή της φαινόμενης κίνησης | 30 |
| 4.4 | Το σύστημα συντεταγμένων | 35 |
| 5 | Διατακτικό βάθος | 37 |
| 5.1 | Εισαγωγή | 37 |
| 5.2 | Η είσοδος | 38 |
| 5.3 | Το μοντέλο σχηματισμού των εικόνων | 40 |
| 5.4 | Υπολογισμός διατακτικού βάθους | 41 |
| 5.5 | Πολικός και λογαριθμικός πολικός μετασχηματισμός | 44 |
| 5.5.1 | Πολικός μετασχηματισμός | 46 |
| 5.5.2 | Λογαριθμικός πολικός μετασχηματισμός | 46 |
| 5.6 | Υλοποίηση | 49 |
| 5.6.1 | Υπολογισμός λογαριθμικού πολικού μετασχηματισμού | 49 |
| 5.6.2 | Υπολογισμός των μερικών παραγώγων | 52 |
| 5.6.3 | Διατακτική αναπαράσταση του βάθους | 54 |
| 6 | Περιγραφή του σχήματος των επιφανειών | 57 |
| 6.1 | Εισαγωγή | 57 |
| 6.2 | Καμπυλότητα καμπυλών και επιφανειών | 57 |
| 6.3 | Υπολογισμός του προσήμου της Gaussian καμπυλότητας | 61 |
| 6.4 | Υλοποίηση και αποτελέσματα | 63 |
| 6.4.1 | Συνθετικά δεδομένα | 63 |
| 6.4.2 | Πραγματικά δεδομένα | 63 |

| | |
|--|----|
| A Σχέσεις μεταξύ των καρτεσιανών και των λογαριθμικών πολικών μερικών παραγώγων | 72 |
| B Πρόσημο της Gaussian καμπυλότητας για την επιφάνεια που ορίζει η $Z(x, y)$ | 75 |
| C Διατακτικό βάθος στην περίπτωση προοπτικής προβολής | 77 |
| C.1 Οι εξισώσεις για την κάθετη ροή στην περίπτωση προοπτικής προβολής | 77 |
| C.2 Σχέση μεταξύ των παραμέτρων U, W, β για ένα fixating ζεύγος | 78 |
| C.3 Διάταξη σημείων με βάση το βάθος σε ειδικές περιπτώσεις | 80 |

Κατάλογος Σχημάτων

| | | |
|-----|--|----|
| 2.1 | Μια μηχανική κεφαλή και οι βαθμοί ελευθερίας της. | 14 |
| 2.2 | Δομικά στοιχεία ενός τελεολογικού συστήματος όρασης. | 17 |
| 3.1 | Ασάφεια στην αντιστοίχιση των κουκίδων. | 22 |
| 4.1 | Μια σκηνή και ένα πεδίο κάθετης ροής γι' αυτήν. | 32 |
| 4.2 | Η φορά της κάθετης ροής περιορίζει την οπτική ροή σε ένα ημιεπίπεδο. | 33 |
| 4.3 | Το πρόβλημα του ανοίγματος. | 34 |
| 4.4 | Το σύστημα συντεταγμένων. | 35 |
| 4.5 | Περιστροφή του σημείου P γύρω από τον άξονα L . | 36 |
| 5.1 | Η γεωμετρία του συστήματος των δύο καμερών. | 38 |
| 5.2 | Disparity σε ένα στερεοσκοπικό σύστημα | 39 |
| 5.3 | Κότοψη των συστημάτων συντεταγμένων για τις δύο κάμερες. | 42 |
| 5.4 | 'Ενα fixating ζεύγος εικόνων. | 45 |
| 5.5 | Διαφορές των εντάσεων για το ζεύγος των εικόνων του σχήματος 5.4. | 46 |
| 5.6 | Πολικός μετασχηματισμός. | 47 |
| 5.7 | Πολικός μετασχηματισμός και απεικόνισή του στο επίπεδο της εικόνας. | 48 |
| 5.8 | Λογαριθμικός πολικός μετασχηματισμός. | 49 |
| 5.9 | Λογαριθμικός πολικός μετασχηματισμός και απεικόνισή του στο επίπεδο της εικόνας. | 50 |

| | |
|--|----|
| 5.10 Διαφορές των εντάσεων για το ζεύγος των εικόνων του σχήματος 5.4, υπολογισμένες στο πεδίο (ξ, η) | 51 |
| 5.11 Διατακτικό βάθος για το ζεύγος του σχήματος 5.4 | 56 |
| | |
| 6.1 Επιφάνειες με διαφορετικό πρόσημο Gaussian καμπυλότητας. | 60 |
| 6.2 Δύο όψεις της γραφικής παράστασης της συνάρτησης 6.12. | 64 |
| 6.3 Πρόσημο της Gaussian καμπυλότητας για την συνάρτηση 6.12. | 65 |
| 6.4 Ιστόγραμμα των G που υπολογίζονται με βάση τις εικόνες του σχήματος 5.4. . | 65 |
| 6.5 Πρόσημο της Gaussian καμπυλότητας για τη σκηνή του σχήματος 5.4 | 66 |
| | |
| C.1 Μετακίνηση του συστήματος $O'X'Y'Z'$ στο $OXYZ$ | 79 |

Κατάλογος Πινάκων

| | |
|---|----|
| 2.1 Σύγκριση των προσεγγίσεων στο πρόβλημα της όρασης. | 13 |
| 6.1 Είδη επιφανειών με βάση το πρόσημο της Gaussian καμπυλότητας. | 61 |

Κεφάλαιο 1

Εισαγωγή

1.1 Υπόβαθρο

Μια από τις πιο σημαντικές ικανότητες ενός έλλογου συστήματος είναι αυτή της πλοήγησης, η ικανότητα δηλαδή της αυτόνομης κίνησης στο περιβάλλον του με βάση τα δεδομένα που προκύπτουν από τις μετρήσεις διαφόρων αισθητήρων. Για να μπορέσει ένα τέτοιο σύστημα να επιτύχει τους εκάστοτε στόχους του, θα πρέπει να αλληλεπιδρά με έναν δυναμικά μεταβαλλόμενο και κάποιες φορές άγνωστο περιβάλλοντα χώρο. Κατά συνέπεια, το σύστημα θα πρέπει να διαθέτει κάποιας μορφής αντίληψη του χώρου, να διαθέτει δηλαδή ικανότητες που να υποστηρίζουν διάφορες λειτουργίες πλοήγησης. Τέτοιες λειτουργίες είναι η εκτίμηση των παραμέτρων κίνησής του ως προς το περιβάλλον, η αποφυγή εμποδίων, η χρήση χαρακτηριστικών τμημάτων του περιβάλλοντος σαν σημεία αναφοράς, ο προγραμματισμός της κίνησης μεταξύ διαφορετικών σημείων του περιβάλλοντος κ.ά. Οι λειτουργίες αυτές έχουν διαφορετικούς βαθμούς πολυπλοκότητας, και απαιτήσεις δεδομένων οι οποίες κυμαίνονται από τοπικές μετρήσεις έως την προσεγγιστική a priori γνώση της τοπολογίας του περιβάλλοντος και των αναπαραστάσεων της μορφής κάποιων τμημάτων του. Αναμφίβολα, ένα πολύ σημαντικό στοιχείο για την αντίληψη του χώρου είναι η αντίληψη του βάθους, της τρίτης διάστασης του κόσμου η οποία χάνεται κατά την προβολή των τμημάτων του σε επίπεδες επιφάνειες.

Διάφορες μέθοδοι έχουν χρησιμοποιηθεί μέχρι σήμερα για τον υπολογισμό του βάθους των σημείων που απεικονίζονται σε μια ή περισσότερες εικόνες. Πληροφορίες που προέρχονται από περισσότερες από μια κάμερα, από την κίνηση μιας κάμερας ως προς το περιβάλλον, την ένταση του φωτός που ανακλούν τα αντικείμενα μιας εικόνας και την κατανομή κάποιων φωτεινών προτύπων (brightness patterns) σε μια εικόνα, αποτέλεσαν την βάση για την ανάπτυξη μεθόδων γνωστών με το γενικό όνομα *structure from X*, όπου X μπορεί να είναι stereo, motion, shading, texture [60]. Παραδοσιακά, οι μέθοδοι αυτές στοχεύουν στον υπολογισμό

1.1 Υπόβαθρο

μετρικού (ποσοτικού) βάθους, δηλαδή στην ακριβή εκτίμηση της απόστασης κάθε σημείου από την κάμερα. Η μεγάλη και μακρόχρονη προσπάθεια που έχει γίνει προς την κατεύθυνση υπολογισμού μετρικού βάθους κατέστησε σαφές ότι ο υπολογισμός αυτός είναι πολύ δύσκολος. Έτσι, παρά την εμφάνιση συστημάτων τα οποία καταφέρνουν να υπολογίζουν μετρικό βάθος σε πολύ ειδικές περιπτώσεις, μέχρι σήμερα δεν έχει γίνει δυνατή η ανάπτυξη τεχνικών ικανών να υπολογίζουν μετρικό βάθος σε γενικά, πραγματικά περιβάλλοντα. Είναι όμως πιθανό ότι η κατανόηση κάποιων χαρακτηριστικών του χώρου δεν απαιτεί κατ' ανάγκη τον υπολογισμό μετρικού βάθους. Συμπεράσματα που έχουν εξαχθεί από διάφορα ψυχοφυσικά πειράματα, αμφισβητούν κατά πόσον τα συστήματα όρασης των βιολογικών συστημάτων υπολογίζουν μετρικό βάθος [40].

Η μελέτη διαφόρων ζώων, αλλά και η προσεκτική παρατήρηση των ανθρώπινων δοσοληψιών με το περιβάλλον, οδηγούν στο συμπέρασμα ότι ένας οργανισμός δεν εκτιμά τα χαρακτηριστικά του περιβάλλοντός του με βάση “εξωγενείς” μονάδες μέτρησης (π.χ. μέτρα, δευτερόλεπτα κ.λπ.) αλλά χρησιμοποιεί μονάδες μέτρησης που σχετίζονται με το σώμα του, τις ικανότητες κίνησής του και την θέση του στο περιβάλλον. Ένας άνθρωπος π.χ., εκτιμά κατά πόσο μπορεί να πιάσει αντικείμενα του περιβάλλοντος χωρίς να μετακινήσει τα υπόλοιπα μέλη του σώματος του με βάση το μήκος των χεριών του, κατά πόσο μπορεί να αποφύγει ένα κινούμενο αντικείμενο που έρχεται προς το μέρος του με βάση την δική του ταχύτητα αντίδρασης κ.λπ. Με άλλα λόγια, οι μονάδες μέτρησης που χρησιμοποιούν οι βιολογικοί οργανισμοί ορίζονται με βάση το σώμα τους και τις μηχανικές ικανότητές (motor capabilities) τους. Επομένως, πολλές από τις πληροφορίες που υπολογίζει ένας οργανισμός για το περιβάλλον του παίρνουν τιμές από ένα σύνολο με σχετικά μικρό πληθικό αριθμό και έχουν ποιοτικό χαρακτήρα. Στο σημείο αυτό θα πρέπει να παρατηρήσουμε ότι δεδομένων κάποιων ποσοτικών πληροφοριών, είναι εύκολη η μετατροπή τους σε μονάδες οριζόμενες με βάση τις μηχανικές ικανότητες κάποιου οργανισμού. Αν όμως οι πληροφορίες αυτές δεν πρόκειται να χρησιμοποιηθούν εκφρασμένες σε “εξωγενείς” μονάδες, ποιός ο λόγος να μην υπολογιστούν εξ’ αρχής αναφορικά με το σώμα και τις μηχανικές ικανότητες κάποιου οργανισμού, χρησιμοποιώντας πιθανότατα πιο εύκολους υπολογισμούς; Ένα παρεμφερές επιχείρημα κατά του υπολογισμού ποσοτήτων εκφρασμένων σε εξωγενείς μονάδες, είναι το γεγονός ότι σε πολλές περιπτώσεις οι ποσότητες αυτές παρέχουν πολύ πιο λεπτομερείς πληροφορίες από ότι θα ήταν αναγκαίες, κατά συνέπεια μπορούν να αντικατασταθούν από ποιοτικές (περιγραφικές) εκτιμήσεις.

Για τους λόγους που αναφέρθηκαν παραπάνω, η παρούσα εργασία ασχολείται με τον υπολογισμό διατακτικού και όχι μετρικού βάθους. Παρουσιάζει επίσης μια πρόταση για τη μελέτη της αντίληψης του χώρου, η οποία θα χρησιμοποιεί πληροφορίες που παρέχονται από το διατακτικό βάθος. Με τον όρο “διατακτικό βάθος” εννοούμε την διάταξη των σημείων μιας σκηνής με βάση το βάθος τους, την απόστασή τους δηλαδή από τον παρατηρητή. Για να καταστεί δυνατή μια τέτοια ταξινόμηση, πρέπει να υπολογιστεί η σχέση ως προς το βάθος ενός

1.1 Υπόβαθρο

οποιουδήποτε ζεύγους σημείων, να αναγνωριστεί δηλαδή το σημείο του ζεύγους που είναι κοντύτερα στον παρατηρητή. Το διατακτικό βάθος παρέχει σαφώς λιγότερες πληροφορίες από ότι το μετρικό, πιστεύουμε όμως ότι μπορεί να είναι πολύ χρήσιμο στην κατανόηση του χώρου. Στο σημείο αυτό κρίνεται σκόπιμο να παρατηρήσουμε ότι ο υπολογισμός μετρικού βάθους είναι ουσιαστικά ο υπολογισμός της ταυτοτικής συνάρτησης

$$f(Z) = Z$$

όπου $Z > 0$ είναι το βάθος, για κάθε σημείο μιας σκηνής. Ο υπολογισμός διατακτικού βάθους αφορά τον υπολογισμό της συνάρτησης

$$g(Z_1, Z_2) = \begin{cases} 1 & \text{αν } Z_1 \geq Z_2 \\ 0 & \text{αν } Z_1 < Z_2 \end{cases}$$

όπου Z_1, Z_2 είναι τα βάθη ενός ζεύγους σημείων. Προφανώς, το διατακτικό βάθος δεν αποτελεί την μόνη εναλλακτική πρόταση αντί του μετρικού βάθους. Αντίθετα, μπορούν να οριστούν διάφορες συναρτήσεις του βάθους ενός ή περισσότερων σημείων, άλλες λιγότερο και άλλες περισσότερο πληροφοριακές ως προς το βάθος. Για παράδειγμα, η συνάρτηση

$$h(Z_1, Z_2) = \frac{Z_1}{Z_2}$$

ορίζει το σχετικό βάθος δύο σημείων μιας σκηνής. Η h παρέχει περισσότερες πληροφορίες από την g ως προς το βάθος, μια και ισχύει η ακόλουθη ισοδυναμία:

$$h(Z_1, Z_2) \geq 1 \Leftrightarrow g(Z_1, Z_2) = 1$$

Η έρευνα στον τομέα της μηχανικής όρασης στοχεύει - στην συντριπτική πλειοψηφία της - στην ανακατασκευή λεπτομερών μετρικών αναπαραστάσεων του κόσμου, ικανών να υποστηρίξουν κάθε λειτουργία που χρειάζεται πληροφορίες βασισμένες σε οπτικά δεδομένα. Η δική μας προσέγγιση διαφέρει αρκετά από τις προηγούμενες κατευθύνσεις, μια και στοχεύει στον υπολογισμό ποιοτικών πληροφοριών, κατάλληλων για την κατανόηση του χώρου και βρίσκεται πιο κοντά στο πνεύμα της τελεολογικής και ποιοτικής ενεργούς όρασης, η οποία παρουσιάζεται με λεπτομέρεια και συγκρίνεται με την ανακατασκευαστική προσέγγιση στο κεφάλαιο 2. Συνοπτικά, μπορούμε να αναφέρουμε ότι η προσέγγιση αυτή στοχεύει στον υπολογισμό ποσοτήτων ικανών να υποστηρίξουν συγκεκριμένους στόχους που έχει ένα σύστημα, ασκώντας κατάλληλο έλεγχο στις παραμέτρους που καθορίζουν την διαδικασία συλλογής των εικόνων έτσι ώστε οι εικόνες που καταγράφονται να διευκολύνουν τους υπολογισμούς που πρέπει να εκτελεστούν.

Η μέθοδος υπολογισμού διατακτικού βάθους που αναπτύχθηκε, χρησιμοποιεί σαν είσοδο ένα fixating ζεύγος εικόνων, ένα ζεύγος δηλαδή το οποίο απεικονίζει μια σκηνή από ελαφρώς διαφορετικά σημεία, με τέτοιο τρόπο ώστε το κέντρο και των δύο εικόνων να απεικονίζει

1.2 Οργάνωση της εργασίας

το ίδιο σημείο της σκηνής. Κάθε ζεύγος εικόνων που καταγράφεται από τις δύο κάμερες ενός στερεοσκοπικού συστήματος, μπορεί να θεωρηθεί ότι προέρχεται από την κίνηση της μιας κάμερας στην θέση της άλλης. Αυτή η υποθετική κίνηση και η προσέγγιση της διαδικασίας σχηματισμού των εικόνων με το μοντέλο της ορθογραφικής προβολής, οδηγεί σε απλές εξισώσεις κίνησης για τα σημεία των εικόνων, οι οποίες επιτρέπουν την διάταξη δύο σημείων με βάση το βάθος. Μια πιθανή χρήση της αναπαράστασης που προκύπτει, αφορά την ποιοτική περιγραφή των σχήματος των επιφανειών που περιέχονται σε μια σκηνή, με βάση το πρόσημο της Gaussian καμπυλότητάς τους.

1.2 Οργάνωση της εργασίας

Το κεφάλαιο 2 παρουσιάζει τις βασικές αρχές της θεωρίας της ενεργούς όρασης και τις αντιδιαστέλλει με αυτές της κλασικής, ανακατασκευαστικής προσέγγισης για την όραση. Στην συνέχεια περιγράφει τα υπολογιστικά πλεονεκτήματα της ενεργούς όρασης και καταλήγει με μια αναφορά στα ερευνητικά θέματα που παρουσιάζουν ενδιαφέρον στα πλαίσια της προσέγγισης αυτής.

Το κεφάλαιο 3 ασχολείται με την οπτική πλοήγηση περιγράφοντας κάποιες από τις απαιτήσεις που θέτει σε ένα σύστημα. Στην συνέχεια, παρουσιάζεται η παραδοσιακή αντιμετώπιση για την επίλυση του προβλήματος της οπτικής πλοήγησης και το κεφάλαιο κλείνει με την παρουσίαση μιας εναλλακτικής προσέγγισης στο πρόβλημα.

Το κεφάλαιο 4 εισάγει την έννοια της κάθετης ροής, της μόνης πληροφορίας σχετικά με την κίνηση η οποία μπορεί να υπολογιστεί από τοπικές μετρήσεις, και παρουσιάζει τις εξισώσεις που την περιγράφουν καθώς και το σύστημα συντεταγμένων που θα χρησιμοποιηθεί στη συνέχεια.

Το κεφάλαιο 5 αναπτύσσει την μέθοδο υπολογισμού του διατακτικού βάθους, με βάση ένα fixating ζεύγος εικόνων. Η προτεινόμενη μέθοδος, εκμεταλλεύεται γεωμετρικούς περιορισμούς που ισχύουν στο περιβάλλον, κάνει χρήση ποιοτικών υπολογισμών, δεν εισάγει απλούστευτικές υποθέσεις για την μορφή του περιβάλλοντος και δεν βασίζεται στην λύση του προβλήματος της αντιστοίχισης (correspondence). Το κεφάλαιο καταλήγει με την περιγραφή μιας υλοποίησης της μεθόδου.

Το κεφάλαιο 6 περιγράφει μια μέθοδο που επιτρέπει τον ποιοτικό χαρακτηρισμό του σχήματος των επιφανειών που απεικονίζονται σε ένα ζεύγος εικόνων, βασισμένη στην πληροφορία για το βάθος που υπολογίζεται στο κεφ. 5 και στην θεωρία επιφανειών του Gauss. Η περιγραφή αυτή βασίζεται σε εγγενή χαρακτηριστικά των επιφανειών, και κατά συνέπεια είναι ανεξάρτητη από την επιλογή του συστήματος αναφοράς.

1.2 Οργάνωση της εργασίας

Το κεφάλαιο 7 παρουσιάζει μερικά συμπεράσματα και προτάσεις για μελλοντική ερευνητική προσπάθεια.

Στα παραρτήματα Α και Β αναπτύσσονται κάποιες σχέσεις που χρησιμοποιούνται στα κεφάλαια 5, 6.

Τέλος, το παράρτημα C παρουσιάζει μερικές περιπτώσεις στις οποίες είναι εφικτός ο υπολογισμός διατακτικού βάθους κάτω από προοπτική προβολή, στοχεύοντας να κάνει φανερό ότι στη περίπτωση αυτή, οι εξισώσεις κίνησης είναι αρκετά περίπλοκες και κατά συνέπεια ο υπολογισμός αυτός είναι δυνατός μόνο σε ειδικές περιπτώσεις.

Κεφάλαιο 2

Τελεολογική και ποιοτική ενεργός οραση

2.1 Εισαγωγή

Ένας από τους πιο φιλόδοξους στόχους της τεχνητής νοημοσύνης, είναι η ανάπτυξη αυτόνομων μηχανών με ικανότητες συγκρίσιμες με αυτές των εξελιγμένων βιολογικών οργανισμών. Δεδομένου ότι η αίσθηση της ορασης είναι πολύ σημαντική για τους τελευταίους, είναι εύκολο να ερμηνευθεί η μεγάλη ερευνητική προσπάθεια στον τομέα της μηχανικής ορασης σαν κλάδο της τεχνητής νοημοσύνης τα τελευταία 40 χρόνια. Παρά τις σημαντικές προόδους που σημειώθηκαν όμως, δεν έχει ακόμα καταστεί δυνατή η κατασκευή μηχανών με ικανότητες ορασης ανάλογες με αυτές που συναντώνται στα έμβια όντα. Στη συνέχεια, θα παρουσιάσουμε τις παραδοσιακές αρχές στις οποίες έχει κινηθεί η έρευνα στον τομέα, θα δώσουμε μια ερμηνεία για την αδυναμία εφαρμογής των αποτελεσμάτων της έρευνας αυτής σε αυτόνομα μηχανικά συστήματα και τέλος θα παρουσιάσουμε μια νέα μεθοδολογική προσέγγιση στο πρόβλημα της ορασης.

2.2 Οι δεσπόζουσες θεωρίες για την οραση

Μια από τις πιο σημαντικές παρατηρήσεις στον τομέα της υπολογιστικής ορασης, ήταν η αναγωγή του προβλήματος της ορασης σε ένα πρόβλημα υπολογισμού συμβολικών πληροφοριών με βάση μια ακολουθία εικόνων [34]. Τα μάτια εστιάζουν εικόνες του κόσμου πάνω στον αμφιβληστροειδή, ο οποίος μέσω του οπτικού νεύρου μεταδίδει τα ερεθίσματα στον εγκέφαλο, όπου και υφίστανται επεξεργασία από εξειδικευμένους νευρώνες. Τα αποτελέσματα της παραπάνω επεξεργασίας, συνιστούν τις συμβολικές περιγραφές των αντικειμένων

2.2 Οι δεσπόζουσες θεωρίες για την όραση

και των ιδιοτήτων τους (χρώμα, σχήμα, υφή κ.λπ.), οι οποίες χρησιμοποιούνται από τον εγκέφαλο για να καθορίσει τις αλληλεπιδράσεις ενός οργανισμού με τον περιβάλλοντα χώρο.

Η παραπάνω περιγραφή όμως, δεν δίνει απαντήσεις σε ερωτήματα όπως ποιά είναι η πληροφορία που περιέχεται στις εικόνες του πραγματικού κόσμου και είναι απαραίτητη για την κατανόησή του, πως εξάγεται η πληροφορία αυτή και πως παριστάνεται στον εγκέφαλο έτσι ώστε να είναι εύχρηστη από διανοητικές διεργασίες ψηλότερου επιπέδου. Τέτοια ερωτήματα είναι ζωτικής σημασίας στην μελέτη της όρασης, αφού καθορίζουν τα προβλήματα που πρέπει να λυθούν ώστε να γίνει εφικτή η κατασκευή ολοκληρωμένων συστημάτων όρασης. Κατά την διάρκεια των δεκαετιών του '60 και '70, οι ερευνητές προσπάθησαν να δώσουν λύσεις στα προβλήματα αυτά, αντιμετωπίζοντας μεγάλες δυσκολίες όσον αφορά την πρακτική εφαρμογή των θεωριών τους. Κατά συνέπεια, η πιο συνηθισμένη τακτική ήταν η διατύπωση υπεραπλουστευτικών υποθέσεων, στην πράξη ανεφάρμοστων, οι οποίες οδηγούσαν στην κατασκευή συστημάτων που δεν βελτίωναν την συνολική κατανόηση του φαινομένου της όρασης.

Ο βασικός λόγος για την αποτυχία των παραπάνω προσπαθειών υπήρξε η έλλειψη μιας πλήρους θεωρίας αντίληψης. Αναγνωρίζοντας το γεγονός αυτό, ο Marr [55] αποπειράθηκε να αναπτύξει μια θεωρία που θα ερμήνευε την λειτουργία της όρασης, με απότερο στόχο την εκμετάλλευση των συμπερασμάτων της για την κατασκευή συστημάτων μηχανικής όρασης. Συγκεκριμένα, πρότεινε ότι οι περισσότερες λειτουργίες που χρησιμοποιούν την όραση, βασίζονται στην επίλυση του ακόλουθου προβλήματος: Με βάση μια ακολουθία εικόνων μιας σκηνής, υπολογίζεται μια ακριβής, τρισδιάστατη γεωμετρική αναπαράσταση της σκηνής και μια ποσοτική περιγραφή των χαρακτηριστικών των αντικειμένων που περιέχονται σ' αυτήν. Οπτικές λειτουργίες όπως είναι η αποφυγή εμποδίων, η πλοήγηση σε πολύπλοκα περιβάλλοντα, η αναγνώριση, το πιάσιμο και η μετακίνηση αντικειμένων κ.ά., εκτελούνται από αλγόριθμους οι οποίοι εκτελούν συμβολικούς υπολογισμούς στα δεδομένα που παρέχει μια τέτοια αναπαράσταση. Με άλλα λόγια, η θεωρία του Marr υπαγορεύει την ανακατασκευή μιας πλήρους, ποσοτικής περιγραφής του κόσμου, ικανής να υποστηρίξει κάθε είδους λειτουργία που απαιτεί οπτικές πληροφορίες. Ο Marr θεώρησε ότι ο τελικός στόχος της μηχανικής όρασης είναι η απεικόνιση του πραγματικού κόσμου στην παραπάνω αναπαράσταση. Κατά συνέπεια, η όραση είναι υπεύθυνη για την κάλυψη της πρώτης συνιστώσας του κύκλου αντίληψη - σκέψη - δράση ενός έλλογου συστήματος, ενώ άλλοι κλάδοι της τεχνητής νοημοσύνης (μάθηση, προγραμματισμός κίνησης, εξαγωγή και αιτιολόγηση συμπερασμάτων κ.ά.) θα ασχοληθούν με τις υπόλοιπες δύο. Η θεωρία αυτή, αντανακλά τις επικρατούσες θεωρίες των νευρολόγων της εποχής, οι οποίες υποστήριζαν ότι οι περιοχές του εγκεφάλου που είναι επιφορτισμένες να υποστηρίζουν την αίσθηση της όρασης, αναλύουν όλη την διαθέσιμη οπτική πληροφορία. Η επεξεργασία από περιοχή σε περιοχή διαφέρει όσον αφορά την πολυπλοκότητα και το επίπεδο αφαίρεσής της.

Ο Marr επεδίωξε να συστηματικοποιήσει την ανάλυση των διαδικασιών αντίληψης,

2.2 Οι δεσπόζουσες θεωρίες για την όραση

βασισμένος στις αρχές της θεωρίας πληροφοριών (information theory). Πρότεινε την εστίαση της προσοχής σε τρία διαφορετικά επίπεδα:

- Το επίπεδο της υπολογιστικής θεωρίας. Με βάση λεπτομερειακή μαθηματική ανάλυση, προκύπτει η σχέση μεταξύ της προς υπολογισμό ποσότητας και των δεδομένων των αισθητήρων. Η ανάλυση αυτή θα καθορίσει κατά πόσον το πρόβλημα έχει μοναδική λύση.
- Το επίπεδο των αλγορίθμων και των δομών δεδομένων. Με βάση την υπολογιστική θεωρία, σχεδιάζονται αλγόριθμοι και κατάληξες δομές δεδομένων για τον υπολογισμό των επιθυμητών ποσοτήτων. Στο επίπεδο αυτό αντιμετωπίζονται θέματα σχετικά με την αποδοτικότητα και την ευρωστία (robustness) των αλγορίθμων αυτών.
- Το επίπεδο της υλοποίησης. Οι αλγόριθμοι που αναπτύχθηκαν στο προηγούμενο επίπεδο, υλοποιούνται σε κάποια σειριακή ή παράλληλη μηχανή.

Η επίδραση της θεωρίας του Marr στον τομέα της όρασης αλλά και της τεχνητής νοημοσύνης γενικότερα, υπήρξε καταλυτική. Ακολουθώντας παγιωμένες τεχνικές σχεδιασμού πολύπλοκων συστημάτων, οι ερευνητές αντιμετώπισαν το πρόβλημα της ανακατασκευής του περιβάλλοντος διακρίνοντας υποσυστήματα τα οποία ήταν αυτόνομα και λειτουργικώς ανεξάρτητα μεταξύ τους, και επιδόθηκαν στην μελέτη κάθε υποσυστήματος χωριστά. Όταν η λειτουργία κάθε υποσυστήματος θα είχε μελετηθεί επαρκώς, θα ήταν δυνατή η υλοποίησή του σε ένα υπολογιστικό σύστημα, η βελτιστοποίηση της απόδοσής του, και τέλος ο καθορισμός πρωτόκολλων επικοινωνίας μεταξύ των υποσυστημάτων, με τελικό στόχο την ολοκλήρωση τους σε ένα πρακτικό σύστημα.

Έτσι, η έρευνα στους διάφορους κλάδους της τεχνητής νοημοσύνης προχώρησε χωριστά, με κάθε κλάδο να μην λαμβάνει υπόψη τις αδυναμίες των υπόλοιπων, αλλά να περιμένει από αυτούς αποτελέσματα που απείχαν πολύ από ότι είχε ήδη επιτευχθεί. Οι ασχολούμενοι με τον προγραμματισμό κίνησης ανέμεναν ότι η έρευνα σχετικά με τους μηχανισμούς αντίληψης θα τους παρείχε ακριβείς περιγραφές του περιβάλλοντος, οι ασχολούμενοι με την όραση ανέμεναν ότι η έρευνα στην μηχανική μάθηση θα τους παρείχε ευέλικτες μεθόδους για την εξερεύνηση του περιβάλλοντος, κ.ο.κ. Μέχρι σήμερα, τέτοιες προσδοκίες δεν έχουν γίνει πραγματικότητα, με αποτέλεσμα την αδυναμία κατανόησης των συστατικών που συνθέτουν το φαινόμενο της νοημοσύνης. Η συντριπτική πλειοψηφία των μελετών που έχουν εμφανιστεί τα τελευταία 20 χρόνια είναι διαποτισμένες από το πνεύμα της θεωρίας του Marr, και παρά τα κομψά θεωρητικά τους αποτελέσματα, αδυνατούν να δώσουν επαρκείς ικανότητες όρασης σε αυτόνομα μηχανικά συστήματα.

Η αδυναμία αυτή μπορεί να αποδοθεί σε διάφορους λόγους: Κατ' αρχήν, η διαδικασία υπολογισμού όλων των χαρακτηριστικών του κόσμου από μια ακολουθία εικόνων έχει τεράστιες υπολογιστικές απαιτήσεις. Για παράδειγμα, περισσότεροι από τους μισούς νευρώνες του

2.3 Μια νέα προσέγγιση

ανθρώπινου εγκεφάλου είναι αφιερωμένοι στην επεξεργασία οπτικών πληροφοριών. Ακόμα και στην περίπτωση όμως που μια πλήρης αναπαράσταση ήταν διαθέσιμη, η υπολογιστική πολυπλοκότητα που συνεπάγεται η εκμετάλλευσή της θα ήταν απαγορευτική. Ένας άλλος λόγος είναι η έλλειψη μιας ολοκληρωμένης θεωρίας αντίληψης. Παρά το γεγονός ότι η θεωρία του Marr έθιξε κάποιες πτυχές της αντίληψης (ερεθισμοί, αισθητήρες και επεξεργασία των δεδομένων των αισθητήρων), αγνόησε άλλες όπως το περιβάλλον του οργανισμού, την φυσιολογία των μηχανικών του μερών και τους στόχους τους οποίους επιδιώκει. Ένας τρίτος λόγος, είναι το γεγονός ότι οι απότελεσματικοί στόχοι της ερευνητικής προσπάθειας ήταν πολύ φιλόδοξοι, στοχεύοντας στην κατασκευή συστημάτων με πολύ εξελιγμένες ικανότητες. Επιπλέον, η μοντελοποίηση κάθε διαδικασίας αντίληψης σαν ένα κλειστό σύστημα, δεν ανταποκρίνεται στην πραγματικότητα. Το περιβάλλον είναι αδύνατο να περιγραφεί και να καθοριστεί πλήρως, γεγονός που αναιρεί την υπόθεση του κλειστού συστήματος. Τέλος, πολλές από τις διαδικασίες ανακατασκευής του περιβάλλοντος είναι πολύ ευαίσθητες όσον αφορά την αριθμητική ακρίβεια των δεδομένων. Κατά συνέπεια, είναι ασταθείς και άρα μη εφαρμόσιμες στην περίπτωση ύπαρξης θορύβου, αβεβαιότητας και ελλιπών δεδομένων.

2.3 Μια νέα προσέγγιση

Επισημαίνοντας τις αδυναμίες της παραδοσιακής μεθοδολογίας για την αντιμετώπιση του προβλήματος της όρασης και την έλλειψη σημαντικών αποτελεσμάτων για την κατανόησή της, αρκετοί ερευνητές άρχισαν από τα μέσα της δεκαετίας του '80 να αναθεωρούν την προσέγγιση του Marr και να αναζητούν εναλλακτικές θεωρίες. Την ίδια περίπου εποχή, ερευνητές από τους τομείς της ψυχοφυσικής και νευροφυσιολογίας [3] κατέληξαν στο συμπέρασμα ότι η όραση όπως και κάθε άλλη αίσθηση, είναι προσαρμοσμένη στις ιδιαιτερότητες του περιβάλλοντος, στους στόχους και στην φυσιολογία του οργανισμού που ζει σε αυτό. Κατά συνέπεια, η όραση μπορεί να μελετηθεί ευκολότερα στα πλαίσια των συμπεριφορών που επιδεικνύει ένας οργανισμός.

Με βάση την παραπάνω διαπίστωση, προέκυψε μια νέα θεωρία για την όραση, η οποία ονομάστηκε ενεργός όραση (*active vision*) [1, 5, 2]. Για την ίδια θεωρία έχουν χρησιμοποιηθεί και τα ονόματα *animate vision* και *utilitarian vision*. Σύμφωνα με αυτήν, ένα σύστημα όρασης θα πρέπει να ελέγχει συνεχώς τις παραμέτρους της διαδικασίας απόκτησης των εικόνων, έτσι ώστε να είναι σε θέση να καταγράφει εικόνες που διευκολύνουν τους διάφορους υπολογισμούς που πρέπει να φέρει σε πέρας. Μια δεύτερη βασική αρχή της θεωρίας αφορά την διάκριση των υποσυστημάτων ενός συστήματος όρασης με βάση τις συμπεριφορές που το σύστημα καλείται να επιδείξει και όχι με βάση την λειτουργική τους σημασία (π.χ. *structure from motion*, *structure from shading*, *structure from texture*, *planning*, *learning* κ.λπ.). Οι συμπεριφορές αυτές υλοποιούνται από ένα σύνολο διεργασιών (*processes*), οι οποίες συνεργάζονται για την επίτευξη των στόχων του συστήματος. Έτσι ο στόχος της όρασης είναι να μετασχηματίσει τα

2.3 Μια νέα προσέγγιση

οπτικά ερεθίσματα σε σειρές από δράσεις που αλληλεπιδρούν με τον κόσμο (συμπεριφορές). Αυτό έρχεται σε πλήρη αντίθεση με την θεωρία του Marr, σύμφωνα με την οποία ο στόχος της όρασης είναι να μετασχηματίσει τα οπτικά ερεθίσματα σε μια γενικού σκοπού, λεπτομερή αναπαράσταση του κόσμου.

Το βασικό ερώτημα στο οποίο θα πρέπει να δοθεί απάντηση πριν οποιαδήποτε μελέτη της όρασης είναι σχετικό με τους στόχους που εξυπηρετεί η τελευταία. Οι έμβιοι οργανισμοί π.χ. χρησιμοποιούν την όραση για να βρίσκουν τροφή, να αποφεύγουν τους εχθρούς τους, να μετακινούνται στο περιβάλλον τους κ.λπ. Από την άλλη πλευρά, ένα μηχανικό σύστημα με κάποια πρακτική αξία, μπορεί να απαιτεί κάποιες από τις παραπάνω ικανότητες αλλά να στερείται πλήρως μερικών άλλων. Αυτό γιατί κάθε σύστημα με την αίσθηση της όρασης (έμβιο ή όχι), έχει μια ειδική σχέση με το περιβάλλον του και η νοημοσύνη του αποτελεί μια ενσάρκωση της σχέσης αυτής. Δεν νοείται αυθύπαρκτη νοημοσύνη, νοημοσύνη δηλαδή που δεν έχει στόχους για να επιδιώξει και δεν διαθέτει κάποιο σώμα για να αλληλεπιδρά με το περιβάλλον. Κάθε σύστημα διαθέτει ένα σύνολο από οπτικές δεξιότητες (visual capabilities), κάθε μια από τις οποίες υλοποιείται από μια εξειδικευμένη διεργασία. Η άποψη αυτή είναι σύμφωνη και με την εξελικτική θεωρία των ειδών: Τα είδη που σε κάποιο στάδιο της εξελικτικής τους πορείας απέκτησαν αισθητήρες ευαίσθητους στο φως, πιθανότατα δεν απέκτησαν ταυτόχρονα όλες τις οπτικές τους ικανότητες, αλλά σε διαφορετικές χρονικές περιόδους βάσει των αναγκών τους. Έτσι δεν αποτελεί έκπληξη η ύπαρξη πολλών διαφορετικών διεργασιών, οι οποίες έχουν διαφορετικά καθήκοντα και πιθανόν υλοποιούνται από διαφορετικές ομάδες νευρώνων. Η έρευνα στον κλάδο της νευροβιολογίας έχει αποδείξει πειραματικά την ύπαρξη εξειδικευμένων διεργασιών στα βιολογικά συστήματα όρασης [85]. Χαρακτηριστικά του ορατού κόσμου όπως η κίνηση, το χρώμα το σχήμα κ.λπ., υπολογίζονται από χωριστές ομάδες νευρώνων στον εγκέφαλο. Αυτό ανακλάται στις διάφορες ποιθήσεις της όρασης, (π.χ. αχρωματοψία), όπου η έλλειψη μιας ή περισσότερων εξειδικευμένων διεργασιών δεν στερεί πλήρως την όραση του ασθενή.

Η τελεολογία των διεργασιών της όρασης επιτρέπει την διατύπωση και επίλυση απλούστερων άρα και ευκολότερων προβλημάτων. Η αποφυγή εμποδίων (obstacle avoidance) για παράδειγμα, είναι δυνατόν να επιτευχθεί όταν υπάρχει διαθέσιμη μια ακριβής, γεωμετρική περιγραφή του περιβάλλοντος. Μέχρι σήμερα όμως, η κατασκευή μιας τέτοιας περιγραφής για περιβάλλοντα γενικής μορφής αποδείχτηκε πολύ δύσκολη, με άμεσο αποτέλεσμα η αποφυγή εμποδίων στα πλαίσια της ανοκατασκευαστικής θεωρίας να παραμένει ένα δύσκολο πρόβλημα στην γενικότητά του. Υπάρχει όμως δυνατότητα παράκαμψης της κατασκευής μιας γενικής περιγραφής του περιβάλλοντος. Εναλλακτικά, μπορεί να τεθούν πολύ συγκεκριμένα προβλήματα σχετικά με το αν π.χ. ένα κινούμενο εμπόδιο πλησιάζει τον παρατηρητή, αν βρίσκεται σε πορεία σύγκρουσης μαζί του, πόσος χρόνος αναφορικά με την ταχύτητα αντίδρασης του παρατηρητή απομένει μέχρι την σύγκρουση, κ.ά. Τα ερωτήματα αυτά δέχονται ένα σχετικά μικρό αριθμό από απαντήσεις, και έτσι επιδέχονται ποιοτικές και όχι ποσοτικές

2.3 Μια νέα προσέγγιση

απαντήσεις. Αν επομένως μπορούν να βρεθούν όμεσες λύσεις σε τέτοια προβλήματα, τότε η γενικού σκοπού περιγραφή του περιβάλλοντος καθίσταται περιττή. Η φυσιολογία ενός συστήματος και το περιβάλλον του, θέτουν περιορισμούς που κάνουν εφικτή την διατύπωση λύσεων σε προβλήματα τα οποία είναι πολύ δύσκολα στην γενικότητά τους. Ο στόχος των αλγορίθμων τους οποίους εφαρμόζουν οι παραπάνω διεργασίες είναι να παρέχουν λύση σε πολύ ειδικά προβλήματα κάτω από γενικές (χαλαρές) υποθέσεις, σε αντιδιαστολή με τους αλγόριθμους που συμφωνούν με την ανακατασκευαστική θεωρία του Marr, οι οποίοι τυπικά στοχεύουν σε πολύ γενικές λύσεις κάνοντας χρήση πολύ περιοριστικών υποθέσεων.

Δεδομένου ότι ένα σύστημα διαθέτει περιορισμένους υπολογιστικούς πόρους, θα πρέπει να κάνει κάποιο συμβιβασμό μεταξύ του όγκου των υπολογισμών που απαιτεί κάποια επεξεργασία και του κέρδους που θα αποκομίσει από αυτήν. Αυτό επιτυγχάνεται με το να είναι όλοι οι υπολογισμοί σκόπιμοι (purposive). Κατά συνέπεια, είναι δυνατή η επιλεκτική επεξεργασία των οπτικών ερεθισμάτων τόσο ως προς τον χώρο και τον χρόνο, όσο και ως προς την λεπτομέρεια. Η επιλεκτική επεξεργασία γίνεται εφικτή μέσω του ελέγχου ενός συνόλου από παραμέτρους που καθορίζουν την διαδικασία συλλογής των οπτικών ερεθισμάτων και την επεξεργασία που αυτά υφίστανται. Οι παράμετροι αυτές διακρίνονται σε τέσσερεις κατηγορίες [66]:

1. Οπτικές παράμετροι

Σ' αυτές περιλαμβάνονται η εστιακή απόσταση, οι παράμετροι του φακού και το μέγεθος του διαφράγματος της κάμερας. Η μεταβολή των παραμέτρων αυτών καθορίζει τον τρόπο σχηματισμού της προβολής μιας απεικονιζόμενης σκηνής πάνω στην φωτοευαίσθητη επιφάνεια. Η εστιακή απόσταση π.χ. καθορίζει το εύρος του οπτικού πεδίου, οι παράμετροι του φακού καθορίζουν την εστίαση, ενώ το μέγεθος του διαφράγματος καθορίζει την ποσότητα του φωτός που εισέρχεται στην κάμερα.

2. Αισθητηριακές παράμετροι

Αυτές καθορίζουν τον τρόπο με τον οποίο δειγματοληπτείται η φωτοευαίσθητη επιφάνεια, έτσι ώστε να σχηματιστεί η αντίστοιχη ψηφιακή εικόνα. Σ' αυτές περιλαμβάνεται το σχήμα της φωτοευαίσθητης επιφάνειας, η κατανομή και η ευαισθησία των φωτοευαίσθητων στοιχείων πάνω σ' αυτήν, κ.λπ.

3. Μηχανικές παράμετροι

Αυτές καθορίζουν την θέση και την ταχύτητα της κάμερας στον τρισδιάστατο κόσμο. Ο έλεγχός τους επιτυγχάνει την μετακίνηση της κάμερας, κάτι που καθίσταται απαραίτητο στις περιπτώσεις που η κάμερα δεν απεικονίζει όλα τα ενδιαφέροντα σημεία του περιβάλλοντος ή δεν έχει ομοιόμορφη διακριτική ικανότητα (βλ. αισθητηριακές παραμέτρους). Με την μεταβολή τους, είναι δυνατή η σταθεροποίηση (stabilization) της εικόνας ενός κινούμενου αντικειμένου και παρακολούθηση του (tracking). Θεωρητικά, η δυνατότητα κίνησης της κάμερας επιτρέπει την καταγραφή όλης της πληροφορίας

2.4 Υπολογιστικά πλεονεκτήματα της μεθοδολογίας της ενεργούς όρασης

που είναι διαθέσιμη στο περιβάλλον του συστήματος.

4. Αλγορίθμικές παράμετροι

Όπως έχει ήδη αναφερθεί, οι συμπεριφορές του συστήματος καθορίζονται από ένα σύνολο διεργασιών, οι οποίες εκτελούνται από κάποιο υπολογιστικό σύστημα. Ο μεγάλος όγκος των υπολογισμών που απαιτούν οι διεργασίες αυτές, επιβάλλει την παράλληλη υλοποίησή τους, έτσι ώστε να επιτευχθεί η επιθυμητή απόδοση. Το είδος του παραλληλισμού όμως διαφέρει από διεργασία σε διεργασία, γεγονός που υπαγορεύει την ύπαρξη πολλών μοντέλων παράλληλου υπολογισμού και την υποστήριξή τους από κάποιο λειτουργικό σύστημα [56]. Εκτός από τον παραλληλισμό που μπορούμε να εκμεταλλευτούμε σε κάθε διεργασία (intratask parallelism), παραλληλισμός υπάρχει και μεταξύ των διεργασιών (intertask parallelism), μια και αυτές εκτελούνται ταυτόχρονα, ανταγωνίζονται για την χρήση των πόρων του και παράγουν αποτελέσματα - πιθανώς αλληλοσυγκρουόμενα - με ρυθμούς που πρέπει να είναι συγκρίσιμοι με τους ρυθμούς μεταβολής του περιβάλλοντος. Απαιτείται λοιπόν κάποιας μορφής συντονισμός μεταξύ τους, ο οποίος θα εξασφαλίζει την αρμονική συνεργασία τους σε πραγματικό χρόνο (real time).

2.4 Υπολογιστικά πλεονεκτήματα της μεθοδολογίας της ενεργούς όρασης

Οι αρχές της ενεργούς όρασης, όπως παρουσιάστηκαν στις προηγούμενες παραγράφους, επιτρέπουν την ανάπτυξη αλγορίθμων με αρκετά πλεονεκτήματα όσον αφορά το υπολογιστικό κόστος και την πολυπλοκότητά τους [6]:

- Η φυσική αναζήτηση είναι εφικτή.
Σε αντιδιαστολή με την αλγορίθμική αναζήτηση σε μια εικόνα, η επιλογή κατάλληλων παραμέτρων μπορεί να επιτρέψει την απόκτηση εικόνων που διευκολύνουν την περαιτέρω επεξεργασία. Όταν π.χ. τα αντικείμενα του ενδιαφέροντος βρίσκονται εκτός του οπτικού πεδίου ή είναι μερικώς ορατά λόγω παρεμβαλλόμενων αντικειμένων, το σύστημα είναι δυνατόν να μετακινήσει τις κάμερες έτσι ώστε να αποκτήσει καλύτερες εικόνες.
- Το σύστημα μπορεί να κάνει γνωστές κινήσεις.
Μια και οι κινήσεις των μηχανικών μερών βρίσκονται κάτω από τον έλεγχο του συστήματος, αυτό μπορεί να κινηθεί με γνωστό τρόπο, και έτσι να εισάγει χρήσιμους περιορισμούς στην διαδικασία σχηματισμού των εικόνων.
- Η χρήση εξωκεντρικών συστημάτων συντεταγμένων είναι εφικτή.
Η ικανότητα εστίασης της προσοχής σε σημεία του κόσμου, επιτρέπει την χρήση

2.5 Ερευνητικά προβλήματα στα πλαίσια της θεωρίας της ενεργού όρασης

| Ανακατασκευαστική προσέγγιση | Ενεργός και τελεολογική προσέγγιση |
|---|---|
| Κύριο μέλημα: κατασκευή αναπαράστασης Χρήση όλων των υποσυστημάτων Παθητική απόκτηση εικόνων Πλήρης επεξεργασία των εικόνων Μέγιστη λεπτομέρεια επεξεργασίας Εγωκεντρικά συστήματα συντεταγμένων Απεριόριστοι πόροι | Κύριο μέλημα: καθορισμός οπτικών ικανοτήτων Χρήση των απαραίτητων υποσυστημάτων Ενεργητική απόκτηση εικόνων Μερική επεξεργασία των εικόνων Ικανή λεπτομέρεια επεξεργασίας Εξωκεντρικά συστήματα συντεταγμένων Περιορισμένοι πόροι |

Πίνακας 2.1: Σύγκριση των προσεγγίσεων στο πρόβλημα της όρασης.

συστημάτων συντεταγμένων στα οποία η αρχή των αξόνων βρίσκεται σε κάποιο σημείο ενός αντικειμένου, αντί για το οπτικό κέντρο της κάμερας το οποίο συνήθως χρησιμοποιεί ένα εγωκεντρικό σύστημα.

- Οι χρησιμοποιούμενοι αλγόριθμοι μπορεί να είναι ποιοτικοί.
Μια και οι ποσότητες που πρέπει να υπολογιστούν μπορούν να πάρουν τιμές από ένα σύνολο με μικρό πληθυκό αριθμό, οι αντίστοιχοι αλγόριθμοι μπορεί να είναι ποιοτικοί, κάτι πολύ σημαντικό στην περίπτωση θορύβου ή αβεβαιότητας.
- Οι χρησιμοποιούμενοι αλγόριθμοι μπορεί να εκμεταλλεύονται χαρακτηριστικά της φυσιολογίας και του περιβάλλοντος του συστήματος.
Εφόσον δεν απαιτείται από τους αλγόριθμους να είναι γενικού σκοπού, αυτοί μπορούν να εκμεταλλεύονται τις ιδιαιτερότητες του συστήματος και του περιβάλλοντός του. Για παράδειγμα, η μέγιστη ταχύτητα με την οποία μπορεί να μετακινηθεί ένα σύστημα εξαρτάται από την κατασκευή των μηχανικών του μερών. Κατά συνέπεια, κάποιος αλγόριθμος μπορεί χρησιμοποιεί ένα όνω φράγμα για την ταχύτητα του συστήματος, εισάγοντας έναν επιπλέον περιορισμό.

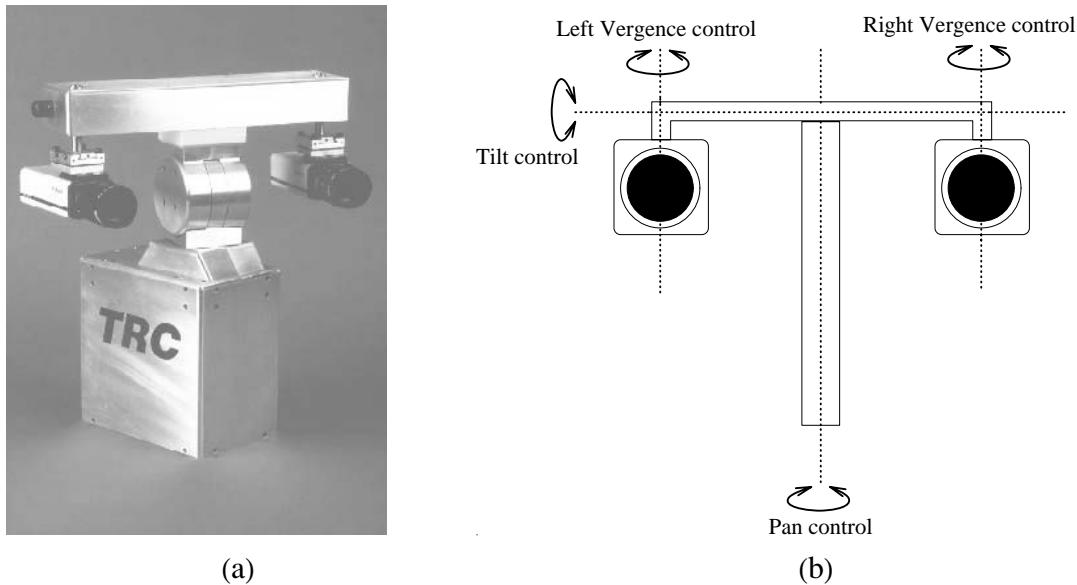
Οι βασικές διαφορές μεταξύ της κλασικής, ανακατασκευαστικής (reconstructive) θεωρίας της όρασης και αυτής της ενεργούς και τελεολογικής (purposive) όρασης συνοψίζονται στον πίνακα 2.1, ο οποίος προέρχεται από το [8].

2.5 Ερευνητικά προβλήματα στα πλαίσια της θεωρίας της ενεργού όρασης

Έχει γίνει φανερό ότι η ενεργός όραση είναι προσανατολισμένη στις συμπεριφορές που θα πρέπει να επιδεικνύει κάποιο σύστημα. Κατά συνέπεια, το πιο ουσιαστικό ερώτημα που προηγείται οποιασδήποτε μελέτης αναφορικά με ένα σύστημα όρασης, αφορά τις ικανότητες και

2.5 Ερευνητικά προβλήματα στα πλαίσια της θεωρίας της ενεργού όρασης

την ευελιξία που θα πρέπει να διαθέτει το σύστημα αυτό. Είναι λοιπόν εύκολο να ερμηνευθεί η ενασχόληση των ερευνητών με συστήματα τα οποία προσομοιάζουν τα βιολογικά όχι τόσο σε πολυπλοκότητα όσο σε επιμέρους ικανότητες. Ιδιαίτερα δημοφιλή είναι ανθρωπομορφικά συστήματα, μια και το ανθρώπινο σύστημα όρασης αποτελεί ίσως τον καλύτερο συνδυασμό ευελιξίας και αποδοτικότητας που απαντάται στην φύση. Οι τεχνολογικές εξελίξεις στους τομείς των οπτικών αισθητήρων (κάμερες CCD), των μικροεπεξεργαστών και μικροελεγκτών και των ηλεκτροκινητήρων, έχουν καταστήσει εφικτή την κατασκευή μηχανικών κεφαλών (head - eye systems), οι οποίες διαθέτουν τους περισσότερους από τους βαθμούς ελευθερίας που καθορίζουν οι παράμετροι των τριών πρώτων κατηγοριών που αναφέρθηκαν παραπάνω. Ένα μοντέλο μιας κεφαλής που είναι διαθέσιμο εμπορικά από την εταιρεία TRC φαίνεται στο σχήμα 2.1.(a). Το σχήμα 2.1.(b) απεικονίζει τους τέσσερεis βαθμούς ελευθερίας που διαθέτει η κεφαλή αυτή: Ανεξάρτητη κίνηση κάθε κάμερας δεξιά-αριστερά (vergence control), κίνηση του “λαιμού” δεξιά-αριστερά (pan control) και κίνηση και των δύο καμερών πάνω-κάτω (tilt control).



Σχήμα 2.1: Μια μηχανική κεφαλή (a) και οι βαθμοί ελευθερίας της (b).

Η ύπαρξη όμως μηχανικών κεφαλών με τις παραπάνω δυνατότητες δεν συνεπάγεται την ύπαρξη συστημάτων ικανών να βλέπουν, μια και οι διαθέσιμες γνώσεις αναφορικά με τους αλγόριθμους που πρέπει να χρησιμοποιηθούν είναι περιορισμένες. Επιπλέον, ο μεγάλος αριθμός των παραμέτρων για τον έλεγχο μιας κεφαλής, αυξάνει δραματικά τον βαθμό πολυπλοκότητας του συστήματος ελέγχου της. Στη συνέχεια παρουσιάζονται κάποια θέματα που παρουσιάζουν ερευνητικό ενδιαφέρον, μερικά από τα οποία είναι ειδικά στην μεθοδολογία της ενεργού όρασης [76].

Τα δεδομένα που παρέχουν οι οπτικοί αισθητήρες χαρακτηρίζονται από τον τεράστιο

2.5 Ερευνητικά προβλήματα στα πλαίσια της θεωρίας της ενεργού όρασης

όγκο τους και τον πλεονασμό στις πληροφορίες που περιέχουν. Ένα σύστημα με ενεργό όραση, διαθέτει μηχανισμούς οι οποίοι του επιτρέπουν να επιλέξει μεταξύ των δεδομένων που συλλέγουν οι αισθητήρες και να καθορίσει την υπολογιστική ισχύ που θα αφιερώσει στην επεξεργασία τους, με βάση τους στόχους στους οποίους έχει επικεντρωθεί. Τέτοιοι μηχανισμοί ονομάζονται μηχανισμοί εστίασης της προσοχής (*attention mechanisms*) και αποτελούν ένα πολύ σημαντικό τμήμα ενός συστήματος ενεργούς όρασης, μια και του επιτρέπουν να είναι τελεολογικό. Οι μηχανισμοί εστίασης της προσοχής στοχεύουν να επιλέξουν νέους στόχους για ένα σύστημα ενεργούς όρασης και να καθοδηγήσουν κατάλληλα το τελευταίο. Υπάρχει μια αναλογία μεταξύ των μηχανισμών εστίασης της προσοχής και του scheduler ενός λειτουργικού συστήματος: Ο scheduler είναι υπεύθυνος για την κατανομή των διαθέσιμων υπολογιστικών πόρων στις προς εκτέλεση εργασίες, ενώ ταυτόχρονα εξυπηρετεί τυχόν διακοπές (*interrupts*) που οφείλονται σε εξωτερικά γεγονότα. Ανάλογα, οι μηχανισμοί εστίασης της προσοχής αποφασίζουν για τους στόχους και τις προτεραιότητές τους και ρυθμίζουν τις παραμέτρους του συστήματος ώστε να διευκολυνθεί η επεξεργασία που υπαγορεύουν οι στόχοι αυτοί.

Τα κριτήρια για την επιλογή νέων στόχων μπορεί να εξαρτώνται από έναν άλλο στόχο ο οποίος έχει ήδη επιλεγεί, από χαρακτηριστικά (features) που προκύπτουν από απλούς υπολογισμούς (γνωστούς ως *preattentive*) ή από το γενικό πλαίσιο που καθορίζεται από τους στόχους του συστήματος, την κατάστασή του και την κατάσταση του περιβάλλοντος. Στην πρώτη περίπτωση, τα κριτήρια είναι ανεξάρτητα από τα αποτελέσματα της οπτικής επεξεργασίας. Π.χ. όταν ένα σύστημα μετακινείται ελέγχοντας για τυχόν εμπόδια, στρέφει μηχανικά την προσοχή του στην ίδια ακολουθία στόχων. Στην δεύτερη περίπτωση, η ύπαρξη κάποιων χαρακτηριστικών όπως κίνηση, χρώμα κ.λπ., που εμφανίζονται τοπικά σε μια εικόνα, προσελκύει την προσοχή του συστήματος. Με αυτόν τον τρόπο, το σύστημα μπορεί να αντιδρά σε ξαφνικά εξωτερικά γεγονότα. Στην τρίτη περίπτωση, διεργασίες ψηλού επιπέδου εισάγουν νέους στόχους. Π.χ. όταν ένα σύστημα θέλει να βγει από ένα δωμάτιο, πρέπει να ψάξει για κάποια πόρτα.

Οι φωτοευαίσθητες επιφάνειες πολλών βιολογικών συστημάτων εμφανίζουν ανομοιόμορφη κατανομή των αισθητηριακών τους στοιχείων, η οποία συνεπάγεται μεταβλητή διακριτική ικανότητα. Συγκεκριμένα, εμφανίζουν μέγιστη διακριτική ικανότητα στο κέντρο του οπτικού πεδίου (ωχρή κηλίδα - *fovea*), ενώ η διακριτική ικανότητα μειώνεται προχωρώντας προς την περιφέρεια. Η κατανομή αυτή, επιτρέπει τον συνδυασμό μιας περιοχής με μεγάλη διακριτική ικανότητα με οπτικό πεδίο μεγάλου εύρους. Αποτελεί μάλιστα μια οικονομική λύση, τόσο ως προς τον αριθμό των αισθητηριακών στοιχείων, όσο και ως προς τον αριθμό των νευρώνων που απαιτούνται για την επεξεργασία των οπτικών δεδομένων. Για παράδειγμα, έχει διατυπωθεί η άποψη ότι αν το ανθρώπινο μάτι είχε την ίδια μεγάλη διακριτική ικανότητα σε όλο το οπτικό πεδίο, τότε ο ανθρώπινος εγκέφαλος θα ζύγιζε 10.000 τόνους. Η ανομοιομορφία στην διακριτική ικανότητα, επιβάλλει στο σύστημα να είναι ενεργό, έτσι

2.5 Ερευνητικά προβλήματα στα πλαίσια της θεωρίας της ενεργού όρασης

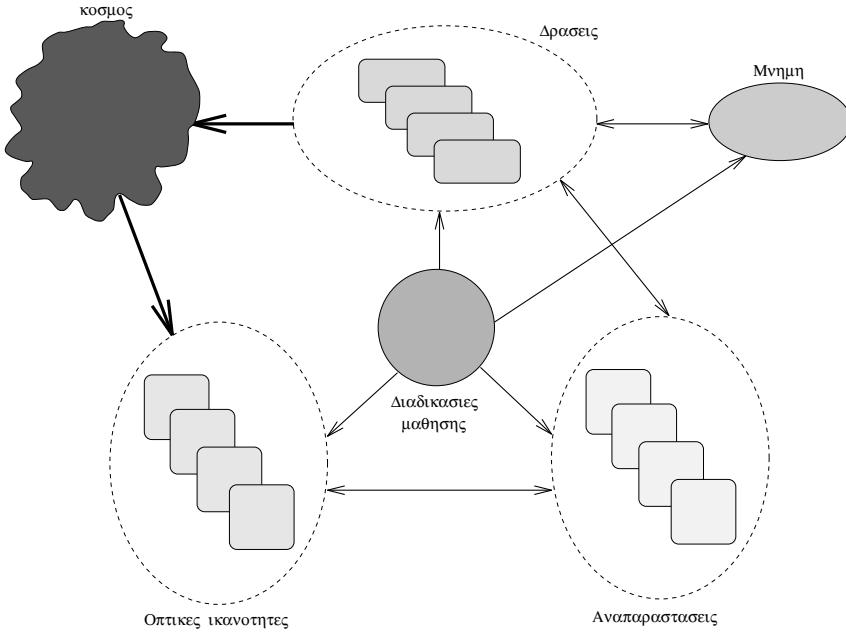
ώστε μέσω κατάλληλου ελέγχου των μηχανικών και οπτικών παραμέτρων του, να μπορεί να φέρνει τα αντικείμενα του ενδιαφέροντος στο κέντρο του οπτικού του πεδίου, ώστε αυτά να απεικονιστούν με την μέγιστη διαθέσιμη ακρίβεια. Ο καθορισμός χαρακτηριστικών τα οποία θα υπολογίζονται από τοπική επεξεργασία στην περιφέρεια του οπτικού πεδίου και θα παρέχουν χρήσιμη πληροφορία παρά την μικρή διακριτική ικανότητα, θα επιτρέψει την χρήση μηχανισμών εστίασης της προσοχής βασισμένους σε αυτά.

Ο έλεγχος των μηχανικών και οπτικών παραμέτρων του συστήματος, γνωστός σαν *gaze control*, εμφανίζεται με δύο μορφές: Την σταθεροποίηση της εικόνας κάποιου αντικειμένου (*gaze stabilization*) και την εστίαση ενός νέου (*gaze change*). Στην πρώτη, γνωστή και ως *fixation*, το σύστημα επιχειρεί να κρατήσει ένα πιθανόν κινούμενο αντικείμενο στο κέντρο του οπτικού του πεδίου. Στην δεύτερη, επιλέγεται ένα νέο αντικείμενο και το σύστημα επιχειρεί να το απεικονίσει στο κέντρο του οπτικού του πεδίου. Το πρόβλημα αυτό είναι γνωστό ως “*where to look next*”. Έχει υποστηριχθεί ότι το παραπάνω πρόβλημα μπορεί να αντιμετωπιστεί ευκολότερα μέσω μιας ακολουθίας έργων εντοπισμού (*location*) και αναγνώρισης (*identification*). Τα πρώτα συνίστανται στον εντοπισμό κάποιου αντικειμένου που βρίσκεται στην περιφέρεια του οπτικού πεδίου με βάση ένα μοντέλο του. Τα δεύτερα, συνίστανται στην αναγνώριση του αντικειμένου που βρίσκεται στο κέντρο του οπτικού πεδίου, με βάση ένα σύνολο μοντέλων γνωστών αντικειμένων.

Η όραση σαν αίσθηση είναι παθητική, με την έννοια ότι δεν μπορεί να ενεργήσει σε αντικείμενα του κόσμου, αλλάζοντας την κατάστασή τους. Για το λόγο αυτό, είναι πολύ ενδιαφέρων ο συνδυασμός ενός συστήματος όρασης με ένα μηχανικό βραχίονα. Ο τελευταίος, θα μπορεί να παρέχει επιπλέον δεδομένα μέσω αισθητήρων διαφόρων ειδών (αφής, θερμότητας, κ.ά.) και θα επιτρέπει την εκτίμηση των αποτελεσμάτων της όρασης, κλείνοντας τον κύκλο μεταξύ αντίληψης και δράσης.

Τα θέματα που παρουσιάστηκαν παραπάνω, πρέπει να μελετηθούν στα πλαίσια ενός φορμαλισμού που θα μοντελοποιεί αφ' ενός τα δομικά στοιχεία ενός συστήματος όρασης και αφ' ετέρου την συνεργατική σχέση του συστήματος με το περιβάλλον. Ένα μοντέλο σύμφωνο με τις αρχές της τελεολογικής όρασης, είναι αυτό του σχήματος 2.2 και προτείνεται από τους Fermüller και Aloimonos [30]. Τα βασικά δομικά στοιχεία του είναι οι οπτικές ικανότητες (*visual competences*), οι τελεολογικές αναπαραστάσεις (*purposive representations*), οι δράσεις (*actions*), οι διαδικασίες μάθησης (*learning procedures*) και η μνήμη (*memory*).

Σε κάθε χρονική στιγμή, το σύστημα έχει ένα σύνολο από στόχους τους οποίους καλείται να εκπληρώσει. Οι οπτικές ικανότητες επεξεργάζονται τα δεδομένα των οπτικών αισθητήρων, επιλέγοντας εκείνα τα οποία σχετίζονται άμεσα με τους στόχους του συστήματος. Οι οπτικές ικανότητες δεν στοχεύουν στην ανακατασκευή του κόσμου, αλλά στην αναγνώριση απόψεων του απαραίτητων για την εκπλήρωση κάποιου στόχου. Τα αποτελέσματα της επεξεργασίας των οπτικών αισθητήρων αποτελούν τις (τελεολογικές) αναπαραστάσεις του συστήματος. Οι



Σχήμα 2.2: Δομικά στοιχεία ενός τελεολογικού συστήματος όρασης.

τελευταίες είναι μερικές όσον αφορά τις πληροφορίες που παρέχουν και έχουν διαφορετικούς βαθμούς πολυπλοκότητας. Οι οπτικές ικανότητες και οι αναπαραστάσεις κωδικοποιούν την νοημοσύνη του συστήματος. Οι αναπαραστάσεις έχουν όμεση πρόσβαση στις δράσεις του συστήματος, οι οποίες είναι είτε εξωτερικές, επιτρέποντας την αλληλεπίδραση του συστήματος με το περιβάλλον, είτε εσωτερικές, επιτρέποντας την αλλαγή της κατάστασης του συστήματος. Οι δράσεις είναι αποθηκευμένες με συσχετιστικό (associative) τρόπο [41], προσπελαύνονται δηλαδή με βάση τα περιεχόμενα των αναπαραστάσεων. Ακολουθίες δράσεων αποτελούν τις συμπεριφορές του συστήματος. Οι αναπαραστάσεις και οι δράσεις του συστήματος καθώς και άλλες χρήσιμες για την λειτουργία του πληροφορίες είναι αποθηκευμένες στην μνήμη. Οι διαδικασίες μάθησης επιτρέπουν στο σύστημα να μαθαίνει κατάλληλες τιμές για τις παραμέτρους του, νέες δράσεις, νέες οπτικές ικανότητες για τον υπολογισμό νέων αναπαραστάσεων και κατάλληλους συνδυασμούς αναπαραστάσεων και δράσεων που διευκολύνουν την επίτευξη κάποιου στόχου. Έτσι, ένα σύστημα θα έχει την ικανότητα προσαρμογής σε νέες συνθήκες χωρίς να απαιτείται εξ' αρχής ο λεπτομερής προγραμματισμός του αναφορικά με κάθε κατάσταση στην οποία μπορεί να βρεθεί. Ειδικά σε συστήματα με επιλεκτική αντίληψη, οι αυξημένες πιθανότητες εμφάνισης των προβλήματος της ελλιπούς αντίληψης (incomplete perception) [58], προσδίδουν στην ικανότητα της μάθησης ιδιαίτερη σημασία.

Σχετικά με το είδος των στόχων που θα πρέπει να έχει ένα σύστημα όρασης, η Fermüller [29] προτείνει μια προσέγγιση που προσομοιάζει στην διαδικασία εξέλιξης των οργανισμών και την ονομάζει συνθετική (synthetic). Σύμφωνα με την συνθετική προσέγγιση, το σύστημα

2.5 Ερευνητικά προβλήματα στα πλαίσια της θεωρίας της ενεργού όρασης

ξεκινά με πολύ απλούς στόχους και προοδευτικά προστίθενται νέοι με μεγαλύτερη πολυπλοκότητα. Η συνθετική προσέγγιση μοιάζει με την ιεραρχική αρχιτεκτονική του Brooks (subsumption architecture) [13]. Η τελευταία όμως πάσχει από σημαντικά μειονεκτήματα, αφού αγνοεί τελείως την φυσιολογία του συστήματος και τους στόχους του καθώς και το θέμα της αναπαράστασης της γνώσης (knowledge representation). Επιπλέον, η αυστηρά ιεραρχική δομή θέτει προβλήματα επικοινωνίας μεταξύ των υποσυστημάτων.

Κεφάλαιο 3

Οπτική πλοήγηση

3.1 Εισαγωγή

Ο όρος “πλοήγηση” για κάποιο σύστημα, αναφέρεται στην ικανότητά του να μετακινείται αυτόνομα στο περιβάλλον του με την χρήση κάποιων αισθητήρων. Ο ορισμός αυτός είναι αρκετά γενικός ώστε να συμπεριλαμβάνει μια πληθώρα βιολογικών και μηχανικών συστημάτων. Για παράδειγμα, σχεδόν όλοι οι ζωντανοί οργανισμοί διαθέτουν ικανότητες που τους επιτρέπουν να αναζητούν την τροφή τους, να ξαναβρίσκουν την φωλιά τους όταν απομακρύνονται από αυτήν και να μετακινούνται με ασφάλεια στον χώρο στον οποίο ζουν. Οι μηχανισμοί πλοήγησης που χρησιμοποιούνται για τον σκοπό αυτό, παρουσιάζουν σημαντικές διαφορές από οργανισμό σε οργανισμό, κάνοντας χρήση διαφόρων αισθητήρων και τεχνικών [19, 72, 20, 36]. Υπάρχουν συστήματα βασισμένα σε αδρανειακούς (inertial) αισθητήρες, σε αισθητήρες υπερήχων (sonar), σε οπτικούς αισθητήρες, σε αισθητήρες που ανιχνεύουν διαταραχές σε ηλεκτρικά ή μαγνητικά πεδία, σε αισθητήρες αφής, κ.ά. Η ικανότητα των ζωντανών οργανισμών να μετακινούνται αυτόνομα μεταξύ δυο σημείων του περιβάλλοντός τους, είναι γνωστή και σαν *homming*, όρος που επινοήθηκε από τους ερευνητές της συμπεριφοράς των ζώων (ethologists), αλλά σήμερα έχει γιοθετηθεί και από τους ερευνητές στον κλάδο της ρομποτικής. Εκτός από τα βιολογικά, υπάρχουν και μερικά μηχανικά συστήματα με κάποιες ικανότητες αυτόνομης πλοήγησης. Π.χ. τα συστήματα πλοήγησης των αεροσκαφών και των πλοίων συνδυάζουν αδρανειακούς αισθητήρες με ραδιοπομπούς των οποίων η θέση είναι γνωστή εκ των προτέρων, οι στρατιωτικοί πύραυλοι διάφορων κατηγοριών κατευθύνονται προς τις πηγές εκπομπής ηλεκτρομαγνητικής ακτινοβολίας - συνήθως υπέρυθρης -, ενώ τα διαστημόπλοια που εξερευνούν το ηλιακό σύστημα βασίζονται στις θέσεις των ουρανίων σωμάτων.

Η μελέτη της ικανότητας αυτόνομης πλοήγησης παρουσιάζει σημαντικό θεωρητικό ενδιαφέρον. Η ικανότητα αυτή μπορεί να οριστεί αναφορικά με διάφορα περιβάλλοντα,

3.2 Ανασκόπηση της περιοχής

κατά συνέπεια μπορεί να θεωρηθεί σαν ένα ειδικό πρόβλημα που πρέπει να λυθεί κάτω από γενικές συνθήκες. Επιπλέον, οι μηχανισμοί αυτόνομης πλοήγησης μπορούν να οργανωθούν σε μια ιεραρχία με βάση τον βαθμό πολυπλοκότητάς τους και έτσι μπορούν να μελετηθούν με την συνθετική προσέγγιση που αναφέρθηκε στο προηγούμενο κεφάλαιο. Μερικές από τις συνιστώσες της αυτόνομης πλοήγησης ενός συστήματος είναι η ικανότητα εκτίμησης των παραμέτρων κίνησής του (egomotion estimation), η ικανότητα αποφυγής εμποδίων (obstacle avoidance) και η ικανότητα μετακίνησής του από σημείο σε σημείο (homing). Τέλος, η ευρεία διάδοση της ικανότητας αυτόνομης πλοήγησης μεταξύ των ζωντανών οργανισμών, υποδηλώνει ότι αυτή αποτελεί στοιχειώδη ικανότητα ενός νοήμονος συστήματος. Ένα τέτοιο σύστημα θα πρέπει να εξάγει και να “απομνημονεύει” κάποια χαρακτηριστικά του περιβάλλοντος και αργότερα να χρησιμοποιεί αυτή την γνώση για την αναγνώριση τμημάτων του περιβάλλοντος. Το γεγονός αυτό συνεπάγεται λύσεις σε προβλήματα αντίληψης, μάθησης, αναπαράστασης γνώσης, αιτιολόγησης και εξαγωγής συμπερασμάτων. Σημαντικό όμως είναι και το πρακτικό ενδιαφέρον του θέματος. Η αυτόνομη πλοήγηση μπορεί να αποτελέσει δομικό στοιχείο σύνθετων συμπεριφορών που έχουν σημαντικές πρακτικές εφαρμογές [59]. Μερικές από αυτές περιλαμβάνουν βιοθήματα για άτομα με ειδικές ανάγκες, συστήματα για την παρακολούθηση και ασφάλεια χώρων, συστήματα για χρήση σε επικίνδυνα για τον άνθρωπο περιβάλλοντα, συστήματα για την βιομηχανία, συστήματα για οικιακή χρήση κ.ά.

Στη συνέχεια, θα επικεντρωθούμε στην οπτική πλοήγηση, στην ικανότητα δηλαδή ενός συστήματος να κινείται αυτόνομα με βάση τις πληροφορίες που του παρέχει μια ακολουθία εικόνων. Το αυξημένο ενδιαφέρον για την οπτική πλοήγηση έναντι των άλλων μεθόδων που χρησιμοποιούνται από τεχνητά και βιολογικά συστήματα οφείλεται σε διάφορους λόγους. Όσον αφορά τα τεχνητά συστήματα, οι υπάρχουσες μέθοδοι χρησιμοποιούνται σε προβλέψιμα περιβάλλοντα, στηρίζονται στην ύπαρξη πληροφοριών που δεν συλλέγονται αυτόματα από το ίδιο το σύστημα (π.χ. χάρτες) και απαιτούν την αλλαγή του περιβάλλοντος κατά τέτοιο τρόπο ώστε να διευκολύνεται η συλλογή κάποιων άλλων πληροφοριών (π.χ. προσθήκη πομπών). Κατά συνέπεια, η χρησιμότητα των τεχνικών αυτών για άγνωστα περιβάλλοντα είναι περιορισμένη. Από την άλλη πλευρά, οι γνώσεις που διαθέτουμε για τους μηχανισμούς μη οπτικής πλοήγησης που χρησιμοποιούν τα βιολογικά συστήματα είναι πολύ λίγες. Αντίθετα, η συστηματικότερη μελέτη της όρασης, ο μεγάλος πλούτος πληροφοριών που μπορούν να εξαχθούν μέσω αυτής και η μεγάλη διάδοση των μηχανισμών οπτικής πλοήγησης στην φύση, κάνουν την χρήση οπτικών μεθόδων για την πλοήγηση αρκετά ελκυστική.

3.2 Ανασκόπηση της περιοχής

Μέχρι σήμερα, η πλειοψηφία των μελετών σχετικά με την οπτική πλοήγηση είναι επηρεασμένες από την θεωρία του Marr, η οποία παρουσιάστηκε στο προηγούμενο κεφάλαιο. Δεδομένου ότι η θεωρία αυτή διαχωρίζει τις διαδικασίες της αντίληψης από αυτές της σκέψης, οι μελέτες

3.2 Ανασκόπηση της περιοχής

αυτές ασχολούνται με την επίλυση δύο προβλημάτων: Το πρώτο είναι ο υπολογισμός μιας τρισδιάστατης, γεωμετρικής περιγραφής του χώρου μέσω μιας ακολουθίας εικόνων που καταγράφει ένας παρατηρητής που κινείται ως προς το περιβάλλον του και το δεύτερο η χρήση αυτής της περιγραφής για τον υπολογισμό μιας ακολουθίας κινήσεων που θα επιτύχει την μετακίνηση ενός μηχανικού συστήματος στον χώρο. Το πρώτο πρόβλημα, είναι γνωστό σαν *structure from motion* και το δεύτερο σαν πρόβλημα προγραμματισμού της κίνησης (*motion planning*). Θα εξετάσουμε κάθε πρόβλημα χωριστά στην συνέχεια.

3.2.1 Structure from motion

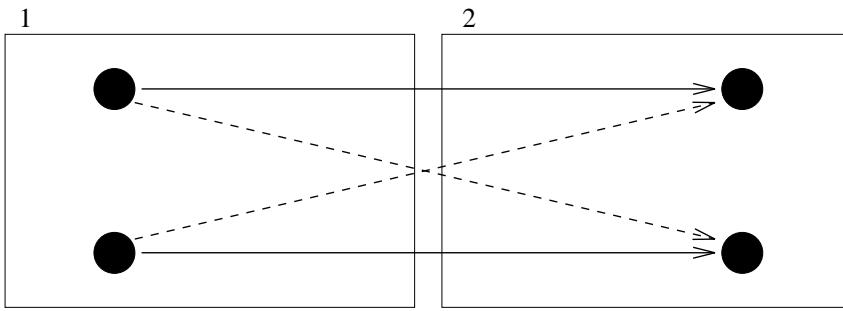
Σύμφωνα με την ανακατασκευαστική θεωρία για την όραση, η ανάλυση της κίνησης μέσω εικόνων γίνεται σε δύο στάδια.

- Αρχικά υπολογίζεται η μετατόπιση των προβολών κάποιων σημείων μεταξύ συνεχόμενων εικόνων, η οποία οφείλεται στην σχετική κίνηση του παρατηρητή ως προς αυτά. Στην περίπτωση που η κίνηση αυτή είναι μεγάλη, σύνολα σημείων ομαδοποιούνται σε χαρακτηριστικά όπως γωνίες και ευθύγραμμα τμήματα, και η μετατόπιση υπολογίζεται μέσω της αντιστοίχισης (*correspondence*) τέτοιων χαρακτηριστικών σε διαδοχικές εικόνες. Η αντιστοίχιση συνίσταται στην αναγνώριση κάποιων χαρακτηριστικών σε διαδοχικές εικόνες, τα οποία αποτελούν απεικονίσεις του ίδιου συνόλου σημείων του πραγματικού κόσμου. Στην περίπτωση μικρής κίνησης, η ακολουθία των εικόνων θεωρείται μια συνεχής συνάρτηση δύο χωρικών και μιας χρονικής μεταβλητής. Οι χωρικές και χρονικές μερικές παράγωγοι της συνάρτησης αυτής σε συνδυασμό με κάποιες επιπλέον υποθέσεις σχετικά με την κίνηση, δίνουν το διάνυσμα της ταχύτητας κάθε σημείου στο επίπεδο της εικόνας (*image plane*).
- Στην συνέχεια, υπολογίζεται η τρισδιάστατη ταχύτητα και η δομή της σκηνής που περιέχεται στην ακολουθία των εικόνων, μέσω των εξισώσεων που συσχετίζουν την διδιάστατη ταχύτητα των σημείων της εικόνας με την τρισδιάστατη κίνηση των σημείων του κόσμου που απεικονίζονται σ' αυτά.

Το πρόβλημα του *structure from motion* έχει μελετηθεί διεξοδικά σε θεωρητικό επίπεδο. Είναι γνωστά αποτελέσματα που αφορούν το είδος των πληροφορίων σχετικά με την δομή και την σχετική κίνηση των σημείων μιας σκηνής που μπορούν να εξαχθούν από μία ακολουθία εικόνων [80], τις συνθήκες που εξασφαλίζουν την μοναδικότητα των πληροφοριών αυτών και την συμπεριφορά των μεθόδων αυτών στην περίπτωση ύπαρξης θορύβου [79, 74]. Σχεδιάστηκαν επίσης επαναληπτικοί και μη επαναληπτικοί αλγόριθμοι για τον υπολογισμό της δομής και της κίνησης μιας σκηνής. Παρά τις προσπάθειες αρκετών ετών όμως, μέχρι σήμερα δεν έχει καταστεί δυνατή η κατασκευή ενός συστήματος ικανού να αντιμετωπίζει με επιτυχία το πρόβλημα του *structure from motion*.

3.2 Ανασκόπηση της περιοχής

Το πρόβλημα αυτό είναι από τη φύση του δύσκολο, μια και η μόνη πληροφορία αναφορικά με την κίνηση που μπορεί να εξαχθεί από μια ακολουθία εικόνων, βασίζεται στις αλλαγές της φωτεινότητας κάποιων σημείων, οι οποίες δεν οφείλονται αποκλειστικά στην σχετική κίνηση του παρατηρητή ως προς τα παρατηρούμενα αντικείμενα (βλ. και κεφ. 4). Εξίσου δύσκολο είναι και το πρόβλημα του υπολογισμού της μετατόπισης κάποιων σημείων που απαιτείται στο πρώτο στάδιο της ανακατασκευαστικής προσέγγισης. Ο βασικός λόγος γι' αυτό είναι ότι οι γεωμετρικοί περιορισμοί που αφορούν αφ' ενός το σύστημα απόκτησης των εικόνων και αφ' ετέρου τα αντικείμενα που απεικονίζονται μέσω αυτού, δεν είναι αρκετά ισχυροί ώστε να εξασφαλίσουν μοναδική λύση. Ένα απλό παράδειγμα για να γίνει αυτό κατανοητό, παρουσιάζεται στο σχήμα 3.1. Στο σχήμα αυτό, φαίνονται δύο διαδοχικές εικόνες που παριστάνουν ένα ζεύγος από κουκίδες, το οποίο είναι μετατοπισμένο από εικόνα σε εικόνα. Η μετατόπιση που μπορεί να αντιστοιχιστεί σε κάθε κουκίδα δεν είναι μοναδική. Οι κουκίδες μπορεί να θεωρηθεί ότι μετατοπίζονται οριζόντια (όπως υποδηλώνουν τα συμπαγή βέλη) ή διαγώνια (διακεκομμένα βέλη). Για να εξασφαλιστεί η μοναδικότητα της αντιστοίχισης, χρησιμοποιούνται επιπλέον περιορισμοί, βασισμένοι σε ειδικές υποθέσεις σχετικά με τα αντικείμενα που απεικονίζονται στις εικόνες. Κατά κανόνα, οι περιορισμοί αυτοί ισχύουν σε περιβάλλοντα με πολύ συγκεκριμένη μορφή, οδηγώντας έτσι σε πολύ εξειδικευμένες μεθόδους οι οποίες έχουν κακή απόδοση όταν εφαρμοστούν σε περιβάλλοντα διαφορετικού είδους. Τέλος, η διαδικασία υπολογισμού των παραμέτρων της κίνησης στο χώρο με βάση την μετατόπιση ή την ταχύτητα των σημείων στο επίπεδο της εικόνας, είναι πολύ ευαίσθητη στον θόρυβο και κατά συνέπεια ασταθής.



Σχήμα 3.1: Ασάφεια στην αντιστοίχιση των κουκίδων.

3.2.2 Προγραμματισμός κίνησης

Ο προγραμματισμός της κίνησης ενός μηχανικού συστήματος (ρομπότ), δεδομένης μιας τρισδιάστατης, γεωμετρικής περιγραφής του περιβάλλοντος, της αρχικής θέσης του συστήματος και της επιθυμητής τελικής του θέσης, συνίσταται στον υπολογισμό ενός ελεύθερου μονοπατιού ανάμεσα στα εμπόδια, το οποίο συνδέει την αρχική με την τελική θέση του συστήματος [39, 45]. Το πρόβλημα του υπολογισμού αυτού του μονοπατιού, είναι γνωστό

3.2 Ανασκόπηση της περιοχής

σαν “piano movers problem” ή “generalized movers problem”. Οι μέθοδοι που έχουν προταθεί, αποτελούν παραλλαγές μερικών βασικών προσεγγίσεων, οι οποίες περιγράφονται συνοπτικά στην συνέχεια.

- **Σκελετός (Skeleton).** Κατά την προσέγγιση αυτή, το σύνολο των εφικτών θέσεων του συστήματος, εκείνων δηλαδή που δεν παρεμποδίζονται από τα αντικείμενα του περιβάλλοντος, απεικονίζεται σε ένα δίκτυο. Τα μονοπάτια του γράφου αυτού αντιστοιχούν σε όλες τις εφικτές ακολουθίες κινήσεων του συστήματος. Έτσι, ο υπολογισμός μιας ακολουθίας κινήσεων ελαττώνεται σε ένα πρόβλημα αναζήτησης σε γράφους. Στην κατηγορία αυτή ανήκουν μέθοδοι που χρησιμοποιούν δομές όπως οι γράφοι ορατότητας (*visibility graphs*) και τα διαγράμματα *Voronoi* (*Voronoi diagrams*) [63, 52, 50, 64, 16].
- **Διαχωρισμός Θέσεων (Cell Decomposition).** Ο ελεύθερος χώρος που περιβάλλει τα αντικείμενα του περιβάλλοντος καλύπτεται από ένα σύνολο απλών γεωμετρικών σχημάτων, π.χ. παραλληλογράμμων, για τα μέλη του οποίου υπολογίζονται σχέσεις γειτνίασης. Ο υπολογισμός μιας ακολουθίας κινήσεων ανάγεται στην αναζήτηση μιας ακολουθίας τέτοιων σχημάτων, ανά δύο γειτονικών, η οποία θα ενώνει τα σχήματα που περιέχουν την αρχική και την τελική θέση του συστήματος. Μέθοδοι της κατηγορίας αυτής είναι αυτές που περιγράφονται στις αναφορές [12, 73, 70, 4].
Οι τεχνικές των δύο παραπάνω κατηγοριών βασίζονται σε τεχνικές δανεισμένες από την υπολογιστική γεωμετρία [68]. Ένα άλλο κοινό χαρακτηριστικό τους, είναι ότι απαιτούν κάποιας μορφής αναζήτηση στην δομή που παριστάνει τον ελεύθερο χώρο. Κλασικές τεχνικές αναζήτησης σε γράφους, όπως π.χ. depth-first, breadth-first, best-first, A^* , χρησιμοποιούνται για τον σκοπό αυτό. Συχνά επίσης εφαρμόζονται τεχνικές ελαχιστοποίησης όπως ο αλγόριθμος του Dijkstra, μέθοδοι simulated annealing, κ.ά.
- **Δυναμικό Πεδίο (Potential Field).** Στην προσέγγιση αυτή, κάθε αντικείμενο του περιβάλλοντος ασκεί απωστικές δυνάμεις στα σημεία του χώρου που το περιβάλλουν, το μέτρο των οποίων εξαρτάται από την απόσταση των σημείων από αυτό. Ο στόχος, η επιθυμητή τελική θέση του συστήματος δηλαδή, ασκεί αντίστοιχες ελκτικές δυνάμεις. Έτσι, σε κάθε σημείο του χώρου εργασίας αντιστοιχεί ένα διάνυσμα το οποίο υπολογίζεται ως η συνισταμένη των ελκτικών και απωστικών δυνάμεων στο σημείο αυτό, ορίζοντας ένα διανυσματικό πεδίο. Το μέτρο της δύναμης είναι ελάχιστο στο σημείο που αντιστοιχεί στον στόχο και αυξάνεται με την απόσταση από αυτόν. Υπολογίζοντας την κλίση (gradient) σε κάποιο σημείο του χώρου, ένα σύστημα μπορεί να επιλέξει την διεύθυνση κατά την οποία πρέπει να κινηθεί ώστε να φτάσει στον στόχο του όσο γίνεται συντομότερα [42, 77]. Το βασικότερο πρόβλημα με την προσέγγιση αυτή, είναι το γεγονός ότι η διανυσματική συνάρτηση που υπολογίζεται έχει συνήθως πολλά τοπικά ελάχιστα, τα οποία είναι πιθανό να “παγιδεύσουν” το σύστημα.

3.2 Ανασκόπηση της περιοχής

- *Μαθηματικός Προγραμματισμός (Mathematical Programming)*. Το πρόβλημα υπολογισμού μιας ακολουθίας κινήσεων, αντιμετωπίζεται σαν πρόβλημα βελτιστοποίησης κάποιας συνάρτησης, κάτω από περιορισμούς που σχετίζονται με τα αντικείμενα του περιβάλλοντος. Μια και οι περιορισμοί αυτοί καθώς και η προς βελτιστοποίηση συνάρτηση είναι μη γραμμικοί, συνήθως χρησιμοποιούνται αριθμητικές μέθοδοι για την εύρεση της βέλτιστης λύσης [27].

Βασικό πλεονέκτημα των παραπάνω προσεγγίσεων είναι το γεγονός ότι τα υπολογιζόμενα μονοπάτια είναι βέλτιστα σύμφωνα με κριτήρια όπως είναι το συνολικό τους μήκος, η απόσταση του μηχανικού συστήματος από τα εμπόδια κ.ά. Η προϋπόθεση όμως ύπαρξης μιας ακριβούς αναπαράστασης του περιβάλλοντος, απαιτεί εξελιγμένες ικανότητες αντίληψης και είναι πολύ δύσκολο αν όχι αδύνατο, να επιτευχθεί στην πράξη. Αντίθετα, η όποια γνώση για το περιβάλλον είναι ελλιπής και χαρακτηρίζεται από αβεβαιότητα λόγω ανακριβών μετρήσεων των αισθητήρων, απρόβλεπτων αλλαγών του περιβάλλοντος και ατελή ελέγχου του συστήματος. Επιπλέον, ο διαχωρισμός του υπολογισμού μιας ακολουθίας κινήσεων από την εκτέλεσή τους, καθώς και το μεγάλο υπολογιστικό τους κόστος, κάνει τις μεθόδους αυτές ανεπαρκείς για δυναμικά περιβάλλοντα. Ένα άλλο πρόβλημα έχει να κάνει με το γεγονός ότι οι μέθοδοι αυτές δεν είναι επεκτάσιμες σε πολύ μεγάλους χώρους εργασίας, μιας και οι απαιτούμενοι υπολογισμοί καθώς και η μνήμη που απαιτείται για την αποθήκευση της αναπαράστασης του κόσμου ξεπερνούν τους διαθέσιμους πόρους. Τα παραπάνω, περιορίζουν την χρησιμότητα των μεθόδων αυτών σε πολύ ειδικές περιπτώσεις, όπου το περιβάλλον είναι στατικό και ελεγχόμενο ενώ η δομή του είναι πλήρως γνωστή εκ των προτέρων¹.

Η αδυναμία κατασκευής μιας αναπαράστασης του κόσμου κατάλληλης να υποστηρίξει αλγόριθμους προγραμματισμού κίνησης όπως αυτοί που παρουσιάστηκαν παραπάνω, καθώς και η αδυναμία εφαρμογής των τελευταίων σε κάποια περιβάλλοντα, αποτέλεσαν το κίνητρο για την εμφάνιση πιο ευέλικτων μεθόδων. Βασικά τους χαρακτηριστικά, είναι η χρήση χαλαρότερων και πιο ρεαλιστικών υποθέσεων για την λεπτομέρεια αναπαράστασης του περιβάλλοντος και την ακρίβεια των χρησιμοποιούμενων αισθητήρων, η προσπάθεια εξαγωγής κάποιας αναπαράστασης του περιβάλλοντος σύμφωνης με τις υποθέσεις αυτές και τέλος η απόπειρα συνδυασμού του προγραμματισμού της κίνησης με την ανάδραση (feedback) που παρέχουν οι αισθητήρες. Οι περισσότερες από αυτές χρησιμοποιούν “χάρτες” του περιβάλλοντος οι οποίοι ενημερώνονται από τις μετρήσεις των αισθητήρων. Η οργάνωση αυτών των χαρτών δανείζεται στοιχεία από τις γεωμετρικές μεθόδους που παρουσιάστηκαν παραπάνω. Διάφορες τεχνικές αναζήτησης, συχνά ευριστικές [17], χρησιμοποιούν τους χάρτες αυτούς για να υπολογίσουν ακολουθίες κινήσεων μεταξύ σημείων [51, 26, 10, 54, 86, 11, 32]. Μερικές μέθοδοι όπως οι [53, 57], είναι πλήρως αναδραστικές (reactive), αμφισβητούν την ανάγκη ύπαρξης οποιασδήποτε αναπαράστασης του περιβάλλοντος και θεωρούν το τελευταίο ως την καλύτερη αναπαράσταση του εαυτού του [14]. Οι κινήσεις που επιλέγονται οι μέθοδοι

¹Μια τέτοια περίπτωση είναι π.χ. ο έλεγχος βιομηχανικών ρομπότ σε μια γραμμή παραγωγής.

3.3 Μια εναλλακτική προσέγγιση

αυτές, εξαρτώνται από τις τρέχουσες ενδείξεις των αισθητήρων και μερικές φορές και από κάποια περιορισμένη μνήμη. Παρά το γεγονός ότι οι αναδραστικές μέθοδοι επιτρέπουν την γρήγορη προσαρμογή του συστήματος σε ένα μεταβαλλόμενο περιβάλλον, πάσχουν από την εγγενή έλλειψη γενικής θεώρησης του χώρου. Οι αισθητήρες που τυπικά χρησιμοποιούνται από τις παραπάνω μεθόδους, υπολογίζουν απ' ευθείας την απόσταση μεταξύ του συστήματος και του πλησιέστερου αντικειμένου σε κάποια διεύθυνση. Τέτοιοι αισθητήρες, γνωστοί ως αισθητήρες απόστασης (rangefinders), βασίζονται σε μη οπτικά ερεθίσματα, όπως οι υπέρηχοι, η υπέρυθρη ακτινοβολία και η ακτινοβολία laser. Η λειτουργία τους στηρίζεται στην μέτρηση του χρόνου που μεσολαβεί μεταξύ της εκπομπής κάποιου κύματος και της λήψης της πρώτης ανάκλασής του λόγω κάποιου αντικειμένου. Η απόσταση του αντικειμένου αυτού, υπολογίζεται ως το μισό του γινομένου του χρόνου που μετρήθηκε επί την ταχύτητα μετάδοσης του κύματος.

Χαρακτηριστικό των μεθόδων που περιγράφηκαν στην προηγούμενη παράγραφο, είναι ότι δίνουν πολύ μεγαλύτερη βαρύτητα στην εξαγωγή ακριβών μετρήσεων, παρά στην οργάνωση της χρησιμοποιούμενης αναπαράστασης και στην χρήση της για την εξαγωγή συμπερασμάτων χρήσιμων για πλοήγηση και καθοδήγηση. Επίσης, το είδος των αισθητήρων που χρησιμοποιούν διευκολύνει μερικές λειτουργίες όπως η αποφυγή εμποδίων, δυσχεραίνει όμως κάποιες άλλες όπως η αναγνώριση τμημάτων του περιβάλλοντος και η χρήση τους για τον υπολογισμό της θέσης του παρατηρητή (localization) [25]. Η ακολουθούμενη συλλογιστική τέλος, συνεχίζει να έχει σημαντικές επιδράσεις από την ανακατασκευαστική θεωρία, στοχεύοντας σε μετρικές αναπαραστάσεις οι οποίες δεν μπορούν να χειριστούν ανακριβείς και ελλιπείς πληροφορίες. Στη συνέχεια, περιγράφουμε έναν τρόπο αντιμετώπισης του προβλήματος της πλοήγησης. Βασικοί του ύπονες είναι αφ' ενός η χρήση μιας ποιοτικής περιγραφής του χώρου η οποία είναι ευκολότερο να κατασκευαστεί από ότι μια ποσοτική αναπαράσταση, και αφ' ετέρου ο προγραμματισμός της κίνησης σε δύο επίπεδα, ώστε να λαμβάνονται υπ' όψη τόσο τοπικοί (local) όσο και καθολικοί (global) παράγοντες.

3.3 Μια εναλλακτική προσέγγιση

Η αποτυχία των παραδοσιακών μεθόδων αντιμετώπισης του προβλήματος της οπτικής πλοήγησης να κατασκευάσουν ένα σύστημα ικανό να λειτουργεί σε πραγματικά περιβάλλοντα, αποτελεί το έναυσμα για την αναθεώρησή τους. Οι αρχές της ενεργούς όρασης, όπως παρουσιάστηκαν στο προηγούμενο κεφάλαιο, μπορούν να αποτελέσουν τις κατευθυντήριες γραμμές για μια νέα αντιμετώπιση. Με βάση αυτές, η Fermüller [29] προτείνει ένα σύνολο ικανοτήτων πλοήγησης με διαφορετικούς βαθμούς πολυπλοκότητας και ταξινομημένων ιεραρχικά έτσι ώστε οι ικανότητες των ψηλότερων επιπέδων να στηρίζονται σε αυτές των χαμηλότερων. Σύμφωνα με αυτήν την ταξινόμηση, η απλούστερη ικανότητα ενός συστήματος είναι η εκτίμηση των παραμέτρων κίνησής του, έτσι ώστε το σύστημα να είναι σε θέση να

3.3 Μια εναλλακτική προσέγγιση

μεταβάλλει κάποιες από αυτές μέσω των μηχανικών του μερών. Ακολουθεί η ικανότητα υπολογισμού μερικών παραμέτρων της σχετικής κίνησης ενός αντικειμένου ως προς το σύστημα, όπως η διεύθυνση της μεταφορικής κίνησης και ο χρόνος που απομένει για τυχόν σύγκρουση του συστήματος με το αντικείμενο αυτό. Έπειτα η ικανότητα αποφυγής εμποδίων, η οποία έχει σαν στόχο να διακρίνει εμπόδια που παρεμβάλλονται στην πορεία του συστήματος και να τα αποφύγει με κατάλληλους ελιγμούς. Τέλος, η πιο πολύπλοκη ικανότητα έχει σαν στόχο την μετακίνηση του συστήματος μεταξύ δύο σημείων του περιβάλλοντός του. Αλγόριθμοι για την υλοποίηση των δύο πρώτων ικανοτήτων δίνονται στο [29], της τρίτης στα [62, 69, 9] ενώ προσπάθειες για την υλοποίηση της τέταρτης σε απλουστευμένα περιβάλλοντα, περιγράφονται στα [61, 22]. Στην συνέχεια θα περιγράψουμε μια προσέγγιση για την μελέτη της οπτικής πλοήγησης, με έμφαση στην τέταρτη ικανότητα (*homming*).

Είναι εύκολο να αντιληφθούμε ότι η διαδικασία μετακίνησης ενός συστήματος μεταξύ σημείων του περιβάλλοντος βασίζεται σε κάποιας μορφής αναπαράσταση του τελευταίου και σε ένα σύνολο μεθόδων οι οποίες θα τροποποιούν την αναπαράσταση αυτή και θα την χρησιμοποιούν για να καθορίσουν κατάλληλες κινήσεις των μηχανικών μερών του συστήματος. Αναφορικά με την οργάνωση της αναπαράστασης αυτής, αποτελέσματα από την μελέτη της συμπεριφοράς εξελιγμένων βιολογικών οργανισμών μπορούν να δώσουν χρήσιμες ιδέες. Οι βιολογικοί οργανισμοί απομνημονεύουν χαρακτηριστικά τμήματα του περιβάλλοντος και τα χρησιμοποιούν σαν σημάδια σε συνδυασμό με τις χωρικές σχέσεις μεταξύ τους για να αναγνωρίσουν τοποθεσίες και να προγραμματίσουν την κίνησή τους στο χώρο. Στο εξής, θα αναφερόμαστε στα παραπάνω χαρακτηριστικά τμήματα του χώρου με το όνομα ορόσημα (*landmarks*).

Μιμούμενο τα βιολογικά, ένα μηχανικό σύστημα μπορεί να χρησιμοποιεί ορόσημα για να φτάνει στον προορισμό του. Σε κάθε ορόσημο αντιστοιχεί κάποια περιγραφή, η οποία απομνημονεύεται από το σύστημα. Η αναγνώριση ενός οροσήμου συνίσταται στον υπολογισμό μιας περιγραφής του με βάση οπτικά δεδομένα και στην συνέχεια ταίριασμά της χρησιμοποιώντας κάποιο μέτρο ομοιότητας, με ένα σύνολο περιγραφών το οποίο αντιστοιχεί στα ορόσημα του περιβάλλοντος που είναι γνωστά στο σύστημα από κάποια παραδείγματα. Έτσι, το πρόβλημα της αναγνώρισης οροσήμων, μετατρέπεται σε πρόβλημα υπολογισμού των περιγραφών τους και αναζήτησης παρόμοιων περιγραφών σε μια συσχετιστική μνήμη. Με βάση την παρατήρηση αυτή, η γνώση του συστήματος για το περιβάλλον καθορίζεται από τα περιεχόμενα μιας μνήμης οπτικών πληροφοριών - ΜΟΠ (visual memory). Η τελευταία περιέχει έναν μεγάλο αριθμό από πρότυπα (patterns), τα οποία αντιστοιχούν στις περιγραφές των οροσήμων που είναι γνωστές στο σύστημα. Κάθε τέτοιο πρότυπο μπορεί να συνδέεται με πληροφορίες που θα καθορίζουν την ενέργεια που πρέπει να εκτελεστεί κάθε φορά που αυτό αναγνωρίζεται. Για τον σκοπό της πλοήγησης τέτοιες πληροφορίες μπορεί να είναι η διεύθυνση της κίνησης προς κάποιον προορισμό, η θέση του αντίστοιχου τμήματος του περιβάλλοντος ως προς κάποιο σύστημα αναφοράς, κ.ά. Η ΜΟΠ περιέχει τοπικές περιγραφές

3.3 Μια εναλλακτική προσέγγιση

του περιβάλλοντος, οι οποίες συνιστούν τις τοποθεσίες που είναι γνωστές στο σύστημα, ως προς συστήματα συντεταγμένων τα οποία ορίζονται μέσω των οροσήμων. Η ΜΟΠ είναι οργανωμένη με τέτοιο τρόπο ώστε να επιτρέπει τον χωρικό (spatial) συσχετισμό αυτών των τοπικών συστημάτων συντεταγμένων με βάση τις κινήσεις που εκτέλεσε το σύστημα καθώς μάθαινε τα αντίστοιχα ορόσημα. Με τον τρόπο αυτό, επιτυγχάνεται η αναπαράσταση της τοπολογίας του περιβάλλοντος.

Στο σημείο αυτό, τίθεται ένα ερώτημα που αφορά το είδος των περιγραφών των οροσήμων και τα χαρακτηριστικά που αυτές θα πρέπει να κωδικοποιούν. Ένα σημαντικό χαρακτηριστικό κάθε τμήματος του χώρου είναι το βάθος (depth) των σημείων που το απαρτίζουν. Σύμφωνα με ψυχοφυσικά πειράματα, οι άνθρωποι δεν είναι σε θέση να υπολογίζουν το ακριβές βάθος των σημείων που βλέπουν [40], αλλά μπορούν να υπολογίζουν το διατακτικό βάθος για ζεύγη σημείων [78]. Επιπλέον, πολλά από τα χαρακτηριστικά του περιβάλλοντος που υπολογίζουν οι βιολογικοί οργανισμοί, εκφράζονται σε μονάδες μέτρησης που σχετίζονται με τις μηχανικές τους ικανότητες και όχι σε κάποιες μονάδες οι οποίες ορίζονται συμβατικά [44]. Με βάση αυτά τα πειραματικά δεδομένα αλλά και το γεγονός ότι ο υπολογισμός απόλυτου (μετρικού) βάθους είναι πολύ δύσκολος, προτείνουμε την χρήση διατακτικού βάθους για τις περιγραφές των οροσήμων. Σε επόμενα κεφάλαια, θα παρουσιάσουμε μεθόδους που επιτρέπουν τον υπολογισμό του διατακτικού βάθους των σημείων μιας σκηνής. Επιπλέον χαρακτηριστικά που μπορούν να χρησιμοποιηθούν για τις περιγραφές των οροσήμων, είναι το χρώμα, το texture, κ.ά. Η ΜΟΠ περιέχει ποιοτικές πληροφορίες, υποκαθιστώντας την άμεση, γεωμετρική αναπαράσταση που χρησιμοποιούν οι παραδοσιακές μέθοδοι. Οι σημαντικότερες προϋποθέσεις για να επιτύχει αυτή η προσέγγιση, είναι η επιλογή χαρακτηριστικών και περιγραφών που θα κάνουν δυνατή την αξιόπιστη αναγνώριση των οροσήμων, καθώς και η ανάπτυξη μεθόδων οι οποίες θα επιτρέπουν την επιλογή οροσήμων που θα αποθηκεύονται στην ΜΟΠ. Χρήση οροσήμων για την πλοιήγηση σε ένα μερικώς γνωστό περιβάλλον έχει γίνει σε διάφορες εργασίες, π.χ. [47, 46], καμιά από αυτές όμως δεν λύνει το πρόβλημα επιλογής και αναγνώρισης των οροσήμων, αλλά υποθέτουν ότι τα τελευταία ή είναι γνωστά a priori είτε διαθέτουν πολύ χαρακτηριστικές ιδιότητες που επιτρέπουν την εύκολη αναγνώρισή τους.

Έχοντας δοσμένη την αναπαράσταση του περιβάλλοντος, το επόμενο στάδιο είναι η ανάπτυξη μιας μεθόδου προγραμματισμού κίνησης. Ως τέτοια προτείνεται μια υβριδική μέθοδος, υπό την έννοια ότι επιδιώκει να συνδυάσει την συνολική αναπαράσταση του περιβάλλοντος, όπως αυτή καθορίζεται από τα περιεχόμενα της ΜΟΠ, με τις τοπικές περιγραφές που παρέχουν οι οπτικοί αισθητήρες. Υβριδικές προσεγγίσεις έχουν χρησιμοποιηθεί και σε άλλα συστήματα, όπως τα [37, 71, 21].

Στη μέθοδο που προτείνουμε, ο προγραμματισμός της κίνησης γίνεται σε δύο επίπεδα. Στο ψηλότερο, ένας στρατηγικός (strategic) planner είναι υπεύθυνος για τον προσδιορισμό μιας ακολουθίας από ενδιάμεσους στόχους, η εκπλήρωση των οποίων θα έχει σαν αποτέλεσμα την εκπλήρωση του συνολικού στόχου. Οι ενδιάμεσοι στόχοι αντιστοιχούν σε ορόσημα τα οποία

3.3 Μια εναλλακτική προσέγγιση

έχουν επιλεχθεί από το σύστημα μέσω μιας διαδικασίας εκπαίδευσης. Ο στρατηγικός planner εκτελεί τους υπολογισμούς του με βάση προσεγγιστικές περιγραφές του περιβάλλοντος και κατά συνέπεια τα σχέδια που υπολογίζει δεν είναι πάντα πλήρως εφικτά. Για το λόγο αυτό, υπάρχει ένας *tactical* planner στο ομέσως χαμηλότερο επίπεδο, ο οποίος αναλαμβάνει να μετακινήσει το σύστημα μεταξύ δύο ενδιάμεσων θέσεων που του υποδεικνύονται από τον στρατηγικό planner, λαμβάνοντας ταυτόχρονα υπόψη την κατάσταση του περιβάλλοντος. Ο τακτικός planner θα συνδυάζει τις τρεις πρώτες ικανότητες πλοήγησης που παρουσιάστηκαν παραπάνω, διαθέτοντας ικανότητες εκτίμησης της κίνησης του συστήματος, μερικής εκτίμησης της κίνησης των αντικειμένων του περιβάλλοντος και αποφυγής εμποδίων.

Η παραπάνω διεπίπεδη διαδικασία προγραμματισμού κίνησης επιτρέπει τον υπολογισμό σχεδίων σε διαφορετικά επίπεδα λεπτομέρειας και αφαίρεσης, με τρόπο που έχει κάποιες ομοιότητες με τις αντίστοιχες τεχνικές που χρησιμοποιούν οι ανεπτυγμένοι οργανισμοί. Μια τέτοια προσέγγιση φαίνεται να συνδυάζει τα πλεονεκτήματα του μακροπρόθεσμου και του βραχυπρόθεσμου σχεδιασμού, επιτρέποντας την εκμετάλλευση συνολικής γνώσης για το περιβάλλον, ενώ ταυτόχρονα λαμβάνει υπόψη της τυχόν αλλαγές του τελευταίου. Οι ακολουθίες κινήσεων που προτείνει δεν είναι κατ' ανάγκη βέλτιστες μια και βασίζεται σε ελλιπή δεδομένα, μπορούν όμως να βρίσκονται αρκετά κοντά στο συνολικό βέλτιστο.

Κεφάλαιο 4

Κάθετη ροή

4.1 Εισαγωγή

Μια ακολουθία εικόνων που παράγεται ενώ μια κάμερα κινείται ως προς τα αντικείμενα που απεικονίζει, μπορεί να δώσει χρήσιμες πληροφορίες σχετικά με τα αντικείμενα αυτά. Οι απεικονίσεις των αντικειμένων μετακινούνται μεταξύ διαδοχικών εικόνων, δίνοντας έτσι την εντύπωση μιας κίνησης γνωστής ως φαινόμενη κίνηση (*apparent motion*). Μέσω του υπολογισμού κάποιων χαρακτηριστικών της κίνησης αυτής, είναι δυνατόν να εξαχθούν γεωμετρικά χαρακτηριστικά μιας σκηνής. Στην συνέχεια, θα παρουσιάσουμε τις βασικές εξισώσεις για την κίνηση, θα εισάγουμε την έννοια της κάθετης ροής, η οποία θα χρησιμεύσει σαν είσοδος σε αλγόριθμους που θα περιγραφούν στο επόμενο κεφάλαιο και τέλος θα παρουσιάσουμε το σύστημα συντεταγμένων που θα χρησιμοποιηθεί.

4.2 Οπτική ροή και πεδίο κίνησης

Η αλλαγή της φωτεινότητας των σημείων σε μια ακολουθία εικόνων, προκαλεί την φαινόμενη κίνηση κάποιων τμημάτων, η οποία ονομάζεται οπτική ροή (*optical flow*). Σε κάθε σημείο των περιοχών αυτών αντιστοιχεί ένα διδιάστατο διάνυσμα ταχύτητας, δίνοντας ένα διανυσματικό πεδίο το οποίο περιγράφει τις μετατοπίσεις των σημείων που παρατηρούνται από εικόνα σε εικόνα. Η κίνηση ενός σημείου μιας σκηνής ως προς την κάμερα, προκαλεί την κίνηση της προβολής του πάνω στην εικόνα, με τρόπο που καθορίζεται από το είδος της προβολής (προοπτική, ορθογραφική, σφαιρική, παραπροοπτική, κ.λπ.). Ορίζεται έτσι ένα δεύτερο διανυσματικό πεδίο για τα σημεία μιας εικόνας, το οποίο αντιστοιχεί σε καθένα από αυτά την ταχύτητα κίνησής του. Το πεδίο αυτό, ονομάζεται πεδίο κίνησης (*motion field*) και έχει μια καθαρά γεωμετρική έννοια, αντίθετα με το πεδίο οπτικής ροής το οποίο ορίζεται από τις αλλαγές φωτεινότητας των εικόνων.

4.3 Ποσοτική περιγραφή της φαινόμενης κίνησης

Διαισθητικά, θα περίμενε κανείς τα δυο πεδία που ορίστηκαν παραπάνω να είναι ίσα. Αν και στις περισσότερες περιπτώσεις η υπόθεση αυτή ισχύει, υπάρχουν άλλες όπου παραβιάζεται. Μια από αυτές περιγράφεται από τον Horn [34]: Έστω μια σφαίρα με ομοιόμορφη επιφάνεια, η οποία περιστρέφεται με σταθερή ταχύτητα γύρω από τον άξονά της. Μια φωτεινή πηγή βρίσκεται σε σταθερή θέση ως προς τον άξονα της σφαίρας. Λόγω της καμπύλης επιφάνειας της σφαίρας, θα υπάρχουν περιοχές της τελευταίας που θα παρουσιάζουν σκιές και “γυαλάδες”. Αυτές οι περιοχές διαφορετικής φωτεινότητας όμως, δεν μετακινούνται μαζί με την επιφάνεια της σφαίρας, και κατά συνέπεια η οπτική ροή είναι μηδέν για κάθε σημείο της εικόνας. Αντίθετα, το πεδίο κίνησης είναι μη μηδενικό, μια και τα σημεία της επιφάνειας της σφαίρας κινούνται. Στην περίπτωση που η σφαίρα είναι ακίνητη αλλά μετακινείται η φωτεινή πηγή, οι περιοχές διαφορετικής φωτεινότητας μετακινούνται πάνω στην επιφάνεια της σφαίρας, έχοντας σαν αποτέλεσμα ένα μη μηδενικό πεδίο οπτικής ροής. Το πεδίο κίνησης όμως, είναι προφανώς μηδενικό. Γενικά, στις περιπτώσεις ύπαρξης σκιών, το πεδίο οπτικής ροής δεν είναι ίσο με το πεδίο κίνησης. Οι Verri και Poggio [82] έδειξαν ότι τα δύο πεδία είναι ακριβώς ίσα μόνο στην περίπτωση που η επιφάνεια που απεικονίζεται είναι τέτοια ώστε να ανακλά το φως ομοιόμορφα προς κάθε κατεύθυνση (Lambertian surface) ενώ η κίνησή της ως προς την κάμερα είναι μόνο μεταφορική.

4.3 Ποσοτική περιγραφή της φαινόμενης κίνησης

Θεωρούμε ότι η κίνηση που μεσολαβεί μεταξύ δύο διαδοχικών εικόνων είναι μικρή, κατά συνέπεια η ακολουθία εικόνων που προκύπτει μπορεί να παρασταθεί από μια συνεχή συνάρτηση $I(x, y, t)$ δύο χωρικών (x, y) και μιας χρονικής μεταβλητής (t) . Η συνάρτηση αυτή μπορεί να αναλυθεί με τεχνικές διαφορικού λογισμού. Έστω ότι ένα σημείο του κόσμου απεικονίζεται στο σημείο (x, y) μιας εικόνας, το οποίο την χρονική στιγμή t έχει ένταση (intensity) $I(x, y, t)$ και διάνυσμα οπτικής ροής ίσο με (u, v) . Την χρονική στιγμή $t + dt$, το σημείο αυτό απεικονίζεται στο σημείο $(x + dx, y + dy)$, όπου $dx = udt$, $dy = vdt$. Υποθέτωντας ότι ένα σημείο του κόσμου απεικονίζεται σε σημεία της εικόνας τα οποία έχουν την ίδια ένταση, οι εντάσεις των σημείων (x, y) και $(x + dx, y + dy)$ τις χρονικές στιγμές t και $t + dt$ θα πρέπει να είναι ίσες, δηλαδή,

$$I(x, y, t) = I(x + udt, y + vdt, t + dt)$$

Αναπτύσσοντας το δεξιό μέλος της παραπάνω εξίσωσης σε μια σειρά Taylor, παίρνουμε

$$I(x, y, t) = I(x, y, t) + dx \frac{\partial I}{\partial x} + dy \frac{\partial I}{\partial y} + dt \frac{\partial I}{\partial t} + e$$

όπου ο προσθεταίος e αντιστοιχεί στους όρους μεγαλύτερου βαθμού και $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, $\frac{\partial I}{\partial t}$ οι μερικές παράγωγοι της I ως προς x, y, t αντίστοιχα. Απαλείφοντας τον όρο $I(x, y, t)$, διαιρώντας δια

4.3 Ποσοτική περιγραφή της φαινόμενης κίνησης

dt και παίρνοντας το όριο της προκύπτουσας εξίσωσης όταν $dt \rightarrow 0$, φτάνουμε στην

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0$$

Η εξίσωση αυτή είναι το ανάπτυγμα της $\frac{dI}{dt} = 0$, όπου $\frac{dI}{dt}$ η ολική παράγωγος της I ως προς το χρόνο. Χρησιμοποιώντας τις συντομογραφίες

$$u = \frac{dx}{dt}, \quad v = \frac{dy}{dt},$$

$$I_x = \frac{\partial I}{\partial x}, \quad I_y = \frac{\partial I}{\partial y}, \quad I_t = \frac{\partial I}{\partial t},$$

καταλήγουμε στην

$$I_x u + I_y v + I_t = 0 \quad (4.1)$$

Η εξίσωση (4.1) είναι γνωστή σαν εξίσωση περιορισμού της οπτικής ροής (*optical flow constraint equation*) μια και εκφράζει έναν περιορισμό για τις συνιστώσες u και v της οπτικής ροής και αναπτύχθηκε αρχικά από τους Horn και Schunck [35]. Γράφοντας την εξίσωση αυτή με μορφή εσωτερικού γινομένου

$$(I_x, I_y) \cdot (u, v) = -I_t \quad (4.2)$$

βλέπουμε ότι επιτρέπει τον υπολογισμό της προβολής της οπτικής ροής στο διάνυσμα της κλίσης της έντασης (intensity gradient), το οποίο είναι κάθετο στην ακμή στο αντίστοιχο σημείο. Η αλγεβρική τιμή της προβολής αυτής μπορεί να υπολογιστεί από την (4.2) διαιρώντας και τα δύο μέλη της με το μέτρο της κλίσης και είναι ίση με

$$-\frac{I_t}{\sqrt{I_x^2 + I_y^2}} \quad (4.3)$$

Η συνιστώσα της οπτικής ροής κατά μήκος της κλίσης, ονομάζεται κάθετη ροή (*normal flow*) και είναι ίση με

$$-\frac{I_t}{\sqrt{I_x^2 + I_y^2}} \left(\frac{I_x}{\sqrt{I_x^2 + I_y^2}}, \frac{I_y}{\sqrt{I_x^2 + I_y^2}} \right)$$

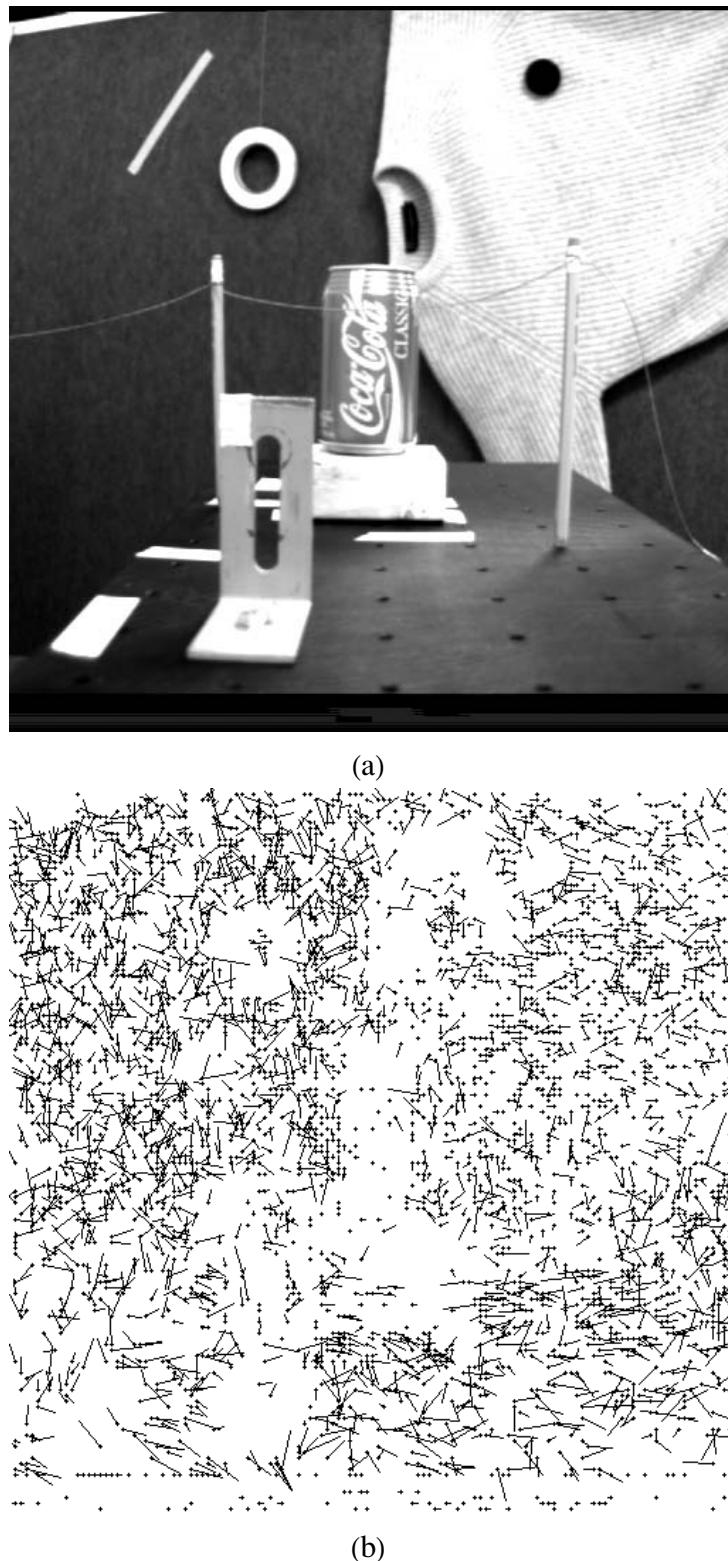
δηλαδή με

$$\left(-\frac{I_t I_x}{\|\nabla I\|^2}, -\frac{I_t I_y}{\|\nabla I\|^2} \right)$$

όπου $\|\nabla I\|$ είναι το μέτρο της κλίσης. Το σχήμα 4.1.(b) απεικονίζει ένα τυπικό πεδίο κάθετης ροής για την σκηνή του σχήματος 4.1.(a). Το μήκος των διανυσμάτων είναι ανάλογο του μέτρου τους.

Η εξίσωση (4.2) περιορίζει το διάνυσμα της οπτικής ροής να βρίσκεται σε ένα από τα δύο ημιεπίπεδα που ορίζει η ακμή. Συγκεκριμένα, αν η κάθετη ροή είναι ομόροπη με την κλίση, τότε η αντίστοιχη οπτική ροή πρέπει να περιέχεται στο ημιεπίπεδο που περιέχει το διάνυσμα

4.3 Ποσοτική περιγραφή της φαινόμενης κίνησης



Σχήμα 4.1: Μια σκηνή (a) και ένα πεδίο κόθετης ροής γι' αυτήν (b).

4.3 Ποσοτική περιγραφή της φαινόμενης κίνησης

της κλίσης (το οποίο είναι κάθετο στην ακμή). Στην περίπτωση που η κάθετη ροή είναι αντίρροπη με την κλίση, τότε η οπτική ροή δεν μπορεί να είναι στο ημιεπίπεδο που περιέχει το διάνυσμα της κλίσης. Ο παραπάνω περιορισμός γίνεται σαφέστερος με το σχήμα 4.2. Σ' αυτό, η χοντρή μαύρη γραμμή είναι η ακμή, η λεπτή το διάνυσμα της κλίσης και η διακεκομμένη το διάνυσμα της κάθετης ροής. Η οπτική ροή πρέπει κάθε φορά να βρίσκεται στο ημιεπίπεδο που περιέχει το διάνυσμα της κάθετης ροής, το οποίο εμφανίζεται γραμμοσκιασμένο στο σχήμα.



Σχήμα 4.2: Η φορά της κάθετης ροής περιορίζει την οπτική ροή σε ένα ημιεπίπεδο.

Η αλγεβρική τιμή του πεδίου κάθετης κίνησης (*normal motion field*), της συνιστώσας δηλαδή του πεδίου κίνησης κατά μήκος της κλίσης είναι ίση με

$$\begin{aligned} \left(\frac{dx}{dt}, \frac{dy}{dt} \right) \cdot \frac{(I_x, I_y)}{\sqrt{I_x^2 + I_y^2}} &= \\ \left(\frac{dx}{dt}, \frac{dy}{dt} \right) \cdot \frac{\nabla I}{\|\nabla I\|} &= \\ \frac{1}{\|\nabla I\|} (I_x \frac{dx}{dt} + I_y \frac{dy}{dt}) \end{aligned} \quad (4.4)$$

Η διαφορά μεταξύ των αλγεβρικών τιμών της κάθετης ροής και του πεδίου κάθετης κίνησης, δίνεται από την διαφορά των εξισώσεων (4.4) και (4.3) και είναι ίση με

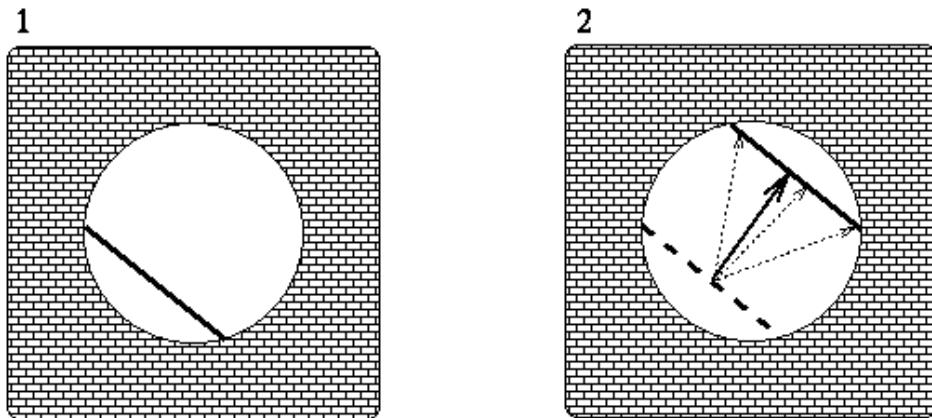
$$\frac{1}{\|\nabla I\|} \frac{dI}{dt}$$

Η τελευταία παράσταση, δείχνει ότι η κάθετη ροή είναι καλή προσέγγιση του πεδίου κάθετης κίνησης στα σημεία που το $\|\nabla I\|$ είναι μεγάλο. Κατά συνέπεια, η κάθετη ροή υπολογίζεται μόνο στα σημεία με μεγάλη κλίση. Πρακτικά αυτό σημαίνει ότι οι αλγόριθμοι υπολογίζουν την κάθετη ροή μόνο στα σημεία στα οποία το μέτρο της κλίσης υπερβαίνει κάποιο κατώφλι (threshold).

Η αδυναμία υπολογισμού της συνιστώσας της οπτικής ροής κατά μήκος μιας κατεύθυνσης διαφορετικής από αυτήν της κλίσης, οφείλεται σε ένα πρόβλημα γνωστό σαν πρόβλημα του ανοίγματος (*aperture problem*). Το πρόβλημα αυτό παρουσιάζεται στο σχήμα 4.3. Έστω ότι παρακολουθούμε ένα χαρακτηριστικό μιας σκηνής (γραμμή στην προκειμένη περίπτωση),

4.3 Ποσοτική περιγραφή της φαινόμενης κίνησης

μέσω ενός ανοίγματος το οποίο έχει μικρότερες διαστάσεις από αυτό. Με βάση δύο διαδοχικές εικόνες μέσα από το άνοιγμα, δεν είναι δυνατόν να καθορίσουμε που κινήθηκε κάθε σημείο του χαρακτηριστικού αυτού. Η μόνη πληροφορία που μπορεί να εξαχθεί από τοπικές μετρήσεις είναι η συνιστώσα της ταχύτητας που είναι κάθετη στο χαρακτηριστικό (στην προκειμένη περίπτωση η συνεχής μαύρη γραμμή με το βέλος στην δεύτερη εικόνα), δηλαδή η κάθετη ροή. Στην περίπτωση που το άνοιγμα είναι τοποθετημένο γύρω από κάποιο σημείο ασυνέχειας του χαρακτηριστικού (π.χ. τέλος μιας γραμμής, γωνία κ.λπ.), το πρόβλημα του ανοίγματος παύει να υφίσταται, αφού είναι η θέση του σημείου αυτού σε δύο χρονικές στιγμές είναι γνωστή με ακρίβεια.

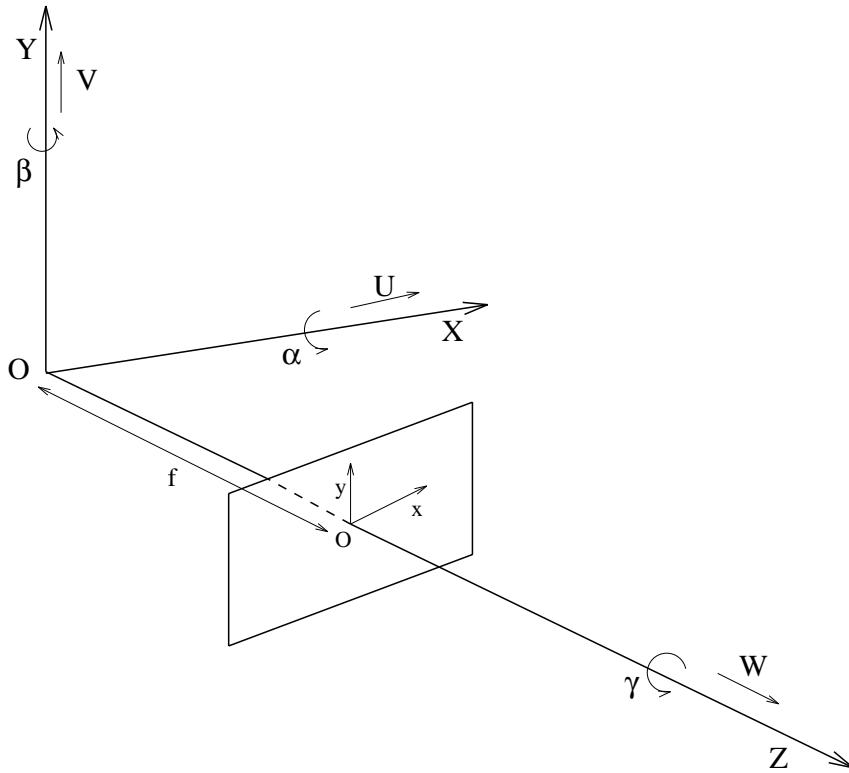


Σχήμα 4.3: Το πρόβλημα του ανοίγματος. Η νέα θέση ενός σημείου στην δεύτερη εικόνα, δεν είναι δυνατό να καθοριστεί με βάση μόνο τοπικές μετρήσεις.

Η εξίσωση (4.1) παρέχει έναν τοπικό περιορισμό για τον υπολογισμό της οπτικής ροής. Το πρόβλημα του ανοίγματος, το οποίο είναι ανεξάρτητο από την μέθοδο που χρησιμοποιείται για τον υπολογισμό της οπτικής ροής, συνεπάγεται ότι ο περιορισμός αυτός είναι ο μόνος που μπορεί να διατυπωθεί αναφορικά με τοπικές μετρήσεις. Δεδομένου όμως ότι υπάρχουν δύο άγνωστοι (u και v), χρειάζεται κάποιος επιπλέον περιορισμός, ο οποίος θα είναι κατ' ανάγκη καθολικός (global). Ένας περιορισμός που έχει χρησιμοποιηθεί κατά κόρον, απαιτεί ότι η οπτική ροή μεταβάλλεται ομαλά σε όλη την εικόνα [35]. Στην πράξη, τέτοιες υποθέσεις οδηγούν σε ένα μη γραμμικό πρόβλημα βελτιστοποίησης, το οποίο είναι πολύ δύσκολο να επιλυθεί και για το λόγο αυτό δεν θα προσπαθήσουμε να υπολογίσουμε την οπτική ροή. Οι μέθοδοι που θα αναπτυχθούν στην συνέχεια θα βασιστούν στην κάθετη ροή, η οποία μπορεί να υπολογιστεί άμεσα από μια ακολουθία εικόνων.

4.4 Το σύστημα συντεταγμένων

Έστω το σύστημα συντεταγμένων $OXYZ$ το οποίο είναι προσαρμοσμένο στο οπτικό κέντρο μιας κάμερας, έτσι ώστε ο άξονας OZ να συμπίπτει με τον οπτικό άξονα. Έστω επίσης ότι η κάμερα κινείται ως προς μια στατική σκηνή με μεταφορική κίνηση $\vec{t} = (U, V, W)$ και περιστροφική κίνηση $\vec{\omega} = (\alpha, \beta, \gamma)$ όπως φαίνεται στο σχήμα 4.4.



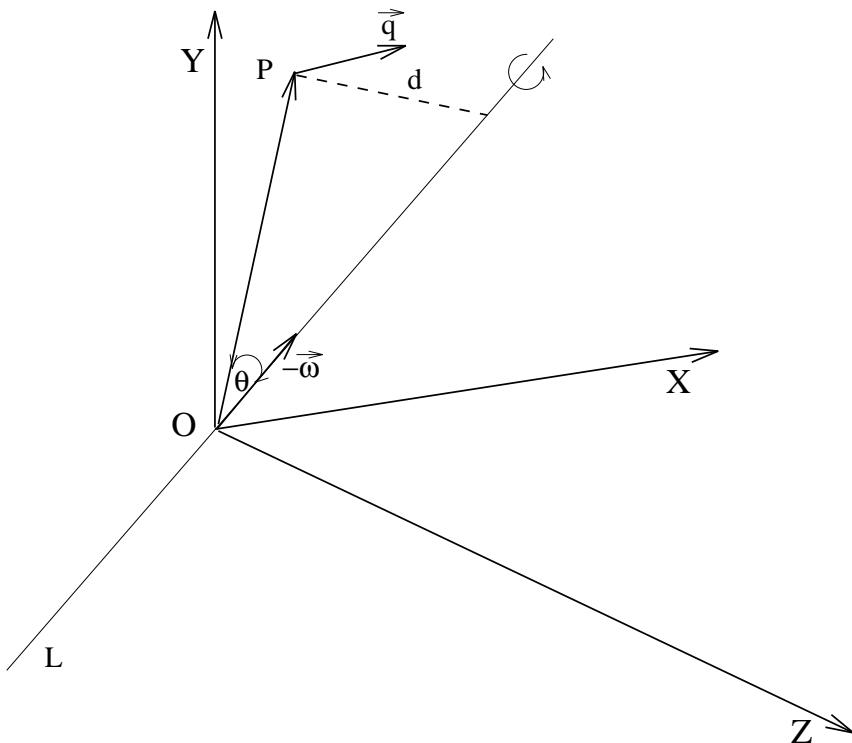
Σχήμα 4.4: Το σύστημα συντεταγμένων.

Η περίπτωση αυτή είναι γεωμετρικά ισοδύναμη με την περίπτωση όπου η κάμερα είναι ακίνητη και η σκηνή κινείται ως προς αυτή με μεταφορική ταχύτητα $-\vec{t}$ και περιστροφική $-\vec{\omega}$. Η ταχύτητα \vec{v} ενός σημείου $P = (X, Y, Z)$ ως προς το σύστημα συντεταγμένων $OXYZ$ υπολογίζεται ως εξής: Έστω L ο άξονας περιστροφής που ορίζεται από το διάνυσμα $-\vec{\omega}$ και d η απόσταση του σημείου P από αυτόν (βλ. σχήμα 4.5). Το P περιστρέφεται γύρω από τον L με φορά αντίθετη εκείνης των δεικτών του ρολογιού. Η \vec{v} είναι το άθροισμα της εφαπτομενικής ταχύτητας λόγω περιστροφής \vec{q} και της μεταφορικής ταχύτητας $-\vec{t}$. Η \vec{q} είναι κάθετη στο επίπεδο των L και P και μέτρο της είναι ίσο με το γινόμενο του μέτρου της γωνιακής ταχύτητας επί την απόσταση του P από το L , δηλαδή $\vec{q} = \| -\vec{\omega} \| d = \| -\vec{\omega} \| \|P\| \sin\theta = -\vec{\omega} \times P$, όπου \times συμβολίζει το εξωτερικό γινόμενο. Έτσι, η \vec{v} είναι ίση με $-\vec{\omega} \times P - \vec{t}$. Κατά συνέπεια, οι τρεις συνιστώσες της ταχύτητας του P είναι ίσες με

$$X' = -U - \beta Z + \gamma Y$$

4.4 Το σύστημα συντεταγμένων

$$\begin{aligned} Y' &= -V - \gamma X + \alpha Z \\ Z' &= -W - \alpha Y + \beta X \end{aligned} \quad (4.5)$$



Σχήμα 4.5: Περιστροφή του σημείου P γύρω από τον άξονα L .

Οι εξισώσεις (4.5) θα χρησιμοποιηθούν στο επόμενο κεφάλαιο για να συσχετίσουν το πεδίο κίνησης με τις παραμέτρους κίνησης \vec{t} και $\vec{\omega}$.

Κεφάλαιο 5

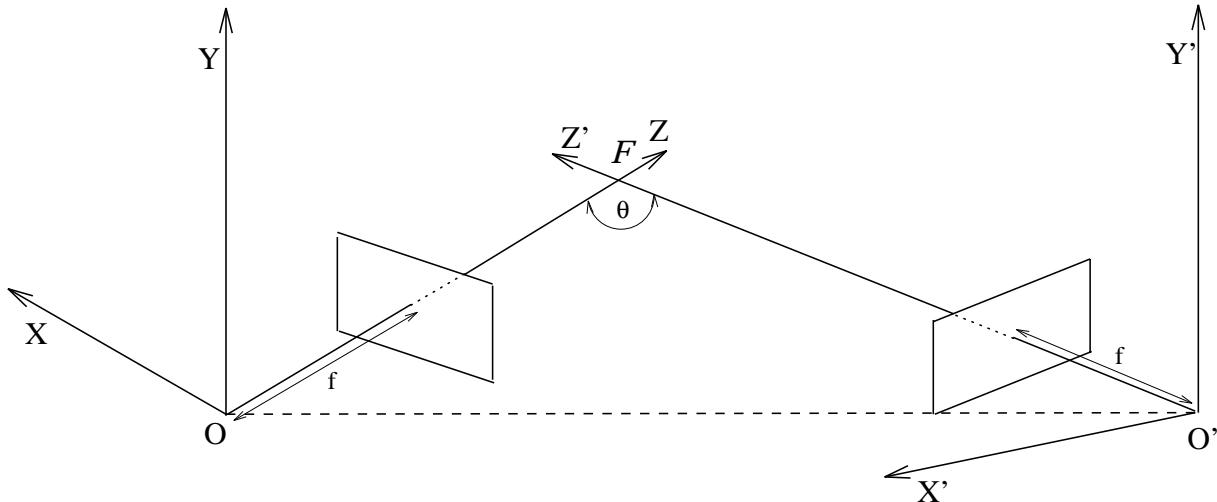
Διατακτικό βάθος

5.1 Εισαγωγή

Στο κεφάλαιο αυτό θα παρουσιάσουμε μια μέθοδο που επιτρέπει την διάταξη των σημείων μιας σκηνής με βάση το βάθος τους, την απόστασή τους δηλαδή από τον παρατηρητή. Ονομάζουμε την αναπαράσταση που προκύπτει διατακτική αναπαράσταση του βάθους. Η διατακτική αναπαράσταση του βάθους είναι μερική, υπό την έννοια ότι δεν κωδικοποιεί όλη την πληροφορία αναφορικά με το βάθος (π.χ. δεν περιλαμβάνει την αριθμητική διαφορά του βάθους δύο σημείων). Πιστεύουμε όμως ότι είναι αρκετή για να χρησιμοποιηθεί σαν είσοδος από μεθόδους υπολογισμού περιγραφών του σχήματος των αντικειμένων και αναπαραστάσεων του χώρου γενικότερα. Η προσέγγιση που θα ακολουθηθεί, κινείται σύμφωνα με τις κατευθύνσεις της μεθοδολογίας για την όραση που παρουσιάστηκε στο κεφάλαιο 2. Συγκεκριμένα, εκμεταλλεύεται καθαρά γεωμετρικούς περιορισμούς, χωρίς να καταφεύγει σε ειδικές υποθέσεις σχετικά με το είδος των αντικειμένων που απεικονίζονται. Επιπλέον, οι χρησιμοποιούμενοι υπολογισμοί είναι ποιοτικοί, βασιζόμενοι στο πρόσημο κάποιων ποσοτήτων και όχι στην αλγεβρική τους τιμή. Στην συνέχεια, περιγράφεται το είδος των εικόνων που χρησιμοποιούνται και η πειραματική διάταξη για την απόκτησή τους, αναπτύσσονται οι εξισώσεις που επιτρέπουν τον υπολογισμό διατακτικού βάθους και τέλος παρουσιάζεται μια υλοποίηση της μεθόδου. Η μέθοδος αυτή έχει ανάλογο πνεύμα με αυτήν που περιγράφεται στο [38], η οποία επιτρέπει σε έναν ενεργό παρατηρητή να υπολογίσει σχετικό βάθος για το σύνολο των σημείων ενός αντικειμένου στα οποία η κλίση της έντασης είναι κάθετη στο διάνυσμα που ορίζει την παράλληλη μεταφορική κίνηση του αντικειμένου αυτού ως προς το εστιακό επίπεδο.

5.2 Η είσοδος

Η μέθοδος που αναπτύχθηκε, χρησιμοποιεί ένα ζεύγος εικόνων, κάθε μια από τις οποίες απεικονίζει μια σκηνή από ελαφρώς διαφορετικό σημείο. Οι εικόνες είναι τέτοιες ώστε το κέντρο τους να απεικονίζει το ίδιο σημείο F της σκηνής. Ένα τέτοιο ζεύγος εικόνων, είναι γνωστό ως *fixating image pair* και μπορεί να καταγραφεί από ένα σύστημα δύο καμερών, οι οποίες είναι στραμμένες έτσι ώστε οι οπτικοί τους άξονες να τέμνονται στο σημείο F , το οποίο ονομάζεται *fixation point*. Η γωνία θ που σχηματίζουν οι δύο οπτικοί άξονες ονομάζεται *vergence angle*. Η διάταξη του συστήματος των δύο καμερών φαίνεται στο σχήμα 5.1, το οποίο δείχνει την σχετική θέση μεταξύ των συστημάτων συντεταγμένων για τις κάμερες. Πρόκειται δηλαδή για ένα στερεοσκοπικό σύστημα (*stereo system*), στο οποίο οι οπτικοί άξονες συγκλίνουν αντί να είναι παράλληλοι όπως συχνά συμβαίνει στα στερεοσκοπικά συστήματα. Η διάταξη αυτή είναι σχετικά εύκολο να επιτευχθεί με μια μηχανική κεφαλή όπως αυτές που περιγράφηκαν στο κεφάλαιο 2.

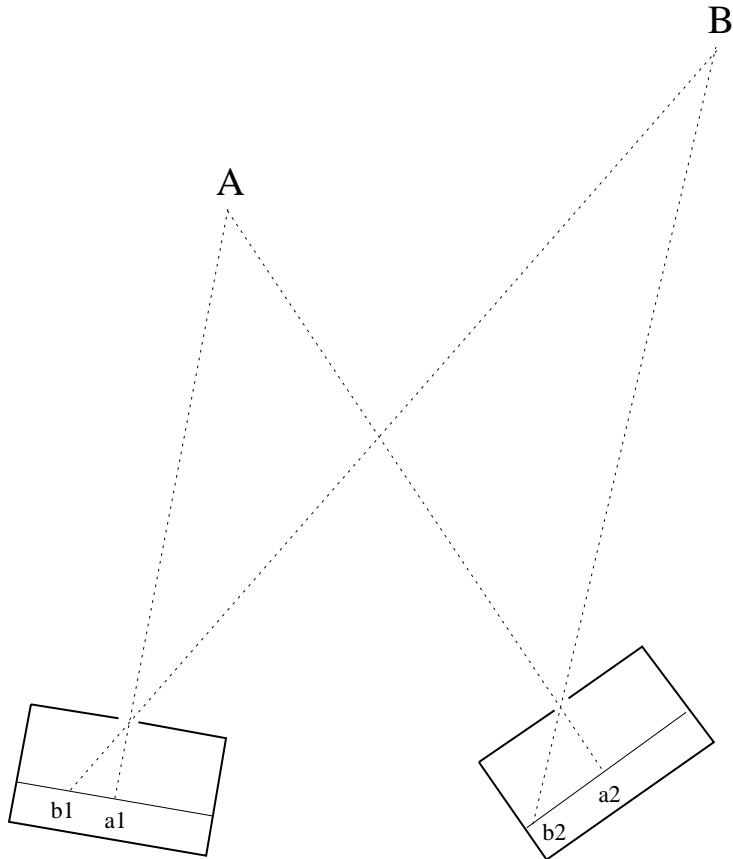


Σχήμα 5.1: Η γεωμετρία του συστήματος των δύο καμερών.

Τυπικά, η επεξεργασία ενός στερεοσκοπικού ζεύγους στοχεύει στον υπολογισμό του (μετρικού) βάθους κάθε σημείου της σκηνής. Η πληροφορία για το βάθος βρίσκεται στις διαφορετικές θέσεις στα επίπεδα των δύο εικόνων στις οποίες απεικονίζεται ένα σημείο της σκηνής. Η διαφορά αυτή, γνωστή σαν *disparity* [7], είναι ίση με την διαφορά $a_2b_2 - a_1b_1$ στο σχήμα 5.2. Η διαφορά αυτή σε συνδυασμό με άλλες παραμέτρους του συστήματος (την εστιακή απόσταση f και το μήκος του τμήματος OO' , γνωστού σαν *baseline*), οι οποίες υπολογίζονται από μια διαδικασία καλιμπραρίσματος (calibration), δίνει εύκολα το βάθος, δηλαδή την Z συντεταγμένη κάθε σημείου: Ένα σημείο του κόσμου πρέπει να βρίσκεται πάνω στην ευθεία γραμμή που ορίζεται από το κέντρο του συστήματος συντεταγμένων μιας κάμερας και το σημείο του επιπέδου της εικόνας στο οποίο προβάλλεται. Για ένα

5.2 Η είσοδος

στερεοσκοπικό σύστημα, ο περιορισμός αυτός ισχύει και για τις δύο κάμερες. Επομένως, ένα σημείο πρέπει να βρίσκεται στην τομή των ευθειών που ορίζονται τα κέντρα των συστημάτων συντεταγμένων στις κάμερες και οι προβολές του στα επίπεδα των εικόνων. Αυτή η διαδικασία υπολογισμού του βάθους ονομάζεται *triangulation*. Στην πράξη, το δυσκολότερο πρόβλημα αφορά τον υπολογισμό του *disparity*, μια και προϋποθέτει την επίλυση του προβλήματος της αντιστοίχισης (βλ. κεφ. 3). Η μέθοδος υπολογισμού διατακτικού βάθους που αναπτύσσεται εδώ, θα ακολουθήσει μια διαφορετική προσέγγιση, η οποία δεν απαιτεί την πλήρη επίλυση του προβλήματος της αντιστοίχισης.



Σχήμα 5.2: Οι δύο κάμερες εστιάζουνται στο σημείο A. Το A απεικονίζεται στο κέντρο του επιπέδου εικόνας και στις δύο κάμερες. Ένα πια μακρινό σημείο B, απεικονίζεται σε διαφορετικά σημεία στα επίπεδα των εικόνων.

Ένα ζεύγος εικόνων που συλλέγεται από ένα στερεοσκοπικό σύστημα, μπορεί να θεωρηθεί σαν μια ακολουθία εικόνων και να αναλυθεί με τεχνικές ανάλυσης κίνησης. Πράγματι, ένα τέτοιο ζεύγος εικόνων είναι το ίδιο με αυτό που θα κατέγραψε μια κινούμενη κάμερα, η οποία την χρονική στιγμή t θα ήταν στην θέση της αριστερής (δεξιάς) κάμερας του στερεοσκοπικού συστήματος και την χρονική στιγμή $t + 1$ στην θέση της δεξιάς (αριστερής). Η κάμερα αυτή κινείται με τέτοιο τρόπο ώστε οι οπτικοί της άξονες σε δυο διαδοχικές θέσεις της να διέρχονται από το ίδιο σημείο μιας σκηνής και χαρακτηρίζεται σαν fixating camera.

5.3 Το μοντέλο σχηματισμού των εικόνων

Η προηγούμενη παρατήρηση, μπορεί να ανάγει κάθε πρόβλημα στερεοσκοπικής ανάλυσης σε πρόβλημα ανάλυσης κίνησης, και ισχύει ανεξάρτητα από την σχετική θέση των καμερών του στερεοσκοπικού συστήματος. Ο υπολογισμός του διατακτικού βάθους θα βασιστεί στην ανάλυση της υποθετικής κίνησης της μιας κάμερας στην θέση της άλλης. Κατά την διάρκεια αυτής της υποθετικής κίνησης, το περιβάλλον είναι στατικό δεδομένου ότι οι εικόνες κάθε fixating ζεύγους καταγράφονται την ίδια χρονική στιγμή.

Οι λόγοι που οδήγησαν στην χρήση ενός fixating ζεύγους σχετίζονται αφ' ενός με το γεγονός ότι οι εξισώσεις που περιγράφουν την κίνηση είναι απλές και αφ' ετέρου με το ότι είναι εφικτή η χρήση διαφορικών τεχνικών. Συγκεκριμένα, η υποθετική κίνηση της κάμερας καθορίζεται από λιγότερες παραμέτρους από ότι η γενική περίπτωση κίνησης που εξετάστηκε στο κεφ. 4. Το γεγονός ότι τα disparities σε μια περιοχή γύρω από το κέντρο των εικόνων είναι μικρά, επιτρέπει τον υπολογισμό της κάθετης ροής με την μέθοδο που παρουσιάστηκε στο κεφ. 4. Επιπλέον, σε μια περιοχή γύρω από τον οπτικό άξονα, είναι εφικτή η προσέγγιση της διαδικασίας σχηματισμού των εικόνων με το μοντέλο της ορθογραφικής προβολής (*orthographic projection*). Τέλος, ένα άλλο κίνητρο για την μελέτη της διαδικασίας του fixation είναι η μεγάλη διάδοσή της μεταξύ των βιολογικών οργανισμών που χρησιμοποιούν στερεοσκοπική όραση. Το ανθρώπινο σύστημα όρασης για παράδειγμα, κάνει συνεχώς fixations, τα οποία συνήθως διαρκούν περίπου 1/4 του δευτερολέπτου. Ανάμεσα σε διαδοχικά fixations, μεσολαβούν πολύ γρήγορες κινήσεις των ματιών (saccadic eye movements), κατά την διάρκεια των οποίων δεν γίνεται καμιά οπτική επεξεργασία.

5.3 Το μοντέλο σχηματισμού των εικόνων

Ένα σημείο P με συντεταγμένες (X, Y, Z) ως προς το σύστημα αναφοράς που είναι προσαρμοσμένο στην κάμερα, προβάλλεται μέσω της προοπτικής προβολής στο σημείο (\tilde{x}, \tilde{y}) στο επίπεδο της εικόνας, με

$$\tilde{x} = \frac{fX}{Z}$$

$$\tilde{y} = \frac{fY}{Z}$$

όπου f είναι η εστιακή απόσταση (focal length). Ένα άλλο μοντέλο προβολής, είναι αυτό της ασθενούς προοπτικής προβολής (weak perspective ή scaled orthographic), σύμφωνα με το οποίο το P απεικονίζεται στο (x, y) με

$$\begin{aligned} x &= sX \\ y &= sY \end{aligned} \tag{5.1}$$

όπου s είναι μια θετική σταθερά, συνήθως ίση με f/Z_0 , και Z_0 το μέσο βάθος του παρατηρούμενου περιβάλλοντος. Η ασθενής προοπτική προβολή οδηγεί σε απλές εξισώσεις για

5.4 Υπολογισμός διατακτικού βάθους

την κίνηση γιατί οι σχέσεις (5.1) είναι γραμμικές ως προς X, Y . Αντίθετα, οι σχέσεις για την προοπτική προβολή περιέχουν το Z στον παρονομαστή, δίνοντας έτσι πιο πολύπλοκες εξισώσεις. Εξετάζουμε στην συνέχεια σε ποιές περιπτώσεις η ασθενής προοπτική προβολή αποτελεί καλή προσέγγιση της προοπτικής προβολής. Έστω E το λάθος για την τετμημένη, όπου $E = |\check{x} - x|$. Αντικαθιστώντας τις τιμές των \check{x} και x παίρνουμε

$$\begin{aligned} E &= \left| \frac{fX}{Z} - sX \right| = \\ &= \left| \frac{fX}{Z} - \frac{fX}{Z_0} \right| = \\ &= \left| \frac{fX}{Z} \right| \left| 1 - \frac{Z}{Z_0} \right| \end{aligned} \quad (5.2)$$

Η ποσότητα $\frac{fX}{Z}$ είναι η τετμημένη της προβολής στο επίπεδο της εικόνας, επομένως η (5.2) γίνεται

$$E = |\check{x}| \left| 1 - \frac{Z}{Z_0} \right| \quad (5.3)$$

Σύμφωνα με την (5.3), το λάθος είναι μικρό για σημεία που βρίσκονται κοντά στον οπτικό άξονα ή έχουν βάθος Z κοντά στο μέσο βάθος Z_0 . Σε ανάλογα συμπεράσματα καταλήγουμε αν υπολογίσουμε το λάθος για την τεταγμένη. Επομένως, στην περίπτωση που χρησιμοποιείται ένα fixating ζεύγος, είναι εφικτή η χρήση της ασθενούς προοπτικής προβολής για σημεία που βρίσκονται κοντά στο κέντρο των εικόνων.

5.4 Υπολογισμός διατακτικού βάθους

Στην συνέχεια θα αναπτύξουμε τις εξισώσεις που επιτρέπουν τον υπολογισμό διατακτικού βάθους. Αρχίζουμε με τις εξισώσεις που συσχετίζουν την κάθετη ροή με τις παραμέτρους κίνησης.

Έστω το σημείο (X, Y, Z) , το οποίο μέσω της ασθενούς προοπτικής προβολής απεικονίζεται στο σημείο (x, y) . Η τιμή (u, v) του πεδίου κίνησης στο σημείο (x, y) είναι εξ' ορισμού ίση με

$$\begin{aligned} u &= \frac{dx}{dt} = s \frac{dX}{dt} \\ v &= \frac{dy}{dt} = s \frac{dY}{dt} \end{aligned} \quad (5.4)$$

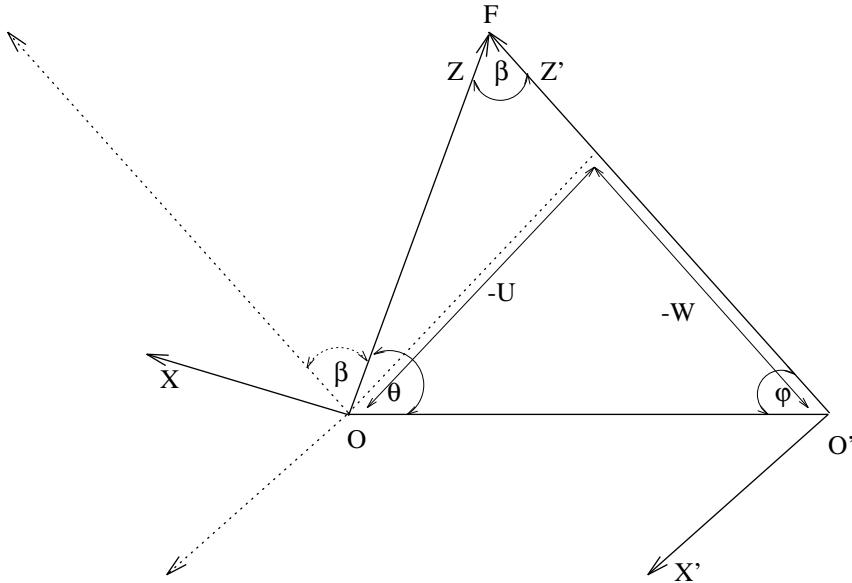
Οι $\frac{dX}{dt}, \frac{dY}{dt}$ δίνονται από τις εξισώσεις (4.5), οπότε οι (5.4) με αντικατάσταση δίνουν

$$\begin{aligned} u &= s(-\beta Z + \gamma Y - U) \\ v &= s(-\gamma X + \alpha Z - V) \end{aligned} \quad (5.5)$$

όπου τα $\alpha, \beta, \gamma, U, V$ είναι παράμετροι κίνησης της κάμερας, οι οποίες ορίστηκαν στο κεφ. 4.

5.4 Υπολογισμός διατακτικού βάθους

Στην συνέχεια θα ακολουθούμε την σύμβαση ότι η υποθετική κίνηση της κάμερας που αναφέρθηκε παραπάνω γίνεται με φορά από την αριστερή κάμερα του στερεοσκοπικού συστήματος προς την δεξιά. Η κίνηση αυτή περιγράφεται με λιγότερες παραμέτρους από τις πέντε που εμφανίζονται στις εξισώσεις της προηγούμενης παραγράφου. Συγκεκριμένα, από το σχήμα 5.3 φαίνεται ότι το σύστημα συντεταγμένων $OXYZ$ μπορεί να μεταφερθεί ώστε να συμπέσει με το $O'X'Y'Z'$ με μια περιστροφή γύρω από τον άξονα OY , μια μεταφορά στον άξονα OX και μια μεταφορά στον άξονα OZ . Οι κινήσεις αυτές περιγράφονται από τις παραμέτρους β, U, W αντίστοιχα. Η περιστροφή γύρω από τον άξονα OY είναι κατά την θετική φορά, επομένως είναι $\beta > 0$. Οι μετατοπίσεις είναι αρνητικές αφού έχουν την αντίθετη φορά από αυτήν των θετικών ημιαξόνων OX, OZ , επομένως είναι $U < 0, W < 0$. Οι παράμετροι α, γ, V είναι ίσες με μηδέν.



Σχήμα 5.3: Κάτοψη των συστημάτων συντεταγμένων για τις δύο κάμερες. Οι άξονες $OY, O'Y'$ είναι κάθετοι στο επίπεδο της σελίδας, με φορά προς τον αναγνώστη.

Με βάση τα παραπάνω, οι εξισώσεις (5.5) απλοποιούνται στις

$$\begin{aligned} u &= s(-\beta Z - U) \\ v &= 0 \end{aligned} \tag{5.6}$$

Η αλγεβρική τιμή του πεδίου κάθετης κίνησης είναι ίση με

$$\begin{aligned} u^n &= (u, v) \left(\frac{I_x}{\|\nabla I\|}, \frac{I_y}{\|\nabla I\|} \right) = \\ &\frac{I_x}{\|\nabla I\|} s(-\beta Z - U) \end{aligned} \tag{5.7}$$

Υποθέτωντας ότι $I_x \neq 0$, η (5.7) μπορεί να γραφτεί και σαν

$$u^n \frac{\|\nabla I\|}{I_x} = -s\beta Z - sU \tag{5.8}$$

5.4 Υπολογισμός διατακτικού βάθους

Προσεγγίζοντας το πεδίο κάθετης κίνησης με το πεδίο κάθετης ροής, μπορούμε να αντικαταστήσουμε το u^n με την αλγεβρική τιμή του πεδίου κάθετης ροής, η οποία όπως έχουμε ήδη δείξει είναι ίση με $-\frac{I_t}{\|\nabla I\|}$. Κατά συνέπεια, η εξίσωση (5.8) γίνεται

$$\begin{aligned} -\frac{I_t}{\|\nabla I\|} \frac{\|\nabla I\|}{I_x} &= -s\beta Z - sU \Leftrightarrow \\ -\frac{I_t}{I_x} &= -s\beta Z - sU \end{aligned} \quad (5.9)$$

Στο σημείο αυτό, παρατηρούμε ότι πετύχαμε τον υπολογισμό μιας συνάρτησης του βάθους της μορφής

$$r(Z) = AZ + B$$

όπου A και B άγνωστες σταθερές. Θεωρούμε τώρα ένα ζεύγος σημείων P_1, P_2 με συντεταγμένες (X_1, Y_1, Z_1) και (X_2, Y_2, Z_2) και αντίστοιχες εντάσεις I^1, I^2 . Η (5.9) δίνει για τα σημεία αυτά

$$-\frac{I_t^1}{I_x^1} = -s\beta Z_1 - sU \quad (5.10)$$

$$-\frac{I_t^2}{I_x^2} = -s\beta Z_2 - sU \quad (5.11)$$

Αφαιρώντας κατά μέλη τις (5.10) και (5.11) έχουμε

$$B_{1,2} \equiv -\frac{I_t^1}{I_x^1} + \frac{I_t^2}{I_x^2} = s\beta(Z_2 - Z_1) \quad (5.12)$$

Δεδομένου ότι $s, \beta > 0$, το πρόσημο της διαφοράς $Z_2 - Z_1$ μπορεί να υπολογιστεί από το πρόσημο του $B_{1,2}$. Έτσι έχουμε:

- Αν $B_{1,2} > 0$, τότε $Z_1 < Z_2$.
- Αν $B_{1,2} < 0$, τότε $Z_1 > Z_2$.
- Αν $B_{1,2} = 0$, τότε $Z_1 = Z_2$.

Η εξίσωση (5.12) επιτρέπει την σύγκριση της διαφοράς του βάθους δύο σημείων με την μέγιστη διαφορά βάθους στην σκηνή. Έστω π.χ. το σημείο m με το ελάχιστο βάθος Z_m στην σκηνή και το σημείο M με το μέγιστο βάθος Z_M . Τα σημεία αυτά είναι το πρώτο και το τελευταίο στοιχείο της ταξινομημένης (κατά αύξουσα τάξη) ακολουθίας των βαθών. Έστω τώρα τα σημεία i, j με αντίστοιχα βάθη Z_i, Z_j . Η (5.12) δίνει

$$B_{M,m} \equiv -\frac{I_t^M}{I_x^M} + \frac{I_t^m}{I_x^m} = s\beta(Z_m - Z_M) \quad (5.13)$$

και

$$B_{i,j} \equiv -\frac{I_t^i}{I_x^i} + \frac{I_t^j}{I_x^j} = s\beta(Z_j - Z_i) \quad (5.14)$$

5.5 Πολικός και λογαριθμικός πολικός μετασχηματισμός

Διαιρώντας κατά μέλη τις δύο προηγούμενες εξισώσεις, φτάνουμε στην

$$\frac{Z_i - Z_j}{Z_M - Z_m} = \frac{B_{i,j}}{B_{M,m}} \quad (5.15)$$

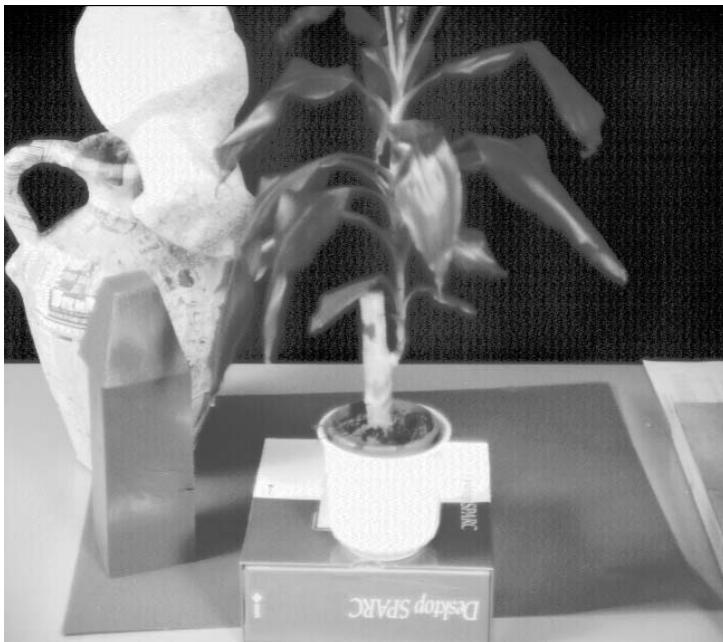
Στη συνέχεια, θα λέμε ότι η εξίσωση (5.15) δίνει την σχετική διαφορά βάθους για το ζεύγος των σημείων i, j , μια και καθορίζει το λόγο της διαφοράς των βαθών τους προς μια σταθερή ποσότητα (ίση με $Z_M - Z_m$ στην προκειμένη περίπτωση). Όταν το σημείο j καθορίζεται από τα συμφραζόμενα, θα μιλάμε για την σχετική διαφορά βάθους του σημείου i .

Δείξαμε επομένως πως είναι δυνατό να υπολογιστεί η σχέση του βάθους δύο σημείων, δεδομένου ενός fixating ζεύγους εικόνων. Η μέθοδος που παρουσιάστηκε βασίζεται στην μερική επίλυση του προβλήματος της αντιστοίχισης, μια και υπολογίζει την μια από τις δύο συνιστώσες της μετατόπισης των σημείων (την κάθετη ροή). Επίσης, δεν απαιτεί τον υπολογισμό του πεδίου οπτικής ροής ούτε κάποιου είδους καλιμπράρισμα του στερεοσκοπικού συστήματος. Τέλος, η μέθοδος είναι ποιοτική, μια και βασίζεται στο πρόσημο των ποσοτήτων $B_{..}$ και όχι στο μέτρο τους. Στην συνέχεια, θα περιγράψουμε μια υλοποίηση της μεθόδου για τον υπολογισμό διατακτικού βάθους. Πριν γίνει αυτό όμως, θα εισάγουμε κάποιους μετασχηματισμούς οι οποίοι θα χρησιμοποιηθούν στην υλοποίηση.

5.5 Πολικός και λογαριθμικός πολικός μετασχηματισμός

Η μέθοδος υπολογισμού διατακτικού βάθους που παρουσιάστηκε παραπάνω, βασίζεται στον υπολογισμό της κάθετης ροής. Δεδομένου ότι ο υπολογισμός αυτός υποθέτει ότι η χρησιμοποιούμενη ακολουθία εικόνων μπορεί να περιγραφεί από μια συνεχή συνάρτηση τριών μεταβλητών, η κίνηση μεταξύ δύο διαδοχικών εικόνων θα πρέπει να είναι μικρή. Κατά συνέπεια, τα disparities μεταξύ δύο διαδοχικών εικόνων θα πρέπει να είναι της τάξης μερικών pixels. Στην περίπτωση ενός fixating ζεύγους, η απαίτηση αυτή ικανοποιείται από τα σημεία που βρίσκονται κοντά στο κέντρο των εικόνων. Προχωρώντας όμως προς την περιφέρεια, τα disparities γίνονται πολύ μεγαλύτερα, με αποτέλεσμα να μην μπορεί να εφαρμοστεί η τεχνική υπολογισμού της κάθετης ροής που παρουσιάστηκε στο κεφ. 4. Αυτό φαίνεται καλύτερα με την βοήθεια των σχημάτων 5.4, 5.5. Οι εικόνες 5.4.(a) και 5.4.(b) είναι η αριστερή και δεξιά εικόνα αντίστοιχα ενός fixating ζεύγους, το οποίο καταγράφηκε από μια κεφαλή που κατασκευάστηκε στο Βασιλικό Ινστιτούτο Τεχνολογίας της Σουηδίας (KTH) [65]. Η εικόνα 5.5 απεικονίζει την απόλυτη τιμή της διαφοράς των εντάσεων των 5.4.(a), 5.4.(b) σε κάθε σημείο. Είναι φανερό ότι σημεία αρκετά μακριά από το κέντρο (π.χ. τα φύλλα του φυτού), παρουσιάζουν μεγάλα disparities. Για να αντιμετωπιστεί το πρόβλημα αυτό, χρησιμοποιείται ο λογαριθμικός πολικός μετασχηματισμός. Πρώτα όμως θα παρουσιάσουμε τον πολικό μετασχηματισμό.

5.5 Πολικός και λογαριθμικός πολικός μετασχηματισμός

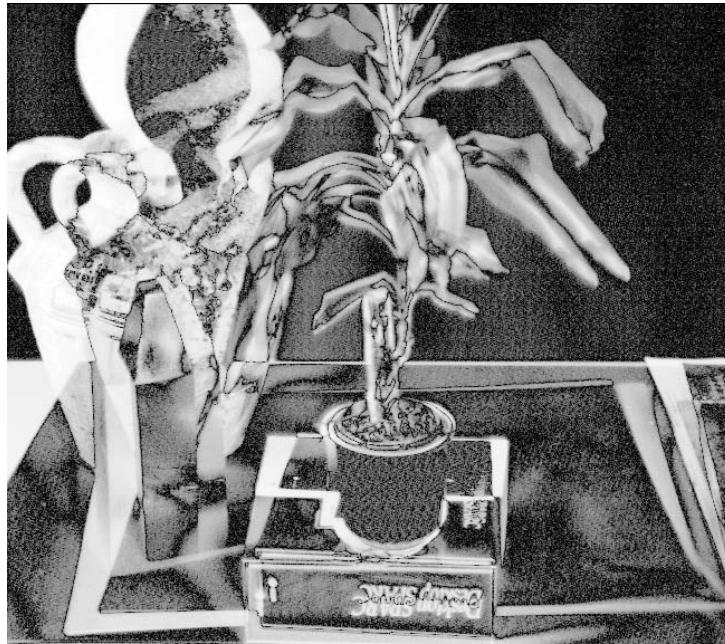


(a)



(b)

Σχήμα 5.4: Η (a) είναι η αριστερή εικόνα ενός fixating ζεύγους και η (b) η δεξιά.



Σχήμα 5.5: Διαφορές των εντάσεων για το ζεύγος των εικόνων του σχήματος 5.4.

5.5.1 Πολικός μετασχηματισμός

Έστω μια εικόνα $I(x, y)$. Το σημείο (x, y) απεικονίζεται με τον πολικό μετασχηματισμό στο σημείο (ρ, η) όπου

$$\begin{aligned}\rho &= \sqrt{x^2 + y^2} \\ \eta &= \arctan\left(\frac{y}{x}\right)\end{aligned}\tag{5.16}$$

Οι παράμετροι (x, y) εκφράζονται συναρτήσει των (ρ, η) από τις

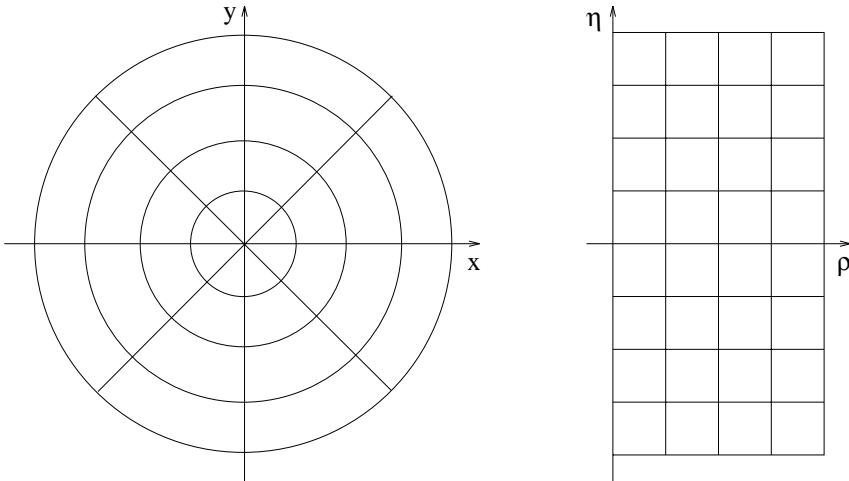
$$\begin{aligned}x &= \rho \cos(\eta) \\ y &= \rho \sin(\eta)\end{aligned}$$

Η απεικόνιση που ορίζουν οι εξισώσεις (5.16) φαίνεται στο σχήμα 5.6.

Στο σχήμα 5.7.(a) φαίνεται ο πολικός μετασχηματισμός της εικόνας 5.4.(a). Στο σχήμα 5.7.(b) φαίνεται ο ίδιος μετασχηματισμός πάνω στο επίπεδο της εικόνας.

5.5.2 Λογαριθμικός πολικός μετασχηματισμός

Ο λογαριθμικός πολικός μετασχηματισμός ορίζεται ανάλογα με τον πολικό μετασχηματισμό, με την διαφορά ότι αντί του ρ , χρησιμοποιείται ο φυσικός λογάριθμός του. Το σημείο (x, y)



Σχήμα 5.6: Πολικός μετασχηματισμός.

δηλαδή, απεικονίζεται στο σημείο (ξ, η) όπου

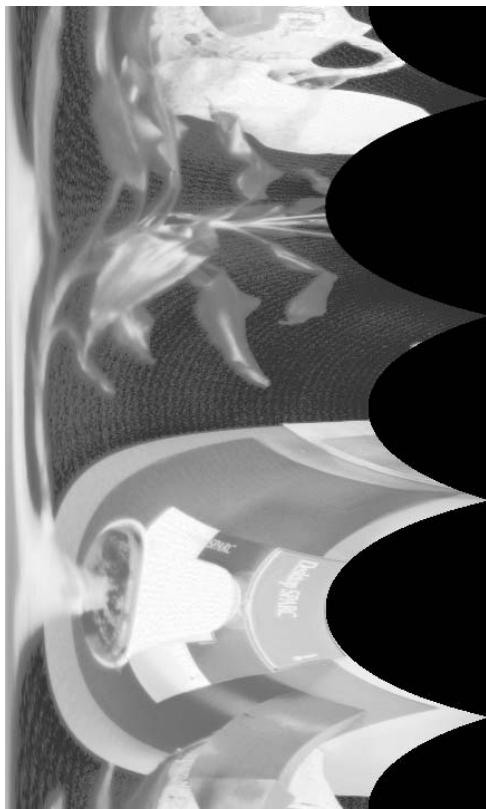
$$\begin{aligned}\xi &= \ln(\sqrt{x^2 + y^2}) \\ \eta &= \arctan\left(\frac{y}{x}\right)\end{aligned}\quad (5.17)$$

Επειδή ο λογάριθμος του 0 δεν ορίζεται, ο λογαριθμικός πολικός μετασχηματισμός δεν εφαρμόζεται στα σημεία μιας εικόνας που βρίσκονται στο εσωτερικό ενός μικρού κυκλικού δίσκου στο κέντρο της (βλ. γραμμοσκιασμένο δίσκο στο σχήμα 5.8). Οι παράμετροι (x, y) εκφράζονται συναρτήσει των (ξ, η) από τις

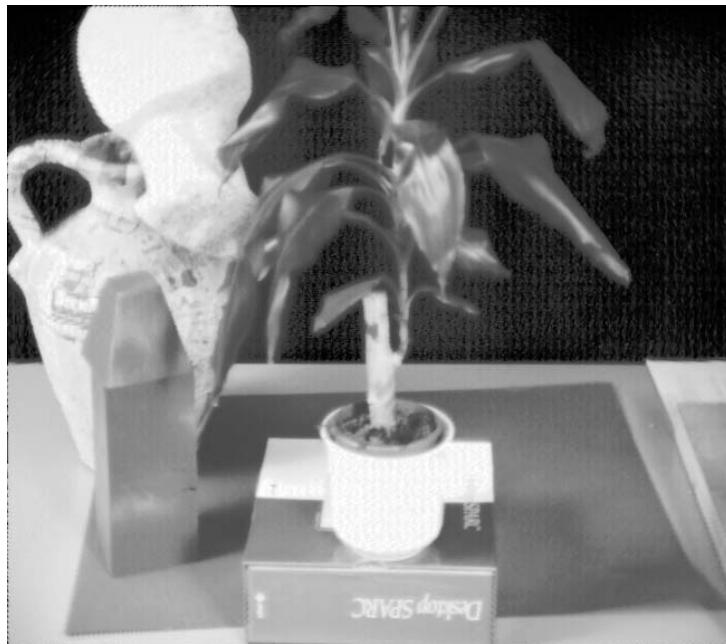
$$\begin{aligned}x &= e^\xi \cos(\eta) \\ y &= e^\xi \sin(\eta)\end{aligned}$$

Όπως αναφέρθηκε στο κεφ. 2, η κατανομή των φωτοευαίσθητων στοιχείων στον αμφιβληστροειδή πολλών ζώων είναι μεταβλητή. Ο λογαριθμικός πολικός μετασχηματισμός αποτελεί μια καλή προσέγγιση της κατανομής των στοιχείων αυτών. Τελευταία μάλιστα, έχουν κατασκευαστεί κάμερες με ανομοιόμορφη κατανομή των φωτοευαίσθητων στοιχείων τους, η οποία καθορίζεται από τον λογαριθμικό πολικό μετασχηματισμό. Η απεικόνιση που ορίζουν οι εξισώσεις (5.17) φαίνεται στο σχήμα 5.8.

Στο σχήμα 5.9.(a) φαίνεται ο λογαριθμικός πολικός μετασχηματισμός της εικόνας 5.4.(a). Στο σχήμα 5.9.(b) φαίνεται ο ίδιος μετασχηματισμός πάνω στο επίπεδο της εικόνας. Ο αριθμός των σημείων (x, y) τα οποία απεικονίζονται σε ένα σημείο (ξ_0, η_0) μέσω του λογαριθμικού πολικού μετασχηματισμού, εξαρτάται από την απόσταση του σημείου αυτού από την αρχή των αξόνων ξ, η . Αυτό γίνεται ευκολότερα κατανοητό με την βοήθεια των σχημάτων 5.8, 5.9.(b) στα οποία φαίνεται ότι το εμβαδόν των περιοχών του πεδίου (x, y) που απεικονίζονται σε ένα σημείο του πεδίου (ξ, η) , αυξάνεται με την απόσταση από το κέντρο της εικόνας. Ο

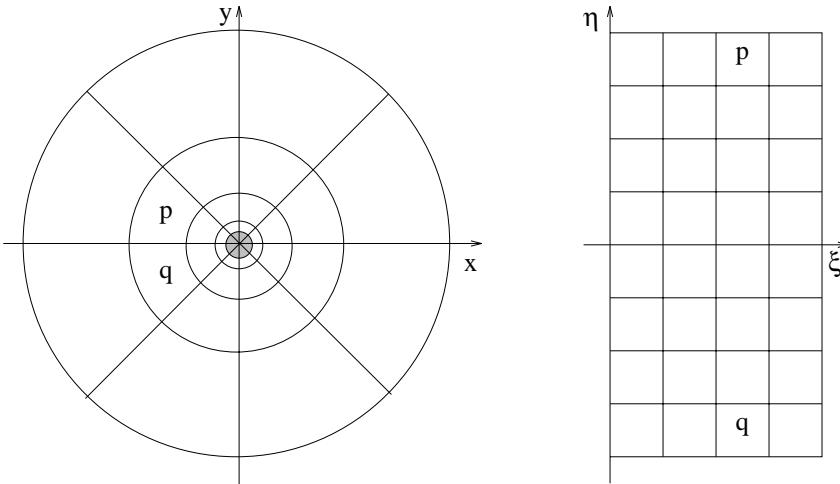


(a)



(b)

Σχήμα 5.7: Πολικός μετασχηματισμός (a) και απεικόνισή του στο επίπεδο της εικόνας (b).



Σχήμα 5.8: Λογαριθμικός πολικός μετασχηματισμός.

βασικός λόγος για αυτό είναι ότι η ακτίνα των ομόκεντρων κύκλων στο σχήμα 5.8, αυξάνεται εκθετικά. Έτσι, τα disparities που μετρώνται στο πεδίο (ξ, η) είναι πολύ μικρότερα από αυτά του πεδίου (x, y) . Το σχήμα 5.10 δείχνει τις απόλυτες διαφορές των εντάσεων για το ζεύγος του σχήματος 5.4, υπολογισμένες από τους λογαριθμικούς πολικούς μετασχηματισμούς των εικόνων αυτών. Με βάση τα παραπάνω, ο υπολογισμός της κάθετης ροής θα γίνει μέσω του λογαριθμικού πολικού μετασχηματισμού. Μπορούμε να αποδείξουμε (βλ. παράρτημα A) ότι οι μερικές παράγωγοι της έντασης ως προς x, y δίνονται συναρτήσει αυτών ως προς ξ, η από τις σχέσεις

$$\frac{\partial I}{\partial x} = \frac{\partial I}{\partial \xi} \frac{1}{\rho} \cos(\eta) - \frac{\partial I}{\partial \eta} \frac{1}{\rho} \sin(\eta) \quad (5.18)$$

$$\frac{\partial I}{\partial y} = \frac{\partial I}{\partial \xi} \frac{1}{\rho} \sin(\eta) + \frac{\partial I}{\partial \eta} \frac{1}{\rho} \cos(\eta) \quad (5.19)$$

5.6 Υλοποίηση

Περιγράφουμε τώρα μια υλοποίηση της μεθόδου υπολογισμού διατακτικού βάθους που αναπτύχθηκε σ' αυτό το κεφάλαιο.

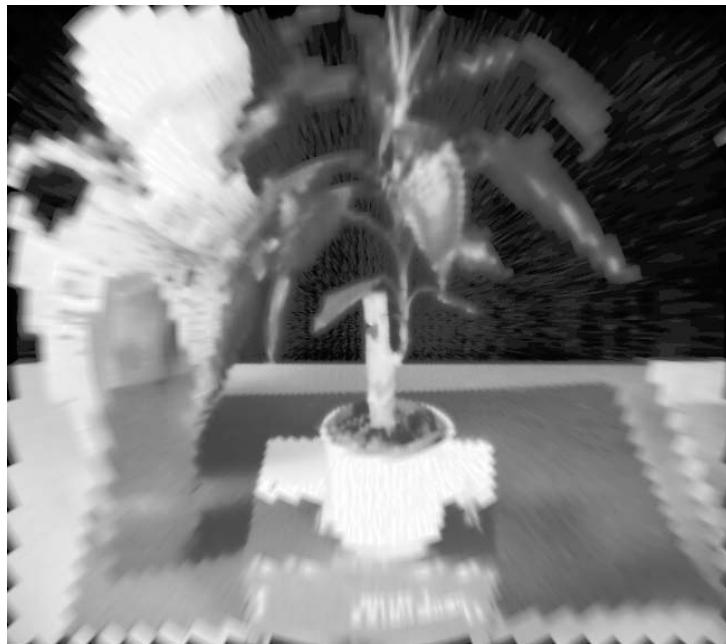
5.6.1 Υπολογισμός λογαριθμικού πολικού μετασχηματισμού

Ο υπολογισμός του διακριτού λογαριθμικού πολικού μετασχηματισμού γίνεται με την χρήση διγραμμικής παρεμβολής (bilinear interpolation). Ο μετασχηματισμός δεν υπολογίζεται για τα σημεία που βρίσκονται μέσα σε ένα δίσκο ακτίνας 5 pixels στο κέντρο των εικόνων.

5.6 Υλοποίηση



(a)



(b)

Σχήμα 5.9: Λογαριθμικός πολικός μετασχηματισμός (a) και απεικόνισή του στο επίπεδο της εικόνας (b).

5.6 Υλοποίηση



Σχήμα 5.10: Διαφορές των εντάσεων για το ζεύγος των εικόνων του σχήματος 5.4, υπολογισμένες στο πεδίο (ξ, η) (σε μεγέθυνση).

5.6 Υλοποίηση

Κάθε δείγμα στον όξονα ξ , απέχει από το κέντρο των εικόνων απόσταση ίση με το 1.05 της αντίστοιχης απόστασης του προηγούμενου δείγματος. Αυτό σημαίνει ότι η ακτίνα καθενός από τους κύκλους του σχήματος 5.8 είναι ίση με το 1.05 της ακτίνας του προηγουμένου. Το βήμα δειγματοληψίας για τον όξονα η είναι ίσο με 0.5° .

5.6.2 Υπολογισμός των μερικών παραγώγων

Πριν αναφερθούμε στην μέθοδο υπολογισμού των μερικών παραγώγων που χρησιμοποιήθηκε, είναι σημαντικό να επισημανθούν οι ιδιαιτερότητες που εμφανίζει η επεξεργασία λογαριθμικών πολικών εικόνων, λόγω της ανομοιόμορφης κατανομής των pixels [83]. Στη συνέχεια, αναφέρουμε τις ιδιαιτερότητες αυτές και περιγράφουμε πως αντιμετωπίστηκαν στην συγκεκριμένη υλοποίηση.

- Κάποια pixels τα οποία είναι γειτονικά στην επιφάνεια του αισθητήρα δεν είναι γειτονικά όταν τα pixels της εικόνας αποθηκευτούν σε ένα πίνακα. Π.χ. τα pixels p , q τα οποία είναι γειτονικά στο αριστερό τμήμα του σχήματος 5.8, δεν είναι γειτονικά στο δεξιό. Αυτό συμβαίνει με όλα τα ζεύγη pixels τα οποία αντιστοιχούν στην μέγιστη και ελάχιστη τιμή του η , για δεδομένη τιμή του ξ . Μια και ο αριθμός των pixels αυτών είναι πολύ μικρός συγκρινόμενος με τον συνολικό αριθμό pixels της εικόνας, το πρόβλημα αυτό αντιμετωπίστηκε απαγορεύοντας κάθε τοπική επεξεργασία με κέντρο κάποιο pixel της πρώτης ή τελευταίας γραμμής του πίνακα. Μια λύση που θα εξάλειφε τελείως το πρόβλημα θα ήταν να αυξήσουμε την κατακόρυφη διάσταση του πίνακα κατά 2 και να αντιγράψουμε την τελευταία γραμμή πάνω από την πρώτη καθώς και την πρώτη κάτω από την τελευταία. Στην συνέχεια, κάθε τοπική επεξεργασία του πίνακα αυτού θα γίνεται με κέντρο μόνο τα pixels που δεν ανήκουν σε κάποια από τις 2 γραμμές που προστέθηκαν.
- Κάποια από τα pixels του επιπέδου ξ , η δεν αντιστοιχούν σε pixels πάνω στην επιφάνεια του αισθητήρα. Αυτό φαίνεται καλύτερα στο σχήμα 5.7.(a), στο οποίο οι 5 ελλειπτικές περιοχές μαύρου χρώματος στο δεξιό μέρος απαρτίζονται από pixels τα οποία αντιστοιχούν σε pixels εκτός του παραλληλογράμμου που ορίζει η εικόνα του σχήματος 5.7.(b). Στην δική μας υλοποίηση τέτοιο πρόβλημα δεν υφίσταται μιας και όλοι οι υπολογισμοί περιορίζονται στο εσωτερικό ενός κυκλικού δίσκου, ο οποίος μετασχηματίζεται σε ένα παραλληλόγραμμο από τον λογαριθμικό πολικό μετασχηματισμό.
- Το εμβαδόν που καλύπτει ένα pixel στην επιφάνεια του αισθητήρα εξαρτάται από την θέση του pixel ως προς το κέντρο του τελευταίου, με άμεση συνέπεια κάθε τοπική επεξεργασία της εικόνας να πρέπει να λαμβάνει υπ' όψην της το μέγεθος των αντίστοιχων pixels. Στην δική μας περίπτωση όμως, το χαρακτηριστικό αυτό δεν αποτελεί πρόβλημα

5.6 Υλοποίηση

αλλά αντιθέτως είναι επιθυμητό, μια και όπως είδαμε παραπάνω επιτρέπει την μείωση των disparities στο επίπεδο ξ, η .

Περιγράφουμε τώρα την μέθοδο υπολογισμού των μερικών παραγώγων. Όπως αποδεικνύεται στο παράρτημα A, η μερική παράγωγος ως προς τον χρόνο είναι η ίδια, ανεξάρτητα από το αν υπολογίζεται στο πεδίο x, y ή στο ξ, η . Έτσι, η I_t υπολογίζεται από τον μέσο όρο των διαφορών των pixels των λογαριθμικών πολικών εικόνων σε μια 3×3 γειτονιά. Για τον υπολογισμό των χωρικών μερικών παραγώγων I_x, I_y , χρησιμοποιείται η μέθοδος του Canny [15], η οποία περιγράφεται περιληπτικά στην συνέχεια. Σημειώνεται πάντως ότι τα κατά πολύ απλούστερα φίλτρα Sobel [33] σε συνδυασμό με Gaussian smoothing, έδωσαν καλά αποτελέσματα.

Η μέθοδος ανίχνευσης ακμών του Canny, ανήκει στην γενική κατηγορία μεθόδων ανίχνευσης ακμών με την χρήση φύλτρων παραγώγισης. Η στρατηγική που ακολουθείται από τις μεθόδους της κατηγορίας αυτής, συνίσταται στην συνέλιξη του προς εξέταση σήματος με το κατάλληλο φίλτρο και στην συνέχεια εύρεση ακμών στα σημεία που το αποτέλεσμα της συνέλιξης παρουσιάζει τοπικό ακρότατο.

Η ανάλυση Canny στοχεύει στον καθορισμό, σύμφωνα με κάποια κριτήρια, ενός βέλτιστου φύλτρου για την ανίχνευση ακμών, αρχικά για σήματα μιας διάστασης. Τα κριτήρια αυτά είναι:

1. Καλή ανίχνευση

Το επιθυμητό φίλτρο πρέπει να ανιχνεύει τις πραγματικές ακμές και να μην δίνει λανθασμένες αποκρίσεις.

2. Ακριβής ανίχνευση

Τα σημεία που ανιχνεύονται σαν ακμές θα πρέπει να είναι όσο το δυνατόν πλησιέστερα στην πραγματική θέση των ακμών, παρά την πιθανή ύπαρξη θορύβου.

3. Μονοδική απόκριση για κάθε ακμή

Κάθε ακμή θα πρέπει να ανιχνεύεται μόνο μία φορά. Δεν θα πρέπει δηλαδή να υπάρχουν πολλαπλές αποκρίσεις.

Ο Canny δίνει μια μαθηματική έκφραση των παραπανών κριτηρίων και χρησιμοποιώντας αριθμητική βελτιστοποίηση, καταλήγει σε βέλτιστα φίλτρα για την ανίχνευση διαφόρων τύπων ακμών. Στην περίπτωση που οι ακμές είναι τύπου βήματος (step edges) και ο θόρυβος είναι προσθετικός Gaussian, το βέλτιστο φίλτρο στο οποίο καταλήγει ο Canny προσεγγίζεται καλά από την πρώτη παράγωγο της Gaussian συνάρτησης

$$g(x) = -\frac{x}{\sigma^3 \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

5.6 Υλοποίηση

Το σείναι μια παράμετρος που καθορίζει το πάχος των ακμών που εντοπίζονται. Στην συγκεκριμένη υλοποίηση έχει τιμή 1. Ένα ενδιάμεσο αποτέλεσμα που παρέχει η μέθοδος του Canny είναι οι χωρικές μερικές παράγωγοι της έντασης, οι οποίες χρησιμοποιούνται εδώ για τον υπολογισμό της κάθετης ροής.

Ο υπολογισμός των I_x, I_y γίνεται μέσω του υπολογισμού των I_ξ, I_η και των εξισώσεων (5.18), (5.19). Ο λόγος που οι I_x, I_y δεν υπολογίζονται από ευθείας στο πεδίο x, y , είναι ότι μια εικόνα στο πεδίο ξ, η είναι μικρότερη από την αντίστοιχη στο πεδίο x, y , και κατά συνέπεια οι υπολογισμοί απαιτούν λιγότερο χρόνο. Οι εικόνες π.χ. που χρησιμοποιήθηκαν, έχουν μέγεθος 650×574 ενώ οι λογαριθμικοί πολικοί μετασχηματισμοί τους είναι 93×720 , περίπου 5.5 φορές μικρότεροι όσον αφορά τον αριθμό των pixels.

5.6.3 Διατακτική αναπαράσταση του βάθους

Από την ανάλυση του 5.4 προκύπτει ότι δεν μπορούμε να αποφανθούμε για την σχέση, ως προς το βάθος, ενός σημείου με άλλα σημεία στις ακόλουθες περιπτώσεις:

- Το σημείο βρίσκεται “μακριά” από το κέντρο της εικόνας.
Στην περίπτωση αυτή, η προσέγγιση της προοπτικής προβολής με την ασθενή προοπτική δεν είναι ακριβής.
- Το μέτρο της κλίσης είναι “μικρό”.
Στην περίπτωση αυτή, δεν ισχύει η προσέγγιση του πεδίου κάθετης κίνησης από το πεδίο κάθετης ροής.
- Η I_x είναι μηδέν.
Στην περίπτωση αυτή, η κλίση είναι κάθετη στον άξονα x και η αλγεβρική τιμή του πεδίου κάθετης κίνησης είναι ίση με μηδέν.

Στην συγκεκριμένη υλοποίηση, ασχοληθήκαμε με τον υπολογισμό διατακτικού βάθους σημείων τα οποία

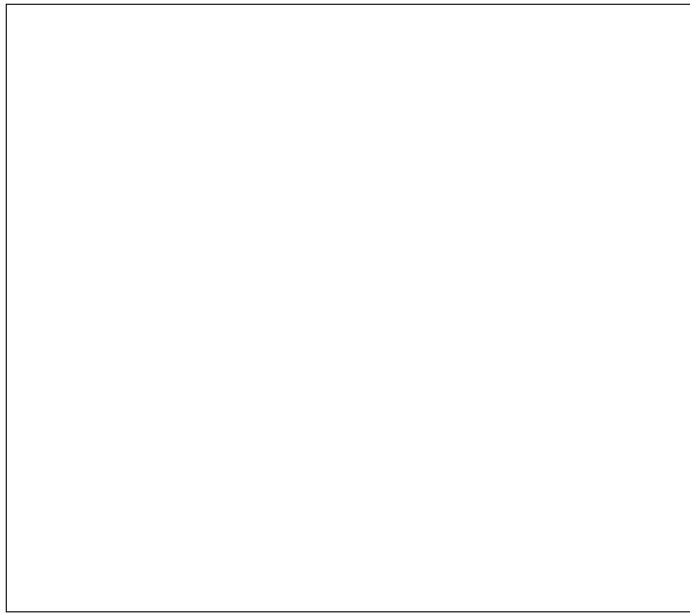
- Βρίσκονται στο εσωτερικό ενός κυκλικού δίσκου ακτίνας 200 pixels και κέντρου ίδιου με το κέντρο της εικόνας.
- Έχουν κλίση της οποίας το μέτρο υπερβαίνει κάποιο κατώφλι.
- Έχουν κλίση τέτοια ώστε η γωνία μεταξύ αυτής και του άξονα x είναι το πολύ 82° ή τουλάχιστον 98° .

Για κάθε ζεύγος σημείων που ικανοποιούν τις παραπάνω προϋποθέσεις, η εξίσωση (5.12) δίνει την σχέση μεταξύ του βάθους των. Τα σημεία αυτά ταξινομούνται κατά αύξουσα

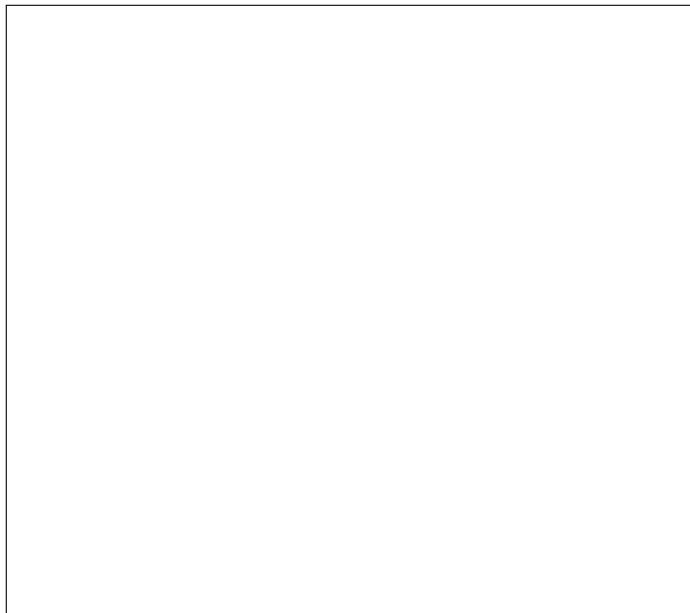
5.6 Υλοποίηση

σειρά ως προς το βάθος, με τον αλγόριθμο *QuickSort*. Το σχήμα 5.11.(a) δείχνει σε μορφή εικόνας την διατακτική αναπαράσταση του βάθους που υπολογίζεται από τις εικόνες του σχήματος 5.4. Οι περιοχές με κόκκινο χρώμα, αντιστοιχούν σε σημεία στα οποία δεν είναι δυνατός ο υπολογισμός διατακτικού βάθους. Για τα υπόλοιπα σημεία στο εσωτερικό του κύκλου, η (5.15) για $j = m$ δίνει την σχέση της απόστασής τους από το κοντινότερο στον παρατηρητή σημείο, με την μέγιστη διαφορά αποστάσεων που υπάρχει στην σκηνή. Αντιστοιχώντας στο κοντινότερο σημείο το άσπρο χρώμα και στο μακρύτερο το μαύρο, οι ενδιάμεσες διαβαθμίσεις του γκρίζου κωδικοποιούν τις αποστάσεις των υπόλοιπων σημείων της σκηνής. Για να διευκολυνθεί η κατανόηση του σχήματος 5.11.(a), το σχήμα 5.11.(b) δείχνει ένα κύκλο ακτίνας 200 pixels που έχει προστεθεί στην αριστερή εικόνα του σχήματος 5.4. Ο αριθμός των pixels στα οποία είναι δυνατόν να υπολογιστεί διατακτικό βάθος είναι περίπου το 25% των συνόλου των pixels που περιέχονται μέσα στον κύκλο.

5.6 Υλοποίηση



(a)



(b)

Σχήμα 5.11: Το (a) δείχνει το διατακτικό βάθος για το ζεύγος του σχήματος 5.4 και το (b) τον κυκλικό δίσκο στον οποίο υπολογίζεται το διατακτικό βάθος.

Κεφάλαιο 6

Περιγραφή των σχήματος των επιφανειών

6.1 Εισαγωγή

Στο κεφάλαιο αυτό θα παρουσιάσουμε μια μέθοδο για την τοπική περιγραφή του σχήματος (shape) των επιφανειών που συνθέτουν μια σκηνή. Η περιγραφή αυτή παρέχει έναν ποιοτικό χαρακτηρισμό των επιφανειών με βάση την μορφή τους στην γειτονιά κάποιου σημείου. Η μέθοδος θα στηριχθεί στην σχετική διαφορά βάθους για ζεύγη σημείων, ο υπολογισμός της οποίας συζητήθηκε στο κεφ. 5. Στο σημείο αυτό πρέπει να δοθεί έμφαση στο γεγονός ότι ο υπολογισμός μιας τοπικής περιγραφής του σχήματος δεν συνιστά την μόνη δυνατή χρήση μη μετρικού βάθους. Η παρουσίαση που θα ακολουθήσει, στοχεύει να δικαιολογήσει τους ισχυρισμούς σχετικά με την χρησιμότητα του μη μετρικού βάθους και να αποτελέσει το έναυσμα για την ανάπτυξη πολυπλοκότερων μεθόδων βασισμένων σ' αυτό.

Πριν ασχοληθούμε λεπτομερώς με τη μέθοδο αυτή, θα παρουσιάσουμε περιληπτικά κάποιες έννοιες από την θεωρία επιφανειών του Gauss, οι οποίες είναι απαραίτητες για την κατανόησή της. Για μια εκτενέστερη παρουσίαση, ο αναγνώστης μπορεί να ανατρέξει σε κάποιο βιβλίο διαφορικής γεωμετρίας [24].

6.2 Καμπυλότητα καμπυλών και επιφανειών

Εισάγουμε αρχικά την έννοια της καμπυλότητας για μια καμπύλη στο χώρο (space curve). Μια καμπύλη στο χώρο ορίζεται σαν μια συνάρτηση $r(s)$ από ένα διάστημα του \mathcal{R} στον \mathcal{R}^3 . Η παράμετρος s είναι η απόσταση που έχει διανυθεί κατά μήκος της καμπύλης, αρχίζοντας από κάποιο σημείο της. Έστω t το εφαπτόμενο διάνυσμα της καμπύλης στο σημείο P , το οποίο

6.2 Καμπυλότητα καμπυλών και επιφανειών

είναι ίσο με την τιμή της παραγώγου της r ως προς s στο P . Δεδομένου ότι για την Ευκλείδεια απόσταση ds μεταξύ δύο πολύ κοντινών σημείων $ds^2 = dr \cdot dr$, το t είναι μοναδιαίο, δηλαδή $t \cdot t = 1$. Παραγωγίζοντας την τελευταία σχέση, έχουμε ότι $t \cdot \hat{t} = 0$, δηλαδή το t είναι κάθετο στο \hat{t} , την παράγωγό του ως προς s στο P . Αν n είναι το μοναδιαίο διάνυσμα στην κατεύθυνση του \hat{t} , τότε υπάρχει κάποιος πραγματικός αριθμός κ , τέτοιος ώστε

$$\hat{t} = \kappa n$$

Ο κ ονομάζεται *καμπυλότητα (curvature)* στο σημείο P της καμπύλης r . Το n ονομάζεται *πρώτη κάθετος (principal normal)* της καμπύλης και το επίπεδο των t, n εγγύτατο επίπεδο (*osculating plane*) της καμπύλης. Το εξωτερικό γινόμενο $t \times n$, ορίζει το διάνυσμα b , το οποίο είναι κάθετο στο εγγύτατο επίπεδο. Το b ονομάζεται *δεύτερη κάθετος (binormal)* της καμπύλης. Έτσι, σε κάθε σημείο της καμπύλης, τα διανύσματα t, n, b ορίζουν ένα τοπικό σύστημα συντεταγμένων, γνωστό ως συνοδεύον τρίακμο της καμπύλης. Οι σχέσεις μεταξύ των διανυσμάτων t, n, b και των παραγώγων τους, δίνονται από τους τύπους Serret - Frenet.

Η έννοια της καμπυλότητας μπορεί να γίνει ευκολότερα κατανοητή στην περίπτωση μιας επίπεδης καμπύλης (*plane curve*), στην οποία η r είναι μια απεικόνιση του \mathcal{R} στο \mathcal{R}^2 . Στην περίπτωση αυτή, η καμπυλότητα είναι ίση με το αντίστροφο της ακτίνας καμπυλότητας (*radius of curvature*), η οποία είναι η ακτίνα του κύκλου γύρω από τον οποίο καμπυλώνεται η r και εφάπτεται σ' αυτόν στο σημείο P . Η έννοια της καμπυλότητας μιας καμπύλης, θα χρησιμοποιηθεί για τον ορισμό της καμπυλότητας μιας επιφάνειας, η οποία περιγράφεται στην συνέχεια.

Μια *επιφάνεια (surface patch)* ορίζεται σαν μια συνάρτηση $r(x, y)$ από ένα ανοιχτό σύνολο του \mathcal{R}^2 στο \mathcal{R}^3 . Το σχήμα μιας επιφάνειας καθορίζεται από δύο δευτεροβάθμιες εξισώσεις, γνωστές σαν *θεμελιώδεις μορφές (fundamental forms)*. Οι μερικές παράγωγοι r_x και r_y της r ως προς x, y αντίστοιχα, σε ένα σημείο P , ορίζουν το εφαπτόμενο επίπεδο $T_P(r)$ της r στο P . Η πρώτη θεμελιώδης μορφή καθορίζει το μήκος ενός διανύσματος που ανήκει σ' αυτό το επίπεδο. Το εξωτερικό γινόμενο των r_x και r_y ορίζει μια κατεύθυνση ζ , κάθετη στο εφαπτόμενο επίπεδο $T_P(r)$ και άρα και στην r . Κάθε κατεύθυνση ξ στο επίπεδο $T_P(r)$, ορίζει ένα επίπεδο κάθετο στην r , το οποίο περιέχει τα ξ, ζ . Η τομή αυτού του επιπέδου με την r , ορίζει μια επίπεδη καμπύλη. Η δεύτερη θεμελιώδης μορφή δίνει μια έκφραση για την καμπυλότητα της καμπύλης αυτής στο P . Η δεύτερη θεμελιώδης μορφή, παρουσιάζει ένα μοναδικό μέγιστο κ_1 και ένα μοναδικό ελάχιστο κ_2 , κατά μήκος δύο κάθετων κατευθύνσεων e_1 και e_2 . Αυτές οι κατευθύνσεις ονομάζονται *κύριες κατευθύνσεις (principal directions)*, ενώ οι κ_1, κ_2 ονομάζονται *κύριες καμπυλότητες (principal curvatures)*. Η καμπυλότητα μιας κάθετης τομής κατά μήκος μιας κατεύθυνσης διαφορετικής από τις κ_1, κ_2 , υπολογίζεται σαν γραμμικός συνδυασμός των τελευταίων, μέσω του τύπου του Euler:

$$\kappa_\phi = \kappa_1 \cos^2 \phi + \kappa_2 \sin^2 \phi$$

6.2 Καμπυλότητα καμπυλών και επιφανειών

όπου ϕ η γωνία μεταξύ της νέας διεύθυνσης και της e_1 . Κατά συνέπεια, η γνώση των κύριων καμπυλοτήτων, επιτρέπει πλήρη τοπικό χαρακτηρισμό του σχήματος της r . Διαισθητικά, η καμπυλότητα μιας κάθετης τομής περιγράφει την μορφή που έχει η επιφάνεια όταν απομακρυνόμαστε από ένα σημείο ακολουθώντας σταθερή κατεύθυνση. Οι κύριες καμπυλότητες αποτελούν εγγενή χαρακτηριστικά της επιφάνειας και κατά συνέπεια είναι ανεξάρτητες από το σύστημα αναφοράς που χρησιμοποιείται για τόν ορισμό της. Ο χαρακτηρισμός του σχήματος που παρέχουν είναι ποσοτικός, άρα περισσότερο ισχυρός από κάποιο ποιοτικό χαρακτηρισμό ανάλογο με αυτούς που φαίνεται να χρησιμοποιεί το ανθρώπινο σύστημα όρασης [75].

Ένα ποιοτικό χαρακτηρισμό του σχήματος παρέχει το πρόσημο της Gaussian καμπυλότητας K , η οποία ορίζεται σαν το γινόμενο των κύριων καμπυλοτήτων κ_1, κ_2 . Όταν η Gaussian καμπυλότητα είναι θετική, τότε η επιφάνεια καμπυλώνεται το ίδιο (είτε προς τα “πάνω” είτε προς τα “κάτω”) και στις δύο κύριες κατευθύνσεις. Η επιφάνεια τότε λέγεται ελλειπτική, και το σχήμα της είναι τοπικά κοίλο ή κυρτό, ανάλογα με το σημείο παρατήρησής της. Ένα παράδειγμα ελλειπτικής επιφάνειας είναι αυτή μιας σφαίρας. Όταν η Gaussian καμπυλότητα είναι αρνητική, τότε η επιφάνεια καμπυλώνεται με διαφορετικό τρόπο στις δύο κύριες κατευθύνσεις. Η επιφάνεια τότε λέγεται υπερβολική και το σχήμα της θυμίζει αυτό μιας σέλλας. Τέλος, όταν Gaussian καμπυλότητα είναι μηδέν, τότε η επιφάνεια μπορεί να είναι είτε επίπεδη (όταν και οι δύο κύριες καμπυλότητες είναι μηδέν) είτε παραβολική (όταν μόνο μια κύρια καμπυλότητα είναι ίση με μηδέν). Η επιφάνεια τότε λέγεται αναπτύξιμη (*developable*). Παράδειγμα επίπεδης επιφάνειας είναι προφανώς η παρούσα σελίδα, ενώ παραβολική επιφάνεια είναι αυτή ενός κυλίνδρου, όπου κατά την κατεύθυνση που είναι παράλληλη με τον άξονα η καμπυλότητα είναι μηδέν, ενώ κατά την κάθετη προς τον άξονα κατεύθυνση είναι μη μηδενική. Το όνομα “αναπτύξιμες” για αυτές τις επιφάνειας προέρχεται από το γεγονός ότι αυτές αντιστοιχούν στις επιφάνειες που μπορούν να σχηματιστούν λυγίζοντας ένα κομμάτι χαρτί, χωρίς αυτό να σχιστεί ή να τσαλακωθεί. Το σχήμα 6.1 παρουσιάζει μερικά παραδείγματα επιφανειών με διάφορα πρόσημα Gaussian καμπυλότητας. Σύμφωνα με το [49] η Gaussian καμπυλότητα δίνεται από την

$$K = \frac{DD'' - (D')^2}{EG - F^2} \quad (6.1)$$

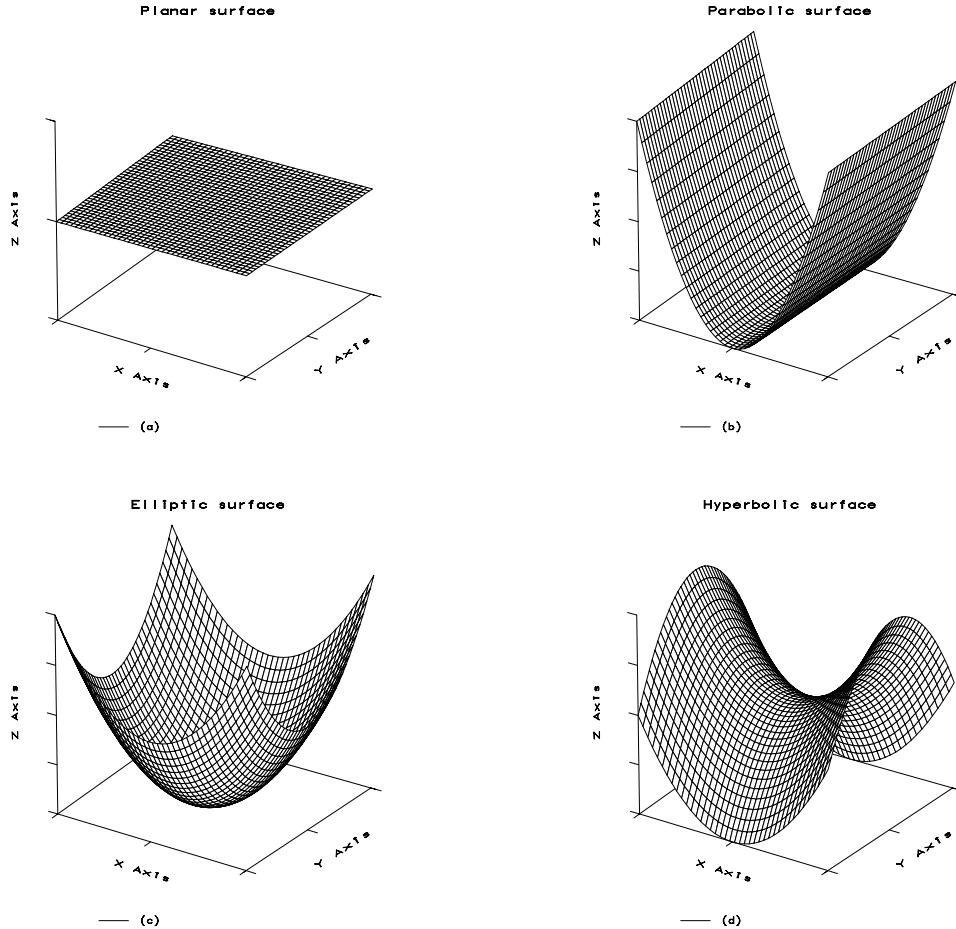
όπου

$$\begin{aligned} E &= \left(\frac{\partial r}{\partial x}\right)^2 \\ F &= \frac{\partial r}{\partial x} \cdot \frac{\partial r}{\partial y} \\ G &= \left(\frac{\partial r}{\partial y}\right)^2 \\ D &= -N \cdot \frac{\partial^2 r}{\partial x^2} \end{aligned}$$

6.2 Καμπυλότητα καμπυλών και επιφανειών

$$\begin{aligned} D' &= -N \cdot \frac{\partial^2 r}{\partial x \partial y} \\ D'' &= -N \cdot \frac{\partial^2 r}{\partial y^2} \end{aligned} \quad (6.2)$$

και $N = \frac{\partial r}{\partial x} \times \frac{\partial r}{\partial y}$.



Σχήμα 6.1: Επιφάνειες με διαφορετικό πρόσημο Gaussian καμπυλότητας:

- (a) $K = 0$, $\kappa_1^2 + \kappa_2^2 = 0$ και επίπεδη,
- (b) $K = 0$, $\kappa_1^2 + \kappa_2^2 \neq 0$ και παραβολική,
- (c) $K > 0$ και ελλειπτική και
- (d) $K < 0$ και υπερβολική.

Στη συνέχεια, θα κατατάξουμε τις επιφάνειες σε κατηγορίες με βάση το πρόσημο της Gaussian καμπυλότητας, όπως συνοψίζει ο πίνακας 6.1, καταλήγοντας έτσι σε κάποια μορφή αναπαράστασης του σχήματος.

6.3 Υπολογισμός του προσήμου της Gaussian καμπυλότητας

| Τιμή της Gaussian καμπυλότητας | Είδος της επιφάνειας |
|--------------------------------|-----------------------------------|
| Θετική | Ελλειπτική (κοίλη ή κυρτή) |
| Αρνητική | Υπερβολική |
| Μηδέν | Αναπτύξιμη (επίπεδη ή παραβολική) |

Πίνακας 6.1: Είδη επιφανειών με βάση το πρόσημο της Gaussian καμπυλότητας.

6.3 Υπολογισμός του προσήμου της Gaussian καμπυλότητας

Έστω (x, y) ένα σημείο μιας εικόνας και $Z(x, y)$ το βάθος του. Έστω επίσης το γράφημα της συνάρτησης $Z(x, y)$, το οποίο ορίζει την επιφάνεια $r : \mathcal{R}^2 \rightarrow \mathcal{R}^3$ με $r(x, y) = (x, y, Z(x, y))$. Όπως αποδεικνύεται στο παράρτημα B, το πρόσημο της Gaussian καμπυλότητας της επιφάνειας r είναι ίδιο με αυτό της ποσότητας L , όπου

$$L = \frac{\partial^2 Z}{\partial x^2} \frac{\partial^2 Z}{\partial y^2} - \left(\frac{\partial^2 Z}{\partial x \partial y} \right)^2 \quad (6.3)$$

Οι παράγωγοι δεύτερης τάξης που εμφανίζονται στην (6.3) μπορούν να υπολογιστούν σαν το άθροισμα των διαφορών των τιμών της συνάρτησης Z στη γειτονιά κάποιου pixel. Το [23] δίνει τις παρακάτω μάσκες, οι συνέλιξη των οποίων με την Z δίνει τις μερικές παραγώγους δεύτερης τάξης της τελευταίας:

$$\frac{\partial^2 Z}{\partial x^2} = \frac{1}{3} \begin{bmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \\ 1 & -2 & 1 \end{bmatrix} * Z \quad (6.4)$$

$$\frac{\partial^2 Z}{\partial y^2} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{bmatrix} * Z \quad (6.5)$$

$$\frac{\partial^2 Z}{\partial x \partial y} = \frac{1}{4} \begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix} * Z \quad (6.6)$$

Στο κεφάλαιο 5, δείξαμε πως μπορεί να υπολογιστεί η σχετική διαφορά βάθους για ζεύγη σημείων. Η εξίσωση (5.15) σε συνδυασμό με το γεγονός ότι το άθροισμα των στοιχείων στις παραπάνω μάσκες είναι ίσο με μηδέν, επιτρέπει τον υπολογισμό του προσήμου των παραγώγων δεύτερης τάξης του βάθους. Για παράδειγμα, η τιμή της $\frac{\partial^2 Z}{\partial x^2}$ στο σημείο (i, j) είναι ίση με

$$1/3(Z(i-1, j-1) - Z(i-1, j) + Z(i-1, j+1) - Z(i-1, j) + \\ Z(i, j-1) - Z(i, j) + Z(i, j+1) - Z(i, j) +$$

6.3 Υπολογισμός του προσήμου της Gaussian καμπυλότητας

$$\begin{aligned}
& Z(i+1, j-1) - Z(i+1, j) + Z(i+1, j+1) - Z(i+1, j)) = \\
& (B_{(i-1,j-1),(i-1,j)} + B_{(i-1,j+1),(i-1,j)} + B_{(i,j-1),(i,j)} + B_{(i,j+1),(i,j)} + \\
& B_{(i+1,j-1),(i+1,j)} + B_{(i+1,j+1),(i+1,j)}) \frac{Z_M - Z_m}{3B_{M,m}} \equiv \\
& C_{i,j}^x \frac{Z_M - Z_m}{3B_{M,m}}
\end{aligned} \tag{6.7}$$

όπου τα B_{\dots} , Z_M , Z_m είναι αυτά που ορίστηκαν στο κεφ. 5. Ανάλογα, η τιμή της $\frac{\partial^2 Z}{\partial y^2}$ στο σημείο (i, j) είναι ίση με

$$C_{i,j}^y \frac{Z_M - Z_m}{3B_{M,m}} \tag{6.8}$$

και της $\frac{\partial^2 Z}{\partial x \partial y}$

$$C_{i,j}^{x,y} \frac{Z_M - Z_m}{4B_{M,m}} \tag{6.9}$$

Κατά συνέπεια, η (6.3) δίνει

$$L = \left(\frac{C_{i,j}^x C_{i,j}^y}{9} - \frac{(C_{i,j}^{x,y})^2}{16} \right) \left(\frac{Z_M - Z_m}{B_{M,m}} \right)^2 \tag{6.10}$$

Έτσι, το πρόσημο της ποσότητας L και άρα και της Gaussian καμπυλότητας είναι ίδιο με αυτό της παράστασης

$$G \equiv \left(\frac{C_{i,j}^x C_{i,j}^y}{9} - \frac{(C_{i,j}^{x,y})^2}{16} \right) \tag{6.11}$$

η τιμή της οποίας μπορεί να υπολογιστεί ακριβώς.

Στο σημείο αυτό θα πρέπει να παρατηρήσουμε ότι ο υπολογισμός της σχετικής διαφοράς βάθους για ζεύγη σημείων, δεν απαιτεί κατ' ανάγκη τον υπολογισμό μιας συνάρτησης του βάθους που έχει την μορφή αυτής που υπολογίστηκε στο κεφ. 5. Για παράδειγμα, αν είναι γνωστό το σχετικό βάθος για ζεύγη σημείων, τότε είναι δυνατό να υπολογιστεί και η σχετική διαφορά βάθους γι' αυτά. Πράγματι, αν με r_{ij} συμβολίσουμε τον λόγο Z_i/Z_j τότε

$$\frac{Z_i - Z_j}{Z_M - Z_m} = \frac{1}{r_{Mi} - r_{mi}} - \frac{1}{r_{Mj} - r_{mj}}$$

Ανακεφαλαιώνοντας, δείξαμε ότι δεδομένης της σχετικής διαφοράς βάθους των σημείων, είναι δυνατό να χαρακτηρίσουμε τοπικά το σχήμα των απεικονιζόμενων επιφανειών. Οι χρησιμοποιούμενοι υπολογισμοί δεν απαιτούν την ύπαρξη κάποιου χάρτη βάθους (depth map), ο οποίος παρέχει την τιμή του βάθους για κάθε σημείο μιας διδιάστατης εικόνας, αλλά βασίζονται σε λιγότερες πληροφορίες για το βάθος, ο υπολογισμός των οποίων παρακάμπτει την επίλυση του προβλήματος structure from motion. Η μέθοδος που παρουσιάστηκε παρέχει ανάλογα αποτελέσματα με αυτήν που περιγράφεται στο [84], χρησιμοποιεί όμως διαφορετικές τεχνικές και δεν απαιτεί τον υπολογισμό ενός πυκνού (dense) πεδίου οπτικής ροής.

6.4 Υλοποίηση και αποτελέσματα

6.4.1 Συνθετικά δεδομένα

Στο σχήμα 6.2 φαίνεται το γράφημα της συνάρτησης

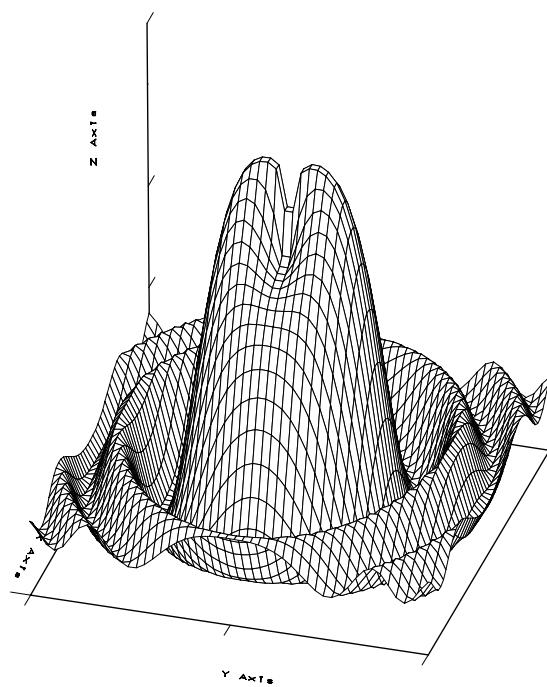
$$Z(x, y) = \frac{\sin(2x^2 + 3y^2)}{x^2 + y^2} \quad (6.12)$$

στο σύνολο $[-2, 2] \times [-2, 2]$, ιδωμένο από δύο διαφορετικές κατευθύνσεις. Δεδομένης της αναλυτικής έκφρασης (6.12) για το βάθος, το πρόσημο της Gaussian καμπυλότητας της επιφάνειας του σχ. 6.2 προκύπτει από την (6.3). Οι μερικές παράγωγοι δεύτερης τάξης της Z υπολογίζονται με τις 3×3 μάσκες που παρουσιάστηκαν παραπάνω. Το αποτέλεσμα του υπολογισμού αυτού παριστάνεται με την μορφή εικόνας στο σχήμα 6.3. Το άσπρο χρώμα αντιστοιχεί σε ελλειπτικές περιοχές, ενώ το μαύρο σε υπερβολικές. Η οριζόντια διάσταση της εικόνας αντιστοιχεί με τον X άξονα του σχ. 6.2.(b) ενώ η κάθετη με τον Y .

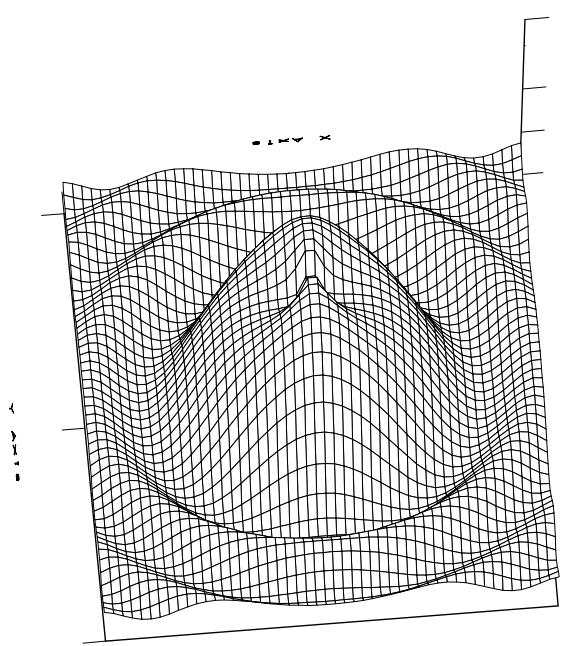
6.4.2 Πραγματικά δεδομένα

Η ίδια μέθοδος εφαρμόστηκε για τα δεδομένα που παρέχει η σχετική διαφορά του βάθους που υπολογίστηκε στο προηγούμενο κεφάλαιο. Μια και η τελευταία δεν μπορεί να υπολογιστεί για δύο οποιαδήποτε σημεία μιας εικόνας, ο υπολογισμός του προσήμου της Gaussian καμπυλότητας είναι δυνατός μόνο σε όσα σημεία είναι τέτοια ώστε η σχετική διαφορά του βάθους και των 9 σημείων της 3×3 γειτονιάς τους είναι γνωστή. Στο σχήμα 6.4 φαίνεται το ιστόγραμμα των G που υπολογίζονται από την (6.11) με βάση τις εικόνες του σχήματος 5.4. Η τιμή $G = 0$ εμφανίζεται σε 1563 σημεία, ενώ οι υπόλοιπες τιμές για το G εμφανίζονται σε μονοψήφιο αριθμό σημείων. Λόγω της μεγάλης διαφοράς μεταξύ της μέγιστης τιμής και των υπολοίπων, η κλίμακα στον κατακόρυφο άξονα του ιστογράμματος είναι λογαριθμική, έτσι ώστε να φαίνονται καλύτερα οι μικρές τιμές. Από το σχήμα αυτό είναι εύκολο να παρατηρήσουμε ότι εμφανίζονται πολλά σημεία με G πολύ κοντά στο μηδέν. Η συσσώρευση αυτή οφείλεται σε λάθη ακρίβειας κατά τον υπολογισμό τόσο της σχετικής διαφοράς βάθους όσο και της ποσότητας G . Έτσι, θα θεωρούμε ότι σημεία με $|G| < G_0$, όπου G_0 κάποιο κατώφλι, αντιστοιχούν σε τοπικά αναπτύξιμες επιφάνειες για τις οποίες $G = 0$. Η προσέγγιση αυτή έχει το μειονέκτημα ότι κάποιες επιφάνειες που τοπικά εμφανίζουν πολύ μικρή καμπυλότητα, είναι πιθανό να χαρακτηριστούν αναπτύξιμες.

Στο σχήμα 6.5.(a) φαίνεται το πρόσημο της Gaussian καμπυλότητας που υπολογίζεται από τις εικόνες του σχήματος 5.4 για $G_0 = 0.05$. Τα σημεία με κόκκινο χρώμα αντιστοιχούν σε αναπτύξιμες επιφάνειες, αυτά με πράσινο σε ελλειπτικές και αυτά με μπλε σε υπερβολικές. Για τα υπόλοιπα σημεία, δεν υπάρχει αρκετή πληροφορία ώστε να είναι δυνατός ο υπολογισμός του προσήμου της Gaussian καμπυλότητας.

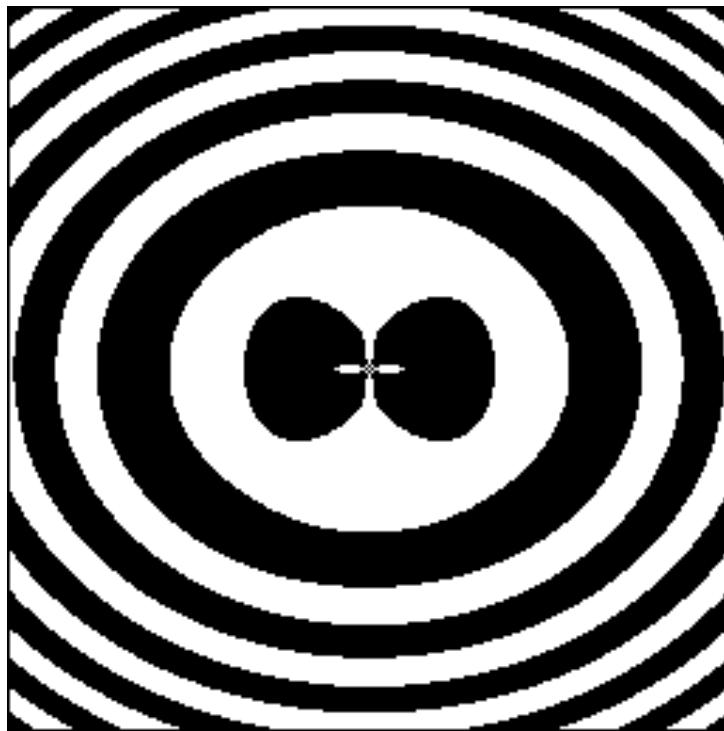


(a)

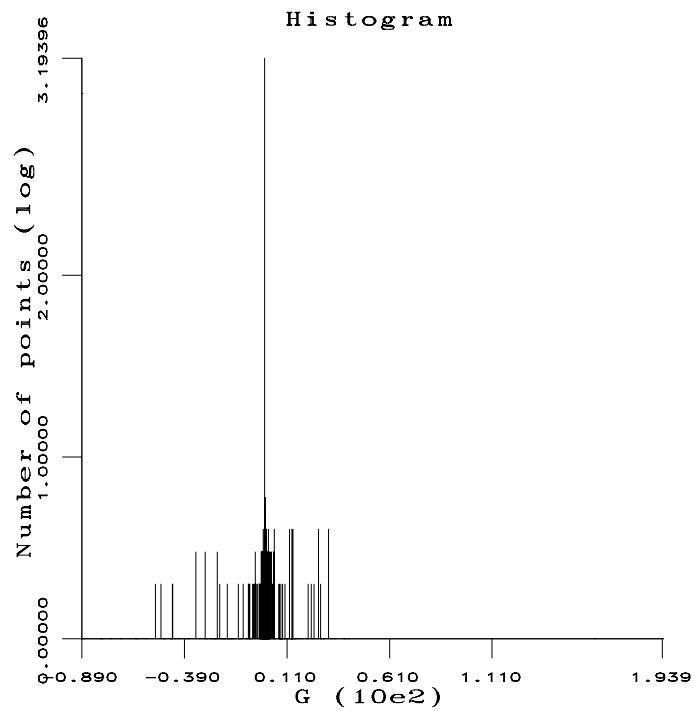


(b)

Σχήμα 6.2: Δύο όψεις της γραφικής παράστασης της συνάρτησης 6.12.

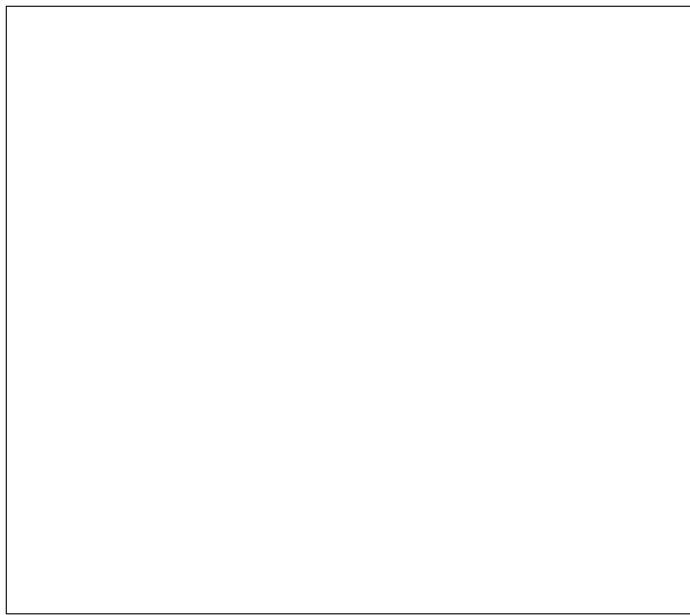


Σχήμα 6.3: Πρόσημο της Gaussian καμπυλότητας για την συνάρτηση 6.12.



Σχήμα 6.4: Ιστόγραμμα των G που υπολογίζονται με βάση τις εικόνες του σχήματος 5.4.

6.4 Υλοποίηση και αποτελέσματα



(a)



(b)

Σχήμα 6.5: Το (a) δείχνει το πρόσημο της Gaussian καμπυλότητας για τη σκηνή του σχήματος 5.4 και το (b) τον κυκλικό δίσκο στον οποίο γίνονται οι υπολογισμοί.

Κεφάλαιο 7

Επίλογος

Στην εργασία αυτή παρουσιάσαμε μια μέθοδο υπολογισμού της σχέσης των σημείων μιας σκηνής ως προς το βάθος καθώς και μια χρήση της σχετικής διαφοράς του βάθους των για τον υπολογισμό μιας απλής περιγραφής του σχήματος των αντικειμένων που περιέχονται στη σκηνή. Ο υπολογισμός του διατακτικού βάθους που παρουσιάστηκε, είναι τοπικός υπό την έννοια ότι αφορά μια μικρή περιοχή γύρω από κάποιο fixation point. Ένα ενδιαφέρον πρόβλημα επομένως, θα ήταν η μελέτη τρόπων συνδυασμού πολλών τέτοιων τοπικών αναπαραστάσεων που προκύπτουν από διαδοχικά fixations, με στόχο κάποια συνολικότερη αναπαράσταση. Το πρόβλημα αυτό μπορεί να θεωρηθεί σαν μια εκδοχή του προβλήματος σταθερής αντίληψης του κόσμου (*stable world perception problem*), το οποίο εξετάζει τον τρόπο με τον οποίο ο πραγματικός κόσμος γίνεται αντιληπτός σαν σταθερός και ενιαίος, παρά το γεγονός ότι ο παρατηρητής ή τα μάτια του βρίσκονται σε κίνηση σχετικά με το περιβάλλον. Ένα εξίσου ενδιαφέρον πρόβλημα είναι η αναζήτηση ιδιοτήτων του διατακτικού βάθους, οι οποίες είναι ανεξάρτητες από την σχετική θέση του παρατηρητή και της παρατηρούμενης σκηνής. Τέτοιες ιδιότητες θα μπορούσαν να χρησιμοποιηθούν για την ακριβέστερη περιγραφή απόψεων του περιβάλλοντος.

Σύμφωνα με την ορολογία του μοντέλου των Fermüller και Aloimonou το οποίο περιγράφεται στο κεφ. 2, η παρούσα εργασία παρουσίασε δύο οπτικές ικανότητες (αυτή του υπολογισμού διατακτικού βάθους και αυτή του ποιοτικού χαρακτηρισμού του σχήματος) και δύο αναπαραστάσεις, αυτές του διατακτικού βάθους και του ποιοτικού σχήματος. Άλλες οπτικές ικανότητες θα μπορούσαν να εκμεταλλευτούν πληροφορίες όπως το texture, οι σκιές (shading) κ.λπ. για να δώσουν επιπλέον περιγραφές του σχήματος, ο κατάλληλος συνδυασμός των οποίων θα διεκόλυνε την επίτευξη των εκάστοτε στόχων κάποιου συστήματος. Τα κίνητρα για την χρήση μιας διατακτικής αναπαράστασης του βάθους βρίσκονται αφ' ενός στα αποτελέσματα ψυχοφυσικών πειραμάτων σχετικών με την στερεοσκοπική όραση και αφ' ετέρου στην αδυναμία των παραδοσιακών, μετρικών μεθόδων να οδηγήσουν στην κατασκευή αυτόνομων συστημάτων, ικανών να λειτουργούν σε πραγματικά περιβάλλοντα.

Οι αλγόριθμοι που αναπτύχθηκαν ακολουθούν τις κατευθυντήριες γραμμές της θεωρίας της ποιοτικής ενεργούς όρασης, οι οποίες παρουσιάστηκαν στο κεφ. 2. Κύρια χαρακτηριστικά τους, είναι ότι στοχεύουν στην επίλυση πολύ συγκεκριμένων και απλών προβλημάτων, εκμεταλλεύονται γεωμετρικούς περιορισμούς που ισχύουν στο περιβάλλον και τέλος κάνουν χρήση ποιοτικών υπολογισμών. Η μόνη υπόθεση που κάνουν όσον αφορά τις ιδιότητες του περιβάλλοντος, είναι ότι το τελευταίο έχει αρκετό texture και έτσι επιτρέπει τον υπολογισμό της κλίσης των εντάσεων των εικόνων, και άρα και της κάθετης ροής. Η υπόθεση αυτή όμως, είναι περισσότερο ρεαλιστική από τις υποθέσεις που συνήθως χρησιμοποιούνται για την διεκόλυνση της ανάλυσης κίνησης. Με την αυστηρή έννοια του όρου, η ακολουθούμενη μεθοδολογία δεν είναι ενεργός, μια και δεν βασίζεται στην επιλογή τιμών για τις παραμέτρους του στερεοσκοπικού συστήματος. Δεδομένου όμως ότι προϋποθέτει την ύπαρξη ενός fixating ζεύγους εικόνων, το οποίο αποκτάται με κατάλληλους χειρισμούς των μηχανικών παραμέτρων του συστήματος, με μια γενικότερη θεώρηση της διαδικασίας απόκτησης των εικόνων και της επεξεργασίας τους είναι δυνατόν να χαρακτηριστεί ενεργός. Τέλος, τόσο ο υπολογισμός του πεδίου κάθετης ροής όσο και ο υπολογισμός του προσήμου της Gaussian καμπυλότητας, εμφανίζουν παραλληλισμό δεδομένων (data parallelism) με άμεση συνέπεια να μπορούν να παραλληλοποιηθούν πολύ εύκολα και να επιτύχουν καλούς χρόνους εκτέλεσης. Επίσης, ο υπολογισμός της διατακτικής αναπαράστασης του βάθους μπορεί να γίνει με την χρήση ενός παράλληλου αλγορίθμου ταξινόμησης. Στα πλαίσια της παρούσας εργασίας όμως, δεν έγινε κάποια παράλληλη υλοποίηση των τεχνικών που παρουσιάστηκαν, μια και η έμφαση δεν είχε δοθεί στην ταχύτητα των χρησιμοποιούμενων αλγορίθμων.

Στο σημείο αυτό, θα πρέπει να είναι σαφές ότι υπάρχει μια ολόκληρη οικογένεια περιγραφών του βάθους, ένα μέλος της οποίας είναι και το διατακτικό βάθος με το οποίο ασχοληθήκαμε. Το τελευταίο δεν είναι κατ' ανάγκη το καταλληλότερο για την κατανόηση του χώρου, κατά συνέπεια θα πρέπει να μελετηθούν και άλλες περιγραφές του βάθους και να κατηγοριοποιηθούν με γνώμονα την χρησιμότητα αλλά και την ευκολία υπολογισμού των. Μια και ο απότερος στόχος μας είναι η οπτική πλοήγηση, οι προσπάθειες θα πρέπει να εστιαστούν στην κατασκευή μιας αναπαράστασης του χώρου, ικανής να υποστηρίξει διάφορες ικανότητες πλοήγησης. Αναφορικά με το είδος μιας τέτοιας αναπαράστασης, τα αποτελέσματα των μελετών των ανθρώπινων ικανοτήτων αντίληψης του χώρου συνιστούν ότι θα πρέπει να είναι ποιοτική, μια και η αντίληψη των μετρικών χαρακτηριστικών του χώρου υπόκειται σε σημαντικά λάθη [67]. Μια ποιοτική αναπαράσταση του χώρου παρουσιάζει σημαντικά υπολογιστικά πλεονεκτήματα έναντι μιας μετρικής, μια και μπορεί ευκολότερα να χειριστεί ανακριβείς και ελλιπείς πληροφορίες και αναμένεται να βασίζεται σε απλούστερους αλγορίθμους για τον υπολογισμό της. Ήδη στο κεφάλαιο 3 έχει αναφερθεί ότι η αναπαράσταση του χώρου θα μπορούσε να χρησιμοποιεί ορόσημα, τα οποία επιτρέπουν την περιγραφή τμημάτων του χώρου μέσω χαρακτηριστικών οπτικών γνωρισμάτων του, καταργώντας την ανάγκη ύπαρξης καθολικών συστημάτων συντεταγμένων. Για να μπορέσει όμως ένα πραγματικό σύστημα να χρησιμοποιήσει τέτοια ορόσημα, θα πρέπει να έχει επιλέξει

κατάλληλες περιγραφές τους, να διαθέτει μεθόδους για τον υπολογισμό και το ταίριασμά τους και τέλος να διαθέτει κάποιας μορφής μνήμη για την αποθήκευση των περιγραφών. Εκτός από τις περιγραφές των οροσήμων, η μνήμη αυτή θα πρέπει να κωδικοποιεί τις χωρικές σχέσεις μεταξύ τους (spatial memory), επιτρέποντας έτσι σε αλγόριθμους προγραμματισμού της κίνησης να επιλέξουν διαδρομές μεταξύ διαφορετικών τοποθεσιών του χώρου. Η μνήμη θα πρέπει να είναι οργανωμένη κατά τέτοιο τρόπο ώστε να υποστηρίζει αποδοτικά διάφορες μορφές αναζήτησης, και επιπλέον να επιτρέπει την μεταβολή των περιεχομένων της από αλγόριθμους μάθησης οι οποίοι θα κωδικοποιούν την εμπειρία του συστήματος.

Μέχρι σήμερα, οι περισσότερες προσπάθειες έχουν ασχοληθεί με την κατανόηση του χώρου μόνο σε συμβολικό (symbolic) επίπεδο. Ο λόγος γι' αυτό είναι ότι μέχρι πρότινος, οι διαδικασίες αντίληψης και δράσης αποτελούσαν χωριστό πεδίο έρευνας από τις διαδικασίες σκέψης. Έτσι, η μελέτη επικεντρώθηκε στην κατανόηση κάτω από μεταβαλλόμενες συνθήκες, των χωρικών σχέσεων μεταξύ αντικειμένων που είχαν ήδη αναγνωριστεί. Ένα αυτόνομο σύστημα όμως, το οποίο βρίσκεται μέσα σε κάποιο περιβάλλον και έχει κάποιους στόχους, αλληλεπιδρά με τον περιβάλλοντα χώρο μέσω αντιλήψεων και ενεργειών. Τα αντικείμενα που περιβάλλονταν το σύστημα έχουν γι' αυτό μια από τις ακόλουθες σημασίες: Είτε πρόκειται για αντικείμενα που χρησιμεύουν σαν ορόσημα, είτε πρόκειται για αντικείμενα που πρέπει να αναχαιτιστούν, είτε τέλος για αντικείμενα που αντιπροσωπεύουν κάποιο κίνδυνο και πρέπει να αποφευχθούν. Επομένως, ένα μεγάλο μέρος της όποιας γνώσης του συστήματος για το χώρο, συνίσταται στην ικανότητά του να δίνει απαντήσεις σε προβλήματα προσδιορισμού της ταυτότητας (τι είναι αυτό που βλέπει) και εντοπισμού (που είναι κάποιο αντικείμενο ενδιαφέροντος) διαφόρων αντικειμένων. Οι λύσεις των προβλημάτων αυτών, απαιτούν κάποιες ικανότητες αναγνώρισης αντικειμένων.

Το πρόβλημα της αναγνώρισης τρισδιάστατων αντικειμένων περιλαμβάνει υποπροβλήματα όπως η κατασκευή μοντέλων για την αναπαράστασή τους, η συλλογή δεδομένων από τους αισθητήρες, η επεξεργασία των δεδομένων αυτών και τέλος το ταίριασμά τους με τα αποθηκευμένα μοντέλα. Το γεγονός ότι η όψη ενός αντικειμένου μεταβάλλεται σημαντικά όταν αλλάζει η σχετική του θέση ως προς τον παρατηρητή, αποτελεί τον κυριότερο λόγο για τον οποίο το πρόβλημα της επιλογής κατάλληλων μοντέλων, και κατ' επέκταση το πρόβλημα της αναγνώρισης τρισδιάστατων αντικειμένων, είναι πολύ δύσκολο. Η κατάσταση περιπλέκεται περισσότερο, δεδομένου ότι οι αλλαγές στο φωτισμό ενός αντικειμένου αλλά και η ύπαρξη αντικειμένων που παρεμβαίνουν μεταξύ του παρατηρητή και του παρατηρούμενου αντικείμενου (occlusion), μπορούν να επηρεάσουν δραματικά την όψη του τελευταίου. Οι προσεγγίσεις στο πρόβλημα της αναγνώρισης αντικειμένων που έχουν ακολουθηθεί μέχρι σήμερα, μπορούν να διακριθούν σε τρεις κατηγορίες, από τις οποίες οι δύο έχουν εισαχθεί σχετικά πρόσφατα.

Η πρώτη και πιο παλιά κατηγορία, βασίζεται σε τρισδιάστατα γεωμετρικά μοντέλα, τα οποία συντίθεται από απλά σχήματα, όπως πολύεδρα, σφαίρες κ.ά. Η αναγνώριση επιτυγ-

χάνεται σε δύο στάδια. Αρχικά, αντιστοιχούνται κάποια χαρακτηριστικά του μοντέλου με κάποια χαρακτηριστικά του παρατηρούμενου αντικειμένου και στη συνέχεια υπολογίζεται η θέση και ο προσανατολισμός του μοντέλου (pose) με βάση την αντιστοίχιση. Η θέση και ο προσανατολισμός του μοντέλου που προκύπτουν από το δεύτερο στάδιο, χρησιμοποιούνται για να επιβεβαιώσουν την ορθότητα της αντιστοίχισης, προβλέποντας την ύπαρξη επιπλέον χαρακτηριστικών και στην συνέχεια εύρεσή τους στο παρατηρούμενο αντικείμενο σύμφωνα με την πρόβλεψη. Οι προσεγγίσεις της κατηγορίας αυτής χαρακτηρίζονται σαν *object centered*. Το βασικότερο μειονέκτημά τους είναι η δυσκολία κατασκευής των μοντέλων, τόσο επειδή αυτά πρέπει να είναι πολύ λεπτομερή όσο και επειδή τα στοιχειώδη σχήματα που συνήθως χρησιμοποιούνται είναι ανεπαρκή για την περιγραφή φυσικών αντικειμένων (π.χ. δέντρα, βουνά κ.λπ.). Επιπλέον, το υπολογιστικό κόστος που απαιτεί η διαδικασία ταιριάσματος είναι πολύ μεγάλο.

Οι προσεγγίσεις της δεύτερης κατηγορίας αναπαριστούν ένα αντικείμενο μέσω ενός συνόλου από διδιάστατες όψεις του. Όπως και στην προηγούμενη προσέγγιση, η αναγνώριση γίνεται αρχικά με την αντιστοίχιση κάποιων χαρακτηριστικών μεταξύ του αντικειμένου και των αποθηκευμένων όψεων του και στην συνέχεια εύρεση ενός άμεσα υπολογιζόμενου μετασχηματισμού που “ευθυγραμμίζει” το μοντέλο με το αντικείμενο. Οι προσεγγίσεις της κατηγορίας είναι γνωστές σαν *viewer centered*. Βασικότερο πλεονέκτημά τους είναι ότι αφ' ενός εξασφαλίζουν την σταθερότητα των αναπαραστάσεων μετά από απλούς γεωμετρικούς μετασχηματισμούς και αφ' ετέρου ότι δεν απαιτούν την άμεση αναπαράσταση της τρισδιάστατης δομής των αντικειμένων. Πειραματικές μαρτυρίες που υποστηρίζουν μια τέτοια στρατηγική για την αναγνώριση αντικειμένων, έχουν προκύψει από την μελέτη του τρόπου με τον οποίο οι πίθηκοι μαθαίνουν να αναγνωρίζουν αντικείμενα που παρουσιάζονται σε μια οθόνη υπολογιστή [48]. Εκτός από τα πειραματικά, τα τελευταία χρόνια έχουν δημοσιευτεί και σημαντικά θεωρητικά αποτελέσματα για την κατηγορία αυτή: Οι Koenderink και van Doorn [43] έδειξαν ότι δύο εικόνες ενός συνόλου σημείων είναι αρκετές για τον υπολογισμό μιας αναπαράστασης που παραμένει σταθερή μετά από έναν οποιονδήποτε αφφινικό (και άρα *non-rigid*) μετασχηματισμό. Οι Ullman και Basri [81] έδειξαν ότι ένα τρισδιάστατο αντικείμενο μπορεί να παρασταθεί σαν γραμμικός συνδυασμός διδιάστατων όψεών του. Οι Clemens και Jacobs [18] μελέτησαν τις απαιτήσεις χρόνου και χώρου που έχει η αναγνώριση αντικειμένων μέσω διδιάστατων όψεων.

Η τρίτη κατηγορία περιλαμβάνει μεθόδους οι οποίες στοχεύουν στον υπολογισμό χαρακτηριστικών τα οποία είναι ανεξάρτητα από την σχετική θέση αντικειμένου - παρατηρητή καθώς και από τον φωτισμό [31]. Ένα τέτοιο χαρακτηριστικό είναι η Gaussian καμπυλότητα που χρησιμοποιήθηκε στο προηγούμενο κεφάλαιο. Η αναγνώριση αντικειμένων είναι δυνατή με απ'ευθείας αναζήτηση των αμετάβλητων χαρακτηριστικών τους στο σύνολο των αποθηκευμένων μοντέλων, εξαλείφοντας το στάδιο “ευθυγράμμισης” του μοντέλου με το αντικείμενο.

Επίλογος

Κάθε μια από τις παραπάνω προσεγγίσεις έχει τα δικά της υπέρ και κατά, χωρίς να είναι προφανές ποιά είναι η καλύτερη. Ανεξάρτητα από αυτό πάντως, φαίνεται ότι η ικανότητα αντίληψης του χώρου που διαθέτουν οι βιολογικοί οργανισμοί κάνει χρήση τόσο εγωκεντρικών όσο και γεωκεντρικών συστημάτων αναφοράς, με διάφορους συνδυασμούς. Τα πρώτα αναφέρονται σε συστήματα που έχουν κέντρο τον οργανισμό και κινούνται μαζί του. Τέτοια π.χ. είναι τα χρησιμοποιούμενα συστήματα στην περίπτωση χρήσης διδιάστατων όψεων των αντικειμένων ή αδρανειακών αισθητήρων. Από την άλλη πλευρά, η ικανότητα υπολογισμού χωρικών σχέσεων μεταξύ τμημάτων του περιβάλλοντος και κωδικοποίησή τους με τη μορφή κάποιου χάρτη, απαιτεί την ύπαρξη γεωκεντρικών συστημάτων αναφοράς, συστημάτων δηλαδή που είναι προσαρμοσμένα σε κάποια σταθερά σημεία του περιβάλλοντος. Τέτοια σημεία είναι τα ορόσημα που περιγράψαμε στο κεφ. 3. Είναι πιθανό ακόμα να γίνεται χρήση αναπαραστάσεων του περιβάλλοντος οι οποίες είναι προσαρμοσμένες σε κάποια ενέργεια (action driven). Επομένως είναι λογικό να συμπεράνουμε ότι ένα σημαντικό τμήμα της ερευνητικής προσπάθειας για την κατανόηση του χώρου πρέπει να εστιαστεί στην μελέτη των συνδυασμού των διάφορων μορφών αναπαράστασής του αλλά και στην μετατροπή των αναπαραστάσεων από την μια μορφή στην άλλη. Κάποιες ιδέες προς την κατεύθυνση αυτή θα μπορούσε να δώσει το μοντέλο που έχει προταθεί από τον Feldman [28].

Παράρτημα Α

Σχέσεις μεταξύ των καρτεσιανών και των λογαριθμικών πολικών μερικών παραγώγων

Στη συνέχεια θα δώσουμε τις εξισώσεις που σχετίζουν τις καρτεσιανές μερικές παραγώγους της έντασης (ως προς x, y) με τις λογαριθμικές πολικές μερικές παραγώγους (ως προς ξ, η). Θα συμβολίζουμε με \hat{f} την $\frac{df}{dt}$ και με f_s την $\frac{\partial f}{\partial s}$. Υπολογίζουμε πρώτα το $\hat{\rho}$.

$$\begin{aligned}
 \hat{\rho} &= \frac{d\rho}{dt} \\
 &= \frac{\partial \rho}{\partial x} \frac{dx}{dt} + \frac{\partial \rho}{\partial y} \frac{dy}{dt} \\
 &= \frac{\hat{x}2x}{2\sqrt{x^2+y^2}} + \frac{2y\hat{y}}{2\sqrt{x^2+y^2}} \\
 &= \frac{\rho \cos(\eta)\hat{x}}{\sqrt{\rho^2 \cos^2(\eta) + \rho^2 \sin^2(\eta)}} \frac{\rho \sin(\eta)\hat{y}}{\sqrt{\rho^2 \cos^2(\eta) + \rho^2 \sin^2(\eta)}} \\
 &= \cos(\eta)\hat{x} + \sin(\eta)\hat{y}
 \end{aligned} \tag{A.1}$$

Η (A.1) σε συνδυασμό με το γεγονός ότι $\frac{d(\arctan(x))}{dx} = \frac{1}{1+x^2}$, δίνει για τα $\hat{\xi}, \hat{\eta}$

$$\begin{aligned}
 \hat{\xi} &= \frac{d\xi}{dt} \\
 &= \frac{\partial \xi}{\partial x} \frac{dx}{dt} + \frac{\partial \xi}{\partial y} \frac{dy}{dt} \\
 &= \frac{1}{\sqrt{x^2+y^2}} \frac{1}{2\sqrt{x^2+y^2}} 2x\hat{x} + \frac{1}{\sqrt{x^2+y^2}} \frac{2y}{2\sqrt{x^2+y^2}} \hat{y} \\
 &= \frac{1}{\sqrt{x^2+y^2}} \left(\frac{x}{\sqrt{x^2+y^2}} \hat{x} + \frac{y}{\sqrt{x^2+y^2}} \hat{y} \right) \\
 &= \frac{1}{\rho} (\cos(\eta)\hat{x} + \sin(\eta)\hat{y})
 \end{aligned}$$

$$= \frac{\hat{\rho}}{\rho} \quad (\text{A.2})$$

$$\begin{aligned}\hat{\eta} &= \frac{d\eta}{dt} \\ &= \frac{\partial\eta}{\partial x} \frac{dx}{dt} + \frac{\partial\eta}{\partial y} \frac{dy}{dt} \\ &= \frac{1}{1 + \frac{y^2}{x^2}} \left(-\frac{1}{x^2} \right) y \hat{x} + \frac{1}{1 + \frac{y^2}{x^2}} \frac{1}{x} \hat{y} \\ &= -\frac{\hat{x}}{x^2 + y^2} y + \frac{\hat{y}}{x + \frac{y^2}{x}} \\ &= -\frac{\rho \sin(\eta) \hat{x}}{\rho^2} + \frac{\hat{y} x}{x^2 + y^2} \\ &= -\frac{\sin(\eta) \hat{x}}{\rho} + \frac{\rho \cos(\eta) \hat{y}}{\rho^2} \\ &= \frac{-\sin(\eta) \hat{x} + \cos(\eta) \hat{y}}{\rho} \end{aligned} \quad (\text{A.3})$$

Για τις μερικές παραγόγους της έντασης ως προς x, y έχουμε

$$\begin{aligned}\frac{\partial I}{\partial x} &= \frac{\partial I}{\partial \xi} \frac{\partial \xi}{\partial x} + \frac{\partial I}{\partial \eta} \frac{\partial \eta}{\partial x} \\ &= I_\xi \frac{1}{\sqrt{x^2 + y^2}} \frac{2x}{2\sqrt{x^2 + y^2}} + I_\eta \frac{1}{1 + \frac{y^2}{x^2}} y \left(-\frac{1}{x^2} \right) \\ &= I_\xi \frac{1}{\rho} \cos(\eta) - I_\eta \frac{1}{\rho} \sin(\eta) \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned}\frac{\partial I}{\partial y} &= \frac{\partial I}{\partial \xi} \frac{\partial \xi}{\partial y} + \frac{\partial I}{\partial \eta} \frac{\partial \eta}{\partial y} \\ &= I_\xi \frac{1}{2\sqrt{x^2 + y^2}} \frac{2y}{\sqrt{x^2 + y^2}} + I_\eta \frac{1}{1 + \frac{y^2}{x^2}} \frac{1}{x} \\ &= I_\xi \frac{1}{\rho} \sin(\eta) + I_\eta \frac{1}{\rho} \cos(\eta) \end{aligned} \quad (\text{A.5})$$

Οι (5.18), (5.19) δίνουν για το μέτρο της κλίσης

$$\begin{aligned}||\nabla I|| &= \sqrt{I_x^2 + I_y^2} \\ &= \sqrt{(I_\xi \frac{1}{\rho} \cos(\eta) - I_\eta \frac{1}{\rho} \sin(\eta))^2 + (I_\xi \frac{1}{\rho} \sin(\eta) + I_\eta \frac{1}{\rho} \cos(\eta))^2} \\ &= \sqrt{(I_\xi \frac{1}{\rho} \cos(\eta))^2 - 2I_\xi \frac{1}{\rho} \cos(\eta) I_\eta \frac{1}{\rho} \sin(\eta) + (I_\eta \frac{1}{\rho} \sin(\eta))^2 + (I_\xi \frac{1}{\rho} \sin(\eta))^2 + 2I_\xi \frac{1}{\rho} \sin(\eta) I_\eta \frac{1}{\rho} \cos(\eta) + (I_\eta \frac{1}{\rho} \cos(\eta))^2} \\ &= \sqrt{I_\xi^2 \frac{1}{\rho^2} + I_\eta^2 \frac{1}{\rho^2}} \end{aligned}$$

Σχέσεις μεταξύ των καρτεσιανών και των λογαριθμικών πολικών μερικών παραγώγων

$$= \frac{1}{\rho} \sqrt{I_\xi^2 + I_\eta^2} \quad (\text{A.6})$$

Η μερική παράγωγος της έντασης ως προς τον χρόνο είναι η ίδια, ανεξάρτητα από το αν υπολογίζεται στο πεδίο x, y ή στο ξ, η . Αυτό μπορεί να αποδειχθεί ως εξής: Έστω I_t^{xy} η μερική παράγωγος που υπολογίζεται στο πεδίο x, y και $I_t^{\xi\eta}$ αυτή που υπολογίζεται στο πεδίο ξ, η . Από την εξίσωση περιορισμού της οπτικής ροής έχουμε

$$I_t^{xy} = -(I_x \hat{x} + I_y \hat{y}) \quad (\text{A.7})$$

Η εξίσωση περιορισμού της οπτικής ροής στο πεδίο ξ, η και οι (A.2), (A.3), δίνουν για την $I_t^{\xi\eta}$

$$\begin{aligned} \frac{dI}{dt} &= \frac{\partial I}{\partial \xi} \frac{d\xi}{dt} + \frac{\partial I}{\partial \eta} \frac{d\eta}{dt} + I_t^{\xi\eta} \\ &= I_\xi \hat{\xi} + I_\eta \hat{\eta} + I_t^{\xi\eta} \\ &= I_\xi \left(\frac{1}{\rho} (\cos(\eta) \hat{x} + \sin(\eta) \hat{y}) \right) + I_\eta \left(\frac{-\sin(\eta) \hat{x} + \cos(\eta) \hat{y}}{\rho} \right) + I_t^{\xi\eta} \\ &= \hat{x} \left(I_\xi \frac{1}{\rho} \cos(\eta) - I_\eta \frac{1}{\rho} \sin(\eta) \right) + \hat{y} \left(I_\xi \frac{1}{\rho} \sin(\eta) + I_\eta \frac{1}{\rho} \cos(\eta) \right) + I_t^{\xi\eta} \\ &= \hat{x} \frac{\partial I}{\partial x} + \hat{y} \frac{\partial I}{\partial y} + I_t^{\xi\eta} \\ &= 0 \Rightarrow \\ I_t^{\xi\eta} &= -(I_x \hat{x} + I_y \hat{y}) \end{aligned} \quad (\text{A.8})$$

Από τις (A.7), (A.8) συνεπάγεται ότι $I_t^{xy} = I_t^{\xi\eta}$.

Παράρτημα Β

Πρόσημο της Gaussian καμπυλότητας για την επιφάνεια που ορίζει η $Z(x, y)$

Έστω $Z(x, y)$ το βάθος του pixel (x, y) και $r : \mathcal{R}^2 \rightarrow \mathcal{R}^3$ με $r(x, y) = (x, y, Z(x, y))$ η επιφάνεια η οποία ορίζεται από το γράφημα της συνάρτησης $Z(x, y)$. Θα δείξουμε ότι το πρόσημο της Gaussian καμπυλότητας για την επιφάνεια r δίνεται από την (6.3).

Παρατηρούμε ότι το πρόσημο του K στην (6.1) είναι το ίδιο με αυτό του αριθμητή, μια και ο παρονομαστής είναι πάντα θετικός. Πράγματι, το τετράγωνο του μέτρου του κάθετου διανύσματος n στην επιφάνεια είναι

$$\begin{aligned} |n|^2 &= \left| \frac{\partial r}{\partial x} \times \frac{\partial r}{\partial y} \right|^2 \\ &= \left(\frac{\partial r}{\partial x} \right)^2 \left(\frac{\partial r}{\partial y} \right)^2 - \left(\frac{\partial r}{\partial x} \cdot \frac{\partial r}{\partial y} \right)^2 \\ &= EG - F^2 > 0 \end{aligned}$$

Είναι ακόμα $N = \frac{n}{|n|}$, οπότε το πρόσημο της Gaussian καμπυλότητας είναι ίσο με αυτό της παράστασης

$$L \equiv (n \cdot \frac{\partial^2 r}{\partial x^2})(n \cdot \frac{\partial^2 r}{\partial y^2}) - (n \cdot \frac{\partial^2 r}{\partial x \partial y})^2 \quad (\text{B.1})$$

Υπολογίζουμε τώρα κάθε ένα από τους όρους της (B.1):

$$\begin{aligned} n &= (1, 0, \frac{\partial Z}{\partial x}) \times (0, 1, \frac{\partial Z}{\partial y}) \\ &= \left(-\frac{\partial Z}{\partial x}, -\frac{\partial Z}{\partial y}, 1 \right) \end{aligned} \quad (\text{B.2})$$

$$\begin{aligned} \frac{\partial^2 r}{\partial x^2} &= \frac{\partial}{\partial x}(1, 0, \frac{\partial Z}{\partial x}) \\ &= (0, 0, \frac{\partial^2 Z}{\partial x^2}) \end{aligned} \quad (\text{B.3})$$

Πρόσημο της Gaussian καμπυλότητας για την επιφάνεια που ορίζει η $Z(x, y)$

$$\begin{aligned}\frac{\partial^2 r}{\partial y^2} &= \frac{\partial}{\partial y}(0, 1, \frac{\partial Z}{\partial y}) \\ &= (0, 0, \frac{\partial^2 Z}{\partial y^2})\end{aligned}\tag{B.4}$$

$$\begin{aligned}\frac{\partial^2 r}{\partial x \partial y} &= \frac{\partial}{\partial x}(0, 1, \frac{\partial Z}{\partial y}) \\ &= (0, 0, \frac{\partial^2 Z}{\partial x \partial y})\end{aligned}\tag{B.5}$$

Αντικαθιστώντας τις παραπάνω εξισώσεις στην (B.1) έχουμε

$$\begin{aligned}L &= ((-\frac{\partial Z}{\partial x}, -\frac{\partial Z}{\partial y}, 1) \cdot (0, 0, \frac{\partial^2 Z}{\partial x^2}))((-\frac{\partial Z}{\partial x}, -\frac{\partial Z}{\partial y}, 1) \cdot (0, 0, \frac{\partial^2 Z}{\partial y^2})) - \\ &\quad ((-\frac{\partial Z}{\partial x}, -\frac{\partial Z}{\partial y}, 1) \cdot (0, 0, \frac{\partial^2 Z}{\partial x \partial y}))^2 \\ &= \frac{\partial^2 Z}{\partial x^2} \frac{\partial^2 Z}{\partial y^2} - (\frac{\partial^2 Z}{\partial x \partial y})^2\end{aligned}\tag{B.6}$$

δηλαδή την (6.3).

Παράρτημα C

Διατακτικό βάθος στην περίπτωση προοπτικής προβολής

Στη συνέχεια θα δώσουμε τις εξισώσεις που σχετίζουν την κάθετη ροή με τις παραμέτρους κίνησης υποθέτωντας προοπτική προβολή και θα περιγράψουμε μερικές περιπτώσεις στις οποίες είναι δυνατός ο υπολογισμός διατακτικού βάθους. Κύριος στόχος μας θα είναι να γίνει κατανοητό ότι στην περίπτωση προοπτικής προβολής, οι εξισώσεις γίνονται αρκετά περίπλοκες, έτσι ώστε να επιτρέπουν τον υπολογισμό διατακτικού βάθους μόνο σε ειδικές περιπτώσεις.

C.1 Οι εξισώσεις για την κάθετη ροή στην περίπτωση προοπτικής προβολής

Έστω το σημείο (X, Y, Z) , το οποίο μέσω της προοπτικής προβολής απεικονίζεται στο σημείο (x, y) . Η τιμή (u, v) του πεδίου κίνησης στο σημείο (x, y) είναι εξ' ορισμού ίση με

$$\begin{aligned} u &= \frac{dx}{dt} = f \frac{d(X/Z)}{dt} \\ v &= \frac{dy}{dt} = f \frac{d(Y/Z)}{dt} \end{aligned} \tag{C.1}$$

δηλαδή

$$\begin{aligned} u &= f \frac{\frac{dX}{dt}Z - X\frac{dZ}{dt}}{Z^2} \\ v &= f \frac{\frac{dY}{dt}Z - Y\frac{dZ}{dt}}{Z^2} \end{aligned} \tag{C.2}$$

Οι $\frac{dX}{dt}, \frac{dY}{dt}, \frac{dZ}{dt}$ δίνονται από τις εξισώσεις (4.5), οπότε οι (C.2) με αντικατάσταση δίνουν

$$u = \frac{-Uf + xW}{Z} + \alpha \frac{xy}{f} - \beta \left(\frac{x^2}{f} + f \right) + \gamma y$$

C.2 Σχέση μεταξύ των παραμέτρων U, W, β για ένα fixating ζεύγος

$$v = \frac{-Vf+yW}{Z} + \alpha\left(\frac{y^2}{f} + f\right) - \beta\frac{xy}{f} - \gamma x \quad (\text{C.3})$$

Δεδομένου ότι $\alpha = \gamma = V = 0$, οι (C.3) γίνονται

$$\begin{aligned} u &= \frac{-Uf+xW}{Z} - \beta\left(\frac{x^2}{f} + f\right) \\ v &= \frac{yW}{Z} - \beta\frac{xy}{f} \end{aligned} \quad (\text{C.4})$$

Η αλγεβρική τιμή του πεδίου κάθετης κίνησης είναι

$$\begin{aligned} v_o^n &= (u, v)\left(\frac{I_x}{\|\nabla I\|}, \frac{I_y}{\|\nabla I\|}\right) = \\ &= (uI_x + vI_y)\frac{1}{\|\nabla I\|} \end{aligned} \quad (\text{C.5})$$

Υποθέτωντας ότι $I_x \neq 0$, η (C.5) μπορεί να γραφτεί και σαν

$$\begin{aligned} -v_o^n \frac{\|\nabla I\|}{I_x} &= -u - v \frac{I_y}{I_x} = \\ &= \frac{Uf - xW}{Z} + \beta\left(\frac{x^2}{f} + f\right) - \frac{yW}{Z} \frac{I_y}{I_x} + \beta \frac{xy}{f} \frac{I_y}{I_x} \end{aligned} \quad (\text{C.6})$$

Προσεγγίζοντας το πεδίο κάθετης κίνησης με το πεδίο κάθετης ροής, μπορούμε να αντικαταστήσουμε το v_o^n με την αλγεβρική τιμή του πεδίου κάθετης ροής, η οποία όπως έχουμε ήδη δείξει είναι ίση με $-\frac{I_t}{\|\nabla I\|}$. Κατά συνέπεια, η εξίσωση (C.6) γίνεται

$$H \equiv \frac{I_t}{I_x} = \left(\frac{Uf}{W} - x - yn\right) \frac{W}{Z} + \beta\left(\frac{x^2}{f} + \frac{xy}{f}n + f\right) \quad (\text{C.7})$$

όπου $n = \frac{I_y}{I_x}$. Το $\frac{Uf}{W}$ είναι η τετμημένη του σημείου $(\frac{Uf}{W}, \frac{Vf}{W})$, το οποίο είναι γνωστό σαν Focus Of Expansion - FOE. Το σημείο αυτό, είναι η προβολή του διανύσματος (U, V, W) στο επίπεδο της εικόνας και ορίζει την διεύθυνση της μεταφορικής κίνησης.

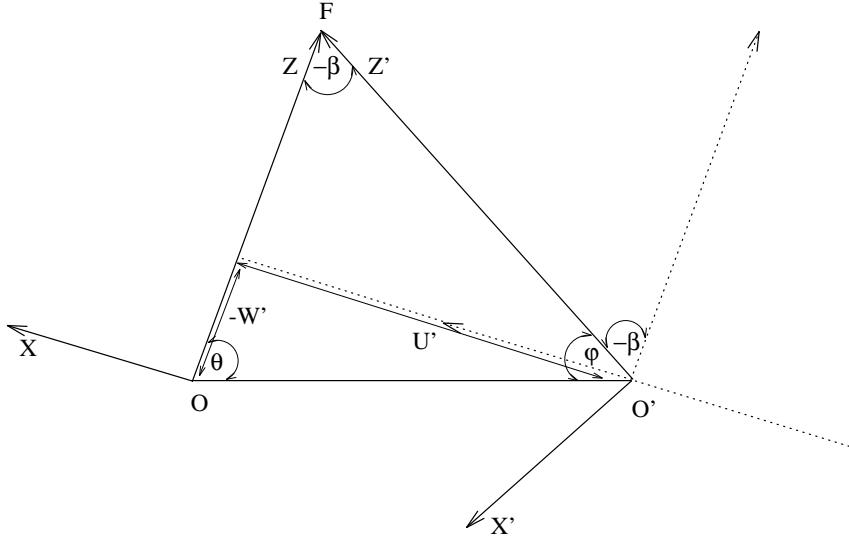
C.2 Σχέση μεταξύ των παραμέτρων U, W, β για ένα fixating ζεύγος

Ειδικά στην περίπτωση ενός fixating ζεύγους εικόνων, μπορούμε να βρούμε μια σχέση μεταξύ των μεταβλητών U, W, β . Από το σχήμα 5.3 είναι φανερό ότι $\tan(\varphi) = \frac{U}{W}$. Ανάλογα, θεωρώντας ότι η υποθετική κίνηση της κάμερας γίνεται με φορά από την δεξιά κάμερα του στερεοσκοπικού συστήματος προς την αριστερή, το σχήμα C.1 δίνει $\tan(\theta) = -\frac{U'}{W'}$. Το μείον στην εξίσωση αυτή μπαίνει γιατί τα U' και W' έχουν αντίθετα πρόσημα ($U' > 0, W' < 0$). Επιπλέον, είναι και $\beta = 2\pi - (\varphi + \theta)$. Δεδομένου ότι τα αντίθετα και τα παραπληρωματικά τόξα έχουν αντίθετες εφαπτομένες, η τελευταία σχέση δίνει

$$\begin{aligned} \beta &= 2\pi - (\varphi + \theta) \Rightarrow \\ \pi - \beta &= -(\pi - (\varphi + \theta)) \Rightarrow \\ \tan(\pi - \beta) &= \tan(-(\pi - (\varphi + \theta))) \Rightarrow \end{aligned}$$

C.2 Σχέση μεταξύ των παραμέτρων U, W, β για ένα fixating ζεύγος

$$\begin{aligned}
-\tan(\beta) &= -\tan(\pi - (\varphi + \theta)) = \tan(\varphi + \theta) \Rightarrow \\
\beta &= \arctan\left(-\frac{\tan(\varphi) + \tan(\theta)}{1 - \tan(\varphi)\tan(\theta)}\right) \Rightarrow \\
\beta &= \arctan\left(-\frac{\frac{U}{W} - \frac{U'}{W'}}{1 + \frac{U}{W}\frac{U'}{W'}}\right)
\end{aligned} \tag{C.8}$$



Σχήμα C.1: Μετακίνηση του συστήματος $O'X'Y'Z'$ στο $OXYZ$.

Υποθέτοντας ότι έχουμε μια μέθοδο που επιτρέπει τον ακριβή υπολογισμό του FOE , τότε η (C.8) μπορεί να χρησιμοποιηθεί για τον υπολογισμό του β και στην συνέχεια η (C.7) μπορεί να χρησιμοποιηθεί για τον υπολογισμό σχετικού βάθους. Πράγματι, χρησιμοποιώντας τους συμβολισμούς

$$\begin{aligned}
C &= \frac{Uf}{W} - x - yn \\
D &= \frac{x^2}{f} + \frac{xy}{f}n + f
\end{aligned} \tag{C.9}$$

η (C.7) γράφεται

$$H = C\frac{W}{Z} + \beta D \tag{C.10}$$

Για δύο σημεία με συντεταγμένες (X_1, Y_1, Z_1) και (X_2, Y_2, Z_2) η παραπάνω εξίσωση δίνει

$$\begin{aligned}
H_1 - \beta D_1 &= C_1 \frac{W}{Z_1} \\
H_2 - \beta D_2 &= C_2 \frac{W}{Z_2}
\end{aligned} \tag{C.11}$$

C.3 Διάταξη σημείων με βάση το βάθος σε ειδικές περιπτώσεις

οπότε διαιρώντας κατά μέλη παίρνουμε

$$\frac{Z_2}{Z_1} = \frac{C_2}{C_1} \frac{H_1 - \beta D_1}{H_2 - \beta D_2} \quad (\text{C.12})$$

Η σύγκριση του δεξιού μέλους της τελευταίας εξίσωσης με τη μονάδα, επιτρέπει τον υπολογισμό του σχετικού βάθους των δύο σημείων. Το σημαντικότερο μειονέκτημα της μέθοδου που περιγράφηκε είναι ότι έμμεσα υποθέτει ακριβή fixations, στα οποία το fixation point απεικονίζεται ακριβώς στο κέντρο κάθε εικόνας του ζεύγους. Κάπι τέτοιο είναι δύσκολο να γίνει, με αποτέλεσμα η τιμή του β που υπολογίζεται από την (C.8) να διαφέρει από την πραγματική.

C.3 Διάταξη σημείων με βάση το βάθος σε ειδικές περιπτώσεις

Θα περιγράψουμε στη συνέχεια μερικές περιπτώσεις στις οποίες είναι δυνατός ο υπολογισμός της σχέσης δύο σημείων αναφορικά με το βάθος τους, χωρίς να εκμεταλλευόμαστε την σχέση μεταξύ των U, W, β . Έστω ένα ζεύγος σημείων P_1, P_2 με συντεταγμένες (X_1, Y_1, Z_1) και (X_2, Y_2, Z_2) και αντίστοιχες προβολές στο επίπεδο της εικόνας (x_1, y_1) και (x_2, y_2) .

Ας υποθέσουμε αρχικά ότι είναι γνωστά κάποια φράγματα για την ποσότητα $\frac{Uf}{W}$, δηλ. $\frac{Uf}{W} \in [\mathcal{F}_{min}, \mathcal{F}_{max}]$, ενώ δεν γνωρίζουμε τίποτα για το β . Διαιρώντας και τα δύο μέλη της (C.7) δια $\frac{x^2}{f} + \frac{xy}{f}n + f$, παίρνουμε

$$L \equiv \frac{H}{\frac{x^2}{f} + \frac{xy}{f}n + f} = M \frac{W}{Z} + \beta \quad (\text{C.13})$$

όπου

$$M = \frac{\frac{Uf}{W} - x - yn}{\frac{x^2}{f} + \frac{xy}{f}n + f}$$

Από τα φράγματα για την ποσότητα $\frac{Uf}{W}$, μπορούμε να υπολογίσουμε φράγματα για την ποσότητα M . Πράγματι, χωρίς περιορισμό της γενικότητας μπορούμε να υποθέσουμε ότι $\frac{x^2}{f} + \frac{xy}{f}n + f > 0$, οπότε

$$M_{min} \equiv \frac{\mathcal{F}_{min} - x - yn}{\frac{x^2}{f} + \frac{xy}{f}n + f} < M < \frac{\mathcal{F}_{max} - x - yn}{\frac{x^2}{f} + \frac{xy}{f}n + f} \equiv M_{max} \quad (\text{C.14})$$

Έτσι για τα σημεία P_1, P_2 έχουμε

$$M_{min}^1 \frac{W}{Z_1} + \beta > M_1 \frac{W}{Z_1} + \beta > M_{max}^1 \frac{W}{Z_1} + \beta \quad (\text{C.15})$$

$$M_{min}^2 \frac{W}{Z_2} + \beta > M_2 \frac{W}{Z_2} + \beta > M_{max}^2 \frac{W}{Z_2} + \beta \quad (\text{C.16})$$

C.3 Διάταξη σημείων με βάση το βάθος σε ειδικές περιπτώσεις

Πολλαπλασιάζοντας τις ανισότητες για το P_2 επί -1 , προσθέτωντας κατά μέλη και απαλείφοντας το W προκύπτει η

$$M_{min}^1 \frac{1}{Z_1} - M_{max}^2 \frac{1}{Z_2} < M_1 \frac{1}{Z_1} - M_2 \frac{1}{Z_2} < M_{max}^1 \frac{1}{Z_1} - M_{min}^2 \frac{1}{Z_2} \quad (\text{C.17})$$

Οι παραπάνω ανισότητες μπορούν σε κάποιες περιπτώσεις να δώσουν το πρόσημο της παράστασης $M_1 \frac{1}{Z_1} - M_2 \frac{1}{Z_2}$. Αν π.χ. το δεξιότερο μέλος είναι αρνητικό, τότε είναι $M_1 \frac{1}{Z_1} - M_2 \frac{1}{Z_2} < 0$, ενώ αν το αριστερότερο μέλος είναι θετικό, τότε $M_1 \frac{1}{Z_1} - M_2 \frac{1}{Z_2} > 0$. Ας υποθέσουμε ότι

$$M_1 \frac{1}{Z_1} - M_2 \frac{1}{Z_2} > 0$$

ή ισοδύναμα

$$M_1 > M_2 \frac{Z_1}{Z_2}$$

Διακρίνουμε τις ακόλουθες περιπτώσεις:

- $M_2 > 0$. Τότε $\frac{M_1}{M_2} > \frac{Z_1}{Z_2}$ και αν $\frac{M_1}{M_2} < 1$ τότε $Z_2 > Z_1$
- $M_2 < 0$. Τότε $\frac{M_1}{M_2} < \frac{Z_1}{Z_2}$ και αν $\frac{M_1}{M_2} > 1$ τότε $Z_2 < Z_1$

Στην περίπτωση που η τιμή της ποσότητας $\frac{Uf}{W}$ είναι γνωστή ακριβώς, η (C.13) δίνει για τα P_1, P_2

$$L_1 = M_1 \frac{W}{Z_1} + \beta \quad (\text{C.18})$$

$$L_2 = M_2 \frac{W}{Z_2} + \beta \quad (\text{C.19})$$

οπότε αφαιρώντας κατά μέλη, έχουμε

$$L_1 - L_2 = W(M_1 \frac{1}{Z_1} - M_2 \frac{1}{Z_2}) \quad (\text{C.20})$$

Οι ποσότητες L_1, L_2, M_1, M_2 είναι τώρα γνωστές, όπως και το πρόσημο του W (αρνητικό). Έτσι, η (C.20) μπορεί να δώσει το πρόσημο της παράστασης $M_1 \frac{1}{Z_1} - M_2 \frac{1}{Z_2}$ και η ανάλυση συνεχίζει όπως παραπάνω.

Στην περίπτωση που εκτός από την ποσότητα $\frac{Uf}{W}$ είναι γνωστά και κάποια φράγματα για το β , δηλαδή $\beta \in [\beta_{min}, \beta_{max}]$, τότε μπορούμε να διατάξουμε ως προς το βάθος επιπλέον ξενήγη σημείων. Διαιρώντας και τα δύο μέλη της (C.7) δια $\frac{Uf}{W} - x - yn$ έχουμε

$$N \equiv \frac{H}{\frac{Uf}{W} - x - yn} = \frac{W}{Z} + \beta S \quad (\text{C.21})$$

όπου

$$S = \frac{\frac{x^2}{f} + \frac{xy}{f} n + f}{\frac{Uf}{W} - x - yn}$$

C.3 Διάταξη σημείων με βάση το βάθος σε ειδικές περιπτώσεις

H προηγούμενη εξίσωση δίνει για τα P_1, P_2

$$N_1 = \frac{W}{Z_1} + \beta S_1 \quad (\text{C.22})$$

$$N_2 = \frac{W}{Z_2} + \beta S_2 \quad (\text{C.23})$$

Αφαιρώντας κατά μέλη, έχουμε

$$N_1 - N_2 = W\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right) + \beta(S_1 - S_2) \quad (\text{C.24})$$

Οι ποσότητες N_1, N_2, S_1, S_2 είναι γνωστές. Χωρίς περιορισμό της γενικότητας, υποθέτουμε ότι $N_1 - N_2 > 0$ Διακρίνουμε τις παρακάτω περιπτώσεις:

- $S_1 = S_2$. Τότε $W\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right) > 0$ και άρα $Z_1 > Z_2$. *H περίπτωση αυτή συμβαίνει όταν οι συντελεστές διεύθυνσης της κλίσης είναι για τα δύο σημεία είναι ίσοι με $n_1 = -\frac{x_1}{y_1}, n_2 = -\frac{x_2}{y_2}$ αντίστοιχα. Τα αντίστοιχα διανύσματα κάθετης ροής δηλαδή, είναι εφαπτόμενα στους κύκλους που ορίζουν τα διανύσματα θέσης των δύο σημείων¹. Είναι σημαντικό να επισημάνουμε ότι στην περίπτωση αυτή, η διάταξη ως προς το βάθος δύο σημείων χρησιμοποιώντας ορθογραφική προβολή, είναι ίδια με αυτή που παίρνουμε κάνοντας χρήση προοπτικής προβολής. Κατά συνέπεια, η χρήση του προσεγγιστικού μοντέλου της ορθογραφικής προβολής, δεν επηρεάζει την ορθότητα της ταξινόμησης σημείων στα οποία τα διανύσματα θέσης είναι κάθετα στα αντίστοιχα διανύσματα κάθετης ροής.*

Πράγματι, για ένα σημείο τέτοιο ώστε $n = -\frac{x}{y}$, η (C.7) γίνεται

$$\frac{I_t}{I_x} = \frac{Uf}{Z} + \beta f$$

και άρα για ένα ξεύγος σημείων είναι

$$\frac{I_t^1}{I_x^1} - \frac{I_t^2}{I_x^2} = Uf\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right)$$

Υποθέτωντας ότι $\frac{I_t^1}{I_x^1} - \frac{I_t^2}{I_x^2} > 0$, και επειδή το U είναι αρνητικό, παίρνουμε $\frac{1}{Z_1} - \frac{1}{Z_2} < 0$ και άρα $Z_1 > Z_2$. Από την άλλη πλευρά, η ποσότητα $-\frac{I_t^1}{I_x^1} + \frac{I_t^2}{I_x^2}$ είναι προφανώς αρνητική, οπότε και η (5.12) θα δώσει $Z_1 > Z_2$.

- $S_1 < S_2$. Τότε $W\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right) > 0$ οπότε $Z_1 > Z_2$.
- $S_1 > S_2$. *Eίναι*

$$T \equiv (N_1 - N_2) - \beta_{min}(S_1 - S_2) \geq (N_1 - N_2) - \beta(S_1 - S_2) \geq (N_1 - N_2) - \beta_{max}(S_1 - S_2) \equiv T'$$

- *Av $T < 0$, τότε $W\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right) < 0$ και $Z_1 < Z_2$.*
- *Av $T' > 0$, τότε $W\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right) > 0$ και $Z_1 > Z_2$.*

¹Αυτό παρατηρήθηκε αρχικά από τον Αντώνη Αργυρό.

Βιβλιογραφία

- [1] J. Aloimonos and A. Badyopadhyay. Active vision. In IEEE 1st Int. Conf. on Computer Vision, pages 35--54, Jun. 1987.
- [2] Y. Aloimonos. Purposive and qualitative active vision. In DARPA Image Understanding Workshop, pages 816--828, 1990.
- [3] M. A. Arbib. Perceptual structures and distributed motor control. In V.B. Brooks, editor, Handbook of Physiology - The nervous system II. Motor Control, pages 1449--1480, Bethesda, MD, 1981. American Physiological Society.
- [4] F. Avnaim, J. D. Boissonnat, and B. Faverjon. A practical exact motion planning algorithm for polygonal objects amidst polygonal obstacles. Technical Report 890, INRIA, Sophia-Antipolis, France, 1988.
- [5] R. Bajcsy. Active perception. Proceedings of the IEEE, 76(8):996--1005, Aug. 1988.
- [6] D. Ballard. Animate vision. Artificial Intelligence, 48(1):57--86, Aug. 1991.
- [7] D. H. Ballard and C. M. Brown. Computer Vision. Prentice Hall, Englewood Cliffs, NJ, 1982.
- [8] D. H. Ballard and C. M. Brown. Principles of animate vision. In Yiannis Aloimonos, editor, Active perception, page 254. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
- [9] Y. Baram and Y. Barniv. Obstacle detection by recognizing binary expansion patterns. Technical Report CIS-9321, Technion, Israel, Sept. 1993.
- [10] J. Borenstein and Y. Koren. Real-time obstacle avoidance for fast mobile robots. IEEE Trans. on Sys. Man and Cybern., SMC-19(5):1179--1187, Sept. - Oct. 1989.
- [11] J. Borenstein and Y. Koren. The vector field histogram - fast obstacle avoidance for mobile robots. IEEE Trans. on Robotics and Autom., 7(3):278--288, Jun. 1991.
- [12] R. A. Brooks. Solving the find-path problem by good representation of free space. IEEE Trans. on Sys. Man and Cybern., SMC-13(3):190--197, Mar. - Apr. 1983.

- [13] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE J. Robotics Auromat.*, RA-2(7):14--23, Apr. 1986.
- [14] R. A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139--160, 1991.
- [15] J. Canny. A computational approach to edge detection. *IEEE Trans. on Pat. Anal. and Mach. Int.*, 8:679--697, 1986.
- [16] J. F. Canny. *The Complexity of Robot Motion Planning*. MIT Press, Cambridge, MA, 1988.
- [17] R. Chattergy. Some heuristics for the navigation of a robot. *Int. J. Robotics Res.*, 4(1):59--66, 1985.
- [18] D. T. Clemens and D. W. Jacobs. Space and time bounds on indexing 3-D models from 2-D images. *IEEE Trans. on Pat. Anal. and Mach. Int.*, PAMI-13(10):1007--1017, Oct. 1991.
- [19] T.S. Collet, E. Dillmann, A. Giger, and R. Wehner. Visual landmarks and route following in desert ants. *Journal of Comparative Physiology A*, 170:435--442, 1992.
- [20] T.S. Collet, S.N. Fry, and R. Wehner. Sequence learning by honeybees. *Journal of Comparative Physiology A*, 172:693--706, 1993.
- [21] I. J. Cox. Blanche: An experiment in guidance and navigation of an autonomous robot vehicle. *IEEE Trans. on Robotics and Autom.*, 7(2):193--204, Apr. 1991.
- [22] B. Crespi, C. Furlanello, and L. Stringa. A memory-based approach to navigation. *Biological Cybernetics*, 69:385--393, 1993.
- [23] A. Cumani. Edge detection in multispectral images. *CVGIP: Graphical models and image processing*, 53(1):40--51, 1991.
- [24] M. P. do Carmo. Differential geometry of curves and surfaces. Prentice Hall, Englewood Cliffs, NJ, 1976.
- [25] M. Drumheller. Mobile robot localization using sonar. *IEEE Trans. on Pat. Anal. and Mach. Int.*, PAMI-9(2):325--332, Mar. 1987.
- [26] A. Elfes. Occupancy Grids: A Probabilistic Framework for Mobile Robot Perception and Navigation. *PhD dissertation, Electrical and Computer Engineering Dept./ Robotics Institute, Carnegie Mellon University*, 1989.
- [27] B. Faverjon and P. Tournassoud. A local based approach for path planning of manipulators with a high number of degrees of freedom. In Proc. of the IEEE Int. Conf. on Robotics and Automation, pages 1152--1159, Raleigh, NC, 1987.

- [28] J. A. Feldman. Four frames suffice: A provisional model of vision and space. *The behavioral and brain sciences*, 88:265--289, 1985.
- [29] C. Fermuller. Basic Visual Capabilities. *PhD dissertation, Center for Automation Research, University of Maryland*, 1993.
- [30] C. Fermuller and Y. Aloimonos. *Design and analysis of vision systems*. 1994. To appear.
- [31] D. Forsyth, J. Mundy, A. Zisserman, C. Coelho, A. Heller, and C. Rothwell. Invariant descriptors for 3-D object recognition and pose. *IEEE Trans. on Pat. Anal. and Mach. Int., PAMI-13(10):971--991*, Oct. 1991.
- [32] G. Fouxf, M. Heymann, and A. Bruckstein. Two-dimensional robot navigation among unknown stationary polygonal obstacles. *IEEE Trans. on Robotics and Autom.*, 9(1), Feb. 1993.
- [33] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, third edition, 1992.
- [34] B. K. P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.
- [35] B. K. P. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185-203, 1981.
- [36] Adrian Horridge. Bee vision of pattern and 3D. *The Bidder Lecture, delivered at the Annual Meeting of the Society for Experimental Biology at Swansea*, Apr. 1994.
- [37] I. D. Horswill. Specialization of Perceptual Processes. *PhD dissertation, Massachusetts Institute of Technology*, 1993.
- [38] L. Huang and Y. Aloimonos. How normal flow constrains relative depth for an active observer. *Image and Vision Computing*, 12(7):435--445, Sept. 1994.
- [39] Yong K. Hwang and Narendra Ahuja. Gross motion planning -- a survey. *ACM Computing Surveys*, 24(3):221--291, Sept. 1992.
- [40] E. Johnston. Systematic distortions of shape from stereopsis. *Vision Research*, 31:1351--1360, 1991.
- [41] Pentti Kanerva. *Sparse distributed memory*. MIT Press, Cambridge, MA, 1988.
- [42] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. *Int. J. Robotics Res.*, 5(1):90--98, 1986.
- [43] J. Koenderink and A. van Doorn. Affine structure from motion. *Journal of the Optical Society of America A*, 8(2):377--385, Feb. 1991.

- [44] Jurgen Konczak. Towards an ecological theory of motor development: The relevance of the Gibsonian approach to vision for motor development research. In J. E. Clark and J. H. Humphrey, editors, *Advances in Motor Development Research*, volume 3, pages 201--224. 1991.
- [45] J. C. Latombe. *Robot Motion Planning*. Kluver Academic Publishers, Boston, MA, 1991.
- [46] A. Lazanas and J. C. Latombe. Motion planning with uncertainty: A landmark approach. To appear in *Artificial Intelligence Journal*, 1994.
- [47] T. S. Levitt and D. T. Lawton. Qualitative navigation for mobile robots. *Artificial Intelligence*, 44(3):305--360, Aug. 1990.
- [48] Nikos K. Logothetis, J. Pauls, and Tomaso Poggio. Viewer-centered object recognition in monkeys. Technical Report AIM-1473, Massachusetts Institute of Technology, Artificial Intelligence laboratory, Apr. 1994.
- [49] E. A. Lord and C. B. Wilson. The mathematical description of shape and form. *Ellis Horwood Ltd., London*, 1984.
- [50] T. Lozano-Perez. Spatial planning: A configuration space approach. *IEEE Trans. on Computers*, C-32(2):108--120, Feb. 1983.
- [51] T. Lozano-Perez and M. T. Mason. Automatic synthesis of fine-motion strategies for robots. *Int. J. Robotics Res.*, 3(1):3--24, 1984.
- [52] T. Lozano-Perez and M. A. Wesley. An algorithm for planning collision-free paths among polyhedral obstacles. *Commun. ACM*, 22(10):560--570, Oct. 1979.
- [53] V. Lumelski and A. Stepanov. Dynamic path planning for a mobile automation with limited information on the environment. *IEEE Trans. on Automat. Control*, CA-31(11), Nov. 1986.
- [54] V. J. Lumelsky, S. Mukhopadhyay, and K. Sun. Dynamic path planning in sensor-based terrain acquisition. *IEEE Trans. on Robotics and Autom.*, 6(4), Aug. 1990.
- [55] David Marr. *Vision*. W. H. Freeman, San Francisco, 1982.
- [56] B. D. Marsh, C. M. Brown, T. J. LeBlanc, M. L. Scott, T. G. Becker, P. Das, J. Karlsson, and C. A. Quiroz. Operating system support for animate vision. Technical Report 374, University of Rochester, Jun. 1991.
- [57] M. J. Mataric. Integration of representation into goal-driven behavior-based robots. *IEEE Trans. on Robotics and Autom.*, 8(3):304--312, Jun. 1992.

- [58] R. A. McCallum. Learning with incomplete selective perception. Technical report, Department of Computer Science, University of Rochester, Mar. 1993.
- [59] H. P. Moravec. The Stanford cart and the CMU rover. Proceedings of the IEEE, 71(7):872--884, Jul. 1983.
- [60] V. S. Nalwa. A Guided tour of Computer Vision, chapter 5-8. Addison-Wesley, 1993.
- [61] R. C. Nelson. Visual Navigation. PhD dissertation, University of Maryland, 1988.
- [62] R. C. Nelson and J. Aloimonos. Obstacle avoidance using flow field divergence. IEEE Trans. on Pat. Anal. and Mach. Int., PAMI-11(10):1102--1106, Oct. 1989.
- [63] N. J. Nilsson. A mobile automation, an application of artificial intelligence techniques. In Int. Joint Conf. Artif. Intell., pages 509--520, 1969.
- [64] C. O'Dunlaing, M. Sharir, and C. K. Yap. Generalized voronoi diagrams for moving a ladder. Technical Report 32, Robotics Laboratory, Courant Institute, New York University, 1984.
- [65] K. Pahlavan and J. O. Eklundh. A head-eye system -- analysis and design. CVGIP: Image Understanding, Special Issue on Purposive, Qualitative Active Vision, 56:41--56, 1992.
- [66] K. Pahlavan, T. Uhlin, and J. O. Eklundh. Active vision as a methodology. In Yiannis Aloimonos, editor, Active perception, chapter 1. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
- [67] H. Pick, A. Yonas, D. Gentile, P. Melendez, D. Wagner, and D. Wegesin. Perceptual aspects of navigation. In DARPA Image Understanding Workshop, pages 481--483, 1993.
- [68] F. P. Preparata and M. I. Shamos. Computational Geometry: An Introduction. Springer-Verlag, New York, NY, 1985.
- [69] D. L. Ringach and Y. Baram. A diffusion mechanism for obstacle detection from size-change information. IEEE Trans. on Pat. Anal. and Mach. Int., PAMI-15(12), Dec. 1993.
- [70] K. D. Rueb and K. C. Wong. Structuring free space as a hypergraph for roving robot path planning and navigation. IEEE Trans. on Pat. Anal. and Mach. Int., PAMI-9(2):263--273, Mar. 1987.
- [71] A. Saffiotti, K. G. Konolige, and E. H. Rusconi. A multivalued logic approach to integrating planning and control. Technical Report 533, SRI Artificial Intelligence Center, Jun. 1993.
- [72] I. Schmidt, T.S. Collet, F.-X. Dillier, and R. Wehner. How desert ants cope with enforced detours on their way home. Journal of Comparative Physiology A, 171:285--288, 1992.

- [73] J. T. Schwartz and M. Sharir. On the ‘Piano Movers’ Problem: II General Techniques for Computing Topological properties of Real Algebraic Manifolds, pages 298–351. *Advances in Applied Mathematics*, Academic Press 4, 1983.
- [74] M. E. Spetsakis and J. Aloimonos. Optimal motion estimation. In IEEE Workshop on visual motion, pages 229–237, 1989.
- [75] K. A. Stevens. The visual interpretation of surface contours. *Artificial Intelligence*, 17:47–73, 1981.
- [76] M. J. Swain and M. A. Stricker. Promising directions in active vision. *International Journal of Computer Vision*, 11(2):109–126, 1993.
- [77] R. B. Tilove. Local obstacle avoidance for mobile robots based on the method of artificial potentials. In Proc. of the IEEE Int. Conf. on Robotics and Automation, pages 566–571, Cincinnati, OH, 1990.
- [78] J. Todd and Reichel. Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychology Review*, 96:643–657, 1989.
- [79] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. on Pat. Anal. and Mach. Int., PAMI-6*:13–27, 1984.
- [80] S. Ullman. The interpretation of structure from motion. In Royal Society, London, B, volume 203, pages 405–426, 1979.
- [81] S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE Trans. on Pat. Anal. and Mach. Int., PAMI-13*(10):992–1006, Oct. 1991.
- [82] A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *IEEE Trans. on Pat. Anal. and Mach. Int., 11*:490–498, 1989.
- [83] R. S. Wallace, P. W. Ong, B. B. Bederson, and E. L. Schwartz. Space variant image processing. *International Journal of Computer Vision*, 13(1):71–90, 1994.
- [84] D. Weinshall. Direct computation of qualitative 3-D shape and motion invariants. *IEEE Trans. on Pat. Anal. and Mach. Int., 13*(12):1236–1240, 1991.
- [85] A. Zeki. The visual image in mind and brain. *Scientific American*, 267(3):69–76, Sep. 1992.
- [86] A. Zelinsky. Environment Exploration and Path Planning Algorithms for Mobile Robot Navigation using Sonar. *PhD dissertation, Computer Science Dept., University of Wollongong, Australia*, 1991.