



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας
Λόγου και Επεξεργασίας Σημάτων

Εικονικός πράκτορας υποβοήθησης συναρμολόγησης αντικειμένου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

JACK HADFIELD

Επιβλέπων : Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2017



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας
Λόγου και Επεξεργασίας Σημάτων

Εικονικός πράκτορας υποβοήθησης συναρμολόγησης αντικειμένου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

JACK HADFIELD

Επιβλέπων : Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18η Ιουλίου 2017.

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

.....
Γεράσιμος Ποταμάνος
Αναπληρωτής Καθηγητής
Παν/μίου Θεσσαλίας

.....
Κωνσταντίνος Τζαφέστας
Επίκουρος Καθηγητής
Ε.Μ.Π.

Αθήνα, Ιούλιος 2017

.....
Jack Hadfield

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

Copyright © Jack Hadfield, 2017.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Στη διπλωματική εργασία αυτή εξετάζουμε το πρόβλημα της υποβοήθησης συναρμολόγησης αντικειμένου αξιοποιώντας τεχνικές όρασης υπολογιστών. Προτείνουμε μια πρωτότυπη προσέγγιση, βασισμένη σε μεθόδους εκτίμησης πόζας αντικειμένων. Χρησιμοποιήσαμε μια συσκευή Kinect για την αυτόματη παρακολούθηση της διαδικασίας της συναρμολόγησης όσο αυτή εξελίσσεται. Οι πόζες των μερών της κατασκευής εκτιμώνται σε δύο στάδια: πρώτα χρησιμοποιώντας έναν αλγόριθμο παρακολούθησης βάση ανίχνευσης πάνω στις εικόνες χρώματος για την παραγωγή αδρών εκτιμήσεων 2D θέσεων στο επίπεδο της εικόνας, και έπειτα με χρήση Rao-Blackwellized Particle Filter για τη βελτίωση των εκτιμήσεων και τον υπολογισμό των υπόλοιπων παραμέτρων των 3D ποζών. Στη συνέχεια, αξιοποιείται η πληροφορία της παρακολούθησης για τον εντοπισμό των ολοκληρωμένων συνδέσεων μεταξύ των μερών και την εκτίμηση της κατάστασης της συναρμολόγησης. Με βάση την κατάσταση αυτή, εμφανίζεται σχετική πληροφορία στον χρήστη του συστήματος. Παρουσιάζουμε τρεις εφαρμογές του συστήματός μας, που αναδεικνύουν την χρησιμότητά του, τη γενικότητα ως προς τα σχήματα των αντικειμένων και τις δυνατότητες λειτουργίας του σε πραγματικό χρόνο.

Λέξεις κλειδιά

Όραση υπολογιστών, σύστημα υποβοήθησης συναρμολόγησης, μοντελοποίηση συναρμολόγησης, εντοπισμός συνδέσεων, 3D παρακολούθηση, MOT, Particle Filter, κάμερα βάθους, ανίχνευση χρώματος, παρακολούθηση-βάση-ανίχνευσης, αλληλεπίδραση ανθρώπου-μηχανής

Abstract

In this thesis we examine the problem of object assembly assistance using computer vision techniques. We propose a novel approach, based on object pose estimation methods. A Kinect device is used to observe the assembly process in action. The poses of the assembly parts are estimated in two stages: first using a colour detection-based-tracker to produce crude 2D position estimates on the image plane, then using a Rao-Blackwellized Particle Filter to refine the estimates and infer the 6DoF poses. The tracking information is then used to detect which connections have been established and to determine the state of the assembly. Based on this state, relevant information is displayed to the system's user. We present three use cases of our system, which demonstrate this work's usefulness, generality with respect to the parts' shape and real time capabilities.

Key words

Computer vision, assembly support system, assembly modeling, connection detection, 3D tracking, MOT, Particle Filter, depth camera, colour detection, tracking-by-detection, human-computer interaction

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή Πέτρο Μαραγκό για την ευκαιρία εκπόνησης της παρούσας διπλωματικής, καθώς και για τον εξαιρετικό ενθουσιασμό και αφοσίωση που εκδηλώνει στις διαλέξεις του, οι οποίες μου κέντρισαν το ενδιαφέρον και με παρακίνησαν να ασχοληθώ με τον τομέα της όρασης υπολογιστών. Ευχαριστώ επίσης την ερευνητική του ομάδα, και ιδιαίτερα τον Βασίλη Πιτσικάλη και τον Πέτρο Κούτρα, για τις κρίσιμες συμβουλές και τη διαθεσιμότητά τους να με υποστηρίξουν καθ' όλη τη διάρκεια της εργασίας.

Θα ήθελα να εκφράσω ένα μεγάλο ευχαριστώ σε όλους τους φίλους μου, συμφοιτητές και μη, για τον ξεχωριστό ρόλο του καθενός στο να κυλήσουν τόσο ευχάριστα τα χρόνια των σπουδών μου. Τέλος, ευχαριστώ την οικογένειά μου για την κάθε λογής στήριξη που αδιάκοπα μου έχει δώσει.

Jack Hadfield,

Αθήνα, 18η Ιουλίου 2017

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
Κατάλογος σχημάτων	15
1. Εισαγωγή	17
1.1 Όραση υπολογιστών	17
1.2 Υποβοήθηση συναρμολόγησης	17
1.3 Οπτική παρακολούθηση	18
1.4 Προτεινόμενη προσέγγιση	20
1.5 Διάρθρωση της εργασίας	22
2. Παλαιότερες εργασίες	25
2.1 Οπτικά συστήματα υποστήριξης συναρμολόγησης	25
2.2 3Δ παρακολούθηση από εικόνες βάθους	26
2.2.1 Ευθυγράμμιση σημειοσυνόλων	27
2.2.2 Οπτική ροή	29
2.2.3 Αναδρομική Μπεϋζιανή Εκτίμηση	29
2.2.4 Ελαχιστοποίηση ενέργειας	30
3. Παρακολούθηση βάση ανίχνευσης σε εικόνες χρώματος	33
3.1 Χρωματικός χώρος HSV	34
3.2 Διαφορετικά εύρη χρωμάτων	35
3.2.1 Αλγόριθμος	35
3.2.2 Βαθμολογίες αξιοπιστίας	36
3.3 Όμοια εύρη χρωμάτων	36
4. 3Δ παρακολούθηση αντικειμένων	41
4.1 Αναπαράσταση πόζας	42
4.2 Φίλτρο Bayes	43
4.3 Particle Filter	45
4.4 Rao-Blackwellized Particle Filter	47
4.5 3Δ στοχαστική παρακολούθηση αντικειμένου	49
4.5.1 Δυναμικό μοντέλο	50
4.5.2 Μοντέλο παρατήρησης	51

4.5.3	Είσοδος	53
4.6	Αλγόριθμος	55
4.7	Παρακολούθηση πολλαπλών αντικειμένων	56
4.7.1	Επέκταση του Particle Tracker	57
4.7.2	Τομές αντικειμένων	58
4.7.3	Βύθιση στην επιφάνεια εργασίας	59
5.	Εργασίες συναρμολόγησης: αναπαράσταση και αξιολόγηση	61
5.1	Αναπαράσταση συναρμολόγησης	62
5.1.1	Αναπαράσταση στατικής γνώσης	62
5.1.2	Αναπαράσταση αλληλουχίας	62
5.2	Αξιολόγηση της συναρμολόγησης	65
5.2.1	Υπολογισμός σχετικής πόζας	65
5.2.2	Ισότητα ποζών	67
5.2.3	Χειρισμός συμμετριών	68
5.3	Αλλαγές στον αλγόριθμο 3Δ παρακολούθησης	71
5.3.1	Κίνητρο	71
5.3.2	Προτεινόμενη τροποποίηση	72
5.3.3	Βαθμολόγηση	73
6.	Υλοποίηση - Πειράματα και αποτελέσματα	75
6.1	Υλοποίηση	75
6.1.1	ROS	75
6.1.2	Αρχιτεκτονική του συστήματος υποστήριξης συναρμολόγησης	76
6.2	Πειράματα	77
6.2.1	1ο Πείραμα - Guidance Mode	77
6.2.2	2ο Πείραμα – Guess Mode	78
6.2.3	3ο Πείραμα – Learning Mode	79
6.2.4	Πειραματική διαδικασία	80
6.3	Αποτελέσματα	81
6.3.1	Ποιοτικά αποτελέσματα	81
6.3.2	Ποσοτικά αποτελέσματα	81
7.	Συμπεράσματα – Μελλοντική εργασία	93
7.1	Μελλοντική έρευνα	94
	Βιβλιογραφία	95
	Παράρτημα	99
	A. Ανίχνευση τραπεζιού	99
	B. Τετραδόνια	103

Κατάλογος αλγορίθμων

1	Ανίχνευση αντικειμένων διαφορετικών χρωμάτων	36
2	Παρακολούθηση αντικειμένων ίδιου χρώματος	39
3	Particle Filter resampling	48
4	6DoF παρακολούθηση ενός αντικειμένου με Particle Filter	56
5	6DoF παρακολούθηση πολλαπλών αντικειμένων με Particle Filter	60
6	Τροποποίηση του αλγορίθμου 3D παρακολούθησης για το συνυπολογισμό συνδέσεων	74

Κατάλογος σχημάτων

1.1	Βασική δομή της προτεινόμενης προσέγγισης.	21
2.1	Αλγόριθμος ICP πάνω σε δύο σημειοσύνολα.	27
2.2	Οπτική ροή.	30
3.1	Χρωματικός χώρος HSV.	34
4.1	Μαρκοβιανό μοντέλο με κρυφή κατάσταση x και παρατήρηση z	43
4.2	Μαρκοβιανό μοντέλο με μεταβλητή εισόδου u	44
4.3	Απεικόνιση της διαδικασίας του Rao-Blackwellisation	48
4.4	Μοντέλο του Μπεϋσιανού δικτύου για το πρόβλημα 3Δ παρακολούθησης	50
4.5	Βοηθητικό μοντέλο παρατήρησης.	52
5.1	Απλή δομή 5 κύβων.	62
5.2	Αναπαράσταση στατικής γνώσης του σχήματος 5.1.	63
5.3	Γράφος αλληλουχιών των εφικτών συνδέσεων του σχήματος 5.1.	64
5.4	Μια δυνατή αναπαράσταση με δυαδικό δέντρο της συναρμολόγησης του σχήματος 5.1.	65
5.5	Διαφορετική μεταχείριση κάθετου άξονα	66
5.6	Παράδειγμα δομής τριών κύβων, με δύο τρόπους σύνδεσης των κύβων που δημιουργούν ισοδύναμα αποτελέσματα.	70
5.7	Μετατόπιση του σταθμισμένου μέσου όρου των particles μετά την επιβολή τιμωρίας στις τομές αντικειμένων.	72
5.8	Μαρκοβιανό μοντέλο του σχήματος 4.2 με προσθήκη του διανύσματος συνδέσεων c	73
6.1	Αρχιτεκτονική ROS του συστήματος.	77
6.2	1ο Πείραμα: Ζητούμενη κατασκευή.	78
6.3	2ο Πείραμα: Επιτρεπτές δομές.	79
6.4	Τρία στιγμιότυπα του 1ου πειράματος σε διαφορετικά στάδια ολοκλήρωσης της κατασκευής.	82
6.5	Τρία στιγμιότυπα του 2ου πειράματος σε διαφορετικά στάδια.	83
6.6	Τρία στιγμιότυπα του 3ου πειράματος σε διαφορετικά στάδια.	84
6.7	Κατανομή συνδέσεων του 1ου πειράματος, για χρονικά περιθώρια 1-4 δευτερολέπτων.	85
6.8	1ο πείραμα: Ποσοστά σωστά ανιχνευθέντων συνδέσεων ως προς το μήκος επιτρεπτού χρονικού περιθωρίου.	86
6.9	2ο πείραμα: Μήτρες σύγχυσης για $m_p = 1\text{cm}$	88
6.10	2ο πείραμα: Μήτρες σύγχυσης $m_p = 1.8\text{cm}$	89
6.11	3ο πείραμα: Ποσοστά σωστά ανιχνευθέντων συνδέσεων ως προς το μήκος επιτρεπτού χρονικού περιθωρίου, για $m_p = 1.8\text{cm}$	90

6.12	3ο πείραμα: Ποσοστά σωστά ανιχνευθέντων συνδέσεων ως προς το μήκος επιτρεπτού χρονικού περιθωρίου, για $m_p = 1\text{cm}$	90
6.13	Κατανομή συνδέσεων του 3ου πειράματος, για περιθώρια 1-4 δευτερολέπτων, για $m_p = 1.8\text{cm}$ και $K = 5$	91
6.14	Κατανομή συνδέσεων του 3ου πειράματος, για περιθώρια 1-4 δευτερολέπτων, για $m_p = 1\text{cm}$ και $K = 3$	91
A.1	Βήματα ανίχνευσης τραπεζιού.	100

Κεφάλαιο 1

Εισαγωγή

1.1 Όραση υπολογιστών

Αυτή η διπλωματική εργασία ανήκει στον κλάδο της όρασης υπολογιστών. Η όραση υπολογιστών είναι ένας επιστημονικός τομέας που ασχολείται με την εξαγωγή, την επεξεργασία και τελικώς την κατανόηση πληροφορίας από μία ή περισσότερες εικόνες. Βαθιά επηρεασμένη από τις ικανότητες του ανθρωπίνου οπτικού συστήματος και αυτού άλλων ζώων, η όραση υπολογιστών επιχειρεί να αναπαράγει αυτές τις ικανότητες με αυτόματο τρόπο, με την ελπίδα ότι θα επιτευχθεί παρόμοια ή και καλύτερη απόδοση μέσω υπολογιστικών αλγορίθμων. Ο απώτερος σκοπός είναι να αναχθεί υψηλού επιπέδου πληροφορία από τις εικόνες, προκειμένου να κατανοηθούν στοιχεία του πραγματικού κόσμου που αυτές απεικονίζουν. Αυτή η πληροφορία μπορεί να περιλαμβάνει, μεταξύ άλλων, ετικέτες αντικειμένων ή σκηνών, θέσεις αντικειμένων, γεγονότα, δράσεις, κείμενο ή τρισδιάστατες δομές.

1.2 Υποβοήθηση συναρμολόγησης

Στην παρούσα εργασία εξετάζουμε το πρόβλημα σχεδιασμού και υλοποίησης ενός εικονικού πράκτορα υποβοήθησης συναρμολόγησης αντικειμένων, ο οποίος θα βοηθήσει έναν χρήστη σε εργασία συναρμολόγησης μέσω οπτικής επίβλεψης. Η εργασία μπορεί να αφορά τη συναρμολόγηση παιχνιδιού, μινιατούρας, επίπλου, βιομηχανικού προϊόντος κλπ. Με τεχνικές όρασης υπολογιστών, ο στόχος του συστήματος είναι να παρακολουθήσει αυτόματα την κατάσταση της συναρμολόγησης κάθε χρονική στιγμή, ώστε να προσφέρει βοήθεια στο χρήστη μέσω εντολών, ή/και να κρίνει τις δεξιότητες του χρήστη σε κάποιες εργασίες.

Επιλέξαμε να επιβλέψουμε τη συναρμολόγηση με τεχνικές οπτικής παρακολούθησης αντικειμένων. Πριν παρουσιάσουμε την προσέγγιση που ακολουθήσαμε, κάνουμε μια αναφορά στο βασικό πρόβλημα της παρακολούθησης από τη σκοπιά της όρασης υπολογιστών καθώς και στις βασικές κατηγοριοποιήσεις των αλγορίθμων παρακολούθησης που υπάρχουν.

1.3 Οπτική παρακολούθηση

Η οπτική παρακολούθηση αφορά τον εντοπισμό ενός ή περισσότερων αντικειμένων σε μια σκηνή μέσα από μια αλληλουχία εικόνων που τα περιλαμβάνει. Προαιρετικά μπορεί να ασχολείται και με την εκτίμηση του προσανατολισμού των αντικειμένων ως προς κάποιο σταθερό σύστημα αναφοράς. Η παρακολούθηση διαφέρει από την ανίχνευση στο γεγονός ότι συνήθως συνυπολογίζει τις παλαιότερες θέσεις των αντικειμένων, θεωρώντας ότι υπάρχει συνέχεια στις τροχιές τους. Η ανίχνευση, από την άλλη, εργάζεται σε κάθε πλαίσιο ξεχωριστά και δε λαμβάνει υπόψη την κίνηση των αντικειμένων.

Λόγω του μεγάλου αριθμού διαφορετικών αλγορίθμων παρακολούθησης που έχουν προταθεί στη βιβλιογραφία, η επιλογή ενός κατάλληλου αλγορίθμου κατά το σχεδιασμό μιας εφαρμογής μπορεί από μόνη της να είναι δύσκολη διαδικασία. Παρουσιάζεται παρακάτω μια σειρά παραγόντων που πρέπει να εξεταστούν. Η λίστα δεν είναι εξαντλητική, αλλά περιλαμβάνει τις βασικές σχεδιαστικές επιλογές που θα πρέπει να ληφθούν.

- **Είδος εικόνας:** Οι εικόνες που χρησιμοποιούνται για την παρακολούθηση μπορεί να είναι γκριζες, χρώματος ή βάθους. Κάθε τύπος εικόνας προσφέρει διαφορετική πληροφορία για τη σκηνή μπροστά στην κάμερα, και μπορεί να αξιοποιηθεί διαφορετικά κατά την παρακολούθηση ενός αντικειμένου. Ορισμένοι αλγόριθμοι αξιοποιούν παραπάνω από ένα είδος εικόνας.
- **2Δ ή 3Δ:** Η οπτική παρακολούθηση μπορεί να δώσει δισδιάστατες ή τρισδιάστατες πόζες, ανάλογα με το σκοπό της. Στις δύο διαστάσεις, η θέση του αντικειμένου περιγράφεται με δύο τιμές, και ο προσανατολισμός του με μια τρίτη. Στις τρεις διαστάσεις, η θέση περιλαμβάνει τρεις μεταβλητές και ο προσανατολισμός άλλες τρεις. Ωστόσο, δεν είναι πάντα απαραίτητες οι μεταβλητές στροφής. Η δισδιάστατη παρακολούθηση είναι συνηθέστερη και υλοποιείται πιο εύκολα, αλλά μπορεί να είναι ανεπαρκής σε ορισμένες περιπτώσεις.
- **Με ή χωρίς σημάδια αναφοράς:** Αν τοποθετηθούν σημάδια αναφοράς πάνω σε ένα αντικείμενο μπορεί να διευκολυνθεί το πρόβλημα. Από την άλλη, μπορεί να είναι μη-πρακτικό ή και αδύνατο να γίνει αυτό σε ορισμένες περιπτώσεις. Για παράδειγμα, κατά την παρακολούθηση ανθρώπων σ' ένα πλήθος, δεν μπορούμε να επέμβουμε τοποθετώντας σημάδια πάνω σε αυτούς.
- **Online ή offline:** Ορισμένες εφαρμογές απαιτούν ρυθμούς επεξεργασίας πραγματικού χρόνου, ενώ άλλες μπορούν να εκτελεστούν offline. Ο όρος "πραγματικού χρόνου" είναι λίγο ασαφής, βέβαια, με κάποιες πηγές να αναφέρουν ρυθμό 5fps ενώ άλλες απαιτούν ρυθμό 30fps ή παραπάνω. Το βασικό χαρακτηριστικό, πάντως, είναι ότι επεξεργάζονται οι εικόνες καθώς λαμβάνονται από την κάμερα, και όχι σε κάποιον ύστερο χρόνο. Το μειονέκτημα των αλγορίθμων πραγματικού χρόνου είναι

ότι τείνουν να παράξουν λιγότερο ακριβείς εξόδους, εξαιτίας των χρονικών περιορισμών υπολογισμού. Επίσης, οι offline αλγόριθμοι μπορούν να είναι μη-αιτιατοί, δηλαδή να χρησιμοποιήσουν και μελλοντικές εισόδους για την εκτίμηση της πόζας, ενώ οι online αλγόριθμοι πρέπει πάντα να είναι αιτιατοί.

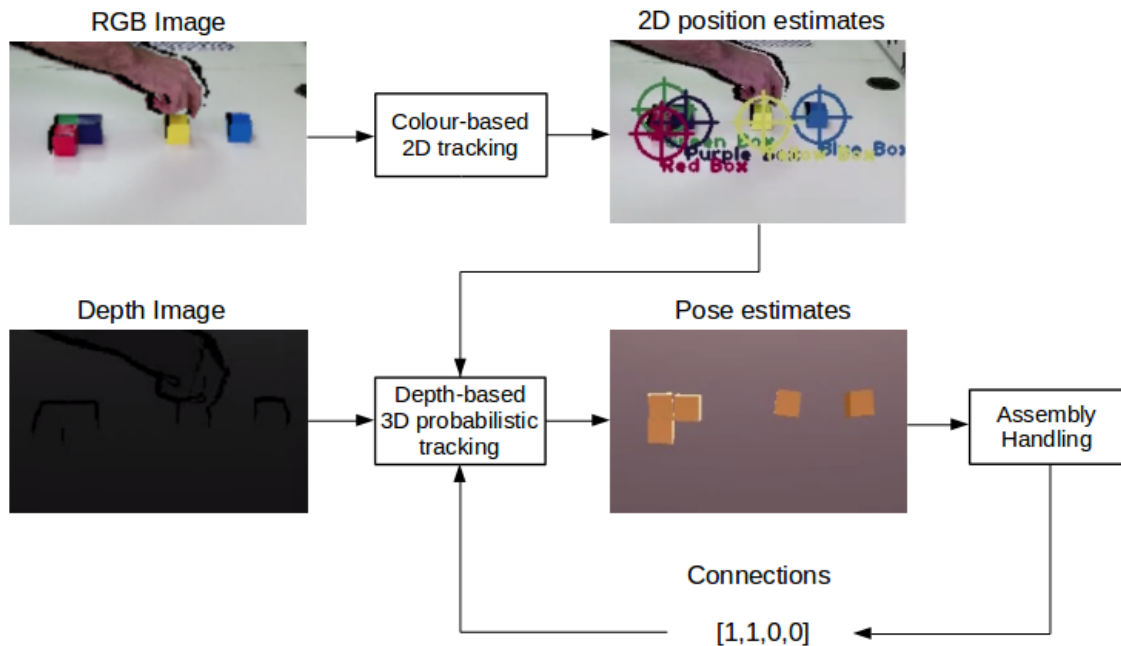
- **Επιθυμητή ακρίβεια:** Η ακρίβεια των αποτελεσμάτων αποτελεί σημαντικό παράγοντα και διαφέρει αρκετά από αλγόριθμο σε αλγόριθμο. Ακριβείς εκτιμήσεις μπορεί να είναι απαραίτητες, αλλά μπορεί και όχι, ανάλογα με τους σκοπούς της παρακολούθησης. Το κόστος της ακρίβειας έγκειται στις αυξημένες υπολογιστικές ανάγκες του αλγορίθμου.
- **Παρακολούθηση με ή χωρίς μοντέλο:** Οι αλγόριθμοι model-based είναι αυτοί που επιχειρούν να εντοπίσουν ένα προκαθορισμένο μοντέλο μέσα σε κάθε εικόνα. Στην περίπτωση της δισδιάστατης παρακολούθησης, το μοντέλο μπορεί να είναι ένα σύνολο χαρακτηριστικών, ένα συγκεκριμένο σχήμα ή οποιοδήποτε άλλο γνώρισμα που διαχωρίζει ένα αντικείμενο από την υπόλοιπη σκηνή. Στην τρισδιάστατη περίπτωση, το μοντέλο είναι συνήθως μια αναπαράσταση CAD (Computer Aided Design) ή ένα σύνολο χαρακτηριστικών που εξάγεται από μια τέτοια αναπαράσταση. Όταν το μοντέλο είναι άγνωστο, οι αλγόριθμοι είτε το κατασκευάζουν την ώρα της παρακολούθησης, είτε παρακολουθούν ένα σύνολο σημείων ενδιαφέροντος μεταξύ δύο πλαισίων.
- **Αρχικοποίηση:** Η αρχικοποίηση της θέσης ή της πόζας μπορεί να γίνει είτε αυτόματα, αξιοποιώντας κάποιον αλγόριθμο ανίχνευσης, είτε χειροκίνητα. Επιπλέον, μπορούμε να διαχωρίσουμε τους αλγορίθμους παρακολούθησης σε αυτούς που απαιτούν ακριβείς αρχικές θέσεις και αυτούς που λειτουργούν αποτελεσματικά και με προσεγγιστική αρχικοποίηση. Στη βιβλιογραφία συναντώνται, επίσης, κάποιοι αλγόριθμοι που δε χρειάζονται καμία αρχικοποίηση.
- **Στερεά ή μη-στερεά αντικείμενα:** Ένα στερεό αντικείμενο είναι ένα αντικείμενο του οποίου τα σημεία διατηρούν ανά δύο τις αποστάσεις τους στο χρόνο. Υπάρχουν αλγόριθμοι παρακολούθησης και για στερεά και για μη-στερεά αντικείμενα. Η στερεότητα απλοποιεί το πρόβλημα, διότι οι πόζες μη-στερεών αντικειμένων έχουν περισσότερες μεταβλητές που πρέπει να εκτιμηθούν.
- **Επεκτασιμότητα σε πολλαπλά αντικείμενα:** Κάποιες εφαρμογές απαιτούν την παρακολούθηση πολλών αντικειμένων. Μερικοί αλγόριθμοι είναι καταλληλότεροι από άλλους για το σκοπό αυτό, και η αποτελεσματικότητά τους εξαρτάται από την ομοιότητα και την ευκολία σύγχυσης των αντικειμένων, το βαθμό αλληλεπίδρασης μεταξύ τους και φυσικά τη θεμελιώδη μέθοδο επίλυσης τέτοιων προβλημάτων.

- **Επικαλύψεις:** Εάν περιμένουμε να παρεμποδίζεται το αντικείμενο από κάτι άλλο σε ορισμένες χρονικές στιγμές, πρέπει να επιλέξουμε αλγόριθμο ικανό να χειριστεί αυτές τις περιπτώσεις. Οι επικαλύψεις (occlusions) μπορεί να είναι μερικές, όπου ένα μέρος μόνο του αντικειμένου είναι ορατό, ή πλήρεις, όπου ολόκληρο το αντικείμενο κρύβεται πίσω από κάποιο άλλο. Οι μερικές επικαλύψεις γενικά χειρίζονται πιο εύκολα, ενώ οι πλήρεις επικαλύψεις ακόμη και αν μοντελοποιηθούν εισάγουν μια μεγαλύτερη αβεβαιότητα στη θέση του αντικειμένου, που μπορεί να εκδηλωθεί ως ολίσθηση της εκτίμησης στο χρόνο. Επιπλέον, όταν παρακολουθούνται πολλαπλά αντικείμενα, συνήθως είναι απαραίτητο να χειριστούν σωστά και οι επικαλύψεις μεταξύ τους.
- **Επαναρχικοποίηση:** Σε ορισμένες περιπτώσεις μπορεί ο αλγόριθμος να χάσει ένα αντικείμενο, επειδή βγήκε εκτός των ορίων της εικόνας, επειδή κινήθηκε γρήγορα ή εξαιτίας κάποιου σφάλματος. Ενώ κάποιοι αλγόριθμοι αδυνατούν να ξαναβρούν το αντικείμενο, άλλοι μπορεί να εκτελέσουν επαναρχικοποίηση της θέσης του όταν ικανοποιούνται συγκεκριμένα κριτήρια. Οι αλγόριθμοι παρακολούθησης μέσω ανιχνευσης, για παράδειγμα, μπορούν να αναθέσουν εκ νέου ένα παρακολουθούμενο αντικείμενο σε ένα ανιχνευμένο, εάν προηγουμένως η θέση του ήταν άγνωστη.

Όπως βλέπουμε, οι σχεδιαστικές επιλογές είναι πολυπληθείς. Συνήθως η βελτίωση της παρακολούθησης ως προς έναν παράγοντα προκαλεί υποβάθμιση ως προς άλλον. Αν, για παράδειγμα, θέλουμε τρισδιάστατη παρακολούθηση σε πραγματικό χρόνο, τότε περιμένουμε να είναι πιο ανακριβής από έναν αλγόριθμο που λειτουργεί offline. Συχνά, θέλουμε να βρούμε μια μέση λύση που θα προσφέρει αποτελεσματικότητα, ευρωστία και μειωμένες υπολογιστικές απαιτήσεις. Μια προσέγγιση που συναντάται συχνά στη βιβλιογραφία περιλαμβάνει τη σύνδεση πολλαπλών σταδίων, με την ελπίδα ότι θα συνδυαστούν τα διαφορετικά πλεονεκτήματα του κάθε σταδίου.

1.4 Προτεινόμενη προσέγγιση

Στην εργασία αυτή σχεδιάσαμε και υλοποιήσαμε ένα σύστημα επίβλεψης συναρμο-λόγησης με βασική δομή αυτή που φαίνεται στο σχήμα 1.1. Χρησιμοποιείται μια συσκευή Kinect για τη λήψη ροής εικόνων RGB και βάθους, ώστε να αποκτηθεί πληροφορία για τις ενέργειες του χρήστη. Το σύστημα βασίζεται στη συνεχή εκτίμηση των θέσεων και προσανατολισμών των μερών που απαρτίζουν το τελικό αντικείμενο, και στη σύγκριση των σχετικών θέσεων που προκύπτουν με αυτές που ορίζουν την εργασία. Αφού εξεταστεί σε κάθε πλαίσιο ποιες συνδέσεις έχουν ολοκληρωθεί, προκύπτει η τελική πληροφορία της κατάστασης της εργασίας, που μπορεί στη συνέχεια να αξιοποιηθεί για την εμφάνιση κατάλληλων μηνυμάτων στο χρήστη της εφαρμογής.



Σχήμα 1.1: Βασική δομή της προτεινόμενης προσέγγισης.

Βασικές προϋποθέσεις που έπρεπε να πληρεί το σύστημα ήταν η ευρωστία σε διάφορες συμπεριφορές χρηστών, συνθήκες φωτισμού κλπ, οι χαμηλές υπολογιστικές απαιτήσεις ώστε να λειτουργήσει σε πραγματικό χρόνο, η δυνατότητα εντοπισμού των αντικειμένων που έχουν χαθεί προσωρινά από την εικόνα και η ικανότητα χειρισμού διαφόρων μορφών κατασκευών, με συμμετρικά και μη αντικείμενα.

Ο αλγόριθμος παρακολούθησης που κατασκευάσαμε ήταν δύο σταδίων, με το πρώτο στάδιο να χρησιμοποιεί τη ροή εικόνων χρώματος του Kinect για την παραγωγή αδρών δισδιάστατων εκτιμήσεων της θέσης των αντικειμένων, και το δεύτερο να αξιοποιεί τις εικόνες βάθους καθώς και την έξοδο του προηγούμενου σταδίου για την εκτίμηση των τρισδιάστατων ποζών με σχετικά καλή ακρίβεια. Η αρχικοποίηση εκτελείται χειροκίνητα, ενώ οι επικαλύψεις χειρίζονται και στα δύο στάδια. Επίσης, και τα δύο στάδια μπορούν να θεωρηθούν model-based, καθώς τα χρώματα των αντικειμένων και τα τρισδιάστατα σχήματά τους καθορίζονται εκ των προτέρων.

Προκειμένου να απλοποιήσουμε το πρόβλημα, κάναμε κάποιες υποθέσεις για τον τρόπο χρήσης του τελικού συστήματος. Πρώτον, θεωρήσαμε ότι η συναρμολόγηση θα πραγματοποιείται πάνω σε κάποια επίπεδη επιφάνεια όπως ένα τραπέζι. Έτσι, γίνεται εύκολα η παραμετροποίηση της επιφάνειας εργασίας για τη μοντελοποίησή του. Θεωρήσαμε ότι η κάμερα θα είναι σε σταθερή θέση ως προς την επιφάνεια εργασίας, ώστε να χρειαστεί να ανιχνευθεί μόνο μία φορά κατά την εκκίνηση της εφαρμογής. Σημειώνεται ότι η θεώρηση αυτή μας επέτρεψε να δοκιμάσουμε μια απλή μέθοδο αφαίρεσης του παρασκη-νίου, αλλά αυτό αποδείχθηκε αχρείαστο στη συνέχεια, οπότε τελικώς απερρίφθη. Επίσης, υποθέσαμε πως δε θα υπάρχουν άλλα αντικείμενα στη σκηνή από αυτά που χρειάζονται

για τη συναρμολόγηση, ώστε να διευκολύνουμε τη διαδικασία της οπτικής παρακολούθησης. Σε διαφορετική περίπτωση, ενδεχομένως να υπήρχε κίνδυνος σύγχυσης των μερών της συναρμολόγησης με άλλα αχρειαστα αντικείμενα.

Υπό τις παραπάνω συνθήκες, το προτεινόμενο σύστημα διαθέτει ορισμένα πλεονεκτήματα έναντι άλλων που προτείνονται στη βιβλιογραφία. Το βασικότερο είναι η γενικότητά του ως προς το σχήμα των μερών της κατασκευής. Ενώ πολλές προσεγγίσεις υποθέτουν ότι τα μέρη θα είναι ορθογώνια, εμείς δεν κάνουμε τέτοια θεώρηση. Άλλο πλεονέκτημα είναι η δυναμική λειτουργία του συστήματος. Η κατάσταση της συναρμολόγησης εκτιμάται διαρκώς, όχι μόνο στο τέλος της. Έτσι, μπορεί να αντιδράσει στις ενέργειες του χρήστη, παραθέτοντας μηνύματα για την ορθότητα των πράξεων του και παρουσιάζοντας το επόμενο βήμα της εργασίας. Τέλος, χάρη στη διακριτή δομή του συστήματος, μπορεί να αντικατασταθεί οποιοδήποτε υποσύστημα χωρίς μεγάλη δυσκολία, ώστε ενδεχομένως να δοκιμαστούν άλλοι αλγόριθμοι σε κάθε στάδιο.

1.5 Διάρθρωση της εργασίας

- Στο Κεφάλαιο 2 παρουσιάζουμε εργασίες που σχετίζονται με την παρούσα, όσων αφορά το γενικότερο σχεδιασμό συστήματος υποβοήθησης συναρμολόγησης, αλλά και τα επιμέρους τμήματα που απαρτίζουν το τελικό μας σύστημα.
- Στο Κεφάλαιο 3 εξετάζουμε το πρόβλημα της παρακολούθησης αντικειμένων με βάση το χρώμα τους, συγκεκριμένα με τεχνικές παρακολούθησης-μέσω-ανίχνευσης. Οι αλγόριθμοι που προτείνονται συγκροτούν το πρώτο στάδιο του τελικού συστήματος.
- Στο Κεφάλαιο 4 μελετάμε μια μέθοδο τρισδιάστατης παρακολούθησης που βασίζεται στο Particle Filtering, καθώς και το γενικότερο πρόβλημα της Αναδρομικής Μπεϋζιανής Εκτίμησης. Στη συνέχεια, προτείνουμε έναν τρόπο επέκτασης του αλγορίθμου για το χειρισμό πολλαπλών αντικειμένων.
- Στο Κεφάλαιο 5 ασχολούμαστε με τις δομές αναπαράστασης των εργασιών συναρμολόγησης, καθώς και τους τρόπους εκτίμησης της κατάστασης μιας εργασίας με δοσμένες τις πόζες των επιμέρους μερών του αντικειμένου που συναρμολογείται. Έπειτα, προτείνουμε μια τροποποίηση του αλγορίθμου που παρουσιάστηκε στο Κεφάλαιο 4, ώστε να ληφθούν υπόψη και οι συνδέσεις κατά τη διαδικασία της παρακολούθησης.
- Στο Κεφάλαιο 6 αναφέρουμε τις λεπτομέρειες της υλοποίησης του συστήματος, καθώς και των πειραμάτων που διεξάγαμε για την αξιολόγηση και την ανάδειξη των πλεονεκτημάτων της προτεινόμενης προσέγγισης. Στη συνέχεια, παραθέτουμε τα αποτελέσματα των πειραμάτων αυτών, τόσο ποιοτικά, όσο και ποσοτικά.

- Στο Κεφάλαιο 7 παρουσιάζουμε τα συμπεράσματα που προέκυψαν μέσω της εργασίας αυτής και προτείνουμε κάποιες κατευθύνσεις προς τις οποίες θα ήταν χρήσιμο να κινηθεί μελλοντική έρευνα.
- Στο Παράρτημα Α περιγράφουμε μια μέθοδο ανίχνευσης της επιφάνειας του τραπεζιού πάνω στην οποία εργάζεται ο χρήστης της εφαρμογής, καθώς και ο τρόπος εξαγωγής των παραμέτρων της εξίσωσης επιπέδου που αυτή ορίζει.

Κεφάλαιο 2

Παλαιότερες εργασίες

Στο κεφάλαιο αυτό παραθέτουμε μια συνοπτική παρουσίαση της βιβλιογραφίας που σχετίζεται με την παρούσα εργασία. Σχετικά με άλλα συστήματα υποστήριξης συναρμολόγησης που βασίζονται σε τεχνικές όρασης υπολογιστών, επικεντρωνόμαστε στις βασικές μεθόδους κατανόησης σκιηνής που χρησιμοποιούνται καθώς και στις πιθανές εφαρμογές των συστημάτων που προτείνονται. Επιπλέον, εξετάζουμε διάφορες μεθόδους τρισδιάστατης παρακολούθησης που χρησιμοποιούν εικόνες βάθους.

2.1 Οπτικά συστήματα υποστήριξης συναρμολόγησης

Χάρη στην αποτελεσματικότητα της υποστήριξης συναρμολόγησης από τον υπολογιστή, μαζί με το αυξημένο ενδιαφέρον στον τομέα της Επαυξημένης Πραγματικότητας (AR), σημαντική, αν και περιορισμένη σε όγκο, έρευνα έχει διεξαχθεί στο σχεδιασμό οπτικών συστημάτων υποβοήθησης συναρμολόγησης.

Παλαιότερες εργασίες στον τομέα αυτό στηρίζονταν αποκλειστικά σε εικόνες χρώματος για την ταυτοποίηση και την παρακολούθηση διαφόρων αντικειμένων που σχετίζονται με τη συναρμολόγηση. Στην διδακτορική του διατριβή [1], ο Molineros J. κατασκεύασε έναν αυτόματο υποστηρικτή συναρμολόγησης, που παρακολουθούσε τη διαδικασία έχοντας τοποθετήσει σημάδια αναφοράς σε κάθε αντικείμενο. Στο [2] προτείνεται ένα εικονικό διαδραστικό εργαλείο για την καθοδήγηση της συναρμολόγησης. Συγκεκριμένα, ένα αντικείμενο με μορφή στυλογράφου παρακολουθείται χρησιμοποιώντας ένα νευρωνικό δίκτυο Restricted Coulomb Energy (RCE). Το αντικείμενο αλληλεπιδρά με ένα εικονικό πάνελ στο οποίο εμφανίζονται οδηγίες στο χρήστη. Σε μια πιο πρόσφατη εργασία [3], οι ερευνητές προτείνουν μια μέθοδο για την υποστήριξη της ολοκλήρωσης ενός επίπεδου παζλ πολυγωνικών κομματιών. Τα μέρη αναγνωρίζονται χρησιμοποιώντας μια τεχνική ταυρίσματος καμπυλών που παρουσιάζεται στο [4].

Με την αύξηση της δημοτικότητας των αισθητήρων βάθους τα τελευταία χρόνια, έχουν σχεδιαστεί και υλοποιηθεί κάποια συστήματα που ενσωματώνουν την πληροφορία βάθους κατά την επίβλεψη της συναρμολόγησης. Ένα τέτοιο σύστημα παρουσιάζεται στο [5], όπου τα μέρη του αντικειμένου αναγνωρίζονται μέσω 3Δ περιγραφικών και παρακολουθούνται με τον αλγόριθμο ICP (βλέπε ενότητα 2.2.1 παρακάτω). Ωστόσο, όπως

αναφέρεται στο άρθρο, το σύστημα δεν είναι εύρωστο σε περιβάλλοντα με αποκρύψεις αντικειμένων, και κατά τη στιγμή της συγγραφής είχε δοκιμαστεί μόνο στη συναρμολόγηση αντικειμένου που αποτελείται δύο μέρη.

Υπό την υπόθεση ότι τα σχήματα των μερών θα είναι ορθογώνια πολύεδρα, κατασκευάστηκαν ορισμένα συστήματα που εκτελούν voxelization του εικονικού χώρου εργασίας. Στο [6] και στη σχετική διδακτορική διατριβή [7], τα μέρη της συναρμολόγησης ήταν τουβλάκια Duplo[®] 2×4 . Συνεπώς, το μοντέλο της συναρμολόγησης θεωρείται ότι ενυπάρχει στον voxelized χώρο, με μέγεθος voxel ίσο με τις διαστάσεις ενός τουβλακιού 1×1 . Επίσης, το γεγονός ότι τα τουβλάκια είναι ορθογώνια καθιστά ευκολότερο τον υπολογισμό της αρχικής στροφής του μοντέλου. Επειδή τα κάθετα διανύσματα σε κάθε σημείο του μοντέλου θα είναι όλα παράλληλα ή κάθετα το ένα προς το άλλο, προκύπτουν τρεις βασικές διευθύνσεις καθέτων, που περιορίζουν τους δυνατούς προσανατολισμούς σε 24. Απ' αυτούς, επιλέγεται ο καλύτερος.

Ομοίως, στο [8] χρησιμοποιούνται τουβλάκια LEGO[®], και ο χώρος τμηματοποιείται σύμφωνα με τις διαστάσεις τους. Αξιοποιείται ο αλγόριθμος Lattice-First [9] για την παρακολούθηση των τουβλακιών, ο οποίος βρίσκει τους προσανατολισμούς των κυρίαρχων καθέτων και στη συνέχεια χωρίζει το χώρο σε voxels, για να βρει ποια voxels περιέχουν τουβλάκι και ποια όχι.

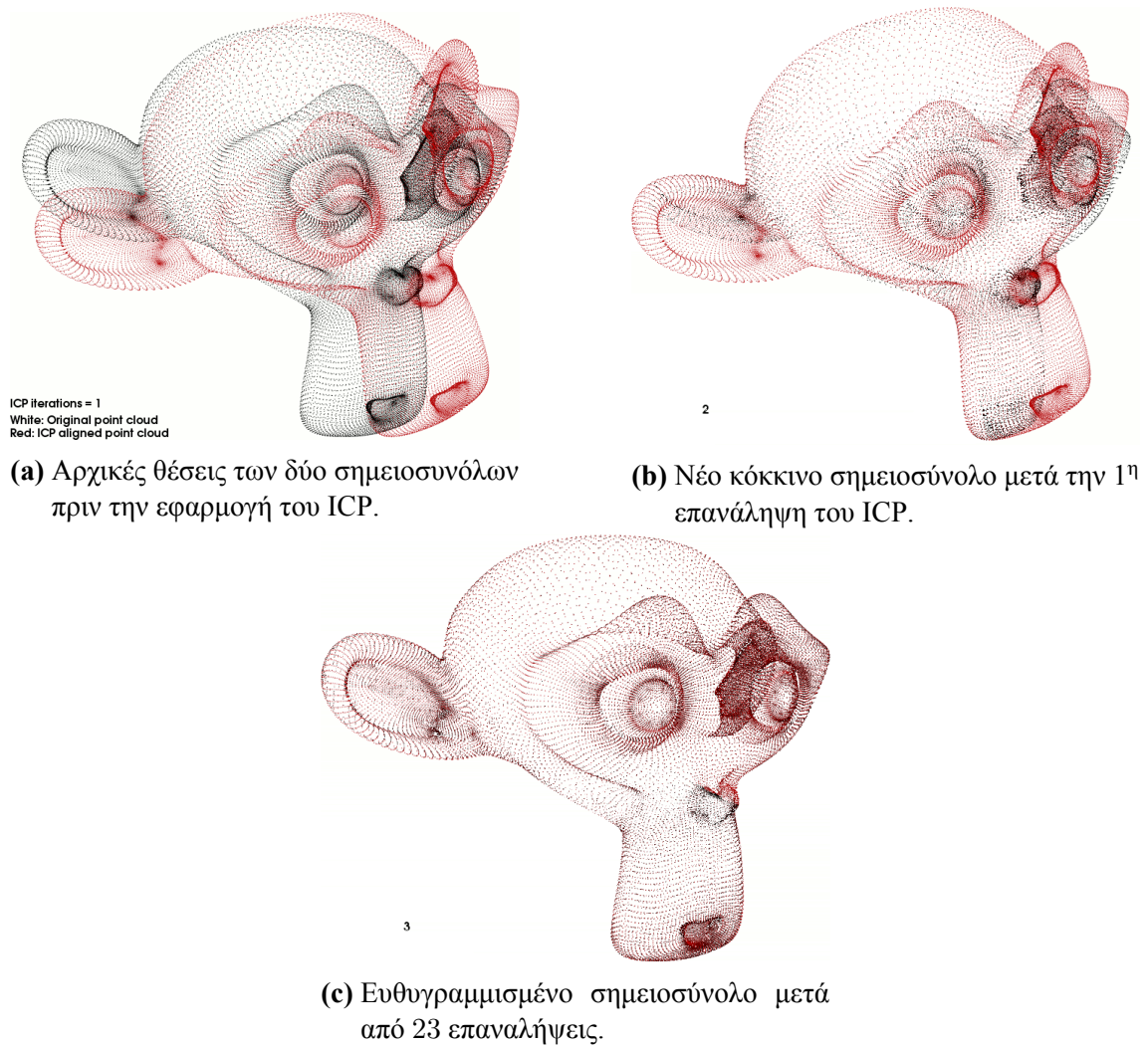
2.2 3Δ παρακολούθηση από εικόνες βάθους

Εξαιτίας της διαθεσιμότητας φθηνών καμερών βάθους όπως το Kinect της Microsoft, το Asus Xtion και το Intel RealSense, πολλές πρόσφατες εργασίες επικεντρώνονται στην ανάπτυξη αλγορίθμων παρακολούθησης που βασίζονται σε εικόνες βάθους. Οι αισθητήρες βάθους προσφέρουν πλούσια τρισδιάστατη πληροφορία που μπορεί να αξιοποιηθεί για την εξαγωγή της θέσης και προσανατολισμού των αντικειμένων στο χώρο. Έχουν, επίσης, το πλεονέκτημα ότι λειτουργούν εξίσου καλά σε διάφορες συνθήκες φωτισμού, αποδίδοντας έτσι μια μεγαλύτερη ευρωστία στους αλγορίθμους. Συγκριτικά με τα συστήματα τύπου motion capture, οι κάμερες βάθους είναι μη επεμβατικές, και επειδή δεν απαιτούν τον ίδιο όγκο εξοπλισμού, είναι φθηνότερες.

Οι περισσότεροι αλγόριθμοι εξαρτώνται από μία ή περισσότερες από τις παρακάτω θεμελιώδεις τεχνικές:

- Ευθυγράμμιση Σημειοσυνόλων (Point Set Registration - PSR)
- Οπτική Ροή
- Αναδρομική Μπεϋζιανή Εκτίμηση
- Λοιπές μέθοδοι ελαχιστοποίησης ενέργειας

2.2.1 Ευθυγράμμιση σημειοσυνόλων



Σχήμα 2.1: Αλγόριθμος ICP πάνω σε δύο σημειοσύνολα. Το κόκκινο σύνολο είναι η πηγή και το μαύρο ο στόχος. (Πηγή εικόνων: http://pointclouds.org/documentation/tutorials/interactive_icp.php)

Ο στόχος των αλγορίθμων ευθυγράμμισης σημειοσυνόλων είναι η εύρεση ενός συμπαγούς μετασχηματισμού που θα ταιριάζει βέλτιστα ένα σύνολο τρισδιάστατων σημείων ή σημειοσύνολο (η **πηγή**) σε ένα άλλο (ο **στόχος**). Ένα τέτοιο σύνολο σημείων ονομάζεται και **νέφος σημείων**. Η πηγή προέρχεται από το πιο πρόσφατο στιγμιότυπο της σκηνής, ενώ ο στόχος μπορεί είτε να προέρχεται από την προηγούμενη σάρωση, είτε από κάποιο μοντέλο CAD ενός αντικειμένου ή μιας σκηνής. Η πρώτη περίπτωση συναντάται όταν ο στόχος της ευθυγράμμισης είναι η κατασκευή αναπαράστασης μιας σκηνής ή ενός αντικειμένου, δοσμένης μιας σειράς μερικών σαρώσεων. Η δεύτερη περίπτωση χρησιμοποιείται όταν θέλουμε να βρούμε την πόζα ενός αντικειμένου γνωστού σχήματος μέσα στη σκηνή, που αφορά, δηλαδή, την τρισδιάστατη παρακολούθηση.

Ο πιο δημοφιλής αλγόριθμος PSR είναι ο επαναληπτικός αλγόριθμος κοντινότερου

σημείου (**Iterative Closest Point - ICP**), που παρουσιάστηκε πρώτη φορά στα άρθρα [10] και [11]. Ο αλγόριθμος αποτελείται από τρία βήματα, τα οποία εκτελούνται επαναληπτικά έως ότου ικανοποιηθεί κάποιο κριτήριο τερματισμού: Ταίριασμα κάθε σημείου του σημειοσυνόλου της πηγής με το εγγύτερο του συνόλου του στόχου, εύρεση της μετατόπισης και της στροφής που ελαχιστοποιούν τη μετρική του μέσου τετραγωνικού σφάλματος και μετασχηματισμός των σημείων της πηγής σύμφωνα με την υπολογισμένη μετατόπιση και στροφή.

Πολλές παραλλαγές του ICP έχουν προταθεί για τη βελτίωση του αρχικού αλγορίθμου. Ορισμένες από αυτές συγκρίνονται στο [12] ως προς την ταχύτητα σύγκλισής τους. Οι αλγόριθμοι κατηγοριοποιούνται ανάλογα με το πώς δειγματοληπτούνται τα αρχικά νέφη σημείων, πώς γίνεται η ταύτιση των σημείων, αν χρησιμοποιείται κάποια τεχνική επιβολής βαρών ή/και απόρριψη ζευγών, ποια μετρική σφάλματος ελαχιστοποιείται και πώς εκτελείται η ελαχιστοποίηση. Σε μια πιο πρόσφατη εργασία [13], παρουσιάζεται ένα πρωτόκολλο σύγκρισης διαφόρων παραλλαγών του ICP, και με αυτό συγκρίνονται ορισμένες από τις δημοφιλέστερες παραλλαγές, ανανεώνοντας έτσι τα αποτελέσματα του [12].

Στις εναλλακτικές του ICP συμπεριλαμβάνονται η Συνάρτηση Προσημασμένης Απόστασης (**Signed Distance Function**), που προτάθηκε για πρώτη φορά στο άρθρο [14], το **Coherent Point Drift (CPD)** [15], που χρησιμοποιεί Μοντέλο Μίγματος Γκαουσιανών για την εκτέλεση στοχαστικού PSR, το **Robust Point Matching (RPM)** [16], που εκτελεί ντετερμινιστική ανόπτηση ενώ χρησιμοποιούνται συνεχείς τιμές αντιστοιχίας μεταξύ των σημείων αντί δυαδικών τιμών, και μέθοδοι που βασίζονται στην **συσχέτιση** (correlation-based), όπως στο [17], όπου το PSR επιτυγχάνεται μεγιστοποιώντας ένα μέτρο συσχέτισης πυρήνα.

Η παρακολούθηση με βασισμένη στο PSR έχει τα εξής μειονεκτήματα:

- Συνήθως απαιτείται μια καλή αρχική εκτίμηση. Αυτή μπορεί να είναι δύσκολο να βρεθεί, ειδικά αν το αντικείμενο κινείται γρήγορα από το ένα πλαίσιο στο επόμενο.
- Σε πολλές εκδοχές δε λογαριάζονται οι επικαλύψεις. Μια μερική επικάλυψη μπορεί να είναι ή να μην είναι προβληματική, αλλά οι ολικές επικαλύψεις σχεδόν πάντα οδηγούν τον αλγόριθμο σε αποτυχία, αφού δεν υπάρχουν έγκυρα σημεία να ταιριαστούν στο μοντέλο CAD. Υπάρχουν, βέβαια, παραλλαγές που λαμβάνουν υπόψη μόνο τα σημεία που βρίσκονται κοντά στην προηγούμενη εκτίμηση, που κατά κάποιο τρόπο χειρίζονται τις επικαλύψεις. Παρ' όλ' αυτά, υπάρχει πάντα ο κίνδυνος ότι θα ταιριαστούν σημεία από το παρασκήνιο σε περίπτωση που χάνεται το αντικείμενο.
- Όταν πολλαπλά αντικείμενα βρίσκονται κοντά το ένα στο άλλο, το μοντέλο CAD συχνά θα αντιστοιχηθεί στο λάθος αντικείμενο μέσα στο σημειοσύνολο.

- Επειδή λαμβάνεται υπόψη μόνο το ταίριασμα των σημειοσυνόλων και όχι η κινηματική των αντικειμένων, οι εκτιμήσεις δεν είναι τόσο ομαλές όσο αυτές που παράγουν άλλες μέθοδοι. Ειδικά σε περιπτώσεις συμμετρικών αντικειμένων, ο εκτιμώμενος προσανατολισμός παρουσιάζει μεγάλη ασυνέχεια λόγω των πολλαπλών τρόπων ταιριάσματος των σημείων.

2.2.2 Οπτική ροή

Η οπτική ροή είναι η κατανομή των εμφανών ταχυτήτων μεταξύ δύο εικόνων. Εκφράζεται ως διανυσματικό πεδίο πάνω στο εύρος των εικόνων. Το πεδίο αυτό μπορεί να είναι είτε δισδιάστατο, όταν προσδιορίζεται η κίνηση κάθε pixel μεταξύ δύο εικόνων, είτε τρισδιάστατο, όταν εκφράζεται η κίνηση των voxels. Η οπτική ροή μπορεί να υπολογιστεί πάνω σε όλα τα pixels των εικόνων, ή σε κάποιο υποσύνολό τους, επιλεγμένα ώστε να περιγράφουν επαρκώς την εμφανή κίνηση μεταξύ των εικόνων (πχ. σε ακμές, γωνίες, σημεία ενδιαφέροντος κλπ.). Συνδυάζοντας ένα σύνολο τοπικών ταχυτήτων παρμένων από το διανυσματικό πεδίο, μπορούμε να εκτιμήσουμε την κίνηση ενός αντικειμένου από το ένα πλαίσιο στο επόμενο. Δημοφιλείς αλγόριθμοι για τον υπολογισμό του πεδίου οπτικής ροής αποτελούν ο Lucas-Kanade [18] και ο Horn-Schunck [19], μαζί με τις πολλές παραλλαγές τους που έχουν προταθεί στη βιβλιογραφία.

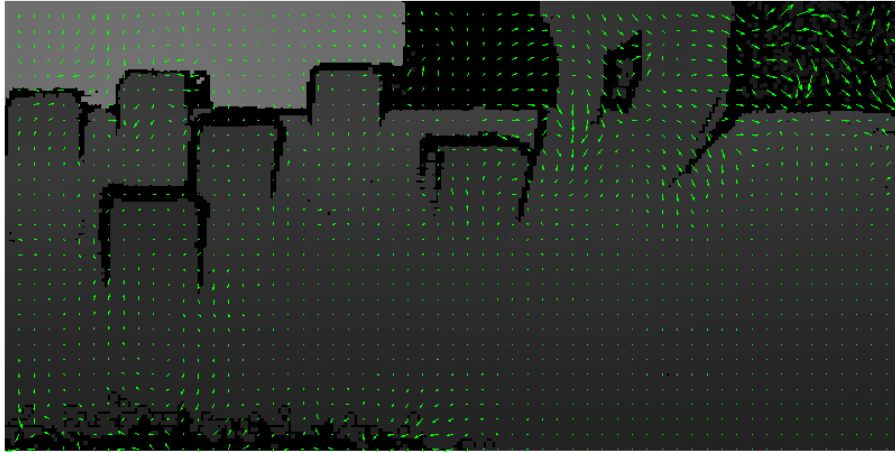
Τυπικά χρησιμοποιούνται εικόνες χρώματος ή έντασης στην οπτική ροή. Ωστόσο, μπορεί να υπολογιστεί και πάνω σε αλληλουχία εικόνων βάθους με τον ίδιο τρόπο. Επίσης, μπορεί να υπολογιστεί και η τρισδιάστατη οπτική ροή, μιας και προσφέρεται πληροφορία για την τρίτη διάσταση. Η τρισδιάστατη οπτική ροή που προέρχεται από μια σειρά εικόνων βάθους ονομάζεται στην αγγλική ορολογία **range flow**. Στο [20] παρουσιάζεται μια μέθοδος εκτίμησής της.

Η οπτική ροή μπορεί να χρησιμοποιηθεί άμεσα για την εκτίμηση της κίνησης ενός αντικειμένου, ή ως πρώτο στάδιο ενός αλγορίθμου παρακολούθησης, όπου οι εκτιμήσεις βελτιώνονται μέσω άλλων μεθόδων. Ένα παράδειγμα της δεύτερης περίπτωσης φαίνεται στο [21], όπου η πληροφορία ροής δίνεται σε ένα Εκτεταμένο Φίλτρο Kalman, ώστε να παραχθεί ακριβής εκτίμηση πόζας.

Μολονότι η οπτική ροή αποτελεί ισχυρό εργαλείο, παρουσιάζει κάποια μειονεκτήματα όταν χρησιμοποιείται μόνο αυτή στην παρακολούθηση. Οι περισσότεροι αλγόριθμοι είναι απαιτητικοί υπολογιστικά, γεγονός που συχνά τους καθιστά ακατάλληλους για παρακολούθηση πραγματικού χρόνου. Επίσης, χρησιμοποιώντας μόνο την οπτική ροή, αγνοούμε πιθανές επικαλύψεις αντικειμένων που μπορεί να υπάρχουν.

2.2.3 Αναδρομική Μπεϋζιανή Εκτίμηση

Ένας αριθμός πρόσφατων εργασιών επικεντρώνονται στη μοντελοποίηση της κίνησης των αντικειμένων με Κρυφό Μαρκοβιανό Μοντέλο, θεωρώντας τις θέσεις και προ-



Σχήμα 2.2: Οπτική ροή.

σανατολισμούς ως μη-παρατηρήσιμες καταστάσεις και την πληροφορία που προέρχεται από τις εικόνες χρώματος και βάθους ως μετρήσεις. Στην Αναδρομική Μπεύζιανή Εκτίμηση ο σκοπός είναι να εκτιμηθούν οι κρυφές καταστάσεις, δεδομένων των τωρινών και προηγούμενων μετρήσεων. Μια λεπτομερέστερη ανάλυση της Αναδρομική Μπεύζιανή Εκτίμησης παρουσιάζεται στην ενότητα 4.2.

Στο [22] οι μετρήσεις περιέχουν πληροφορία σχήματος και σιλουέτας, υπολογισμένη από τα δεδομένα RGB-D, ενώ λύνεται το πρόβλημα της εκτίμησης κατάστασης με ένα Unscented Kalman Filter. Όπως αναφέρθηκε προηγουμένως, οι αλγόριθμοι οπτικής ροής μπορούν επίσης να προσφέρουν ένα σύνολο μετρήσεων, που έπειτα μπορούν να χρησιμοποιηθούν για την εκτίμηση των κρυφών καταστάσεων, όπως στο [21].

Σε άλλες εργασίες οι εικόνες βάθους χρησιμοποιούνται άμεσα ως μετρήσεις, χωρίς να εξαχθεί καμία υψηλότερου επιπέδου πληροφορία εκ των προτέρων. Εξαιτίας της μη-Γκαουσιανής μορφής του θορύβου που υπάρχει στις εικόνες βάθους, οι τεχνικές τύπου Kalman Filter δεν μπορούν να εφαρμοστούν. Αντί αυτών, μια προσέγγιση Particle Filter προτάθηκε στο [23], ενώ στο [24] παρουσιάστηκε μια μέθοδος αντικατάστασης των πραγματικών μετρήσεων με εικονικές μετρήσεις ώστε να μπορεί να χρησιμοποιηθεί Gaussian Filter. Στο [25] προτείνεται τρόπος εκτίμησης ταυτόχρονα και των 3Δ θέσεων και του αριθμού των αντικειμένων από γκριζες ή έγχρωμες εικόνες, που με μικρή τροποποίηση θα μπορούσε να λειτουργήσει πάνω σε εικόνες βάθους.

2.2.4 Ελαχιστοποίηση ενέργειας

Το πρόβλημα της παρακολούθησης μπορεί να θεωρηθεί πρόβλημα ελαχιστοποίησης ενέργειας, όπου η συνάρτηση ενέργειας εξαρτάται από την απόκλιση των πραγματικών παρατηρήσεων από τις αναμενόμενες παρατηρήσεις που δίνει η εκάστοτε υπόθεση θέσης/πόζας. Ο στόχος είναι η εύρεση της εκτίμησης πόζας που θα ελαχιστοποιήσει τη

συνάρτηση αυτή. Σημειώνεται ότι οι παραπάνω μέθοδοι αποτελούν ειδικές περιπτώσεις προβλημάτων ελαχιστοποίησης. Εδώ, παρουσιάζουμε προσεγγίσεις που δεν ανήκουν στις περιπτώσεις αυτές.

Στο άρθρο [26] η παρακολούθηση βασίζεται στην κινηματική των αντικειμένων σύμφωνα με το 2ο νόμο του Νεύτωνα. Η ενέργεια που ελαχιστοποιείται περιέχει δύο όρους, έναν που αφορά τη διαφορά της αναμενόμενης εικόνας από την πραγματική και έναν για τον περιορισμό των υπολογισμένων δυνάμεων που επιδρούν στα αντικείμενα. Στο [27] η ενέργεια που περιλαμβάνει έναν όρο περιορισμού των διαφορών στην παρατηρούμενη εικόνα βάθους με την αναμενόμενη κι έναν όρο τιμωρίας των τομών των αντικειμένων. Η ελαχιστοποίηση γίνεται με Particle Swarm Optimization (PSO), οπότε η κινηματική των αντικειμένων συμπεριλαμβάνεται στη συμπεριφορά των particles.

Κεφάλαιο 3

Παρακολούθηση βάση ανίχνευσης σε εικόνες χρώματος

Ένα σημαντικό πρόβλημα στον τομέα της όρασης υπολογιστών αφορά την παρακολούθηση ενός ή παραπάνω αντικειμένων στο δισδιάστατο επίπεδο της εικόνας. Εξαιτίας της προόδου στην ανίχνευση αντικειμένων τα τελευταία χρόνια, αρκετή έρευνα στην παρακολούθηση επικεντρώνεται σε μεθόδους που βασίζονται στην ανίχνευση (**tracking-by-detection**). Η ανίχνευση περιλαμβάνει τον εντοπισμό αντικειμένων μέσα σε μια εικόνα, με βάση κάποια χαρακτηριστικά που τα διαφοροποιούν από τα υπόλοιπα στοιχεία της εικόνας, όπως το χρώμα, η υφή, το σχήμα, το μέγεθος, τα τοπικά γνωρίσματα κλπ. Ο αλγόριθμος ανίχνευσης βρίσκει πιθανές θέσεις αντικειμένων εντός της σκηνής, οι οποίες πρέπει στη συνέχεια να αντιστοιχιστούν με τα αντικείμενα που παρακολουθούνται. Έτσι, το πρόβλημα της παρακολούθησης γίνεται πρόβλημα συσχέτισης δεδομένων.

Μια τέτοια προσέγγιση μπορεί να είναι κατάλληλη για διάφορους λόγους. Η ανίχνευση μπορεί να γίνει σε διάφορα επίπεδα ακρίβειας, ανάλογα με τις υπολογιστικές απαιτήσεις της εφαρμογής. Μια υπολογιστικά φτηνή ανίχνευση είναι εφικτή εάν τα αντικείμενα που παρακολουθούνται έχουν εύκολα διακρίσιμα χαρακτηριστικά. Συχνά σε τέτοιες περιπτώσεις, τα αντικείμενα μπορούν να ανιχνευτούν σε πραγματικό χρόνο. Επίσης, μέσω μεθόδων που βασίζονται στην ανίχνευση, μπορούν να ξαναεντοπιστούν αντικείμενα που βγαίνουν εκτός σκηνής ή αποκρύπτονται, όταν αυτά εμφανιστούν πάλι στην εικόνα.

Στην εργασία αυτή εξετάσαμε αλγόριθμους παρακολούθησης βάση ανίχνευσης για την παραγωγή αδρών εκτιμήσεων θέσης, που θα δοθούν στη μονάδα 3Δ παρακολούθησης του επόμενου κεφαλαίου. Χρησιμοποιήσαμε έναν απλό ανιχνευτή χρώματος για τον εντοπισμό των πιθανών θέσεων. Ανάλογα με την διαφορετικότητα των αντικειμένων ως προς το χρώμα τους, προσαρμόζουμε αντίστοιχα τον τρόπο συσχέτισης των πιθανών θέσεων με της ζητούμενες θέσεις. Στη συνέχεια του κεφαλαίου παρουσιάζονται οι αλγόριθμοι που χρησιμοποιήθηκαν.

3.1 Χρωματικός χώρος HSV

Συχνά, θέλουμε να ξεχωρίσουμε τις συνιστώσες χρωματικού περιεχομένου από την φωτεινότητα του χρώματος. Σε διαφορετικές συνθήκες φωτισμού περιμένουμε μεγαλύτερες αλλαγές στη φωτεινότητα απ' ότι στις υπόλοιπες συνιστώσες. Συνεπώς, όταν επιχειρούμε να εντοπίσουμε κάποιο αντικείμενο με βάση το χρώμα του, εξετάζουμε κυρίως τις χρωματικές συνιστώσες και δίνουμε μικρότερη έμφαση την ένταση του χρώματος. Ο διαχωρισμός μπορεί να χρησιμεύει και για άλλους σκοπούς, όπως για την αύξηση της αντίθεσης στην εικόνα, τη διόρθωση του ιστογράμματος κλπ. Εδώ, βέβαια, μας ενδιαφέρει μόνο η πρώτη εφαρμογή.

Υπάρχουν διάφοροι χρωματικοί χώροι που το επιτυγχάνουν το διαχωρισμό αυτό, όπως οι HSV, YCbCr, YUV, Lab κ.α. Ο πιο συνηθισμένος είναι ο HSV, διότι η μετατροπή από RGB είναι εύκολη, οπότε συναντώνται πολλές υλοποιήσεις της μετατροπής σε ελεύθερο κώδικα. Εξαιτίας αυτού, επιλέξαμε να χρησιμοποιήσουμε τον χώρο HSV και στην παρούσα εργασία. Ο HSV έχει επίσης το πλεονέκτημα ότι αντιπροσωπεύει αρκετά καλά τον τρόπο που αντιλαμβάνεται ο άνθρωπος το χρώμα.

Τα αρχικά HSV προκύπτουν από τις λέξεις hue (απόχρωση), saturation (κορεσμός) και value (τιμή/ένταση). Η απόχρωση του χρώματος δηλώνει σε ποιο από τα κύρια χρώματα είναι πιο κοντά (κόκκινο, πράσινο και μπλε) και σε τι βαθμό, ο κορεσμός δείχνει πόσο χρωματικό είναι (δηλαδή πόσο απέχει από το γκρι) και η ένταση πόσο φωτεινό είναι.



(a) Μεταβολή της απόχρωσης (για $S = 1$ και $V = 1$). (b) Μεταβολή του κορεσμού (για $H = 0$ και $V = 0.5$). (c) Μεταβολή της έντασης (για $H = 0$ και $S = 1$).

Σχήμα 3.1: Χρωματικός χώρος HSV.

Οι μετατροπή από RGB σε HSV γίνεται με βάση τις σχέσεις

$$H = \begin{cases} 0 & \text{αν } M = m \\ 60^\circ \left(\frac{G-B}{M-m} \bmod 6 \right) & \text{αν } M = R \\ 60^\circ \left(\frac{B-R}{M-m} + 2 \right) & \text{αν } M = G \\ 60^\circ \left(\frac{R-G}{M-m} + 4 \right) & \text{αν } M = B \end{cases} \quad (3.1)$$

$$S = \begin{cases} 0 & \text{αν } M = 0 \\ \frac{M-m}{M} & \text{αλλιώς} \end{cases} \quad (3.2)$$

$$V = M \quad (3.3)$$

, όπου $m = \min \{R, G, B\}$ και $M = \max \{R, G, B\}$.

3.2 Διαφορετικά εύρη χρωμάτων

Αν τα αντικείμενα που αναζητούμε διαφέρουν χρωματικά σε σημαντικό βαθμό, τόσο μεταξύ τους όσο και από το παρασκήνιο, τότε το πρόβλημα γίνεται αρκετά ευκολότερο. Σε κάθε αντικείμενο αντιστοιχείται ένα διαφορετικό εύρος στον χρωματικό χώρο HSV. Το εύρος μπορεί να καθοριστεί εκ των προτέρων, ή να υπολογιστεί online επιλέγοντας χειροκίνητα την περιοχή της εικόνας που περιέχει το αντικείμενο. Στη δεύτερη περίπτωση, το εύρος επιλέγεται γύρω από τη μέση τιμή των δειγμάτων του χρώματος στην επιλεγμένη περιοχή, με πλάτος ανάλογο της τυπικής απόκλισης των δειγμάτων. Αν, δηλαδή, συμβολίσουμε $\mathbf{c}_i = (\mathbf{H}_i, \mathbf{S}_i, \mathbf{V}_i)$ το i -οστό δείγμα χρώματος, με $i \in [1, N_C]$, τότε λαμβάνουμε το εύρος

$$H_{range} = \begin{cases} [0^\circ, \mu_H + \kappa\sigma_H] \cup [\mu_H - \kappa\sigma_H, 360^\circ] & \text{αν } \mu_H \notin [\kappa\sigma_H, 360^\circ - \kappa\sigma_H] \\ [\mu_H - \kappa\sigma_H, \mu_H + \kappa\sigma_H] & \text{αλλιώς} \end{cases} \quad (3.4)$$

$$S_{range} = [\mu_S - \kappa\sigma_S, \mu_S + \kappa\sigma_S] \quad (3.5)$$

$$V_{range} = [\mu_V - \kappa\sigma_V, \mu_V + \kappa\sigma_V] \quad (3.6)$$

, όπου μ_H, μ_S, μ_V οι μέσες τιμές και $\sigma_H, \sigma_S, \sigma_V$ οι τυπικές αποκλίσεις των $\{H_i\}_{i=1}^{N_C}$, $\{S_i\}_{i=1}^{N_C}$ και $\{V_i\}_{i=1}^{N_C}$ αντίστοιχα. Ο συντελεστής κ πολλαπλασιάζεται επί την τυπική απόκλιση για τον καθορισμό του εύρους, με μεγάλη τιμή να δίνει μεγαλύτερο εύρος. Πειραματικά βρήκαμε ότι μια τιμή κοντά στο 2.5 δίνει ικανοποιητική ευρωστία σε μεταβολές της φωτεινότητας, χωρίς να συμπεριληφθούν χρώματα του παρασκηνίου.

3.2.1 Αλγόριθμος

Για κάθε αντικείμενο εκτελούμε τα ακόλουθα:

- Παράγουμε μια δυαδική εικόνα, της οποίας τα pixels τίθενται στο 1 εκεί που το χρώμα HSV είναι εντός του προκαθορισμένου εύρους, και 0 αλλού.
- Περνάμε τη δυαδική εικόνα μέσα από μορφολογικό φίλτρο για να απαλείψουμε το θόρυβο. Πρώτα, εκτελούμε μορφολογικό opening για να αφαιρέσουμε τις μικρές ομάδες pixel που προκύπτουν από το θόρυβο της αρχικής εικόνας. Μετά, εκτελούμε closing στην εικόνα, για να γεμίσουν τυχόν μικρές τρύπες που μπορεί να υπάρχουν. Επιλέξαμε τετραγωνικό μορφολογικό πυρήνα με μήκος πλευράς 4 pixels και στις δύο πράξεις.

- Βρίσκουμε τα περιγράμματα στην προκύπτουσα δυαδική εικόνα. Το περίγραμμα της περιοχής με το μεγαλύτερο εμβαδό αντιστοιχεί στην προβολή του αντικειμένου. Το κέντρο της εσώκλειστης περιοχής επιλέγεται για την επιθυμητή εκτίμηση θέσης.

Algorithm 1 Ανίχνευση αντικειμένων διαφορετικών χρωμάτων

Input: Colour image I , HSV ranges $\{(H_{range}^k, S_{range}^k, V_{range}^k)\}_{k=1}^K$ (K : number of objects)

Output: Object positions $\{\mathbf{p}^k = (x^k, y^k)\}_{k=1}^K$

- 1: Convert I to HSV colour space $\rightarrow (H, S, V)$
 - 2: **for** $k = 1 : K$ **do**
 - 3: $M_1^k \leftarrow (H \in H_{range}^k) \wedge (S \in S_{range}^k) \wedge (V \in V_{range}^k)$
 - 4: $M_2^k \leftarrow (M_1^k \circ B_1) \bullet B_2$, where B_1, B_2 are morphological masks and the operators \circ, \bullet denote morphological opening and closing respectively
 - 5: Find contours of M_2^k : $\{C_i^k\}_{i=1}^{N_k}$
 - 6: Choose contour with largest area: C_m^k , where $m = \operatorname{argmax}_i \operatorname{Area}(C_i^k)$
 - 7: $\mathbf{p}^k \leftarrow \operatorname{Center}(C_m^k)$
 - 8: **end for**
-

3.2.2 Βαθμολογίες αξιοπιστίας

Αργότερα στην εργασία αυτή, θα αποδειχθεί χρήσιμος ο υπολογισμός ενός μέτρου της αξιοπιστίας των εκτιμήσεων που λαμβάνονται στο κεφαλαίο αυτό. Υπάρχουν διάφοροι τρόποι για να γίνει αυτό. Μπορούμε να αξιοποιήσουμε πληροφορία για την ομοιότητα του σχήματος των αντικειμένων με το σχήμα του επιλεχθέντος περιγράμματος. Μπορούμε, επίσης, να υπολογίσουμε πόσο αποκλίνουν τα χρώματα των pixels εντός του περιγράμματος από το εύρος HSV.

Επιλέξαμε να δώσουμε μια απλή βαθμολογία σε κάθε εκτίμηση θέσης, σύμφωνα με τη σχέση

$$score = \frac{\max_{i=1}^N A_i}{\sum_{i=1}^N A_i} \quad (3.7)$$

, όπου το A_i είναι το εμβαδόν της εσώκλειστης περιοχής του περιγράμματος i . Με άλλα λόγια, η βαθμολογία δίνεται από το λόγο του εμβαδού της επιλεγμένης περιοχής δια το άθροισμα των εμβαδών όλων των ανιχνευθέντων περιοχών. Αν βρεθεί ένα μοναδικό περιγραμμά, τότε η εκτίμηση θέσης είναι αξιόπιστη. Αν υπάρχει πολύς θόρυβος στην εικόνα, τότε εμφανίζονται περισσότερα περιγράμματα, και κατά συνέπεια μειώνεται η αξιοπιστία.

3.3 Όμοια εύρη χρωμάτων

Όταν τα αντικείμενα που παρακολουθούνται έχουν παρόμοια ή ίδια χρώματα, τότε ο προηγούμενος αλγόριθμος είναι ακατάλληλος. Στην περίπτωση αυτή, δεν μπορούμε απλά

να επιλέξουμε το περίγραμμα με τη μεγαλύτερη εσώκλειστη επιφάνεια, αφού άλλα περιγράμματα μπορεί να αντιστοιχούν σε άλλα αντικείμενα. Ενώ υπάρχουν πολλές προσεγγίσεις στο πρόβλημα αυτό, με την καθεμία να διαφέρει ως προς την πολυπλοκότητα και την αποτελεσματικότητα, εμείς επιλέξαμε μια σχετικά απλή επίλυση, που λειτουργεί ικανοποιητικά υπό συγκεκριμένες συνθήκες. Το πλεονέκτημα της προτεινόμενης μεθόδου έναντι άλλων είναι το μικρό της υπολογιστικό κόστος, αλλά εις βάρος της ευρωστίας και της απόδοσής της. Επειδή, ωστόσο, ο σκοπός του σταδίου παρακολούθησης βάση χρώματος είναι να παρέχει αδρές εκτιμήσεις των θέσεων, η ακρίβεια δεν είναι αποφασιστικής σημασίας για την απόδοση του συνολικού συστήματος που περιγράφεται στην παρούσα εργασία.

Επειδή μοιάζουν χρωματικά τα αντικείμενα που αναζητούμε, πρέπει να λάβουμε υπόψη τη χρονική συνέχεια της τροχιάς του καθενός για να τα ξεχωρίσουμε. Η δυσκολία έγκειται στο ότι οι προβολές των αντικειμένων πάνω στην οθόνη μπορεί να επικαλύπτονται, ή μπορεί να μετακινηθούν γρήγορα από το ένα πλαίσιο στο άλλο. Για να λύσουμε το πρόβλημα κρατάμε την προηγούμενη θέση και ταχύτητα του κάθε αντικειμένου και αναζητούμε το αντικείμενο σε περιοχή κοντά σ' αυτήν που αναμένουμε να έχει βρεθεί στο νέο πλαίσιο.

Ας υποθέσουμε ότι έχουμε ένα μοναδικό εύρος HSV, εντός του οποίου ανήκουν τα χρώματα όλων των αντικειμένων που παρακολουθούνται. Όπως και στην ενότητα 3.2, παράγουμε μια δυαδική εικόνα που αντιστοιχεί στο εύρος αυτό, αλλά τώρα για όλα τα αντικείμενα ταυτόχρονα. Εκτελούμε το ίδιο μορφολογικό φιλτράρισμα με προηγουμένως. Έπειτα, βρίσκουμε όλα τα περιγράμματα που υπάρχουν στη δυαδική εικόνα. Κάθε περίγραμμα περικυκλώνει ένα **blob**. Στόχος μας είναι η αντιστοίχιση των K αντικειμένων με τα N blobs. Γενικά $K \neq N$, με κάποια blobs να αντιστοιχούν σε ένα αντικείμενο, κάποια σε πολλά αντικείμενα και άλλα σε κανένα.

Αναθέτουμε σε κάθε αντικείμενο ένα πλαίσιο οριοθέτησης (bounding box). Με δεδομένα τις προηγούμενες θέσεις $\mathbf{p}_{t-1}^k = [x_{t-1}^k, y_{t-1}^k]$, μεγέθη $\mathbf{s}_{t-1}^k = [w_{t-1}^k, h_{t-1}^k]$ (πλάτος και ύψος αντίστοιχα) και ταχύτητες $\mathbf{v}_{t-1}^k = [v_{t-1,x}^k, v_{t-1,y}^k]$, προχωρούμε ως εξής:

- Μετακινούμε τα πλαίσια οριοθέτησης σύμφωνα με τις προηγούμενες ταχύτητές τους: $\mathbf{p}_t^{k*} = \mathbf{p}_{t-1}^k + \mathbf{v}_{t-1}^k$.
- Αντιστοιχούμε τα αντικείμενα με τα blobs:
 - Αν ένα πλαίσιο οριοθέτησης περιέχει ένα μοναδικό blob, τότε το αντικείμενο που αντιπροσωπεύει αντιστοιχίζεται με το blob αυτό.
 - Αν ένα πλαίσιο περιέχει πολλά blobs, τότε επιλέγεται το blob με τη μεγαλύτερη επιφάνεια εντός του πλαισίου.
 - Αν δεν υπάρχει κανένα blob εντός του πλαισίου, τότε θεωρούμε ότι το αντικείμενο θα είναι επικαλυπτόμενο.

- Αν κάποια blobs έχουν ανατεθεί σε παραπάνω από ένα αντικείμενο, τότε τα χωρίζουμε σε μικρότερα blobs, χρησιμοποιώντας τον αλγόριθμο k-means πάνω στο σύνολο των σημείων τους. Στη συνέχεια αντιστοιχούμε τα νέα blobs με τα αντικείμενα, ανάλογα με την απόσταση του καθενός.
- Τα πλαίσια οριοθέτησης ανανεώνονται, ώστε να περιέχουν τα αντίστοιχα blobs (με ένα προαιρετικό περιθώριο σε κάθε πλευρά). Τα πλαίσια των επικαλυπτόμενων αντικειμένων παραμένουν στάσιμα, αλλά μεγαλώνουν λίγο σε μέγεθος λόγω της αβεβαιότητας που αντιπροσωπεύουν. Από τη διαδικασία αυτή προκύπτουν οι ανανεωμένες τιμές των \mathbf{p}_t^k και \mathbf{s}_{t-1}^k .
- Οι ταχύτητες ανανεώνονται σύμφωνα με τον κανόνα $\mathbf{v}_t^k = c\Delta\mathbf{p}_t^k + (1-c)\mathbf{v}_{t-1}^k$, όπου το c είναι ο ρυθμός ανανέωσης της ταχύτητας και το $\Delta\mathbf{p}_t^k$ δηλώνει τη μεταβολή της θέσης σε σχέση την προηγούμενη επανάληψη.

Algorithm 2 Παρακολούθηση αντικειμένων ίδιου χρώματος

Input: Colour image I , HSV ranges $(H_{range}, S_{range}, V_{range})$, previous bounding box positions, sizes and velocities $\{(\mathbf{p}_{t-1}^k, \mathbf{s}_{t-1}^k, \mathbf{v}_{t-1}^k)\}_{k=1}^K$

Output: Updated bounding box positions, sizes and velocities $\{(\mathbf{p}_t^k, \mathbf{s}_t^k, \mathbf{v}_t^k)\}_{k=1}^K$

```
1: Convert  $I$  to HSV colour space  $\rightarrow (H, S, V)$ 
2:  $M_1 \leftarrow (H \in H_{range}) \wedge (S \in S_{range}) \wedge (V \in V_{range})$ 
3:  $M_2 \leftarrow (M_1 \circ B_1) \bullet B_2$ , where  $B_1, B_2$  are morphological masks and the operators  $\circ, \bullet$  denote morphological opening and closing respectively
4: Find blobs of  $M_2$ :  $\{C_i\}_{i=1}^N$ 
5: Initialize sets  $T_i \leftarrow \emptyset, i = 1, \dots, N \quad \triangleright T_i$  contains the objects assigned to blob  $i$ 
6: for  $k = 1 : K$  do
7:    $\mathbf{p}_t^{k*} \leftarrow \mathbf{p}_{t-1}^k + \mathbf{v}_{t-1}^k$ 
8:   /*Calculate the approximate area of each blob within each bounding box*/
9:   for  $i = 1 : N$  do
10:     $x \leftarrow \max \{x_t^{k*}, \min \{x : x \in C_i\}\}$ 
11:     $y \leftarrow \max \{y_t^{k*}, \min \{y : y \in C_i\}\}$ 
12:     $w \leftarrow \min \{x_t^{k*} + w_{t-1}^k, \max \{x : x \in C_i\}\} - x$ 
13:     $h \leftarrow \min \{x_t^{k*} + h_{t-1}^k, \max \{y : y \in C_i\}\} - y$ 
14:     $A_i^k \leftarrow w * h$ 
15:   end for
16:   if  $A_i^k = 0 \forall i$  then
17:      $\mathbf{p}_t^k \leftarrow \mathbf{p}_t^{k*} - g \quad \triangleright g$ : growth rate for occluded objects
18:      $\mathbf{s}_t^k \leftarrow \mathbf{s}_{t-1}^k + 2g$ 
19:      $\mathbf{v}_t^k \leftarrow k_v \mathbf{v}_{t-1}^k$ 
20:   else
21:      $m \leftarrow \operatorname{argmax}_i A_i^k$ 
22:   end if
23:    $T_m \leftarrow T_m \cup k$ 
24: end for
25: for  $i = 1 : N$  do
26:   if  $|T_i| = 1$  then
27:      $k \leftarrow \text{single element of } T_i$ 
28:      $\mathbf{p}_t^k \leftarrow (\min \{x : x \in C_i\} - a, \min \{y : y \in C_i\} - a) \quad \triangleright a$ : margin
29:      $\mathbf{s}_t^k \leftarrow (\max \{x : x \in C_i\} + a, \max \{y : y \in C_i\} + a) - \mathbf{p}_t^k$ 
30:      $\mathbf{v}_t^k \leftarrow k_v \mathbf{v}_{t-1}^k + (1 - k_v)(\mathbf{p}_t^k - \mathbf{p}_{t-1}^k)$ 
31:   else if  $|T_i| > 1$  then
32:     Split  $C_i$  into  $|T_i|$  smaller blobs using k-means:  $\{C_{i,j}\}_{j=1}^{|T_i|}$ 
33:     for all  $k \in T_i$  do
34:       Choose closest  $C_{i,j}$  that hasn't already been chosen
35:        $\mathbf{p}_t^k \leftarrow (\min \{x : x \in C_{i,j}\} - a, \min \{y : y \in C_{i,j}\} - a)$ 
36:        $\mathbf{s}_t^k \leftarrow (\max \{x : x \in C_{i,j}\} + a, \max \{y : y \in C_{i,j}\} + a) - \mathbf{p}_t^k$ 
37:        $\mathbf{v}_t^k \leftarrow k_v \mathbf{v}_{t-1}^k + (1 - k_v)(\mathbf{p}_t^k - \mathbf{p}_{t-1}^k)$ 
38:     end for
39:   end if
40: end for
```

Κεφάλαιο 4

3Δ παρακολούθηση αντικειμένων

Η τρισδιάστατη οπτική παρακολούθηση, όπως αναφέρθηκε στο εισαγωγικό κεφάλαιο, αφορά την εκτίμηση της θέσης και του προσανατολισμού ενός ή περισσοτέρων αντικειμένων στον 3Δ χώρο, δοσμένης μιας σειράς εικόνων που περιέχουν την κίνηση των αντικειμένων. Οι εικόνες αυτές μπορεί να είναι εικόνες χρώματος, έντασης, βάθους κλπ.

Η 3Δ παρακολούθηση παρουσιάζει ορισμένα πλεονεκτήματα έναντι της 2Δ περίπτωσης. Σε κάποια προβλήματα η εκτίμηση της 3Δ πόζας μπορεί να είναι αναγκαία. Στην παρούσα εργασία, για παράδειγμα, απαιτούμε πληροφορία 3Δ πόζας για να ελέγξουμε αν οι συνδέσεις της συναρμολόγησης έχουν ολοκληρωθεί. Η απαίτηση αυτή προκύπτει επειδή δύο αντικείμενα μπορεί να φαίνονται συνδεδεμένα στην 2Δ εικόνα, αλλά στην πραγματικότητα να είναι σε απόσταση μεταξύ τους κατά μήκος του κάθετου προς την εικόνα άξονα. Επιπλέον, ο χειρισμός των επικαλύψεων μεταξύ δύο αντικειμένων γίνεται αρκετά ευκολότερος, διότι μπορούμε να τα διαχωρίσουμε βάση των διαφορετικών τους αποστάσεων από την κάμερα.

Βέβαια, η 3Δ παρακολούθηση εμπεριέχει ορισμένες δυσκολίες. Η τρίτη διάσταση μας δίνει συνολικά έξι μεταβλητές που πρέπει να εκτιμηθούν (τρεις μεταβλητές θέσης και τρεις στροφής) σε αντίθεση με τις τρεις στην 2Δ περίπτωση (δύο μεταβλητές θέσης και μία στροφής). Αυτό αυξάνει το υπολογιστικό κόστος των αλγορίθμων. Επίσης, η 2Δ παρακολούθηση έχει μελετηθεί πολύ εκτενέστερα από την 3Δ, επειδή είναι απλούστερη, μπορεί να γίνει αποδοτικά σε έγχρωμες εικόνες που είναι συχνότερες από άλλους τύπους εικόνων, και συχνά είναι επαρκής.

Στην παρούσα εργασία εργαστήκαμε με εικόνες βάθους από τον αισθητήρα του Kinect για την παραγωγή εκτιμήσεων πόζας. Οι εικόνες βάθους είναι γενικά κατάλληλες για την τρισδιάστατη παρακολούθηση, αφού περιέχουν πληροφορία και στους τρεις άξονες. Αν ένα αντικείμενο μετακινηθεί κατά μήκος του άξονα που είναι κάθετος στο επίπεδο προβολής, ο μόνος τρόπος ανίχνευσης της κίνησης σε εικόνες RGB είναι μέσω της μεταβολής των διαστάσεων της προβολής του αντικειμένου στην οθόνη, και σε πολύ ειδικές περιπτώσεις μέσω της αλλαγής στη φωτεινότητα λόγω ανακλαστικών ιδιοτήτων του αντικειμένου. Από την άλλη, η κάμερα βάθους διευκολύνει το πρόβλημα, αφού προσφέρει άμεσα την απόσταση του αντικειμένου από την κάμερα σε κάθε πλαίσιο.

Στο κεφάλαιο 2 παρουσιάσαμε μια επισκόπηση πρόσφατων εργασιών στην 3Δ πα-

ρακολούθηση από εικόνες βάθους. Επιλέξαμε να ακολουθήσουμε τη μέθοδο που περιγράφεται στο [23], για διάφορους λόγους, που θα ξεκαθαριστούν αργότερα στην εργασία αυτή:

- Ο αλγόριθμος παρακολούθησης είναι εύρωστος και αποδοτικός, ικανός να τρέξει σε πραγματικό χρόνο.
- Οι μερικές επικαλύψεις χειρίζονται αποτελεσματικά, ενώ οι πλήρεις επικαλύψεις προκαλούν ολίσθηση των εκτιμήσεων μόνο σε μικρό βαθμό.
- Ο αλγόριθμος είναι κατάλληλος για τον χειρισμό των δεδομένων βάθους του Kinect, επειδή τα μοντέλα που παρουσιάζονται περιέχουν θόρυβο που προσεγγίζει αρκετά καλά τον πραγματικό θόρυβο του αισθητήρα.
- Η παρακολούθηση μπορεί να επεκταθεί σε πολλαπλά αντικείμενα, εάν μοιραστούν τα particles μεταξύ των αντικειμένων.
- Συμπεριλαμβάνοντας μια είσοδο στο μοντέλο του συστήματος, τροφοδοτούμε τον αλγόριθμο με μια αδρή εκτίμηση που προέρχεται από το προηγούμενο στάδιο (κεφάλαιο 3).
- Τα particles καθιστούν εύκολη τη μετατροπή του αλγορίθμου ώστε να εφαρμοστούν οι περιορισμοί των συνδέσεων μεταξύ των αντικειμένων (ενότητα 5.3).

Στις ακόλουθες ενότητες αναφέρουμε συνοπτικά πώς μπορούν να αναπαρασταθούν οι πόζες των αντικειμένων, εισάγουμε το πρόβλημα της Αναδρομικής Μπεϋζιανής Εκτίμησης, περιγράφουμε τη λειτουργία του Particle Filter γενικά και παρουσιάζουμε τη μέθοδο που προτείνεται στο [23]. Τέλος, εξηγούμε πώς αυτή μπορεί να επεκταθεί και σε περισσότερα αντικείμενα.

4.1 Αναπαράσταση πόζας

Η θέση του αντικειμένου ως προς ένα πλαίσιο αναφοράς αναπαριστάται με ένα τρισδιάστατο διάνυσμα, που περιέχει τις συντεταγμένες x , y και z . Ο προσανατολισμός μπορεί να αναπαρασταθεί με διάφορους τρόπους, όπως με πίνακα (μήτρα) στροφής, ζεύγη άξονα-γωνίας, τετραδόνια (quaternions) και στροφές Euler. Γενικά, τρεις τιμές είναι απαραίτητες και ικανές για την περιγραφή του προσανατολισμού. Ένας πίνακας στροφής περιέχει 9 στοιχεία, αλλά επειδή πρέπει να είναι ορθογώνιος ($R^T = R^{-1}$) με ορίζουσα ίση με 1, οι βαθμοί ελευθερίας μειώνονται στους 3. Τα τετραδόνια περιέχουν 4 μεταβλητές, αλλά ο περιορισμός της μοναδιαίας νόρμας $\|q\| = 1$ πάλι μειώνει τους βαθμούς ελευθερίας στους 3. Ομοίως, τα ζεύγη άξονα και γωνίας αποτελούνται από 4 τιμές, 3 για τον άξονα

και 1 για τη γωνία, αλλά η απαίτηση ότι ο άξονας θα είναι μοναδιαίο διάνυσμα μειώνει κατά 1 τις ελεύθερες μεταβλητές.

Στο κεφάλαιο αυτό χρησιμοποιούμε την αναπαράσταση ζεύγους άξονα-γωνίας, ενώ στο κεφάλαιο 5 χρησιμοποιούμε τετραδόνια (βλέπε παράρτημα Β). Το θεώρημα στροφής του Euler δηλώνει ότι οποιαδήποτε αλληλουχία στροφών ενός στερεού αντικειμένου γύρω από ένα σταθερό σημείο ισοδυναμεί με μία στροφή γύρω από έναν σταθερό άξονα. Έστω e ο μοναδιαίος άξονας στροφής Euler και θ η γωνία στροφής. Κατασκευάζουμε το λεγόμενο διάνυσμα Euler πολλαπλασιάζοντας τη γωνία με τον άξονα:

$$\boldsymbol{\theta} = \theta e \quad (4.1)$$

Επειδή έχουμε τον περιορισμό $\|e\| = 1$, το διάνυσμα $\boldsymbol{\theta}$ εμπεριέχει όλη την πληροφορία της στροφής. Πράγματι, μπορούμε από το διάνυσμα Euler να ξαναβρούμε το διάνυσμα και τη γωνία στροφής, από τις σχέσεις:

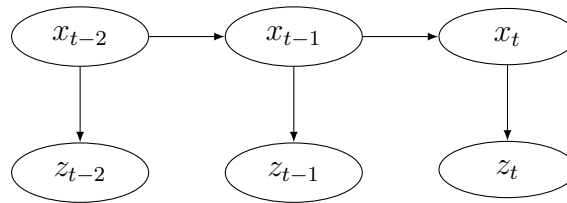
$$\theta = \|\boldsymbol{\theta}\| \quad (4.2)$$

$$e = \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|} \quad (4.3)$$

Συνδυάζοντας τις τρεις μεταβλητές θέσης με τις τρεις ανεξάρτητες μεταβλητές προσανατολισμού, κατασκευάζουμε ένα διάνυσμα έξι διαστάσεων που περιγράφει πλήρως την πόζα του αντικειμένου στον 3Δ χώρο.

4.2 Φίλτρο Bayes

Έστω ένα Κρυφό Μαρκοβιανό Μοντέλο (Hidden Markov Model - HMM), με κρυφή κατάσταση x_t και παρατηρήσιμη μέτρηση z_t σε κάθε χρονικό βήμα t . Γνωστά είναι το **δυναμικό μοντέλο** (ή μοντέλο μετάβασης) $p(x_t|x_{t-1})$ και το **μοντέλο παρατήρησης** $p(z_t|x_t)$. Εξαιτίας της υπόθεσης Markov, είναι $p(x_t|x_{t-1}) = p(x_t|x_{1:t-1})$ και $p(z_t|x_t) = p(z_t|x_{1:t})$.



Σχήμα 4.1: Μαρκοβιανό μοντέλο με κρυφή κατάσταση x και παρατήρηση z .

Το φίλτρο Bayes, γνωστό και ως **αναδρομική Μπεϋζιανή εκτίμηση** (recursive Bayesian estimation), είναι ένας αλγόριθμος εκτίμησης της συνάρτησης πυκνότητας πιθανότητας (ΣΠΠ) του x_t δεδομένων όλων των προηγούμενων μετρήσεων $z_{1:t}$, δηλαδή του

$p(x_t|z_{1:t})$. Η κατανομή εκτιμάται σε δύο βήματα: το βήμα πρόβλεψης, στο οποίο χρησιμοποιείται η προηγούμενη εκτίμηση $p(x_{t-1}|z_{1:t-1})$ και το δυναμικό μοντέλο ώστε να βρεθεί η αναμενόμενη κατάσταση, και το βήμα ανανέωσης, όπου διορθώνεται η πρόβλεψη με βάση την τελευταία μέτρηση z_t . Από τον κανόνα του Bayes, έχουμε ότι

$$p(x_t|z_{1:t}) = p(x_t|z_t, z_{1:t-1}) = \frac{p(z_t|x_t, z_{1:t-1})p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})} = \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})} \quad (4.4)$$

Η πιθανότητα αυτή είναι η πιθανότητα ανανέωσης της εκτίμησης, δεδομένης της πιο πρόσφατης μέτρησης. Επειδή ο παρονομαστής είναι σταθερός ως προς το x_t , μπορούμε να τον αγνοήσουμε στους υπολογισμούς μας, και απλά να κανονικοποιήσουμε τον όρο $p(z_t|x_t)p(x_t|z_{1:t-1})$. Ο όρος $p(z_t|x_t)$ είναι το μοντέλο παρατήρησης του συστήματος, και μπορεί να υπολογιστεί δοσμένων των z_t και x_t . Η πιθανότητα $p(x_t|z_{1:t-1})$ αντιστοιχεί στην πρόβλεψη του x_t δεδομένων των προηγούμενων τιμών του z . Αυτή η πιθανότητα μπορεί να υπολογιστεί ολοκληρώνοντας πάνω στην προηγούμενη κατάσταση x_{t-1} , χρησιμοποιώντας την εξίσωση Charman–Kolmogorov:

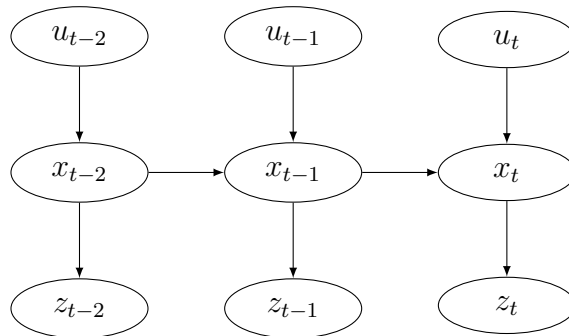
$$p(x_t|z_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|z_{1:t-1}) dx_{t-1} \quad (4.5)$$

Ο πρώτος όρος του ολοκληρώματος αντιστοιχεί στο μοντέλο μετάβασης κατάστασης. Ο δεύτερος όρος είναι η posterior κατανομή που υπολογίστηκε την προηγούμενη χρονική στιγμή $t - 1$.

Προσθέτοντας μια είσοδο u_t στο σύστημα, οι εξισώσεις (4.4-4.5) γίνονται:

$$p(x_t|z_{1:t}, u_{1:t}) = \frac{p(z_t|x_t, u_{1:t})p(x_t|z_{1:t-1}, u_{1:t})}{p(z_t|z_{1:t-1}, u_{1:t})} = \frac{p(z_t|x_t)p(x_t|z_{1:t-1}, u_{1:t})}{p(z_t|z_{1:t-1}, u_{1:t})} \quad (4.6)$$

$$p(x_t|z_{1:t-1}, u_{1:t}) = \int p(x_t|x_{t-1}, u_t)p(x_{t-1}|z_{1:t-1}, u_{1:t-1}) dx_{t-1} \quad (4.7)$$



Σχήμα 4.2: Μαρκοβιανό μοντέλο με μεταβλητή εισόδου u .

Τώρα το μοντέλο λαμβάνει υπόψη και την είσοδο. Το μοντέλο παρατήρησης, ωστόσο,

παραμένει αμετάβλητο. Αυτό σημαίνει ότι η παρουσία μιας μεταβλητής εισόδου δεν καθιστά πιο περίπλοκο το πρόβλημα του Μπεϋσιανού φιλτραρίσματος.

Εάν και το δυναμικό μοντέλο και το μοντέλο παρατήρησης είναι γραμμικά και περιλαμβάνουν Γκαουσιανό θόρυβο, τότε μια βέλτιστη αναλυτική λύση μπορεί να βρεθεί χρησιμοποιώντας ένα φίλτρο Kalman. Αν τα μοντέλα είναι γενικά μη-γραμμικά, αλλά μπορούν να θεωρηθούν γραμμικά τοπικά στην παρούσα εκτίμηση κατάστασης, τότε η λύση μπορεί να βρεθεί προσεγγιστικά χρησιμοποιώντας Γκαουσιανό φίλτρο. Συνηθισμένοι τύποι Γκαουσιανών φίλτρων είναι το Εκτεταμένο Φίλτρο Kalman (Extended Gaussian Filter - EKF) και το Unscented Kalman Filter (UKF). Στην περίπτωση Γκαουσιανού φίλτρου, η απαίτηση ότι όλοι οι θόρυβοι είναι Γκαουσιανοί παραμένει.

Αν είτε το δυναμικό μοντέλο είτε το μοντέλο παρατήρησης περιλαμβάνουν μη-Γκαουσιανό θόρυβο, τότε τα Γκαουσιανά φίλτρα δε λειτουργούν. Οι εξισώσεις (4.4), (4.5) δεν μπορούν να επιλυθούν αναλυτικά, οπότε δεν μπορεί να βρεθεί βέλτιστη αναλυτική εκτίμηση. Μια συνηθισμένη προσέγγιση σ' αυτές τις περιπτώσεις είναι η χρήση Particle Filter (PF), όπως περιγράφεται παρακάτω. Το Particle Filter είναι μια μη-παραμετρική μέθοδος που βασίζεται σε αρχές Monte Carlo για την εκτίμηση της ζητούμενης πυκνότητας πιθανότητας αριθμητικά.

4.3 Particle Filter

Υποθέτοντας ότι γνωρίζουμε τα μοντέλα μετάβασης και παρατήρησης και τις κατανομές θορύβου τους, μπορούμε να βρούμε μια προσεγγιστική λύση του προβλήματος που περιγράφηκε στην προηγούμενη ενότητα. Επιλέγουμε N δείγματα, ή particles, από την posterior συνάρτηση πυκνότητας πιθανότητας $p(x_{t-1}|z_{1:t-1}, u_{1:t-1})$, που υπολογίστηκε (κατά προσέγγιση) στο προηγούμενο χρονικό βήμα $t - 1$. Ας συμβολίσουμε τα δείγματα αυτά $\{x_{t-1}^{(i)}\}_{i=1}^N$. Εάν το N είναι επαρκώς μεγάλο, τότε

$$p(x_{t-1}|z_{1:t-1}, u_{1:t-1}) \approx \frac{1}{N} \sum_{i=1}^N \delta_{x_{t-1}^{(i)}}(x_{t-1}) \quad (4.8)$$

, όπου το $\delta_\alpha(x)$ συμβολίζει τη συνάρτηση δέλτα γύρω από το σημείο α . Το Particle Filter χρησιμοποιεί τα δείγματα για να εκτιμήσει την $p(x_t|z_{1:t}, u_{1:t})$. Το φίλτρο αποτελείται από δύο στάδια, το στάδιο πρόβλεψης και το στάδιο ανανέωσης.

Κατά το στάδιο **πρόβλεψης**, τα δείγματα προωθούνται μέσα από το μοντέλο μετάβασης κατάστασης $x_t = f(x_{t-1}, u_t, v_t)$, όπου v_t είναι ο θόρυβος του συστήματος, δειγματοληπτημένος από την αντίστοιχη ΣΠΠ. Αυτό οδηγεί σε ένα σύνολο prior δειγμάτων $\{x_t^{(i)*}\}_{i=1}^N$, με κάθε δείγμα ίσο με

$$x_t^{(i)*} = f(x_{t-1}^{(i)}, u_t, v_t^{(i)}) \quad (4.9)$$

Τα νέα αυτά δείγματα ακολουθούν προσεγγιστικά την πρότερη κατανομή του (4.7). Πράγματι,

$$\begin{aligned}
p(x_t|z_{1:t-1}, u_{1:t}) &\approx \int p(x_t|x_{t-1}, u_t) \frac{1}{N} \sum_{i=1}^N \delta_{x_{t-1}^{(i)}}(x_{t-1}) dx_{t-1} \\
&= \frac{1}{N} \sum_{i=1}^N p(x_t|x_{t-1}^{(i)}, u_t) \delta_{x_{t-1}^{(i)}}(x_{t-1}) \\
&\approx \frac{1}{N} \sum_{i=1}^N \delta_{x_t^{(i)*}}(x_t)
\end{aligned} \tag{4.10}$$

Στο στάδιο της **ανανέωσης**, υπολογίζουμε την πιθανοφάνεια κάθε πρότερου δείγματος βάση του μοντέλου παρατήρησης και του παρόντος διανύσματος μέτρησης, και αναθέτουμε σε κάθε δείγμα ένα βάρος ίσο με αυτήν την πιθανοφάνεια:

$$\tilde{w}_t^{(i)} = p(z_t|x_t^{(i)*}) \tag{4.11}$$

Έπειτα, τα βάρη χρησιμοποιούνται για την κατασκευή μιας προσέγγισης της posterior ΣΠΠ της (4.6). Σημειώνουμε ότι τα βάρη πρέπει πρώτα να κανονικοποιηθούν, ώστε να έχουν μοναδιαίο άθροισμα: $w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}$. Έτσι, η (4.6) γίνεται:

$$p(x_t|z_{1:t}, u_{1:t}) \approx \sum_{i=1}^N w_t^{(i)} \delta_{x_t^{(i)*}}(x_t) \tag{4.12}$$

Η παρούσα εκτίμηση της κατάστασης ισούται με

$$\tilde{x}_t = \sum_{i=1}^N w_t^{(i)} x_t^{(i)*} \tag{4.13}$$

Το σύνολο των particles $\{x_t^{(i)*}\}_{i=1}^N$ μπορεί στη συνέχεια να δοθεί στην επόμενη επανάληψη του αλγορίθμου.

Διαισθητικά, τα particles αντιστοιχούν σε διακριτές εκτιμήσεις της κατάστασης, τα οποία προωθούνται μέσα από το δυναμικό μοντέλο για να βρεθεί η αναμενόμενη επόμενη κατάσταση. Η πιθανοφάνεια του κάθε particle δηλώνει το βαθμό συμφωνίας της νέας μέτρησης με την αναμενόμενη μέτρηση που προκύπτει από την προβλεπόμενη κατάσταση. Η επίδραση κάθε particle εξαρτάται από την πιθανοφάνειά του, με την τελική κατάσταση να προκύπτει ως σταθμισμένο μέσο όρο όλων των προβλέψεων.

Ένα πρόβλημα που προκύπτει μετά από μερικές επαναλήψεις είναι ότι τα βάρη των particles έχουν την τάση να αποκλίνουν, που σημαίνει ότι μετά από κάποιο χρόνο μερικά δείγματα έχουν αμελητέα επίδραση στην εκτίμηση της κατάστασης. Για να αποφευ-

χθεί αυτό, όταν η απόκλιση των βαρών υπερβεί ένα άνω όριο, επιλέγονται καινούργια δείγματα. Δειγματοληπώντας εκ νέου την εκτιμώμενη posterior κατανομή, κατασκευάζουμε ένα καινούργιο σύνολο particles $\{x_t^{(i)}\}_{i=1}^N$, των οποίων τα βάρη διανέμονται πιο ομοιόμορφα. Ουσιαστικά, αντικαθιστούμε κάποια από τα particles που έχουν χαμηλά βάρη με άλλα που έχουν υψηλότερα βάρη. Η μέθοδος αυτή οδηγεί στο φίλτρο Sampling Importance Resampling (SIR), που προτάθηκε στο [28].

Ένα συνηθισμένο μέτρο για την αποτίμηση της απόκλισης των βαρών είναι η απόκλιση Kullback-Leibler (KL). Για δύο διακριτές κατανομές πιθανότητας P και Q , η απόκλιση KL ορίζεται ως

$$D_{KL}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (4.14)$$

Αν οι κατανομές ταυτίζονται, τότε $D_{KL}(P\|Q) = 0$. Διαφορετικά, μια μεγάλη τιμή του $D_{KL}(P\|Q)$ δείχνει ότι οι P και Q διαφέρουν σημαντικά.

Η νέα δειγματοληψία μπορεί να γίνει όταν η απόκλιση KL της κατανομής των βαρών από μια ομοιόμορφη κατανομή υπερβεί ένα συγκεκριμένο κατώφλι:

$$D_{KL}(P\|U) > D_{KL,max} \quad (4.15)$$

όπου U είναι η ομοιόμορφη κατανομή. Από την (4.14), έχουμε

$$\begin{aligned} D_{KL}(P\|U) &= \sum_{i=1}^N w_t^{(i)} \log \frac{w_t^{(i)}}{1/N} \\ &= \sum_{i=1}^N w_t^{(i)} \log w_t^{(i)} - \log \left(\frac{1}{N} \right) \sum_{i=1}^N w_t^{(i)} \\ &= \sum_{i=1}^N w_t^{(i)} \log w_t^{(i)} + \log N \end{aligned} \quad (4.16)$$

4.4 Rao-Blackwellized Particle Filter

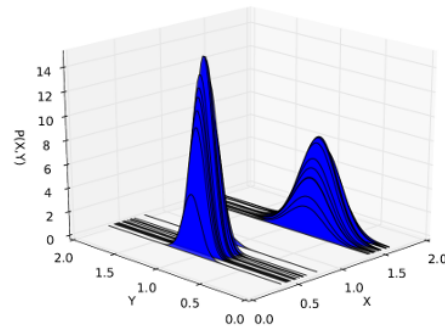
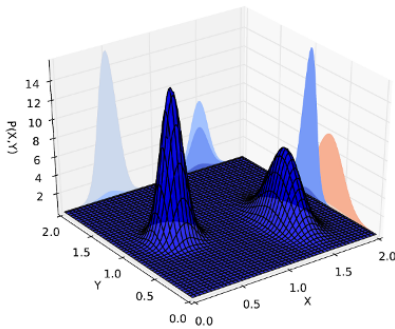
Τα Particle filters λειτουργούν καλά όταν οι κρυφές καταστάσεις είναι λίγων διαστάσεων. Όσο αυξάνονται οι διαστάσεις, ωστόσο, ο αριθμός των particles που χρειάζονται για την παραγωγή ακριβούς εκτίμησης αυξάνεται εκθετικά. Εναλλακτικά, διατηρώντας τον ίδιο αριθμό particles, η απόδοση του φίλτρου φθίνει καθώς αυξάνονται οι διαστάσεις του διανύσματος κατάστασης. Προκειμένου να αντιμετωπίσουμε το πρόβλημα αυτό, μπορούμε σε ορισμένες περιπτώσεις να αξιοποιήσουμε μια τεχνική που ονομάζεται Rao-Blackwellization, για να μειώσουμε τις διαστάσεις της κατάστασης και να βελτιώσουμε την ακρίβεια του φίλτρου. Το προκύπτον φίλτρο ονομάζεται **Rao-Blackwellized Particle Filter** (RBPF). Οι μέθοδοι που περιγράφονται στις επόμενες παράγραφοι προτάθηκαν από

Algorithm 3 Particle Filter resampling

Input: $\{x_t^{(i)*}\}_{i=1}^N, \{w_t^{(i)*}\}_{i=1}^N$ **Output:** $\{x_t^{(i)}\}_{i=1}^N, \{w_t^{(i)}\}_{i=1}^N$

```
1: Sample  $p \sim \mathcal{N}(0, 1)$ 
2:  $u \leftarrow \frac{1}{2}(1 + \text{erf}(\frac{p}{\sqrt{2}}))$ 
3: for  $i = 1 : N$  do
4:    $j \leftarrow 1$ 
5:   while  $\sum_{n=1}^j w_t^{(n)*} < u$  do
6:      $j \leftarrow j + 1$ 
7:   end while
8:    $x_t^{(i)} \leftarrow x_t^{(j)*}$ 
9:    $w_t^{(i)} \leftarrow w_t^{(j)*}$ 
10: end for
11:  $s \leftarrow \sum_{i=1}^N w_t^{(i)}$ 
12: for  $i = 1 : N$  do
13:    $w_t^{(i)} \leftarrow w_t^{(i)} / s$ 
14: end for
```

▷ Normalize weights



(a) Από κοινού συνάρτηση κατανομής των X και Y (b) Δεσμευμένες κατανομές $p(Y|X)$

Σχήμα 4.3: Απεικόνιση της διαδικασίας του Rao-Blackwellisation. Η από κοινού συνάρτηση πυκνότητας πιθανότητας αποτελείται προσεγγιστικά από δύο Γκαουσιανές. Μπορούμε να δειγματοληπτήσουμε μόνο τη διάσταση X , εκμεταλλευόμενοι την Γκαουσιανή μορφή των ΣΠΠ $p(Y|X)$. (Πηγή: [29])

τους Doucet et. al. στο άρθρο τους [30].

Το θεώρημα Rao-Blackwell δηλώνει ότι με δεδομένα έναν εκτιμητή $\delta(X)$ μιας अपαρατήρητης παραμέτρου θ από παρατηρήσιμα δεδομένα X , και μια επαρκή στατιστική $T(X)$ υπολογισμένη από τα X , ο εκτιμητής Rao-Blackwell

$$\delta_1(X) = E(\delta(X)|T(X)) \quad (4.17)$$

θα είναι τουλάχιστον το ίδιο καλός με τον $\delta(X)$. Μια στατιστική $T(X)$ είναι επαρκής εάν καμία άλλη στατιστική υπολογισμένη από τα δεδομένα X δεν προσφέρει παραπάνω πληροφορία για την άγνωστη παράμετρο θ .

Ας υποθέσουμε ότι μπορούμε να διαιρέσουμε τις κρυφές μεταβλητές x_t του μοντέλου Markov σε δύο κατηγορίες, μια κατηγορία μεταβλητών x_t^l που ακολουθούν αναλυτικά υπολογίσιμες κατανομές, και μια άλλη x_t^n που δεν μπορούν να εκτιμηθούν αναλυτικά. Ο κανόνας αλυσίδας μας δίνει

$$p(x_t|z_{1:t}, u_{1:t}) = p(x_t^l, x_t^n|z_{1:t}, u_{1:t}) = p(x_t^l|x_t^n, z_{1:t}, u_{1:t})p(x_t^n|z_{1:t}, u_{1:t}) \quad (4.18)$$

Το πρώτο μέρος μπορεί να υπολογιστεί αναλυτικά, ανάλογα με τη μορφή της ΣΠΠ. Για παράδειγμα, αν $x_t^l \sim \mathcal{N}(\mu, \sigma)$, τότε μπορούμε να χρησιμοποιήσουμε φίλτρο Kalman για τον υπολογισμό της βέλτιστης λύσης. Το δεύτερο μέρος μπορεί να βρεθεί προσεγγιστικά με κατανομή particles όπως περιγράψαμε στην προηγούμενη ενότητα:

$$p(x_t^n|z_{1:t}, u_{1:t}) \approx \sum_{i=1}^N w_t^{n(i)} \delta_{x_t^{n(i)}}(x_t^n) \quad (4.19)$$

Επομένως,

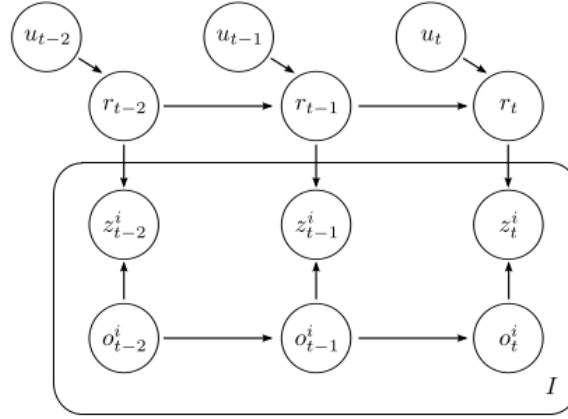
$$p(x_t|z_{1:t}, u_{1:t}) \approx \sum_{i=1}^N p(x_t^l|x_t^{n(i)}, z_{1:t}, u_{1:t})w_t^{n(i)} \delta_{x_t^{n(i)}}(x_t^n) \quad (4.20)$$

Αυτό σημαίνει πως έχουμε N ΣΠΠ κλειστής μορφής, με την καθεμία να αντιστοιχεί σε ένα βάρος $w_t^{n(i)}$. Το θεώρημα Rao-Blackwell μας λέει ότι ένας τέτοιος εκτιμητής θα είναι τουλάχιστον το ίδιο καλός με τον εκτιμητή του απλού Particle Filter με δειγματοληψία πάνω σε όλες τις διαστάσεις της κρυφής κατάστασης. Επειδή οι ΣΠΠ των x_t^l είναι γνωστές, μπορούμε να εκτιμήσουμε την κατάσταση περιθωριοποιώντας τις μεταβλητές αυτές. Οι ΣΠΠ μπορούν να περιγραφούν με μια επαρκής στατιστική. Στην περίπτωση Γκαουσιανών κατανομών, για παράδειγμα, οι μέσες τιμές και οι διασπορές αποτελούν επαρκείς στατιστικές των μεταβλητών x_t^l , που μας δίνουν την κλειστή μορφή των κατανομών.

4.5 3Δ στοχαστική παρακολούθηση αντικειμένου

Για το πρόβλημα της παρακολούθησης, η κρυφή κατάσταση x_t περιλαμβάνει την 6DoF πόζα του αντικειμένου, η παρατήρηση z_t είναι η εικόνα βάθους και η είσοδος u_t είναι ένα προαιρετικό σύνολο μεταβλητών που μπορούν να προσφέρουν μια αδρή εκτίμηση ενός ή περισσότερων μεταβλητών πόζας, πχ. οι συντεταγμένες εικόνας του αντικειμένου που δίνονται από έναν υπολογιστικά φθινό ανιχνευτή χρώματος. Επειδή η εικόνα βάθους περιέχει θόρυβο βαριάς ουράς, τα Gaussian Filters είναι ακατάλληλα. Επομένως, χρησιμοποιούμε ένα Particle Filter σαν αυτό που περιγράψαμε προηγουμένως. Οι ιδέες της ενότητας αυτής είναι αυτές που παρουσιάζονται στο άρθρο [23].

Η πόζα \mathbf{r}_t του αντικειμένου περιέχει τρεις μεταβλητές θέσης $\mathbf{p}_t = (p_{t,x}, p_{t,y}, p_{t,z})$ και τρεις μεταβλητές προσανατολισμού $\boldsymbol{\theta}_t = (\theta_{t,x}, \theta_{t,y}, \theta_{t,z})$ (του διανύσματος Euler). Επειδή



Σχήμα 4.4: Μοντέλο του Μπεϋσιανού δικτύου για το πρόβλημα της 3Δ παρακολούθησης (σχήμα από [23]).

ο προσανατολισμός δεν ανήκει σε Ευκλείδειο χώρο, αναπαριστούμε την πόζα ως διαφορική μεταβολή θέσης $\delta \mathbf{p}_t$ και προσανατολισμού $\delta \boldsymbol{\theta}_t$ από μια αρχική πόζα $\mathbf{r}_t^0 = (\mathbf{p}_t^0, \boldsymbol{\theta}_t^0)$. Το \mathbf{r}_t^0 ανανεώνεται σε κάθε επανάληψη σύμφωνα με την προηγούμενη κατάσταση. Υποθέτουμε ότι οι μεταβολές θα είναι αρκετά μικρές ώστε να θεωρήσουμε προσεγγιστικά ότι ανήκουν στον Ευκλείδειο χώρο. Επίσης, εκτιμούμε την ταχύτητα $\mathbf{v}_t = (\mathbf{v}_t, \boldsymbol{\omega}_t)$, όπου \mathbf{v}_t η γραμμική ταχύτητα και η $\boldsymbol{\omega}_t$ γωνιακή. Το τελικό διάνυσμα κατάστασης γράφεται

$$\mathbf{x}_t = (\delta \mathbf{r}_t, \mathbf{v}_t) = (\delta \mathbf{p}_t, \delta \boldsymbol{\theta}_t, \mathbf{v}_t, \boldsymbol{\omega}_t) \quad (4.21)$$

Ένα σύνολο δυαδικών μεταβλητών $o_t = \{o_t^i\}$ χρησιμοποιείται για να μοντελοποιηθούν οι επικάλυψεις, με το δείκτη i να δηλώνει το pixel. Υπάρχει επικάλυψη στο pixel i τη χρονική στιγμή t αν $o_t^i = 1$, ενώ αν $o_t^i = 0$ τότε το αντικείμενο είναι ορατό στο pixel αυτό. Οι μεταβλητές αυτές είναι επίσης μη-παρατηρήσιμες. Επομένως, η κρυφή κατάσταση περιλαμβάνει την πόζα του αντικειμένου και τις μεταβλητές επικάλυψης. Οι δύο ομάδες μεταβλητών διαφέρουν στο ότι οι μεταβλητές επικάλυψης μπορούν να περιθωριοποιηθούν, ενώ οι μεταβλητές πόζας όχι, όπως θα φανεί παρακάτω.

4.5.1 Δυναμικό μοντέλο

Αν το σύστημα ήταν ντετερμινιστικό, τότε θα μπορούσαμε να γράψουμε

$$\mathbf{r}_t = \mathbf{r}_{t-1} + k_v \mathbf{v}_{t-1} + b \mathbf{u}_t \quad (4.22)$$

$$\mathbf{v}_t = k_v \mathbf{v}_{t-1} + b \mathbf{u}_t \quad (4.23)$$

, όπου k_v είναι ο συντελεστής της ταχύτητας και b ο συντελεστής εισόδου. Το \mathbf{u}_t βασίζεται στην εκτίμηση 2Δ θέσης του αντικειμένου από την ανίχνευση χρώματος του κεφαλαίου 3 (βλέπε ενότητα 4.5.3).

Επειδή στην πραγματικότητα η ταχύτητα μεταβάλλεται με ακανόνιστο τρόπο, υποθέτουμε ότι το μοντέλο περιέχει Γκαουσιανό θόρυβο. Με την προσθήκη του θορύβου, οι εξισώσεις (4.22), (4.23) γίνονται

$$\mathbf{r}_t = \mathbf{r}_{t-1} + k_v \mathbf{v}_{t-1} + b \mathbf{u}_t + \mathbf{w}_t \quad (4.24)$$

$$\mathbf{v}_t = k_v \mathbf{v}_{t-1} + b \mathbf{u}_t + \mathbf{w}_t \quad (4.25)$$

, όπου $\mathbf{w}_t \sim \mathcal{N}(0, \Sigma)$. Επειδή είναι $\mathbf{r}_{t-1} = \mathbf{r}_{t-1}^0 + \delta \mathbf{r}_{t-1}$ και $\mathbf{r}_t = \mathbf{r}_{t-1}^0 + \delta \tilde{\mathbf{r}}_t$, οι παραπάνω σχέσεις γράφονται

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{u}_t, \mathbf{w}_t) = \begin{bmatrix} \mathbf{I}_6 & k_v \mathbf{I}_6 \\ \mathbf{0}_6 & k_v \mathbf{I}_6 \end{bmatrix} \mathbf{x}_{t-1} + b \begin{bmatrix} \mathbf{u}_t \\ \mathbf{u}_t \end{bmatrix} + \begin{bmatrix} \mathbf{w}_t \\ \mathbf{w}_t \end{bmatrix} \quad (4.26)$$

Επειδή οι μεταβλητές επικάλυψης είναι δυαδικές, το δυναμικό μοντέλο $p(o_t^i | o_{t-1}^i)$ έχει δύο παραμέτρους. Αυτές είναι

- Η πιθανότητα $p(o_t^i = 0 | o_{t-1}^i = 0)$ ενός ορατού pixel i να παραμείνει ορατό.
- Η πιθανότητα $p(o_t^i = 0 | o_{t-1}^i = 1)$ ενός μη-ορατού pixel i να γίνει ορατό.

Οι υπόλοιπες περιπτώσεις είναι συμπληρωματικές, οπότε

$$p(o_t^i = 1 | o_{t-1}^i = 0) = 1 - p(o_t^i = 0 | o_{t-1}^i = 0) \quad (4.27)$$

και

$$p(o_t^i = 1 | o_{t-1}^i = 1) = 1 - p(o_t^i = 0 | o_{t-1}^i = 1) \quad (4.28)$$

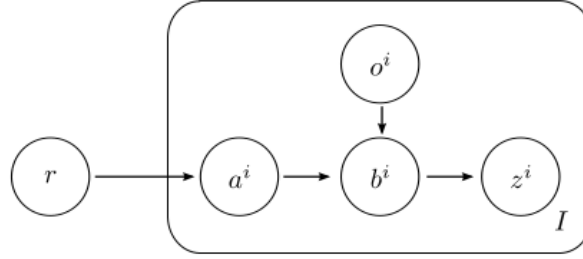
4.5.2 Μοντέλο παρατήρησης

Στο άρθρο, το μοντέλο παρατήρησης $p(z_t^i | x_t, o_t^i)$ για κάθε pixel i εκφράζεται χρησιμοποιώντας δύο βοηθητικές μεταβλητές a^i και b^i , με τρόπο που φαίνεται στο γράφο του σχήματος 4.5. Η μεταβλητή a^i είναι η απόσταση μέχρι το αντικείμενο στο σημείο που αντιστοιχεί στο pixel i , ενώ το b^i είναι η απόσταση που φαίνεται στο pixel i όταν συνυπολογίζονται οι επικαλύψεις. Αν το αντικείμενο είναι ορατό, τότε $a^i = b^i$. Διαφορετικά, $a^i > b^i$.

Οι βοηθητικές μεταβλητές εξαφανίζονται κατά την ολοκλήρωση, όταν υπολογίζεται το μοντέλο παρατήρησης. Από τις Μαρκοβιανές υποθέσεις του σχήματος 4.5 έχουμε:

$$p(z^i | r, o^i) = \int_{a,b} p(z^i | b^i) p(b^i | a^i, o^i) p(a^i | r) \quad (4.29)$$

Καθένας από τους όρους του ολοκληρώματος εκφράζεται αναλυτικά στο άρθρο.



Σχήμα 4.5: Βοηθητικό μοντέλο παρατήρησης (σχήμα από [23]).

Η απόσταση a^i στο αντικείμενο κατά μήκος της δέσμης του pixel i , δοσμένης της πόζας του αντικειμένου r , μπορεί να υπολογιστεί με τεχνικές ray tracing. Το μοντέλο του αντικειμένου πιθανώς να περιέχει σφάλματα, οπότε ενσωματώνουμε θόρυβο στην πιθανότητα:

$$p(a^i|r) = \mathcal{N}(a^i|d^i(r), \sigma_m) \quad (4.30)$$

, όπου $d^i(r)$ είναι η υπολογισμένη απόσταση στο pixel i .

Η πυκνότητα πιθανότητας της απόστασης b^i που φαίνεται στο pixel i , δοσμένης της πραγματικής απόστασης a^i και της τιμής επικάλυψης o^i , εκφράζεται ως:

$$p(b^i|a^i, o^i) = \begin{cases} \delta(b^i - a^i) & \text{if } o^i = 0 \\ I(b^i > 0 \wedge b^i < a^i) \frac{\lambda e^{-\lambda b^i}}{1 - e^{-\lambda a^i}} & \text{if } o^i = 1 \end{cases} \quad (4.31)$$

, όπου δ είναι η συνάρτηση δέλτα και I συμβολίζει τη δείκτρια συνάρτηση (indicator function) που ισούται με ένα όταν η συνθήκη είναι αληθής και μηδέν διαφορετικά. Αυτό σημαίνει ότι αν το αντικείμενο είναι ορατό στο pixel i , τότε το b^i λαμβάνει μια τυχαία τιμή μεταξύ του 0 και του a^i .

Η πραγματική μέτρηση z^i της απόστασης διαφέρει από το b^i εξαιτίας του θορύβου στην εικόνα από την κάμερα του Kinect. Ο θόρυβος αυτός δεν είναι Γκαουσιανός, αλλά μπορεί να προσεγγιστεί από το άθροισμα μιας Γκαουσιανής και μιας ομοιόμορφης κατανομής, ώστε να προστεθεί επιπλέον βάρος στις ουρές της Γκαουσιανής καμπύλης. Έτσι,

$$p(z^i|b^i) = (1 - \beta)\mathcal{N}(z^i|b^i, \sigma_c) + \beta \frac{I(z^i > 0 \wedge z^i < m)}{m} \quad (4.32)$$

, όπου το m είναι το μέγιστο βάθος που μπορεί να μετρήσει η κάμερα, το I είναι η δείκτρια συνάρτηση και το β είναι το βάρος των ουρών. Συνδυάζοντας τις (4.30), (4.31) και (4.32)

και θεωρώντας ότι $0 < z^i < m$, η εξίσωση (4.29) δίνει:

$$\begin{aligned}
p(z^i|r, o^i = 0) &= \int_{a,b} \left[(1 - \beta)\mathcal{N}(z^i|b^i, \sigma_c) + \beta \frac{I(z^i > 0 \wedge z^i < m)}{m} \right] \\
&\quad \delta(b^i - a^i)\mathcal{N}(a^i|d^i(r), \sigma_m) \\
&= \int_{-\infty}^{\infty} \left[(1 - \beta)\mathcal{N}(z^i|a^i, \sigma_c) + \frac{\beta}{m} \right] \mathcal{N}(a^i|d^i(r), \sigma_m) da^i \\
&= (1 - \beta)\mathcal{N}(z^i|d^i(r), \sqrt{\sigma_c^2 + \sigma_m^2}) + \frac{\beta}{m}
\end{aligned} \tag{4.33}$$

$$\begin{aligned}
p(z^i|r, o^i = 1) &= \int_{a,b} \left[(1 - \beta)\mathcal{N}(z^i|b^i, \sigma_c) + \frac{\beta}{m} \right] I(b^i > 0 \wedge b^i < a^i) \\
&\quad \frac{\lambda e^{-\lambda b^i}}{1 - e^{-\lambda a^i}} \mathcal{N}(a^i|d^i(r), \sigma_m) \\
&= \int_0^{a^i} \int_0^{\infty} \left[(1 - \beta)\mathcal{N}(z^i|b^i, \sigma_c) + \frac{\beta}{m} \right] \frac{\lambda e^{-\lambda b^i}}{1 - e^{-\lambda a^i}} \mathcal{N}(a^i|d^i(r), \sigma_m) da^i db^i \\
&\stackrel{(\sigma_m \simeq 0)}{=} (1 - \beta) \frac{\lambda e^{\frac{1}{2}\lambda(2d^i(r) - 2z^i + \lambda\sigma_c^2)} (1 + \operatorname{erf}(\frac{d^i(r) - z^i + \lambda\sigma_c^2}{\sqrt{2}\sigma_c}))}{2e^{\lambda d^i(r)} - 2} + \frac{\beta}{m}
\end{aligned} \tag{4.34}$$

4.5.3 Είσοδος

Η μεταβλητή εισόδου της εξίσωσης μετάβασης προσφέρει έναν τρόπο εξωτερικού ελέγχου της συμπεριφοράς των particles. Εδώ, εξηγούμε τη μορφή της εισόδου που επιλέξαμε για να ενσωματώσουμε τις 2Δ θέσεις των αντικειμένων όπως αυτές εκτιμούνται από την επεξεργασία των εικόνων χρώματος. Όπως εμφανίζεται στην εξίσωση (4.24), το διάνυσμα εισόδου έχει έξι στοιχεία, που επιδρούν πάνω στις έξι μεταβλητές πόζας. Από αυτές, χρησιμοποιούμε μόνο τις πρώτες τρεις, που επηρεάζουν τις μεταβλητές θέσης. Μηδενίζουμε τις άλλες τρεις, διότι δε διαθέτουμε πρότερη πληροφορία για τον προσανατολισμό του αντικειμένου. Επειδή ο ανιχνευτής χρώματος δίνει δισδιάστατες συντεταγμένες εικόνας, πρέπει να τις μετατρέψουμε σε συντεταγμένες του τρισδιάστατου χώρου.

Και ο αισθητήρας βάθους και η κάμερα χρώματος του Kinect είναι προοπτικές κάμερες, που σημαίνει ότι οι εικόνες που λαμβάνονται έχουν σχηματιστεί μέσω **προοπτικής προβολής** της τρισδιάστατης σκηνής στο επίπεδο της εικόνας. Ένα σημείο $P = (X, Y, Z)$ στον τρισδιάστατο χώρο προβάλλεται στο σημείο $p = (x, y)$ της εικόνας, με τρόπο τέτοιο, ώστε:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X/Z \\ Y/Z \\ 1 \end{bmatrix} \tag{4.35}$$

, όπου (c_x, c_y) είναι το κύριο σημείο (principal point) και f_x, f_y είναι οι εστιακές αποστά-

σεις κατά τους άξονες x και y . Ο πίνακας

$$C = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4.36)$$

είναι η **μήτρα κάμερας** της προβολής και οι τιμές c_x, c_y, f_x, f_y είναι οι **εσωτερικές παράμετροι** της κάμερας.

Λύνοντας ως προς X, Y , λαμβάνουμε:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = C^{-1} \begin{bmatrix} xZ \\ yZ \\ Z \end{bmatrix} = \begin{bmatrix} 1/f_x & 0 & -c_x/f_x \\ 0 & 1/f_y & -c_y/f_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} xZ \\ yZ \\ Z \end{bmatrix} \quad (4.37)$$

Συνεπώς, μπορούμε να υπολογίσουμε τη θέση ενός σημείου, δεδομένων των συντεταγμένων του στο επίπεδο της εικόνας και την απόστασή του από την κάμερα.

Έστω $\mathbf{p} = [p_x, p_y]$ οι συντεταγμένες εικόνας που δίνει ο αλγόριθμος παρακολούθησης βάση χρώματος και s η βαθμολογία αξιοπιστίας της σχέσης (3.7). Χρησιμοποιώντας τη σχέση (4.37), υπολογίζουμε τις πρώτες δύο μεταβλητές εισόδου u_x, u_y ως εξής:

$$u_x = \left(\frac{(p_x - c_x)r_z}{f_x} - r_x \right) * s \quad (4.38)$$

$$u_y = \left(\frac{(p_y - c_y)r_z}{f_y} - r_y \right) * s \quad (4.39)$$

Με άλλα λόγια, υπολογίζουμε την 3Δ θέση που αντιστοιχεί στο σημείο μέσω της σχέσης (4.37) και στη συνέχεια αφαιρούμε την παρούσα εκτιμώμενη θέση. Στο σημείο αυτό φαίνεται η χρησιμότητα της βαθμολογίας αξιοπιστίας των 2Δ εκτιμήσεων. Εάν το s είναι μεγάλο (κοντά στο 1), οι μεταβλητές εισόδου θα έχουν σχετικά μεγάλη επίδραση στο δυναμικό μοντέλο. Αντιθέτως, μια μικρή τιμή αξιοπιστίας θα μειώσει την επίδραση των μεταβλητών εισόδου, οπότε η εκτίμηση της πόζας στο στάδιο της πρόβλεψης θα βασιστεί κυρίως στην έως τώρα κίνηση του αντικειμένου και όχι στην ανίχνευση χρώματος.

Η τρίτη μεταβλητή εισόδου u_z μπορεί να ληφθεί από την πιο πρόσφατη εικόνα βάθους. Μπορούμε να χρησιμοποιήσουμε την τιμή βάθους $z = d^p$ του pixel \mathbf{p} και προαιρετικά αυτή των γειτονικών pixels ($z = \frac{\sum_{(i: \|i-p\| \leq R)} d^i}{N}$, όπου N ο αριθμός των σημείων που λαμβάνονται και R η ακτίνα). Τότε

$$u_z = (z - r_z) * \lambda_z * s \quad (4.40)$$

, όπου $0 \leq \lambda_z \leq 1$ είναι ένας περιοριστικός παράγοντας που αντιπροσωπεύει τη μεγαλύτερη αβεβαιότητα κατά τον άξονα z . Βέβαια, πειραματικά βρέθηκε αχρείαστος ο παράγο-

ντας αυτός, καθώς και για $\lambda_z = 1$ είχαμε ικανοποιητικά αποτελέσματα. Βρήκαμε ότι στην περίπτωση που το αντικείμενο δεν μετακινείται πολύ κατά τον κάθετο άξονα, μπορούμε να θέσουμε το z σε μια σταθερή τιμή, σύμφωνα με τη μέση αναμενόμενη απόσταση του αντικειμένου, αρκεί να μειώσουμε αντίστοιχα την τιμή του λ_z .

4.6 Αλγόριθμος

Έχοντας περιγράψει τα μοντέλα που χρησιμοποιούνται για την παρακολούθηση, παραμένει να εξετάσουμε πώς θα χρησιμοποιηθούν οι τεχνικές των ενοτήτων 4.3, 4.4 για την εκτίμηση του διανύσματος κατάστασης. Κατά συμφωνία με τη σχέση (4.18), γράφουμε

$$p(\mathbf{x}_t, o_t^i | \mathbf{x}_{t-1}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) = p(o_t^i | \mathbf{x}_{t-1:t}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \quad (4.41)$$

Ο δεύτερος παράγοντας προσεγγίζεται από ένα σύνολο δειγμάτων $\{\mathbf{x}_t^{(j)}\}_{j=1}^N$, ώστε

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \approx \sum_{j=1}^N w_t^{(j)} \delta_{\mathbf{x}_t^{(j)}}(\mathbf{x}_t) \quad (4.42)$$

Τα βάρη $w_t^{(j)}$ ισούνται με

$$w_t^{(j)} = \frac{p(\mathbf{z}_t | \mathbf{x}_{1:t}, \mathbf{z}_{1:t-1})}{\sum_{j=1}^N p(\mathbf{z}_t | \mathbf{x}_{1:t}^{(j)}, \mathbf{z}_{1:t-1})} \quad (4.43)$$

Η πιθανότητα $p(\mathbf{z}_t | \mathbf{x}_{1:t}, \mathbf{z}_{1:t-1})$ προκύπτει ως γινόμενο όλων των $p(\mathbf{z}_t^i | \mathbf{x}_{1:t}, \mathbf{z}_{1:t-1})$ για κάθε pixel i . Με βάση τις σχέσεις ανεξαρτησίας του σχήματος 4.4, έχουμε

$$p(\mathbf{z}_t | \mathbf{x}_{1:t}, \mathbf{z}_{1:t-1}) = \prod_i \sum_{o_t^i, o_{t-1}^i} [p(\mathbf{z}_t^i | \mathbf{x}_t, o_t^i) p(o_t^i | o_{t-1}^i) p(o_{t-1}^i | \mathbf{x}_{1:t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1})] \quad (4.44)$$

Ο τρίτος παράγοντας της παραπάνω σχέσης είναι η εκτίμηση του o^i που προέρχεται από την προηγούμενη επανάληψη του αλγορίθμου. Ανανεώνεται με βάση την αναδρομική σχέση

$$p(o_t^i | \mathbf{x}_{1:t}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) = \frac{\sum_{o_{t-1}^i} [p(\mathbf{z}_t^i | \mathbf{x}_t, o_t^i) p(o_t^i | o_{t-1}^i) p(o_{t-1}^i | \mathbf{x}_{1:t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1})]}{\sum_{o_t^i, o_{t-1}^i} [p(\mathbf{z}_t^i | \mathbf{x}_t, o_t^i) p(o_t^i | o_{t-1}^i) p(o_{t-1}^i | \mathbf{x}_{1:t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1})]} \quad (4.45)$$

Algorithm 4 6DoF παρακολούθηση ενός αντικειμένου με Particle Filter

Input: $\left\{ \mathbf{x}_{t-1}^{(j)} \right\}_{j=1}^N$, $\left\{ \left\{ p(o_{t-1}^i | \mathbf{x}_{1:t-1}^{(j)}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1}) \right\}_{i=1}^{N_{\text{pixels}}} \right\}_{j=1}^N$, \mathbf{z}_t , \mathbf{u}_t

Output: $\left\{ \mathbf{x}_t^{(j)} \right\}_{j=1}^N$, $\left\{ \left\{ p(o_t^i | \mathbf{x}_{1:t}^{(j)}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \right\}_{i=1}^{N_{\text{pixels}}} \right\}_{j=1}^N$

```
1: for  $j = 1 : N$  do
2:   Sample  $\mathbf{w}_t^{(j)} \sim \mathcal{N}(0, \Sigma)$ 
3:    $\mathbf{x}_t^{(j)*} \leftarrow f(\mathbf{x}_{t-1}^{(j)}, \mathbf{u}_t, \mathbf{w}_t^{(j)})$  (from (4.26))
4:   for  $i = 1 : N_{\text{pixels}}$  do
5:     Render  $d^i(r)$ 
6:     Calculate  $p(\mathbf{z}_t^i | \mathbf{x}_t^{(j)*}, o_t^i)$  from (4.33), (4.34)
7:     Calculate  $p(z_t^i | \mathbf{x}_t^{(j)*}, \mathbf{z}_{1:t-1})$  from (4.44)
8:     Calculate  $p(o_t^i | \mathbf{x}_t^{(j)*}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t})$  from (4.45)
9:   end for
10: end for
11: Calculate  $\left\{ w_t^{(j)} \right\}_{j=1}^N$  from (4.43)
12: Calculate  $D_{KL}(P||U)$  over  $\left\{ w_t^{(j)} \right\}_{j=1}^N$  from (4.16)Z
13: if  $D_{KL}(P||U) > D_{KL,max}$  then
14:   From algorithm 3: Resample  $\left\{ \mathbf{x}_t^{(j)} \right\}_{j=1}^N$  using  $\left\{ \mathbf{x}_t^{(j)*} \right\}_{j=1}^N$ ,  $\left\{ w_t^{(j)} \right\}_{j=1}^N$ 
15: else
16:    $\left\{ \mathbf{x}_t^{(j)} \right\}_{j=1}^N \leftarrow \left\{ \mathbf{x}_t^{(j)*} \right\}_{j=1}^N$ 
17: end if
```

4.7 Παρακολούθηση πολλαπλών αντικειμένων

Έως τώρα επικεντρωθήκαμε στην παρακολούθηση ενός μόνο αντικειμένου. Σε πολλές εφαρμογές, συμπεριλαμβανομένης της δικής μας, επιθυμούμε να παρακολουθήσουμε τις τροχιές παραπάνω αντικειμένων ταυτόχρονα. Το πρόβλημα αυτό είναι γνωστό στον τομέα της όρασης υπολογιστών ως **Multiple Object Tracking (MOT)**.

Υπάρχουν δύο θεμελιώδεις προσεγγίσεις στην επέκταση των αλγορίθμων παρακολούθησης ενός αντικειμένου σε περισσότερα. Η μία αφορά την ανεξάρτητη παρακολούθηση κάθε αντικειμένου, ενώ η δεύτερη την ταυτόχρονη παρακολούθηση θεωρώντας ένα μοναδικό διάνισμα κατάστασης μεγαλύτερης διάστασης. Η ανεξάρτητη προσέγγιση έχει το πλεονέκτημα της επεκτασιμότητας, αφού η πολυπλοκότητα είναι γραμμική ως προς τον αριθμό των αντικειμένων. Αν η παρακολούθηση του ενός αντικειμένου είναι ακριβής, τότε μπορούμε να περιμένουμε πως υπό συγκεκριμένες συνθήκες θα είναι ακριβής και η παρακολούθηση περισσότερων αντικειμένων. Όταν, όμως, δυο όμοια αντικείμενα βρεθούν κοντά το ένα στο άλλο, γίνεται δύσκολο για την κάθε μονάδα παρακολούθησης να ακολουθήσει το σωστό αντικείμενο. Από την άλλη, η από κοινού παρακολούθηση δεν παρουσιάζει τέτοιο πρόβλημα, διότι λογαριάζει τις θέσεις και των δύο αντικειμένων

ταυτόχρονα. Το μειονέκτημα τέτοιων μεθόδων, ωστόσο, έγκειται στη μη-γραμμική πολυπλοκότητα ως προς τον αριθμό των αντικειμένων. Εντούτοις, στην παρούσα εργασία ακολουθήσαμε τη δεύτερη προσέγγιση, αφού από τη φύση του προβλήματος περιμένουμε να αλληλεπιδρούν τα αντικείμενα μεταξύ τους.

4.7.1 Επέκταση του Particle Tracker

Ενώ το πρόβλημα MOT δεν αναφέρεται στο [23], υποστηρίζεται στον κώδικα που προσφέρουν οι συγγραφείς¹. Εδώ, παρουσιάζουμε τη μέθοδο που χρησιμοποιείται στον κώδικα για την επέκταση του αλγορίθμου της ενότητας 4.5 σε περισσότερα αντικείμενα.

Έστω K ο αριθμός των αντικειμένων και $\mathbf{X}_t = [{}^1\mathbf{x}_t, {}^2\mathbf{x}_t, \dots, {}^K\mathbf{x}_t]$ το συνολικό διάνυσμα κατάστασης που αποτελείται από τα διανύσματα κατάστασης του κάθε αντικειμένου (k είναι ο δείκτης του αντικειμένου). Επειδή το \mathbf{X}_t έχει μεγάλο αριθμό διαστάσεων, δεν είναι πρακτικό να δειγματοληπτήσουμε την $p(\mathbf{X}_t | \mathbf{z}_{1:t}, \mathbf{u}_{1:t})$ ταυτόχρονα ως προς όλες τις διαστάσεις. Αν το κάναμε αυτό, τα particles θα ήταν πολύ λίγα για να βρεθεί ακριβής εκτίμηση. Ομως, είναι

$$p(\mathbf{X}_t | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \approx \prod_{k=1}^K p({}^k\mathbf{x}_t | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) \quad (4.46)$$

για μικρές μεταβολές των θέσεων και προσανατολισμών των αντικειμένων. Δηλαδή, τα διανύσματα κατάστασης των αντικειμένων είναι υπό συνθήκες ανεξάρτητα. Με βάση την παρατήρηση αυτή, ακολουθεί η παρακάτω διαδικασία.

Τα N δείγματα διαιρούνται σε K τμήματα. Για κάθε τμήμα k τα δείγματα προωθούνται μέσω του δυναμικού μοντέλου, προσθέτοντας θόρυβο μόνο στις μεταβλητές πόζας του αντίστοιχου k -οστού αντικειμένου, διατηρώντας τους θορύβους των προηγούμενων $k - 1$ αντικειμένων από το προηγούμενο βήμα και μηδενίζοντας τους υπόλοιπους θορύβους. Οι πιθανοφάνειες υπολογίζονται όπως προηγουμένως, από τις οποίες λαμβάνουμε ένα σύνολο βαρών $\{w_t^{(i)}\}_{i=1}^{N/K}$ για κάθε τμήμα. Υπολογίζεται η απόκλιση KL των βαρών αυτών από την ομοιόμορφη κατανομή, με βάση τη σχέση (4.16). Αφού γίνει αυτό για όλα τα τμήματα, τα τελικά βάρη λαμβάνονται από το τελευταίο σύνολο $\{w_t^{(i)}\}_{i=1}^{N/K}$.

Πρακτικά, η διαφορά της παραπάνω διαδικασίας από τη δειγματοληψία όλων των κρυφών καταστάσεων ταυτόχρονα είναι ότι μετά την εκτίμηση της πόζας κάθε αντικειμένου ελέγχεται η κατανομή των βαρών πιθανοφάνειας, ώστε να απορριφθούν οι λιγότερο πιθανές εκτιμήσεις και να μην αχρηστευθούν για τα επόμενα βήματα τα particles που τις αντιπροσωπεύουν.

¹ <https://github.com/bayesian-object-tracking>

Η δυναμική εξίσωση (4.26) γράφεται

$$\mathbf{X}_t = f(\mathbf{X}_{t-1}, \mathbf{U}_t, \mathbf{W}_t) = \begin{bmatrix} \mathbf{I}_6 & k_v \mathbf{I}_6 \\ \mathbf{0}_6 & k_v \mathbf{I}_6 \\ \dots & \dots \\ \mathbf{I}_6 & k_v \mathbf{I}_6 \\ \mathbf{0}_6 & k_v \mathbf{I}_6 \\ \dots & \dots \\ \vdots & \vdots \\ \dots & \dots \\ \mathbf{I}_6 & k_v \mathbf{I}_6 \\ \mathbf{0}_6 & k_v \mathbf{I}_6 \end{bmatrix} \mathbf{X}_{t-1} + b \begin{bmatrix} {}^1 \mathbf{u}_t \\ {}^1 \mathbf{u}_t \\ \dots \\ {}^2 \mathbf{u}_t \\ {}^2 \mathbf{u}_t \\ \dots \\ \vdots \\ {}^K \mathbf{u}_t \\ {}^K \mathbf{u}_t \end{bmatrix} + \begin{bmatrix} {}^1 \mathbf{w}_t \\ {}^1 \mathbf{w}_t \\ \dots \\ {}^2 \mathbf{w}_t \\ {}^2 \mathbf{w}_t \\ \dots \\ \vdots \\ {}^K \mathbf{w}_t \\ {}^K \mathbf{w}_t \end{bmatrix} \quad (4.47)$$

, όπου $\mathbf{U}_t = [{}^1 \mathbf{u}_t, {}^2 \mathbf{u}_t, \dots, {}^K \mathbf{u}_t]$ και $\mathbf{W}_t = [{}^1 \mathbf{w}_t, {}^2 \mathbf{w}_t, \dots, {}^K \mathbf{w}_t]$ τα συνολικά διανύσματα εισόδου και θορύβου που αποτελούνται από τα αντίστοιχα διανύσματα κάθε αντικειμένου.

Οι εξισώσεις του μοντέλου παρατήρησης δεν αλλάζουν, απλά γίνεται rendering όλων τα αντικειμένων για τον υπολογισμό των αποστάσεων $d^i({}^1 \mathbf{r}, {}^2 \mathbf{r}, \dots, {}^K \mathbf{r})$ σε κάθε pixel. Οι εξισώσεις (4.41-4.45) θα είναι ίδιες, αλλά με αντικατάσταση των x_t με \mathbf{X}_t παντού.

4.7.2 Τομές αντικειμένων

Όταν δύο αντικείμενα βρίσκονται κοντά το ένα στο άλλο, ο αλγόριθμος παρακολούθησης μπορεί να δυσκολευτεί να τα ξεχωρίσει, ειδικά όταν έχουν παρόμοια σχήματα. Συχνά, οι εκτιμώμενες θέσεις οδηγούν σε μη-επιτρεπτές διατάξεις, με τα 3Δ μοντέλα να τέμνονται μεταξύ τους. Ιδανικά, θα θέλαμε να αναζητήσουμε μόνο επιτρεπτές διατάξεις. Ωστόσο, εξαιτίας της στοχαστικής φύσεως του αλγορίθμου μπορούμε να ελέγξουμε την εγκυρότητα του κάθε particle αφού πρώτα το περάσουμε μέσα από το δυναμικό μοντέλο. Επομένως, στο στάδιο του υπολογισμού της πιθανοφάνειας κάθε particle τιμωρούμε αυτά που αντιστοιχούν σε μη-επιτρεπτές πόζες.

Επειδή ο έλεγχος της τομής δύο αντικειμένων είναι αρκετά απλούστερος από την εύρεση του βάθους της τομής, επιλέξαμε να ενσωματώσουμε στην πιθανοφάνεια τον απλό παράγοντα

$$c_i = \begin{cases} \lambda_i & \text{αν οποιαδήποτε δύο μοντέλα αντικειμένων τέμνονται} \\ 1 & \text{αλλιώς} \end{cases} \quad (4.48)$$

Πειραματικά βρήκαμε ότι η τιμή $\lambda_i = 0.01$ παράγει ικανοποιητικά αποτελέσματα. Σημειώνεται ότι δεν θα μπορεί να τεθεί $\lambda_i = 0$, διότι κάποιες φορές αναπόφευκτα θα υπάρχει τομή αντικειμένων για όλα τα particles, οπότε στις περιπτώσεις αυτές θα μηδενιζόταν η ζητούμενη ΣΠΠ της ζητούμενης κατάστασης.

4.7.3 Βύθιση στην επιφάνεια εργασίας

Δοκιμάζοντας τον αλγόριθμο παρακολούθησης σε κάποια βίντεο βάθους, παρατηρήσαμε ότι οι εκτιμήσεις των θέσεων μερικές φορές κατέληγαν κάτω από την επιφάνεια του τραπέζιου. Αυτό συνέβαινε ειδικά όταν τα μοντέλα των αντικειμένων περιείχαν επίπεδες έδρες, επειδή είχαν την τάση να ευθυγραμμίζονται με την επιφάνεια του τραπέζιου, παράγοντας υψηλές τιμές πιθανοφάνειας για τα αντίστοιχα particles. Εισάγαμε μια απλή επίλυση του προβλήματος αυτού που περιλαμβάνει την "τιμωρία" των particles που αντιπροσωπεύουν τέτοιες θέσεις.

Για κάθε particle βρίσκουμε την χαμηλότερη κορυφή όλων των mesh model των αντικειμένων, τοποθετημένων στο σύστημα αναφοράς του τραπέζιου σύμφωνα με τις πόζες που αντιπροσωπεύει το particle. Έστω h_{min} το ύψος της κορυφής αυτής από το τραπέζι. Αν $h_{min} > 0$, τότε καμία εκτιμώμενη θέση δε διαπερνά την επιφάνεια του τραπέζιου. Αν $h_{min} < 0$, τότε επιβάλλουμε τιμωρία στο particle μειώνοντας την πιθανοφάνειά του. Συγκεκριμένα, πολλαπλασιάζουμε την πιθανοφάνεια με έναν παράγοντα ίσο με

$$c_h = \begin{cases} e^{\lambda_h h_{min}} & \text{αν } h_{min} < 0 \\ 1 & \text{αλλιώς} \end{cases} \quad (4.49)$$

, όπου λ_h μια παράμετρος που ρυθμίζει την ευαισθησία της τιμωρίας. Μετά από σύντομο πειραματισμό, τέθηκε $\lambda_h = 1000$. Αυτό σημαίνει πως η πιθανοφάνεια υποδεκαπλασιάζεται σε βάθος διείσδυσης περίπου $2.3mm$.

Συμπεριλαμβάνοντας τις σχέσεις (4.48) και (4.49), η (4.43) γίνεται

$$w_t^{(j)} = c_i c_h \frac{p(\mathbf{z}_t | \mathbf{X}_{1:t}^{(j)}, \mathbf{z}_{1:t-1})}{\sum_{j=1}^N p(\mathbf{z}_t | \mathbf{X}_{1:t}^{(j)}, \mathbf{z}_{1:t-1})} \quad (4.50)$$

Algorithm 5 6DoF παρακολούθηση πολλαπλών αντικειμένων με Particle Filter

Input: $\{\mathbf{X}_{t-1}^{(j)}\}_{j=1}^N$, $\{p(o_{t-1}^i | \mathbf{X}_{1:t-1}^{(j)}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1})\}_{i=1}^{N_{\text{pixels}}}$, \mathbf{z}_t , \mathbf{U}_t

Output: $\{\mathbf{X}_t^{(j)}\}_{j=1}^N$, $\{p(o_t^i | \mathbf{X}_{1:t}^{(j)}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t})\}_{i=1}^{N_{\text{pixels}}}\}_{j=1}^N$

```

1: Initialize  ${}^k\mathbf{w}_t^{(j)} = \mathbf{0}$ ,  $k = 1, \dots, K$ ,  $j = 1, \dots, N$ 
2: for  $k = 1 : K$  do
3:   for  $j = 1 : N$  do
4:     Sample  ${}^k\mathbf{w}_t^{(j)*} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ 
5:      $\mathbf{X}_t^{(j)*} \leftarrow f(\mathbf{X}_{t-1}^{(j)}, \mathbf{U}_t, \mathbf{W}_t^{(j)})$  (from (4.47))
6:     for  $i = 1 : N_{\text{pixels}}$  do
7:       Render  $d^i(\mathbf{1}_r, \mathbf{2}_r, \dots, \mathbf{K}_r)$ 
8:       Calculate  $p(\mathbf{z}_t^i | \mathbf{X}_t^{(j)*}, o_t^i)$  from (4.33), (4.34)
9:       Calculate  $p(\mathbf{z}_t^i | \mathbf{X}_t^{(j)*}, \mathbf{z}_{1:t-1})$  from (4.44)
10:      if  $k = K$  then
11:        Calculate  $p(o_t^i | \mathbf{X}_t^{(j)*}, \mathbf{z}_{1:t}, \mathbf{u}_{1:t})$  from (4.45)
12:      end if
13:    end for
14:  end for
15:  Calculate  $\{{}^k w_t^{(j)}\}_{j=1}^N$  from (4.50)
16:  Calculate  $D_{KL}(P||U)$  over  $\{{}^k w_t^{(j)}\}_{j=1}^N$  from (4.16)
17:  if  $D_{KL}(P||U) > D_{KL,max}$  then
18:    From algorithm 3: Resample  $\{\mathbf{X}_t^{(j)}\}_{j=1}^N$  using  $\{\mathbf{X}_t^{(j)*}\}_{j=1}^N$ ,  $\{{}^k w_t^{(j)}\}_{j=1}^N \rightarrow$ 
19:     $\rightarrow$  New noises:  $\{{}^k \mathbf{w}_t^{(j)}\}_{j=1}^N$  (that correspond to  $\{\mathbf{X}_t^{(j)}\}_{j=1}^N$ )
20:  else
21:     $\{\mathbf{X}_t^{(j)}\}_{j=1}^N \leftarrow \{\mathbf{X}_t^{(j)*}\}_{j=1}^N$ 
22:     $\{{}^k \mathbf{w}_t^{(j)}\}_{j=1}^N \leftarrow \{{}^k \mathbf{w}_t^{(j)*}\}_{j=1}^N$ 
23:  end if
24: end for

```

Κεφάλαιο 5

Εργασίες συναρμολόγησης: αναπαράσταση και αξιολόγηση

Μια εργασία συναρμολόγησης αποτελείται από μια σειρά ενεργειών πάνω σε διακριτά μέρη, που οδηγούν στη συγκρότηση ενός σύνθετου αντικειμένου. Οι ενέργειες αυτές σχηματίζουν ένα σύνολο *συνδέσεων* μεταξύ των μερών. Οι συνδέσεις είναι φυσικοί περιορισμοί μεταξύ δύο ή περισσότερων αντικειμένων, που μειώνουν τους βαθμούς ελευθερίας της κίνησης που ειδιάλλως θα είχαν τα αντικείμενα. Μια σύνδεση μεταξύ δύο μερών έχει ως αποτέλεσμα την απώλεια από 1 έως 6 βαθμών ελευθερίας, από τους αρχικούς 12 που διαθέτουν τα μέρη αθροιστικά. Δύο ή περισσότερα συνδεδεμένα αντικείμενα αποτελούν ένα *υποσυγκρότημα* της κατασκευής.

Οι συνδέσεις μπορούν να κατηγοριοποιηθούν ανάλογα με τον τύπο περιορισμού που εισάγουν. Στο [31] παρουσιάζονται τέσσερις τύποι περιορισμών, όπως χρησιμοποιούνται στα σχεδιαστικά προγράμματα Autocad Inventor και Mechanical Desktop: Ταύτιση (σύμπτωση σημείων, αξόνων, επιπέδων ή εδρών), Εισαγωγή (ταίριασμα αξόνων και επιπέδων δύο κύκλων), Ευθυγράμμιση (παραλληλία δύο επιπέδων) και Γωνία (ρύθμιση συγκεκριμένης γωνίας μεταξύ δύο διανυσμάτων ή δύο επιπέδων). Μπορούμε, επίσης, να ομαδοποιήσουμε τις συνδέσεις κατ' αναλογία με την κίνηση που απαιτείται. Μια απλή περίπτωση μπορεί απλά να περιλαμβάνει την τοποθέτηση ενός αντικειμένου δίπλα σε ένα άλλο, ικανοποιώντας μια συγκεκριμένη σχετική πόζα. Άλλες συνδέσεις μπορεί να απαιτούν πιο σύνθετες ενέργειες, όπως βίδωμα, δέσιμο ή κόλληση, μεταξύ άλλων. Στην παρούσα εργασία επικεντρωθήκαμε μόνο στην κατηγορία "τοποθέτησης", ώστε να αποφύγουμε την αναγνώριση σύνθετων δράσεων.

Σ' αυτό το κεφάλαιο περιγράφουμε τις δομές δεδομένων με τις οποίες αναπαριστούμε τις συναρμολογήσεις στην εργασία αυτή, καθώς και τις μεθόδους αξιολόγησης της διαδικασίας συναρμολόγησης. Στη συνέχεια παρουσιάζουμε τις τροποποιήσεις στο tracker με σκοπό την καλύτερη διαχείριση αλληλοσυνδεόμενων αντικειμένων.

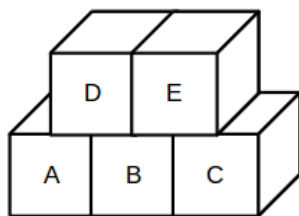
5.1 Αναπαράσταση συναρμολόγησης

Στην ενότητα αυτή περιγράφουμε τους τρόπους με τους οποίους μια εργασία συναρμολόγησης μπορεί να αναπαρασταθεί. Για τους σκοπούς της παρούσας εργασίας, μας ενδιαφέρουν δύο ειδών αναπαραστάσεις: η αναπαράσταση στατικής γνώσης και η αναπαράσταση αλληλουχίας. Η πρώτη περιλαμβάνει όλη την πληροφορία της τελικής διάταξης, χωρίς να λογαριάζει τα επιμέρους βήματα που απαιτούνται για να επιτευχθεί αυτή. Η δεύτερη αφορά το σύνολο (ή αλληλουχία) των δράσεων που οδηγούν στην ολοκλήρωση της συναρμολόγησης.

5.1.1 Αναπαράσταση στατικής γνώσης

Η στατική γνώση μιας συναρμολόγησης μπορεί να αναπαρασταθεί χρησιμοποιώντας έναν μη-κατευθυνόμενο γράφο $G_P = (P, C)$, του οποίου οι κόμβοι $P = \{P_1, P_2, \dots, P_N\}$ αντιπροσωπεύουν τα μέρη της σύνδεσης και οι ακμές $C = \{C_1, C_2, \dots, C_L\}$ υποδηλώνουν τις συνδέσεις μεταξύ τους. Οι ακμές μπορούν να συσχετισθούν με οποιαδήποτε πληροφορία που περιγράφει τις αντίστοιχες συνδέσεις (περιορισμοί ποζών, απαιτούμενες κινήσεις κλπ.).

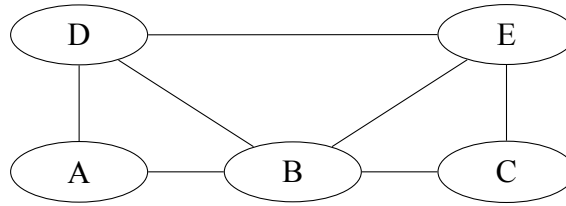
Ας θεωρήσουμε τη συναρμολόγηση της δομής του σχήματος 5.1. Ο γράφος G_P της εργασίας αυτής φαίνεται στο σχήμα 5.2. Υποθέσαμε σύνδεση μεταξύ οποιωνδήποτε δύο αντικειμένων που έχουν επαπτόμενες επιφάνειες. Αξίζει να σημειωθεί ότι μερικές από τις συνδέσεις αυτές φαίνονται περιττές, αφού η ίδια δομή μπορεί να επιτευχθεί χωρίς αυτές. Μάλιστα, στο παράδειγμα αυτό, οποιοσδήποτε συνεκτικός υπογράφος του G_P που περιέχει όλους τους κόμβους του P θα αναπαριστά το ίδιο αποτέλεσμα. Ωστόσο, συχνά είναι χρήσιμο να διατηρηθούν όλες οι συνδέσεις του G_P στην αναπαράσταση, διότι οι στατικές αναπαραστάσεις πρέπει να διατηρούν τη γενικότητα ως προς τον τρόπο ολοκλήρωσης της κατασκευής.



Σχήμα 5.1: Απλή δομή 5 κύβων.

5.1.2 Αναπαράσταση αλληλουχίας

Για την αναπαράσταση των αλληλουχιών των δράσεων που παράγουν την τελική συναρμολόγηση, εξετάζουμε όλα τα δυνατά υποσυγκροτήματα που μπορεί να προκύψουν



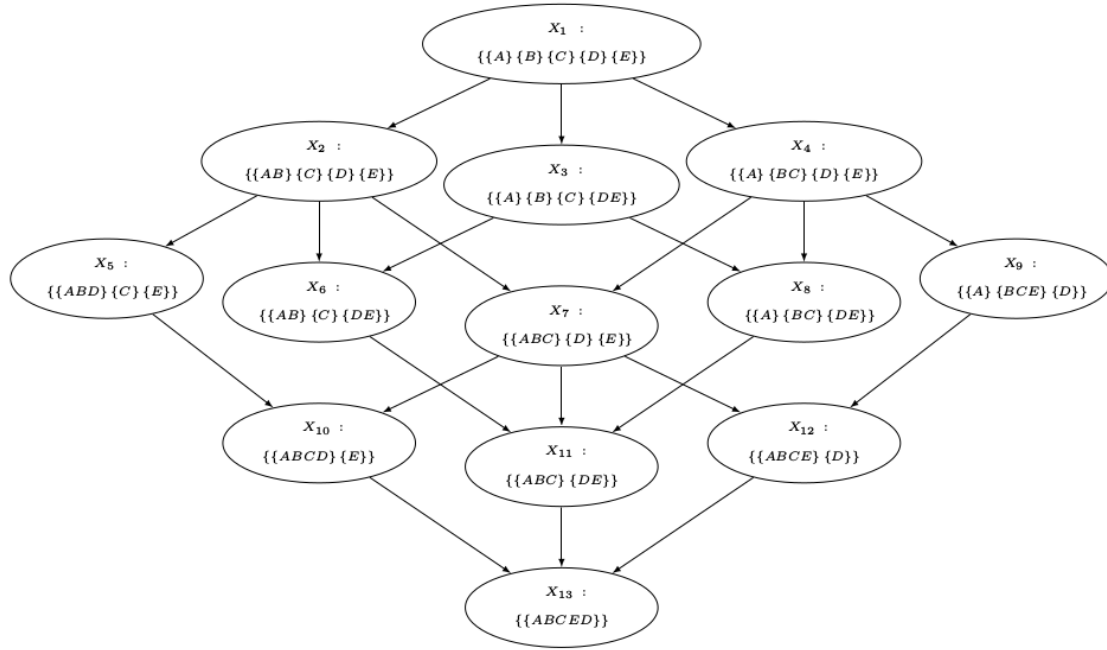
Σχήμα 5.2: Αναπαράσταση στατικής γνώσης του σχήματος 5.1.

κατά τη διαδικασία, μαζί με τη σειρά που θα εμφανιστούν. Ενώ θεωρητικά κάθε χρονική μετάθεση των συνδέσεων θα επαρκούσε για την ολοκλήρωση της συναρμολόγησης, κάποιες από τις αλληλουχίες αυτές δεν είναι εφικτές στον πραγματικό κόσμο. Μας ενδιαφέρουν μόνο οι αλληλουχίες που είναι γεωμετρικά και μηχανικά εφαρμόσιμες. Απαιτούμε, επίσης, απ' όλα τα υποσυγκροτήματα να είναι ευσταθή, δηλαδή να παραμένουν συνδεδεμένα τα μέρη τους αφού συνδεθούν μία φορά. Στο [32] παρουσιάζονται τέσσερις τρόποι αναπαράστασης των αλληλουχιών, συγκεκριμένα ο κατευθυνόμενος γράφος, ο γράφος AND/OR, οι συνθήκες εξασφάλισης και οι σχέσεις προήγησης. Από τους τέσσερις αυτούς τρόπους, μας ενδιαφέρει ο πρώτος.

Κατασκευάζουμε έναν κατευθυνόμενο γράφο $G_X = (X, T)$ για την αναπαράσταση όλων των δυνατών αλληλουχιών. Οι κόμβοι $X = \{X_1, X_2, \dots, X_M\}$ του γράφου αντιστοιχούν στους ευσταθείς συνδυασμούς υποσυγκροτημάτων, ενώ οι μεταβάσεις $T = \{T_1, T_2, \dots, T_K\}$ αντιπροσωπεύουν τις μεμονωμένες συνδέσεις που οδηγούν στην αλλαγή κατάστασης. Οι κόμβοι μπορούν επίσης να θεωρηθούν διαμερισμοί του P . Μονάχα οι εφικτές μεταβάσεις συμπεριλαμβάνονται στο T . Ο κόμβος X_1 αντιστοιχεί στην αρχική κατάσταση της συναρμολόγησης, με όλα τα μέρη ασύνδετα, ενώ ο X_M αντιστοιχεί στην τελική κατάσταση, στην οποία το αντικείμενο έχει συναρμολογηθεί.

Δυστυχώς, όσο πληθαίνουν οι εφικτές συνδέσεις, ο αριθμός των καταστάσεων στον G_X αυξάνεται γρήγορα. Ο γράφος αλληλουχιών της εργασίας του σχήματος 5.1 φαίνεται στο σχήμα 5.3. Με μόνο πέντε μέρη που πρέπει να συνδεθούν, ο γράφος έχει 13 κόμβους. Σε περιπτώσεις όπου ο υπολογιστής είναι υπεύθυνος για την παροχή ενός συνόλου προκαθορισμένων εντολών σε έναν ανθρώπινο ή ρομποτικό χειριστή, μπορούμε να απλοποιήσουμε το γράφο επιλέγοντας οποιαδήποτε διαδρομή που να οδηγεί από το X_1 στο X_M και απορρίπτοντας τις υπόλοιπες καταστάσεις. Αυτό μας δίνει μια λίστα δράσεων που μπορεί να παρουσιαστεί στο χειριστή της εργασίας. Ο νέος κατευθυνόμενος γράφος που προκύπτει προσφέρει έναν απλό τρόπο αξιολόγησης της συναρμολόγησης, αφού σε κάθε χρονική στιγμή χρειάζεται να ελεγχθεί μόνο μία σύνδεση.

Η διαδρομή που επιλέγεται μπορεί να χρησιμοποιηθεί για το σχηματισμό δυαδικού δέντρου, του οποίου οι κόμβοι-παιδιά αντιστοιχούν στα επιμέρους στοιχεία μιας σύνδεσης και οι κόμβοι-πατεράδες στο υποσυγκρότημα που κατασκευάζεται συνδέοντας τα παιδιά τους. Το δυαδικό δέντρο είναι ένας διαισθητικός τρόπος αναπαράστασης μιας πιθανής αλληλουχίας της συναρμολόγησης. Για παράδειγμα, ας θεωρήσουμε τη διαδρομή $X_1 \rightarrow$

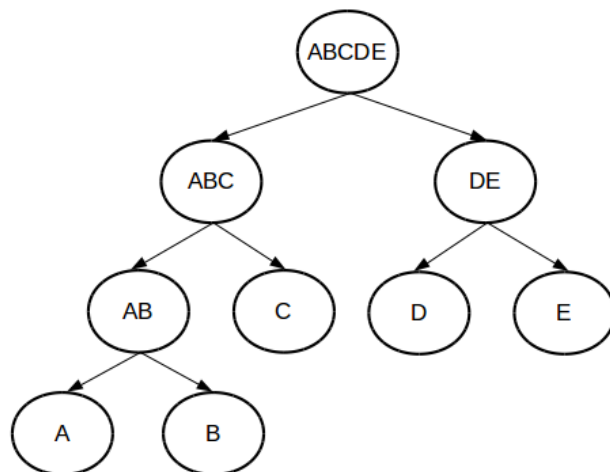


Σχήμα 5.3: Γράφος αλληλουχιών των εφικτών συνδέσεων του σχήματος 5.1. Καταστάσεις όπως η $\{\{AD\}\{B\}\{C\}\{E\}\}$ δεν συμπεριλήφθηκαν διότι δεν είναι ευσταθείς.

$X_2 \rightarrow X_6 \rightarrow X_{11} \rightarrow X_{13}$. Το δυαδικό δέντρο που ορίζει η διαδρομή αυτή φαίνεται στο σχήμα 5.4. Σημειώνεται ότι ένα δυαδικό δέντρο μπορεί να απεικονίζει παραπάνω από μία διαδρομή του G_X . Στο δοσμένο παράδειγμα, το ίδιο δέντρο μπορεί να προκύψει και από τις διαδρομές $X_1 \rightarrow X_2 \rightarrow X_7 \rightarrow X_{11} \rightarrow X_{13}$ και $X_1 \rightarrow X_3 \rightarrow X_6 \rightarrow X_{11} \rightarrow X_{13}$.

Σε άλλες περιπτώσεις ενδεχομένως να θέλουμε να επιτρέψουμε στον χρήστη να ολοκληρώσει τις συνδέσεις σε οποιαδήποτε σειρά, αρκεί να επιτευχθεί το σωστό τελικό αποτέλεσμα. Αυτό σημαίνει πως πρέπει να εξετάσουμε όλες τις διαδρομές από το X_1 στο X_M , οπότε δεν μπορούμε να κατασκευάσουμε ένα μοναδικό δέντρο που να αντικατοπτρίζει όλες τις δυνατές αλληλουχίες.

Κατά τη διαδικασία της συναρμολόγησης, μερικές από τις απαραίτητες συνδέσεις θα είναι ολοκληρωμένες και οι υπόλοιπες όχι. Μπορούμε να αναθέσουμε μια δυαδική τιμή σε κάθε ακμή του γράφου G_P , που θα δείχνει αν η αντίστοιχη σύνδεση είναι εξασφαλισμένη ή όχι. Οι τιμές αυτές συγκροτούν ένα δυαδικό διάνυσμα v_e . Το v_e θα χρησιμοποιηθεί στη συνέχεια για την αναπαράσταση της κατάστασης της εργασίας. Σημειώνουμε ότι το ποσοστό των άσων στο διάνυσμα μας δίνει μια εκτίμηση του βαθμού ολοκλήρωσης της κατασκευής.



Σχήμα 5.4: Μια δυνατή αναπαράσταση με δυαδικό δέντρο της συναρμολόγησης του σχήματος 5.1.

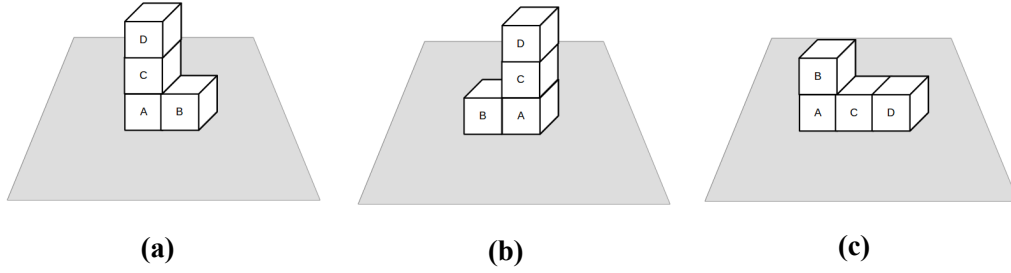
5.2 Αξιολόγηση της συναρμολόγησης

Έχοντας καθορίσει τους τρόπους αναπαράστασης της εργασίας συναρμολόγησης, πρέπει τώρα να εξετάσουμε πώς ο υπολογιστής θα αναγνωρίσει ποιες συνδέσεις έχουν σχηματιστεί και πώς θα συμπεράνει την παρούσα κατάσταση της εργασίας. Επειδή ασχοληθήκαμε μόνο με συνδέσεις "τοποθέτησης", και όχι πιο σύνθετες συνδυαστικές κινήσεις, το μόνο που χρειάζεται για να αξιολογηθεί η διαδικασία της συναρμολόγησης είναι το σύνολο των πόζες των αντικειμένων, οι οποίες προέρχονται από τον 3Δ tracker. Οι πόζες μετασχηματίζονται σε σχετικές πόζες μεταξύ των συνδεόμενων μερών, που στη συνέχεια συγκρίνονται με τις επιθυμητές σχετικές πόζες που ορίζουν την εργασία.

5.2.1 Υπολογισμός σχετικής πόζας

Μια σχετική θέση δύο αντικειμένων μπορεί να υπολογιστεί είτε στο γενικό σύστημα συντεταγμένων, είτε στο σύστημα ενός εκ των δύο αντικειμένων. Και οι δύο τρόποι μπορεί να είναι αναγκαίοι, ανάλογα με τη φύση της κατασκευής. Αν επιθυμούμε απλά να συνδέσουμε δυο μέρη, χωρίς να λάβουμε υπόψη τις θέσεις τους ως προς το χώρο εργασίας, τότε μπορούμε να υπολογίσουμε τη σχετική θέση στο σύστημα αναφοράς του πρώτου αντικειμένου. Αν, από την άλλη, θέλουμε να τοποθετήσουμε το ένα μέρος πάνω στο άλλο, τότε πρέπει να συνυπολογίσουμε τις θέσεις τους στο γενικό σύστημα αναφοράς. Στις περισσότερες κατασκευές από τουβλάκια, ο κάθετος άξονας πρέπει να είναι συγκεκριμένος, ενώ οι άλλοι δύο άξονες μπορούν να τοποθετηθούν με οποιοδήποτε τρόπο, αρκεί να παραμείνουν κάθετοι μεταξύ τους. Ένα παράδειγμα αυτού φαίνεται στο σχήμα 5.5.

Έστω \mathbf{p}_n^O το διάνυσμα θέσης του n -οστού μέρους και R_n^O ο πίνακας στροφής ως προς το γενικό σύστημα αναφοράς. Η στροφή ενός δεύτερου m -οστού μέρους (με $m \neq n$) ως



Σχήμα 5.5: Οι τρεις κατασκευές έχουν την ίδια δομή, αλλά μόνο οι (a) και (b) θεωρούνται ισοδύναμες, αφού η (c) διαφέρει ως προς τον κάθετο άξονα.

προς το σύστημα αναφοράς του n -οστού ισούται με:

$$R_m^n = R_O^n R_m^O = (R_n^O)^{-1} R_m^O = (R_n^O)^T R_m^O \quad (5.1)$$

Σημειώνεται ότι $R^{-1} = R^T$ για οποιοδήποτε πίνακα στροφής. Όταν χρησιμοποιούνται τετραδόνια για την αναπαράσταση των στροφών, η σχέση (5.1) γίνεται:

$$\mathbf{q}_m^n = \mathbf{q}_O^n \mathbf{q}_m^O = (\mathbf{q}_n^O)^{-1} \mathbf{q}_m^O \quad (5.2)$$

Η θέση του μέρους m ως προς το σύστημα συντεταγμένων του n -οστού υπολογίζεται ως εξής:

$$\mathbf{p}_m^n = R_n^O (\mathbf{p}_m^O - \mathbf{p}_n^O) = (R_n^O)^{-1} (\mathbf{p}_m^O - \mathbf{p}_n^O) = (R_n^O)^T (\mathbf{p}_m^O - \mathbf{p}_n^O) \quad (5.3)$$

Χρησιμοποιώντας ομογενείς συντεταγμένες, η σχέση (5.3) μπορεί να γραφτεί:

$$\mathbf{p}_m^n = T_{O^n}^n \mathbf{p}_m^O = (T_n^O)^{-1} \mathbf{p}_m^O \quad (5.4)$$

, όπου T_i^j είναι η μήτρα μετασχηματισμού από το σύστημα i στο σύστημα j :

$$T_i^j = \begin{bmatrix} R_i^j & \vdots & \mathbf{p}_i^j \\ \dots\dots\dots & & \\ 0 & 0 & 0 & \vdots & 1 \end{bmatrix} \quad (5.5)$$

Η αντίστροφη μιας μήτρας μετασχηματισμού T_i^j ισούται με:

$$T_j^i = (T_i^j)^{-1} = \begin{bmatrix} (R_i^j)^T & \vdots & -(R_i^j)^T \mathbf{p}_i^j \\ \dots\dots\dots & & \\ 0 & 0 & 0 & \vdots & 1 \end{bmatrix} \quad (5.6)$$

Όταν μια σύνδεση εκφράζεται ως σχετική πόζα στο σύστημα αναφοράς ενός εκ των δύο μερών, μπορούν να χρησιμοποιηθούν οι παραπάνω εξισώσεις. Αν θέλουμε να υπο-

λογίσουμε μια σχετική θέση στο γενικό σύστημα αναφοράς, τότε γράφουμε:

$$\mathbf{p}_{m,n} = \mathbf{p}_m^O - \mathbf{p}_n^O \quad (5.7)$$

Η σχετική στροφή παραμένει ίδια στην περίπτωση αυτή (σχέση (5.1)).

Ωστόσο, στην παρούσα εργασία, όταν εργαζόμαστε με κατασκευές με τουβλάκια, δε θέλουμε να εξετάσουμε τις πόζες στο γενικό σύστημα συντεταγμένων, αλλά σε ένα σύστημα που κάθετα πάνω στην επιφάνεια εργασίας του χρήστη, με τον άξονα y ευθυγραμμισμένο με το κάθετο διάνυσμα $\mathbf{N}_{\text{table}}$ του επιπέδου (στο παράρτημα Α παρουσιάζονται λεπτομέρειες για τον τρόπο υπολογισμού του διανύσματος αυτού). Στην περίπτωση αυτή, υπολογίζουμε τη στροφή R_O^{table} που απαιτείται ώστε να ευθυγραμμιστεί ο άξονας y του γενικού συστήματος αναφοράς με το διάνυσμα $\mathbf{N}_{\text{table}}$, και περιστρέφουμε τις θέσεις των αντικειμένων σύμφωνα με τη στροφή αυτή. Το R_O^{table} δεν ορίζεται με μοναδικό τρόπο, αλλά επειδή μας ενδιαφέρει μόνο να ευθυγραμμίσουμε τον κάθετο άξονα, οποιαδήποτε από τις επιλογές του R_O^{table} αρκεί. Υπολογίζουμε την στροφή λαμβάνοντας τον άξονα που προκύπτει από το εξωτερικό γινόμενο $\hat{\mathbf{y}}_o \times \mathbf{N}_{\text{table}}$, και εκτελώντας περιστροφή γύρω από τον άξονα αυτό ίση με τη γωνία μεταξύ του $\hat{\mathbf{y}}_o$ και του $\mathbf{N}_{\text{table}}$. Η εξίσωση (5.7) τώρα γίνεται:

$$\mathbf{p}_{m,n} = R_O^{\text{table}} \mathbf{p}_m^O - R_O^{\text{table}} \mathbf{p}_n^O = R_O^{\text{table}} (\mathbf{p}_m^O - \mathbf{p}_n^O) \quad (5.8)$$

5.2.2 Ισότητα ποζών

Για τον έλεγχο κάθε βήματος της διαδικασίας της συναρμολόγησης, χρησιμοποιούμε την πληροφορία πόζας που είναι διαθέσιμη για κάθε μέρος. Κάθε σύνδεση k σχετίζεται με μια επιθυμητή σχετική πόζα $\mathbf{r}_{des}^k = (\mathbf{p}_{des}^k, \mathbf{q}_{des}^k)$, όπου το \mathbf{p}_{des}^k είναι η επιθυμητή σχετική θέση και το \mathbf{q}_{des}^k η σχετική στροφή. Όταν αυτή ταυτίζεται με την πραγματική σχετική πόζα $\mathbf{r}_{real}^k = (\mathbf{p}_{real}^k, \mathbf{q}_{real}^k)$ των δύο μερών που εμπλέκονται (υπολογισμένη μέσω των εξισώσεων (5.2),(5.3)), η σύνδεση θεωρείται ολοκληρωμένη. Επιτρέπουμε, επίσης, ένα μικρό περιθώριο απόκλισης μεταξύ της πραγματικής και της επιθυμητής πόζας. Οι δύο πόζες είναι ίσες όταν είναι ίσες οι μετατοπίσεις και οι στροφές τους.

Ίσες μετατοπίσεις σημαίνει απλά ότι

$$\mathbf{p}_{des}^k = \mathbf{p}_{real}^k. \quad (5.9)$$

Δοσμένου ενός περιθωρίου σφάλματος m_p , αυτό γίνεται

$$\|\mathbf{p}_{des}^k - \mathbf{p}_{real}^k\|_2 < m_p. \quad (5.10)$$

Εναλλακτικά, μπορεί να θέλουμε να δώσουμε διαφορετικό περιθώριο απόκλισης σε

κάθε διάσταση, δηλαδή $m_{p,x}$, $m_{p,y}$, $m_{p,z}$. Τότε, απαιτούμε

$$\begin{cases} |\mathbf{p}_{des,x}^k - \mathbf{p}_{real,x}^k| < m_{p,x} \\ |\mathbf{p}_{des,y}^k - \mathbf{p}_{real,y}^k| < m_{p,y} \\ |\mathbf{p}_{des,z}^k - \mathbf{p}_{real,z}^k| < m_{p,z} \end{cases} \quad (5.11)$$

Η σύγκριση δύο στροφών εξαρτάται από τον τρόπο αναπαράστασης. Με τετραδόνια αποτελεί απλή διαδικασία: Αν $\mathbf{q}_{des}^k = \mathbf{q}_{real}^k$ ή $\mathbf{q}_{des}^k = -\mathbf{q}_{real}^k$, τότε τα δύο τετραδόνια αντιπροσωπεύουν την ίδια στροφή. Διαφορετικά, αυτό γράφεται

$$|\mathbf{q}_{des}^k \cdot \mathbf{q}_{real}^k| = 1 \quad (5.12)$$

Αν θέλουμε να προσθέσουμε ένα περιθώριο σφάλματος m_q στη στροφή, τότε η (5.12) μπορεί να γίνει

$$|\mathbf{q}_{des}^k \cdot \mathbf{q}_{real}^k| \geq 1 - m_q \quad (5.13)$$

Τα στοιχεία του διανύσματος κατάστασης των συνδέσεων \mathbf{v}_c (ενότητα 5.1.2) καθορίζονται βάση του κανόνα

$$\mathbf{v}_c^k = \begin{cases} 1 & \text{αν (5.11) } \wedge \text{ (5.13)} \\ 0 & \text{αλλιώς} \end{cases} \quad (5.14)$$

5.2.3 Χειρισμός συμμετριών

Σε κάποιες περιπτώσεις, εξαιτίας των ιδιοτήτων συμμετρίας των μερών, πρέπει να εξεταστούν περισσότερες από μία σχετικές πόζες. Ένα αντικείμενο με κυλινδρική συμμετρία, πχ. βίδα, άξονας κλπ., μπορεί να τοποθετηθεί σε οποιαδήποτε γωνία γύρω από τον κεντρικό του άξονα. Τα κυβικά αντικείμενα που δεν έχουν κάποια διαφοροποίηση στις 6 έδρες τους μπορούν να τοποθετηθούν με 24 διαφορετικούς τρόπους, που όλοι παράγουν το ίδιο αποτέλεσμα. Επομένως, δεν αρκεί να συγκριθεί απλά η παρατηρούμενη σχετική πόζα με την επιθυμητή· πρέπει να εξεταστεί ένα σύνολο επιτρεπτών διατάξεων.

Στην εργασία αυτή επικεντρωθήκαμε μόνο στην κυβική συμμετρία. Ένα κυβικό αντικείμενο μπορεί να υποστεί οποιοδήποτε συνδυασμό στροφών 90 και 180 μοιρών γύρω από τους βασικούς του άξονες, χωρίς να επηρεαστεί το τελικό αποτέλεσμα. Υπάρχουν συνολικά 24 τέτοιοι συνδυασμοί που παράγουν διαφορετικές στροφές. Οι στροφές αυτές φαίνονται στον ακόλουθο πίνακα:

Αρχικά, πρέπει να ληφθούν όλες οι παραπάνω στροφές και για τα δύο μέρη μιας σύνδεσης. Όταν, όμως, δυο κύβοι συνδεθούν μεταξύ τους, μειώνονται οι επιτρεπτές επιλογές σε σχέση με κάποιον τρίτο κύβο. Ας εξετάσουμε, για παράδειγμα την απλή δομή του σχήματος 5.6.

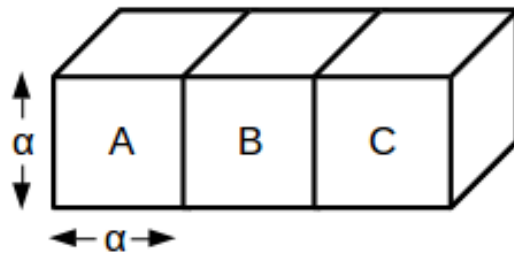
Κυβικές Στροφές			
Τετραδόνιο	Γωνίες Euler (y-z'-x'')	Τετραδόνιο	Γωνίες Euler (y-z'-x'')
[0, 0, 0, 1]	[0, 0, 0]	$[-\frac{\sqrt{2}}{2}, 0, 0, \frac{\sqrt{2}}{2}]$	$[0, 0, -\frac{\pi}{2}]$
[1, 0, 0, 0]	[0, 0, π]	$[0, -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0]$	$[\pi, 0, \frac{\pi}{2}]$
[0, 1, 0, 0]	$[\pi, 0, 0]$	$[0, -\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}]$	$[-\frac{\pi}{2}, 0, 0]$
[0, 0, 1, 0]	[0, $\pi, 0]$	$[0, 0, -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]$	$[0, -\frac{\pi}{2}, 0]$
$[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0, 0]$	$[\pi, \frac{\pi}{2}, 0]$	[0.5, 0.5, 0.5, 0.5]	$[\frac{\pi}{2}, \frac{\pi}{2}, 0]$
$[\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}, 0]$	$[-\frac{\pi}{2}, 0, \pi]$	[-0.5, 0.5, 0.5, 0.5]	$[\frac{\pi}{2}, 0, -\frac{\pi}{2}]$
$[\frac{\sqrt{2}}{2}, 0, 0, \frac{\sqrt{2}}{2}]$	$[0, 0, \frac{\pi}{2}]$	[0.5, -0.5, 0.5, 0.5]	$[-\frac{\pi}{2}, 0, \frac{\pi}{2}]$
$[0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0]$	$[\pi, 0, -\frac{\pi}{2}]$	[0.5, 0.5, -0.5, 0.5]	$[\frac{\pi}{2}, 0, \frac{\pi}{2}]$
$[0, \frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}]$	$[\frac{\pi}{2}, 0, 0]$	[-0.5, -0.5, -0.5, 0.5]	$[-\frac{\pi}{2}, 0, -\frac{\pi}{2}]$
$[0, 0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]$	$[0, \frac{\pi}{2}, 0]$	[-0.5, -0.5, 0.5, 0.5]	$[-\frac{\pi}{2}, \frac{\pi}{2}, 0]$
$[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0, 0]$	$[\pi, -\frac{\pi}{2}, 0]$	[-0.5, 0.5, -0.5, 0.5]	$[\frac{\pi}{2}, -\frac{\pi}{2}, 0]$
$[-\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}, 0]$	$[\frac{\pi}{2}, 0, \pi]$	[0.5, -0.5, -0.5, 0.5]	$[-\frac{\pi}{2}, -\frac{\pi}{2}, 0]$

Ο κύβος B μπορεί να τοποθετηθεί σε οποιαδήποτε από της θέσεις $[a, 0, 0]$, $[-a, 0, 0]$, $[0, a, 0]$, $[0, -a, 0]$, $[0, 0, a]$ ή $[0, 0, -a]$, ως προς το σύστημα αναφοράς του κύβου A, σε οποιαδήποτε από τις κυβικές στροφές που παρουσιάστηκαν παραπάνω. Κατά την αξιολόγηση της πρώτης σύνδεσης δεν έχει σημασία ποια από αυτές τις εναλλακτικές επιλέγεται. Η δεύτερη σύνδεση, ωστόσο, εξαρτάται από την πρώτη, επειδή κάθε εναλλακτική μας αναγκάζει να απαιτήσουμε διαφορετική σχετική πόζα. Αν, για παράδειγμα, $\mathbf{p}_{real}^1 = [a, 0, 0]$, με $\mathbf{q}_{real}^1 = [0, 0, 0, 1]$, τότε πρέπει $\mathbf{p}_{des}^2 = [a, 0, 0]$ (σχήμα 5.6b). Αν, από την άλλη, $\mathbf{p}_{real}^1 = [0, 0, -a]$, με την ίδια σχετική στροφή, τότε απαιτούμε $\mathbf{p}_{des}^2 = [0, 0, -a]$ (σχήμα 5.6c) για το σχηματισμό της δομής που φαίνεται στο σχήμα 5.6a.

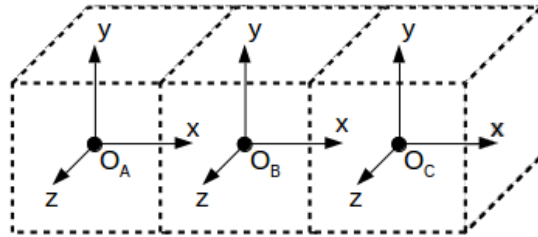
Όπως βλέπουμε, οι επιτρεπτές διατάξεις είναι πολυπληθείς και δεν μπορούν όλες να εξεταστούν ταυτόχρονα σε κάθε καινούργιο πλαίσιο. Προτείνουμε μια λύση του προβλήματος, που περιορίζει τους ελέγχους που πρέπει να γίνουν σε κάθε επανάληψη. Κατά την περιγραφή της εργασίας συναρμολόγησης, ορίζουμε τη θεμελιώδη σχετική πόζα κάθε σύνδεσης, υποθέτοντας ότι έχουν ολοκληρωθεί και οι προηγούμενες συνδέσεις σύμφωνα με τις θεμελιώδεις πόζες τους. Στο σχήμα 5.6, για παράδειγμα, ορίζουμε τις σχετικές θέσεις και προσανατολισμούς των συνδέσεων ως:

$$\mathbf{p}_{real}^1 = [a, 0, 0], \mathbf{q}_{real}^1 = [0, 0, 0, 1], \mathbf{p}_{des}^2 = [a, 0, 0], \mathbf{q}_{real}^2 = [0, 0, 0, 1]$$

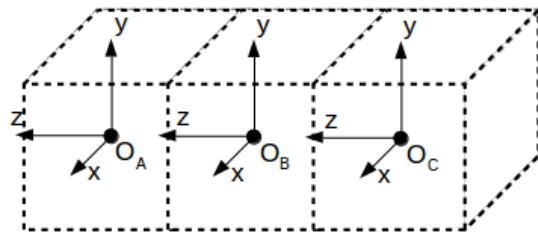
, ενώ οι υπόλοιπες επιτρεπτές επιλογές εννοούνται. Κατά την εξέταση των συνδέσεων, διακρίνουμε τέσσερις περιπτώσεις, ανάλογα με το αν οι κύβοι είναι ήδη συνδεδεμένοι με άλλα μέρη ή όχι:



(a) Δομή



(b) Μία επιλογή τρόπου σύνδεσης



(c) Άλλη επιλογή σύνδεσης

Σχήμα 5.6: Παράδειγμα δομής τριών κύβων, με δύο τρόπους σύνδεσης των κύβων που δημιουργούν ισοδύναμα αποτελέσματα.

- Αν κανένα από τα δύο μέρη δεν έχει συνδεθεί με κάποιο τρίτο, τότε και τα δύο μπορούν να υποστούν οποιαδήποτε κυβική στροφή. Υπολογίζουμε ποια στροφή είναι πιο πιθανή για το καθένα, με τον ακόλουθο τρόπο. Η πρώτη στροφή \mathbf{Q}_1 προκύπτει περιστρέφοντας την παρατηρούμενη σχετική θέση ώστε να ταιριάζει καλύτερα στην επιθυμητή, ελαχιστοποιώντας δηλαδή το εσωτερικό τους γινόμενο. Από τις 24 δυνατές κυβικές στροφές, μέχρι 4 μπορούν να οδηγήσουν σε εξίσου μικρά εσωτερικά γινόμενα. Απ' αυτές, επιλέγουμε αυτήν που ευθυγραμμίζει καλύτερα τον άξονα y του κύβου με το κάθετο διάνυσμα του τραπέζιου. Για τη στροφή \mathbf{Q}_2 του δεύτερου αντικειμένου, λαμβάνουμε τον προσανατολισμό του ως προς το σύστημα αναφοράς του πρώτου και τον πολλαπλασιάζουμε επί \mathbf{Q}_1 , δηλαδή $\mathbf{Q}_2 = \mathbf{Q}_1 \mathbf{q}_2^1$. Αποθηκεύουμε τις αντίστροφες των στροφών αυτών στις αντίστοιχες **πρότερες περιστροφές** των κύβων, όπως αυτά χρησιμοποιούνται στις υπόλοιπες τρεις περιπτώσεις. Οι στροφές αυτές δηλώνουν την απαιτούμενη περιστροφή ώστε να μεταφέρουμε τους προσανατολισμούς των κύβων στη θεμελιώδη κατάσταση και να έχει νόημα η σύγκριση της απαιτούμενης πόζας με την πραγματική.

- Αν ο πρώτος κύβος έχει ήδη συνδεθεί με κάποιον τρίτο, τότε τον περιστρέφουμε κατά την πρότερη περιστροφή του, που υπολογίστηκε προηγουμένως. Στην περίπτωση αυτή μόνο το δεύτερο μέρος μπορεί να περιστραφεί ελεύθερα. Αποθηκεύουμε την πιθανή κυβική στροφή του δεύτερου κύβου όπως προηγουμένως: $Q_2 = Q_1 q_2^1$.
- Αν ο δεύτερος κύβος έχει ήδη συνδεθεί με κάποιον άλλον, τότε τον περιστρέφουμε κατά την πρότερη περιστροφή του, και υπολογίζουμε την κυβική στροφή του πρώτου από τη σχέση $Q_1 = Q_2 q_1^2$.
- Αν και τα δύο μέρη είναι συνδεδεμένα με άλλα, τότε τα περιστρέφουμε και τα δύο κατά τις αντίστοιχες πρότερες περιστροφές τους.

Σε κάθε περίπτωση, μετά την εφαρμογή των αντίστοιχων πρότερων περιστροφών, συγκρίνουμε την προκύπτουσα σχετική πόζα με αυτήν που περιγράφει τη σύνδεση, όπως εξηγήθηκε στην ενότητα 5.2.2.

5.3 Αλλαγές στον αλγόριθμο 3Δ παρακολούθησης

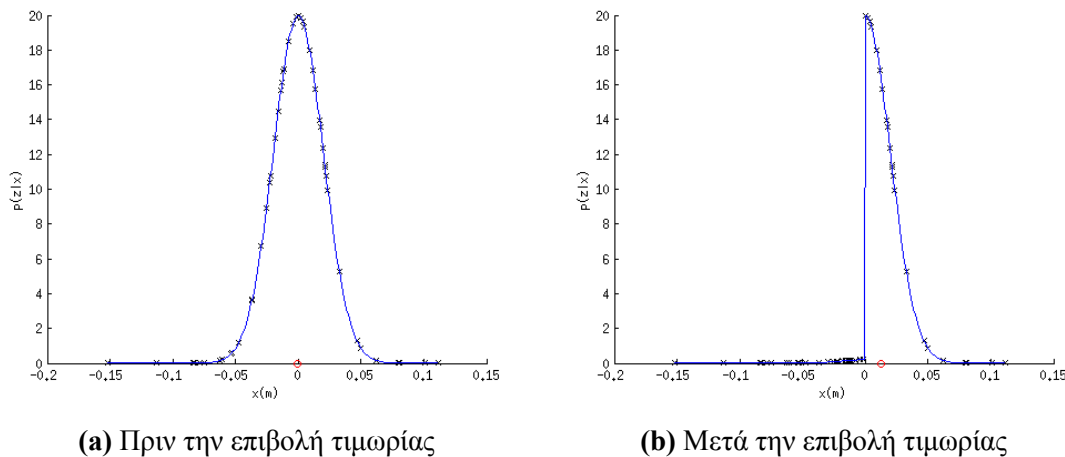
Στα προηγούμενα κεφάλαια περιγράψαμε μια μέθοδο παρακολούθησης πολλαπλών αντικειμένων στον τρισδιάστατο χώρο. Τώρα θα προτείνουμε μια τροποποίηση του αλγορίθμου, για να χειριστεί καλύτερα το πρόβλημα της επίβλεψης συναρμολόγησης. Η τροποποίηση βασίζεται στην παρατήρηση ότι καθώς ο χρήστης συνδέει διάφορα μέρη μεταξύ τους, οι βαθμοί ελευθερίας της κίνησής τους μειώνονται, οπότε ο αλγόριθμος παρακολούθησης μπορεί να επιβάλει τις συνδέσεις κατά την εκτίμηση των ποζών των μερών.

5.3.1 Κίνητρο

Όταν είμαστε σχεδόν σίγουροι πως μια σύνδεση έχει ολοκληρωθεί μεταξύ δύο μερών, μπορούμε να δώσουμε εντολή στον αλγόριθμο παρακολούθησης να υποθέσει μια "κλειδωμένη" σχετική θέση μεταξύ των μερών. Στην περίπτωση αυτή εξαναγκάζουμε τα particles να αντιπροσωπεύσουν την επιθυμητή σχετική πόζα της σύνδεσης. Ο βασικός λόγος να το κάνουμε αυτό είναι ότι απλοποιούμε το πρόβλημα παρακολούθησης, μειώνοντας τον αριθμό των μεταβλητών που πρέπει να εκτιμηθούν, ενώ ταυτόχρονα εξασφαλίζουμε πως οι υπολογισμένες σχετικές θέσεις δε θα απομακρυνθούν η μία από την άλλη. Σε άλλες περιπτώσεις, μπορεί να είναι χρήσιμο να αναθέσουμε σ' ένα μικρότερο υποσύνολο των particles την επιβολή της σύνδεσης, αν δεν είμαστε ακόμη σίγουροι για την ολοκλήρωσή της. Το κίνητρο πίσω απ' αυτή τη διαδικασία θα φανεί στη συνέχεια.

Στην ενότητα 4.7.2 εισάγαμε μια τεχνική επιβολής τιμωρίας στα particles που αντιπροσωπεύουν σε θέσεις που εισχωρούν η μία στην άλλη. Η τεχνική αυτή είναι αποτελεσματική

στην αποφυγή παράνομων διατάξεων, αλλά προκαλεί ένα μικρό σφάλμα στις εκτιμώμενες θέσεις. Το γεγονός αυτό φαίνεται πιο καθαρά στο σχήμα 5.7. Δείχνουμε τις πιθανοφάνειες των particles συναρτήσεις της ελάχιστης απόστασης x μεταξύ δύο αντικειμένων. Αρνητική τιμή απόστασης σηματοδοτεί βάθος εισχώρησης $-x$. Σημειώνεται ότι για σκοπούς οπτικοποίησης χρησιμοποιήσαμε Γκαουσιανή καμπύλη, αλλά αυτό δε συμβαίνει στην πραγματικότητα.



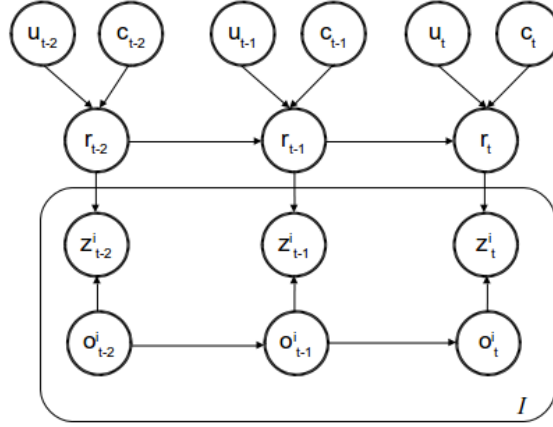
Σχήμα 5.7: Εάν δεν τιμωρούσαμε τις τομές αντικειμένων, οι πιθανοφάνειες των particles (σημειωμένες με μαύρο x) θα ακολουθούσαν την μπλε καμπύλη αριστερά. Μετά την επιβολή τιμωρίας, οι πιθανοφάνειες μειώνονται για $x < 0$, όπως φαίνεται στα δεξιά. Αυτό οδηγεί στη μετατόπιση προς τα δεξιά του σταθμισμένου μέσου όρου του x (κόκκινος κύκλος).

Βλέπουμε ότι μετά την τιμωρία των particles για τα οποία $x < 0$, ο σταθμισμένος μέσος όρος $\frac{\sum_{i=1}^N p(z|x_i)x_i}{\sum_{i=1}^N p(z|x_i)}$ μετατοπίζεται προς τα δεξιά. Αυτό σημαίνει πως όταν δύο μέρη συνδέονται, οι εκτιμώμενες θέσεις τους θα είναι πιο απομακρυσμένες απ' ό,τι θα έπρεπε. Ως αποτέλεσμα, οι συνθήκες που απαιτούνται για την αναγνώριση της σύνδεσης μπορεί να μην ικανοποιηθούν.

Όπως αναφέρθηκε προηγουμένως, μπορούμε να επιβάλουμε την επιθυμητή σχετική πόζα σε ένα μικρό αριθμό particles. Αυτό ισοδυναμεί με το να θέσουμε $x = 0$. Στα particles αυτά θα αποδοθεί μεγάλη τιμή πιθανοφάνειας όταν ολοκληρωθεί η σύνδεση, οπότε ο σταθμισμένος μέσος όρος θα μετακινηθεί πιο κοντά στο μηδέν.

5.3.2 Προτεινόμενη τροποποίηση

Έστω ότι θέλουμε να δεσμεύσουμε N_k particles σε κάθε σύνδεση k . Όπως προαναφέραμε, κάθε σύνδεση σχετίζεται με μια σχετική πόζα $\mathbf{r}_{des}^k = (\mathbf{p}_{des}^k, \mathbf{q}_{des}^k)$, ή με χρήση πίνακα στροφής $\mathbf{r}_{des}^k = (\mathbf{p}_{des}^k, R_{des}^k)$. Επιβάλλουμε στα N_k πρώτα particles την \mathbf{r}_{des}^k αμέσως μετά την εφαρμογή της εξίσωσης μετάβασης, ενώ τα υπόλοιπα αφήνονται ελεύθερα όπως προηγουμένως. Για να λάβουμε υπόψη τη δυναμική και των δύο μερών, χωρίζουμε



Σχήμα 5.8: Μαρκοβιανό μοντέλο του σχήματος 4.2 με προσθήκη του διανύσματος συνδέσεων c . Το c δείχνει ποιες συνδέσεις είναι ολοκληρωμένες. Κάθε σύνδεση εισάγει έναν περιορισμό στην κατάσταση r .

το N_k στα δύο, όπου για την πρώτη ομάδα θέτουμε $\mathbf{r}_2 = (\mathbf{p}_1 + R_1 R_{des}^k, R_1 R_{des}^k)$ και για τη δεύτερη $\mathbf{r}_1 = (\mathbf{p}_2 - R_2 (R_{des}^k)^T \mathbf{p}_{des}^k, R_2 (R_{des}^k)^T)$.

5.3.3 Βαθμολόγηση

Όταν μια σύνδεση βρίσκεται κοντά στην ολοκλήρωσή της, υπολογίζουμε μια βαθμολογία ενδεικτική του πόσο όμοια είναι η παρατηρούμενη σχετική πόζα με την επιθυμητή. Η βαθμολογία αυτή μπορεί να χρησιμοποιηθεί για τον υπολογισμό του αριθμού των particles που θα ανατεθούν στη συγκεκριμένη σύνδεση. Χρησιμοποιήσαμε τιμή βαθμολογίας ίση με

$$score = \frac{1}{2} \left(\frac{1}{d(\mathbf{p}_{des}, \mathbf{p}_{real})} + \frac{1}{d(\mathbf{p}'_{des}, \mathbf{p}'_{real})} \right) \quad (5.15)$$

, όπου:

$$d(\mathbf{p}_1, \mathbf{p}_2) = \sqrt{\max \left\{ 0, K \left[\left(\frac{p_{1,x} - p_{2,x}}{m_{p,x}} \right)^2 + \left(\frac{p_{1,y} - p_{2,y}}{m_{p,y}} \right)^2 + \left(\frac{p_{1,z} - p_{2,z}}{m_{p,z}} \right)^2 - 3 \right] \right\} + 1} \quad (5.16)$$

και $\mathbf{p}'_{des}, \mathbf{p}'_{real}$ οι αντίστοιχες σχετικές πόζες εκφρασμένες στο σύστημα αναφοράς του δεύτερου αντικειμένου αντί του πρώτου, δηλαδή (από σχέση (5.6))

$$\mathbf{p}'_{des} = -\mathbf{q}_{des}^* \mathbf{p}_{des} \quad (5.17)$$

$$\mathbf{p}'_{real} = -\mathbf{q}_{real}^* \mathbf{p}_{real} \quad (5.18)$$

Στη σχέση (5.16) το K είναι παράμετρος που επηρεάζει την ευαισθησία της βαθμολογίας. Όσο μεγαλύτερο είναι το K , τόσο μικρότερη θα είναι η βαθμολογία όταν η επιθυμητή σχετική θέση διαφέρει από την πραγματική σε βαθμό μεγαλύτερο από τα επιτρεπόμενα

Algorithm 6 Τροποποίηση του αλγορίθμου 3Δ παρακολούθησης για το συνυπολογισμό συνδέσεων

Input: Number of particles to assign: $\{N_k\}_{k=1}^{N_{\text{connections}}}$, previous poses: $\left\{ \left\{ \mathbf{r}_i^{(j)} \right\}_{i=1}^K \right\}_{j=1}^{N_{\text{samples}}}$,

relative poses to enforce: $\left\{ \mathbf{r}_{des}^k \right\}_{k=1}^{N_{\text{connections}}}$

Output: New poses: $\left\{ \left\{ \mathbf{r}_i^{(j)} \right\}_{i=1}^K \right\}_{j=1}^{N_{\text{samples}}}$

```

1: for  $k = 1 : N_{\text{connections}}$  do
2:   for  $j = 1 : 2 : N_k$  do
3:      $a \leftarrow$  First part of connection  $k$ 
4:      $b \leftarrow$  Second part of connection  $k$ 
5:      $\mathbf{r}_b^{(j)} = (\mathbf{p}_a^{(j)} + R_a^{(j)} R_{des}^k, R_a^{(j)} R_{des}^k)$ 
6:   end for
7: end for
8: for  $k = N_{\text{connections}} : -1 : 1$  do
9:   for  $j = 2 : 2 : N_k$  do
10:     $a \leftarrow$  First part of connection  $k$ 
11:     $b \leftarrow$  Second part of connection  $k$ 
12:     $\mathbf{r}_a^{(j)} = (\mathbf{p}_b^{(j)} - R_b^{(j)} (R_{des}^k)^T \mathbf{p}_{des}^k, R_b^{(j)} (R_{des}^k)^T)$ 
13:   end for
14: end for

```

περιθώρια. Όταν οι σχετικές θέσεις διαφέρουν εντός των περιθωρίων, τότε η βαθμολογία θα είναι ίση με 1. Όσο αυξάνεται η διαφορά, η βαθμολογία μειώνεται και τείνει στο 0.

Ο λόγος που προσθέσαμε το δεύτερο όρο στη σχέση (5.15) είναι ότι μπορεί μια στροφή του πρώτου αντικειμένου να αλλάξει σημαντικά την σχετική θέση ως προς το σύστημα αναφοράς του, ενώ στην πραγματικότητα τα δύο αντικείμενα δεν απέχουν πολύ από τη σύνδεσή τους. Άρα, συγκρίνουμε και τις αντίστροφες σχετικές πόζες και λαμβάνουμε το μέσο όρο των δύο τιμών.

Σημειώνεται ότι δεν συμπεριλάβαμε πληροφορία στροφής στη σχέση (5.16), διότι γενικά οι προσανατολισμοί που υπολογίζονται είναι πιο ανακριβείς από τις θέσεις, και προσφέρουν λιγότερη πληροφορία σχετικά με το πόσο κοντά στην ολοκλήρωση είναι η σύνδεση. Ωστόσο, μια απλή τροποποίηση για να συμπεριληφθεί η πληροφορία αυτή θα ήταν η πρόσθεση ενός όρου, που θα βασίζεται στο εσωτερικό γινόμενο του παρατηρούμενου και του επιθυμητού σχετικού προσανατολισμού.

Υποθέτοντας ότι ο αλγόριθμος παρακολούθησης χρησιμοποιεί συνολικά N_{all} particles, από τα οποία είμαστε πρόθυμοι να διαθέσουμε μέχρι N_{max} particles σε οποιαδήποτε σύνδεση, καθορίζουμε τον αριθμό N_k των particles που αναθέτονται στη μη ολοκληρωμένη σύνδεση k :

$$N_k = \lfloor score * N_{max} \rfloor \quad (5.19)$$

, όπου $\lfloor a \rfloor$ δηλώνει την τιμή του a στρογγυλοποιημένη στον κοντινότερο ακέραιο.

Κεφάλαιο 6

Υλοποίηση - Πειράματα και αποτελέσματα

Στα προηγούμενα κεφάλαια προτείναμε μια μέθοδο εκτίμησης της προόδου μιας εργασίας συναρμολόγησης καθώς αυτή εκτελείται από ένα χρήστη, χρησιμοποιώντας τη ροή εικόνων που προσφέρει ο αισθητήρας του Kinect. Εδώ, θα δείξουμε πώς αυτή εφαρμόστηκε πρακτικά. Πρώτα περιγράφουμε τον τρόπο κατασκευής του βασικού σκελετού του συστήματος (ενότητα 6.1), που περιλαμβάνει τα δύο στάδια της παρακολούθησης και τη μονάδα ελέγχου των συνδέσεων, και στη συνέχεια παρουσιάζουμε τρεις ολοκληρωμένες εφαρμογές που υλοποιήσαμε (ενότητα 6.2).

6.1 Υλοποίηση

Το σύστημα υποβοήθησης συναρμολόγησης υλοποιήθηκε χρησιμοποιώντας το λογισμικό Robot Operating System (ROS) [33]. Μολονότι η εφαρμογή δεν ανήκει αυστηρά στον τομέα της ρομποτικής, πολλά από τα εργαλεία που προσφέρει το ROS αποδείχθηκαν κατάλληλα για το χειρισμό δεδομένων από το Kinect, πληροφορίας πόζας, ασύγχρονης επικοινωνίας μεταξύ των διεργασιών και οπτικοποίηση της εξόδου. Ακολουθεί μια συνοπτική περιγραφή του ROS και των βασικών του στοιχείων.

6.1.1 ROS

Το ROS είναι ένα δωρεάν, ανοικτού κώδικα σύνολο εργαλείων και βιβλιοθηκών σχεδιασμένου για τη διευκόλυνση της ανάπτυξης και του ελέγχου των ρομπότ. Ονομάζεται λειτουργικό σύστημα, διότι όπως και ένα λειτουργικό σύστημα υπολογιστή, το ROS προσφέρει υπηρεσίες όπως είναι η διαχείριση μνήμης, το hardware abstraction, η διαχείριση διεργασιών και ο παραλληλισμός. Σε ένα υψηλότερο επίπεδο, το ROS υποστηρίζει την επικοινωνία μεταξύ των διεργασιών και περιλαμβάνει ένα πλήθος εργαλείων για την υποβοήθηση τυπικών ρομποτικών εργασιών όπως τον εντοπισμό, την χαρτογράφηση, την πλοήγηση και την κατανόηση οπτικών δεδομένων. Είναι modular, αφού κάθε μονάδα λογισμικού περιέχεται σε ένα *πακέτο*, που σημαίνει ότι μπορούν να συνεργαστούν εύκολα ειδικοί σε διάφορους σχετικούς επιστημονικούς τομείς για την κατασκευή σύνθετων ρομποτικών συστημάτων. Στις βασικές έννοιες του ROS περιλαμβάνονται οι κόμβοι, τα

θέματα, τα μηνύματα και οι υπηρεσίες.

Κόμβοι (Nodes)

Ο κόμβος αποτελεί στιγμιότυπο μιας διεργασίας. Οι κόμβοι μπορεί να προσφέρουν δεδομένα αισθητήρων, να εκτελούν διάφορους αλγορίθμους ή να ενεργοποιούν επενεργητές.

Μηνύματα (Messages)

Οι κόμβοι του ROS επικοινωνούν μεταξύ τους μέσω μηνυμάτων. Τα μηνύματα είναι απλές δομές δεδομένων, που περιλαμβάνουν πρωτόγονους τύπους δεδομένων (ακεραίους, λογικές μεταβλητές, χαρακτήρες, αριθμούς κινητής υποδιαστολής κλπ.) και άλλα μηνύματα, που σχηματίζουν εμφωλευμένες δομές. Τα μηνύματα μπορούν να μεταβιβαστούν ασύγχρονα μέσω θεμάτων, ή σύγχρονα μέσω κλήσεων υπηρεσιών.

Θέματα (Topics)

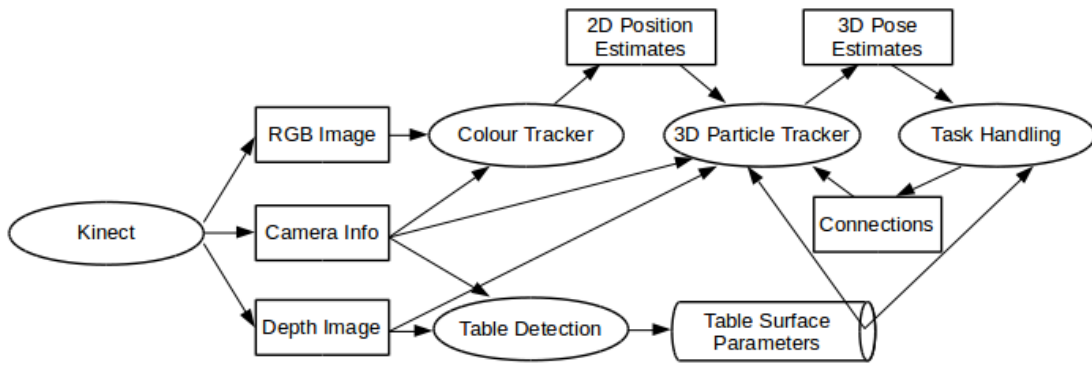
Τα θέματα είναι δίαυλοι μεταφοράς δεδομένων, μέσω των οποίων οι κόμβοι μπορούν να ανταλλάξουν μηνύματα. Ένας ή περισσότεροι κόμβοι μπορούν να δημοσιεύσουν μηνύματα σε ένα θέμα. Τα μηνύματα αυτά διαβάζονται από άλλους κόμβους που έχουν εγγραφεί στο θέμα. Η μεταφορά δεδομένων γίνεται ασύγχρονα, που σημαίνει ότι το μήνυμα μπορεί να δημοσιευτεί όποτε υπάρχει διαθέσιμη καινούργια πληροφορία, και να διαβαστεί σε οποιαδήποτε άλλη χρονική στιγμή από κάποιον άλλο κόμβο.

Υπηρεσίες (Services)

Όταν απαιτείται σύγχρονη μεταφορά μηνυμάτων, οι κόμβοι θα πρέπει να χρησιμοποιήσουν υπηρεσίες και όχι θέματα. Μια υπηρεσία λειτουργεί όπως μια απομακρυσμένη κλήση διαδικασίας. Χρησιμοποιείται ένα ζεύγος μηνυμάτων, ένα για το αίτημα και ένα για την απάντηση. Ο κόμβος που προσφέρει την υπηρεσία την διαφημίζει, και ένας πελάτης καλεί την υπηρεσία αποστέλλοντας ένα μήνυμα αιτήματος. Όταν η υπηρεσία ολοκληρώνει την επεξεργασία του αιτήματος, στέλνει ένα μήνυμα απάντησης στον πελάτη.

6.1.2 Αρχιτεκτονική του συστήματος υποστήριξης συναρμολόγησης

Όπως προαναφέρθηκε, υλοποιήσαμε τον πράκτορα υποβοήθησης συναρμολόγησης πάνω στο σκελετό του ROS. Δημιουργήσαμε διαφορετικό κόμβο για κάθε υποσύστημα: έναν για την ανίχνευση του τραπεζιού, έναν για την παρακολούθηση βάση χρωμάτων, έναν για την 3D παρακολούθηση και έναν τελικό για το χειρισμό της εργασίας συναρμολόγησης. Η αρχιτεκτονική του συστήματος φαίνεται στο σχήμα 6.1. Οι κόμβοι αναπαρίστανται με ελλείψεις, τα θέματα με ορθογώνια και η υπηρεσία με κύλινδρο. Όπως



Σχήμα 6.1: Αρχιτεκτονική ROS του συστήματος.

φαίνεται, ο κόμβος που αντιστοιχεί στον αισθητήρα του Kinect δημοσιεύει τις εικόνες και τις σχετικές πληροφορίες των καμερών (αναλύσεις και μήτρες προβολής) στα αντίστοιχα θέματα, που στη συνέχεια διαβάζονται από τους υπόλοιπους κόμβους του συστήματος. Ο κόμβος παρακολούθησης βάση χρώματος δημοσιεύει τις συντεταγμένες των pixels στα οποία φαίνονται τα αντικείμενα, οι οποίες διαβάζονται από τον κόμβο 3D παρακολούθησης. Το υποσύστημα της 3D παρακολούθησης διαβάζει, επίσης, τις πληροφορίες συνδέσεων (βλέπε ενότητα 5.3) που δημοσιεύονται από τον κόμβο χειρισμού της εργασίας. Όταν ολοκληρώσει την εκτίμηση πόζας των αντικειμένων, ο κόμβος παρακολούθησης δημοσιεύει την πληροφορία στο αντίστοιχο θέμα, για ανάγνωση από τον τελικό κόμβο. Τέλος, ο κόμβος ανίχνευσης του τραπέζιού υπολογίζει τις παραμέτρους της επιφάνειας του τραπέζιού και προσφέρει μια υπηρεσία για μοιραστεί τις παραμέτρους αυτές με άλλους κόμβους. Χρησιμοποιήθηκε υπηρεσία έναντι θέματος, για να αποφευχθούν μη απαραίτητες δημοσιεύσεις, αφού οι παράμετροι χρειάζονται ανάγνωση μόνο μία φορά από τους άλλους κόμβους (υποθέτοντας σταθερή θέση κάμερας).

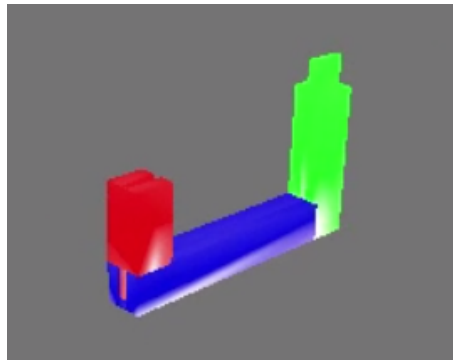
6.2 Πειράματα

Έχοντας κατασκευάσει τον σκελετό της επίβλεψης συναρμολόγησης, μπορούμε να δείξουμε έναν αριθμό εφαρμογών του συστήματός μας. Διεξάγαμε τρία πειράματα, το καθένα με άλλο σκοπό. Τα πειράματα περιγράφονται λεπτομερώς στη συνέχεια μαζί με το σκοπό του καθενός.

6.2.1 1ο Πείραμα - Guidance Mode

Στο πείραμα αυτό αξιολογούμε την ικανότητα του συστήματος να εκτιμήσει δυναμικά την κατάσταση μιας εργασίας συναρμολόγησης. Προκειμένου να δικαιολογήσουμε τις σχεδιαστικές επιλογές σε κάθε στάδιο του συστήματος, επιλέξαμε ένα αντικείμενο που

αποτελείται από μη-κυβικά μέρη. Τα μέρη αυτά είναι ράβδοι που προέρχονται από 3Δ εκτύπωση και χρησιμοποιούνται στο πρόγραμμα BabyRobot (σχήμα 6.2).



Σχήμα 6.2: 1ο Πείραμα: Ζητούμενη κατασκευή.

Όπως φαίνεται στο σχήμα, τα αντικείμενα είναι διαφορετικών χρωμάτων, οπότε χρησιμοποιήσαμε τον αλγόριθμο της ενότητας 3.2 και όχι αυτόν της ενότητας 3.3.

Μετά από επίδειξη των βημάτων της συναρμολόγησης, ζητήθηκε από τους χρήστες να εκτελέσουν αυτά τα βήματα μπροστά από το Kinect. Το σύστημα επιχείρησε να εκτιμήσει πότε κάθε βήμα είχε ολοκληρωθεί, εμφανίζοντας σχετικό μήνυμα κάθε φορά.

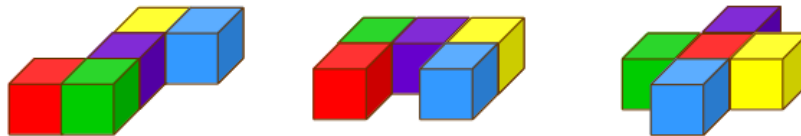
6.2.2 2ο Πείραμα – Guess Mode

Εδώ ο χρήστης έχει τη δυνατότητα, χρησιμοποιώντας 5 τουβλάκια, διαστάσεων $3.6 \times 3.6 \times 3.6$ cm, να κατασκευάσει μία από τις τρεις δομές που φαίνονται στο σχήμα 6.3. Ο χρήστης επιλέγει μία από τις δομές, και όσο την κατασκευάζει το σύστημα επιχειρεί να μαντέψει ποια εκ των τριών δομών είναι πιο πιθανό να έχει επιλεγεί.

Ο σκοπός του πειράματος είναι να δειχθεί πώς θα μπορούσε το υλοποιημένο σύστημα να εκτελέσει κάποιας μορφής **αναγνώριση πρόθεσης**. Στο [34], η αναγνώριση πρόθεσης ορίζεται ως "το πρόβλημα της αναγνώρισης των προθέσεων ενός πράκτορα αναλύοντας μερικές ή όλες τις δράσεις του ή/και αναλύοντας τις αλλαγές στην κατάσταση του περιβάλλοντός του που προκαλούνται από τις δράσεις του". Ο στόχος της αναγνώρισης πρόθεσης είναι η ενίσχυση της συνεργασίας του υπολογιστή με τον άνθρωπο, κατανοώντας τους σκοπούς των ενεργειών του ανθρώπου. Στο συγκεκριμένο πείραμα, η πρόθεση του χρήστη αφορά την επιλογή της κατασκευής που επιχειρεί να φτιάξει.

Μια σημαντική διαφορά ανάμεσα σ' αυτό το πείραμα και το προηγούμενο είναι ότι η σειρά με την οποία ο χρήστης εκτελεί κάθε βήμα είναι αυθαίρετη. Αυτό σημαίνει πως πρέπει να ελέγξουμε όλες τις επιτρεπτές συνδέσεις σε κάθε επανάληψη, και όχι απλά την παρούσα σύνδεση όπως προηγουμένως. Για κάθε εργασία χρησιμοποιούμε το δυαδικό διάνυσμα v_c της ενότητας 5.1.2. Όταν αναγνωριστεί μια σύνδεση, μεταβάλλεται η αντίστοιχη τιμή του διανύσματος από 0 σε 1.

Τα δυαδικά διανύσματα που προκύπτουν μπορούν να χρησιμοποιηθούν για να μαντέψει το σύστημα ποια δομή επιχειρεί να κατασκευάσει ο χρήστης. Μετρώντας τον αριθμό των άσπων στο κάθε διάνυσμα, προκύπτει ο αριθμός των συνδέσεων κάθε δομής που έχουν ολοκληρωθεί. Επειδή κάθε δομή έχει τον ίδιο αριθμό απαιτούμενων συνδέσεων (τέσσερις), αυτή με τις περισσότερες ολοκληρωμένες συνδέσεις επιλέγεται ως πιο πιθανή επιλογή του χρήστη. Σε περίπτωση ισοβαθμίας, επιλέγεται αυτή που ήταν πιο πιθανή στο τελευταίο χρονικό βήμα που δεν υπήρχε ισοβαθμία.



Σχήμα 6.3: 2ο Πείραμα: Επιτρεπτές δομές.

6.2.3 3ο Πείραμα – Learning Mode

Στο πείραμα αυτό δείχνουμε πώς το σύστημά μας μπορεί να αξιοποιηθεί για εκπαιδευτικούς σκοπούς. Όπως προηγουμένως, ο χρήστης έχει στη διάθεσή του έναν αριθμό από τουβλάκια. Παρουσιάζεται στο χρήστη μια πρόταση στα Ελληνικά, και καλείται να σχηματίσει σωστά την αντίστοιχη πρόταση στα Ισπανικά. Κάθε τουβλάκι αντιστοιχεί σε μία λέξη των Ισπανικών. Ο στόχος του χρήστη είναι να τοποθετήσει τα τουβλάκια στη σωστή σειρά, ώστε να κατασκευάσει την πρόταση που δίνεται. Οι λέξεις προβάλλονται στις αναπαραστάσεις των κύβων στην οθόνη. Έτσι, πολλές διαφορετικές προτάσεις μπορούν να εμφανιστούν χωρίς να χρειαστεί αλλαγή στον φυσικό κόσμο, πχ. κολλώντας και ξεκολλώντας αυτοκόλλητα ή χρησιμοποιώντας διαφορετικό σύνολο τουβλακιών για κάθε πρόταση.

Ο αριθμός των κύβων μπορεί να είναι μεγαλύτερος από τον αριθμό των λέξεων που εμφανίζονται στις προτάσεις. Στην περίπτωση αυτή ανατίθενται στα υπόλοιπα τουβλάκια λέξεις που δε σχετίζονται με την πρόταση ή είναι γραμματικά λανθασμένες. Τα τουβλάκια αυτά δεν χρειάζεται να συνδεθούν στα υπόλοιπα για να ολοκληρωθεί η εργασία, και δε λαμβάνονται υπόψη κατά την αξιολόγηση της προόδου του χρήστη. Η ανάθεση των λέξεων στους κύβους γίνεται με τυχαίο τρόπο κάθε φορά, στην αρχή του πειράματος.

Οι προτάσεις επιλέχθηκαν προσεκτικά ώστε να μην επιτραπούν πολλαπλές σωστές απαντήσεις. Βέβαια, αυτό θα μπορούσε να αντιμετωπιστεί εύκολα ελέγχοντας έναν αριθμό εργασιών ταυτόχρονα, με την κάθε εργασία να αναπαριστά μια διαφορετική σωστή απάντηση. Ωστόσο, αυτό δεν εμπίπτει στους σκοπούς του πειράματος, οπότε δεν υλοποιήθηκε.

Τα τουβλάκια ήταν πάλι διαφορετικών χρωμάτων, οπότε χρησιμοποιήσαμε τον αντίστοιχο αλγόριθμο ανίχνευσης. Δοκιμάσαμε το πείραμα και με κύβους ίδιου χρώματος, αλλά τα αποτελέσματα δεν ήταν ιδιαίτερα ικανοποιητικά, οπότε τα παραλείψαμε από την εργασία αυτή. Σημειώνεται ότι χρησιμοποιήθηκαν τα ίδια τουβλάκια με το 2ο πείραμα, δηλαδή μεγέθους $3.6 \times 3.6 \times 3.6\text{cm}$.

6.2.4 Πειραματική διαδικασία

Έχει προαναφερθεί ότι ένα από τα χαρακτηριστικά του συστήματός μας είναι η ικανότητά του να λειτουργεί online. Ωστόσο, προκειμένου να αξιολογήσουμε το σύστημα σε διάφορες παραλλαγές και με διαφορετικές λειτουργικές παραμέτρους, ήταν αναγκαίο να καταγράψουμε μια σειρά από βίντεο, με τα οποία θα μπορούσαμε να πειραματιστούμε offline. Φροντίσαμε ώστε η επεξεργασία των καταγραφών να γίνεται πάντα σε ρυθμούς πραγματικού χρόνου, για να βεβαιωθούμε ότι το τελικό βελτιστοποιημένο σύστημα θα είναι ικανό να λειτουργήσει επίσης σε πραγματικό χρόνο.

Καλέσαμε έξι άτομα για την εκτέλεση των πειραμάτων που περιγράψαμε παραπάνω. Το πρώτο πείραμα εκτελέστηκε 3 φορές από τον καθένα, το δεύτερο πάλι 3 (μία για κάθε κατασκευή) και το τρίτο 2 (με προτάσεις 4 και 5 λέξεων). Με τον τρόπο αυτό, μπορέσαμε να εξετάσουμε διαφορετικούς τρόπους με τους οποίους οι χρήστες αντιδρούν στο περιβάλλον των πειραμάτων και εκτελούν τις εργασίες, ενώ ταυτόχρονα κατασκευάσαμε ένα σχετικά ποικιλόμορφο σύνολο δεδομένων, πάνω στο οποίο πειραματιστήκαμε στη συνέχεια. Το αποτέλεσμα είναι ένα σύστημα πιο γενικό ως προς την αντιμετώπιση διαφορετικών συμπεριφορών των χρηστών. Κατά την αξιολόγηση του συστήματος παρακάτω, παραλείψαμε τις καταγραφές κατά τις οποίες ο χρήστης έκανε λάθος, οπότε τελικά εργαστήκαμε σε 15, 16 και 12 καταγραφές στα τρία πειράματα αντίστοιχα.

Οι παράμετροι που μεταβλήθηκαν ήταν: Ο συντελεστής εισόδου b και ο συντελεστής ταχύτητας k_v της εξίσωσης (4.26), η τυπική απόκλιση γραμμικής ταχύτητας σ_v και γωνιακής ταχύτητας σ_ω του θορύβου ω της (4.26), τα περιθώρια σφάλματος $m_{p,x}$, $m_{p,y}$ και $m_{p,z}$ καθώς και το K της (5.15), το μέγιστο αριθμό particles N_{max} που μπορούν να ανατεθούν σε μη ολοκληρωμένη σύνδεση και τον αριθμό συνεχόμενων πλαισίων N_{checks} στα οποία πρέπει να ικανοποιείται μια σχετική πόζα, ώστε να θεωρηθεί ολοκληρωμένη η αντίστοιχη σύνδεση.

Ο συνολικός αριθμός των particles που χρησιμοποιεί ο αλγόριθμος παρακολούθησης καθορίστηκε μειώνοντας τον, μέχρι το σημείο που δεν χάνονται πλαίσια εικόνων του Kinect λόγω υψηλών υπολογιστικών απαιτήσεων. Συγκεκριμένα, βρέθηκε ότι για $N = 225$ μπορούν να αξιοποιηθούν όλες οι εικόνες σε πραγματικό χρόνο.

6.3 Αποτελέσματα

Εδώ παρουσιάζουμε τα αποτελέσματα των πειραμάτων που εκτελέστηκαν πάνω στις καταγραφές. Μας ενδιαφέρουν τα ποιοτικά και τα ποσοτικά αποτελέσματα, με τα πρώτα να μας προσφέρουν μια διαίσθηση των ικανοτήτων και των αδυναμιών του συστήματος, και τα δεύτερα μια πιο τυπική αξιολόγησή του.

6.3.1 Ποιοτικά αποτελέσματα

Εδώ παρουσιάζουμε κάποια ενδεικτικά στιγμιότυπα του συστήματος και για τα τρία πειράματα. Στο σχήμα 6.4 φαίνεται το πρώτο πείραμα, στο σχήμα 6.5 το δεύτερο και στο σχήμα 6.6 το τρίτο. Σε γενικές γραμμές η θέσεις των αντικειμένων εκτιμούνται σωστά, ενώ οι προσανατολισμοί είναι λίγο δυσκολότερο να εκτιμηθούν, διότι το πρώτο στάδιο της παρακολούθησης δίνει πληροφορία μόνο για τις θέσεις. Ωστόσο, ακόμη κι έτσι οι συνδέσεις αναγνωρίζονται με ικανοποιητική ακρίβεια.

6.3.2 Ποσοτικά αποτελέσματα

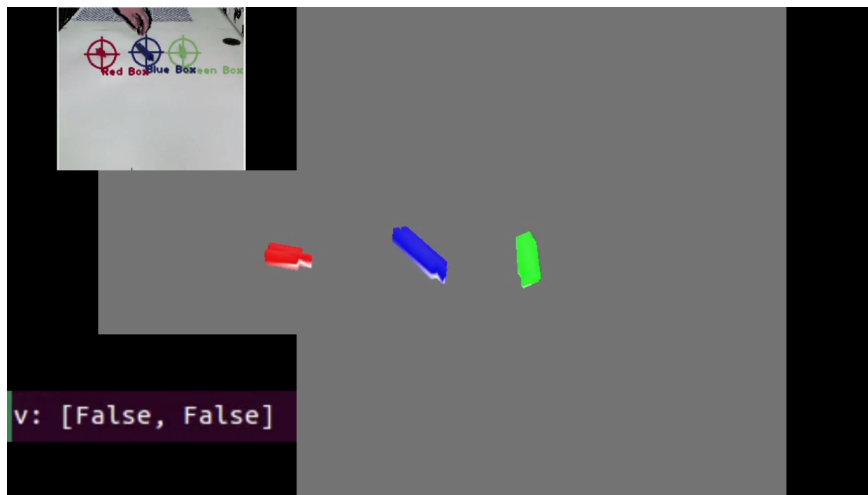
Άλλες σχετικές εργασίες εξετάζουν το βαθμό στον οποίο η αυτόματη επίβλεψη διευκολύνει το χρήστη, έναντι άλλων μεθόδων (γραπτές οδηγίες, βίντεο επίδειξης κλπ.). Εδώ επικεντρωθήκαμε στην ικανότητα του συστήματος να κατανοήσει την εξέλιξη της εργασίας, και όχι στη χρησιμότητά του για το χρήστη. Το δεύτερο έχει ήδη μελετηθεί, και δεν αναμένουμε σημαντικές διαφορές και στο δικό μας σύστημα. Αυτό που αξίζει να διερευνηθεί είναι η αποτελεσματικότητα του ίδιου του συστήματος.

Χειροκίνητα καταγράψαμε τις χρονικές στιγμές ολοκλήρωσης όλων των συνδέσεων σε όλες τις καταγραφές. Αυτές έπειτα συγκρίθηκαν με τους χρόνους που εντόπισε το σύστημά μας. Επειδή πρέπει να υπάρχει μια ελαστικότητα στο χρόνο αναγνώρισης κάθε σύνδεσης, υπολογίζουμε το ποσοστό των συνδέσεων που εντοπίστηκαν σωστά εντός ενός χρονικού περιθωρίου.

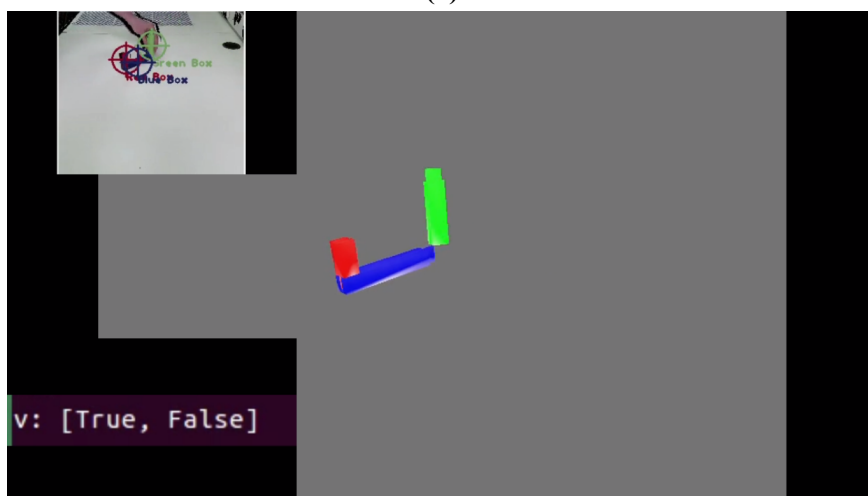
1ο Πείραμα

Μετά από πειραματισμό, το βέλτιστο αποτέλεσμα επιτεύχθη για: $b = 0.5$, $k_v = 0.3$, $\sigma_v = 1\text{cm}/\beta\eta\mu\alpha$, $\sigma_W = 0.1\text{rad}/\beta\eta\mu\alpha$, $m_p = 4\text{cm}$, $D_{KL,max} = 0.5$, $K = 7$, $N_{max} = 12$ και $N_{checks} = 4$. Οι παράμετροι της 3Δ παρακολούθησης επιλέχθηκαν ώστε να είναι όσο το δυνατόν ακριβέστερες οι εκτιμήσεις θέσης και προσανατολισμού. Στο σχήμα 6.7 φαίνεται ο αριθμός των συνδέσεων που αναγνωρίστηκαν νωρίς, εντός χρονικού περιθωρίου, αργά και ποτέ, για 4 διαφορετικά περιθώρια.

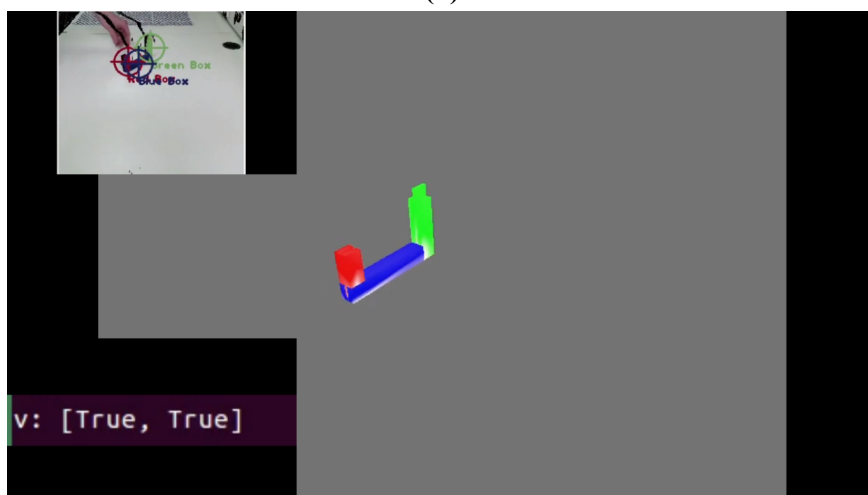
Στον πίνακα 6.1 καταγράφονται κάποια στοιχεία και χαρακτηριστικά κάθε καταγραφής. Ο χρόνος ολοκλήρωσης υπολογίζεται από την αρχή της καταγραφής μέχρι τη στιγμή της τελευταίας σύνδεσης. Αναφέρεται ο βαθμός επικάλυψων των αντικειμένων, αν τα



(a)

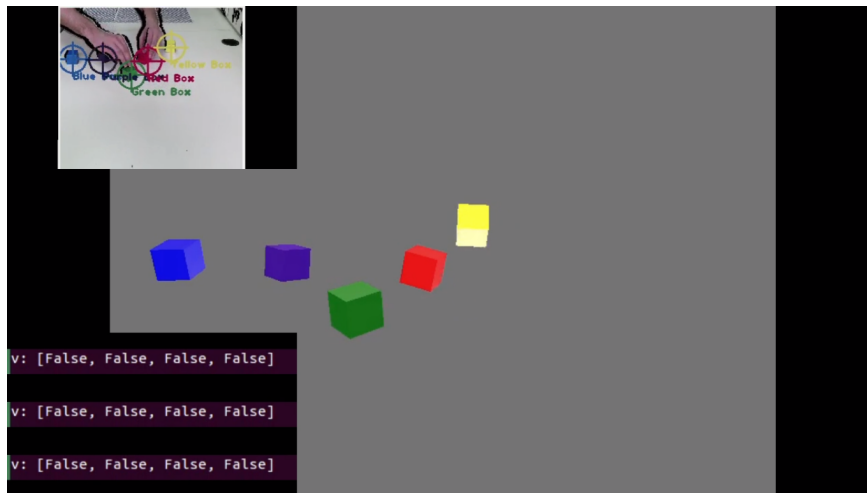


(b)

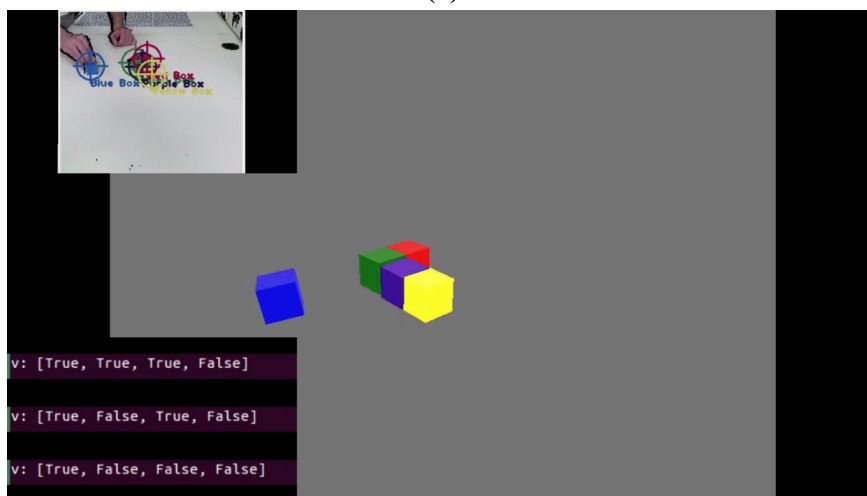


(c)

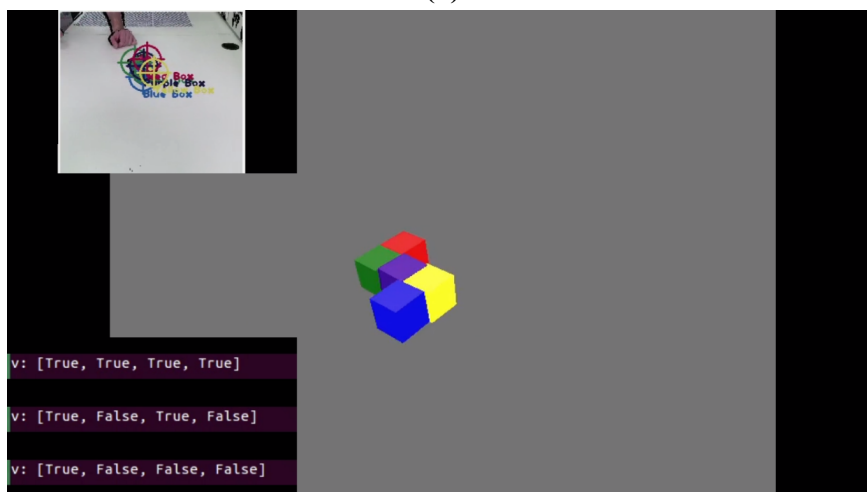
Σχήμα 6.4: Τρία στιγμιότυπα του 1ου πειράματος σε διαφορετικά στάδια ολοκλήρωσης της κατασκευής: (a) στην αρχική κατάσταση, (b) μετά την πρώτη σύνδεση, (c) μετά την ολοκλήρωση της εργασίας. Πάνω αριστερά φαίνεται το αποτέλεσμα της ανίχνευσης χρώματος, στο κέντρο οι εκτιμώμενες πόζες και κάτω αριστερά η κατάσταση των συνδέσεων σε κάθε στιγμιότυπο.



(a)



(b)

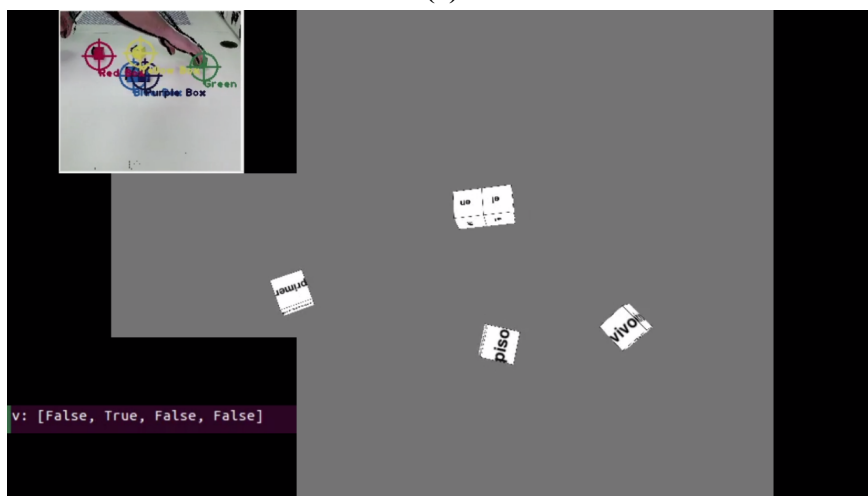


(c)

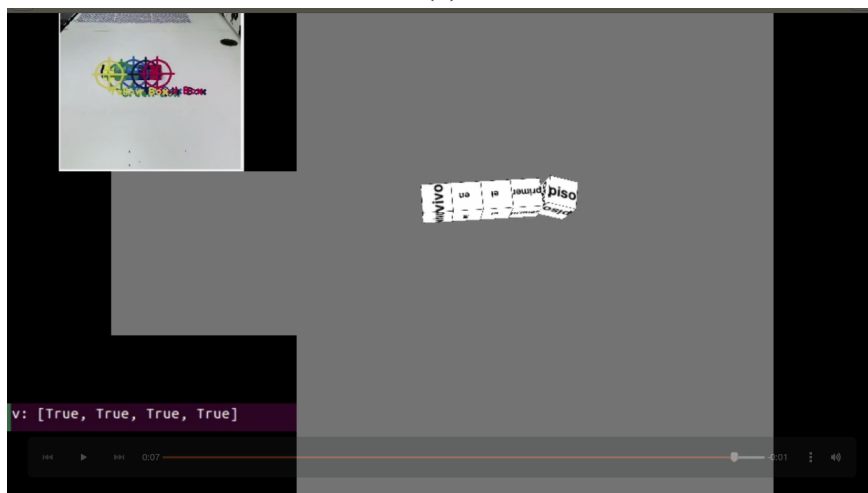
Σχήμα 6.5: Τρία στιγμιότυπα του 2ου πειράματος σε διαφορετικά στάδια: (a) πριν πραγματοποιηθεί η πρώτη σύνδεση, (b) μετά την τρίτη σύνδεση, (c) μετά την ολοκλήρωση της εργασίας. Ο χρήστης επέλεξε την πρώτη κατασκευή, η οποία αναγνωρίζεται σωστά. Πάνω αριστερά φαίνεται το αποτέλεσμα της ανίχνευσης χρώματος, στο κέντρο οι εκτιμώμενες πόζες και κάτω αριστερά τα διανύσματα συνδέσεων των τριών πιθανών κατασκευών σε κάθε στιγμιότυπο.



(a)

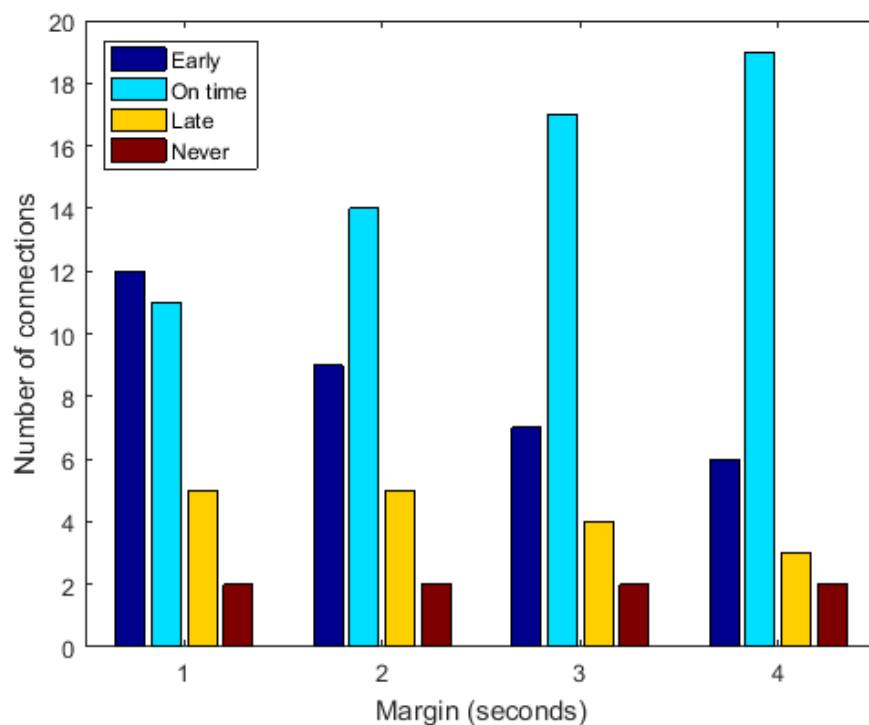


(b)



(c)

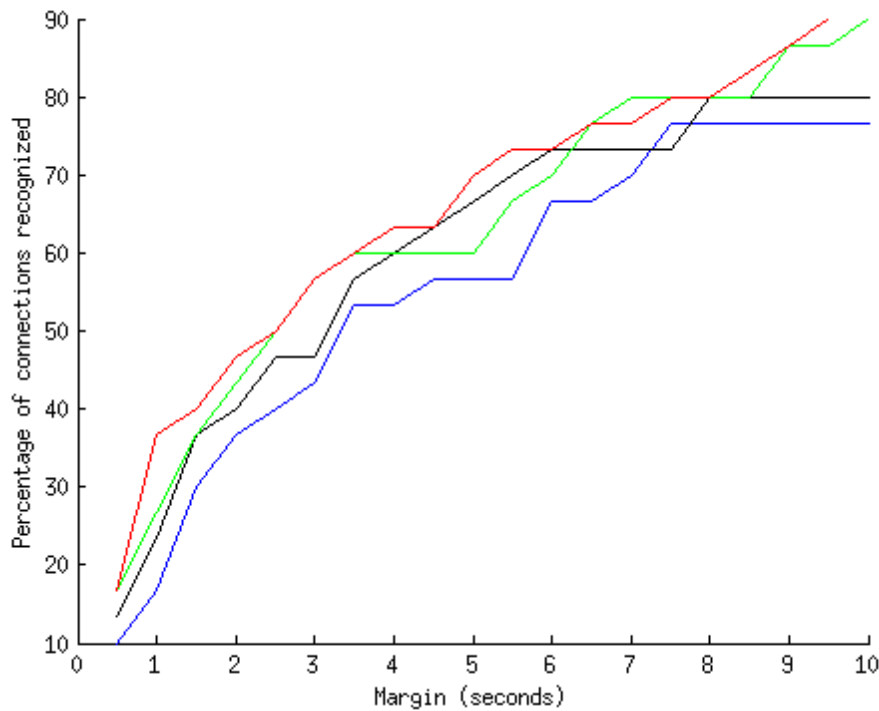
Σχήμα 6.6: Τρία στιγμιότυπα του 3ου πειράματος σε διαφορετικά στάδια: (a) στην αρχική κατάσταση, (b) μετά την πρώτη σύνδεση, (c) μετά την ολοκλήρωση της εργασίας. Η πρόταση που καλείται ο χρήστης να κατασκευάσει είναι "Vivo en el primo piso" ("μένω στον πρώτο όροφο"). Πάνω αριστερά φαίνεται το αποτέλεσμα της ανίχνευσης χρώματος, στο κέντρο η αναπαράσταση της σκηνής με τις λέξεις να προβάλλονται στα κουτιά, και κάτω αριστερά η κατάσταση των συνδέσεων σε κάθε στιγμιότυπο.



Σχήμα 6.7: Κατανομή συνδέσεων του 1ου πειράματος, για χρονικά περιθώρια 1-4 δευτερολέπτων.

Χρήστης	Καταγραφή	Χρόνος ολοκλήρωσης (sec)	Αποκρούσεις	Εκτός εικόνας	Ταχύτητα κινήσεων	Αναγνωρισμένες συνδέσεις	Μέση χρονική απόκλιση (sec)
1	1	10.97	Λίγες	Όχι	Μικρή	2	-1.17
1	2	10.94	Λίγες	Όχι	Μικρή	2	+0.03
1	3	13.35	Λίγες	Όχι	Μικρή	2	-1.88
2	1	12.31	Λίγες	Όχι	Μεγάλη	2	-1.52
2	2	11.37	Λίγες	Όχι	Μεγάλη	2	-0.05
2	3	10.16	Λίγες	Όχι	Μεσαία	2	-0.57
3	1	8.02	Αρκετές	Όχι	Μεσαία	1	-1.64
3	2	11.44	Αρκετές	Όχι	Μεγάλη	2	-1.64
3	3	5.06	Αρκετές	Όχι	Μεγάλη	2	+2.09
4	1	6.53	Πολλές	Όχι	Μεγάλη	1	+0.42
5	1	17.68	Λίγες	Όχι	Μικρή	2	-1.15
5	2	15.84	Αρκετές	Όχι	Μικρή	2	-4.61
5	3	22.44	Λίγες	Όχι	Μικρή	2	-4.09
6	1	13.71	Αρκετές	Ναι	Μεσαία	2	-5.91
6	2	5.96	Λίγες	Όχι	Μικρή	2	-0.71

Πίνακας 6.1: Στοιχεία καταγραφών του 1ου πειράματος.



Σχήμα 6.8: 1ο πείραμα: Ποσοστά σωστά ανιχνευθέντων συνδέσεων ως προς το μήκος επιτρεπτού χρονικού περιθωρίου. Κόκκινο: $K = 7$, $N_{max} = 12$, $N_{checks} = 4$. Πράσινο: $K = 9$, $N_{max} = 12$, $N_{checks} = 4$. Μαύρο: $K = 7$, $N_{max} = 8$, $N_{checks} = 4$. Μπλε: $K = 7$, $N_{max} = 12$, $N_{checks} = 6$.

αντικείμενα βγήκαν προσωρινά εκτός ορίων της εικόνας και η μέση ταχύτητα κίνησης των αντικειμένων, στην 4η, 5η και 6η στήλη αντίστοιχα. Οι δύο τελευταίες στήλες προέκυψαν για τις παραμέτρους που αναφέραμε παραπάνω και αναφέρονται στον αριθμό των συνδέσεων που αναγνωρίστηκαν μέχρι τη λήξη της καταγραφής (μέγιστο 2) και τη μέση διαφορά των εκτιμώμενων χρόνων σύνδεσης από τους πραγματικούς.

Όπως φαίνεται στο σχήμα 6.7 και στον πίνακα, κατά κύριο λόγο τα λάθη οφείλονται σε συνδέσεις που αναγνωρίστηκαν πιο νωρίς απ' ό,τι έπρεπε. Αυτό προέκυψε από το σχετικά μεγάλο περιθώριο σφάλματος m_p κατά τη σύγκριση των σχετικών θέσεων, πράγμα αναγκαίο, όμως, ώστε να μην αγνοηθούν υπαρκτές συνδέσεις. Στο σχήμα 6.8 φαίνεται το διάγραμμα ποσοστών αναγνώρισης συνδέσεων ως προς το επιτρεπτό χρονικό περιθώριο πριν και μετά τους πραγματικούς χρόνους, για $b = 0.5$, $k_v = 0.3$, $\sigma_v = 1\text{cm/βήμα}$, $\sigma_W = 0.1\text{rad/βήμα}$, $m_p = 4\text{cm}$ και $D_{KL,max} = 0.5$. Δείχνουμε, δηλαδή, ότι όταν ο έλεγχος των συνδέσεων γίνεται πιο αυστηρός, τα ποσοστά σωστά αναγνωρισμένων συνδέσεων μειώνονται.

Στο πείραμα αυτό τα ποσοστά των συνδέσεων που εντοπίστηκαν σωστά είναι χαμηλότερα απ' ό,τι στα άλλα δύο πειράματα, όπως θα φανεί παρακάτω. Αυτό συνέβη διότι οι ράβδοι που χρησιμοποιήθηκαν πρέπει να συνδεθούν σε κατάλληλες γωνίες, ενώ ήταν δύσκολο για το σύστημα να εκτιμήσει τους προσανατολισμούς τους σε κάποιες περιπτώσεις.

Ο προσανατολισμός είναι πιο δύσκολος στην εκτίμηση διότι έχει μικρότερη επίδραση στην αναμενόμενη παρατήρηση απ' ότι η θέση κάθε αντικειμένου. Στην περίπτωση των κύβων οι περιστροφές τους είναι μικρότερης σημασίας, καθώς λόγω συμμετρίας ελέγχονται ταυτόχρονα πολλές ισοδύναμες διατάξεις.

2ο Πείραμα

Στον πίνακα 6.2 παραθέτουμε τα στοιχεία κάθε καταγραφής του δεύτερου πειράματος, όπως προηγουμένως. Εδώ μας ενδιαφέρει ποια κατασκευή έχει αναγνωριστεί από το σύστημα σε κάθε βίντεο και η σύγκρισή της με την πραγματική επιλογή του χρήστη.

Στα σχήματα 6.9 και 6.10 παρουσιάζονται οι μήτρες σύγχυσης για $m_p = 1\text{cm}$ και $m_p = 1.8\text{cm}$ αντίστοιχα και για διάφορες τιμές του K . Ο βέλτιστος συνδυασμός βρέθηκε για $m_p = 1\text{cm}$ και $K = 3$, με ποσοστό επιτυχίας 93.75% (15/16). Στη μία καταγραφή που υπήρξε λάθος, το σύστημα βρήκε την ίδια πιθανότητα για τις κατασκευές 1 και 2 και επέλεξε τυχαία.

Οι υπόλοιπες παράμετροι επιλέχθηκαν ποιοτικά, με βάση την ακρίβεια της μονάδας παρακολούθησης. Συγκεκριμένα, λάβαμε $b = 0.3$, $k_v = 0.2$, $\sigma_v = 0.5\text{cm/βήμα}$, $\sigma_W = 0.05\text{grad/βήμα}$, $D_{KL,max} = 1$, $N_{max} = 9$ και $N_{checks} = 2$.

Χρήστης	Κατασκευή	Χρόνος ολοκλήρωσης (sec)	Αποκρύψεις	Εκτός εικόνας	Ταχύτητα κινήσεων	Αναγνωρισμένη κατασκευή
1	1	15.59	Λίγες	Όχι	Μικρή	1
1	2	15.25	Λίγες	Όχι	Μικρή	2
1	3	17.29	Αρκετές	Όχι	Μικρή	3
2	1	20.60	Πολλές	Όχι	Μεσαία	1
2	2	23.31	Λίγες	Όχι	Μεσαία	2
2	3	11.48	Αρκετές	Όχι	Μεγάλη	3
3	1	18.10	Πολλές	Όχι	Μεσαία	1
3	2	13.26	Αρκετές	Όχι	Μεσαία	2
3	3	12.70	Λίγες	Όχι	Μεσαία	3
4	1	15.56	Πολλές	Όχι	Μεγάλη	1
4	2	16.18	Πολλές	Όχι	Μεσαία	2
4	3	13.87	Πολλές	Όχι	Μεσαία	3
5	1	21.32	Λίγες	Ναι	Μικρή	2
5	2	20.46	Λίγες	Ναι	Μικρή	2
5	3	22.16	Λίγες	Ναι	Μικρή	3
6	1	13.32	Λίγες	Όχι	Μικρή	1

Πίνακας 6.2: Στοιχεία καταγραφών του 2ου πειράματος.

Confusion Matrix

Output Class	1	5 31.2%	1 6.2%	2 12.5%	62.5% 37.5%
	2	1 6.2%	4 25.0%	0 0.0%	80.0% 20.0%
	3	0 0.0%	0 0.0%	3 18.8%	100% 0.0%
		83.3% 16.7%	80.0% 20.0%	60.0% 40.0%	75.0% 25.0%
		1	2	3	
		Target Class			

(a) $K = 2$

Confusion Matrix

Output Class	1	5 31.2%	0 0.0%	0 0.0%	100% 0.0%
	2	1 6.2%	5 31.2%	0 0.0%	83.3% 16.7%
	3	0 0.0%	0 0.0%	5 31.2%	100% 0.0%
		83.3% 16.7%	100% 0.0%	100% 0.0%	93.8% 6.2%
		1	2	3	
		Target Class			

(b) $K = 3$

Confusion Matrix

Output Class	1	2 12.5%	1 6.2%	1 6.2%	50.0% 50.0%
	2	3 18.8%	4 25.0%	0 0.0%	57.1% 42.9%
	3	1 6.2%	0 0.0%	4 25.0%	80.0% 20.0%
		33.3% 66.7%	80.0% 20.0%	80.0% 20.0%	62.5% 37.5%
		1	2	3	
		Target Class			

(c) $K = 5$

Confusion Matrix

Output Class	1	2 12.5%	1 6.2%	1 6.2%	50.0% 50.0%
	2	4 25.0%	4 25.0%	1 6.2%	44.4% 55.6%
	3	0 0.0%	0 0.0%	3 18.8%	100% 0.0%
		33.3% 66.7%	80.0% 20.0%	60.0% 40.0%	56.2% 43.8%
		1	2	3	
		Target Class			

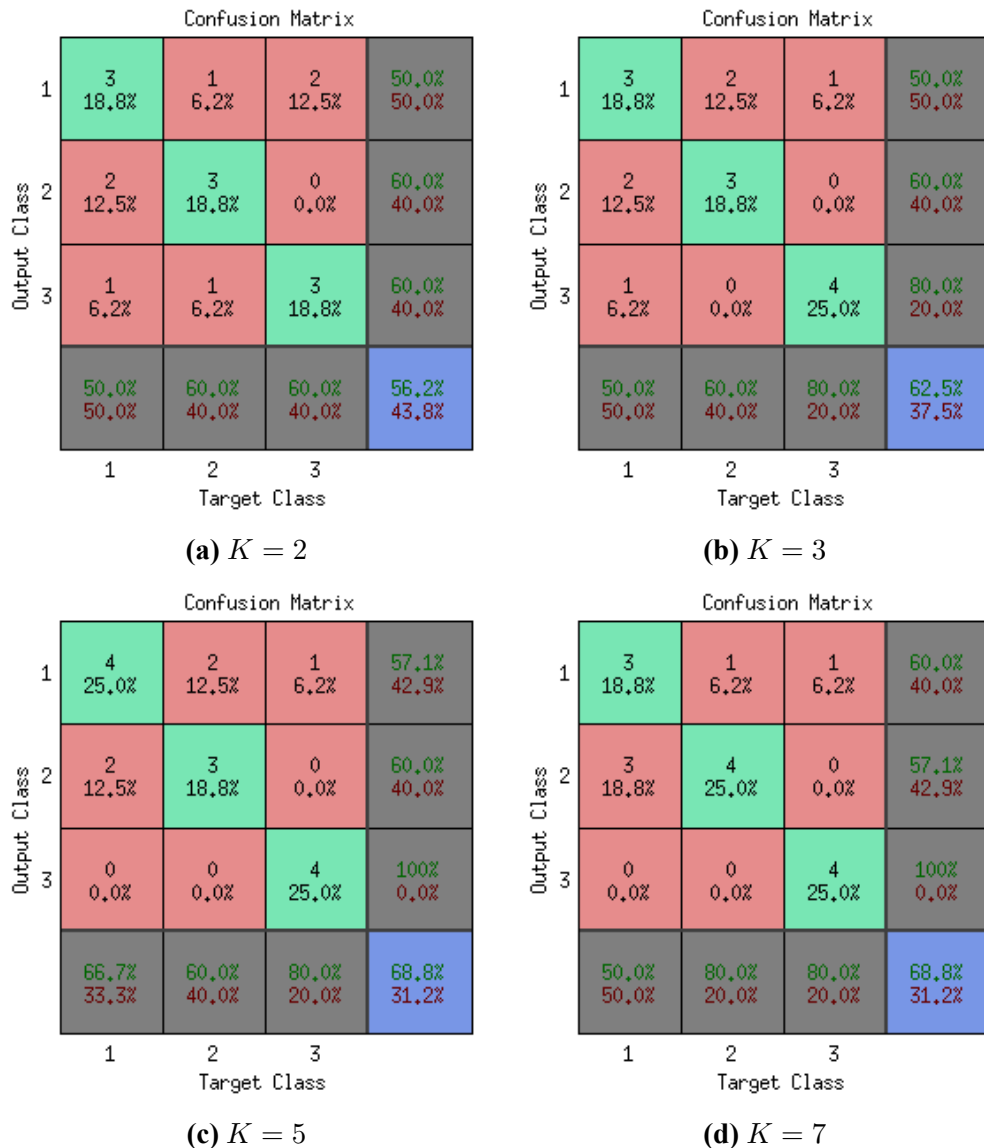
(d) $K = 7$

Σχήμα 6.9: 2ο πείραμα: Μήτρες σύγχυσης για $m_p = 1\text{cm}$.

3ο Πείραμα

Όπως και στα προηγούμενα πειράματα, παραθέτουμε τα χαρακτηριστικά κάθε καταγραφής στον πίνακα 6.3. Συγκρίνουμε πάλι τους υπολογισμένους χρόνους συνδέσεων με τους πραγματικούς. Στα σχήματα 6.11 και 6.12 φαίνονται τα ποσοστά ορθά αναγνωρισμένων συνδέσεων για $m_p = 1.8\text{cm}$ και $m_p = 1\text{cm}$ αντίστοιχα, για διάφορες τιμές του K . Και στις δύο περιπτώσεις είναι $b = 0.3$, $k_v = 0.2$, $\sigma_v = 0.5\text{cm/βήμα}$, $\sigma_W = 0.05\text{grad/βήμα}$, $D_{KL,max} = 1$, $N_{max} = 9$ και $N_{checks} = 2$. Σημειώνουμε ότι οι τιμές αυτές είναι οι ίδιες που βρέθηκαν βέλτιστες και στο 2ο πείραμα. Αυτό είναι λογικό, αφού χρησιμοποιούνται τα ίδια αντικείμενα και στα δύο πειράματα.

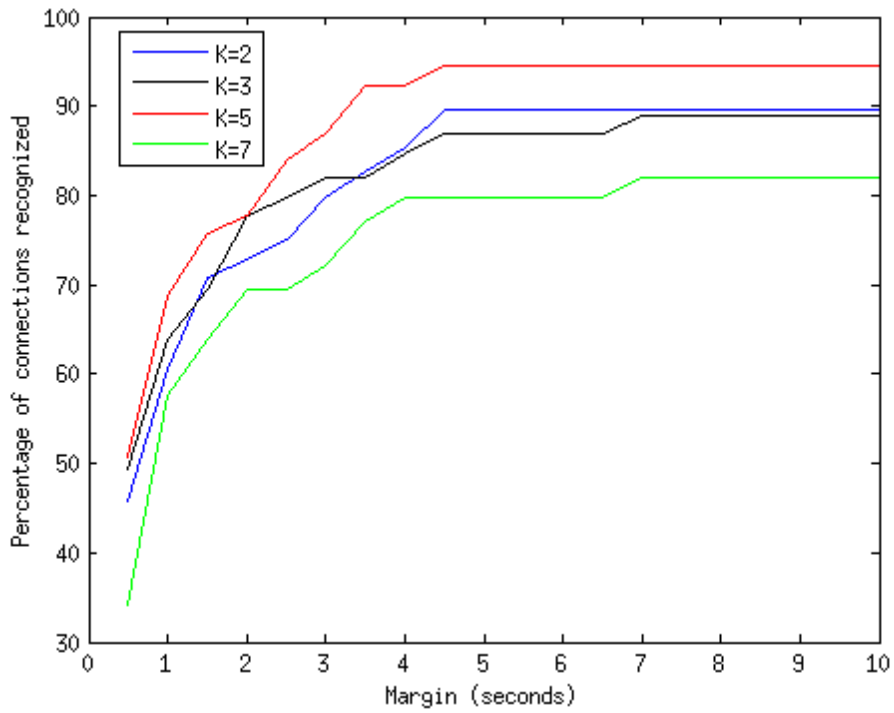
Στο σχήμα 6.13 παρουσιάζεται η κατανομή των συνδέσεων για $m_p = 1.8\text{cm}$ και $K = 5$, ενώ στο 6.14 για $m_p = 1\text{cm}$ και $K = 3$. Φαίνεται ότι ο δεύτερος συνδυασμός είναι γενικά καλύτερος.



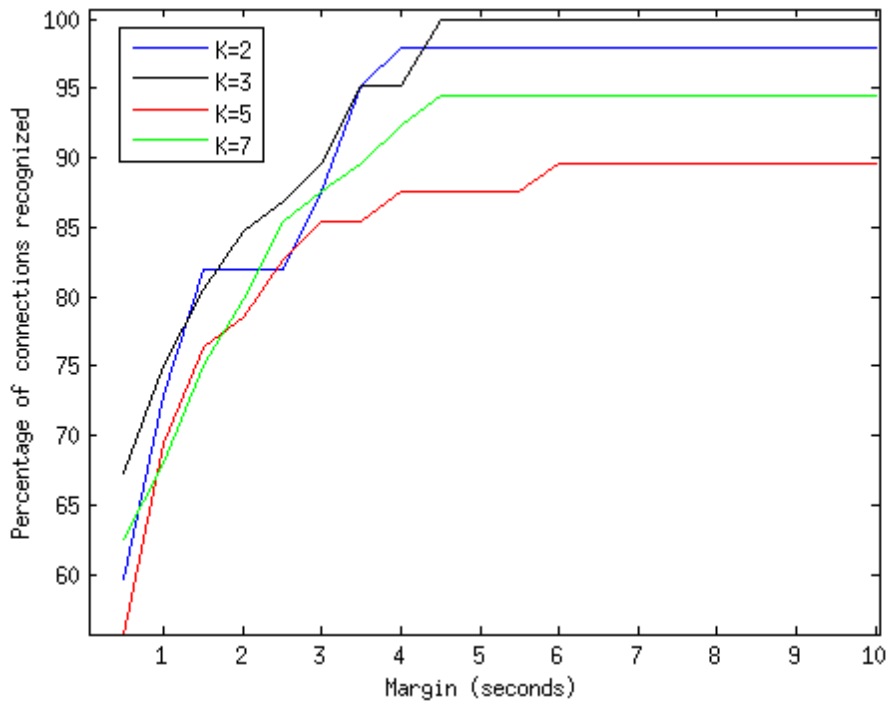
Σχήμα 6.10: 2ο πείραμα: Μήτρες σύγχυσης $m_p = 1.8\text{cm}$.

Χρήστης	Αριθμός λέξεων	Χρόνος ολοκλήρωσης (sec)	Αποκρύψεις	Εκτός εικόνας	Ταχύτητα κινήσεων	Αναγνωρισμένες συνδέσεις	Μέση χρονική απόκλιση (sec)
1	5	10.53	Λίγες	Όχι	Μικρή	4/4	-0.29
1	4	11.98	Λίγες	Όχι	Μεσαία	3/3	-0.49
2	5	11.77	Λίγες	Όχι	Μεσαία	4/4	-0.19
2	4	8.53	Λίγες	Όχι	Μεσαία	2/3	-0.16
3	5	9.55	Αρκετές	Όχι	Μεγάλη	4/4	-0.17
3	4	7.58	Λίγες	Όχι	Μεγάλη	2/3	-0.48
4	5	9.21	Αρκετές	Όχι	Μεγάλη	4/4	+0.05
4	4	8.06	Πολλές	Όχι	Μεγάλη	3/3	-1.04
5	5	24.88	Λίγες	Ναι	Μικρή	4/4	-1.41
5	4	12.89	Λίγες	Ναι	Μικρή	3/3	-0.07
6	5	8.84	Λίγες	Όχι	Μεσαία	4/4	-0.89
6	4	8.24	Λίγες	Όχι	Μικρή	3/3	+0.20

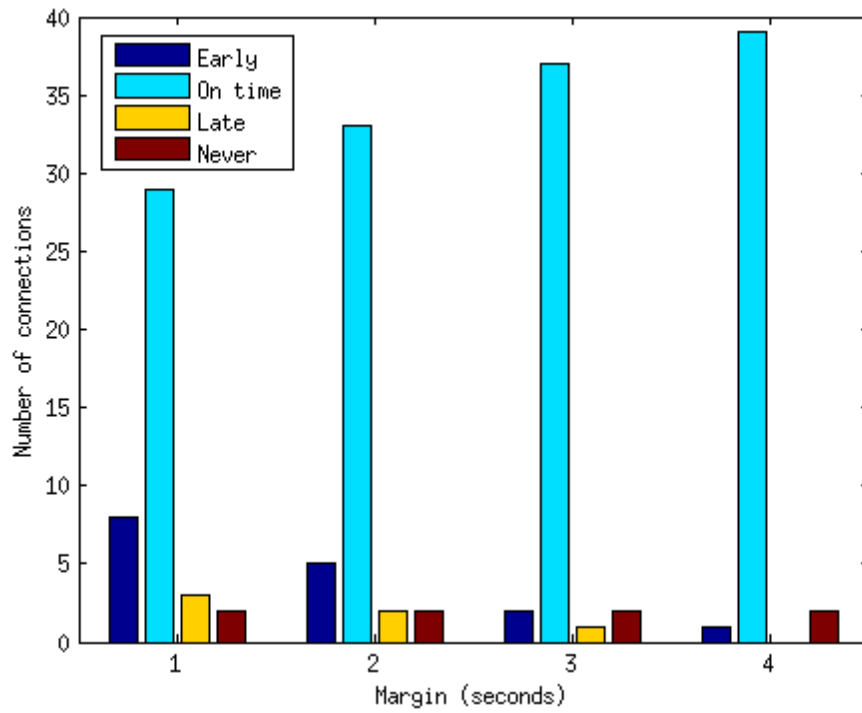
Πίνακας 6.3: Στοιχεία καταγραφών του 3ου πειράματος.



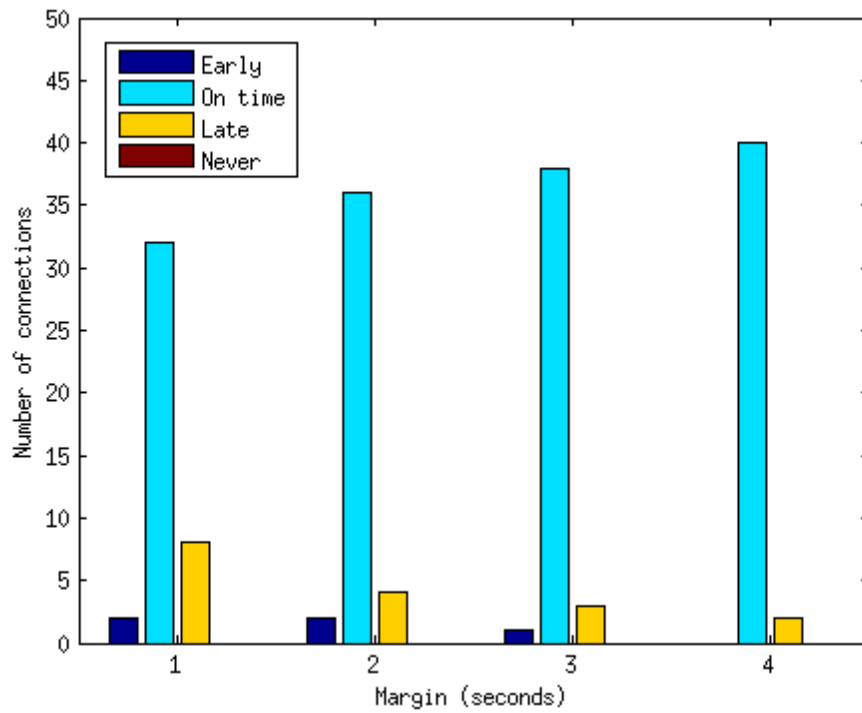
Σχήμα 6.11: 3ο πείραμα: Ποσοστά σωστά ανιχνευθέντων συνδέσεων ως προς το μήκος επιτρεπτού χρονικού περιθωρίου, για $m_p = 1.8\text{cm}$.



Σχήμα 6.12: 3ο πείραμα: Ποσοστά σωστά ανιχνευθέντων συνδέσεων ως προς το μήκος επιτρεπτού χρονικού περιθωρίου, για $m_p = 1\text{cm}$.



Σχήμα 6.13: Κατανομή συνδέσεων του 3ου πειράματος, για περιθώρια 1-4 δευτερόλεπτων, για $m_p = 1.8\text{cm}$ και $K = 5$.



Σχήμα 6.14: Κατανομή συνδέσεων του 3ου πειράματος, για περιθώρια 1-4 δευτερόλεπτων, για $m_p = 1\text{cm}$ και $K = 3$.

Κεφάλαιο 7

Συμπεράσματα – Μελλοντική εργασία

Στην εργασία αυτή καταφέραμε να σχεδιάσουμε και να κατασκευάσουμε ένα σύστημα οπτικής επίβλεψης εργασιών συναρμολόγησης. Όπως φαίνεται από τα αποτελέσματα του προηγούμενου κεφαλαίου, το σύστημα λειτουργεί αρκετά ικανοποιητικά, υπό τις συνθήκες των πειραμάτων που διεξήχθησαν. Ιδιαίτερα για εύκολα διακρίσιμα αντικείμενα, η παρακολούθηση είναι επιτυχής ακόμη και όταν αποκρύπτονται τα αντικείμενα για σημαντικά χρονικά διαστήματα. Η πλειοψηφία των συνδέσεων αναγνωρίζονται εντός αποδεκτών χρονικών ορίων. Έτσι, αποδεικνύεται η εγκυρότητα των σχεδιαστικών επιλογών που λάβαμε.

Η συμβολή της εργασίας συνοψίζεται στα εξής:

- Μελετήσαμε την εφαρμογή αλγορίθμων οπτικής παρακολούθησης για το σκοπό της υποβοήθησης συναρμολόγησης, δοκιμάζοντας αλγορίθμους που δεν έχουν ξαναχρησιμοποιηθεί για το σκοπό αυτό. Τα τελικά αποτελέσματα έδειξαν ότι τέτοιες μέθοδοι μπορούν να είναι επιτυχείς, με αξιοπιστία ανάλογη της ακρίβειας της παρακολούθησης.
- Προτείναμε έναν τρόπο συνδυασμού πληροφορίας χρώματος και βάθους για την εκτίμηση των θέσεων και προσανατολισμών πολλαπλών αντικειμένων σε πραγματικό χρόνο. Όσο γνωρίζουμε, η ενσωμάτωση 2Δ πληροφορίας θέσεων προερχόμενης από ανίχνευση χρώματος στη μεταβλητή εισόδου του δυναμικού μοντέλου κίνησης των αντικειμένων, για την εκτίμηση 6DoF ποζών με Particle Filter, αποτελεί πρωτότυπη ιδέα της εργασίας αυτής. Το σκεπτικό αυτό θα μπορούσε να εφαρμοστεί και σε άλλα παρόμοια προβλήματα με το δικό μας, που λύνονται με παρακολούθηση αντικειμένων.
- Συνοψίσαμε τις βασικές δομές αναπαράστασης εργασιών συναρμολόγησης και τους τρόπους ελέγχου των συνδέσεων μέσω των σχετικών θέσεων και προσανατολισμών των μερών της κατασκευής.
- Παρουσιάσαμε μια τροποποίηση του Particle Filter ώστε να λαμβάνει υπόψη τις συνδέσεις αντικειμένων, επιβάλλοντας τους φυσικούς περιορισμούς των ολοκλη-

ρωμένων συνδέσεων στα particles και μειώνοντας έτσι τις διαστάσεις του ζητούμενου διανύσματος κατάστασης.

- Σχεδιάσαμε τρεις απλές εφαρμογές που βασίζονται στις μεθοδολογίες που μελετήθηκαν, που αναδεικνύουν τη χρησιμότητα των συστημάτων επίβλεψης συναρμολόγησης γενικά, αλλά και συγκεκριμένα του συστήματος που υλοποιήσαμε στην παρούσα εργασία.

7.1 Μελλοντική έρευνα

Επειδή το σύστημα αποτελείται από διακριτές μονάδες, θα ήταν εύκολο να δοκιμαστούν άλλοι αλγόριθμοι παρακολούθησης ανεξάρτητα για τα δύο στάδια. Ιδιαίτερα στο πρώτο στάδιο, υπάρχουν σημαντικά περιθώρια βελτίωσης όταν εμπλέκονται αντικείμενα όμοιων χρωμάτων. Η βελτίωση θα μπορούσε να επιτευχθεί με χρήση πιο σύνθετου αλγορίθμου ανίχνευσης ή/και συσχέτισης δεδομένων, με χρήση τεχνικών μηχανικής μάθησης, συνελκτικών δικτύων κλπ. Αυτό θα ήταν χρήσιμο σε περιπτώσεις όπου το σχήμα ή η υφή των αντικειμένων παίζει πιο καθοριστικό ρόλο στην ανίχνευσή τους απ' ότι το χρώμα τους.

Στο κομμάτι της 3D παρακολούθησης, θα μπορούσαν να δοκιμαστούν άλλες μέθοδοι, ή να ενσωματωθούν κι άλλες τεχνικές στο υπάρχον σύστημα για να βελτιωθεί η εκτίμηση πόζας. Για παράδειγμα, ίσως η προσθήκη ενός τελικού σταδίου τύπου ICP να προσέδιδε μεγαλύτερη ακρίβεια στις εκτιμήσεις. Επίσης, θα μπορούσε ενδεχομένως να βρεθεί τρόπος επεξεργασίας των εικόνων βάθους ώστε το φίλτρο Bayes να λειτουργήσει με μετρήσεις υψηλότερου επιπέδου. Με κάποια άλλη μέθοδο θα μπορούσε ίσως να γίνει εφικτή η παρακολούθηση περισσότερων μερών.

Μια άλλη ενδιαφέρουσα κατεύθυνση έρευνας θα περιελάμβανε τη μελέτη άλλων τρόπων συνδυασμού των πληροφορίας του χρώματος και του βάθους. Εδώ προτείναμε την αξιοποίηση των εκτιμήσεων του πρώτου σταδίου παρακολούθησης στην φάση της πρόβλεψης του Μπεϋσιανού φίλτρου του δεύτερου σταδίου. Διαφορετικά, θα μπορούσε να ενσωματωθεί το χρώμα στο μοντέλο παρατήρησης του particle filter, και να συνυπολογιστεί στην πιθανοφάνεια κάθε δείγματος η διαφορά του μετρούμενου και του προβλεπόμενου χρώματος σε όλα τα pixels. Επίσης, θα μπορούσε να χρησιμοποιηθεί κάποια μέθοδος data fusion για να συγχωνευτούν οι εκτιμήσεις των δύο σταδίων, οι οποίες εκτιμήσεις στην περίπτωση αυτή θα προέκυπταν ανεξάρτητα. Υποθέσαμε στην εργασία αυτή ότι ο επιλεγμένος τρόπος θα είναι καταλληλότερος, αλλά ενδεχομένως με πειραματισμό να αποδειχθεί καλύτερη κάποια άλλη μέθοδος.

Τέλος, θα ήταν ιδιαίτερα χρήσιμο να διερευνηθούν άλλες πιθανές εφαρμογές των συστημάτων υποβοήθησης συναρμολόγησης όπως αυτού που υλοποιήσαμε.

Βιβλιογραφία

- [1] J. M. Molineros, “Computer vision and augmented reality for guiding assembly,” Ph.D. dissertation, Dept. CSE, The Pennsylvania State University, 2002.
- [2] M. Yuan, S. Ong, and A. Y. Nee, “Assembly guidance in augmented reality environments using a virtual interactive tool,” *Innovation in Manufacturing Systems and Technology*, 2005.
- [3] A. Nishihara and J. Okamoto, “Object recognition in assembly assisted by augmented reality system,” in *SAI Intelligent Systems Conference (IntelliSys)*. IEEE, 2015, pp. 400–407.
- [4] D. Cakmakov and E. Celakoska, “Estimation of curve similarity using turning functions,” *International Journal of Applied Mathematics*, vol. 15, pp. 403–416, 2004.
- [5] R. Radkowski, “Object tracking with a range camera for augmented reality assembly assistance,” *Journal of Computing and Information Science in Engineering*, vol. 16, no. 1, p. 011004, 2016.
- [6] A. Gupta, D. Fox, B. Curless, and M. Cohen, “Duplotrack: a real-time system for authoring and guiding duplo block assembly,” in *Proc. 25th annual ACM symposium User interface software and technology*. ACM, 2012, pp. 389–402.
- [7] A. Gupta, “Interactive playspaces for object assembly and digital storytelling,” Ph.D. dissertation, Dept. CSE, University of Washington, 2013.
- [8] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura, “The effectiveness of an AR-based context-aware assembly support system in object assembly,” in *Virtual Reality (VR)*. IEEE, 2014, pp. 57–62.
- [9] A. Miller, B. White, E. Charbonneau, Z. Kanzler, and J. J. LaViola Jr, “Interactive 3D model acquisition and tracking of building block structures,” *IEEE transactions on visualization and computer graphics*, vol. 18, no. 4, pp. 651–659, 2012.
- [10] P. J. Besl and N. D. McKay, “Method for registration of 3-D shapes,” in *Robotics-DL tentative*. International Society for Optics and Photonics, 1992, pp. 586–606.

- [11] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and vision computing*, vol. 10, no. 3, pp. 145–155, 1992.
- [12] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *2001. Proceedings. Third International Conference on 3-D Digital Imaging and Modeling*. IEEE, 2001, pp. 145–152.
- [13] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, "Comparing ICP variants on real-world data sets," *Autonomous Robots*, vol. 34, no. 3, pp. 133–148, 2013.
- [14] A. W. Fitzgibbon, "Robust registration of 2D and 3D point sets," *Image and Vision Computing*, vol. 21, no. 13, pp. 1145–1153, 2003.
- [15] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.
- [16] S. Gold, A. Rangarajan, C.-P. Lu, S. Pappu, and E. Mjolsness, "New algorithms for 2D and 3D point matching: Pose estimation and correspondence," *Pattern recognition*, vol. 31, no. 8, pp. 1019–1031, 1998.
- [17] Y. Tsing and T. Kanade, "A correlation-based approach to robust point set registration," in *European conference on computer vision*. Springer, 2004, pp. 558–569.
- [18] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision," in *Proc. Int'l Joint Conf. on Artificial Intelligence*, 1981, pp. 674–679.
- [19] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [20] H. Spies, B. Jähne, and J. L. Barron, "Range flow estimation," *Computer Vision and Image Understanding*, vol. 85, no. 3, pp. 209–231, 2002.
- [21] O. S. Gedik and A. A. Alatan, "3-D rigid body tracking using vision and depth sensors," *IEEE transactions on cybernetics*, vol. 43, no. 5, pp. 1395–1405, 2013.
- [22] P. Hebert, N. Hudson, J. Ma, T. Howard, T. Fuchs, M. Bajracharya, and J. Burdick, "Combined shape, appearance and silhouette for simultaneous manipulator and object tracking," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2405–2412.

- [23] M. Wüthrich, P. Pastor, M. Kalakrishnan, J. Bohg, and S. Schaal, “Probabilistic object tracking using a range camera,” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2013, pp. 3195–3202.
- [24] J. Issac, M. Wüthrich, C. G. Cifuentes, J. Bohg, S. Trimpe, and S. Schaal, “Depth-based object tracking using a robust gaussian filter,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 608–615.
- [25] B.-N. Vo, B.-T. Vo, N.-T. Pham, and D. Suter, “Joint detection and estimation of multiple objects from image observations,” *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5129–5141, 2010.
- [26] M. Salzmann and R. Urtasun, “Physically-based motion models for 3D tracking: A convex formulation,” in *ICCV*. IEEE, 2011, pp. 2064–2071.
- [27] N. Kyriazis and A. Argyros, “Scalable 3d tracking of multiple interacting objects,” in *CVPR*, 2014, pp. 3430–3437.
- [28] N. J. Gordon, D. J. Salmond, and A. F. Smith, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation,” in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2. IET, 1993, pp. 107–113.
- [29] J. Schwendner, “Embodied Localisation and Mapping,” Ph.D. dissertation, Universität Bremen, Bremen, 2013.
- [30] A. Doucet, N. De Freitas, K. Murphy, and S. Russell, “Rao-Blackwellised particle filtering for dynamic Bayesian networks,” in *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2000, pp. 176–183.
- [31] K. Nassar, W. Thabet, and Y. Beliveau, “Building assembly detailing using constraint-based modeling,” *Automation in Construction*, vol. 12, no. 4, pp. 365–379, 2003.
- [32] L. H. De Mello and A. C. Sanderson, “Representations of mechanical assembly sequences,” *IEEE transactions on Robotics and Automation*, vol. 7, no. 2, pp. 211–227, 1991.
- [33] M. Quigley, J. Faust, T. Foote, and J. Leibs, “ROS: an open-source Robot Operating System,” in *IEEE International Conference on Robotics and Automation (ICRA), Workshop on Open Source Software*, 2009.
- [34] F. Sadri, “Logic-based approaches to intention recognition,” in *Handbook of Research on Ambient Intelligence and Smart Environments: Trends and Perspectives*. IGI Global, 2011, pp. 346–375.

Παράρτημα Α

Ανίχνευση τραπεζιού

Στη ενότητα 4.7 παρουσιάσαμε μια μέθοδο να αποφύγουμε τη βύθιση των εκτιμώμενων θέσεων των αντικειμένων κάτω από το ύψος του τραπεζιού. Αυτό έγινε κατόπιν υπόθεσης ότι οι παράμετροι της επιπέδου του τραπεζιού θα είχαν ήδη υπολογιστεί. Επίσης, στο κεφάλαιο 5 χρησιμοποιήσαμε τις ίδιες αυτές παραμέτρους για να μετασχηματίσουμε τις πόζες των αντικειμένων στο σύστημα αναφοράς του τραπεζιού. Εδώ, περιγράφουμε τον αλγόριθμο που χρησιμοποιήθηκε για την ανίχνευση του επιπέδου του τραπεζιού και για τον υπολογισμό της εξίσωσής του.

Ξεκινάμε με την υπόθεση ότι το τραπέζι θα έχει τη μεγαλύτερη ομαλή επιφάνεια στην εικόνα βάθους, κάτι που είναι αληθές στις περισσότερες περιπτώσεις. Με βάση την παρατήρηση αυτή, υπολογίζουμε την κλίση G της εικόνας I . Η κλίση υπολογίζεται από τη σχέση

$$G = \nabla I = \sqrt{I_x^2 + I_y^2} = \sqrt{(S_x * I)^2 + (S_y * I)^2} \quad (\text{A.1})$$

, όπου S_x, S_y οι πυρήνες συνέλιξης για την παραγωγή των μερικών παραγώγων ως προς τις διαστάσεις x και y αντίστοιχα. Χρησιμοποιήσαμε τους απλούς πυρήνες

$$S_x = \begin{bmatrix} -1 & +1 \end{bmatrix}, \quad S_y = \begin{bmatrix} -1 \\ +1 \end{bmatrix} \quad (\text{A.2})$$

Εφαρμόζουμε τιμή κατωφλίου στην κλίση:

$$M_1 = G < t_1 \quad (\text{A.3})$$

Η M_1 έχει τιμή 1 στις ομαλές περιοχές της εικόνας, και τιμή 0 στις ακμές και στις γωνίες. Εφαρμόζουμε το μετασχηματισμό απόστασης στην M_1 :

$$D = \text{dist}(M_1) \quad (\text{A.4})$$

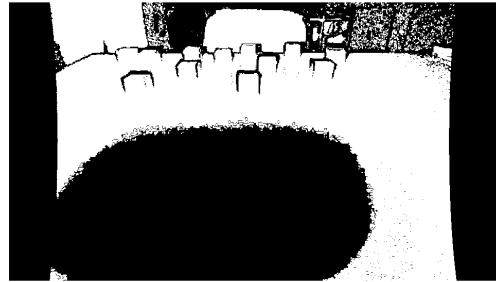
Με κατωφλίωση στην D , παράγουμε μια δυαδική εικόνα, με τιμή 1 σε περιοχές που απέχουν αρκετά από οποιαδήποτε ακμή, και 0 στις περιοχές κοντά στις ακμές:

$$M_2 = D < t_2 \quad (\text{A.5})$$

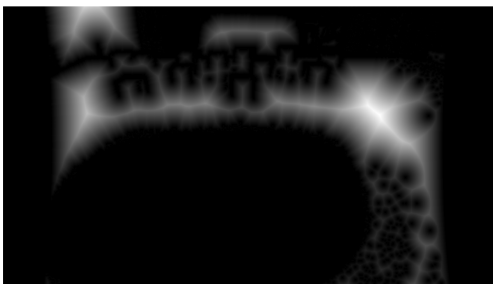
Στη συνέχεια, βρίσκουμε τα contours της M_2 . Επιλέγουμε το contour με το μεγαλύτερο εμβαδό εσώκλειστης επιφάνειας, διότι κατά πάσα πιθανότητα αυτό θα περιέχει σημεία που ανήκουν στην επιφάνεια του τραπεζιού.



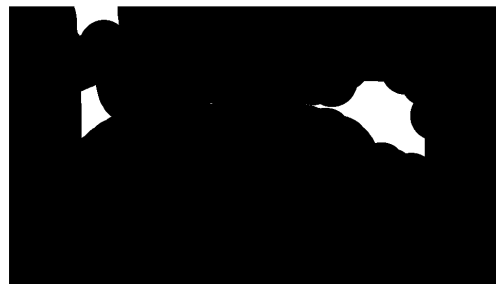
(a) Εικόνα βάθους I .



(b) Μάσκα M_1 .



(c) Μετασχηματισμός απόστασης D .



(d) Μάσκα M_2 .

Σχήμα A.1: Βήματα ανίχνευσης τραπεζιού.

Η γενική μορφή της εξίσωσης ενός επιπέδου είναι

$$a'x + b'y + c'z + d' = 0 \quad (\text{A.6})$$

Διαιρώντας δια b' (για $b' \neq 0$), λαμβάνουμε μια εξίσωση τριών παραμέτρων:

$$y = ax + bz + c \quad (\text{A.7})$$

, όπου $a = -\frac{a'}{b'}$, $b = -\frac{c'}{b'}$ και $c = -\frac{d'}{b'}$. Σημειώνουμε ότι αν $b' = 0$ τότε η εξίσωση (A.7) δεν μπορεί να χρησιμοποιηθεί, αλλά η περίπτωση αυτή αντιστοιχεί σε κάθετο προσανατολισμό της κάμερας, που δε χρησιμοποιήθηκε ποτέ στα πειράματά μας.

Από το contour που προσδιορίστηκε προηγουμένως, διαθέτουμε ένα σύνολο σημείων $\{\mathbf{p}_i(x_i, y_i, z_i)\}_{i=1}^N$ που ανήκουν στην επιφάνεια του τραπεζιού. Από τα σημεία αυτά εκτιμούμε τις παραμέτρους (a, b, c) της εξίσωσης (A.7), επιλύοντας το σύστημα:

$$\begin{bmatrix} x_1 & z_1 & 1 \\ x_2 & z_2 & 1 \\ \vdots & \vdots & \vdots \\ x_N & z_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad (\text{A.8})$$

Το παραπάνω σύστημα είναι υπερορισμένο, οπότε λύνουμε το πρόβλημα ελαχίστων τετραγώνων με decomposition QR.

Έχοντας υπολογίσει τους συντελεστές της εξίσωσης (A.7) του επιπέδου, μπορούμε να κατασκευάσουμε το κάθετο διάνυσμα $\mathbf{N}_T = \frac{(a, -1, b)}{\|(a, -1, b)\|}$, που δείχνει προς τα πάνω. Σε περίπτωση που χρειάζεται κάποιο σημείο σε μηδενικό ύψος, μπορούμε να πάρουμε το σημείο $(0, c, 0)$.

Παράρτημα Β

Τετραδόνια

Τα τετραδόνια είναι μια επέκταση των μιγαδικών αριθμών σε τέσσερις διαστάσεις. Ορίζονται ως

$$\bar{q} = q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k} + q_4 \quad (\text{B.1})$$

όπου τα \mathbf{i} , \mathbf{j} και \mathbf{k} είναι υπερφανταστικοί αριθμοί που ικανοποιούν τις σχέσεις

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1 \quad (\text{B.2})$$

Από τις σχέσεις αυτές μπορεί ναδειχθεί ότι

$$-\mathbf{ij} = \mathbf{ji} = \mathbf{k}, \quad -\mathbf{jk} = \mathbf{kj} = \mathbf{i}, \quad -\mathbf{ki} = \mathbf{ik} = \mathbf{j} \quad (\text{B.3})$$

Ο πολλαπλασιασμός δύο τετραδονίων ορίζεται ως

$$\begin{aligned} \bar{q} \otimes \bar{p} &= (q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k} + q_4)(p_1\mathbf{i} + p_2\mathbf{j} + p_3\mathbf{k} + p_4) \\ &= (q_1p_4 - q_2p_3 + q_3p_2 + q_4p_3)\mathbf{i} + (q_1p_3 + q_2p_4 - q_3p_1 + q_4p_2)\mathbf{j} \\ &\quad + (-q_1p_2 + q_2p_1 + q_3p_4 + q_4p_3)\mathbf{k} + (-q_1p_1 - q_2p_2 - q_3p_3 + q_4p_4) \end{aligned} \quad (\text{B.4})$$
$$= \begin{bmatrix} q_1p_4 - q_2p_3 + q_3p_2 + q_4p_3 \\ q_1p_3 + q_2p_4 - q_3p_1 + q_4p_2 \\ -q_1p_2 + q_2p_1 + q_3p_4 + q_4p_3 \\ -q_1p_1 - q_2p_2 - q_3p_3 + q_4p_4 \end{bmatrix}$$

Όπως φαίνεται, η σχέση αυτή δεν είναι αντιμεταθετική, δηλαδή $\bar{q} \otimes \bar{p} \neq \bar{p} \otimes \bar{q}$.

Ο συζυγής ενός τετραδονίου ισούται με

$$\bar{q}^* = -q_1\mathbf{i} - q_2\mathbf{j} - q_3\mathbf{k} + q_4 \quad (\text{B.5})$$

Η νόρμα ενός τετραδονίου ορίζεται

$$\|\bar{q}\| = \sqrt{\bar{q}\bar{q}^*} = \sqrt{q^*\bar{q}} = \sqrt{q_1^2 - q_2^2 + q_3^2 + q_4^2} \quad (\text{B.6})$$

Τα τετραδόνια χρησιμεύουν στην αναπαράσταση στροφών. Εάν κατασκευάσουμε το

διάνυσμα

$$\vec{u} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} k_x \sin(\theta/2) \\ k_y \sin(\theta/2) \\ k_z \sin(\theta/2) \end{bmatrix} = \hat{\mathbf{k}} \sin(\theta/2) \quad (\text{B.7})$$

τότε μπορούμε να θεωρήσουμε ότι το \bar{q} αντιπροσωπεύει μια στροφή γύρω από τον άξονα \vec{u} , κατά γωνία θ . Για να είναι μοναδιαίο το τετραδόνιο \bar{q} (δηλαδή $\|\bar{q}\| = 1$), παίρνουμε

$$q_4 = \cos(\theta/2) \Rightarrow \theta = 2 \arccos(q_4) \quad (\text{B.8})$$

και

$$|\hat{\mathbf{k}}| = 1 \quad (\text{B.9})$$

Συνεπώς, το διανυσματικό μέρος του τετραδονίου αντιστοιχεί στον άξονα στροφής και το βαθμωτό μέρος στο συνημίτονο της μισής γωνίας στροφής. Για κάθε τετραδόνιο στροφής \bar{q} το αντίθετό του $-\bar{q} = -q_1\mathbf{i} - q_2\mathbf{j} - q_3\mathbf{k} - q_4$ θα αντιπροσωπεύει την ίδια στροφή.

Από τις σχέσεις (B.7), (B.8) βλέπουμε ότι ο συζυγής \bar{q}^* δηλώνει την αντίθετη στροφή του \bar{q} , δηλαδή στροφή γωνίας $-\theta$ γύρω από τον ίδιο άξονα \vec{u} .

Αν \bar{q} και \bar{p} δυο στροφές, τότε το γινόμενό τους $\bar{q} \otimes \bar{p}$ (σχέση (B.4)) δίνει την τελική στροφή που προκύπτει εκτελώντας πρώτα τη στροφή \bar{p} και μετά την \bar{q} .

Η αναπαράσταση στροφών με τετραδόνια έχει κάποια πλεονεκτήματα έναντι άλλων τρόπων. Σε αντίθεση με τα ζεύγη άξονα-γωνίας Euler, δεν παρουσιάζουν ιδιομορφίες. Σε σχέση με τους πίνακες στροφής, είναι πιο συμπαγή, και άρα έχουν μικρότερο βαθμό πλεονάζουσας πληροφορίας, αφού χαρακτηρίζονται από 4 μεταβλητές και όχι 9.

Η μετατροπή από τετραδόνιο σε πίνακα στροφής γίνεται με τη σχέση

$$R = \begin{bmatrix} 1 - 2q_2^2 - 2q_3^2 & 2q_1q_2 - 2q_3q_4 & 2q_1q_3 + 2q_2q_4 \\ 2q_1q_2 + 2q_3q_4 & 1 - 2q_1^2 - 2q_3^2 & 2q_2q_3 - 2q_1q_4 \\ 2q_1q_3 - 2q_2q_4 & 2q_2q_3 + 2q_1q_4 & 1 - 2q_1^2 - 2q_2^2 \end{bmatrix} \quad (\text{B.10})$$

Ένα διάνυσμα Euler θ μετατρέπεται σε τετραδόνιο με βάση τη σχέση

$$\bar{q} = \frac{e_x \sin(\frac{\theta}{2})}{\sqrt{e_x^2 + e_y^2 + e_z^2}} \mathbf{i} + \frac{e_y \sin(\frac{\theta}{2})}{\sqrt{e_x^2 + e_y^2 + e_z^2}} \mathbf{j} + \frac{e_z \sin(\frac{\theta}{2})}{\sqrt{e_x^2 + e_y^2 + e_z^2}} \mathbf{k} + \cos(\frac{\theta}{2}) \quad (\text{B.11})$$

, όπου $\theta = \|\theta\|$ και $e = \frac{\theta}{\|\theta\|}$

Η αντίστροφη διαδικασία δίνεται

$$\theta = \left(\frac{2q_1 \arccos(q_4)}{\sqrt{1 - q_4^2}}, \frac{2q_2 \arccos(q_4)}{\sqrt{1 - q_4^2}}, \frac{2q_3 \arccos(q_4)}{\sqrt{1 - q_4^2}} \right) \quad (\text{B.12})$$