



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Εκπαίδευση οπτικά καθοδηγούμενης ρομποτικής λαβής μέσω
πρωτογενών χαρακτηριστικών σχήματος: Εφαρμογή στο
ανθρωπόμορφο ρομπότ ΝΑΟ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Ι. Μαυράκης

Επιβλέπων : Κωνσταντίνος Σ. Τζαφέστας
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2014



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

**Εκπαίδευση οπτικά καθοδηγούμενης ρομποτικής λαβής μέσω
πρωτογενών χαρακτηριστικών σχήματος: Εφαρμογή στο
ανθρωπόμορφο ρομπότ ΝΑΟ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Ι. Μαυράκης

Επιβλέπων : Κωνσταντίνος Σ. Τζαφέστας

Επίκουρος καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12^η Σεπτεμβρίου 2014.

.....
Κωνσταντίνος Τζαφέστας
Επίκουρος καθηγητής Ε.Μ.Π.

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

.....
Νικόλαος Μαράτος
Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2014

.....
Νικόλαος Ι. Μαυράκης

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

Copyright © Νικόλαος Ι. Μαυράκης, 2014.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Κύριος σκοπός της επιστήμης της Ρομποτικής είναι η παροχή βοήθειας στον άνθρωπο από ειδικά σχεδιασμένες για τον σκοπό αυτό, μηχανές. Ένα ρομπότ πρέπει να είναι σχεδιασμένο έτσι ώστε να αλληλεπιδρά με το περιβάλλον του, να χειρίζεται αντικείμενα με επιδεξιότητα και να μπορεί να τα χρησιμοποιήσει για την επίτευξη της εκάστοτε εργασίας. Το βασικό πεδίο μελέτης της παρούσας διπλωματικής εργασίας είναι ο σχεδιασμός αλγορίθμου εκπαίδευσης ρομποτικού συστήματος με χρήση οπτικών χαρακτηριστικών, με σκοπό την εύρεση κατάλληλων σημείων επάνω σε ένα καθημερινό αντικείμενο με βάση τα οποία το ρομπότ μπορεί να εξάγει τις τελικές επιτρεπτές λαβές για την χρήση του. Η εκπαίδευση είναι επιβλεπόμενη, δηλαδή παρέχοντας στο ρομπότ κάποια πρωταρχικά σχήματα με γνωστά επιλεγμένα σημεία λαβής, αυτό αναπτύσσει την ικανότητα γενίκευσης για άγνωστα, πιο πολύπλοκα σχήματα.

Η διπλωματική εργασία αποτελείται από την θεωρητική παρουσίαση της μεθόδου και την πειραματική εκτέλεσή της. Κατά την θεωρητική παρουσίαση αναφέρεται αναλυτικά η μεθοδολογία που χρησιμοποιήθηκε για κάθε φάση στην υλοποίηση του συστήματος. Αυτές επιγραμματικά είναι η βαθμονόμηση κάμερας, η εξαγωγή οπτικού βάθους με τεχνικές τρισδιάστατης όρασης, η εξαγωγή του διανύσματος εκπαίδευσης με γεωμετρικές τεχνικές, η περιγραφή του νευρωνικού δικτύου που χρησιμοποιήθηκε για την εκπαίδευση και η επίλυση του ανάστροφου κινηματικού προβλήματος του δεξιού βραχίονα του ρομπότ. Ιδιαίτερη έμφαση δίνεται στην εξαγωγή του περιγραφέα για την εκπαίδευση και στα αποτελέσματα που επιτυγχάνει. Το πειραματικό μέρος περιλαμβάνει την εφαρμογή του συστήματος αναγνώρισης στο ανθρωποειδές ρομπότ ΝΑΟ. Με χρήση μικρών, καθημερινών αντικειμένων δοκιμάζεται η ευστοχία και η αποτελεσματικότητα του αλγορίθμου, ενώ δείχνεται ένας βασικός τρόπος εφαρμογής του. Τέλος, παρουσιάζονται συμπεράσματα ενώ υποδεικνύονται πιθανές βελτιώσεις και μελλοντικές επεκτάσεις του αλγορίθμου.

Λέξεις κλειδιά: Βαθμονόμηση κάμερας, τρισδιάστατη όραση, διάνυσμα εκπαίδευσης, τοπικός r -θ περιγραφέας, PCA, ρομποτικές λαβές ,Μηχανή διανυσμάτων υποστήριξης (SVM), ανάστροφο κινηματικό πρόβλημα, ΝΑΟ.

Πίνακας περιεχομένων

1. Εισαγωγή	12
1.1 Ρομποτική	12
1.2 Αλληλεπίδραση ρομπότ με το περιβάλλον	14
1.3 Το γενικό πρόβλημα της ρομποτικής λαβής	15
1.4 Εκπαίδευση ρομπότ	16
1.5 Σύνοψη διπλωματικής εργασίας	18
2. Περιγραφείς αναγνώρισης εικόνας και σχήματος	21
2.1 Εισαγωγή	21
2.2 Προεπεξεργασία έγχρωμης εικόνας	22
2.3 Περιγραφείς περιοχών	24
2.3.1 Ορισμένοι απλοί περιγραφείς περιοχών	25
2.3.2 Ροπές εικόνας	26
2.4 Περιγραφείς περιγράμματος	28
2.4.1 Ορισμένοι απλοί περιγραφείς περιγράμματος	28
2.4.2 Περιγραφείς Fourier	29
2.5 Ακτινικοί περιγραφείς	32
2.5.1 Ανάλυση Κύριων Συνιστωσών	32
2.5.2 Ολικοί ακτινικοί περιγραφείς	35
2.5.3 Τοπικοί ακτινικοί περιγραφείς	37
2.6 Παραδείγματα	40
3. Βαθμονόμηση κάμερας και εύρεση τρισδιάστατων σημείων	42
3.1 Εισαγωγή	42

3.2	Μοντέλο κάμερας οπής	43
3.3	Βαθμονόμηση κάμερας	46
3.3.1	Ομογράφημα	47
3.3.2	Αλγόριθμος βαθμονόμησης	48
3.4	Εκτίμηση πόζας και εξαγωγή 3Δ σημείων	51
3.4.1	Εκτίμηση πόζας της κάμερας σχετικά με το επίπεδο	51
3.4.2	Εξαγωγή συντεταγμένων 3Δ σημείου στο επίπεδο	53
3.4.3	Εφαρμογή	54
4.	Νευρωνικά δίκτυα και μηχανές διανυσμάτων υποστήριξης (SVMs)	56
4.1	Εισαγωγή	56
4.2	Το πρόβλημα της ταξινόμησης σε κλάσεις	58
4.3	Βέλτιστη ταξινόμηση με SVMs	61
4.3.1	Διανύσματα υποστήριξης και βέλτιστο υπερεπίπεδο	61
4.3.2	Εύρεση του βέλτιστου υπερεπιπέδου	64
4.4	Μη γραμμικές μηχανές διανυσμάτων υποστήριξης	67
4.4.1	Γραμμική διαχωρισιμότητα προτύπων	67
4.4.2	Ταξινόμηση μη γραμμικά διαχωρίσιμων προτύπων	68
5.	Διαδικασία εκπαίδευσης ρομπότ και πειραματικά αποτελέσματα	73
5.1	Εισαγωγή	73
5.2	Διάνυσμα εκπαίδευσης	74
5.3	Διαδικασία εκπαίδευσης και αναγνώρισης βασικών σχημάτων	76
5.3.1	Διαδικασία εκπαίδευσης	76
5.3.2	Διαδικασία αναγνώρισης βασικών σχημάτων	78
5.4	Επαλήθευση αποτελεσμάτων	79

5.5	Πειραματικές εκτελέσεις	80
6.	Εντοπισμός και σχηματισμός λαβής	94
6.1	Εισαγωγή	94
6.2	Ρομπότ NAO	94
6.3	Ορθή κινηματική ανάλυση	97
6.3.1	Μέθοδος Denavit – Hartenberg	98
6.3.2	Τοποθέτηση πλαισίων και πίνακας DH για το ρομπότ NAO	100
6.4	Αντίστροφη κινηματική και διαφορική ανάλυση	102
6.5	Συνολική διαδικασία εντοπισμού και σχηματισμού λαβών	106
6.6	Πειραματική εκτέλεση για διάφορα αντικείμενα	112
7.	Συμπεράσματα και μελλοντικές επεκτάσεις της εργασίας	115
7.1	Συμπεράσματα	115
7.2	Βελτιώσεις και μελλοντικές επεκτάσεις	117

Κατάλογος Εικόνων

Εικόνα 1: Παραδείγματα ρομποτικών αρπαγών	16
Εικόνα 2: Παράδειγμα προτύπου	21
Εικόνα 3: Έγχρωμη εικόνα και εικόνα σε κλίμακα του γκρι	22
Εικόνα 4: Εικόνα φόντου (αριστερά), φόντου με αντικείμενο (μεσαία) και εικόνα αντικειμένου μετά από κατάτμηση και κατωφλίωση (δεξιά).....	24
Εικόνα 5: Εικόνες με αριθμό Euler 1, -1 και 0 αντίστοιχα	26
Εικόνα 6: Ανακατασκευή αρχικού περιγράμματος (επάνω δεξιά) χρησιμοποιώντας λιγότερους περιγραφείς Fourier	31
Εικόνα 7: PCA για δισδιάστατα σημεία. Από τους άξονες x_1, x_2 τα δεδομένα προβάλλονται στην νέα βάση z_1, z_2	33
Εικόνα 8: Πρωτεύοντες άξονες και ολικός ακτινικός περιγραφέας $r-\theta$ για τον κύκλο.....	40
Εικόνα 9: Πρωτεύοντες άξονες και ολικός ακτινικός περιγραφέας $r-\theta$ για το τετράγωνο ...	41
Εικόνα 10: Οπτική απεικόνιση τοπικού ακτινικού περιγραφέα.....	41
Εικόνα 11.1 : Μοντέλο κάμερας οπής (προοπτική προβολή).....	43
Εικόνα 12: Διαδοχικές λήψεις αντικειμένου βαθμονόμησης από διαφορετικές οπτικές γωνίες	55
Εικόνα 13: Μοντέλο νευρώνα McCulloch - Pitts	57
Εικόνα 14: Συναρτήσεις ενεργοποίησης για το μοντέλο νευρώνα McCulloch - Pitts.....	58
Εικόνα 15: Ταξινόμηση ψαριών.....	59
Εικόνα 16: Διανύσματα υποστήριξης, βέλτιστο υπερεπίπεδο και μέγιστο περιθώριο διαχωρισμού	64
Εικόνα 17: Γραμμικά (αριστερά) και μη γραμμικά (δεξιά) διαχωρίσιμα σύνολα	68
Εικόνα 18: Προβολή των διανυσμάτων εισόδου σε χώρο ανώτερης διαστατικότητας.....	69
Εικόνα 19: Απεικόνιση των διανυσμάτων εισόδου στον χώρο ανώτερης τάξης με τη συνάρτηση ϕ	71
Εικόνα 21: Περιγράμματα αντικειμένων προς εκπαίδευση και αναγνώριση	74
Εικόνα 22: Τιμές αναφοράς για τα σχήματα εκπαίδευσης και επαλήθευσης (με πράσινο τα σημεία με τιμή 1 και με κόκκινο τα σημεία με τιμή 0)	78
Εικόνα 23: Ρομπότ NAO.....	95
Εικόνα 24: Βαθμοί ελευθερίας του δεξιού χεριού του ρομπότ NAO	97
Εικόνα 25: Τοποθέτηση πλαισίων σε ρομποτικό βραχίονα	98

Εικόνα 26: Ιδιόμορφο σχήμα 116

ΚΕΦΑΛΑΙΟ 1

1 Εισαγωγή

1.1 Ρομποτική

Η ρομποτική είναι η επιστήμη που μελετά και σχεδιάζει ρομπότ, δηλαδή τις μηχανές που είναι σε θέση να αλληλεπιδρούν με το περιβάλλον, να επεξεργάζονται πληροφορίες, να μαθαίνουν και να εκτελούν εργασίες αντί του ανθρώπου. Ένα ρομπότ είναι μία μηχανητρονική διάταξη αποτελούμενη από επενεργητές, αισθητήρες, μικροεπεξεργαστές και ηλεκτρικά κυκλώματα η οποία είναι ειδικά σχεδιασμένη για να εκτελεί κάποια συγκεκριμένη εργασία. Τα ρομπότ λειτουργούν αυτόνομα και εκτελούν εργασίες οι οποίες είναι είτε δύσκολες είτε επικίνδυνες για τον άνθρωπο.

Ιστορικά, οι πρώτες αναφορές σε *αυτόματα*, δηλαδή μηχανές που λειτουργούσαν αυτόνομα και εκτελούσαν εργασίες έγιναν στην αρχαία Κίνα. Έκτοτε, υπάρχουν πλήθος αναφορών σε αυτόματα στην αρχαία Ελλάδα, κατά το Μεσαίωνα, κατά την Αναγέννηση και σε άλλες ιστορικές περιόδους. Αποκορύφωμα αποτελεί ο 20ος αιώνας, με την τεράστια τεχνολογική άνθηση και την εφεύρεση των ηλεκτρονικών υπολογιστών. Η ρομποτική με την μορφή που την γνωρίζουμε σήμερα υπάρχει από την δεκαετία του 1970.

Υπάρχουν πολλές κατηγορίες ρομπότ, ανάλογα με την μηχανική τους διάταξη, το περιβάλλον στο οποίο χρησιμοποιούνται και την εργασία που εκτελούν. Ορισμένες είναι τα βιομηχανικά ρομπότ, τα ανθρωπόμορφα ρομπότ, τα αυτόνομα οχήματα, τα χειρουργικά ρομπότ, τα νανορομπότ, τα διαστημικά ρομπότ, τα ρομποτικά υποβρύχια και τα ρομποτικά ζώα.

Το πλήθος των ρομποτικών εφαρμογών είναι τεράστιο. Στις μέρες μας ρομπότ χρησιμοποιούνται στην βιομηχανία ως κατασκευαστές, στην εξερεύνηση άγονων για τον άνθρωπο περιοχών, στην βοήθεια σε άτομα με ειδικές ανάγκες, στην ιατρική, στην

εξερεύνηση του διαστήματος, σε πολεμικές συρράξεις, ως κατοικίδια ή μέσα ψυχαγωγίας και σε πολλές άλλες περιπτώσεις.

Η επιστήμη της ρομποτικής έχει κάνει την εμφάνισή της σε διαφορετικές μορφές τέχνης. Στην λογοτεχνία, ξεχωρίζουν τα μυθιστορήματα των Isaac Asimov και Arthur Clarke, στα οποία τα ρομπότ παίζουν πρωταγωνιστικό ρόλο. Ανάλογες εμφανίσεις υπάρχουν στον κινηματογράφο, με πλήθος ταινιών επιστημονικής φαντασίας να έχουν ως πρωταγωνιστή ή ως αντικείμενο τα ρομπότ και τις εφαρμογές των.

Ένα ρομποτικό σύστημα αποτελείται από τα παρακάτω βασικά μέρη:

- *Αισθητήρες*, για την συλλογή δεδομένων από το περιβάλλον του. Οι αισθητήρες μπορούν να ανιχνεύσουν αλλαγές στο περιβάλλον του ρομπότ (π.χ. απόσταση από εμπόδιο, θερμοκρασία, πίεση που ασκείται στο ρομπότ) και να μετατρέψουν την αλλαγή αυτή σε ηλεκτρικό σήμα για τροφοδότηση σε άλλα μέρη του ρομπότ.
- *Ενεργοποιητές*, για την κίνηση των μηχανικών μερών του ρομπότ. Οι ενεργοποιητές δέχονται ηλεκτρικά σήματα και παράγουν μηχανικό έργο με τη μορφή κίνησης ή περιστροφής.
- *Μηχανικά μέρη*, για την κίνηση, την λαβή και τις λοιπές λειτουργίες που αποτελούν την αλληλεπίδραση του ρομπότ με το περιβάλλον του. Αυτά μπορεί να είναι τροχοί, κινηματικές αλυσίδες (βραχίονες), συστήματα προώθησης κ.α.
- *Σύστημα ελέγχου*, για τον συντονισμό των ηλεκτρικών σημάτων. Το σύστημα ελέγχου στην πλειονότητα των περιπτώσεων αποτελείται από μικροεπεξεργαστή και προγραμματίζεται ανάλογα με τη λειτουργία που πρέπει να εκτελέσει το ρομπότ. Υπάρχουν κλάδοι της ρομποτικής στους οποίους το σύστημα ελέγχου δεν έχει μικροεπεξεργαστή και αποτελείται από απλές καλωδιακές συνδέσεις, διακόπτες και ηλεκτρονικά στοιχεία.

Γίνεται εύκολα αντιληπτό πως η επιστήμη της ρομποτικής περιλαμβάνει έννοιες από πολλές επιστήμες, όπως η μηχανική, η ηλεκτρονική, η πληροφορική, τα μαθηματικά, ο αυτόματος έλεγχος και πολλές άλλες.

1.2 Αλληλεπίδραση ρομπότ με το περιβάλλον

Όπως έχει αναφερθεί, ένα ρομπότ είναι ικανό να αλληλεπιδρά με το περιβάλλον του, μέσω των αισθητήρων του, των επενεργητών του και των μηχανικών μερών του. Η αλληλεπίδραση του ρομπότ με τον περίγυρό του αποτελείται από την συλλογή δεδομένων από το περιβάλλον που θα χρησιμοποιηθούν για την εκτέλεση της ζητούμενης εργασίας και από την κίνηση των μηχανικών μερών του, η οποία θα προκαλέσει μία αλλαγή στην κατάσταση του ρομπότ ή στο περιβάλλον του. Οι δύο αυτές διαδικασίες μπορούν να εκτελεστούν τόσο σειριακά (off line), όσο και παράλληλα (on line).

Αναλυτικότερα, για να επιτελέσει ένα ρομπότ μία επιθυμητή λειτουργία αρχικά συλλέγει πληροφορίες από τους αισθητήρες εισόδου, δηλαδή τις κάμερες, τα μικρόφωνα, τους αισθητήρες αφής, δύναμης, θερμοκρασίας και άλλους. Οι αισθητήρες αυτοί έχουν την δυνατότητα να μετατρέπουν κάποιο φυσικό μέγεθος σε ηλεκτρικό σήμα, το οποίο στη συνέχεια διοχετεύεται στην μονάδα επεξεργασίας του ρομπότ ή και απ' ευθείας στους επενεργητές. Η επεξεργασία των ηλεκτρικών σημάτων γίνεται στον επεξεργαστή του ρομπότ ή σε τυχόν αναλογικά κυκλώματα που θα διαθέτει.

Αφού γίνει η επεξεργασία των σημάτων, ανάλογα με το αποτέλεσμα, το ρομπότ δίνει εντολή στους επενεργητές να κινήσουν ή να σταματήσουν τα μηχανικά μέρη, ώστε να προέλθει μία αλλαγή στην θέση, ταχύτητα ή στην γενικότερη κατάσταση του ρομπότ. Για παράδειγμα, η ανίχνευση ενός εμποδίου με αισθητήρα απόστασης, θα οδηγήσει στο φρενάρισμα των τροχών του ρομπότ. Επιπλέον, το ρομπότ μπορεί να προκαλέσει μεταβολή στο περιβάλλον του, όπως στην περίπτωση ενός ρομποτικού βραχίονα που θα μετακινήσει ένα αντικείμενο του χώρου του μόλις το ανιχνεύσει στην κάμερά του.

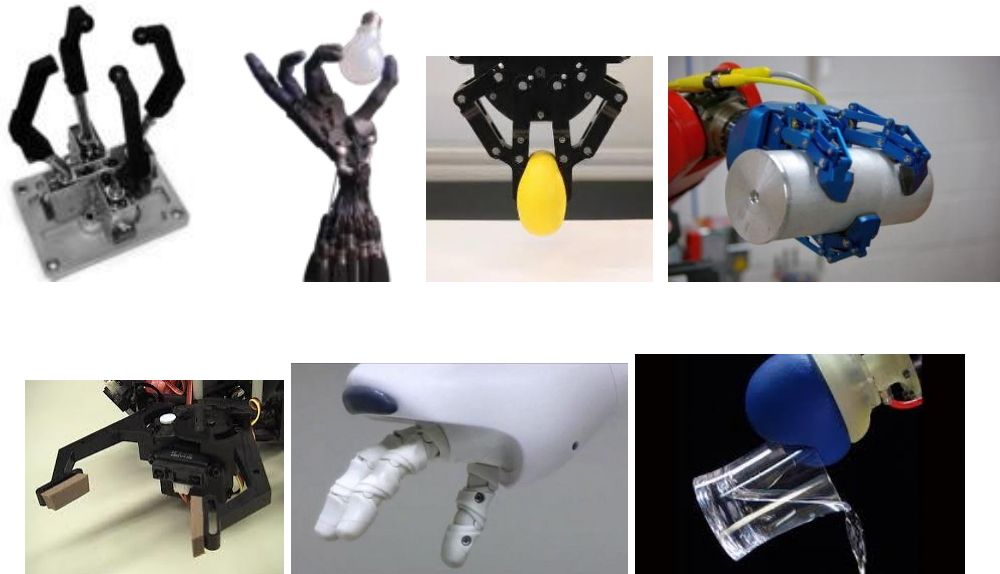
Η αλληλεπίδραση του ρομπότ με το περιβάλλον ξεκινά από απλά δεδομένα και ενέργειες και καταλήγει σε εκτέλεση πολυσύνθετων λειτουργιών. Με αυτή την ιδιότητα ένα ρομπότ μπορεί να εκτελέσει περισσότερες και πιο πολύπλοκες εργασίες.

1.3 Το γενικό πρόβλημα της ρομποτικής λαβής

Μία από τις βασικότερες μεθόδους αλληλεπίδρασης με το περιβάλλον είναι η αφή. Στον άνθρωπο επιτυγχάνεται με τη χρήση των παλαμών, οι οποίες περιέχουν πολλούς βαθμούς ελευθερίας και αντιστακτό αντίχειρα, επιτρέποντας έτσι πλήθος κινήσεων και μεγάλη επιδεξιότητα. Οι νευρικές απολήξεις στα δάχτυλα κάνουν ακόμα πιο εύκολη την αίσθηση της αφής, παρέχοντας μεγαλύτερη δυνατότητα αλληλεπίδρασης με το περιβάλλον. Ο άνθρωπος λοιπόν είναι σε θέση να χρησιμοποιήσει τα χέρια του τόσο για να αισθανθεί, όσο και για να πιάσει αντικείμενα. Έτσι μπορεί με μεγάλη άνεση να χειριστεί αντικείμενα, να τα μεταφέρει και να πιάσει τραχιές και πολύπλοκες επιφάνειες με άνεση.

Οι λαβές που μπορεί να εκτελέσει ο άνθρωπος, θα μπορούσαν να ομαδοποιηθούν σε αρκετές κατηγορίες. Ανάλογα με τη λειτουργία υπάρχουν οι *λαβές δύναμης* και τις *λαβές ακριβείας*. Άλλες κατηγορίες μπορούν να προκύψουν με κριτήριο τις δυνάμεις που ασκούν τα δάκτυλα, δηλαδή δυνάμεις μεταξύ των δακτύλων, δυνάμεις μεταξύ του αντίχειρα και της παλάμης και δυνάμεις μεταξύ του αντίχειρα και του δείκτη. Η ανθρώπινη λαβή σχηματίζεται ανάλογα με τη φυσιομορφία του αντικειμένου προς χρήση και τον την λειτουργία την οποία θα επιτελέσει.

Στην επιστήμη της ρομποτικής είναι μεγάλη η ανάγκη για χρήση λαβών που θα μιμούνται τις ανθρώπινες, θα μπορούν να μετακινούν και να χειρίζονται αντικείμενα και θα παρέχουν πλήθος λειτουργιών. Στην περίπτωση των βιομηχανικών, ανθρωποειδών, αλλά και σε πλήθος άλλων ρομποτικών εφαρμογών, οι λαβές εκτελούνται με *αρπάγες* που βρίσκονται προσαρτημένες στην άκρη των βραχιόνων. Οι αρπάγες έχουν διαφορετική μορφολογία και σχεδιασμό, ανάλογα με την εργασία που χρειάζεται να εκτελέσουν. Οι κατηγορίες ξεκινούν από αρπάγες 2 επαφών που ουσιαστικά ‘τσιμπούν’ το αντικείμενο μέχρι πολύπλοκα ρομποτικά χέρια με αρθρώσεις σε κάθε δάκτυλο που μπορούν να προσομοιώσουν σε εκπληκτικό βαθμό την ανθρώπινη λαβή. Ουσιαστικά αποτελούνται από πολυαρθρωτούς ρομποτικούς μηχανισμούς που λειτουργούν παράλληλα και επιτελούν απλές ή πιο ακριβείς λειτουργίες.



Εικόνα 1: Παραδείγματα ρομποτικών αρπαγών

Η ρομποτική λαβή είναι ουσιαστικά ένα σύνολο *επαφών* σε ένα αντικείμενο οι οποίες χαρακτηρίζονται από κάποιες ιδιότητες. Η βασικότερη ιδιότητα είναι η *ευστάθεια ως προς τη δύναμη (force closure)*, σύμφωνα με την οποία η λαβή είναι ικανή να αντισταθμίζει τυχόν εξωτερικές δυνάμεις στο αντικείμενο, με την άσκηση κατάλληλων δυνάμεων από τα δάκτυλα. Άλλη ιδιότητα είναι η επιδεξιότητα της λαβής. Μία λαβή θεωρείται επιδέξια εάν οι κινηματικές σχέσεις μεταξύ των αρθρώσεων του δαχτύλου και του αντικειμένου επιτρέπουν την κίνηση του με τρόπο συμβατό ως προς την ζητούμενη εργασία. Οι ιδιότητες αυτές μπορούν να εξακριβωθούν τόσο θεωρητικά, όσο και πειραματικά.

1.4 Εκπαίδευση ρομπότ

Για να είναι σε θέση να εκτελέσει μία ζητούμενη, προκαθορισμένη εργασία, το ρομπότ θα πρέπει να εκπαιδευτεί πάνω σε αυτήν έτσι ώστε να την φέρει εις πέρας στο μέλλον. Η εκπαίδευση μπορεί να γίνει είτε στον πραγματικό κόσμο, είτε σε κατάλληλο εικονικό περιβάλλον.

Στον πραγματικό κόσμο, η εκπαίδευση ενός ρομπότ γίνεται με μεθόδους επίδειξης από τον άνθρωπο, ή ακόμα και από ένα άλλο ρομπότ. Συνηθέστερος τρόπος εκπαίδευσης

είναι η καθοδήγηση του κινητού μέρους του ρομπότ άμεσα από τον άνθρωπο, ή μέσω τηλεχειρισμού. Άλλος τρόπος είναι η εκτέλεση της κίνησης ή της εργασίας από τον άνθρωπο, η καταγραφή των κινήσεων του ανθρώπου με αισθητήρες και η μίμηση αυτών των κινήσεων από το ρομπότ. Η 1^η περίπτωση έχει το πλεονέκτημα της λειτουργίας απ' ευθείας στον χώρο αρθρώσεων του ρομπότ, συνεπώς η τροχιά των αρθρώσεων είναι ευκολότερο να αποθηκευτεί και να επανεκτελεστεί. Αντίθετα, στη 2^η περίπτωση τα δεδομένα που θα συλλεχθούν από τους αισθητήρες και την κάμερα του ρομπότ, πρέπει να υποστούν επεξεργασία με κατάλληλους αλγορίθμους για να παραχθεί η επιθυμητή τροχιά αρθρώσεων. Παρ' όλα αυτά, ο άνθρωπος δεν είναι σε θέση να επιδεικνύει στο ρομπότ άμεσα μία εκτέλεση εργασίας, οπότε ο 2^{ος} τρόπος είναι πιο διαδομένος.

Οι δυσκολίες επίδειξης στον πραγματικό κόσμο (κόστος προμήθειας χειριστών, λάθη στον χειρισμό που μπορεί να οδηγήσουν σε καταστροφή υλικού κ.α.) οδήγησαν στην ανάπτυξη υπολογιστικών τεχνικών και λογισμικών προσομοίωσης. Σε αυτά τα λογισμικά, οι άνθρωποι μπορούν να εφαρμόζουν θεωρητικούς αλγορίθμους και τεχνικές επίλυσης των εργασιών αυτών με ασφάλεια και περισσότερη ακρίβεια.

Ένα ρομπότ πρέπει να είναι προετοιμασμένο να εκτελεί ένα μεγάλο πλήθος λειτουργιών, οι οποίες μπορεί να του έχουν επιδειχθεί ή και όχι. Επιπλέον, πολλές φορές ο χρόνος εκπαίδευσης του ρομπότ είναι περιορισμένος, με αποτέλεσμα να μην μπορέσει να διδαχθεί όλες τις απαιτούμενες κινήσεις. Οι λόγοι αυτοί έχουν οδηγήσει στην σύσταση μαθηματικών τεχνικών ή συναρτήσεων, στις οποίες τροφοδοτούνται ως είσοδοι τα δεδομένα εκπαίδευσης και λαμβάνεται μία έξοδος η επιθυμητή συμπεριφορά του ρομπότ. Οι μέθοδοι ποικίλλουν, ανάλογα με το αν η είσοδος έχει συνεχείς ή διακριτές τιμές, εάν το ρομπότ είναι σε θέση να επεξεργάζεται νέα δεδομένα κατά την διάρκεια της κίνησής του, εάν τα δεδομένα εισόδου ανανεώνονται και πολλούς άλλους παράγοντες. Οι δύο μεγάλες κατηγορίες μεθόδων εκπαίδευσης είναι οι *μέθοδοι ταξινόμησης* και οι *μέθοδοι παλινδρόμησης*.

Στις μεθόδους ταξινόμησης, το πρόβλημα έγκειται στην ταξινόμηση των δεδομένων εισόδου σε δύο ή περισσότερες κλάσεις. Ένα μεγάλο μέρος των δεδομένων χρησιμοποιείται για εκπαίδευση και προσδιορισμό των κλάσεων, ενώ ένα μικρότερο χρησιμοποιείται για επαλήθευση. Έπειτα από τον προσδιορισμό των κλάσεων, το ρομπότ είναι σε θέση να

ταξινομεί τα νέα δεδομένα εισόδου που θα λάβει σε κάποια από τις κλάσεις που δημιούργησε. Οι πιο γνωστές μέθοδοι ταξινόμησης είναι τα *Κρυφά Μαρκοβιανά Μοντέλα* (HMMs), ο αλγόριθμος *K-means*, τα *GMMs* (Gaussian Mixture Models), ο αλγόριθμος *EM* (Expectation Maximization) και τα *SVMs* (Support Vector Machines).

Στις μεθόδους παλινδρόμησης, τα δεδομένα προσεγγίζονται στον συνεχή χώρο μέσω επαναληπτικών διαδικασιών, χωρίς να επιχειρείται η ομαδοποίησή τους. Επειδή οι συνεχείς τιμές της εξόδου είναι αποτέλεσμα συνδυασμού διαφόρων δεδομένων εκπαίδευσης, η παλινδρόμηση εφαρμόζεται σε χαμηλού επιπέδου κινήσεις. Οι μέθοδοι παλινδρόμησης ποικίλλουν ανάλογα με το εάν η προσέγγιση της συνάρτησης αντιστοίχισης εκτελείται κατά την κίνηση του ρομπότ ή πριν από αυτή. Πιο γνωστές μέθοδοι παλινδρόμησης είναι η *Οκνηρή Μάθηση* (kNN), η *Τοπικά Σταθμισμένη Παλινδρόμηση* (LWR) και τα *Νευρωνικά Δίκτυα* (NN).

1.6 Σύνοψη διπλωματικής εργασίας

Για να μπορέσει ένα ρομπότ να αλληλεπιδρά με το περιβάλλον του και να χειρίζεται αντικείμενα με τρόπο ιδανικά όμοιο με αυτόν ενός ανθρώπου, πρέπει να σχηματίσει ένα σύνολο λαβών με το τελικό στοιχείο δράσης του. Επειδή η μίμηση των ανθρώπινων λαβών είναι ένα δύσκολο πρόβλημα, οι θεωρητικές και πειραματικές μέθοδοι που έχουν επιχειρηθεί για την λύση του προβλήματος συμπεριλαμβάνουν υποχρεωτικά αρκετούς περιορισμούς και παραδοχές.

Απαραίτητο στοιχείο για την εύρεση λαβών και την τελική αλληλεπίδραση με κάποιο αντικείμενο, είναι η γνώση της γεωμετρίας του αντικειμένου. Κάποιες μέθοδοι περιλαμβάνουν γνώση τρισδιάστατης πληροφορίας, κάποιες περιλαμβάνουν δισδιάστατες πληροφορίες, ενώ υπάρχουν και υβριδικές μέθοδοι, με γνώση και σύνθεση τόσο τρισδιάστατης όσο και δισδιάστατης πληροφορίας. Για το σκοπό αυτό μπορεί να χρησιμοποιηθεί ανακατασκευή του τρισδιάστατου μοντέλου του αντικειμένου, χρήση μοναδικών σημείων στην δισδιάστατη επιφάνειά του, ή ακόμα και γενικότερη δισδιάστατη πληροφορία.

Στην παρούσα διπλωματική εργασία εξετάζεται η περίπτωση της εύρεσης πρωτογενών λαβών για ένα σύνολο καθημερινών αντικειμένων που βρίσκονται τοποθετημένα σε οριζόντιο επίπεδο σχετικά με το ρομπότ. Το ρομπότ που χρησιμοποιείται είναι το ρομπότ NAO της Aldebaran Robotics. Το ρομπότ λαμβάνει οπτικά δεδομένα από την κάμερά του, τις οποίες χρησιμοποιεί για να εξάγει την τελική λαβή.

Εφ' όσον το ρομπότ δεν διαθέτει κινητήρες και ελεγκτές για κάθε άρθρωση των δακτύλων, ως τελική λαβή ορίζεται το σημείο επαφής του αντίχειρα με το αντικείμενο, η μεταφορά και η περιστροφή του καρπού του χεριού. Επιπλέον, για ευκολία χρησιμοποιείται μόνο το δεξί χέρι του ρομπότ στα πειράματα. Η διαδικασία μπορεί πολύ εύκολα να επεκταθεί έτσι ώστε να περιλαμβάνει και τα δύο χέρια.

Για την εύρεση του σημείου επαφής με τον αντίχειρα χρησιμοποιούνται δισδιάστατα δεδομένα περιγράμματος του αντικειμένου. Για αυτό το τελικό σημείο χρησιμοποιούνται μέθοδοι αναλυτικής γεωμετρίας και βρίσκεται η τρισδιάστατη θέση του στο επίπεδο. Έτσι το ρομπότ τοποθετεί τον αντίχειρά του σε αυτήν την τρισδιάστατη θέση.

Η εργασία περιέχει συνολικά επτά κεφάλαια:

- *Κεφάλαιο 1:* Εισαγωγικές έννοιες γύρω από την επιστήμη της ρομποτικής, θεμελιοποίηση και σύντομη περιγραφή των ρομποτικών λαβών, περιγραφή τρόπων εύρεσης ρομποτικών λαβών χρησιμοποιώντας στοιχεία μάθησης μηχανών και γεωμετρίας αντικειμένων και σύνοψη της διπλωματικής εργασίας.
- *Κεφάλαιο 2:* Περιγράφονται τα στοιχεία εκείνα των προς χειρισμό αντικειμένων που θα χρησιμοποιηθούν ως δεδομένα εκπαίδευσης του ρομπότ. Το κεφάλαιο περιλαμβάνει σύντομη αναφορά σε τρόπους εξαγωγής δισδιάστατων χαρακτηριστικών καθώς επίσης και σε τρόπους αναπαράστασης δισδιάστατου σχήματος. Αναλύονται στη συνέχεια οι περιγραφείς σχήματος που χρησιμοποιήθηκαν στην εργασία, ενώ παρουσιάζονται παραδείγματα εφαρμογής των.

- *Κεφάλαιο 3:* Το κεφάλαιο περιέχει γενικές παρουσίαση του μοντέλου κάμερας οπής το οποίο χρησιμοποιεί ως μοντελοποίηση το ρομπότ NAO, ενώ αναλύεται η μεθοδολογία και ο αλγόριθμος βαθμονόμησης της κάμεράς του. Στη συνέχεια επεξηγείται ο τρόπος εύρεσης των τρισδιάστατων συντεταγμένων του τελικού σημείου τοποθέτησης του αντίχειρα επάνω στο αντικείμενο.
- *Κεφάλαιο 4:* Στο κεφάλαιο αυτό παρουσιάζεται ένα βασικό πρόβλημα στην θεωρία μάθησης μηχανών, αυτό της ταξινόμησης ενός διανύσματος σε κλάσεις. Έπειτα περιγράφεται η μεθοδολογία βέλτιστης ταξινόμησης με χρήση *Μηχανών Διανυσμάτων Υποστήριξης (Support Vector Machines, SVMs)*, καθώς επίσης και μία ακριβέστερη εκδοχή τους, τα μη γραμμικά SVMs, ενώ παρουσιάζονται παραδείγματα εκτέλεσης.
- *Κεφάλαιο 5:* Επεξηγείται αναλυτικά η διαδικασία εύρεσης του διανύσματος εκπαίδευσης για το SVM και η διαδικασία εκπαίδευσης του ρομπότ. Στη συνέχεια, ορίζεται η το ποσοστό επιτυχίας και σχηματίζεται ο πίνακας που το δείχνει για κάθε αντικείμενο με τη βοήθεια παραδειγμάτων.
- *Κεφάλαιο 6:* Αναλύεται η διαδικασία σχηματισμού πρωτογενούς λαβής για τα δεδομένα αντικείμενα. Το κεφάλαιο περιέχει την ορθή και αντίστροφη κινηματική ανάλυση του βραχίονα του ρομπότ, ενώ έπειτα επεξηγείται ο γενικός τρόπος εύρεσης και σχηματισμού της λαβής. Το κεφάλαιο ολοκληρώνεται με την παρουσίαση πειραματικών αποτελεσμάτων για το σύνολο των αντικειμένων που χρησιμοποιούνται.
- *Κεφάλαιο 7:* Στο τελικό κεφάλαιο γίνεται σύντομη ανακεφαλαίωση της εργασίας, ενώ παράλληλα παρουσιάζονται συμπεράσματα και χαράσσονται μελλοντικές κατευθύνσεις που επεκτείνουν την παρούσα εργασία.

Στις τελευταίες σελίδες παρατίθεται η βιβλιογραφία που χρησιμοποιήθηκε για την σύλληψη, υλοποίηση και πειραματική εκτέλεση της εργασίας.

ΚΕΦΑΛΑΙΟ 2

2 Περιγραφείς αναγνώρισης εικόνας και σχήματος

2.1 Εισαγωγή

Σε πολλές εφαρμογές ρομποτικής, όρασης είναι απαραίτητη η εξαγωγή και η αναπαράσταση πληροφορίας της εικόνας που επεξεργάζεται το ρομπότ, για την διεξαγωγή της προκαθορισμένης λειτουργίας που πρέπει να εκτελέσει. Η πληροφορία αυτή λαμβάνει πολλές μορφές: μπορεί να σχετίζεται με το σχήμα του αντικειμένου, τη μορφολογία του περιγράμματός του, την υφή του, κ.ο.κ. . Η αναπαράσταση αυτής της πληροφορίας μπορεί επίσης να γίνει με πολλούς τρόπους οι οποίοι θα εξεταστούν στο παρόν κεφάλαιο. Όλες οι αναπαραστάσεις έχουν ένα κοινό: η τελική μορφή τους είναι ένα διάνυσμα από αριθμούς, οι οποίοι έχουν διαφορετικό νόημα ανάλογα με την επεξεργασία που έχει προηγηθεί στην εικόνα. Ένα τέτοιο διάνυσμα ονομάζεται *περιγραφέας*, ή γενικότερα *πρότυπο*.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

Εικόνα 2: Παράδειγμα προτύπου

Τα πρότυπα εισόδου, μπορούν να αναπαραριστούν οτιδήποτε είναι χρήσιμο για την αναπαράσταση του προς εξέταση αντικειμένου. Χρώμα, υφή, σχήμα, αλλά και περιγραφείς φωτομετρίας μπορούν να χρησιμοποιηθούν. Στην παρούσα διπλωματική εργασία, χρησιμοποιήθηκαν περιγραφείς σχήματος. Στο παρόν κεφάλαιο, γίνεται αναφορά και

ανάλυση σε δύο από τις διασημότερες και χρησιμότερες αναπαραστάσεις σχήματος, τους περιγραφείς περιοχών (*region descriptors*) και οι περιγραφείς περιγράμματος (*contour descriptors*). Οι περιγραφείς που προκύπτουν από οποιαδήποτε από τις δύο αναπαραστάσεις, χρησιμεύουν ως είσοδοι σε Νευρωνικά δίκτυα για την ρομποτική εκπαίδευση. Δίνεται ιδιαίτερη έμφαση στους ακτινικούς περιγραφείς, οι οποίοι χρησιμοποιήθηκαν στην παρούσα εργασία.

2.2 Προεπεξεργασία έγχρωμης εικόνας

Για την εξαγωγή των περιγραφών μίας έγχρωμης εικόνας, είναι συνηθέστερο να μετατρέπεται αρχικά σε εικόνα *grayscale* και στη συνέχεια σε *δυναδική εικόνα (binary image)*. Ο λόγος για αυτόν τον μετασχηματισμό οφείλεται στην μικρότερη πολυπλοκότητα που εμφανίζουν τέτοιου είδους εικόνες.



Εικόνα 3: Έγχρωμη εικόνα και εικόνα σε κλίμακα του γκρι

Η έγχρωμη εικόνα θεωρείται ως 3Δ πίνακας μεγέθους $M \times N \times 3$, όπου M ο αριθμός των εικονοστοιχείων στον οριζόντιο άξονα και N ο αριθμός των εικονοστοιχείων στον κάθετο άξονα. Κάθε εικονοστοιχείο αναπαρίσταται με μία τιμή στον RGB χρωματικό χώρο, εξ' ου και ο πολλαπλασιασμός με το 3. Το τελικό αποτέλεσμα είναι μία εικόνα στην κλίμακα του γκρι, μεγέθους $M \times N$ και τιμές εικονοστοιχείων που δίνονται από τον τύπο

$$Gr(x, y) = 0.2126R(x, y) + 0.7152G(x, y) + 0.0722B(x, y) \quad (2.1)$$

Το επόμενο βήμα είναι η εξαγωγή δυαδικής εικόνας από την grayscale εικόνα. Οι εικόνες αυτές μπορούν να προκύψουν από χρήση αλγορίθμων *κατάτμησης εικόνας*, με αποτέλεσμα τον διαχωρισμό των εκάστοτε αντικειμένων από το φόντο και γενικότερα των διαφορετικών περιοχών της εικόνας. Στη συνέχεια, τα χρήσιμα μέρη υφίστανται *κατωφλίωση*, δηλαδή τη διαδικασία κατά την οποία η τιμή των εικονοστοιχείων τους τίθεται ίση με 1, ενώ η τιμή των υπόλοιπων μερών της εικόνας τίθεται ίση με το 0.

Η κατάτμηση εικόνας είναι μία αρκετά δύσκολη εργασία που έχει ως αποτέλεσμα τον διαχωρισμό των διαφορετικών μερών της έγχρωμης εικόνας και την αποτύπωση τους πάνω σε αυτή. Επισημαίνονται κάποιοι βασικοί και διάσημοι τρόποι κατάτμησης έγχρωμης εικόνας.

- *Χρωματική κατάτμηση*: Σε αυτήν τη μέθοδο, επιλέγονται τα εικονοστοιχεία συγκεκριμένου χρώματος ή αποχρώσεων του, ενώ τα υπόλοιπα τίθενται ίσα με 0. Η μέθοδος προτιμάται εάν το προς αναγνώριση αντικείμενο έχει γνωστό και συγκεκριμένο χρώμα, διαφορετικό από το περιβάλλον του.
- *Κατάτμηση με αλγόριθμο watershed*: Η εικόνα εκλαμβάνεται σαν τρισδιάστατος τοπογραφικός χάρτης, με τα σκοτεινότερα στην κλίμακα του γκρι εικονοστοιχεία να θεωρούνται σημεία χαμηλότερου υψομέτρου. Στη συνέχεια, ο χάρτης 'πλημμυρίζει', δηλαδή τα εικονοστοιχεία τίθενται σε κάποια σταθερή τιμή από τα χαμηλότερα ύψη στα ψηλότερα. Έτσι, ο χάρτης χωρίζεται στις *λεκάνες* και στα *φράγματα*. Η διαδικασία συνεχίζεται μέχρι να υπάρξει σαφής διαχωρισμός των περιοχών της εικόνας, με τα φράγματα να αποτελούν τα όρια.
- *Κατάτμηση με αφαίρεση φόντου*: Αρχικά λαμβάνεται μία εικόνα ως εικόνα στατικού φόντου. Οι επόμενες εικόνες με ακίνητη κάμερα περιέχουν τα αντικείμενα, οπότε η αφαίρεση των δύο εικόνων θα δώσει θετικές τιμές μόνο στα εικονοστοιχεία των αντικειμένων. Στην περίπτωση αναγνώρισης σε στατικό

περιβάλλον, η μέθοδος δίνει αρκετά καλά αποτελέσματα. Η τελική εικόνα φιλτράρεται για να αφαιρεθεί τυχόν θόρυβος ή κηλίδες.

Στη συνέχεια, η κατατμημένη εικόνα υφίσταται κατωφλίωση. Σε αυτήν τη διαδικασία, τα εικονοστοιχεία που κρίνονται χρήσιμα ανάλογα με το αποτέλεσμα της κατάτμησης, τίθενται ίσα με 1, ενώ τα υπόλοιπα ίσα με 0. Η χρησιμότητα των εικονοστοιχείων μεταφράζεται στο εάν η τιμή τους ξεπερνά κάποιο κατώφλι. Το κατώφλι μπορεί να θεωρηθεί αυθαίρετα, ή να υπολογιστεί από το *ιστόγραμμα της εικόνας*. Σε απλούστερα προβλήματα, η διαδικασία της κατωφλίωσης μπορεί να εκτελεστεί απ' ευθείας, χωρίς να χρειαστεί κατάτμηση.

Σε κάθε περίπτωση, το αποτέλεσμα των παραπάνω διαδικασιών είναι μία ασπρόμαυρη εικόνα που περιέχει το δισδιάστατο σχήμα του προς εξέταση αντικειμένου. Αυτή η εικόνα επεξεργάζεται στη συνέχεια για να προκύψει πληροφορία σχετικά με το σχήμα.



Εικόνα 4: Εικόνα φόντου (αριστερά), φόντου με αντικείμενο (μεσαία) και εικόνα αντικειμένου μετά από κατάτμηση και κατωφλίωση (δεξιά)

2.3 Περιγραφείς περιοχών

Με τον όρο *περιοχή* μίας εικόνας, ορίζεται ένα σύνολο συνδεδεμένων εικονοστοιχείων, η οποία δεν περιέχει οπές και είναι συνεκτική. Η αναπαράσταση των περιοχών μίας δυαδικής εικόνας και η εξαγωγή περιγραφέων μπορεί να γίνει με πολλούς τρόπους. Οι πιο κλασσικοί εξετάζονται παρακάτω.

2.3.1 Ορισμένοι απλοί περιγραφείς περιοχών

Μία πρώτη προσέγγιση στην εξαγωγή περιγραφών μπορεί να γίνει με τρόπους αρίθμησης εικονοστοιχείων. Οι μέθοδοι αυτές, έχουν το πλεονέκτημα της χαμηλής πολυπλοκότητας, αλλά ταυτόχρονα δεν αποτελούν ισχυρή αναπαράσταση της εκάστοτε περιοχής.

Το *εμβαδόν* E μίας περιοχής εικόνας ορίζεται ως ο αριθμός των εικονοστοιχείων που εμπεριέχονται σε αυτή. Παράλληλα, η *περίμετρος* P ορίζεται ως το μήκος του περιγράμματός της. Οι δύο αυτοί πρωταρχικοί περιγραφείς χρησιμοποιούνται σπάνια ως ξεχωριστοί περιγραφείς. Ο λόγος είναι πως εμφανίζουν καλύτερη απόδοση σε περιπτώσεις που το μέγεθος της ζητούμενης περιοχής είναι αμετάβλητο. Για να ξεπεραστεί αυτό το εμπόδιο, χρησιμοποιείται ο *βαθμός συμπίκνωσης* C που ορίζεται ως $C = \frac{P^2}{E}$. Μία τροποποίηση αυτού του περιγραφέα, είναι ο *λόγος κυκλικότητας* R_c που ορίζεται σαν $R_c = \frac{4\pi E}{P^2}$, δηλαδή σαν το εμβαδόν μίας περιοχής προς το εμβαδόν ενός κύκλου με τιμή περιμέτρου ίση με αυτή της περιοχής. Ο βαθμός συμπίκνωσης και ο λόγος κυκλικότητας αποτελούν αδιάστατα μεγέθη και είναι ανεπηρέαστα από μεταβολές στην κλίμακα και τον προσανατολισμό.

Παρόμοιος περιγραφέας είναι η *ορθογωνιότητα* μίας περιοχής. Έχοντας ως δεδομένη μία περιοχή σε διαφορετικούς προσανατολισμούς k και το ορθογώνιο που την περιβάλλει, υπολογίζεται το εμβαδόν της περιοχής προς το εμβαδόν του περιβάλλοντος ορθογωνίου για κάθε διαφορετικό προσανατολισμό. Το μέγιστο που θα προκύψει, αποτελεί την ορθογωνιότητα της περιοχής.

Στην περίπτωση των δυαδικών εικόνων, εάν ορίσουμε σαν *συνδεδεμένη συνιστώσα* ένα πλήθος εικονοστοιχείων που συνδέονται με 4 ή 8 συνδεσιμότητα, και σαν *οπή* ένα αντίστοιχο πλήθος αντίθετης τιμής, μπορεί να εξαχθεί ο *αριθμός Euler* ως το πλήθος των συνδεδεμένων συνιστωσών, μείον το πλήθος των οπών.



Εικόνα 5: Εικόνες με αριθμό Euler 1, -1 και 0 αντίστοιχα

2.3.2 Ροπές εικόνας

Στη στατιστική και στη μηχανική χρησιμοποιούνται οι ροπές μίας συνάρτησης, προκειμένου να περιγράψουν μεγέθη όπως πυκνότητα πιθανότητας συνάρτησης και κατανομή μάζας αντικειμένου. Οι ροπές μπορούν να χρησιμοποιηθούν και στην περίπτωση των δισδιάστατων εικόνων για να περιγράψουν σχήματα.

Εφ' όσον η εικόνα θεωρείται σαν δισδιάστατη συνάρτηση $I(x, y)$, μπορούν να υπολογιστούν οι ροπές της ως μετρικές για το σχήμα της, ή το σχήμα μίας περιοχής της.

Η ροπή τάξης $(p + q)$ μίας δισδιάστατης συνεχούς συνάρτησης υπολογίζεται από τον παρακάτω τύπο

$$M_{pq} = \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy \quad (2.2)$$

Στην περίπτωση εικόνας $I(x, y)$, η σχέση διακριτοποιείται ως εξής

$$M_{pq} = \sum_x \sum_y x^p y^q I(x, y) \quad (2.3)$$

Οι παράμετροι p, q παίρνουν ακέραιες μη αρνητικές τιμές. Από αυτές τις ροπές μπορεί να προκύψει το *εμβαδόν* και οι *συντεταγμένες του κεντροειδούς* της εικόνας. Οι σχέσεις που δίνουν τα δύο αυτά μεγέθη είναι $A = M_{00}$ και $(x_c, y_c) = (\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}})$. Εάν οι εικόνες αποτελούνται από την δυαδική εικόνα ενός μόνο αντικειμένου, το κεντροειδές της εικόνας και το εμβαδόν αναπαριστούν το κεντρικό σημείο του αντικειμένου και την έκτασή του στον δισδιάστατο χώρο.

Οι *κεντρικές ροπές* ορίζονται ως

$$m_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q I(x, y) \quad (2.4)$$

και οι *κανονικοποιημένες κεντρικές ροπές* ορίζονται ως

$$\eta_{pq} = \frac{m_{pq}}{m_{00}^\gamma}, \quad \text{με } \gamma = \frac{p+p}{2} + 1 \quad (2.5)$$

Από τις κανονικοποιημένες κεντρικές ροπές μπορούν να οριστούν οι αναλλοίωτες ροπές Hu φ , οι οποίες χρησιμοποιούνται πολύ συχνά σε εφαρμογές αναγνώρισης σχημάτων και αντικειμένων. Κύριο χαρακτηριστικό τους είναι το αναλλοίωτο σε μετασχηματισμούς μεταφοράς, αλλαγής κλίμακας, κατοπτρισμού και περιστροφής μεταξύ των εικόνων.

$$\begin{aligned}
\varphi_1 &= (\eta_{20} + \eta_{02}) \\
\varphi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\varphi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
\varphi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
\varphi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
\varphi_6 &= (\eta_{20} + \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\varphi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \tag{2.6}$$

2.4 Περιγραφείς περιγράμματος

Μία περιοχή της εικόνας περιβάλλεται από το *περίγραμμα* της, δηλαδή το συνδεδεμένο σύνολο των εξωτερικών εικονοστοιχείων της περιοχής. Η πληροφορία που μπορεί να εξαχθεί από το περίγραμμα μίας εικόνας χρησιμοποιείται για την δημιουργία ισχυρών αναπαραστάσεων σχήματος. Στην παρούσα διπλωματική χρησιμοποιήθηκε ακτινικός περιγραφέας περιγράμματος ως αναπαράσταση των τοπικών μορφολογικών ιδιοτήτων του αντικειμένου. Η εξαγωγή του και η λειτουργία του αναλύονται σε αυτήν την ενότητα, ενώ παρουσιάζονται και κάποιοι πιο κλασικοί περιγραφείς περιγράμματος.

2.4.1 Ορισμένοι απλοί περιγραφείς περιγράμματος

Όπως και στην περίπτωση των περιγραφέων περιοχών, ο απλούστερος τρόπος περιγραφής ενός περιγράμματος είναι μέσω απλών υπολογισμών απευθείας στα

εικονοστοιχεία του. Το μήκος ενός περιγράμματος είναι ένας πολύ απλός περιγραφέας. Διατρέχοντας τα εικονοστοιχεία του περιγράμματος και θεωρώντας πως τα οριζόντια ή κάθετα εικονοστοιχεία έχουν μήκος 1 και τα διαγώνια μήκος $\sqrt{2}$, μπορεί να υπολογιστεί το συνολικό μήκος του περιγράμματος.

Η διάμετρος ενός περιγράμματος ορίζεται ως

$$Diam = \max_{i,j} \{D(p_i, p_j)\} \quad (2.7)$$

όπου p_i, p_j δύο σημεία της εικόνας με συντεταγμένες $(x_i, y_i), (x_j, y_j)$ αντίστοιχα και $D(p_i, p_j)$ η απόσταση μεταξύ των δύο σημείων. Ως μέτρο απόστασης μπορεί να χρησιμοποιηθεί η Ευκλείδεια απόσταση, η Manhattan απόσταση, η απόσταση σκακιέρας κ.α.

Η τιμή της διαμέτρου και ο προσανατολισμός του διανύσματος που ενώνει τα δύο ακραία σημεία της διαμέτρου, αποτελούν χρήσιμους περιγραφείς περιγράμματος. Το διάνυσμα αυτό λέγεται *κύριος άξονας* του περιγράμματος. Ο άξονας που περνά από το κεντροειδές της περιοχής που περικλείει το περίγραμμα και είναι κάθετος στον κύριο ονομάζεται *δευτερεύων άξονας*. Επιπλέον, ένας ακόμα βασικός περιγραφέας είναι η *εκκεντρότητα* του σχήματος και προκύπτει ως ο λόγος του μήκους του κύριου προς τον δευτερεύων άξονα.

2.4.2 Περιγραφείς Fourier

Οι περιγραφείς Fourier αποτελούν μία πολύ βασική αναπαράσταση περιγράμματος, λόγω του αναλλοίωτου σε μετασχηματισμούς. Έστω ένα περίγραμμα μίας περιοχής εικόνας, στο οποίο έχει γίνει δειγματοληψία με αποτέλεσμα να αποτελείται από K σημεία στο επίπεδο xy . Οι συντεταγμένες στο επίπεδο xy μπορούν να θεωρηθούν σαν το πραγματικό και το φανταστικό μέρος ενός μιγαδικού αριθμού $s(k) = x(k) + jy(k)$. Η

αναπαράσταση αυτή βοηθάει στην αντιμετώπιση του προβλήματος σε μία διάσταση, απλοποιώντας το.

Για την μιγαδική συνάρτηση $s(k)$ μπορεί να υπολογιστεί ο Διακριτός Μετασχηματισμός Fourier (DFT) της ως εξής

$$a(u) = \sum_{k=0}^{K-1} s(k) e^{-j \frac{2\pi uk}{K}} \quad (2.8)$$

Οι μιγαδικοί συντελεστές $a(u)$ είναι γνωστοί ως περιγραφείς Fourier του περιγράμματος. Ο αντίστροφος μετασχηματισμός Fourier που οδηγεί πίσω στην συνάρτηση $s(k)$ είναι

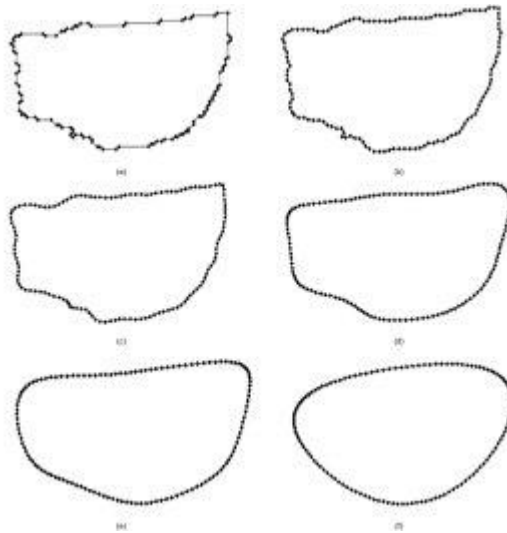
$$s(k) = \frac{1}{K} \sum_{u=0}^{K-1} a(u) e^{j \frac{2\pi uk}{K}} \quad (2.9)$$

Γίνεται έτσι αντιληπτό πως οι περιγραφείς Fourier αποτελούν έναν άλλο τρόπο αναπαράστασης του περιγράμματος στο πεδίο της συχνότητας. Το πλεονέκτημά τους έγκειται στην απεικόνιση των διδιάστατων σημείων του περιγράμματος σε μία μονοδιάστατη συνάρτηση, μειώνοντας την υπολογιστική πολυπλοκότητα.

Σύμφωνα με την παραπάνω ανάλυση, χρειάζονται K περιγραφείς για την ακριβή απεικόνιση και ανακατασκευή του αρχικού περιγράμματος. Έστω ότι χρησιμοποιούνται P περιγραφείς, με $P < K$, θέτοντας μηδενική τιμή στους $P - K$ υπόλοιπους. Τότε, η σχέση (2.9) γίνεται

$$\hat{s}(k) = \frac{1}{P} \sum_{u=0}^{P-1} a(u) e^{j \frac{2\pi u k}{P}} \quad (2.10)$$

Η (2.10) δείχνει πως για την αναπαράσταση και την ανακατασκευή του κάθε σημείου του νέου περιγράμματος, μπορούν να χρησιμοποιηθούν μόνο P όροι της συνάρτησης $a(u)$. Το προσεγγιστικό περίγραμμα που προκύπτει, υστερεί σε λεπτομέρειες έναντι του αρχικού εάν ο αριθμός P είναι πολύ μικρός.



Εικόνα 6: Ανακατασκευή αρχικού περιγράμματος (επάνω δεξιά) χρησιμοποιώντας λιγότερους περιγραφείς Fourier

Ένα ακόμα βασικό πλεονέκτημα των περιγραφέων Fourier είναι η διατήρησή τους σε μεταφορά, περιστροφή, κλιμάκωση και διαφορετικό σημείο εκκίνησης. Ο παρακάτω πίνακας συνοψίζει αυτές τις ιδιότητες.

Μετασχηματισμός	Περίγραμμα	Περιγραφέας Fourier
Ταυτότητα	$s(k)$	$a(u)$
Περιστροφή	$s_r(k) = s(k)e^{j\theta}$	$a_r(u) = a(u)e^{j\theta}$
Κλιμάκωση	$s_s(k) = bs(k)$	$a_s(u) = ba(u)$
Μεταφορά	$s_t(k) = s(k) + \Delta_x + j\Delta_y$	$a_t(u) = a(u) + \Delta_x + j\Delta_y$
Σημείο εκκίνησης	$s_p(k) = s(k - k_0)$	$a_p(u) = a(u)e^{-j2\pi k_0 u/K}$

Πίνακας 1: Ιδιότητες περιγραφέων Fourier

2.5 Ακτινικοί περιγραφείς

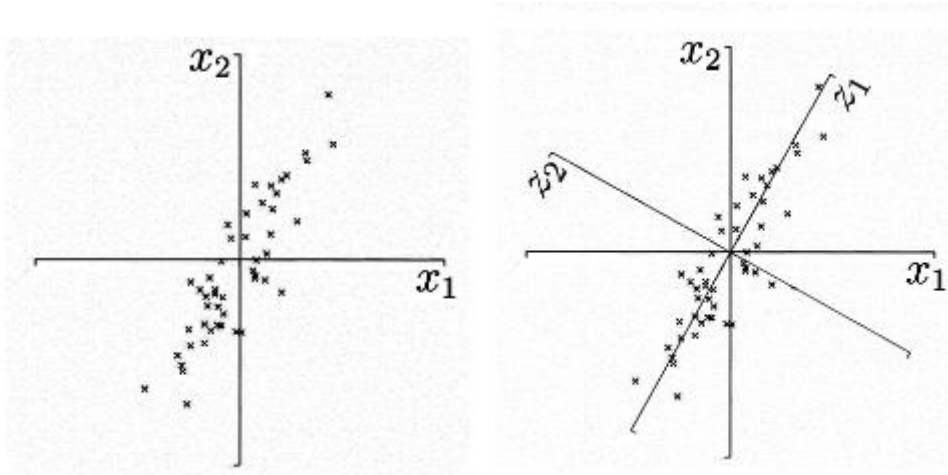
Στην παρούσα εργασία χρησιμοποιήθηκαν ακτινικοί περιγραφείς δύο τύπων, οι *ολικοί* και οι *τοπικοί*. Η κύρια ιδιότητά τους έγκειται στην χρησιμοποίηση ακτίνας που περιστρέφεται και σαρώνει το περίγραμμα του σχήματος. Ο περιγραφέας αποτελείται από την τιμή της ακτίνας σε κάθε γωνία σάρωσης, κανονικοποιημένη στην μέγιστη ακτίνα. Για την εξαγωγή των περιγραφέων χρησιμοποιήθηκε *Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis, PCA)* η οποία παρουσιάζεται στην υποενότητα.

2.5.1 Ανάλυση Κύριων Συνιστωσών

Η Ανάλυση Κύριων Συνιστωσών (PCA) είναι η στατιστική διαδικασία που μετατρέπει ένα σύνολο πιθανώς συσχετισμένων παρατηρήσεων σε ένα σύνολο γραμμικά ασυσχέτιστων τιμών, τις *κύριες συνιστώσες*. Οι κύριες συνιστώσες μπορούν να είναι ίσου ή μικρότερου πλήθους από τις αρχικές.

Στην περίπτωση της επεξεργασίας εικόνων, η PCA μπορεί να χρησιμοποιηθεί για την προβολή των σημείων της εικόνας σε ένα σύστημα συντεταγμένων έτσι ώστε τα νέα

δεδομένα να είναι ασυσχέτιστα μεταξύ τους. Το σύστημα συντεταγμένων που προκύπτει έχει ως αρχή το κεντροειδές της εικόνας και οι άξονές του είναι οι κύρια και η δευτερεύουσα συνιστώσα της εικόνας.



Εικόνα 7: PCA για διοδιάστατα σημεία. Από τους άξονες x_1, x_2 τα δεδομένα προβάλλονται στην νέα βάση z_1, z_2 .

Για την εφαρμογή της PCA είναι απαραίτητη η απεικόνιση των σημείων της εικόνας σε μονοδιάστατο διάνυσμα \mathbf{x} . Η απεικόνιση αυτή για δυαδικές εικόνες γίνεται παίρνοντας κατά στήλη τα δεδομένα κάθε γραμμής της εικόνας. Έτσι, εάν η εικόνα έχει μέγεθος $M \times N$, το διάνυσμα έχει $M \times N$ στοιχεία. Στη συνέχεια υπολογίζεται η μέση τιμή και ο πίνακας συμμεταβλητότητας των δειγμάτων

$$\mathbf{m}_x = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \tag{2.11}$$

$$\mathbf{C}_x = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^T - \mathbf{m}_x \mathbf{m}_x^T$$

Ο πίνακας \mathbf{C}_x είναι συμμετρικός πίνακας πραγματικών τιμών. Κατά συνέπεια, είναι εφικτή η εύρεση ενός συνόλου n ορθοκανονικών ιδιοδιανυσμάτων \mathbf{e}_i . Θεωρείται ο πίνακας

\mathbf{A} ο οποίος αποτελείται από τα ιδιοδιανύσματα \mathbf{e}_i κατά τέτοιο τρόπο ώστε η 1^η γραμμή του \mathbf{A} να είναι το ιδιοδιάνυσμα \mathbf{e}_i που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή του \mathbf{C}_x και η τελευταία γραμμή να είναι το ιδιοδιάνυσμα που αντιστοιχεί στη μικρότερη ιδιοτιμή. Το διάνυσμα \mathbf{x} μετασχηματίζεται στο διάνυσμα \mathbf{y} στην νέα βάση, σύμφωνα με τον μετασχηματισμό *Hotelling*

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_x) \quad (2.12)$$

Η μέση τιμή των διανυσμάτων \mathbf{y} είναι ίση με 0. Ο πίνακας συμμεταβλητότητας \mathbf{C}_y δίνεται από τη σχέση $\mathbf{C}_y = \mathbf{A}\mathbf{C}_x\mathbf{A}^T$. Οι πίνακες $\mathbf{C}_y, \mathbf{C}_x$ έχουν τις ίδιες ιδιοτιμές. Για την ανακατασκευή του διανύσματος \mathbf{x} χρησιμοποιείται ο τύπος

$$\mathbf{x} = \mathbf{A}^T\mathbf{y} + \mathbf{m}_x \quad (2.13)$$

Εάν χρησιμοποιηθεί ο πίνακας \mathbf{A}_k που περιέχει τα πρώτα k ιδιοδιανύσματα του \mathbf{C}_x , το ανακατασκευασμένο διάνυσμα είναι το

$$\hat{\mathbf{x}} = \mathbf{A}_k^T\mathbf{y} + \mathbf{m}_x \quad (2.14)$$

Το μέσο τετραγωνικό σφάλμα ανάμεσα στα δύο ανακατασκευασμένα διανύσματα είναι

$$e_{ms} = \sum_{j=1}^n \lambda_j - \sum_{j=1}^k \lambda_j = \sum_{j=k+1}^n \lambda_j \quad (2.15)$$

όπου λ_j οι ιδιοτιμές των πινάκων C_y, C_x

2.5.2 Ολικοί ακτινικοί περιγραφείς

Ένας ολικός ακτινικός περιγραφέας αποτελεί την διακριτή συναρτησιακή αναπαράσταση ενός περιγράμματος. Έτσι, για την απεικόνιση του περιγράμματος χρειάζεται μόνο η μονοδιάστατη ακτινική συνάρτηση αντί των δισδιάστατων σημείων του περιγράμματος. Ονομάζεται επίσης και *υπογραφή* σχήματος.

Κύρια ιδέα του ολικού περιγραφέα είναι η αναπαράσταση της απόστασης κάθε σημείου του περιγράμματος από το κεντροειδές, ξεκινώντας από ένα σημείο και κινούμενοι με φορά αντίθετη των δεικτών του ρολογιού, προς την γωνία που σχηματίζει αυτή η απόσταση με τον πρωτεύοντα άξονα του σχήματος. Το αποτέλεσμα είναι μία συνάρτηση $r - \theta$ της ακτίνας του σχήματος με την αντίστοιχη γωνία. Η συνάρτηση αυτή είναι διακριτή, καθώς η επεξεργασία γίνεται ανά εικονοστοιχείο. Η δημιουργία της υπογραφής μπορεί να γίνει και αντίστροφα, επιλέγοντας σταθερές τιμές για τη γωνία και υπολογίζοντας την αντίστοιχη απόσταση.

Ένα μεγάλο πλεονέκτημα της χρήσης ολικού περιγραφέα είναι το αναλλοίωτο ως προς τη μετατόπιση. Οπουδήποτε και αν βρίσκεται το προς περιγραφή περίγραμμα μέσα στην εικόνα, η υπογραφή του θα είναι ίδια. Το αναλλοίωτο ως προς την περιστροφή και ως προς την κλιμάκωση δεν είναι εγγυημένο με την χρήση της βασικής μεθόδου της υπογραφής, όμως μπορεί να επιτευχθεί με τροποποίηση της μεθόδου.

Ως προς την περιστροφή, η εκκίνηση από ένα σταθερό σημείο μπορεί να οδηγήσει στο αναλλοίωτο, ανεξάρτητα από τον προσανατολισμό του σχήματος. Μία πρώτη

προσέγγιση είναι η εκκίνηση πάντα από το σημείο της μεγαλύτερης απόστασης από το κεντροειδές. Θεωρητικά, αυτή η μέθοδος είναι σωστή, πρακτικά όμως η παρουσία θορύβου είναι ικανή να αλλοιώσει τα αποτελέσματα. Επιπλέον, η κατάτμηση της εικόνας και η εξαγωγή ενός περιγράμματος σε συμμετρικό αντικείμενο, μπορεί να οδηγήσει στην ιδιόμορφη περίπτωση της ύπαρξης παραπάνω του ενός σημείου που θα έχει αυτή τη μέγιστη απόσταση από το κεντροειδές του σχήματος. Για τους λόγους αυτούς ως σημείο εκκίνησης επιλέγεται το σημείο τομής του κύριου άξονα με το περίγραμμα, διότι ο κύριος άξονας του σχήματος είναι σταθερός και ανεξάρτητος της περιστροφής. Αυτή η μέθοδος απαιτεί περισσότερους υπολογισμούς, αλλά είναι πιο εύρωστη και εγγυάται την ανεξαρτησία από τον προσανατολισμό του σχήματος.

Εφ' όσον η υπογραφή είναι ομοιόμορφη ως προς την κλιμάκωση και την περιστροφή, οι μεταβολές στο μέγεθος ενός αντικειμένου προκαλούν μεταβολές στις τιμές του πλάτους του ολικού περιγραφέα. Αναμένεται λοιπόν μετακίνηση της συνάρτησης προς τα επάνω ή προς τα κάτω, ανάλογα εάν γίνεται μετατροπή σε μεγαλύτερη ή μικρότερη κλίμακα αντίστοιχα. Για να αντιμετωπιστεί αυτή η εξάρτηση, συνηθέστερος τρόπος είναι η κανονικοποίηση του μήκους κάθε ακτίνας, έτσι ώστε αυτό να βρίσκεται πάντοτε εντός κάποιου προκαθορισμένου εύρους τιμών, π.χ. στο διάστημα $[0,1]$. Ως βάση κανονικοποίησης μπορεί να χρησιμοποιηθεί η τιμή της μέγιστης ακτίνας, δίνοντας πολύ καλά αποτελέσματα, τα οποία όμως μπορεί να έχουν εξάρτηση από πιθανό θόρυβο στην εικόνα. Ένας πιο αξιόπιστος τρόπος είναι η κανονικοποίηση ως προς τη μεταβλητότητα της ακτινικής συνάρτησης, εάν υποθεθεί πως η μεταβλητότητα αυτή δεν είναι μηδενική. Παρ' όλες τις υπολογιστικές δυσκολίες της, η μέθοδος αυτή δημιουργεί έναν μεταβλητό παράγοντα κλιμάκωσης που είναι αντιστρόφως ανάλογος των μεταβολών του μεγέθους.

Για τη δημιουργία της υπογραφής, στο δεδομένο περίγραμμα εφαρμόζεται PCA για την εύρεση των αξόνων του περιγράμματος. Εφαρμόζοντας τον μετασχηματισμό Hotelling της σχέσης (2.12), τα σημεία του περιγράμματος προβάλλονται στον κύριο και τον δευτερεύοντα άξονα, με το κεντροειδές του σχήματος ως κέντρο. Δημιουργείται έτσι ένα νέο σύστημα συντεταγμένων y_1, y_2 , με τις τιμές κάθε εικονοστοιχείου να δίνονται από την (2.12). Σε αυτό το σύστημα συντεταγμένων, ο υπολογισμός των r, θ γίνεται ως εξής

$$r = \sqrt{y_1^2 + y_2^2} \tag{2.16}$$

$$\theta = \tan^{-1}\left(\frac{y_2}{y_1}\right)$$

Στη συνέχεια οι νέες ακτίνες προκύπτουν από τη σχέση $r = \frac{r}{r_{\max}}$, για να επιτευχθεί η κανονικοποίηση τους. Το αποτέλεσμα είναι ένας πίνακας που περιέχει στην 1^η στήλη τις ακτίνες και στην 2^η τις αντίστοιχες γωνίες, οι στήλες του οποίου ταξινομούνται κατά αύξουσα γωνία. Ο τελικός, κανονικοποιημένος και ταξινομημένος πίνακας ξεκινά πάντα από το σημείο τομής του πρωτεύοντα άξονα με το περίγραμμα, είναι ανεξάρτητος ως προς την κλιμάκωση, την περιστροφή, τη μεταφορά και την κλίμακα και αποτελεί την ολική ακτινική περιγραφή του σχήματος.

2.5.3 Τοπικοί ακτινικοί περιγραφείς

Παρόμοια με τους ολικούς ακτινικούς περιγραφείς, η φιλοσοφία των τοπικών ακτινικών περιγραφέων είναι η αναπαράσταση απόστασης προς γωνία. Η διαφορά έγκειται στο γεγονός πως στους τοπικούς ακτινικούς περιγραφείς το κέντρο ως προς το οποίο υπολογίζονται οι αποστάσεις δεν είναι σταθερό, αλλά μετακινείται επάνω στο περίγραμμα. Παράλληλα, ο υπολογισμός των αποστάσεων γίνεται για σταθερή γωνία και όχι για όλα τα σημεία του περιγράμματος.

Οι τοπικοί ακτινικοί περιγραφείς αποτελούν μια μαθηματική αναπαράσταση της μορφολογίας του αντικειμένου, σε κάθε γειτονιά του. Η αναπαράσταση είναι μονοδιάστατη διακριτή συνάρτηση $r - \theta$, για δεδομένες κάθε φορά γωνίες, μειώνοντας έτσι την απαιτούμενη υπολογιστική πολυπλοκότητα και παρέχοντας μία απλή αναπαράσταση της τοπικής μορφολογίας του περιγράμματος. Έτσι, μπορούν να ανιχνεύονται σε κάθε περιοχή του αντικειμένου γωνίες, καμπύλες, προεξοχές, ή οποιουδήποτε άλλου είδους διαβαθμίσεις στο περίγραμμα του αντικειμένου, οι οποίες χαρακτηρίζονται από παρόμοιο τοπικό περιγραφέα. Οπτικά αυτό αποτυπώνεται σε αλλαγές στην κλίση της τοπικής ακτινικής

συνάρτησης. Η προσέγγιση αυτή δίνει την δυνατότητα στο ρομπότ να τοποθετήσει το χέρι του σε διαφορετικά σημεία του περιγράμματος του αντικειμένου, προτιμώντας έτσι σημεία που έχουν εύκολη δυνατότητα λαβής.

Οι τοπικοί ακτινικοί περιγραφείς είναι αναλλοίωτοι ως προς τη μεταφορά και την περιστροφή, λόγω του τρόπου υπολογισμού τους, ο οποίος γίνεται τοπικά και χρησιμοποιώντας παράγοντες ανεξάρτητους από αυτούς τους μετασχηματισμούς. Η ανεξαρτητοποίηση από την κλιμάκωση επιτυγχάνεται διαιρώντας όλες τις αποστάσεις με την τοπική μέγιστη ακτίνα, αντιστοιχίζοντας τον περιγραφέα στο διάστημα $[0,1]$. Το αναλλοίωτο ως προς τους μετασχηματισμούς κάνει τους τοπικούς ακτινικούς περιγραφείς πολύ χρήσιμο εργαλείο στην ανίχνευση μορφολογίας του σχήματος.

Για τον υπολογισμό των τοπικών ακτινικών περιγραφέων, αρχικά το περίγραμμα υφίσταται δειγματοληψία ανά n δείγματα. Τα δειγματελημμένα σημεία του περιγράμματος θα αποτελέσουν στην συνέχεια τα σημεία βάσης από τα οποία γίνεται ο υπολογισμός των τοπικών αποστάσεων. Κάθε δειγματελημμένο σημείο βρίσκεται λοιπόν εντός μίας γειτονιάς σημείων του περιγράμματος με $n - 1$ σημεία στα δεξιά του και $n - 1$ σημεία στα αριστερά του. Λαμβάνοντας το σύνολο των $2n - 1$ σημείων, συμπεριλαμβανομένου του σημείου βάσης υπολογίζεται η ευθεία που παρεμβάλλει καλύτερα τα σημεία με τη μέθοδο των *ελαχίστων τετραγώνων*. Η ευθεία που προκύπτει έχει την διανυσματική μορφή

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}t \quad (2.17)$$

όπου \mathbf{r} το διάνυσμα του εκάστοτε σημείου της γραμμής, \mathbf{r}_0 το διάνυσμα του εκάστοτε σημείου βάσης, \mathbf{v} ένα διάνυσμα παράλληλο στην γραμμή και t βαθμωτή μεταβλητή.

Στη συνέχεια υπολογίζεται η κάθετη της ευθείας που ξεκινά από το αντικείμενο και επεκτείνεται προς το εσωτερικό του αντικειμένου. Ο υπολογισμός της κάθετης απαιτεί τον υπολογισμό του διανύσματος \mathbf{u} για το οποίο ισχύει $\mathbf{u} \cdot \mathbf{v} = 0$. Τα διανύσματα που ικανοποιούν αυτή την εξίσωση είναι τα $\mathbf{u} = \begin{pmatrix} \mp \mathbf{v}_1 \\ \pm \mathbf{v}_2 \end{pmatrix}$. Επιλέγεται κάθε φορά το διάνυσμα εκείνο που η φορά του είναι προς το εσωτερικό του περιγράμματος. Η εξίσωση της κάθετου λοιπόν είναι ίδιας μορφής με την (2.17)

$$\mathbf{r}_n = \mathbf{r}_0 + \mathbf{u}t \quad (2.17)$$

Προχωρώντας στην κατεύθυνση αυτής της ευθείας, ευρίσκεται το σημείο τομής της με το περίγραμμα του αντικειμένου.

Με τον ίδιο τρόπο υπολογίζονται ευθείες που ξεκινούν από το σημείο βάσης και σχηματίζουν γωνία θ με την κάθετη. Αναζητείται το διάνυσμα \mathbf{n} προς το εσωτερικό του περιγράμματος για το οποίο ισχύει

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{n}}{|\mathbf{u}| |\mathbf{n}|} \quad (2.18)$$

Για ευκολία, τα διανύσματα \mathbf{n} μπορούν να θεωρηθούν μοναδιαίου μέτρου, οπότε οι συνιστώσες τους προκύπτουν από την λύση του συστήματος των εξισώσεων

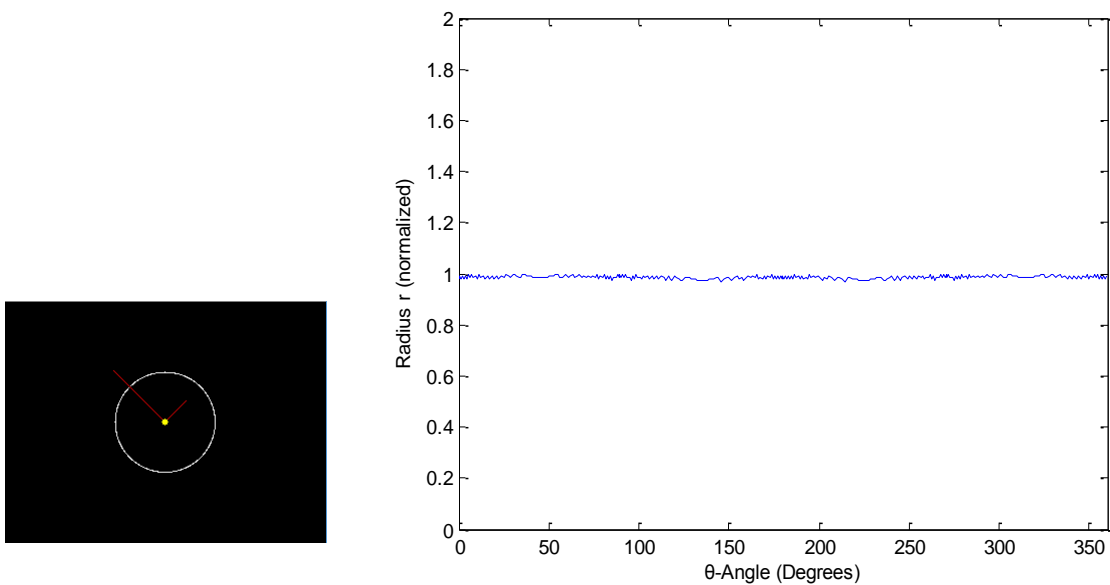
$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{n}}{|\mathbf{u}| |\mathbf{n}|} \quad (2.19)$$

$$|\mathbf{n}| = 1$$

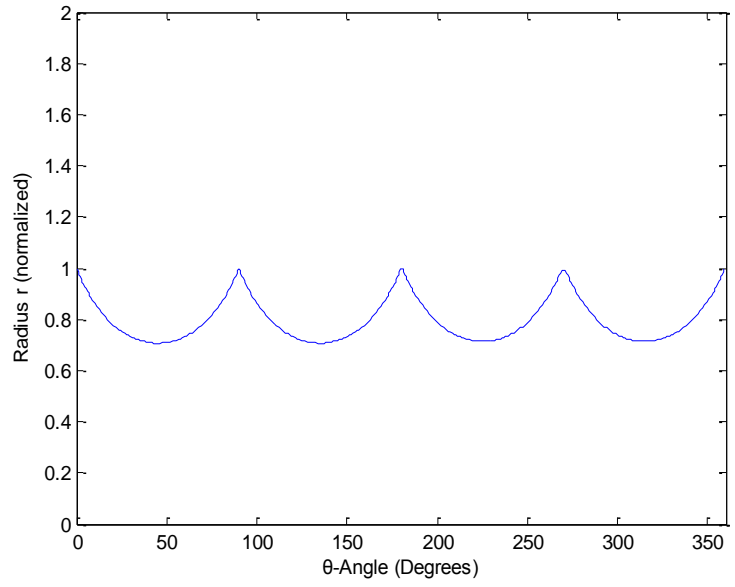
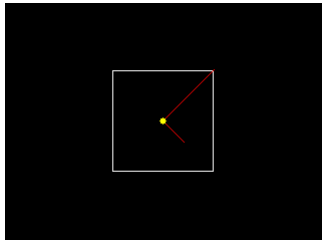
Η γωνία θ μπορεί να μεταβάλλεται ελεύθερα. Εκτελείται λοιπόν μία διαδικασία σάρωσης για γωνίες θ αριστερά της κάθετης και θ δεξιά της κάθετης. Βρίσκοντας κάθε φορά τα σημεία τομής της ακτίνας για διαφορετικές γωνίες με τα σημεία του περιγράμματος, υπολογίζοντας το μήκος της ακτίνας σύμφωνα με την σχέση (2.16) και κανονικοποιώντας ως προς την μέγιστη ακτίνα, δημιουργείται ο τοπικός ακτινικός περιγραφέας για το εκάστοτε δειγματελημένο σημείο ως ένας πίνακας που στην 1^η στήλη περιέχει τις κανονικοποιημένες τιμές της ακτίνας και στην 2^η στήλη τις διακριτές γωνίες από $-\theta$ έως θ .

2.6 Παραδείγματα

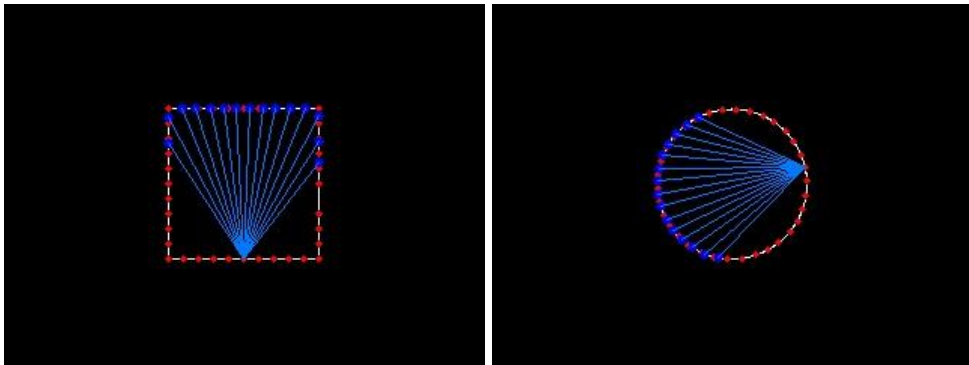
Στις επόμενες εικόνες φαίνονται παραδείγματα ακτινικών περιγραφών για δύο περιγράμματα, ένα ενός κύκλου και ένα ενός τετραγώνου. Άξιο αναφοράς είναι το γεγονός πως ο ολικός περιγραφέας του κύκλου είναι μία ευθεία γραμμή, με τυχόν παρεκκλίσεις να οφείλονται σε σφάλματα ανάλυσης κάμερας και εικονοστοιχείων, ενώ ο ολικός περιγραφέας του τετραγώνου είναι περιοδικός. Αυτές οι ιδιότητες είναι μέτρο της ομοιομορφίας και της συμμετρίας των σχημάτων. Ο τοπικός περιγραφέας στην περίπτωση του κύκλου εμφανίζει ελαφριά κλίση, όπως και στην περίπτωση του τετραγώνου. Στην γωνία του τετραγώνου, υπάρχει μία μεγάλη αυξομείωση του πλάτους της συνάρτησης. Με βάση τις παραπάνω παρατηρήσεις γίνεται αντιληπτή η χρησιμότητα των ακτινικών περιγραφών στην χαρτογράφηση της μορφολογίας ενός σχήματος.



Εικόνα 8: Πρωτεύοντες άξονες και ολικός ακτινικός περιγραφέας r - θ για τον κύκλο



Εικόνα 9: Προτεύοντες άξονες και ολικός ακτινικός περιγραφείας $r-\theta$ για το τετράγωνο



Εικόνα 10: Οπτική απεικόνιση τοπικού ακτινικού περιγραφείας

ΚΕΦΑΛΑΙΟ 3

3 *Βαθμονόμηση κάμερας και εύρεση τρισδιάστατων σημείων*

3.1 *Εισαγωγή*

Με την ανάπτυξη της τεχνολογίας των καμερών, αλλά και με την εμφάνιση υπολογιστών που εκτελούν ταχύτερα περισσότερες πράξεις και επεξεργάζονται περισσότερα δεδομένα σε μικρό χρονικό διάστημα, δημιουργήθηκαν και εξελίχθηκαν οι κλάδοι της *Όρασης υπολογιστών* και της *Επεξεργασίας εικόνας*. Ένα από τα βασικά πεδία ενδιαφέροντος των επιστημών αυτών, είναι η ανάκτηση τρισδιάστατης πληροφορίας από δισδιάστατες εικόνες κάμερας.

Η ρομποτική όραση είναι πολύ σημαντική για την δημιουργία πληρέστερων ρομποτικών συστημάτων. Με την χρήση οπτικών μέσων, τα ρομπότ είναι σε θέση να διαισθανθούν το περιβάλλον τους και να αποκομίσουν από αυτό πληροφορίες, κατά τρόπο παρόμοιο με εκείνον που βλέπει ο άνθρωπος. Τέτοιες πληροφορίες είναι το βάθος των εκάστοτε τοπίων, σκηνικών ή αντικειμένων, το χρώμα τους, την υφή τους και τον όγκο τους. Οι πληροφορίες αυτές ανατροφοδοτούνται στο σύστημα ελέγχου και επεξεργάζονται ώστε να προκύψει μία επιθυμητή δράση για την απαιτούμενη εργασία.

Βασικό στοιχείο για την λήψη οπτικών πληροφοριών είναι η *κάμερα*. Με τον όρο *κάμερα*, εννοείται το μαθηματικό μοντέλο που μετατρέπει 3D σημεία του χώρου σε 2D σημεία στο επίπεδο της εικόνας. Στο κεφάλαιο αυτό θα αναλυθεί το απλούστερο μοντέλο κάμερας, το μοντέλο *κάμερας οπής* (*Pinhole camera*).

Στη συνέχεια θα γίνει αναφορά στους τρόπους *βαθμονόμησης κάμερας*, δηλαδή στις μεθόδους υπολογισμού των παραμέτρων του μοντέλου της κάμερας, έχοντας ως δεδομένα

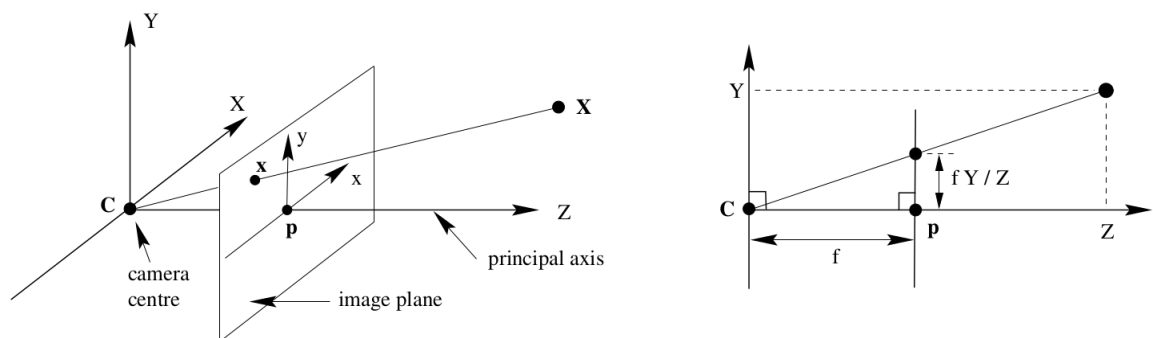
σημεία στο χώρο και την αντιστοίχισή τους στις εκάστοτε εικόνες. Ο βασικός αλγόριθμος, που χρησιμοποιείται ευρέως είναι ο αλγόριθμος του Zhang [1].

Τέλος, θα παρουσιαστεί ο αλγόριθμος εξαγωγής ενός 3D σημείου από μία εικόνα. Ο αλγόριθμος αυτός εκμεταλλεύεται στοιχεία του χώρου και με χρήση στοιχείων αναλυτικής γεωμετρίας χρησιμοποιείται για τον προσδιορισμό στο χώρο των τελικών σημείων τοποθέτησης του αντίχειρα του ρομπότ.

3.2 Μοντέλο κάμερας οπής

Η λειτουργία της κάμερας έγκειται στην μέτρηση του φωτός που ανακλάται από ένα αντικείμενο και την απεικόνισή του σε ένα *εικονοστοιχείο (pixel)* το οποίο μετατρέπει την ενέργεια του φωτός σε ηλεκτρικό ρεύμα. Μία κάμερα λοιπόν αποτελείται από έναν πίνακα εικονοστοιχείων, με διαφορετικές τιμές έντασης φωτός.

Με τον όρο κάμερα οπής, εννοείται ένα μαθηματικό μοντέλο το οποίο μετατρέπει 3D σημεία του επιπέδου σε 2D σημεία, προβάλλοντάς τα στο *επίπεδο εικόνας* μέσω ακτίνων που περνούν από το *κέντρο της κάμερας*. Το μοντέλο απεικονίζεται στην εικόνα 3.1



Εικόνα 11.1 : Μοντέλο κάμερας οπής (προοπτική προβολή)

Όπως φαίνεται, ένα σημείο του χώρου με συντεταγμένες $P = (X, Y, Z)$ ως προς το σύστημα συντεταγμένων της κάμερας, απεικονίζεται στο σημείο $p = (x, y, f)$ στο σύστημα

συντεταγμένων πάνω στο επίπεδο της κάμερας, μέσω μίας ακτίνας που περνά από το κέντρο της κάμερας C . Η απόσταση f είναι σταθερή και ονομάζεται *εστιακό βάθος* της κάμερας. Η ευθεία που περνά από το κέντρο της κάμερας και είναι κάθετη στο επίπεδο της εικόνας καλείται *πρωτεύων άξονας* της κάμερας. Το σημείο τομής του επιπέδου της εικόνας και του πρωτεύοντος άξονα ονομάζεται *πρωτεύων σημείο*.

Με χρήση της μεθόδου των όμοιων τριγώνων στο δεύτερο σχήμα της εικόνας 3.1, προκύπτουν οι εξισώσεις *προοπτικής προβολής* για το σημείο \mathbf{p} (Forsyth & Ponce, 2011)

$$x = f \frac{X}{Z} \tag{3.1}$$

$$y = f \frac{Y}{Z} \tag{3.2}$$

Οι εξισώσεις προοπτικής προβολής για το σημείο \mathbf{p} εκφράζουν τις συντεταγμένες του ως προς το πλαίσιο της κάμερας, δηλαδή το πλαίσιο με αρχή το κέντρο της κάμερας όπως φαίνεται στην Εικόνα 3.1. Η λειτουργία όμως της κάμερας χρησιμοποιεί pixels για την απεικόνιση σημείων. Αυτό σημαίνει πως το δεδομένο σημείο στο πλαίσιο της κάμερας (x, y) , πρέπει να εκφραστεί ως προς το πλαίσιο των εικονοστοιχείων, δίνοντας το σημείο (u, v) . Στο μοντέλο της κάμερας οπής, τα pixels θεωρούνται ίδιας κλίμακας και στις δύο αξονικές κατευθύνσεις, δηλαδή τετράγωνα. Αυτό δεν είναι πάντα ορθό σε εφαρμογές, όπου σε κάποιες κάμερες υπάρχει η πιθανότητα μη τετραγωνικών εικονοστοιχείων, οδηγώντας σε άνισες κλίμακες κατά μήκος των αξονικών κατευθύνσεων. Εάν ορίσουμε σαν α, β τον αριθμό pixel ανά μονάδα απόστασης μετρημένη σε εικονικές συντεταγμένες κατά μήκος των κατευθύνσεων x, y αντίστοιχα, ο μετασχηματισμός που συνδέει τα σημεία (x, y) και (u, v) δίνεται από τις παρακάτω σχέσεις

$$u = a \frac{X}{Z} + x_0 \quad (3.3)$$

$$v = a \frac{Y}{Z} + y_0 \quad (3.4)$$

Οι μεταβλητές x_0, y_0 υποδηλώνουν την μετακίνηση του κεντρικού pixel του επιπέδου της εικόνας ως προς τον πρωτεύον άξονα.

Εάν θεωρήσουμε ομογενείς συντεταγμένες για τα σημεία (u, v) και (x, y, f) , δηλαδή εάν τα γράψουμε στη μορφή $(u, v, 1)$ και $(X, Y, Z, 1)$, οι σχέσεις (3.1) – (3.4) μπορούν να γραφτούν σε μητρική μορφή

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{K} \begin{pmatrix} I_{3 \times 3} & \mathbf{0} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.5)$$

όπου

$$\mathbf{K} = \begin{pmatrix} f\alpha & 0 & X_o \\ 0 & f\beta & Y_o \\ 0 & 0 & 1 \end{pmatrix}$$

Ο πίνακας \mathbf{K} ονομάζεται *πίνακας εσωτερικών παραμέτρων της κάμερας*. Αποτελεί τον μετασχηματισμό ενός 3Δ σημείου εκφρασμένο στο πλαίσιο της κάμερας, στο 2Δ σημείο εκφρασμένο σε συντεταγμένες εικονοστοιχείων της κάμερας. Τα σημεία αναπαρίστανται σε ομογενείς συντεταγμένες.

Το ομογενές σημείο $(X, Y, Z, 1)$ ως προς το σύστημα συντεταγμένων της κάμερας, έχει μοναδική έκφραση $(X_w, Y_w, Z_w, 1)$ στο πλαίσιο του χώρου. Ο μετασχηματισμός που συνδέει τα πλαίσια της κάμερας και του χώρου είναι της μορφής

$$A_w^c = (R_{3 \times 3} \quad T_{3 \times 1}) \quad (3.6)$$

Όπου R, T είναι οι πίνακες περιστροφής και μεταφοράς του πλαισίου της κάμερας ως προς το πλαίσιο του χώρου. Τα στοιχεία αυτών των πινάκων ονομάζονται *εξωτερικές παράμετροι της κάμερας*. Με βάση τον παραπάνω μετασχηματισμό, οι ομογενείς συντεταγμένες ενός σημείου του 3Δ χώρου, εκφρασμένες στο πλαίσιο της κάμερας δίνονται από τη σχέση

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = (R_{3 \times 3} \quad T_{3 \times 1}) \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (3.7)$$

Η τελική λειτουργία της κάμερας οπής, δηλαδή η αντιστοίχιση 3Δ σημείων σε 2Δ σημεία εικονοστοιχείων, μοντελοποιείται πλήρως αντικαθιστώντας την (3.7) στην (3.5)

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K (R_{3 \times 3} \quad T_{3 \times 1}) \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (3.8)$$

Όπου s ένας αυθαίρετος παράγοντας κλίμακας.

3.3 Βαθμονόμηση κάμερας

Όπως αναφέρθηκε στην προηγούμενη ενότητα, για την λειτουργία της κάμερας είναι απαραίτητο να είναι γνωστές τόσο οι εξωτερικές όσο και οι εσωτερικές παράμετροι της κάμερας. Σε πολλές εφαρμογές ρομποτικής όρασης αυτό επιτυγχάνεται με την διαδικασία της βαθμονόμησης κάμερας. Η *βαθμονόμηση κάμερας* αποτελεί τη διαδικασία κατά την οποία προσδιορίζονται οι παράμετροι της κάμερας, με βάση δεδομένα σημεία στον 3Δ χώρο και τις αντιστοιχίσεις τους στον χώρο εικονοστοιχείων. Η βαθμονόμηση

επιτυγχάνεται με τη βοήθεια δισδιάστατων αντικειμένων βαθμονόμησης, ή σπανιότερα με τρισδιάστατα ή μονοδιάστατα αντικείμενα. Στην παρούσα ενότητα αναλύεται η μέθοδος βαθμονόμησης με ομογραφήματα, η οποία οφείλεται στον Zhang [Zhang,2000].

3.3.1 Ομογράφημα

Στην επιστήμη της όρασης υπολογιστών, με τον όρο *ομογράφημα* εννοείται ο μετασχηματισμός που συνδέει δύο διαφορετικές εικόνες που απεικονίζουν το ίδιο επίπεδο. Μαθηματικά, το ομογράφημα δύο εικόνων ενός επιπέδου ορίζεται ως ένας πίνακας 3x3 με βάση την παρακάτω σχέση

$$p_a = H p_b \quad (3.9)$$

όπου p_a, p_b είναι τα 3Δ σημεία του επιπέδου σε ομογενείς συντεταγμένες ως προς τις δύο κάμερες.

Στην περίπτωση της εξίσωσης κάμερας οπής, είναι βολικό να θεωρήσουμε πως το επίπεδο είναι το $Z = 0$. Εάν ορίσουμε τα διανύσματα r_1, r_2, r_3 ως τα διανύσματα-στήλες του πίνακα R , η εξίσωση (3.8) γίνεται

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \begin{pmatrix} r_1 & r_2 & r_3 & T \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{pmatrix} \quad (3.8)$$

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \begin{pmatrix} r_1 & r_2 & T \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix} \quad (3.9)$$

Ο πίνακας ομογραφήματος, που συνδέει τα ομογενή 3Δ σημεία πάνω στο επίπεδο και τις ομογενείς προβολές τους στο επίπεδο εικονοστοιχείων είναι ο

$$H = K (r_1 \ r_2 \ T) \quad (3.10)$$

3.3.2 Αλγόριθμος βαθμονόμησης

Εάν το ομογράφημα γραφτεί ως

$$H = (h_1 \ h_2 \ h_3) \quad (3.11)$$

με h_1, h_2, h_3 τα διανύσματα στήλες του πίνακα H , και χρησιμοποιώντας την πληροφορία ορθοκανονικότητας των r_1, r_2 , παίρνουμε τους περιορισμούς [Zhang,2000]

$$h_1^T K^{-T} K^{-1} h_2 = 0 \quad (3.12)$$

$$h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2 \quad (3.13)$$

Οι δύο αυτοί περιορισμοί είναι πολύ βασικοί για την περίπτωση ενός ομογραφήματος, όσον αφορά τις εσωτερικές παραμέτρους της κάμερας. Επειδή ένα ομογράφημα έχει 8 βαθμούς ελευθερίας ενώ υπάρχουν 6 εξωτερικές παράμετροι (3 για περιστροφή και 3 για μεταφορά), είναι δυνατόν να έχουμε 2 περιορισμούς στις εσωτερικές παραμέτρους. Στην περίπτωση που το μοντέλο βαθμονόμησης στη 2^η θέση είναι παράλληλο σε σχέση με αυτό της 1^{ης} θέσης, το 2^ο ομογράφημα δεν παρέχει επιπλέον γεωμετρικούς περιορισμούς. Γι αυτό, για την εύρεση του ομογραφήματος χρειαζόμαστε σημεία από ένα

επίπεδο, από τουλάχιστον δύο μη παράλληλες λήψεις. Επιπλέον λήψεις απλά προσθέτουν περιορισμούς, οδηγώντας σε πιο ακριβή αποτελέσματα.

Για την εύρεση αναλυτικής λύσης του ομογραφήματος ορίζουμε τον αντισυμμετρικό πίνακα

$$\begin{aligned}
 \mathbf{B} = \mathbf{K}^{-T} \mathbf{K}^{-1} &= \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} \\ \mathbf{B}_{12} & \mathbf{B}_{22} & \mathbf{B}_{23} \\ \mathbf{B}_{13} & \mathbf{B}_{23} & \mathbf{B}_{33} \end{pmatrix} \\
 &= \begin{pmatrix} \frac{1}{\alpha^2} & 0 & \frac{-X_0}{\alpha^2} \\ 0 & \frac{1}{\beta^2} & \frac{-Y_0}{\beta^2} \\ \frac{-X_0}{\alpha^2} & \frac{-Y_0}{\beta^2} & 1 + \frac{X_0}{\alpha^2} + \frac{Y_0}{\beta^2} \end{pmatrix}
 \end{aligned} \tag{3.14}$$

Με αυτόν τον τρόπο, οι εξισώσεις (3.12) – ((3.13) μπορούν να γραφτούν στην μορφή

$$\mathbf{A} \mathbf{b} = \mathbf{0} \tag{3.15}$$

Όπου \mathbf{A} ο πίνακας 2x6 με συντελεστές που εξαρτώνται από τα $\mathbf{h}_1, \mathbf{h}_2$ που προκύπτει εάν εκτελεστούν οι πράξεις και \mathbf{b} ο πίνακας που ορίζεται ως

$$\mathbf{b} = (\mathbf{B}_{11} \quad \mathbf{B}_{12} \quad \mathbf{B}_{22} \quad \mathbf{B}_{13} \quad \mathbf{B}_{23} \quad \mathbf{B}_{33})^T \tag{3.16}$$

Εάν υπάρχουν k εικόνες του αντικειμένου βαθμονόμησης, προκύπτουν $2k$ εξισώσεις της μορφής (3.16). Για $k \geq 3$, οι εξισώσεις έχουν μοναδική λύση της μορφής $\gamma \mathbf{b}$ όπου γ συντελεστής κλιμάκωσης. Οι τελικές εσωτερικές παράμετροι της κάμερας προσδιορίζονται ως εξής [Zhang, 2000]

$$Y_0 = \frac{(B_{12}B_{13} - B_{11}B_{23})}{(B_{11}B_{22} - B_{12}^2)} \quad (3.17)$$

$$\gamma = B_{33} - [B_{13}^2 + \gamma(B_{12}B_{13} - B_{11}B_{23})]/B_{11} \quad (3.18)$$

$$\alpha = \sqrt{\frac{\gamma}{B_{11}}} \quad (3.19)$$

$$\beta = \sqrt{\frac{\gamma B_{11}}{(B_{11}B_{22} - B_{12}^2)}} \quad (3.20)$$

$$X_0 = -\frac{B_{13}\alpha^2}{\gamma} \quad (3.21)$$

Έχοντας υπολογίσει τον πίνακα \mathbf{K} των εσωτερικών παραμέτρων, ο προσδιορισμός των εξωτερικών παραμέτρων γίνεται ως εξής [Zhang,2000]

$$\mathbf{r}_1 = \gamma \mathbf{K}^{-1} \mathbf{h}_1, \mathbf{r}_2 = \gamma \mathbf{K}^{-1} \mathbf{h}_2, \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2, \mathbf{T} = \gamma \mathbf{K}^{-1} \mathbf{h}_3 \quad (3.22)$$

3.4 Εκτίμηση πόζας και εξαγωγή 3Δ σημείων

Σε πολλές ερευνητικές εφαρμογές ρομποτικής όρασης και εκπαίδευσης, γίνεται η παραδοχή του επίπεδου τραπέζιου, δηλαδή τα προς εξέταση αντικείμενα τοποθετούνται σε τραπέζι, η επιφάνεια του οποίου θεωρείται επίπεδη. Το ρομπότ έχει την κάμερά του στραμμένη προς το τραπέζι και παρατηρεί τα αντικείμενα. Αυτή η προσέγγιση χρησιμεύει στην ευκολότερη εύρεση των 3Δ συντεταγμένων ενός σημείου του αντικειμένου που ακουμπά στο τραπέζι. Έχοντας γνωστά 4 μη συγγραμικά σημεία του επιπέδου, υπολογίζεται η εξίσωσή του. Το 3Δ σημείο ευρίσκεται ως η τομή μίας ακτίνας από την κάμερα με το επίπεδο. Στην ενότητα αυτή αναλύεται η μεθοδολογία.

3.4.1 Εκτίμηση πόζας της κάμερας σχετικά με επίπεδο

Για να είναι το ρομπότ σε θέση να εξάγει τις 3Δ συντεταγμένες του αντικειμένου, πρέπει να γνωρίζει που βρίσκεται ως προς το σύστημα συντεταγμένων του χώρου. Η διαδικασία κατά την οποία επιτυγχάνεται αυτό λέγεται *εκτίμηση πόζας* της κάμερας. Με την εκτίμηση πόζας, είναι δυνατή η εκτίμηση των πινάκων \mathbf{R}, \mathbf{T} της κάμερας, έχοντας ως γνωστές τις εσωτερικές παραμέτρους \mathbf{K} . Η προσέγγιση αυτή δεν πρέπει να συγχέεται με την βαθμονόμηση που περιεγράφηκε σε προηγούμενη ενότητα. Κατά τη διαδικασία της βαθμονόμησης, προσδιορίζονται οι παράμετροι $\mathbf{K}, \mathbf{R}, \mathbf{T}$ με τις εξωτερικές παραμέτρους να ισχύουν για την συγκεκριμένη θέση της κάμερας εκείνη τη στιγμή. Εάν η κάμερα υποστεί μετατόπιση ή περιστροφή (ή και τα δύο) οι εξωτερικές παράμετροι αλλάζουν. Σε περίπτωση που οι μετατοπίσεις είναι άγνωστες, οι νέες εξωτερικές παράμετροι προσδιορίζονται με την εκτίμηση πόζας. Ταυτόχρονα, οι εσωτερικές παράμετροι παραμένουν ίδιες.

Είναι σαφές πως με τον όρο *πόζα* εννοείται το σύνολο $\{\mathbf{R}, \mathbf{T}\}$ των εξωτερικών παραμέτρων της κάμερας σε σχέση με το σταθερό σύστημα συντεταγμένων του χώρου.

Για να την εκτίμηση πόζας σε σχέση με επίπεδο χρησιμοποιούνται οι σχέσεις (3.9) και (3.10). Είναι λοιπόν απαραίτητος ο υπολογισμός του ομογραφήματος \mathbf{H} . Οι δύο αυτές σχέσεις ξαναγράφονται ως εξής

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \mathbf{H} \begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix} = \begin{pmatrix} h_1^T \\ h_2^T \\ h_3^T \end{pmatrix} \mathbf{P}_W = \begin{pmatrix} h_1^T \mathbf{P}_W \\ h_2^T \mathbf{P}_W \\ h_3^T \mathbf{P}_W \end{pmatrix} \quad (3.123)$$

$$s = h_3^T \mathbf{P}_W \quad (3.24)$$

$$\begin{aligned} u h_3^T \mathbf{P}_W - h_1^T \mathbf{P}_W &= 0 \\ v h_3^T \mathbf{P}_W - h_2^T \mathbf{P}_W &= 0 \end{aligned} \quad (3.25)$$

$$\begin{aligned} u (h_{31} X_w + h_{32} Y_w + h_{33}) - (h_{11} X_w + h_{12} Y_w + h_{13}) &= 0 \\ v (h_{31} X_w + h_{32} Y_w + h_{33}) - (h_{21} X_w + h_{22} Y_w + h_{23}) &= 0 \end{aligned} \quad (3.26)$$

Εφόσον ένα ομογράφημα έχει 8 βαθμούς ελευθερίας, χρειάζονται τουλάχιστον 8 εξισώσεις της μορφής (3.26) για να υπολογιστεί το ομογράφημα. Συνεπώς, πρέπει να ευρεθούν οι συντεταγμένες στο χώρο τουλάχιστον 4 σημείων του επιπέδου, καθώς και οι αντιστοιχίσεις τους στον χώρο των εικονοστοιχείων. Σε κάθε περίπτωση, προκύπτει ένα σύστημα εξισώσεων της μορφής

$$\mathbf{Lh} = \mathbf{0} \quad (3.27)$$

Γενικά συνίσταται η χρησιμοποίηση περισσότερων των 4 σημείων, για την εξάλειψη του θορύβου στις παρατηρήσεις των εικονοστοιχείων και την εύρεση ακριβέστερων αποτελεσμάτων. Έτσι, το σύστημα (3.27) γίνεται υπερπλήρες και μία λύση του, υποκείμενη στον περιορισμό $\|\mathbf{h}\| = 1$, είναι το ιδιοδιάνυσμα του πίνακα $\mathbf{L}^T \mathbf{L}$ που αντιστοιχεί στην

μικρότερη ιδιοτιμή. Έχοντας υπολογίσει το ομογράφημα H και με γνωστές τις εσωτερικές παραμέτρους K , υπολογίζεται η τελική πόζα της κάμερας σε σχέση με το επίπεδο.

3.4.2 Εξαγωγή συντεταγμένων 3Δ σημείου στο επίπεδο

Έχοντας γνωστά 4 (ή περισσότερα) 3Δ σημεία του δεδομένου επιπέδου και έχοντας βρει την πόζα της κάμερας όπως δείχτηκε στην προηγούμενη ενότητα, είναι δυνατό να τα μετασχηματίσουμε στο πλαίσιο της κάμερας σύμφωνα με την εξίσωση (3.7). Στη συνέχεια, υπολογίζεται η εξίσωση του επιπέδου, εκφρασμένη στο σύστημα συντεταγμένων της κάμερας

$$\varepsilon : aX_{pc} + bY_{pc} + cZ_{pc} + d = 1 \quad (3.28)$$

Για τον υπολογισμό των συντεταγμένων ενός 3Δ σημείου του αντικειμένου που ακουμπά στο επίπεδο (άρα είναι και ταυτόχρονα σημείο του επιπέδου), υπολογίζεται η εξίσωση της ακτίνας από το κέντρο της κάμερας στην προβολή του σημείου στο επίπεδο της εικόνας, όπως φαίνεται στην εικόνα 3.1. Το σημείο που προβάλλεται πάνω στην εικόνα έχει συντεταγμένες (X_c, Y_c, f) , ως προς την κάμερα. Η εξίσωση της ακτίνας, εκφρασμένη στο πλαίσιο της κάμερας έχει την παραμετρική μορφή

$$\begin{aligned} x &= At + x_0 \\ y &= Bt + y_0 \\ z &= Ct + z_0 \end{aligned} \quad (3.29)$$

Εφ' όσον η ακτίνα ξεκινά από το σημείο $(0,0,0)$ της κάμερας, ισχύει $x_0 = y_0 = z_0 = 0$. Οι μεταβλητές A, B, C έχουν τιμές $X_c - 0, Y_c - 0, Z_c - 0$ αντίστοιχα. Αντικαθιστώντας τα παραπάνω στην (3.28) βρίσκεται η τιμή για το t

$$t = \frac{-d}{aX_c + bY_c + cZ_c} \quad (3.30)$$

Η αντικατάσταση αυτής της τιμής στην (3.29) δίνει τις συντεταγμένες, στο πλαίσιο της κάμερας, του σημείου του αντικειμένου. Η έκφρασή του σε συντεταγμένες του χώρου, μπορεί να γίνει από την σχέση (3.7), υπολογίζοντας τον αντίστροφο του πίνακα.

3.4.3 Εφαρμογή

Στην παρούσα διπλωματική, εκτελέστηκε βαθμονόμηση της κάμερας του ρομπότ ΝΑΟ για την λήψη οπτικών δεδομένων. Σύμφωνα με την παραπάνω ανάλυση, χρησιμοποιήθηκαν 20 λήψεις μίας επίπεδης σκακιέρας 6x8, από διαφορετική γωνία λήψης. Οι εικόνες της σκακιέρας δεν είναι ομοεπίπεδες.



Εικόνα 12: Διαδοχικές λήψεις αντικειμένου βαθμονόμησης από διαφορετικές οπτικές γωνίες

Τα αποτελέσματα της βαθμονόμησης για τις εσωτερικές παραμέτρους της κάμερας του NAO είναι τα εξής

$$K = \begin{pmatrix} 378.64049 & 0 & 159.5 \\ 0 & 380.07872 & 119.5 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.31)$$

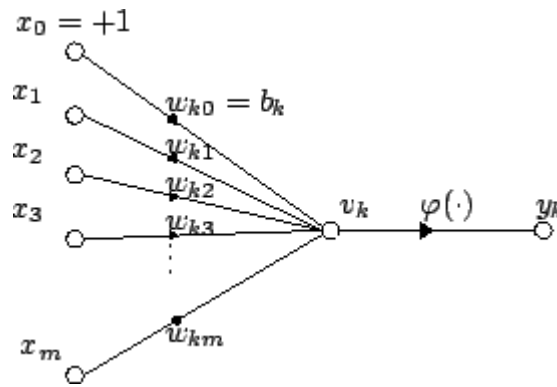
ΚΕΦΑΛΑΙΟ 4

4 Νευρωνικά δίκτυα και μηχανές διανυσμάτων υποστήριξης (SVMs)

4.1 Εισαγωγή

Η ανάγκη για κατανόηση της λειτουργίας του ανθρώπινου εγκεφάλου, οδήγησε στην ανάπτυξη προγραμμάτων και μηχανών, ικανές να προσομοιώσουν τη λειτουργία του σε μεγάλο βαθμό. Ο ανθρώπινος εγκέφαλος λειτουργεί με τη βοήθεια νευρώνων και συνάψεων μεταξύ τους. Οι συνάψεις είναι ουσιαστικά λειτουργικές μονάδες που παίζουν διαμεσολαβητικό ρόλο κατά τις αλληλεπιδράσεις μεταξύ των νευρώνων. Οι έξοδοι των νευρώνων είναι σήματα παλμών τάσης σύντομης διάρκειας, τα οποία μεταδίδονται μέσω των αξόνων των νευρώνων. Οι νευρώνες είναι σχετικά αργοί, με τα συμβάντα να λαμβάνουν χώρα στην κλίμακα του χιλιοστού του δευτερολέπτου. Η ιδιότητα αυτή αντισταθμίζεται με το τεράστιο πλήθος των νευρώνων και των συνάψεων (περίπου 10 τρισεκατομμύρια και 60 τρισεκατομμύρια αντίστοιχα). Με τη βοήθεια των νευρώνων και των συνάψεων, ο άνθρωπος είναι σε θέση να αντιλαμβάνεται το περιβάλλον του, να δέχεται και να επεξεργάζεται ερεθίσματα, να αντιδρά σε αυτά, να μαθαίνει και να διαμορφώνει συμπεριφορές.

Ήταν λοιπόν φυσικό να αναζητηθεί ένας τρόπος μοντελοποίησης της λειτουργίας του εγκεφάλου, για την ανάπτυξη λογισμικού που θα προσομοιώνει όσο το δυνατόν καλύτερα την λειτουργία του. Μια βασική μαθηματική μοντελοποίηση του νευρώνα δόθηκε από τους McCulloch και Pitts. Σύμφωνα με αυτό η έξοδος του νευρώνα είναι το αποτέλεσμα μίας συνάρτησης ενεργοποίησης $\varphi(u)$. Η τιμή αυτής της συνάρτησης ενεργοποίησης εξαρτάται από τις τιμές εισόδου x_i , σταθμισμένες με βάρη w_i . Η είσοδος x_0 είναι σταθερή και ίση με +1. Το βάρος της w_0 συμβολίζεται με b και ονομάζεται *πόλωση*. Μία σχηματική απεικόνιση του μοντέλου του νευρώνα k απεικονίζεται στην εικόνα 5.



Εικόνα 13: Μοντέλο νευρώνα McCulloch - Pitts

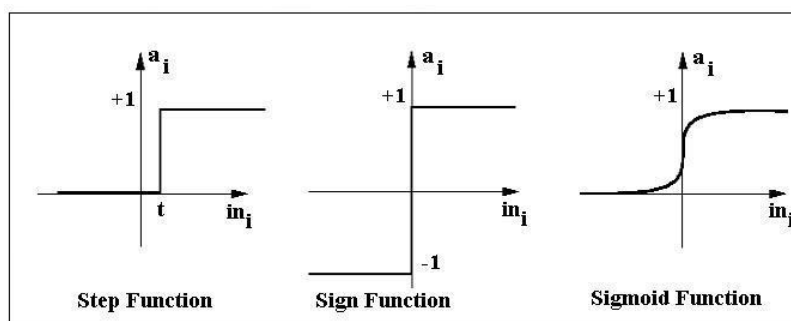
Μαθηματικά, το μοντέλο του νευρώνα k εκφράζεται ως εξής

$$u_k = \sum_{j=1}^m w_{kj} x_j + b \tag{4.1}$$

$$y_k = \varphi(u_k + b_k)$$

με x_m τα σήματα εισόδου, w_{km} τα βάρη των σημάτων εισόδου, φ τη συνάρτηση ενεργοποίησης, u την έξοδο του συνδυαστή των εισόδων, b την πόλωση και y την έξοδο του νευρώνα.

Η συνάρτηση ενεργοποίησης μπορεί να έχει αρκετές μορφές. Συνηθέστερες είναι η συνάρτηση κατωφλίου, η σιγμοειδής συνάρτηση και η συνάρτηση προσήμου.



Εικόνα 14: Συναρτήσεις ενεργοποίησης για το μοντέλο νευρώνα McCulloch - Pitts

Ένα νευρωνικό δίκτυο αποτελείται από ένα σύνολο νευρώνων τα οποία δέχονται εισόδους, και έχουν ως έξοδο το αποτέλεσμα της συνάρτησης ενεργοποίησης. Κύριο χαρακτηριστικό του νευρωνικού δικτύου, είναι η ικανότητά του να εκπαιδεύεται και να μαθαίνει, ανάλογα με τα χαρακτηριστικά που τροφοδοτούνται για εκπαίδευση. Ουσιαστικά, μέσω επαναληπτικών διαδικασιών, το νευρωνικό δίκτυο είναι σε θέση να προσαρμόζει τα βάρη του. Έτσι, μέσω αυτής της διαδικασίας είναι δυνατός ο διαχωρισμός ερεθισμάτων, παρόμοια με τη λειτουργία του ανθρώπινου μυαλού.

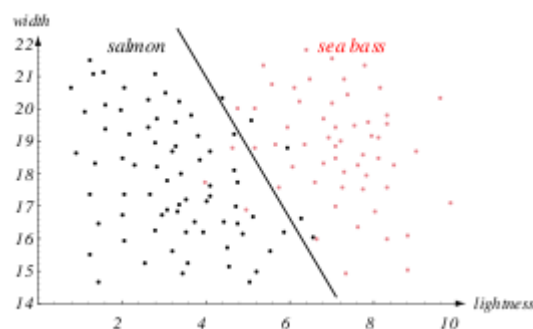
Στο παρόν κεφάλαιο αναλύεται η γενικότερη διαδικασία εκπαίδευσης των νευρωνικών δικτύων. Δίνεται ιδιαίτερη έμφαση στις μηχανές διανυσμάτων υποστήριξης, ενός τύπου δικτύου με πολύ υψηλή ικανότητα εκμάθησης, ενώ παρουσιάζονται παραδείγματα των ιδιοτήτων τους.

4.2 Το πρόβλημα της ταξινόμησης σε κλάσεις

Το βασικότερο ίσως πρόβλημα στην μάθηση μηχανών είναι το πρόβλημα της ταξινόμησης (ή του διαχωρισμού) σε κλάσεις. Δοθέντος ενός αντικειμένου, το πρόβλημα υπάγεται στο να αναγνωριστούν κάποιες ιδιότητές ή χαρακτηριστικά του, έτσι ώστε να υπάρξει τελική απόφαση σχετικά με τη φύση του. Ο άνθρωπος μπορεί να επιτελέσει αυτήν τη διεργασία σχεδόν αυτόματα και ασυναίσθητα. Η βασική αρχή λειτουργίας όμως είναι η ίδια, τόσο στον άνθρωπο όσο και στην επιστήμη της μάθησης μηχανών. Όπως γίνεται αντιληπτό, χρειάζεται να δοθεί ιδιαίτερη προσοχή στον ορισμό της φύσης του αντικειμένου,

καθώς επίσης και στη μαθηματική μοντελοποίηση των χαρακτηριστικών που εξετάζονται για να παρθεί το τελικό συμπέρασμα. Για παράδειγμα, στο πρόβλημα του διαχωρισμού ψαριών, ο άνθρωπος εξετάζει χαρακτηριστικά όπως μέγεθος, χρώμα και υφή για να αποφανθεί για το είδος του. Όπως θα φανεί παρακάτω, με κατάλληλη μοντελοποίηση αυτών των χαρακτηριστικών, ένα ρομπότ μπορεί να εκτελέσει αυτήν την εργασία σε αρκετά ικανοποιητικό βαθμό.

Έστω το πρόβλημα της αναγνώρισης είδους ψαριών που αναφέρθηκε παραπάνω. Για ευκολία στην αναπαράσταση, θα θεωρηθεί πως εξετάζονται μόνο δύο χαρακτηριστικά, το μήκος και η λαμπρότητά του. Οι δύο κλάσεις ταξινόμησης είναι ‘λαβράκι’ ή ‘σολομός’. Μετρώντας τα χαρακτηριστικά των ψαριών προς ταξινόμηση σχηματίζεται το παρακάτω διάγραμμα, όπου αναπαρίστανται με μαύρο οι σολομοί και με κόκκινο τα λαβράκια.



Εικόνα 15: Ταξινόμηση ψαριών

Το πρόβλημα της ταξινόμησης έγκειται στην κατασκευή μίας ευθείας η οποία διαχωρίζει τις δύο κλάσεις. Κατά την εξέταση νέων ψαριών, εάν αυτά βρεθούν δεξιά από την ευθεία θα ταξινομηθούν ως λαβράκια, ενώ εάν βρεθούν αριστερά της ευθείας θα ταξινομηθούν ως σολομοί. Μαθηματικά, εάν ορίσουμε τον χώρο ή διάνυσμα χαρακτηριστικών ως $\mathbf{X} = [x_1 \ x_2]$, και τις κλάσεις ως C_1, C_2 αντίστοιχα, η εξίσωση της ευθείας δίνεται από μία απλουστευμένη μορφή της εξίσωσης (4.1), δηλαδή

$$w_1x_1 + w_2x_2 + b = 0 \quad (4.2)$$

Έχοντας ως γνωστά τα βάρη w_i , και την πόλωση b , εάν το δεξί μέλος της εξίσωσης είναι μεγαλύτερο του μηδενός, το ψάρι αναγνωρίζεται σαν λαβράκι, ενώ εάν είναι μικρότερο του μηδενός, το ψάρι αναγνωρίζεται σαν σολομός. Η περίπτωση που είναι ίσο με μηδέν μπορεί να επιλυθεί είτε επιλέγοντας τυχαία μία από τις δύο κλάσεις, είτε τοποθετώντας το ψάρι σε μία προκαθορισμένη κλάση.

Η διαδικασία κατά την οποία το νευρωνικό δίκτυο δίνει τιμές στα βάρη λέγεται *εκπαίδευση*. Για την εκπαίδευση, είναι απαραίτητο να δοθεί στο νευρωνικό δίκτυο ένα σύνολο με αντικείμενα για τα οποία είναι γνωστά η κλάση που ανήκουν. Το σύνολο αυτό λέγεται *σύνολο εκπαίδευσης*. Μετρώντας τα χαρακτηριστικά τους, το δίκτυο είναι σε θέση να δώσει κάποιες αρχικές εκτιμήσεις για τα βάρη. Οι εκτιμήσεις διορθώνονται και επανεκδίδονται για όλο το μέγεθος του συνόλου εκπαίδευσης. Στο τέλος του συνόλου εκπαίδευσης, έχει δημιουργηθεί η ευθεία ταξινόμησης, οπότε το δίκτυο είναι σε θέση να ταξινομή αντικείμενα που φτάνουν από ένα *σύνολο ταξινόμησης*. Εάν τα αποτελέσματα δεν είναι ικανοποιητικά, ίσως να χρειάζεται επαναπροσδιορισμός των βαρών, με διαφορετικό σύνολο εκπαίδευσης.

Όλη η διαδικασία που περιεγράφηκε, μπορεί να επεκταθεί σε περισσότερες από δύο διαστάσεις. Για το παράδειγμά μας, αυτό θα μπορούσε να επιτευχθεί εξετάζοντας εκτός από το μήκος και τη λαμπρότητα, το πλάτος των ψαριών, το μέγεθος των λεπιών τους, το χρώμα τους, το που αλιεύθηκαν κλπ. Για m μεταβλητές, ο χώρος χαρακτηριστικών γίνεται m -διάστατος, ενώ το δίκτυο έχει m βάρη προς προσδιορισμό. Έτσι, το δίκτυο δεν χτίζει ευθείες, αλλά υπερεπίπεδα σε διανυσματικούς χώρους ανώτερης τάξης.

4.3 Βέλτιστη ταξινόμηση με SVMs

Τα νευρωνικά δίκτυα που περιεγράφηκαν στην προηγούμενη ενότητα έχουν ένα μειονέκτημα: δεν μπορούν να εγγυηθούν την βέλτιστη ευθεία διαχωρισμού των κλάσεων. Έτσι, υπάρχει πιθανότητα η ευθεία να τοποθετηθεί κατά τέτοιο τρόπο που να βρίσκεται πολύ κοντά σε μία κλάση (ή και στις δύο), ταξινομώντας λάθος ένα διάνυσμα που θα βρεθεί ακριβώς δίπλα από την ευθεία. Για να επιλυθεί αυτό το ζήτημα, χρησιμοποιείται μία βελτιωμένη εκδοχή του απλού νευρωνικού δικτύου, η *Μηχανή Διανυσμάτων Υποστήριξης* (*Support Vector Machine, SVM*). Όπως θα φανεί παρακάτω, ένα SVM είναι σε θέση να βρίσκει το βέλτιστο υπερεπίπεδο διαχωρισμού των κλάσεων.

4.3.1 Διανύσματα υποστήριξης και βέλτιστο υπερεπίπεδο

Έστω το δείγμα εκπαίδευσης $\{\mathbf{x}_i, d_i\}$ με $i = 1 \dots N$ και \mathbf{x}_i, d_i το διάνυσμα εισόδου και την επιθυμητή απόκριση αντίστοιχα. Οι τιμές της απόκρισης είναι +1 ή -1, ανάλογα με την κλάση στην οποία ανήκει το δείγμα εισόδου. Η εξίσωση του υπερεπιπέδου απόφασης είναι η (4.1) η οποία για ευκολία γράφεται σε μητρική μορφή

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (4.3)$$

Για τις δεδομένες τιμές αποκρίσεων, η απόφαση γίνεται με βάση τον κανόνα

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + b &\geq 0 && \text{για } d_i = +1 \\ \mathbf{w}^T \mathbf{x}_i + b &< 0 && \text{για } d_i = -1 \end{aligned} \quad (4.4)$$

Για ένα δεδομένο διάνυσμα βαρών \mathbf{w} και δεδομένη πόλωση b , ο διαχωρισμός μεταξύ του υπερεπιπέδου (4.3) και του πλησιέστερου σημείου δεδομένων εισόδου λέγεται

περιθώριο διαχωρισμού και συμβολίζεται με ρ . Ένα SVM έχει ως σκοπό την εύρεση του υπερεπιπέδου που μεγιστοποιεί το περιθώριο ρ για όλες τις κλάσεις. Η υπερεπιφάνεια αυτή ονομάζεται βέλτιστο υπερεπίπεδο.

Εάν συμβολιστούν τα βέλτιστα βάρη και η βέλτιστη πόλωση με \mathbf{w}_o, b_o αντίστοιχα, τότε το βέλτιστο υπερεπίπεδο δίνεται από τη σχέση

$$\mathbf{w}_o^T \mathbf{x} + b_o = 0 \quad (4.5)$$

Για ευκολία, μπορεί να οριστεί η συνάρτηση

$$g(\mathbf{x}) = \mathbf{w}_o^T \mathbf{x} + b_o \quad (4.6)$$

η οποία δίνει το αλγεβρικό μέτρο της απόστασης από του \mathbf{x} από το βέλτιστο υπερεπίπεδο. Αυτό γίνεται αντιληπτό εάν γραφεί το \mathbf{x} ως

$$\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}_o}{\|\mathbf{w}_o\|} \quad (4.7)$$

όπου \mathbf{x}_p Η κανονική προβολή του \mathbf{x} στο βέλτιστο υπερεπίπεδο και r η επιθυμητή αλγεβρική απόσταση. Εφ' όσον εξ' ορισμού ισχύει $g(\mathbf{x}_p) = 0$, έπεται ότι

$$g(\mathbf{x}) = \mathbf{w}_o^T \mathbf{x} + b_o = r \|\mathbf{w}_o\| \quad (4.8)$$

ή ισοδύναμα

$$r = \frac{g(\mathbf{x})}{\|\mathbf{w}_o\|} \quad (4.9)$$

Ειδικότερα, για το σημείο αρχής $\mathbf{x} = \mathbf{0}$ η απόστασή του από το βέλτιστο υπερεπίπεδο είναι $\frac{b_o}{\|\mathbf{w}_o\|}$. Από την τιμή του b_o μπορεί να βρεθεί η τοποθέτηση του υπερεπιπέδου σε σχέση με την αρχή των αξόνων, δηλαδή εάν αυτό είναι από τη θετική πλευρά, την αρνητική πλευρά η εάν περνά από το σημείο αρχής.

Τα διανύσματα για τα οποία ισχύει $g(\mathbf{x}) = \pm 1$, για $d = \pm 1$ αντίστοιχα, ονομάζονται *διανύσματα υποστήριξης* (*support vectors*). Στην ουσία, τα διανύσματα υποστήριξης είναι τα σημεία που βρίσκονται πιο κοντά στο βέλτιστο υπερεπίπεδο και, κατ' επέκταση είναι τα πιο δύσκολα να ταξινομηθούν.

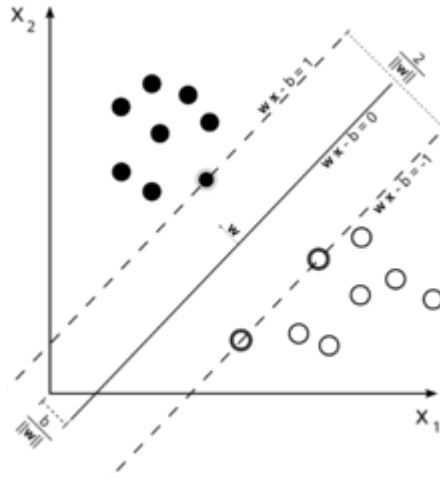
Η απόσταση ενός διανύσματος υποστήριξης \mathbf{x}_s από το βέλτιστο υπερεπίπεδο είναι

$$r = \frac{g(\mathbf{x}_s)}{\|\mathbf{w}_o\|} = \pm \frac{1}{\|\mathbf{w}_o\|} \quad \text{για } d_s = \pm 1 \text{ αντίστοιχα} \quad (4.10)$$

Έτσι, το περιθώριο διαχωρισμού του βέλτιστου υπερεπιπέδου, δίνεται από τον τύπο

$$\rho = 2r = \frac{2}{\|\mathbf{w}_o\|} \quad (4.11)$$

Το αποτέλεσμα αυτό υποδηλώνει πως, η μεγιστοποίηση του περιθωρίου διαχωρισμού μεταξύ δυαδικών κλάσεων απαιτεί την ελαχιστοποίηση της ευκλείδειας νόρμας του διανύσματος βαρών του βέλτιστου υπερεπιπέδου. Το βέλτιστο υπερεπίπεδο παρέχει το μέγιστο περιθώριο διαχωρισμού μεταξύ των διανυσμάτων των δύο κλάσεων. Τα αποτελέσματα της παραπάνω ανάλυσης συνοψίζονται στην παρακάτω εικόνα



Εικόνα 16: Διανύσματα υποστήριξης, βέλτιστο υπερεπίπεδο και μέγιστο περιθώριο διαχωρισμού

4.3.2 Εύρεση του βέλτιστου υπερεπιπέδου

Για την εύρεση του βέλτιστου υπερεπιπέδου, είναι χρήσιμο να επαναδιατυπωθούν οι συνθήκες του προβλήματος.

Δοθέντος ενός δείγματος εκπαίδευσης $[(\mathbf{x}_i, d_i)]_{i=1}^N$, οι βέλτιστες τιμές του διανύσματος βαρών \mathbf{w} και της πόλωσης b ικανοποιούν τους περιορισμούς

$$d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad \text{για } i = 1, 2, \dots, N \quad (4.12)$$

Χρήσιμο είναι επίσης να οριστεί η ακόλουθη συνάρτηση κόστους, την οποία θα ελαχιστοποιεί το διάνυσμα βαρών

$$\Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} \quad (4.13)$$

Το πρόβλημα (4.12) – (4.13) είναι πρόβλημα βελτιστοποίησης. Εφ' όσον οι περιορισμοί (4.12) είναι γραμμικοί ως προς το \mathbf{w} , και η $\Phi(\mathbf{w})$ είναι μία κυρτή συνάρτηση του \mathbf{w} , το πρόβλημα μπορεί να λυθεί με τη βοήθεια των πολλαπλασιαστών Lagrange. Κατασκευάζεται λοιπόν η συνάρτηση Lagrange

$$J(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N a_i [d_i (\mathbf{w}^T \mathbf{x}_i + b) - 1] \quad (4.14)$$

Οι μεταβλητές a_i είναι μη αρνητικές και ονομάζονται *πολλαπλασιαστές Lagrange*. Η συνάρτηση Lagrange ικανοποιεί τις παρακάτω συνθήκες βελτιστότητας

$$\begin{aligned} \frac{\partial J(\mathbf{w}, b, \mathbf{a})}{\partial \mathbf{w}} &= \mathbf{0} \\ \frac{\partial J(\mathbf{w}, b, \mathbf{a})}{\partial b} &= 0 \end{aligned} \quad (4.15)$$

οι οποίες εάν εφαρμοστούν στην συνάρτηση δίνουν τις σχέσεις

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^N a_i d_i \mathbf{x}_i \\ \sum_{i=1}^N a_i d_i &= 0 \end{aligned} \quad (4.16)$$

Παρ' όλο που η λύση είναι μοναδική λόγω της κυρτότητας της συνάρτησης Lagrange, δεν μπορεί να ειπωθεί το ίδιο για τους πολλαπλασιαστές. Για την εύρεσή τους,

χρειάζεται να κατασκευαστεί το δυικό πρόβλημα (dual) του αρχικού. Το δυικό πρόβλημα έχει ίδια βέλτιστη τιμή με το αρχικό, αλλά οι πολλαπλασιαστές Lagrange παρέχουν την βέλτιστη λύση.

Για την διατύπωση του δυικού προβλήματος, αρχικά επεκτείνεται η συνάρτηση Lagrange

$$J(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N a_i d_i \mathbf{w}^T \mathbf{x}_i - b \sum_{i=1}^N a_i d_i + \sum_{i=1}^N a_i \quad (4.17)$$

Λόγω της 2^{ης} συνθήκης (4.16), ο 3^{ος} όρος στην δεξιά μεριά μηδενίζεται. Από την 1^η συνθήκη (4.16), μπορεί να γραφεί

$$\frac{1}{2} \mathbf{w}^T \mathbf{w} = \frac{1}{2} \sum_{i=1}^N a_i d_i \mathbf{w}^T \mathbf{x}_i = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j \quad (4.18)$$

Αντικαθιστώντας στην $J(\mathbf{w}, b, \mathbf{a}) = Q(\mathbf{a})$, η τελική συνάρτηση Lagrange γράφεται

$$Q(\mathbf{a}) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j \quad (4.19)$$

Το δυικό πρόβλημα, έγκειται στην μεγιστοποίηση της συνάρτησης (4.19), υπό τους περιορισμούς

$$\sum_{i=1}^N a_i d_i = 0 \quad (4.20)$$

$$a_i \geq 0$$

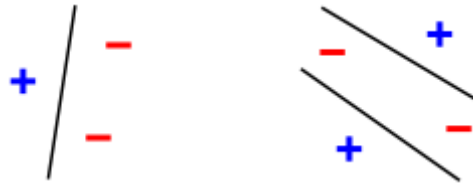
Το πλεονέκτημα του δυικού προβλήματος είναι πως είναι μετασχηματισμένο κατά τέτοιο τρόπο ώστε να βασίζεται στα δεδομένα εκπαίδευσης. Έχοντας λύσει το δυικό πρόβλημα, οι βέλτιστοι συντελεστές $\alpha_{o,i}$ που βρίσκονται, αντικαθιστούνται στην (4.16) και έτσι υπολογίζεται το διάνυσμα βαρών w_0 του βέλτιστου υπερεπιπέδου. Για την βέλτιστη πόλωση, αντικαθίσταται το διάνυσμα βαρών στην εξίσωση επιπέδου ενός διανύσματος υποστήριξης.

4.4 Μη γραμμικές μηχανές διανυσμάτων υποστήριξης

Οι μηχανές που περιεγράφηκαν στην προηγούμενη ενότητα, είναι βέλτιστες στην περίπτωση γραμμικά διαχωρίσιμων δεδομένων εισόδου. Σε αυτήν την ενότητα αναλύεται η έννοια της γραμμικής διαχωρισιμότητας και επεξηγείται η χρήση μη γραμμικών SVMs για αυτήν την περίπτωση.

4.4.1 Γραμμική διαχωρισιμότητα προτύπων

Έστω δύο σύνολα σημείων (πρότυπα) X_0, X_1 σε έναν n-διάστατο Ευκλείδειο χώρο. Τα δύο σύνολα είναι γραμμικά διαχωρίσιμα, εάν υπάρχουν n+1 πραγματικοί αριθμοί w_1, w_2, \dots, w_n, k τέτοιοι ώστε να ισχύει $\sum_{i=1}^N w_i x_i > k$ για κάθε $x \in X_0$ και $\sum_{i=1}^N w_i x_i < k$ για κάθε $x \in X_1$. Γίνεται εύκολα αντιληπτό πως αυτός ο μαθηματικός ορισμός υποδηλώνει την ύπαρξη ευθείας που διαχωρίζει τα δύο σύνολα, εξ' ου και ο όρος γραμμική διαχωρισιμότητα. Σύνολα προτύπων για τα οποία δεν ισχύει η παραπάνω ιδιότητα, ονομάζονται μη γραμμικά διαχωρίσιμα. Στην εικόνα φαίνονται παραδείγματα γραμμικά και μη γραμμικά διαχωρίσιμων συνόλων.

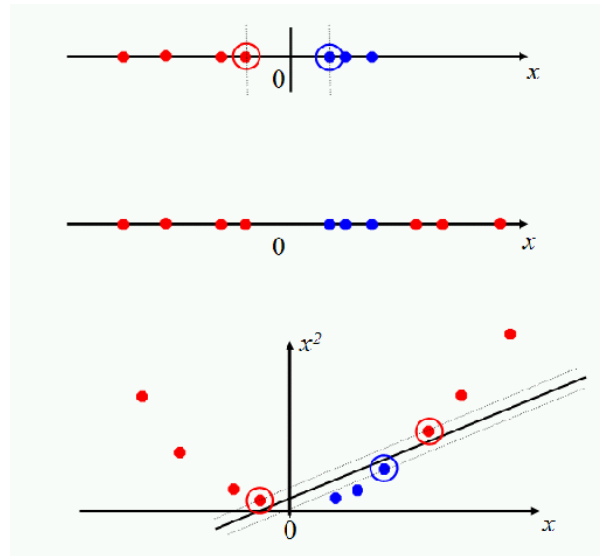


Εικόνα 17: Γραμμικά (αριστερά) και μη γραμμικά (δεξιά) διαχωρίσιμα σύνολα

Η γραμμική διαχωρισιμότητα είναι δύσκολο να προσδιοριστεί, ειδικά σε δεδομένα μεγάλου πλήθους και διαστατικότητας. Συνηθέστερος τρόπος για μικρότερο πλήθος δεδομένων είναι η εκπαίδευση ενός απλού γραμμικού νευρωνικού δικτύου ή SVM σε όλα τα δεδομένα εισόδου. Εάν η έξοδος συμφωνεί επακριβώς με την είσοδο, τότε τα δεδομένα είναι γραμμικά διαχωρίσιμα.

4.4.2 Ταξινόμηση μη γραμμικά διαχωρίσιμων προτύπων

Στην περίπτωση μη γραμμικά διαχωρίσιμων προτύπων, συνηθέστερη μέθοδος αντιμετώπισης του προβλήματος είναι η αντιστοίχιση των δεδομένων εισόδου σε έναν χώρο μεγαλύτερης διαστατικότητας, στον οποίο τα νέα δεδομένα είναι γραμμικά διαχωρίσιμα και μπορούν να ταξινομηθούν με τους τρόπους που περιεγράφηκαν στις προηγούμενες ενότητες. Η διαδικασία αυτή φαίνεται στην παρακάτω εικόνα



Εικόνα 18: Προβολή των διανυσμάτων εισόδου σε χώρο ανώτερης διαστατικότητας

Όπως φαίνεται στην εικόνα, τα αρχικά δεδομένα, όταν πολλαπλασιάστηκαν με την συνάρτηση x^2 , αντιστοιχίστηκαν σε έναν ανώτερο χώρο, με αποτέλεσμα να είναι εφικτός ο διαχωρισμός τους γραφικά.

Γενικότερα, εάν θεωρηθεί το διάνυσμα εισόδου \mathbf{x} και το σύνολο μη γραμμικών συναρτήσεων $\{\varphi_j(\mathbf{x})\}_{j=1}^{\infty}$, το οποίο μετασχηματίζει το διάνυσμα εισόδου σε έναν χώρο άπειρης διαστατικότητας, μπορεί να οριστεί ένα υπερεπίπεδο απόφασης ως εξής

$$\sum_{j=1}^{\infty} \mathbf{w}_j \varphi_j(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) = 0 \quad (4.21)$$

όπου η πόλωση παραλείπεται για ευκολία στην γραφή. Εφ' όσον μπορεί να εκτελεστεί γραμμικός διαχωρισμός σε αυτόν τον χώρο, είναι χρήσιμο να υπολογιστούν τα βάρη του υπερεπιπέδου όπως στην γραμμική περίπτωση, δηλαδή

$$\mathbf{w} = \sum_{i=1}^{N_s} a_i d_i \boldsymbol{\varphi}(\mathbf{x}_i) \quad (4.22)$$

με $\boldsymbol{\varphi}(\mathbf{x}_i) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_j(\mathbf{x})]^T$, το διάνυσμα χαρακτηριστικών στον καινούριο χώρο και N_s τον αριθμό των διανυσμάτων υποστήριξης. Αντικαθιστώντας την (4.22) στην (4.21) ευρίσκεται το υπερεπίπεδο διαχωρισμού

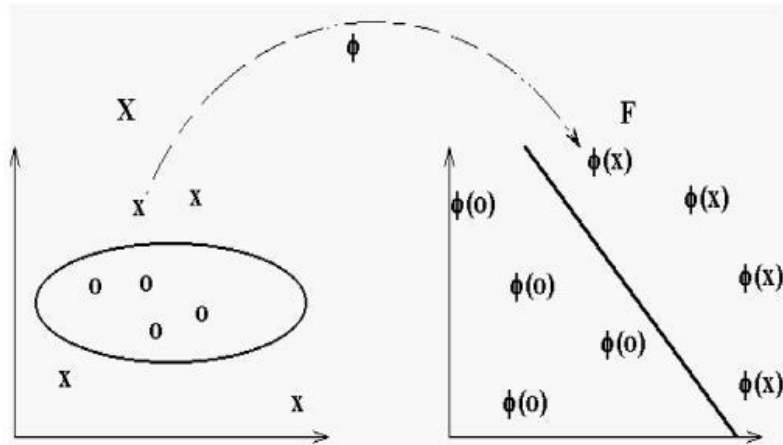
$$\sum_{i=1}^{N_s} a_i d_i \boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}) = 0 \quad (4.23)$$

Ο όρος εσωτερικού γινομένου $\boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x})$ συμβολίζεται με την *συνάρτηση πυρήνα*

$$k(\mathbf{x}, \mathbf{x}_i) = \boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}) \quad (4.24)$$

και η επιφάνεια απόφασης ξαναγράφεται

$$\sum_{i=1}^{N_s} a_i d_i k(\mathbf{x}, \mathbf{x}_i) = 0 \quad (4.25)$$



Εικόνα 19: Απεικόνιση των διανυσμάτων εισόδου στον χώρο ανώτερης τάξης με τη συνάρτηση ϕ

Το πλεονέκτημα αυτής της μεθόδου, είναι πως δεν είναι απαραίτητος ο υπολογισμός της προβολής κάθε αρχικού διανύσματος εισόδου στον νέο χώρο και στη συνέχεια του εσωτερικού γινομένου. Αντίθετα, χρησιμοποιώντας μία συνάρτηση πυρήνα, ο υπολογισμός του εσωτερικού γινομένου στον ανώτερο χώρο, μπορεί να γίνει υπολογίζοντας την τιμή της συνάρτησης στον αρχικό χώρο. Αυτή η ιδιότητα είναι πολύ χρήσιμη στην επεξεργασία μεγάλου πλήθους δεδομένων εισόδου, καθώς είναι αποδοτικότερη από πλευράς υπολογιστικού χρόνου και κόστους.

Ως συναρτήσεις πυρήνα, μπορούν να χρησιμοποιηθούν οι συναρτήσεις που ικανοποιούν το *θεώρημα Mercer*. Το *θεώρημα Mercer* αναφέρει πως μία συνάρτηση $(\mathbf{x}, \mathbf{x}')$, ορισμένη στο διάστημα $\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$, μπορεί να αναπτυχθεί στη σειρά

$$k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{\infty} \lambda_i \varphi_i(\mathbf{x}) \varphi_i(\mathbf{x}') \quad (4.26)$$

με θετικούς συντελεστές λ_i και να συγκλίνει, εάν η συνθήκη

$$\int_a^b \int_a^b k(\mathbf{x}, \mathbf{x}') \psi(\mathbf{x}) \psi(\mathbf{x}') d\mathbf{x} d\mathbf{x}' \geq 0 \quad (4.27)$$

ισχύει για όλες τις $\psi(\mathbf{x})$ που ικανοποιούν το κριτήριο

$$\int_a^b \psi^2(\mathbf{x}) d\mathbf{x} < \infty \quad (4.28)$$

Στην ουσία, το θεώρημα Mercer, υποδεικνύει πότε μία συνάρτηση είναι ικανή να χρησιμοποιηθεί σαν πυρήνας και ταυτόχρονα παρέχει έναν φραγμό στις πιθανές συναρτήσεις πυρήνα που μπορούν να χρησιμοποιηθούν.

Σε εφαρμογές, συνηθέστεροι πυρήνες που χρησιμοποιούνται είναι οι *πολυωνυμικοί* και οι *πυρήνες ακτινικής βάσης (RBFs)*. Οι δύο αυτοί τύποι, παρουσιάζονται κατά σειρά παρακάτω

$$k_p(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^p \quad \text{με } p \text{ καθορισμένο} \quad (4.29)$$

$$k_{RBF}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}'\|^2\right)$$

ΚΕΦΑΛΑΙΟ 5

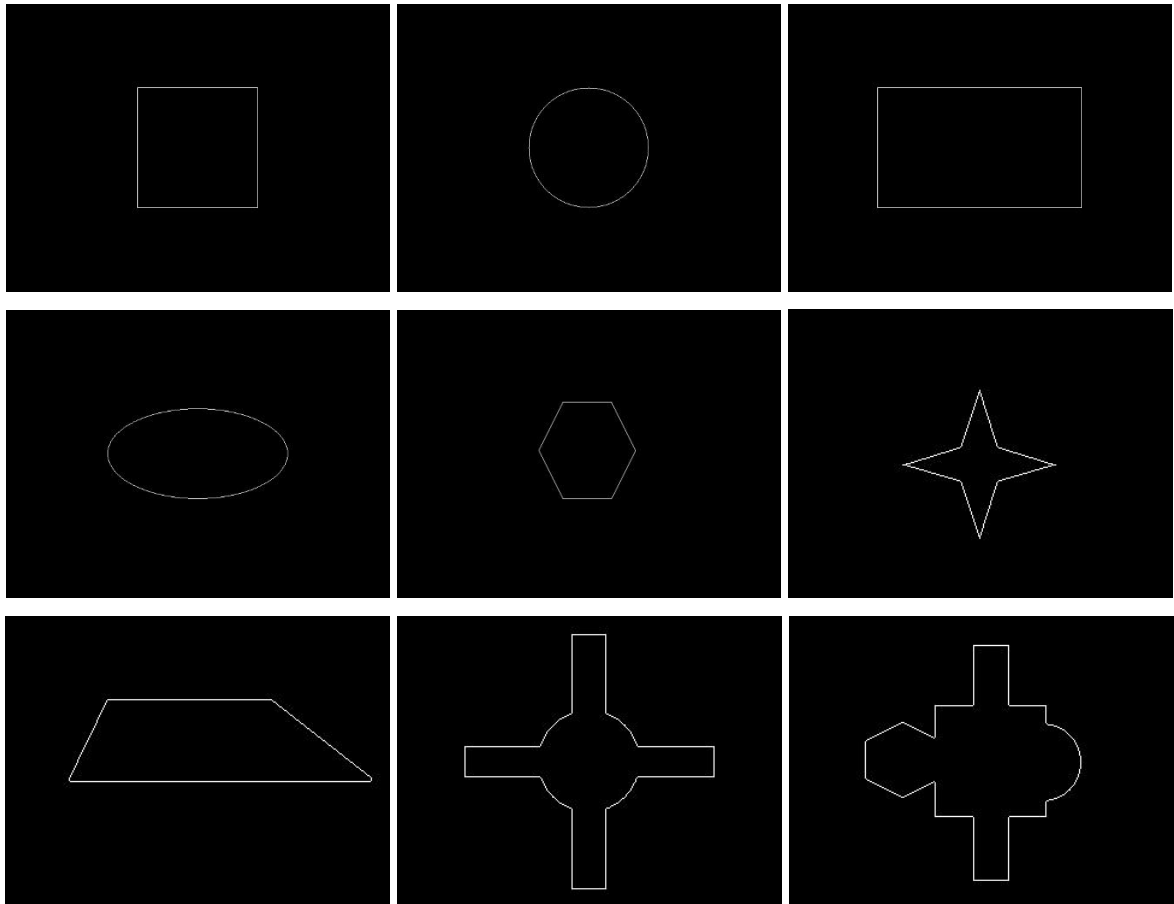
5 *Διαδικασία εκπαίδευσης ρομπότ και πειραματικά αποτελέσματα*

5.1 Εισαγωγή

Η ανάλυση που προηγήθηκε στα προηγούμενα κεφάλαια περιέγραψε την λειτουργία κάθε σταδίου ενός συστήματος αναγνώρισης αντικειμένων με βάση πιθανούς περιγραφείς που προέκυψαν ως αποτέλεσμα της επεξεργασίας εικόνων. Στο παρόν κεφάλαιο τα στάδια αυτά συντίθενται και δημιουργούν ένα πλήρες σύστημα αναγνώρισης, το οποίο χρησιμοποιήθηκε για την εξαγωγή σημείων από αντικείμενα. Τα σημεία αυτά θα χρησιμεύσουν ως σημεία τοποθέτησης του αντίχειρα του ρομπότ, για την εκτέλεση της λαβής και την επίλυση του προβλήματος.

Στο παρόν κεφάλαιο περιγράφεται διεξοδικά η πειραματική διαδικασία εκπαίδευσης ενός μη γραμμικού SVM για την αναγνώριση τοπικών μορφολογιών σε αντικείμενα. Κατά τη διαδικασία εκπαίδευσης, χρησιμοποιούνται σημεία από περιγράμματα διαφόρων σχημάτων.

Τα περιγράμματα περιλαμβάνουν από αντικείμενα μικρής και μεσαίας πολυπλοκότητας, μέχρι πολύπλοκα αντικείμενα ως σύνθεση απλούστερων. Για κάθε σχήμα ελέγχεται η ικανότητα ταξινόμησης με τη βοήθεια του ποσοστού των σημείων του περιγράμματος που ταξινομήθηκαν σωστά και του ποσοστού των σημείων που ταξινομήθηκαν λάθος, καθώς επίσης και τον αριθμό των λαβών που ανιχνεύθηκαν. Σε κάθε εκτέλεση εξετάζεται και η ικανότητα γενίκευσης του συστήματος, με την αυξομείωση των σημείων εκπαίδευσης και την εναλλαγή μεταξύ των σχημάτων που περιέχονται σε αυτό .



Εικόνα 20: Περιγράμματα αντικειμένων προς εκπαίδευση και αναγνώριση

5.2 Διάνυσμα εκπαίδευσης

Στην διαδικασία της εκπαίδευσης δίνονται στο δίκτυο πρότυπα εκπαίδευσης για να είναι σε θέση να δημιουργήσει τα βάρη του υπερεπιπέδου διαχωρισμού προτύπων. Συνεπώς, η διαδικασία εκπαίδευσης είναι το βασικότερο κομμάτι της αναγνώρισης με νευρωνικό δίκτυο. Στην βάση αυτού του κομματιού βρίσκεται η επιλογή ενός κατάλληλου περιγραφέα, ως πρότυπο τροφοδότησης του δικτύου. Η επιλογή του περιγραφέα αντικατοπτρίζει την ουσία του προβλήματος που πρέπει να επιλυθεί και του στόχου που πρέπει να επιτευχθεί.

Στην παρούσα διπλωματική ως διανύσματα εισόδου εκπαίδευσης, χρησιμοποιήθηκαν οι τοπικοί ακτινικοί περιγραφείς του εκάστοτε περιγράμματος εκπαίδευσης. Σύμφωνα με την ανάλυση της υποενότητας 2.5.3 υπολογίστηκαν οι τοπικοί

ακτινικοί περιγραφείς για τα δειγματολημμένα σημεία του κάθε περιγράμματος με συχνότητα δειγματοληψίας 10 σημεία και με γωνία σάρωσης θ μεταβαλλόμενη από -35° έως 35° . Η επιλογή για τη γωνία έγινε με γνώμονα τη διάταξη των δακτύλων του ρομπότ, τα οποία σχηματίζουν γωνία περίπου 60° ($+10^\circ$ περιθώριο για το πάχος των δακτύλων του).

Πολλά σημεία βάσης των περιγραμμάτων των σχημάτων εκπαίδευσης τυχαίνει να έχουν τον ίδιο, ή παρόμοιο τοπικό περιγραφέα, ανάλογα με τη θέση που βρίσκονται στο σχήμα. Για παράδειγμα, δύο σημεία του τετραγώνου που βρίσκονται κοντά στη μέση της πλευράς του έχουν τον ίδιο περιγραφέα, ενώ δύο σημεία που βρίσκονται κοντά σε μία γωνία έχουν παραπλήσιο περιγραφέα. Επιπλέον, οι περιγραφείς όλων των σημείων του κύκλου είναι σχεδόν ίδιοι, λόγω της συμμετρίας του κύκλου σε όλες τις κατευθύνσεις. Για το λόγο αυτό δεν χρησιμοποιήθηκαν όλα σημεία από το κάθε σχήμα στη εκπαίδευση. Το σύνολο εκπαίδευσης αλλάζει ανάλογα με τους στόχους που πρέπει να επιτευχθούν. Η εκλογή ενός σημείου ως ικανοποιητικό ή μη ικανοποιητικό για λαβή αναπαρίσταται από την τιμή 1 και 0 αντίστοιχα. Συνεπώς, οι κλάσεις για αναγνώριση είναι μόνο δύο: Ικανοποιητικό και μη ικανοποιητικό σημείο. Η ύπαρξη μόνο δύο κλάσεων διευκολύνει τον διαχωρισμό και την λειτουργία του SVM που υλοποιήθηκε ενώ προσφέρει αποδοτικότητα και ευκολία στον υπολογισμό του αποτελέσματος.

Η χρησιμοποίηση των πλευρών των δύο αυτών βασικών σχημάτων ως πρότυπα εισόδου, έχει πολύ μεγάλη σημασία όσον αφορά την ανίχνευση της μορφολογίας του περιγράμματος. Οι αλλαγές στην κλίση της συνάρτησης $r - \theta$ του περιγραφέα υποδεικνύουν αλλαγές στην ομαλότητα της πλευράς του αντικειμένου. Πράγματι, αντικείμενα με ομαλές, λείες πλευρές έχουν περιγραφέα με σταθερή κλίση, χωρίς έντονους κυματισμούς στην γραφική παράσταση της συνάρτησης. Είναι λοιπόν προφανής η ανάγκη τροφοδότησης στο ρομπότ με συναρτήσεις που αναπαριστούν λείες πλευρές, ως αρχική περίπτωση. Δεδομένου λοιπόν ενός αντικειμένου με πιθανές προεξοχές και ανάγλυφο περίγραμμα, το ρομπότ θα είναι σε θέση να ξεχωρίσει τις ομαλότερες περιοχές του περιγράμματός του, οι οποίες θα μοιάζουν όσο το δυνατόν περισσότερο σε πλευρά τετραγώνου ή σε τμήμα περιγράμματος κύκλου, παρέχοντας έτσι μία υποψήφια περιοχή για λαβή στο αντικείμενο. Τα σημεία με τέτοιους περιγραφείς τροφοδοτούνται στο ρομπότ με την τιμή κλάσης 1 (ικανοποιητικό σημείο). Με την ίδια λογική, η χρησιμοποίηση

διανυσμάτων που έχουν συνάρτηση με έντονες μεταβολές στην κλίση και κυματισμούς, δίνει μία μοντελοποίηση των σημείων που δεν είναι επαρκή για σχηματισμό λαβής.

Ως δεδομένα εκπαίδευσης μπορούν να χρησιμοποιηθούν περισσότερα σημεία από διαφορετικά σχήματα. Γενικότερα, η χρήση σημείων εκπαίδευσης έχει ως σκοπό την ανάπτυξη της ικανότητας του ρομπότ να συσχετίζει το εκάστοτε δοσμένο σημείο αναγνώρισης με κάποιο από αυτά που του έχουν δοθεί και να κρίνει εάν αυτό είναι ικανοποιητικό ή όχι ως προς τον σχηματισμό της λαβής. Η ποικιλομορφία των συναρτήσεων που δίνονται στο ρομπότ ως εκπαίδευση και κατ' επέκταση των τμημάτων περιγράμματος που αντιστοιχούν σε αυτές τις συναρτήσεις, βοηθά το ρομπότ να ξεχωρίζει πλήθος επιφανειών για σχηματισμό λαβής πάνω σε ένα μορφολογικά πολύπλοκο αντικείμενο, με αποτέλεσμα την αύξηση της ικανότητας αλληλεπίδρασης με αυτό.

5.3 Διαδικασία εκπαίδευσης και αναγνώρισης βασικών σχημάτων

Οι περιγραφείς που αναφέρθηκαν στην προηγούμενη ενότητα χρησιμοποιήθηκαν ως είσοδοι σε δίκτυο SVM για την δημιουργία του συστήματος αναγνώρισης περιγραμμάτων του ρομπότ. Μετά την εκπαίδευση, εκτελέστηκε και επαλήθευση στο σύνολο των σχημάτων της Εικόνας 21.

5.3.1 Διαδικασία εκπαίδευσης

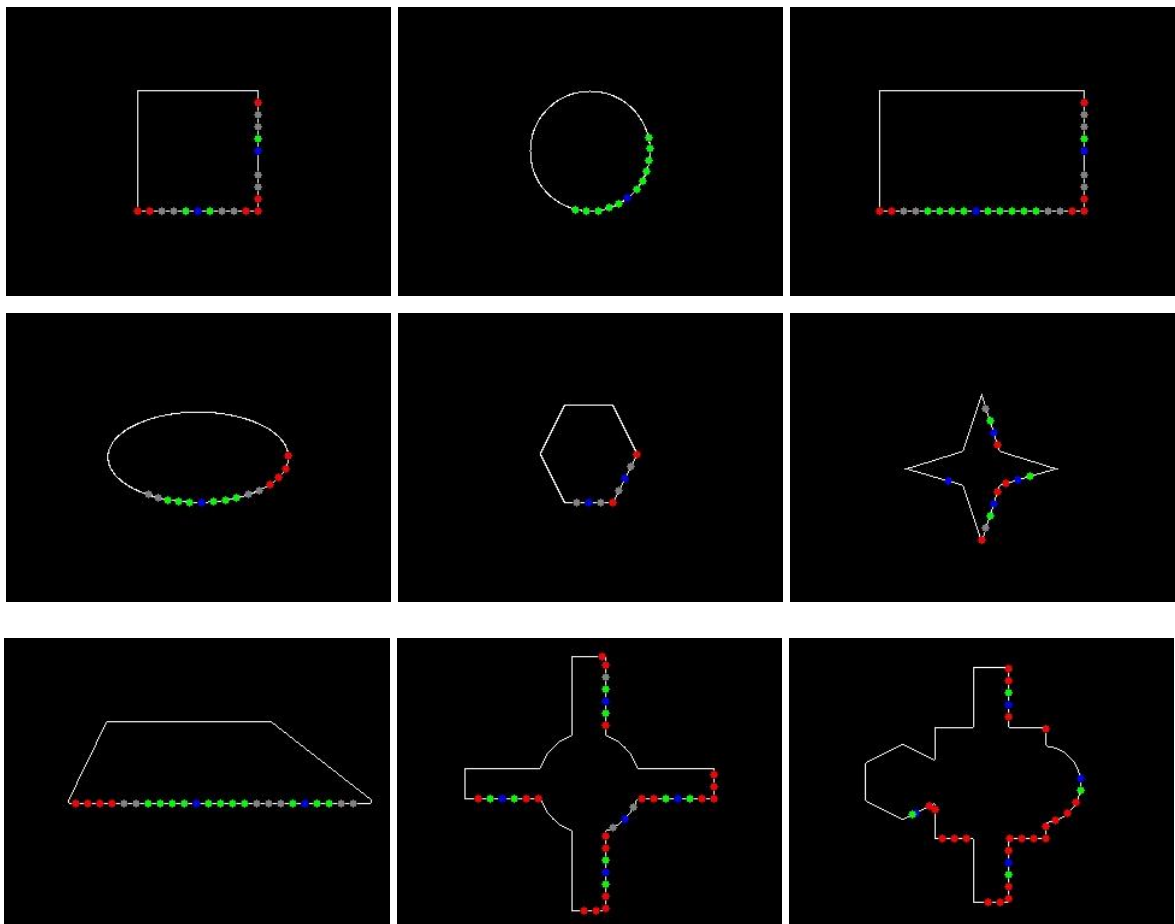
Για την ανάπτυξη του συστήματος αναγνώρισης, χρησιμοποιήθηκε νευρωνικό δίκτυο SVM. Η χρήση του SVM εγγυάται την εύρεση του βέλτιστου υπερεπίπεδου διαχωρισμού μεταξύ των δύο κλάσεων. Για να υπάρξουν ακόμα καλύτερα αποτελέσματα, αλλά και λόγω του ότι δεν είναι δυνατή η γνώση εκ των προτέρων της γραμμικής διαχωρισιμότητας των προτύπων εισόδου, χρησιμοποιήθηκε μη γραμμικό SVM με πολυωνυμικό πυρήνα. Το SVM προβάλλει τα διανύσματα εισόδου σε έναν γραμμικό χώρο, με χρήση πολυωνυμικής συνάρτησης πυρήνα και έτσι μπορεί να συνθέσει το υπερεπίπεδο σε αυτόν τον νέο χώρο, σύμφωνα με την ανάλυση του κεφαλαίου 4.

Κατά την εκπαίδευση δόθηκαν στο δίκτυο οι περιγραφείς των σημείων που αναφέρθηκαν στην προηγούμενη υποενότητα, οι οποίοι αναπαριστούν διάφορα σημεία του περιγράμματος του αντικειμένου και το κατά πόσο οι αντιδιαμετρικές πλευρές του περιγράμματος μπορούν να αποτελέσουν ικανοποιητικά τμήματα για τοποθέτηση του χεριού του ρομπότ και δημιουργία λαβής.

Για ευκολία στην δημιουργία λαβής, λόγω ελλειπούς ικανότητας του χεριού του ρομπότ στο να φτάσει σε συγκεκριμένες θέσεις του χώρου εργασίας του, από κάθε σχήμα εκπαίδευσης επιλέχθηκαν τα σημεία εκείνα του περιγράμματος που ο κάθετος άξονάς τους (όπως υπολογίστηκε στο Κεφάλαιο 2) σχηματίζει γωνία με τον οριζόντιο άξονα της εικόνας εντός του διαστήματος $[80^\circ - 190^\circ]$. Ουσιαστικά το ρομπότ ανιχνεύει σημεία στο κάτω μέρος του αντικειμένου και στις πλευρές του, που να είναι βολικές για να πιαστούν από το δεξί του χέρι. Η παραδοχή αυτή είναι λογική λόγω της φύσης του ρομποτικού βραχίονα του ΝΑΟ. Έχοντας 5 βαθμούς ελευθερίας, τα σημεία του περιγράμματος του πάνω μέρους του αντικειμένου θα οδηγούσαν τον βραχίονα σε σημείο ιδιομορφίας, καθιστώντας μη δυνατό τον σχηματισμό κατάλληλης λαβής.

Στα περιγράμματα των σχημάτων που δόθηκαν για εκπαίδευση και επαλήθευση, δόθηκαν τιμές ανάλογα με το κατά πόσο αυτά μπορούν να αποτελέσουν κατάλληλα σημεία για σχηματισμό λαβής. Η επιλογή των τιμών των σημείων έγινε με βάση την θέση τους στο περίγραμμα και το κατά πόσο αυτά βρίσκονται κοντά σε γωνίες ή σε σημεία με έντονη καμπυλότητα. Έτσι, τα σημεία κοντά στο κέντρο των αντικειμένων μακριά από γωνίες και έντονες καμπυλώσεις έχουν τιμή 1, ενώ τα σημεία κοντά στις προεξοχές και στις καμπυλώσεις έχουν τιμή 0. Μία οπτική απεικόνιση των δοσμένων τιμών δίνεται στην παρακάτω εικόνα, στην οποία τα σημεία με τιμή 1 αναπαρίστανται με πράσινο χρώμα, ενώ τα σημεία με τιμή 0 αναπαρίστανται με κόκκινο χρώμα. Τα σημεία τα οποία εμφανίζουν ιδιομορφίες που θα μπορούσαν να τα τοποθετήσουν και στις δύο κατηγορίες, καθώς επίσης και τα σημεία για τα οποία δεν ενδιαφέρει η τιμή τους, έχουν αναπαρασταθεί με γκρι χρώμα. Για κάθε περιοχή σημείων με τιμή 1, αναπαρίσταται το κεντρικό σημείο με μπλε χρώμα, δείχνοντας έτσι την τελικό σημείο λαβής που θα βρεθεί επάνω στο περίγραμμα, ως μεσαίο σημείο μιας γειτονιάς ικανοποιητικών σημείων. Τονίζεται ότι η επιλογή της τιμής των σημείων δεν είναι απόλυτη, ούτε υπόκειται σε περιορισμούς. Αυτό σημαίνει πως οι τιμές μπορούν να αλλάξουν και να επανατροφοδοτηθούν στο ρομπότ, ανάλογα με την

κρίση του εκπαιδευτή και το κατά πόσο αυτός κρίνει πως το εκάστοτε σημείο μπορεί να αποτελέσει ικανοποιητικό σημείο λαβής. Σε κάθε περίπτωση, οι τιμές αυτές των σημείων αντιμετωπίζονται σαν τιμές αναφοράς (*ground truth*) και χρησιμεύουν ως μέτρο σύγκρισης για τη διαδικασία επαλήθευσης.



Εικόνα 21: Τιμές αναφοράς για τα σχήματα εκπαίδευσης και επαλήθευσης (με πράσινο τα σημεία με τιμή 1 και με κόκκινο τα σημεία με τιμή 0)

5.3.2 Διαδικασία αναγνώρισης βασικών σχημάτων

Μετά την ολοκλήρωση της εκπαίδευσης του SVM με το σύνολο των περιγραμμάτων εκπαίδευσης δίνονται ως είσοδοι στο νευρωνικό δίκτυο όλα τα σχήματα είτε αυτά έχουν χρησιμοποιηθεί ως δεδομένα εκπαίδευσης είτε όχι. Σκοπός αυτής της προσέγγισης είναι να εξεταστεί κατά πόσο μπορεί το σύστημα να αναγνωρίσει τα σημεία

στα οποία έχει εκπαιδευθεί, αλλά και το να φανεί ο βαθμός ανταπόκρισης σε καινούρια περιγράμματα και η ικανότητα γενίκευσής του. Στην επόμενη ενότητα δείχνονται τα αποτελέσματα και τα μέτρα ποσοτικοποίησης αυτών.

5.4 Επαλήθευση αποτελεσμάτων

Για να γίνει εφικτή η μέτρηση της ικανότητας ταξινόμησης του νευρωνικού δικτύου SVM χρειάζονται κάποιες μετρικές ή κάποια μεγέθη που να ποσοτικοποιούν τα αποτελέσματα και να παρέχουν μία αξιολογητική σύγκριση. Οι μετρικές αυτές μπορεί να έχουν οποιαδήποτε μορφή. Κατά καιρούς χρησιμοποιούνται εικόνες, συναρτήσεις, αριθμητικές τιμές και ποσοστά ως μεγέθη σύγκρισης και μέτρησης της αποδοτικότητας της εκάστοτε μεθόδου μάθησης. Στην παρούσα διπλωματική, χρησιμοποιήθηκαν δύο μετρικές, το ποσοστό επιτυχίας Σωστό – Λάθος και ο αριθμός των λαβών που υπολογίστηκαν σωστά σε κάθε σχήμα. Το ποσοστό Σωστό - Λάθος υπολογίζεται ως ο λόγος του αριθμού των σημείων που ταξινομήθηκαν σωστά (Σωστό) προς τον συνολικό αριθμό σημείων του σχήματος αναφοράς. Το ποσοστό Λάθος δίνεται από τον τύπο

$$\text{Ποσοστό Λάθος} = 100\% - \text{Ποσοστό Σωστό} \quad (5.1)$$

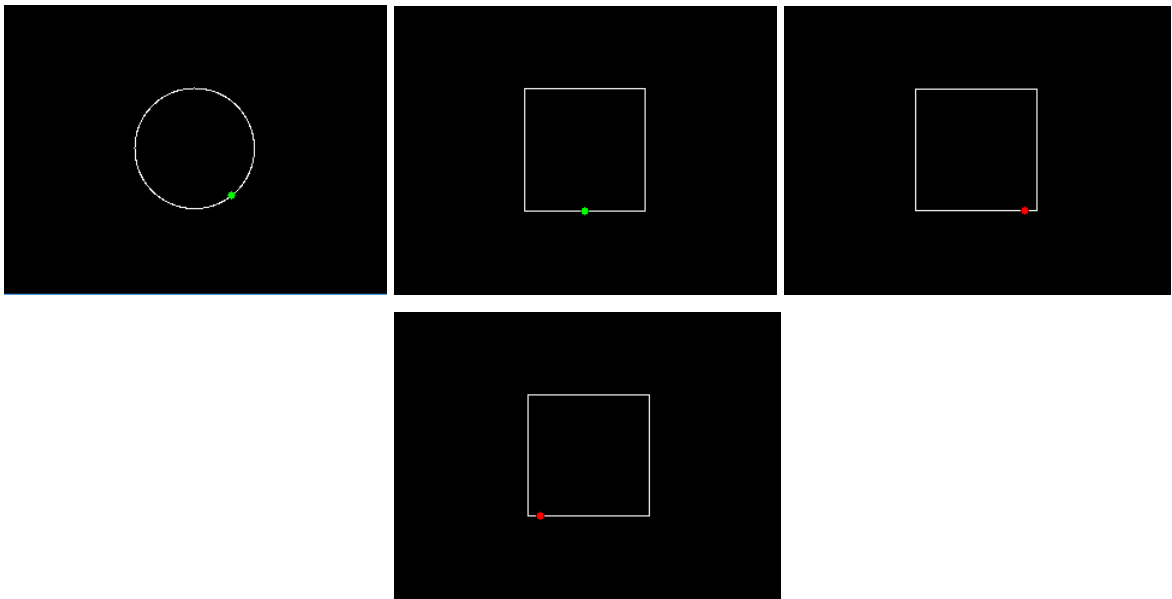
Ως σωστή ταξινόμηση για τον υπολογισμό του ποσοστού Σωστό ορίζεται η ταξινόμηση ενός σημείου σύμφωνα με τις τιμές αναφοράς, είτε το σημείο έχει τιμή 1 είτε τιμή 0. Αντίθετα, εάν ένα σημείο έχει τιμή διαφορετική από αυτή του σχήματος αναφοράς, συμπεριλαμβάνεται στον υπολογισμό του ποσοστού Λάθος.

Για κάθε σχήμα σημειώνεται ο αριθμός των τελικών λαβών που ανιχνεύθηκαν σε αυτό, ως ο αριθμός των κεντρικών σημείων σε κάθε γειτονιά σημείων τιμής 1. Ο αριθμός αυτός πρέπει να είναι ιδανικά ίδιος με τον αριθμό μπλε σημείων του σχήματος αναφοράς. Σε περίπτωση που μειωθεί, ο αλγόριθμος αποτυγχάνει να βρει μία λαβή για το εκάστοτε σχήμα.

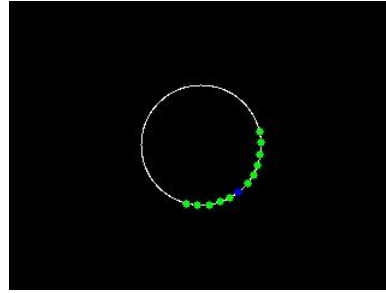
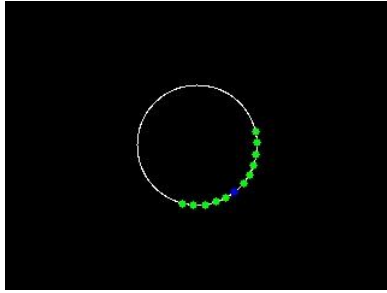
5.5 Πειραματικές εκτελέσεις

Σε αυτήν την υποενότητα εκτελέστηκε ο αλγόριθμος και εξετάστηκαν τα αποτελέσματα για κάθε σχήμα, χρησιμοποιώντας διαφορετικό σύνολο εκπαίδευσης για το SVM.

Σε πρώτη φάση δόθηκαν στο ρομπότ ως σύνολο εκπαίδευσης τα παρακάτω σημεία



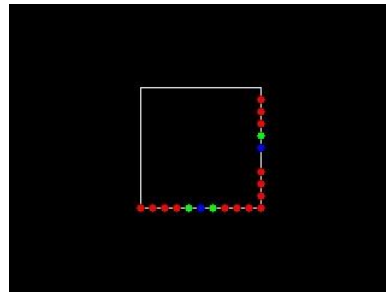
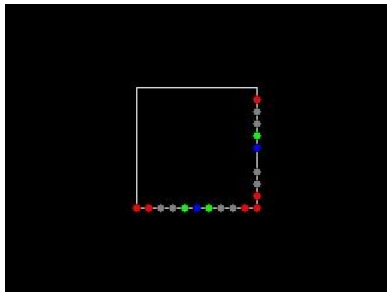
Όπως αναφέρθηκε, η επιλογή έγινε ως πρώτη προσέγγιση στο πρόβλημα της εκπαίδευσης, δίνοντας στο ρομπότ ένα σημείο του κύκλου για αναγνώριση κυκλικών επιφανειών, ένα σημείο με τιμή 1 στο κέντρο του τετραγώνου για αναγνώριση ευθειών και δύο σημεία με τιμή 2 στα άκρα του τετραγώνου, για αποφυγή αναγνώρισης σημείων κοντά σε γωνίες. Τα αποτελέσματα για κάθε σχήμα φαίνονται στις παρακάτω εικόνες. Η αριστερή είναι η εικόνα αναφοράς ενώ η δεξιά είναι η εικόνα που προέκυψε κατά την επαλήθευση



Ποσοστό Σωστό = 100 %

Ποσοστό Λάθος = 0 %

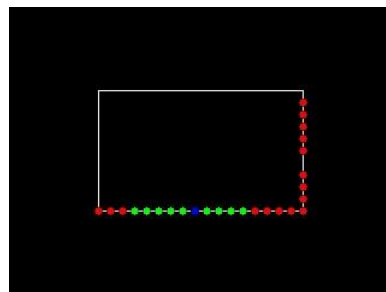
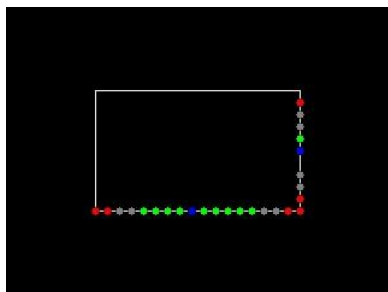
Αριθμός λαβών = 1/1



Ποσοστό Σωστό = 100 %

Ποσοστό Λάθος = 0 %

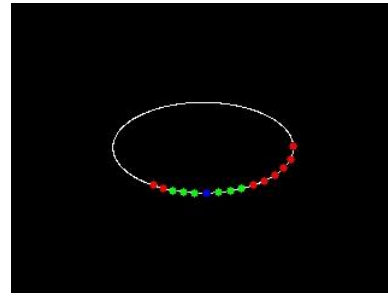
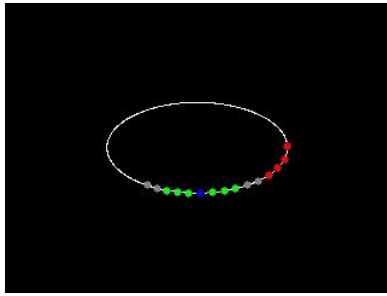
Αριθμός λαβών = 2/2



Ποσοστό Σωστό = 88.88%

Ποσοστό Λάθος = 11.12 %

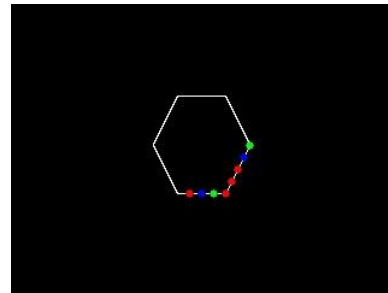
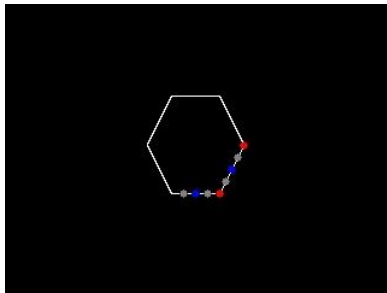
Αριθμός λαβών = 1/2



Ποσοστό Σωστό = 100 %

Ποσοστό Λάθος = 0 %

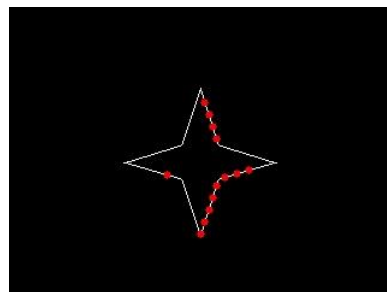
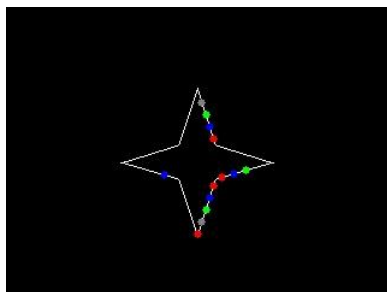
Αριθμός λαβών = 1/1



Ποσοστό Σωστό = 50 %

Ποσοστό Λάθος = 50 %

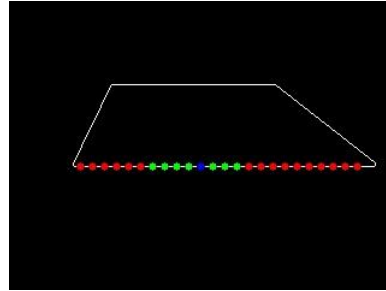
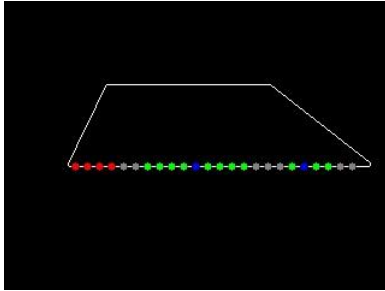
Αριθμός λαβών = 2/2



Ποσοστό Σωστό = 36.36 %

Ποσοστό Λάθος = 63.64 %

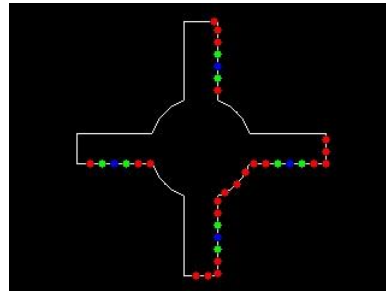
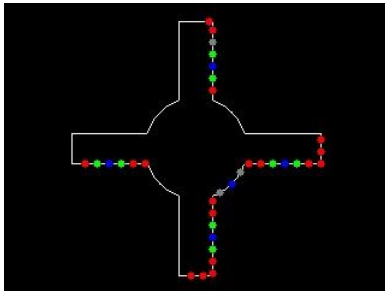
Αριθμός λαβών = 0/4



Ποσοστό Σωστό = 76.48 %

Ποσοστό Λάθος = 23.52 %

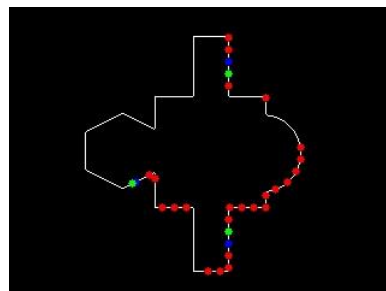
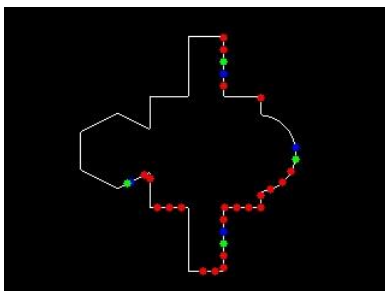
Αριθμός λαβών = 1/2



Ποσοστό Σωστό = 96.78 %

Ποσοστό Λάθος = 3.22 %

Αριθμός λαβών = 4/5



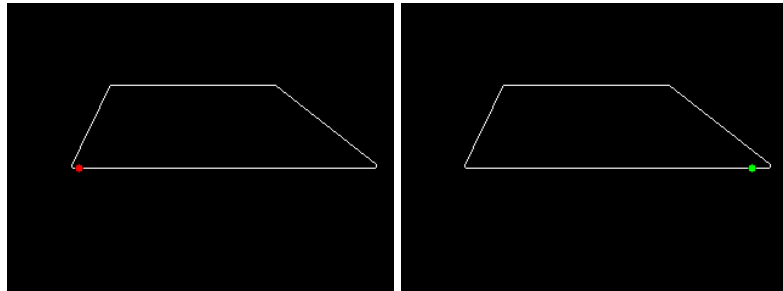
Ποσοστό Σωστό = 93.33%

Ποσοστό Λάθος = 6.67 %

Αριθμός λαβών = 3/4

Τα αποτελέσματα είναι πολύ καλά στην περίπτωση των απλών σχημάτων (κύκλος, τετράγωνο, έλλειψη, παραλληλόγραμμο), με υψηλά ποσοστά και αναγνώριση του μεγαλύτερου μέρους των λαβών. Σε αυτά τα σχήματα το νευρωνικό δίκτυο δεν είχε πρόβλημα να ταξινομήσει σωστά τα περισσότερα σημεία, εκτός από ελάχιστες εξαιρέσεις. Οι λαβές έχουν ανιχνευθεί στα σωστά σημεία, με εξαίρεση την δεξιά λαβή του παραλληλογράμμου. Για την περίπτωση του τραπεζίου και του εξάγωνου, το σύστημα παρουσίασε μεγαλύτερη δυσκολία σωστής ταξινόμησης των σημείων του περιγράμματος. Ενώ το ποσοστό παραμένει σε γενικές γραμμές καλό (50% και άνω) υπάρχουν κάποια σημεία ιδιομορφίας, όπως η γωνία του εξάγωνου, που μπορεί να δημιουργήσουν πρόβλημα στην περίπτωση που το ρομπότ αποφασίσει να σχηματίσει λαβή σε αυτά. Ταυτόχρονα, δεν έχει ανιχνευθεί η λαβή που αντιστοιχεί στην πλευρά μικρής κλίσης του τραπεζίου. Στα πολυπλοκότερα σχήματα ο αλγόριθμος εμφανίζει επίσης υψηλά ποσοστά αναγνώρισης, καθώς και ανίχνευση των περισσότερων λαβών σε αυτά. Η χειρότερη επίδοση συναντάται στην περίπτωση του αστεριού, όπου ο αλγόριθμος δεν έχει ανιχνεύσει κανένα σημείο με τιμή 1 και κατ' επέκταση κανένα σημείο λαβής. Είναι λοιπόν φανερό η δυσκολία του αλγορίθμου σε σχήματα που έχουν κλίση στις πλευρές τους, όπου δεν μπόρεσε να διαχωρίσει σημεία με μικρή κλίση, που είναι ικανοποιητικά για λαβή, από σημεία με μεγάλη κλίση. Στις υπόλοιπες περιπτώσεις, το ρομπότ έχει ταξινομήσει σωστά τα σημεία που πλησιάζουν στο μέσο κάθε αντικειμένου, τα οποία εμφανίζουν συμμετρία στο σχήμα, παρέχοντας έτσι τη δυνατότητα για σχηματισμό σθεναρής λαβής στο μέσο του αντικειμένου. Παρατηρείται επιπλέον η συνάφεια μεταξύ των ανιχνευμένων σημείων με τιμή 1 και του συνόλου εκπαίδευσης, με επιτυχή αναγνώριση των σημείων που είναι όμοια με αυτά στα οποία εκπαιδεύθηκε. Σε γενικές γραμμές, η λειτουργία του αλγορίθμου με σύνολο εκπαίδευσης 5 σημεία μπορεί να βελτιωθεί. Παρακάτω, θα εξεταστούν άλλες 2 περιπτώσεις που το σύνολο εκπαίδευσης άλλαξε και θα αξιολογηθούν τα αποτελέσματα.

Στην επόμενη φάση, προστίθενται 2 σημεία από τις πλευρές του τραπέζιου, ένα που αντιστοιχεί σε αντιδιαμετρική πλευρά με χαμηλή κλίση και ένα σε πλευρά με υψηλή κλίση.



Η επιλογή έγινε για να μπορεί το ρομπότ να αποφαίνεται σε αντικείμενα με πλευρές που έχουν κλίση και να βελτιωθούν τα αποτελέσματα της προηγούμενης ενότητας. Το αποτέλεσμα της εκτέλεσης φαίνεται παρακάτω.



Ποσοστό Σωστό = 100%

Ποσοστό Λάθος = 0 %

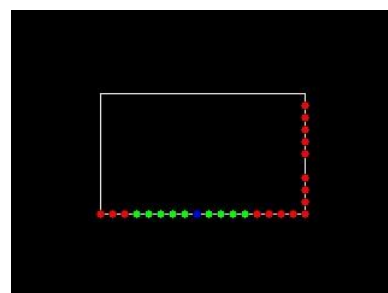
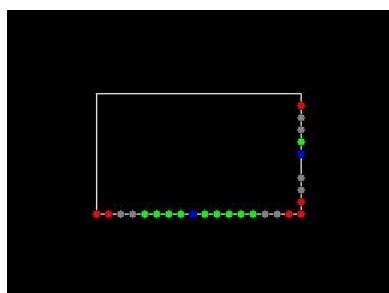
Αριθμός λαβών = 1/1



Ποσοστό Σωστό = 100 %

Ποσοστό Λάθος = 0 %

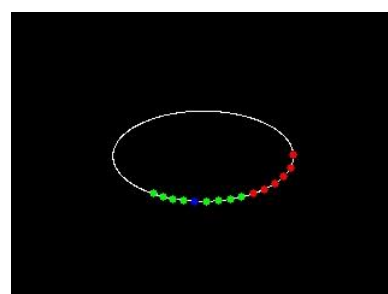
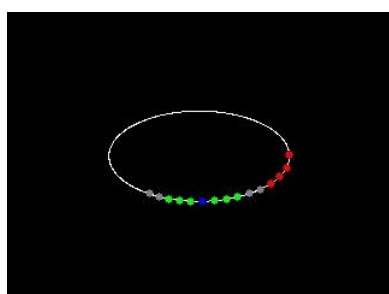
Αριθμός λαβών = 2/2



Ποσοστό Σωστό = 88.88%

Ποσοστό Λάθος = 11.12%

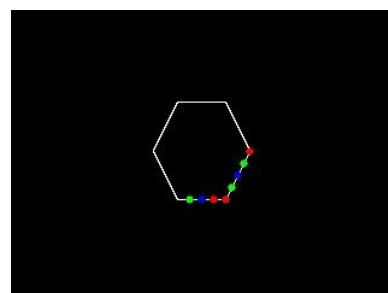
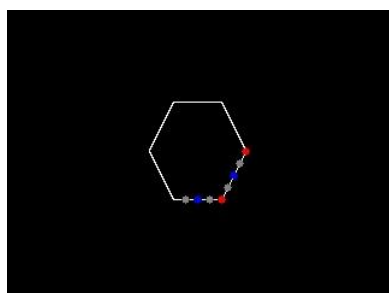
Αριθμός λαβών = 1/2



Ποσοστό Σωστό = 100 %

Ποσοστό Λάθος = 0%

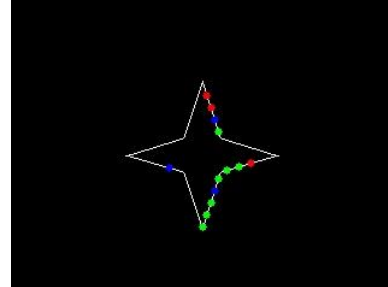
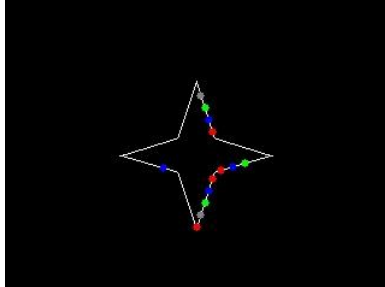
Αριθμός λαβών = 1/1



Ποσοστό Σωστό = 100%

Ποσοστό Λάθος = 0 %

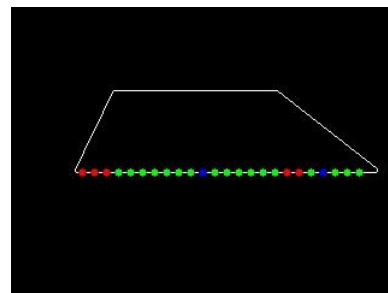
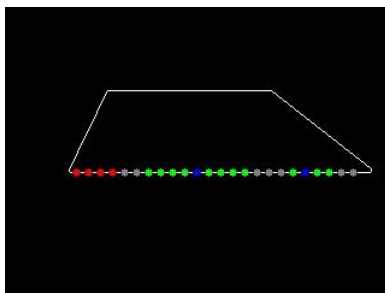
Αριθμός λαβών = 2/2



Ποσοστό Σωστό = 45.45%

Ποσοστό Λάθος = 54.55 %

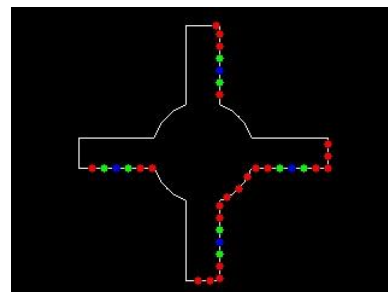
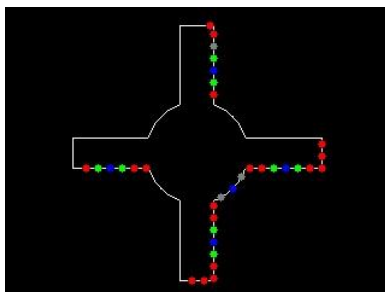
Αριθμός λαβών = 3/4



Ποσοστό Σωστό = 88.24%

Ποσοστό Λάθος = 11.76%

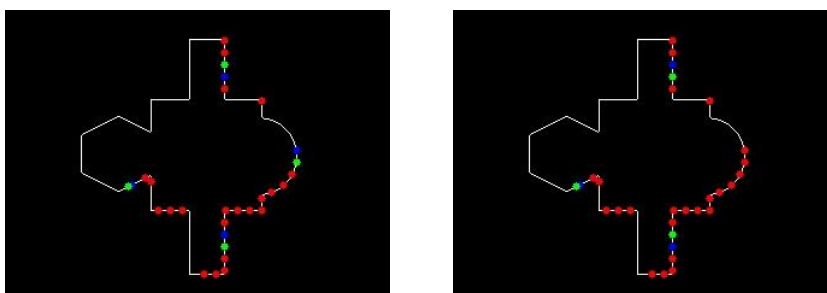
Αριθμός λαβών = 2/2



Ποσοστό Σωστό = 96.78 %

Ποσοστό Λάθος = 3.22 %

Αριθμός λαβών = 4/5



Ποσοστό Σωστό = 93.33%

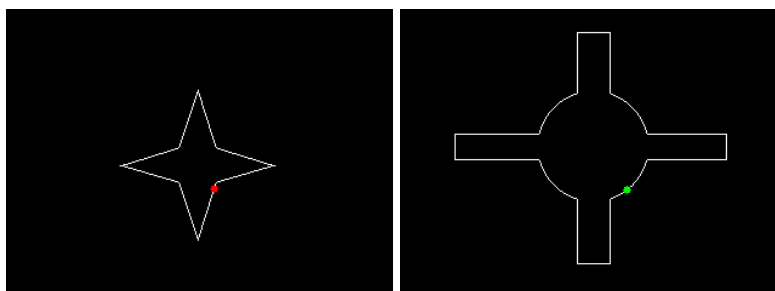
Ποσοστό Λάθος = 6.67 %

Αριθμός λαβών = 3/4

Συγκρινόμενα με αυτά του αρχικού συνόλου εκπαίδευσης, τα αποτελέσματα είναι εμφανώς βελτιωμένα, αν και υπάρχει επιπλέον περιθώριο βελτίωσης. Όσον αφορά τα βασικά σχήματα, το αποτέλεσμα δεν έχει αλλάξει καθόλου. Τα ποσοστά παραμένουν στις ίδιες υψηλές τιμές και οι λαβές είναι όλες αναγνωρισμένες, με εξαίρεση την δεξιά λαβή του παραλληλογράμμου.. Αυτό συμβαίνει διότι οι πλευρές τους είναι πιο ομαλές χωρίς πολλές προεξοχές και διαφοροποιήσεις και ανταποκρίνονται ακριβώς στα δεδομένα εκπαίδευσης, με αποτέλεσμα η αναγνώριση να είναι πιο εύκολη. Στα σχήματα με κλίση στις πλευρές τους, τα ποσοστά είναι πολύ καλύτερα σε σχέση με την εκτέλεση με βασικό σύνολο εκπαίδευσης, πράγμα που είναι λογικό διότι τα δύο σημεία που προστέθηκαν στο σύνολο είναι σχεδιασμένα να αντιπροσωπεύουν κλίση στις αντιδιαμετρικές περιοχές τους. Για το τραπέζιο έχει αναγνωριστεί η δεξιά λαβή, ενώ στο αστερί έχουν αναγνωριστεί 3 λαβές. Στα πολύπλοκα αντικείμενα ο αλγόριθμος έχει παραμείνει στα ίδια αποτελέσματα, με τα ίδια ποσοστά και τις ίδιες ανιχνεύσιμες λαβές. Η προσθήκη λοιπόν δύο σημείων κλίσης στο σύνολο εκπαίδευσης, οδηγεί σε καλύτερα αποτελέσματα σε όλα τα σχήματα, σε μεγαλύτερη δυνατότητα αναγνώρισης πιο σύνθετης μορφολογίας και σε αυξημένη ικανότητα γενίκευσης του συστήματος.

Εύκολα φαίνεται η αδυναμία του αλγορίθμου να ταξινομήσει σωστά σημεία που βρίσκονται σε εσωτερικό κοιλωμάτων, όπως στην περίπτωση του αστεριού και των

πολυπλοκότερων σχημάτων. Το εσωτερικό του αστεριού ταξινομείται ως σύνολο ικανοποιητικών σημείων, ενώ το εσωτερικό του σχήματος σταυρού με κύκλο ταξινομείται ως μη ικανοποιητικό για αναγνώριση. Για το λόγο αυτό, για την εκτέλεση της επόμενης δοκιμής, δίνονται στο ρομπότ 2 επιπλέον σημεία εκπαίδευσης από το εσωτερικό των αντικειμένων, τα οποία αντιπροσωπεύουν περιοχές που δεν είναι καλές ως προς τον σχηματισμό λαβής. Το ένα σημείο είναι στο εσωτερικό του αστεριού και το άλλο στο κυκλικό σημείο του σχήματος σταυρού με κύκλο, τα οποία θεωρούνται σημεία τιμής 0 και 1 αντίστοιχα.



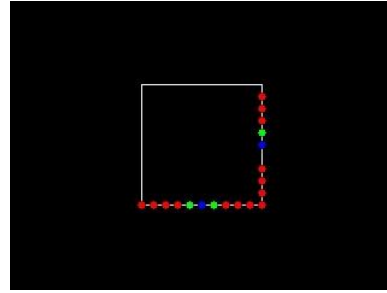
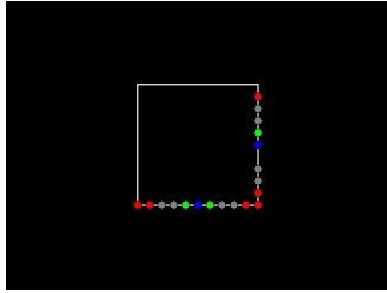
Η προσπάθεια αυτή γίνεται για να αυξηθούν ακόμα περισσότερο οι λαβές και για να υπάρξει μεγαλύτερη ακρίβεια σε περιοχές κοιλιοτήτων. Τα αποτελέσματα της εκτέλεσης φαίνονται παρακάτω.



Ποσοστό Σωστό = 100%

Ποσοστό Λάθος = 0%

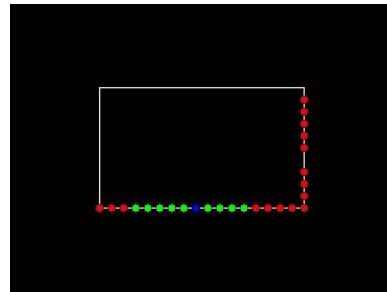
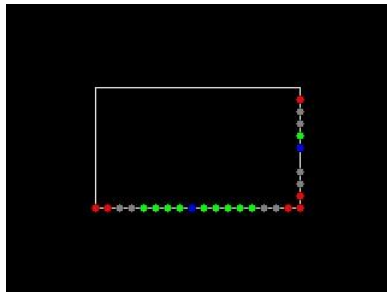
Αριθμός λαβών = 1/1



Ποσοστό Σωστό = 100%

Ποσοστό Λάθος = 0%

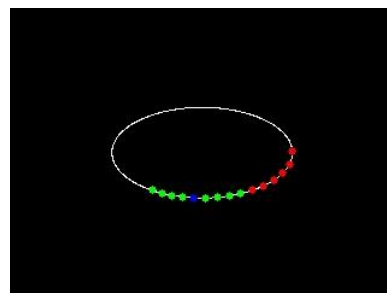
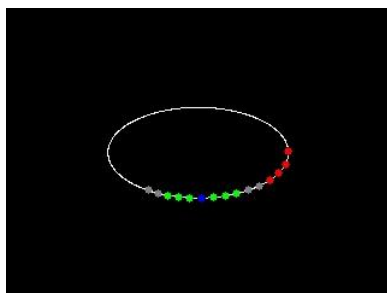
Αριθμός λαβών = 2/2



Ποσοστό Σωστό = 88.88%

Ποσοστό Λάθος = 11.12%

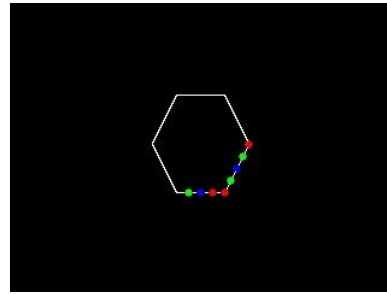
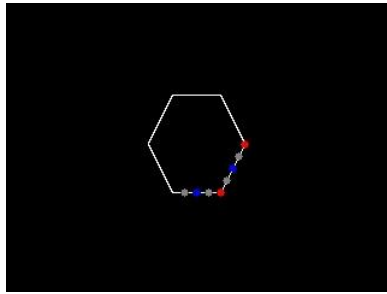
Αριθμός λαβών = 1/2



Ποσοστό Σωστό = 100 %

Ποσοστό Λάθος = 0%

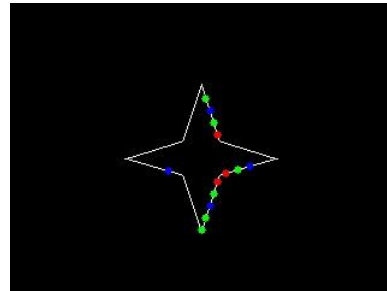
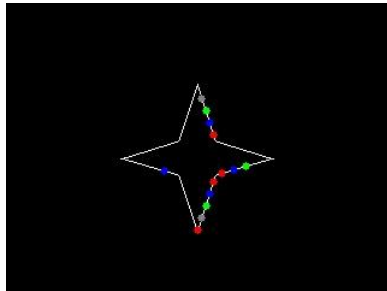
Αριθμός λαβών = 1/1



Ποσοστό Σωστό = 100%

Ποσοστό Λάθος = 0%

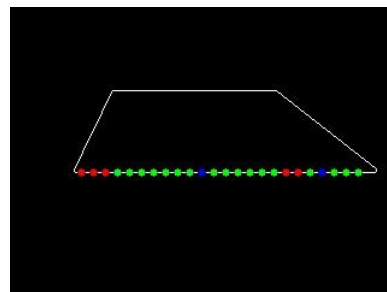
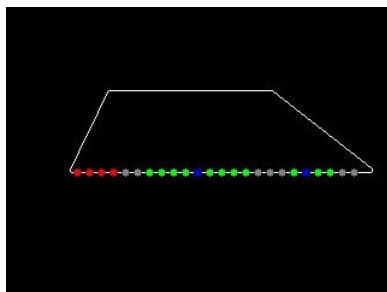
Αριθμός λαβών = 2/2



Ποσοστό Σωστό = 81.81%

Ποσοστό Λάθος = 18.19%

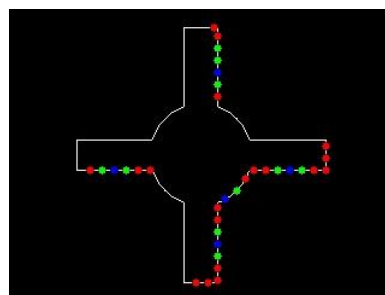
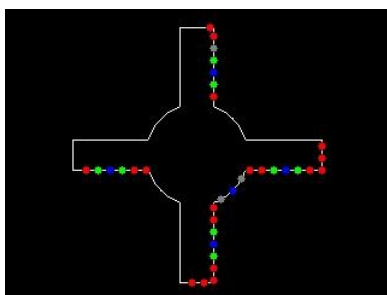
Αριθμός λαβών = 4/4



Ποσοστό Σωστό = 88.23 %

Ποσοστό Λάθος = 11.76%

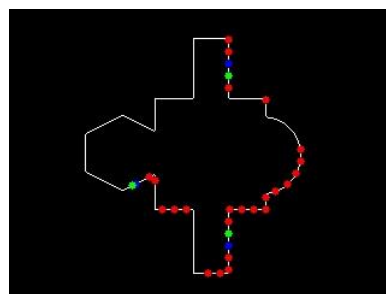
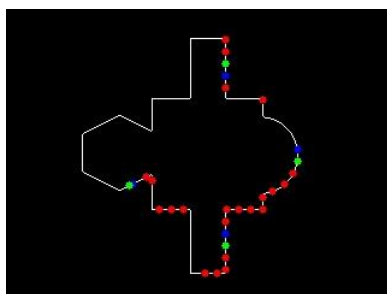
Αριθμός λαβών = 2/2



Ποσοστό Σωστό = 100 %

Ποσοστό Λάθος = 0%

Αριθμός λαβών = 5/5



Ποσοστό Σωστό = 93.33%

Ποσοστό Λάθος = 6.67 %

Αριθμός λαβών = 3/4

Η χρησιμοποίηση των επιπλέον 2 σημείων κοιλοτήτων έδωσε τα καλύτερα αποτελέσματα για το σύνολο αναγνώρισης. Οι εικόνες για κάθε σχήμα έχουν μεγάλη συνάφεια με τις εικόνες αναφοράς, ενώ τα ποσοστά αναγνώρισης είναι σε μεγάλο βαθμό υψηλά. Οι περισσότερες λαβές έχουν ανιχνευθεί, ενώ τα δύο σημεία που προστέθηκαν στο σύνολο εκπαίδευσης ταξινομήθηκαν ορθά. Σε κάποια σχήματα υπάρχει ολοκληρωτική ταύτιση της εικόνας αναφοράς με αυτήν που παρουσίασε ο αλγόριθμος. Τα ποσοστά αναγνώρισης προέρχονται από όλους τους τύπους σημείων, ενώ και στα πολυπλοκότερα σχήματα, ο αλγόριθμος έχει παρουσιάσει πολύ καλή λειτουργία. Οι μόνες λαβές που δεν

αναγνωρίστηκαν είναι η δεξιά λαβή του παραλληλογράμμου και η δεξιά λαβή του τελευταίου σχήματος. Όσον αφορά το παραλληλόγραμμο, λόγω του μεγάλου μήκους του η μη αναγνωρισμένη λαβή έχει περιγραφέα παρόμοιο με αυτόν του σημείου του τετραγώνου κοντά στη δεξιά γωνία. Η δεξιά λαβή του τελευταίου σχήματος έχει περιγραφέα που μοιάζει με γωνία αντικείμενου, δηλαδή σχετικά ομαλό στα άκρα και απότομο στο κέντρο. Η πρόσθεση λοιπόν αυτών των δύο σημείων στο σύνολο εκπαίδευσης θα αλλοίωσε κατά πολύ όλα τα υπόλοιπα σχήματα καθώς θα οδηγούσε σε λανθασμένη ταξινόμηση των γωνιών τους. Για το λόγο αυτό οι δύο αυτές οι λαβές δεν μπορούν να αναγνωριστούν. Τέλος, τονίζεται πως χρήση σχετικά λίγων σημείων για εκπαίδευση είναι πολύ σημαντική, καθώς συντελεί στην ανάπτυξη της ικανότητας γενίκευσης του συστήματος για πολύπλοκα σχήματα. Με χρήση μόλις 9 αντιπροσωπευτικών για κάθε μορφολογία σημείων, το σύστημα παρουσίασε μία πολύ ικανοποιητική εικόνα για την πλειονότητα των σχημάτων. Κατά συνέπεια, το τελευταίο σύνολο δεδομένων εκπαίδευσης, θα χρησιμοποιηθεί στο επόμενο κεφάλαιο για την εκτέλεση πειραμάτων σε πραγματικά αντικείμενα.

ΚΕΦΑΛΑΙΟ 6

6 Εντοπισμός και σχηματισμός λαβής

6.1 Εισαγωγή

Έχοντας αναλύσει τη διαδικασία εκπαίδευσης του ρομπότ και έχοντας καταλήξει σε κάποιο σημείο ικανό για την δημιουργία λαβής, το επόμενο βήμα είναι η εκτέλεση της λαβής από το ρομπότ NAO. Η εκτέλεση γίνεται με τη βοήθεια των μηχανικών μερών του ρομπότ και συγκεκριμένα το δεξί του χέρι. Για τον σχηματισμό και την εκτέλεση της λαβής απαιτείται ο συνδυασμός πληροφορίας και μεθόδων που έχουν αναλυθεί στα Κεφάλαια 2,3 και 5. Στο παρόν κεφάλαιο παρουσιάζεται αναλυτικότερα το ρομπότ NAO και οι δυνατότητες του όσον αφορά την εκτέλεση λαβών, ευρίσκονται οι εξισώσεις της ορθής κινηματικής ανάλυσης, καθώς επίσης και το ανάστροφο κινηματικό μοντέλο που παίζει βασικό ρόλο στην κίνηση του χεριού, περιγράφεται η συνολική διαδικασία εντοπισμού και δημιουργίας της λαβής, ενώ δίνονται παραδείγματα εκτέλεσης για το σύνολο αντικειμένων της Εικόνας 23.

6.2 Ρομπότ NAO

Στην παρούσα διπλωματική, το ρομποτικό σύστημα που χρησιμοποιήθηκε είναι το ανθρωποειδές ρομπότ NAO. Το ρομπότ NAO δημιουργήθηκε από την εταιρία Aldebaran Robotics και η κατασκευή και ο εύκολος προγραμματισμός του οδήγησαν στην ευρεία χρήση του ως ρομπότ ψυχαγωγίας, παροχής βοήθειας σε οικιακό περιβάλλον, παροχής βοήθειας σε άτομα που υποφέρουν από αυτισμό και άλλες. Αποτελεί βασική πλατφόρμα έρευνας σε πανεπιστήμια και ερευνητικά κέντρα παγκοσμίως, με σημαντικό ερευνητικό έργο να έχει δοκιμαστεί σε αυτό.



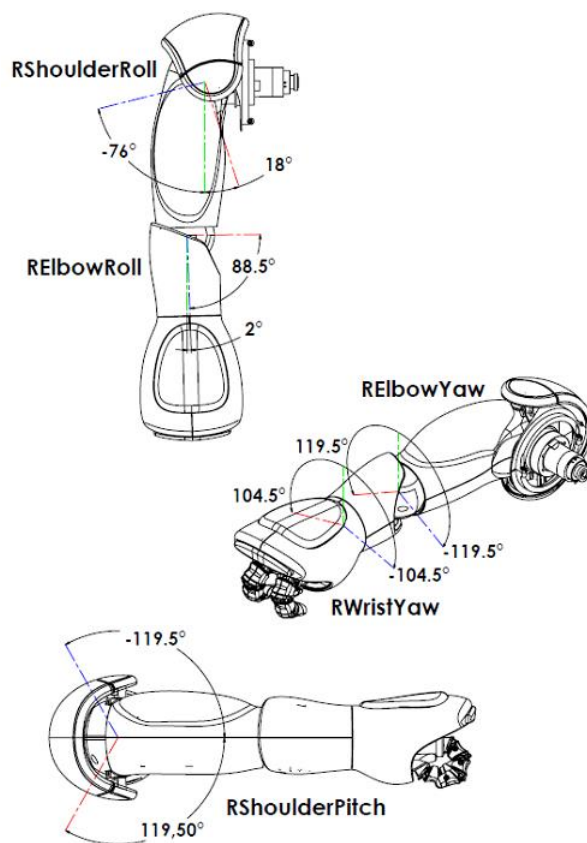
Εικόνα 22: Ρομπότ NAO.

Το ρομπότ NAO έχει 14, 21 ή 25 βαθμούς ελευθερίας ανάλογα με το μοντέλο. Είναι εφοδιασμένο με αισθητήρες αφής στο κεφάλι, αισθητήρες συγκρούσεων στα πόδια και στα χέρια, αισθητήρες απόστασης με χρήση υπερήχων, επιταχυνσιόμετρο, μικρόφωνο, ηχεία και 2 κάμερες υψηλής ευκρίνειας που βρίσκονται στο κεφάλι του. Στον παρακάτω πίνακα φαίνονται κάποια επιπλέον τεχνικά χαρακτηριστικά του ρομπότ.

Nao Next Gen (2011)	
Height	58 centimetres (23 in)
Weight	4.3 kilograms (9.5 lb)
Autonomy	60 minutes (active use), 90 minutes (normal use)
Degrees of freedom	21 to 25
CPU	Intel Atom @ 1.6 GHz
Built-in OS	Linux
Compatible OS	Windows , Mac OS , Linux
Programming languages	C++ , Python , Java , MATLAB , Urbi , C , .Net
Vision	Two HD 1280x960 cameras
Connectivity	Ethernet , Wi-Fi

Ο προγραμματισμός του ρομπότ γίνεται με γραμμένα προγράμματα από τον χρήστη, τα οποία είτε τρέχουν στο ρομπότ, είτε συνδέονται με το ρομπότ μέσω Ethernet ή WiFi. Υπάρχουν έτοιμες συναρτήσεις που επικοινωνούν με το υλικό και τους επενεργητές του ρομπότ, με αποτέλεσμα η χρήση τους να καθιστά την ανάπτυξη ρομποτικών εφαρμογών πιο εύκολη και άμεση.

Τα χέρια του ρομπότ έχουν 5 βαθμούς ελευθερίας, επιτρέποντάς του ένα μεγάλο εύρος κινήσεων και αλληλεπίδρασης με αντικείμενα εντός του χώρου εργασίας του.



Εικόνα 23: Βαθμοί ελευθερίας του δεξιού χεριού του ρομπότ NAO

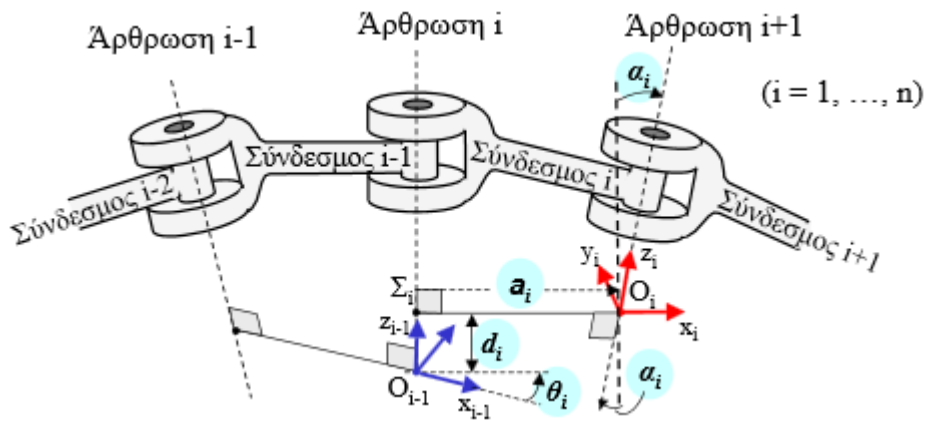
6.3 Ορθή κινηματική ανάλυση

Με τον όρο *κινηματική αλυσίδα* εννοείται οποιοδήποτε σύστημα στερεών σωμάτων που συνδέονται μεταξύ τους με *αρθρώσεις*. Οι κινηματικές αλυσίδες έχουν *βαθμούς ελευθερίας*, δηλαδή έναν αριθμό ανεξάρτητων μεταβλητών οι οποίοι χρησιμεύουν για την περιγραφή της διάταξης της αλυσίδας στο χώρο. Το χέρι του ρομπότ αποτελεί μία ανοικτή κινηματική αλυσίδα με 5 βαθμούς ελευθερίας, ενώ στις αρθρώσεις του βρίσκονται ηλεκτρικοί κινητήρες συνεχούς ρεύματος που περιστρέφουν κάθε μέλος της αλυσίδας και το οδηγούν σε κάποια θέση, προκαθορισμένη ή όχι. Η εύρεση της θέσης του τελικού στοιχείου του ρομποτικού βραχίονα, έχοντας ως γνωστές τις περιστροφές των αρθρώσεων του είναι αντικείμενο της *ορθής κινηματικής ανάλυσης* του βραχίονα και αποτελεί το αντικείμενο αυτής της ενότητας. Για την ορθή κινηματική ανάλυση απαιτείται να τοποθετηθεί σε κάθε άρθρωση ένα ξεχωριστό πλαίσιο, όπως επίσης και στη βάση και στο τελικό στοιχείο του ρομπότ, έτσι ώστε με συνεχείς μετασχηματισμούς μεταξύ των πλαισίων, να δοθεί μία περιγραφή που θα συνδέει τη βάση με το τελικό στοιχείο δράσης και

θα δίνει μία μαθηματική σχέση μεταξύ των γωνιών των αρθρώσεων και της θέσης και του προσανατολισμού του τελικού στοιχείου δράσης. Ένας εύκολος και άμεσος τρόπος να διατυπωθεί αυτή η σχέση, γίνεται με την Μέθοδο Denavit – Hartenberg, η οποία περιγράφεται παρακάτω.

6.3.1 Μέθοδος Denavit - Hartenberg

Η μέθοδος DH αποσκοπεί στον ορισμό της θέσης και του προσανατολισμού μεταξύ διαδοχικών συνδέσμων (Denavit- Hartenberg, 1964). Για να είναι αυτό εφικτό, είναι απαραίτητη η τοποθέτηση πλαισίων συντεταγμένων σε κάθε άρθρωση. Η μέθοδος έχει ως βασικό στοιχείο τον ορισμό πλαισίων με γνώμονα τέσσερις βασικούς κανόνες.



Εικόνα 24: Τοποθέτηση πλαισίων σε ρομποτικό βραχίονα

1. Επιλέγεται ο άξονας z_i κατά μήκος του άξονα της άρθρωσης $i + 1$.
2. Το κέντρο O_i του πλαισίου τοποθετείται στην τομή του άξονα z_i με την κοινή κάθετο των αξόνων z_{i-1} και z_i .
3. Ο άξονας x_i επιλέγεται κατά μήκος της κοινής κάθετου των αξόνων z_i και z_{i-1} , με φορά από την άρθρωση i στην άρθρωση $i + 1$.

4. Επιλέγεται ο άξονας y_i έτσι ώστε να σχηματίζεται δεξιόστροφο ορθοκανονικό σύστημα συντεταγμένων για το πλαίσιο i .

Για την περιγραφή της κινηματικής αλυσίδας, η μέθοδος χρησιμοποιεί τέσσερις βασικές παραμέτρους, οι οποίες ορίζονται ως εξής:

d_i : Η απόσταση κατά μήκος του άξονα z_{i-1} από το σημείο O_{i-1} μέχρι το σημείο τομής των αξόνων x_i και z_{i-1} .

θ_i : Η γωνία μεταξύ του άξονα x_i και του άξονα x_{i-1} , γύρω από τον άξονα z_{i-1} κατά την δεξιόστροφη φορά.

a_i : Η απόσταση του σημείου τομής των αξόνων x_i και z_{i-1} μέχρι το σημείο O_i , κατά μήκος του άξονα x_i .

α_i : Η γωνία μεταξύ των αξόνων z_{i-1} και z_i γύρω από τον άξονα x_i , κατά τη δεξιόστροφη φορά.

Έχοντας ορίσει λοιπόν τα πλαίσια των αρθρώσεων, όπως στην Εικόνα 26 και έχοντας βρει τις παραμέτρους της μεθόδου, εκτελούνται τα παρακάτω βήματα για τον ορισμό της σχετικής θέσης και του προσανατολισμού του πλαισίου i ως προς το πλαίσιο $i + 1$.

Βήμα 1: Περιστροφή του πλαισίου $i - 1$ γύρω από τον άξονα z_{i-1} κατά γωνία θ_i .

Βήμα 2: Μετατόπιση d_i του πλαισίου $i - 1$ κατά μήκος του άξονα z_{i-1} .

Βήμα 3: Μετατόπιση a_i κατά τον άξονα x_i , όσο και το μήκος της κοινής καθέτου.

Βήμα 4: Περιστροφή γύρω από τον άξονα x_i κατά γωνία α_i .

Είναι προφανές πως για την περιγραφή γίνονται κατά σειρά μία περιστροφή, δύο μετατοπίσεις και άλλη μία περιστροφή. Εάν χρησιμοποιηθούν *μήτρες μετασχηματισμού* , από τα Βήματα 1 και 2 προκύπτει

$$A_1 = \begin{pmatrix} \cos \theta_i & -\sin \theta_i & 0 & 0 \\ \sin \theta_i & \cos \theta_i & 0 & 0 \\ 0 & 0 & 1 & d_i \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6.1)$$

ενώ από τα Βήματα 3 και 4

$$A_2 = \begin{pmatrix} 1 & 0 & 0 & a_i \\ 0 & \cos \alpha_i & -\sin \alpha_i & 0 \\ 0 & \sin \alpha_i & \cos \alpha_i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6.2)$$

Τελικά, η μήτρα μετασχηματισμού από το πλαίσιο i στο πλαίσιο $i + 1$ δίνεται από τον τύπο $A_i^{i-1} = A_1 A_2$ και ισούται με

$$A_i^{i-1} = \begin{pmatrix} \cos \theta_i & -\sin \theta_i \cos \alpha_i & \sin \theta_i \sin \alpha_i & a_i \cos \theta_i \\ \sin \theta_i & \cos \theta_i \cos \alpha_i & -\cos \theta_i \sin \alpha_i & a_i \sin \theta_i \\ 0 & \sin \alpha_i & \cos \alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6.3)$$

6.3.2 Τοποθέτηση πλαισίων και πίνακας DH για το ρομπότ NAO

Για το δεξί χέρι, και για συγκεκριμένη αρχική διάταξη, η τοποθέτηση των πλαισίων σύμφωνα με τη μέθοδο DH φαίνεται στο παρακάτω σχήμα

Ακολουθώντας τη μεθοδολογία της μεθόδου DH προέκυψε ο πίνακας παραμέτρων DH για το δεξί χέρι του NAO

Πλαίσιο	α_i	a_i	d_i	θ_i
1	-90°	0	0	q_1
2	90°	0	0	$q_2 + 90^\circ$
3	-90°	0	l_1	q_3
4	90°	0	0	q_4
E	0	0	$l_2 + l_3$	q_5

Πίνακας 2: Πίνακας παραμέτρων Denavit - Hartenberg για το δεξί χέρι του ρομπότ ΝΑΟ

Ενώ οι μήτρες μετασχηματισμού είναι (με $c_i = \cos(i)$ και $s_i = \sin(i)$)

$$A_0^1 = \begin{pmatrix} c_1 & 0 & -s_1 & 0 \\ s_1 & 0 & c_1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$A_1^2 = \begin{pmatrix} -s_2 & 0 & c_2 & 0 \\ c_2 & 0 & s_2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$A_2^3 = \begin{pmatrix} c_3 & 0 & -s_3 & 0 \\ s_3 & 0 & c_3 & 0 \\ 0 & -1 & 0 & l_1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$A_3^4 = \begin{pmatrix} c_4 & 0 & -s_4 & 0 \\ s_4 & 0 & c_4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$A_4^5 = \begin{pmatrix} c_5 & -s_5 & 0 & 0 \\ s_5 & c_5 & 0 & 0 \\ 0 & 0 & 1 & l_2 + l_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Ο πίνακας A_0^5 προκύπτει από τη σχέση $A_0^5 = A_0^1 A_1^2 A_2^3 A_3^4 A_4^5$

$$A_0^5 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

όπου

$$\begin{aligned} a_{11} &= -c_5[c_4(s_1s_3 + c_1c_3s_2) + c_1c_2s_4] - c_1s_2s_3 \\ a_{12} &= s_5[c_4(s_1s_3 + c_1c_3s_2) + c_1c_2s_4] - c_1s_2s_3 \\ a_{13} &= -s_4(s_1s_3 + c_1c_3s_2) + c_1c_2c_4 \\ a_{14} &= l_1c_1c_2 - (l_2 + l_3)[s_4(s_1s_3 + c_1c_3s_2) - c_1c_2c_4] \\ a_{21} &= c_5[c_4(c_1s_3 - c_1c_3s_2) - c_1c_2s_4] + s_5(c_1c_3 + c_1s_2s_3) \\ a_{22} &= -s_5[c_4(c_1s_3 - c_1c_3s_2) - c_1c_2s_4] + c_5(c_1c_3 + c_1s_2s_3) \\ a_{23} &= s_4(c_1s_3 - c_1c_3s_2) + c_1c_2c_4 \\ a_{24} &= l_1c_1c_2 + (l_2 + l_3)[s_4(c_1s_3 - c_1c_3s_2) + c_1c_2c_4] \\ a_{31} &= c_5(s_2s_4 - c_2c_3c_4) + c_2s_3s_5 \\ a_{32} &= -s_5(s_2s_4 - c_2c_3c_4) + c_2s_3c_5 \\ a_{33} &= -s_2c_4 - c_2c_3s_4 \\ a_{34} &= -l_1s_2 - (l_2 + l_3)(c_4s_2 + c_2c_3c_4) \end{aligned}$$

6.4 Αντίστροφη κινηματική και διαφορική ανάλυση

Κατά την αντίστροφη κινηματική ανάλυση του ρομποτικού βραχίονα, αναζητείται η εύρεση των τιμών των μεταβλητών q_i έχοντας ως γνωστή την θέση και τον

προσανατολισμό του τελικού στοιχείου δράσης. Η διαφορά με την ορθή κινηματική ανάλυση, είναι πως στην ορθή ανάλυση η λύση που προκύπτει για την θέση και τον προσανατολισμό της αρπάγης είναι μοναδική. Στην ανάστροφη κινηματική ανάλυση, η λύση που προκύπτει όχι μόνο δεν είναι μοναδική, αλλά μπορεί και να μην υπάρχει.

Σε περίπτωση που υπάρχει λύση, αυτή δίνεται στη μορφή μήτρας μετασχηματισμού

$$\mathbf{T}_E^0 = \begin{pmatrix} x_{E,x} & y_{E,x} & z_{E,x} & p_{E,x} \\ x_{E,y} & y_{E,y} & z_{E,y} & p_{E,y} \\ x_{E,z} & y_{E,z} & z_{E,z} & p_{E,z} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6.4)$$

Η εξίσωση του πίνακα (6.4) με τον πίνακα μετασχηματισμού \mathbf{A}_E^0 που προέκυψε από τις μεταβλητές DH, δίνει τις 12 εξισώσεις που πρέπει να λυθούν για την εύρεση των τιμών των γωνιών των αρθρώσεων q_i .

Για την περίπτωση του δεξιού χεριού του NAO, προέκυψε το παρακάτω σύστημα μη γραμμικών εξισώσεων .

$$\begin{aligned} -c_5[c_4(s_1s_3 + c_1c_3s_2) + c_1c_2s_4] - c_1s_2s_3 &= x_{E,x} \\ c_5[c_4(c_1s_3 - c_1c_3s_2) - c_1c_2s_4] + s_5(c_1c_3 + c_1s_2s_3) &= x_{E,y} \\ c_5(s_2s_4 - c_2c_3c_4) + c_2s_3s_5 &= x_{E,z} \\ s_5[c_4(s_1s_3 + c_1c_3s_2) + c_1c_2s_4] - c_1s_2s_3 &= y_{E,x} \\ -s_5[c_4(c_1s_3 - c_1c_3s_2) - c_1c_2s_4] + c_5(c_1c_3 + c_1s_2s_3) &= y_{E,y} \\ -s_5(s_2s_4 - c_2c_3c_4) + c_2s_3c_5 &= y_{E,z} \\ -s_4(s_1s_3 + c_1c_3s_2) + c_1c_2c_4 &= z_{E,x} \\ s_4(c_1s_3 - c_1c_3s_2) + c_1c_2c_4 &= z_{E,y} \\ -s_2c_4 - c_2c_3s_4 &= z_{E,z} \\ l_1c_1c_2 - (l_2 + l_3)[s_4(s_1s_3 + c_1c_3s_2) - c_1c_2c_4] &= p_{E,x} \\ l_1c_1c_2 + (l_2 + l_3)[s_4(c_1s_3 - c_1c_3s_2) + c_1c_2c_4] &= p_{E,y} \\ -l_1s_2 - (l_2 + l_3)(c_4s_2 + c_2c_3c_4) &= p_{E,z} \end{aligned}$$

Για το σύστημα αυτό είναι δύσκολο να ευρεθεί μία μαθηματική λύση. Για το λόγο χρησιμοποιείται *ανάστροφη διαφορική ανάλυση* για την εύρεση των γωνιών των αρθρώσεων. Η μέθοδος αυτή χρησιμοποιείται και από τους κατασκευαστές του ρομπότ, για τον έλεγχο των αρθρώσεων.

Το ευθύ διαφορικό μοντέλο του ρομπότ δίνεται από τη σχέση

$$\dot{\mathbf{p}} = \mathbf{J}(\mathbf{q})\dot{\mathbf{q}} \quad (6.5)$$

όπου $\dot{\mathbf{q}}$ η γωνιακή ταχύτητα των αρθρώσεων του βραχίονα, $\dot{\mathbf{p}}$ το διάνυσμα γραμμικής και γωνιακής ταχύτητας του τελικού στοιχείου δράσης και $\mathbf{J}(\mathbf{q})$ η Ιακωβιανή μήτρα του ρομπότ. Η μήτρα αυτή είναι διαστάσεων $6 \times n$, όπου n ο αριθμός βαθμών ελευθερίας του ρομποτικού βραχίονα. Χωρίζεται σε 2 υποπίνακες $3 \times n$, οι οποίοι εκφράζουν την επίδραση του κάθε μέλους της αλυσίδας στην γραμμική και στη γωνιακή ταχύτητα του τελικού στοιχείου δράσης. Ο υπολογισμός κάθε γραμμής του πίνακα γραμμικών ταχυτήτων γίνεται με παραγωγή των p_x, p_y, p_z ως προς $q_i, i = 1 \dots n$. Η κάθε στήλη \mathbf{b}_i του πίνακα γωνιακών ταχυτήτων ισούται με την 3^η στήλη του πίνακα \mathbf{R}_0^i [Τζαφέστας, 2003]. Με βάση τα παραπάνω, η Ιακωβιανή μήτρα για το χέρι του ΝΑΟ είναι

$$\mathbf{J}(\mathbf{q}) = \begin{pmatrix} j_{11} & j_{12} & j_{13} & j_{14} & j_{15} \\ j_{21} & j_{22} & j_{23} & j_{24} & j_{25} \\ j_{31} & j_{32} & j_{33} & j_{34} & j_{35} \\ j_{41} & j_{42} & j_{43} & j_{44} & j_{45} \\ j_{51} & j_{52} & j_{53} & j_{54} & j_{55} \\ j_{61} & j_{62} & j_{63} & j_{64} & j_{65} \end{pmatrix}$$

όπου

$$j_{11} = -l_1 s_1 c_2 - (l_2 + l_3)[s_4(c_1 s_3 - s_1 c_3 s_2) + s_1 c_2 c_4]$$

$$j_{12} = -l_1 c_1 s_2 - (l_2 + l_3)[s_4 c_1 c_3 c_2 + c_1 s_2 c_4]$$

$$\begin{aligned}
j_{13} &= -(l_2 + l_3)s_4(c_1s_3 - s_1s_3s_2) \\
j_{14} &= -(l_2 + l_3)[c_4(s_1s_3c_2 + c_1c_3s_2) + c_1c_2s_4] \\
j_{14} &= 0 \\
j_{21} &= -l_1s_1c_2 + (l_2 + l_3)[s_4(-s_1s_3 + s_1c_3s_2) - s_1c_2c_4] \\
j_{22} &= -l_1c_1s_2 - (l_2 + l_3)(s_4c_1c_3c_2 - c_1s_2c_4) \\
j_{23} &= (l_2 + l_3)s_4(c_1c_3 + c_1s_3s_2) \\
j_{24} &= (l_2 + l_3)c_4(c_1s_3 - c_1c_3s_2) - c_1c_2s_4 \\
j_{25} &= 0 \\
j_{31} &= 0 \\
j_{32} &= -l_1c_2 - (l_2 + l_3)(c_4c_2 - s_2c_3c_4) \\
j_{33} &= (l_2 + l_3)(c_2s_3c_4) \\
j_{34} &= -(l_2 + l_3)(-s_4c_2 - s_2c_3s_4) \\
j_{35} &= 0 \\
j_{41} &= 0 \\
j_{42} &= -s_1 \\
j_{43} &= c_1c_2 \\
j_{44} &= c_1s_3s_2 - s_1c_3 \\
j_{45} &= -s_4(s_1s_3 + c_1c_3s_2) + c_1c_2c_4 \\
j_{51} &= 0 \\
j_{52} &= c_1 \\
j_{53} &= c_1c_2 \\
j_{54} &= c_1s_3s_2 + c_1c_3 \\
j_{55} &= s_4(c_1s_3 - c_1c_3s_2) + c_1c_2c_4 \\
j_{61} &= 1 \\
j_{62} &= 0 \\
j_{63} &= -s_2 \\
j_{64} &= c_2s_3 \\
j_{65} &= -c_4s_2 - c_3c_2s_4
\end{aligned}$$

Για την εύρεση των γωνιών των αρθρώσεων επιλύεται το ανάστροφο διαφορικό πρόβλημα, δηλαδή η σχέση

$$\dot{\mathbf{q}} = \mathbf{J}(\mathbf{q})^{-1}\dot{\mathbf{p}} \quad (6.6)$$

Σε περίπτωση που η Ιακωβιανή μήτρα δεν είναι αντιστρέψιμη, επιλύεται το πρόβλημα (6.6), αλλά χρησιμοποιείται η Moore – Penrose ψευδοαντίστροφη της Ιακωβιανής, που δίνεται από τη σχέση $\mathbf{J}(\mathbf{q})^+ = \mathbf{J}(\mathbf{q})^T (\mathbf{J}(\mathbf{q})\mathbf{J}(\mathbf{q})^T)^{-1}$. Η χρήση αυτής της μήτρας, δίνει την βέλτιστη ελαχίστων τετραγώνων λύση του συστήματος διαφορικών εξισώσεων. Στο ρομπότ, η λύση των εξισώσεων προσεγγίζεται με αριθμητικές μεθόδους.

6.5 Συνολική διαδικασία εντοπισμού και σχηματισμού λαβών

Στην παρούσα ενότητα αναλύεται βήμα προς βήμα η εκτέλεση του συνολικού αλγορίθμου σχηματισμού λαβών, ενώ φαίνεται η επίδραση και ο ρόλος όσων αναφέρθηκαν στις προηγούμενες ενότητες και κεφάλαια. Παρουσιάζεται λοιπόν η λειτουργία ολόκληρου το ρομποτικού συστήματος, όπως αυτό συντέθηκε τόσο θεωρητικά, όσο και πειραματικά.

Η διάταξη του περιβάλλοντος του ρομπότ φαίνεται στην παρακάτω εικόνα. Το ρομπότ είναι καθισμένο χαμηλά μπροστά από ένα επίπεδο ώστε να μπορεί να φτάσει αλλά και να δει τα αντικείμενα που τοποθετούνται μπροστά του. Η θέση αυτή επιλέχθηκε ως η βασικότερη και αμεσότερη τοποθέτηση, χωρίς να χρειάζεται προγραμματισμός για επιπλέον κινήσεις του σώματος του ρομπότ. Μπροστά στο ρομπότ, σε κοντινή απόσταση, τοποθετούνται ένα-ένα τα αντικείμενα που φαίνονται στην Εικόνα 22. Στη συνέχεια, περιγράφονται τα βήματα που εκτελεί το ρομπότ για την εύρεση του σημείου λαβής και την εκτέλεση της λαβής.

ΒΗΜΑ 1: Κατάτμηση εικόνας με αφαίρεση φόντου

Σε πρώτη φάση, το ρομπότ παίρνει μία φωτογραφία του επιπέδου χωρίς αντικείμενο και την αποθηκεύει για όλη τη διάρκεια του πειράματος, χρησιμοποιώντας την ως βασική εικόνα φόντου $F(x, y)$. Στη συνέχεια, ειδοποιεί τον χρήστη να τοποθετήσει ένα αντικείμενο μπροστά του, μέσω κατάλληλης φωνητικής εντολής. Όταν γίνει αυτό, το ρομπότ παίρνει φωτογραφία ($I(x, y)$) το αντικείμενο μαζί με το φόντο, ενώ μετατρέπει τις δύο φωτογραφίες σε εικόνες grayscale, αντιστοιχώντας τις τιμές από τα 3 κανάλια της εικόνας, σε ένα με τιμή στο διάστημα $[0-256]$ σύμφωνα με τη σχέση (2.1), όπου η τιμή 0 αναπαριστά μαύρο χρώμα και η τιμή 256 άσπρο. Η κατατμημένη εικόνα προκύπτει από τον τύπο

$$I_{seg}(x, y) = abs(F(x, y) - I(x, y))$$

Η κατατμημένη εικόνα είναι μία εικόνα grayscale, συνεπώς υφίσταται κατωφλίωση και φιλτράρισμα για να αφαιρεθεί ο θόρυβος και να προκύψει η τελική ασπρόμαυρη εικόνα του αντικειμένου. Στην εικόνα αυτή εφαρμόζεται ανίχνευση ακμών (Canny,), το αποτέλεσμα της οποίας είναι μία ασπρόμαυρη εικόνα με τιμές 1 στα σημεία του περιγράμματος του αντικειμένου και 0 στα υπόλοιπα σημεία. Η εικόνα αυτή αποτελεί την απεικόνιση του περιγράμματος του αντικειμένου, πάνω στην οποία γίνεται όλη η ακόλουθη επεξεργασία που περιγράφεται στα επόμενα βήματα.

Για αποτελεσματικότερη κατάτμηση, μπορούν να χρησιμοποιηθούν άλλοι αλγόριθμοι κατάτμησης, όπως αυτοί που περιεγράφηκαν στο Κεφάλαιο 2. Ο αλγόριθμος αφαίρεσης φόντου προτιμήθηκε λόγω της στατικότητας του πειράματος αλλά και του ρομπότ.

ΒΗΜΑ 2: Εξαγωγή περιγραφών

Το επόμενο βήμα της κατάτμησης είναι η επεξεργασία της ασπρόμαυρης, δυαδικής εικόνας περιγράμματος για την εξαγωγή των ακτινικών περιγραφών, ολικού ακτινικού και τοπικού ακτινικού, όπως έχουν αναλυθεί στο Κεφάλαιο 2. Η εικόνα του περιγράμματος του αντικείμενου υφίσταται δειγματοληψία, ανά 10 σημεία. Με αυτόν τον τρόπο προκύπτει ένας μεγάλος βαθμός σημείων προς εξέταση, αποφεύγοντας την εξέταση όλου του περιγράμματος. Κάτι τέτοιο θα ήταν άσκοπο, λόγω του ότι γειτονικά εικονοστοιχεία αναπαριστούν πολύ κοντινά μέρη στο αντικείμενο και η εξέτάσή τους θα ήταν προβληματική από πλευράς χρόνου και πολυπλοκότητας. Έτσι, το περίγραμμα υφίσταται δειγματοληψία, για να υπάρχει μία όσο το δυνατόν ακριβέστερη αναπαράσταση του περιγράμματος, εξοικονομώντας χρόνο και υπολογιστικό χώρο.

Στο δειγματελημμένο περίγραμμα εφαρμόζεται PCA για τον υπολογισμό του πρωτεύοντος και του δευτερεύοντος άξονα. Οι άξονες αυτοί χρησιμεύουν ως άξονες αναφοράς, με την προσπέλαση και την επεξεργασία των στοιχείων να γίνεται ξεκινώντας από τον πρωτεύοντα άξονα. Στη συνέχεια υπολογίζεται ο ολικός ακτινικός περιγραφέας του δειγματελημμένου περιγράμματος, στη μορφή πίνακα που περιέχει στην 1^η στήλη τις κανονικοποιημένες ακτίνες r , στην δεύτερη τις γωνίες θ εκφρασμένες σε μοίρες και στην 3^η και στην 4^η τις συντεταγμένες εικόνας του σημείου από το οποίο προήλθαν. Έπειτα, ο περιγραφέας ταξινομείται κατά αύξουσα γωνία και διαμορφώνεται κατά τέτοιο τρόπο, ώστε το 1^ο στοιχείο να είναι το σημείο του περιγράμματος που τέμνει τον πρωτεύοντα άξονα, ή, σε περίπτωση που υπάρχουν σφάλματα στις αριθμητικές τιμές, βρίσκεται πιο κοντά σε αυτόν. Με αυτόν τον τρόπο, η προσπέλαση των σημείων του περιγράμματος γίνεται ξεκινώντας από τον πρωτεύοντα άξονα και συνεχίζοντας κατά αύξουσα γωνία (ανθωρολογιακά). Ο ολικός ακτινικός περιγραφέας δεν έχει κάποια χρησιμότητα όσον αφορά την εκπαίδευση του δικτύου, παρέχει όμως μία χρήσιμη αναπαράσταση του περιγράμματος σε ταξινομημένη μορφή καθιστώντας πιο εύκολη την περεταίρω επεξεργασία των σημείων.

Έχοντας υπολογίσει τον ολικό ακτινικό περιγραφέα, το επόμενο βήμα είναι να υπολογιστεί ο τοπικός ακτινικός περιγραφέας για κάθε δειγματελημμένο σημείο. Σε αντίθεση με τον ολικό ακτινικό περιγραφέα, ο τοπικός περιγραφέας είναι το διάνυσμα

εισόδου του SVM και χρησιμεύει στην τελική ταξινόμηση ενός σημείου ως ικανοποιητικό ή μη ικανοποιητικό. Ο υπολογισμός του γίνεται ως εξής: τα σημεία εξετάζονται κατά αύξουσα γωνία όπως αναφέρθηκε προηγουμένως. Για κάθε διαφορετικό σημείο που εξετάζεται λαμβάνονται τα 20 γειτονικά του σημεία, 10 στα αριστερά και 10 στα δεξιά του. Η γειτονιά αυτή των 20 σημείων δεν είναι τα σημεία του δειγματελιημμένου περιγράμματος, αλλά του αρχικού. Στα 21 συνολικά σημεία παρεμβάλλεται μία γραμμή που ελαχιστοποιεί την τετραγωνική τους απόσταση από αυτή, σύμφωνα με την μέθοδο των ελαχίστων τετραγώνων. Τα 21 σημεία λοιπόν προσεγγίζονται με μία γραμμή που περνά από το δειγματελιημμένο σημείο. Για την ευθεία αυτή λαμβάνεται το διάνυσμα κατεύθυνσης της και εξετάζονται τα διανύσματα που σχηματίζουν γωνία 90° με αυτό και βρίσκονται στο εσωτερικό του περιγράμματος. Τα κάθετα διανύσματα είναι πολύ σημαντικά, καθώς στην κατεύθυνσή τους γίνεται η τελική τοποθέτηση του χεριού του ρομπότ. Ξεκινώντας λοιπόν από το εκάστοτε δειγματελιημμένο σημείο και προχωρώντας στην κατεύθυνση του κάθετου διανύσματος, ευρίσκεται το σημείο τομής με το περίγραμμα. Η διαδικασία αυτή επαναλαμβάνεται για διανύσματα που ξεκινούν από το δειγματελιημμένο σημείο και σχηματίζουν γωνίες από -35° έως 35° με το κάθετο διάνυσμα. Συνεπώς προκύπτουν τα σημεία του περιγράμματος για ένα εύρος 70° και υπολογίζονται οι αποστάσεις τους από το δειγματελιημμένο σημείο και κανονικοποιούνται ως προς τη μέγιστη απόσταση. Για κάθε σημείο λαμβάνεται ο περιγραφέας του, η γωνία που σχηματίζει με το κάθετο διάνυσμα, οι συντεταγμένες του και οι συντεταγμένες του απέναντι σημείου του, δηλαδή του σημείου τομής του περιγράμματος με το κάθετο διάνυσμα. Οι περιγραφείς χρησιμεύουν ως διάνυσμα εισόδου στο SVM, ενώ τα απέναντι σημεία χρησιμεύουν στην εύρεση της ευθείας στην οποία θα τοποθετηθεί το χέρι του ρομπότ.

BHMA 3: Αναγνώριση περιγραφέων και εξαγωγή 2Δ σημείων λαβής

Εφ' όσον έχουν υπολογιστεί οι περιγραφείς για κάθε δειγματελιημμένο σημείο του περιγράμματος, πρέπει να τροφοδοτηθούν στο SVM του ρομπότ ώστε αυτό να αποφανθεί για το εάν αυτά αποτελούν ικανοποιητικά σημεία σχηματισμού λαβής ή όχι. Το SVM που χρησιμοποιήθηκε είναι μη γραμμικό, με πολυωνυμικό πυρήνα και εκπαιδεύεται με τα 4 αρχικά σημεία εκπαίδευσης, σε συμφωνία με την ανάλυση του Κεφαλαίου 5. Τα σημεία που προκύπτουν έχουν κάθετο διάνυσμα που σχηματίζει γωνία από 90° έως 180° με τον

άξονα x της εικόνας. Έτσι, απορρίπτονται τα σημεία εκείνα που το ρομπότ δεν μπορεί να φτάσει χρησιμοποιώντας το δεξί του χέρι. Τονίζεται πως ο περιορισμός αυτός ελήφθη καθαρά για ευκολία στην εκτέλεση της πειραματικής διαδικασίας. Ανάλογα με τη γωνία της καθέτου, μπορεί να χρησιμοποιηθεί και το αριστερό χέρι, ώστε να εξεταστούν περισσότερα σημεία στο περίγραμμα. Το αποτέλεσμα είναι μία εικόνα της μορφής της Εικόνας 23, με τα σημεία που ταξινομήθηκαν ως ικανοποιητικά για λαβή να σημειώνονται με πράσινο χρώμα στην εικόνα και τιμή 1 στην κλάση ταξινόμησης, ενώ τα σημεία που ταξινομήθηκαν ως μη ικανοποιητικά σημειώνονται με κόκκινο χρώμα στην εικόνα και με τιμή 0 στην κλάση ταξινόμησης.

Τα σημεία που προέκυψαν από την αναγνώριση υφίστανται περαιτέρω επεξεργασία. Σε πρώτη φάση, τα σημεία εξετάζονται έτσι ώστε να εφαρμοστεί σε αυτά μια διαδικασία εξάλειψης αυτών που βρίσκονται ανάμεσα σε πλήθος σημείου αντίθετης τιμής. Συγκεκριμένα ακολουθώντας ανθρωπολογική φορά, εάν κάποιο σημείο βρίσκεται ανάμεσα σε 2 (1 δεξιά και 1 αριστερά του) σημεία αντίθετης τιμής, τότε η τιμή του αλλάζει. Με αυτόν τον τρόπο μπορούν να εξαλειφονται τυχόν ατέλειες της διαδικασίας αναγνώρισης και να ομαλοποιείται περισσότερο η τελική μορφολογία του περιγράμματος. Το αποτέλεσμα είναι διακριτές περιοχές για εκτέλεση ή μη λαβής επάνω στο περίγραμμα, που έχουν ως ελάχιστο πλήθος 3 σημεία.

Εφ' όσον κάθε ικανοποιητικό σημείο μπορεί να αποτελέσει τελικό σημείο λαβής, δημιουργείται η ανάγκη περαιτέρω μείωσης του αριθμού τους επάνω στο περίγραμμα. Πράγματι, η ύπαρξη τόσων πολλών επιλογών λαβής, τόσο κοντά μεταξύ τους, δεν προσφέρει κάτι επιπρόσθετο στην πειραματική διαδικασία. Για το λόγο αυτό, εκλέγεται για κάθε γειτονιά ένα αντιπροσωπευτικό σημείο. Για κάθε διακριτή γειτονιά ικανοποιητικών σημείων, εκλέγεται ένα τυχαίο σημείο της γειτονιάς, ως σημείο λαβής. Με αυτόν τον τρόπο, θα προκύψουν τόσα τελικά σημεία, όσα και οι γειτονιές. Κάθε σημείο αναπαριστά διαφορετική λαβή, οδηγώντας έτσι σε ελάχιστο πλήθος λαβών, που αναπαριστά καλύτερα την κάθε γειτονιά.

ΒΗΜΑ 4: Σχηματισμός και εκτέλεση τελικής λαβής

Η τελική φάση του συστήματος είναι ο σχηματισμός και η εκτέλεση της λαβής. Στο τέλος του προηγούμενου βήματος προέκυψε ένας αριθμός σημείων για σχηματισμό λαβής. Στα σημεία αυτά γίνεται η τοποθέτηση του αντίχειρα του ρομπότ, έτσι ώστε η λαβή να περιέχει όλο το αντικείμενο. Το τελικό σημείο που θα εκτελεστεί η λαβή επιλέγεται τυχαία μέσα από τα τυχαία επιλεγμένα σημεία κάθε γειτονιάς.

Στο επιλεγμένο σημείο εφαρμόζεται ο αλγόριθμος που περιεγράφηκε στο Κεφάλαιο 3, για την εύρεση των 3Δ συντεταγμένων του. Αρχικά, τοποθετείται ένα μικρό φύλλο χαρτιού με εντυπωμένη σκακιέρα στο επίπεδο εργασίας. Με χρήση αλγορίθμου εύρεσης γωνιών, εκλέγονται οι 4 γωνίες της σκακιέρας και δίνονται σε αυτές 3Δ τιμές. Συγκεκριμένα, η επάνω αριστερά γωνία ορίζεται ως το σημείο αρχής $O_w = (0,0,0)$ του συστήματος συντεταγμένων του χώρου. Έχοντας γνωστό το μήκος των τετραγώνων της σκακιέρας, γίνεται ανάθεση 3Δ τιμών στις γωνίες της. Σημειώνεται πως η συνιστώσα Z για κάθε σημείο είναι 0, σε συμφωνία με την παραδοχή του Κεφαλαίου 3. Ακολουθώντας την ανάλυση της ενότητας 3.4, ευρίσκεται η πόζα (R,T) της κάμερας του ρομπότ από την αρχή του συστήματος συντεταγμένων O_w , η εξίσωση του επιπέδου σε συντεταγμένες κάμερας και η εξίσωση της ευθείας από το κέντρο της κάμερας στο σημείο προβολής του επιλεγμένου προς λαβή σημείου, επάνω στο επίπεδο της εικόνας. Η εύρεση των 3Δ συντεταγμένων γίνεται ως η τομή του επιπέδου με την ευθεία.

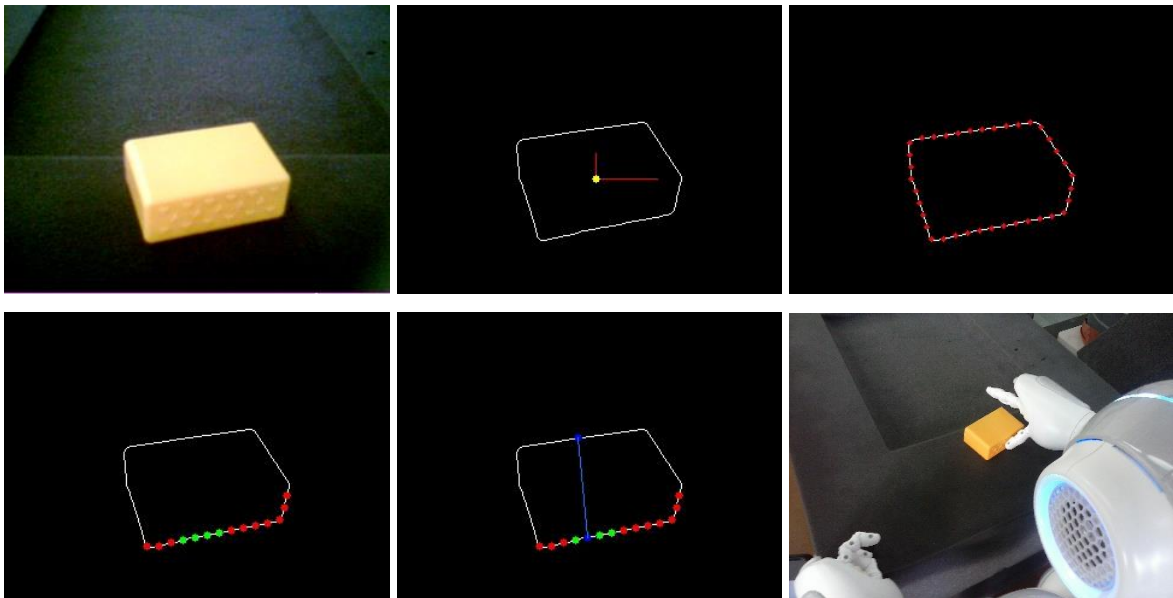
Μαζί με το επιλεγμένο σημείο, το ρομπότ υπολογίζει το σημείο τομής της καθέτου (όπως υπολογίστηκε στο Κεφάλαιο 2) με το περίγραμμα του αντικειμένου, καθώς επίσης και τη γωνία φ που αυτή σχηματίζει με τον άξονα της εικόνας. Για περισσότερη ευκολία, και επειδή το ρομπότ στην πειραματική διάταξη είναι στραμμένο κατά τέτοιο τρόπο έτσι ώστε να βλέπει το αντικείμενο από ψηλά, εξετάζεται μόνο η περιστροφή στον κάθετο προς το επίπεδο εργασίας άξονα. Η περιστροφή αυτή, αντιστοιχεί στην γωνία φ .

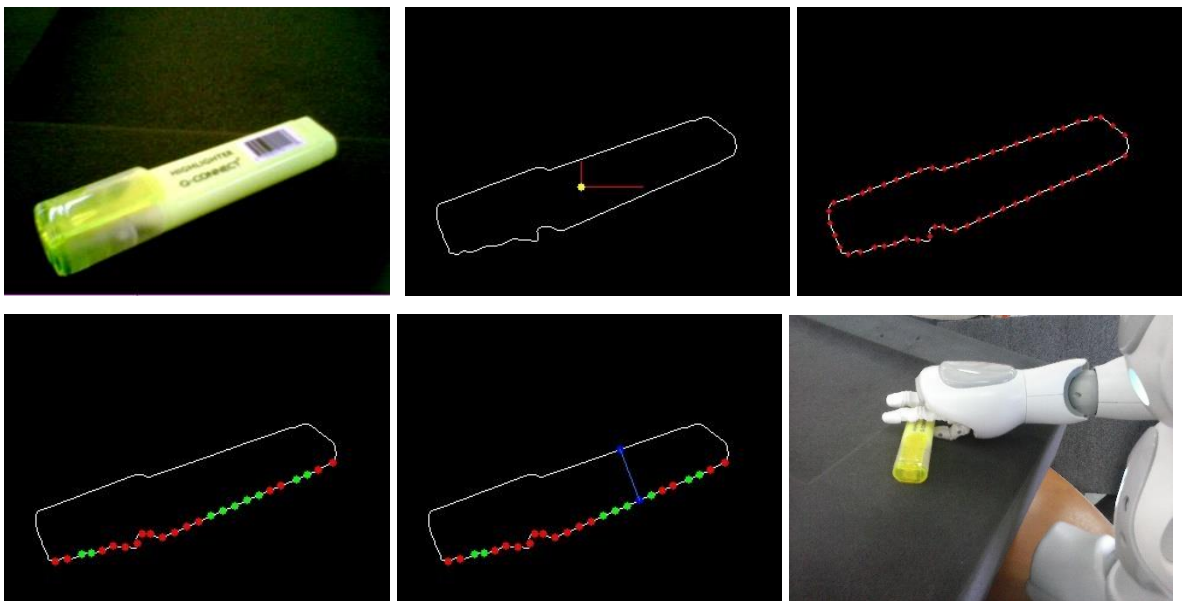
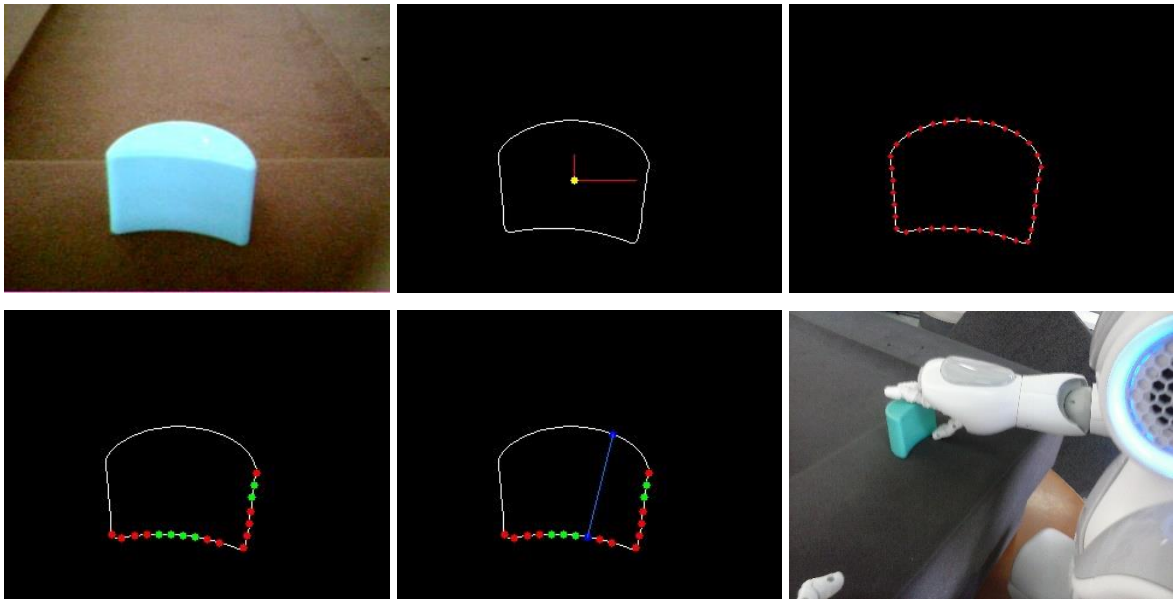
Έχοντας υπ' όψη τα παραπάνω, τελικά το ρομπότ χρησιμοποιεί ένα 3Δ σημείο και έναν προσανατολισμό για την δημιουργία της λαβής. Το αποτέλεσμα αυτό τροφοδοτείται στις εξισώσεις ανάστροφου διαφορικού προβλήματος του ρομπότ, για την εύρεση των γωνιών που πρέπει να σχηματίζουν οι αρθρώσεις, έτσι ώστε να καταστεί δυνατή η εκτέλεση της

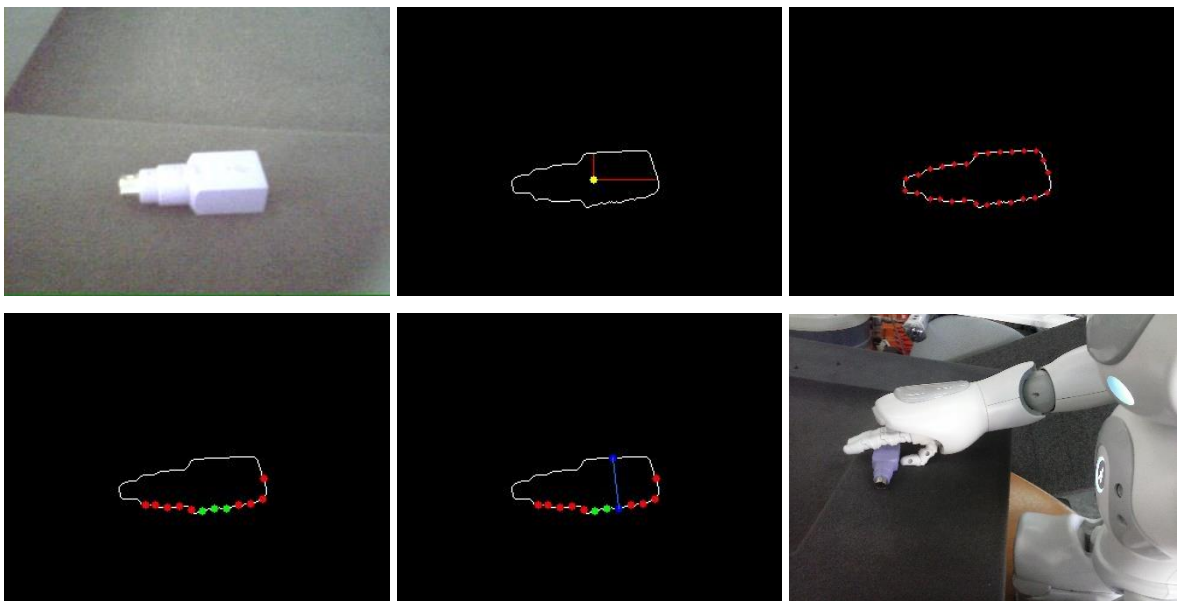
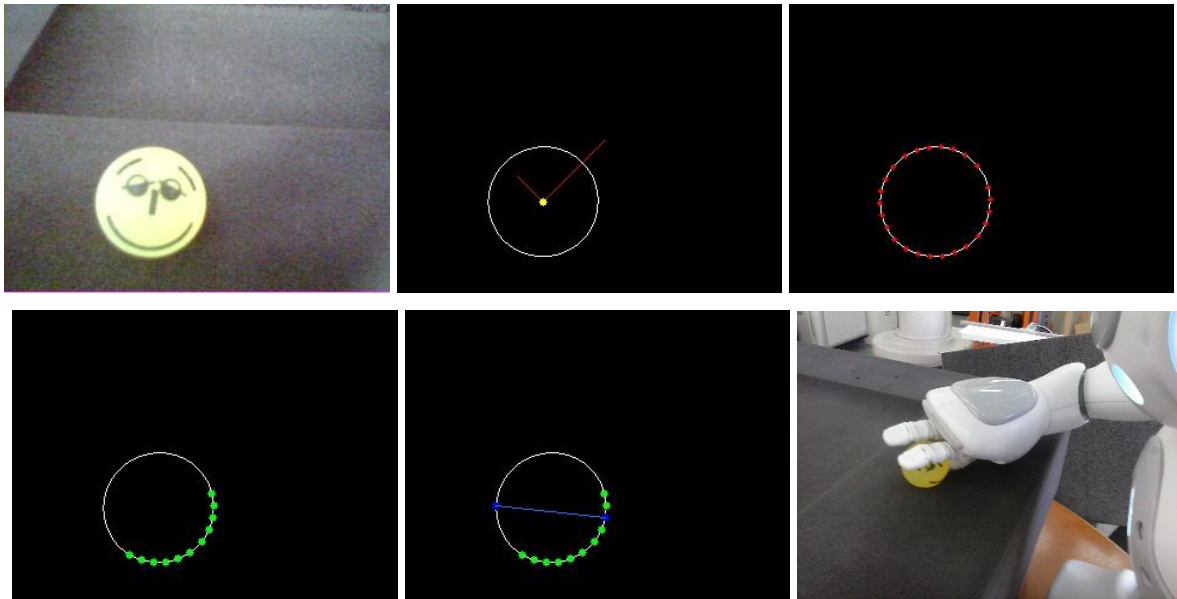
λαβής. Τελικό βήμα είναι η εκτέλεση της λαβής, όπου τοποθετείται ο αντίχειρας του ρομπότ στο επιλεγμένο σημείο λαβής, ενώ η υπόλοιπη παλάμη έχει προσανατολισμό φ .

6.6 Πειραματική εκτέλεση για διάφορα αντικείμενα

Στην ενότητα αυτή παρουσιάζονται τα αποτελέσματα της εκτέλεσης του συνολικού πειράματος μέσα από εικόνες που δημιούργησε ενδιάμεσα το ρομπότ, καθώς και από φωτογραφίες που ελήφθησαν εξωτερικά κατά την εκτέλεση.







Είναι φανερή η σωστή λειτουργία του αλγορίθμου, ο οποίος είναι σε θέση να βρει σημεία λαβής επάνω στο περίγραμμα και να οδηγήσει το χέρι επιτυχώς σε αυτά, με σωστό προσανατολισμό και στο σωστό 3Δ σημείο, με σφάλμα της τάξεως του χιλιοστού.

ΚΕΦΑΛΑΙΟ 7

7 *Συμπεράσματα και μελλοντικές επεκτάσεις της εργασίας*

7.1 *Συμπεράσματα*

Στην παρούσα διπλωματικής εργασία δημιουργήθηκε σύστημα αναγνώρισης αντικειμένων και εκτέλεσης λαβής σε αυτά, με βάση την μορφολογία του περιγράμματός τους, των τοπικών γεωμετρικών χαρακτηριστικών του και τις οπτικές πληροφορίες από την κάμερα του ρομπότ, κρίθηκε η αποτελεσματικότητα του αλγορίθμου μέσα από συγκρίσεις με διαφορετικά σχήματα και εκτελέστηκε μία σειρά πειραμάτων εκτέλεσης λαβής σε πραγματικά, καθημερινά αντικείμενα. Το ρομποτικό σύστημα χρησιμοποιεί οπτικά δεδομένα από την κάμερα και αλγορίθμους επεξεργασίας εικόνας, καθώς επίσης και γεωμετρικές και αλγεβρικές τεχνικές για την εξαγωγή πληροφορίας για το περίγραμμα. Η πληροφορία αυτή τροφοδοτείται σε ένα δίκτυο SVM , το οποίο αποφαινεται για την καταλληλότητα ή μη των σημείων του περιγράμματος ως προς τον σχηματισμό λαβής. Στη συνέχεια, με χρήση της κάμερας αποκτάται τρισδιάστατη πληροφορία, για να είναι ακριβέστερος ο σχηματισμός της τελικής λαβής. Τέλος, με χρήση τεχνικών από την επιστήμη της ρομποτικής, το ρομπότ είναι σε θέση να εκτελεί πραγματικές λαβές σε δεδομένα αντικείμενα.

Η εικόνα που παρέχεται στο σύστημα είναι στη μορφή RGB και μετατρέπεται σε εικόνα grayscale. Ο αλγόριθμος κατάτμησης βασίζεται στην αφαίρεση δύο εικόνων στον grayscale χώρο, είναι πολύ απλός και χρησιμεύει στην περίπτωση στατικής εικόνας, όπως είναι και στην παρούσα διπλωματική. Με χρήση φίλτρων και τεχνικών επεξεργασίας εικόνας, παράγεται ένα ικανοποιητικό αποτέλεσμα για την συγκεκριμένη περίπτωση. Βελτιωμένοι και πιο αποτελεσματικοί αλγόριθμοι θα μπορούσαν να οδηγήσουν σε καλύτερο αποτέλεσμα και να χρησιμοποιηθούν σε περιπτώσεις αλλαγής του περιβάλλοντος, όπως επίδραση στον φωτισμό, μετακίνηση της κάμερας και άλλες.

Κύριος σκοπός της εργασίας είναι η διερεύνηση της αποτελεσματικότητας του τοπικού ακτινικού περιγραφέα ως διάνυσμα εκπαίδευσης και η εξαγωγή συμπερασμάτων από τη χρησιμοποίησή του. Ο περιγραφέας αποτελεί μία συναρτησιακή, μονοδιάστατη απεικόνιση της τοπικής μορφολογίας του αντικειμένου, ιδιαίτερα εύχρηστη για την αναγνώριση αντικειμένων. Συνεπώς, η χρησιμοποίησή του ως είσοδο στο SVM, επιτρέπει στο ρομπότ να εκπαιδεύεται απ' ευθείας στις τοπικές ιδιότητες ενός αντικειμένου και να εξάγει χρήσιμες πληροφορίες για αυτό. Επιπλέον, ο περιγραφέας είναι αναλλοίωτος ως προς τη μετατόπιση, την περιστροφή και την κλιμάκωση του αντικειμένου, παρέχοντας έτσι μία ακόμα πιο εύρωστη αναπαράσταση της μορφολογίας του αντικειμένου. Η προσέγγιση αυτή καθιστά το σύστημα πιο ρεαλιστικό, με ικανότητα για εφαρμογή σε πραγματικά αντικείμενα, όπως και έγινε.

Με τη χρησιμοποίηση του τοπικού περιγραφέα, είναι δυνατή η αναγνώριση μορφολογίας απλών αλλά και πιο σύνθετων σχημάτων, χωρίς να χρειάζεται κάποιος αλγόριθμος κατάτμησης σχήματος. Αυτό συμβαίνει διότι ο τοπικός περιγραφέας ξεκινά από ένα σημείο στο περίγραμμα και εξετάζει μία περιοχή απέναντι από αυτό το σημείο. Έτσι, μπορεί να ανιχνεύει προεξοχές, κοιλότητες και άλλες ιδιομορφίες στα αντικείμενα. Πρέπει να σημειωθεί πως η χρήση του περιγραφέα με τη μορφή που δημιουργείται στην παρούσα εργασία, αποκλείει κάποιες λαβές σε σχήματα που έχουν παραπάνω από μία τιμή r σε κάποια δεδομένη γωνία θ . Για παράδειγμα, έστω το παρακάτω σχήμα



Εικόνα 25: Ιδιόμορφο σχήμα

Εάν ο περιγραφέας ξεκινήσει από το κόκκινο σημείο σημείο κάτω αριστερά στο σχήμα και επεκταθεί προς τα δεξιά, θα εξετάσει την αμέσως δεξιά πλευρά του αντικειμένου

που σημειώνεται με πράσινο χρώμα. Όμως, μία ικανοποιητική λαβή για το αντικείμενο θα μπορούσε να είναι αυτή που ξεκινά από το κόκκινο σημείο και καταλήγει στην γαλάζια πλευρά και περιέχει ολόκληρο το αντικείμενο. Με ελαφριά τροποποίηση της μεθόδου εξαγωγής περιγραφέα, μπορούν να βρεθούν παραπάνω από ένας περιγραφείς για κάθε σημείο εφ' όσον υπάρχουν και να δίνονται όλοι στο δίκτυο για εκπαίδευση και αναγνώριση.

Η τρισδιάστατη πληροφορία που λαμβάνεται είναι αποτέλεσμα αλγορίθμων γραμμικής άλγεβρας και όρασης υπολογιστών. Η παραδοχή επίπεδου εργασίας με εξίσωση $Z = 0$ επιτρέπει την λήψη της πληροφορίας με αρκετά ακριβή αποτελέσματα. Σε πραγματικές συνθήκες όμως, η παραδοχή αυτή δεν είναι ισχυρή, καθώς η Z συνιστώσα μπορεί να μεταβάλλεται ελεύθερα. Για εφαρμογές στον πραγματικό κόσμο πρέπει να αναζητηθούν άλλοι αλγόριθμοι εξαγωγής 3D πληροφορίας, όπως τριγωνοποίηση, ανόρθωση εικόνας και άλλοι. Παρ' όλα αυτά σε εργοστασιακές ή οικιακές συνθήκες, υπάρχουν περιπτώσεις που ένα ρομπότ μπορεί να κληθεί να αλληλεπιδράσει με αντικείμενα σε κάποιο επίπεδο, οπότε ο αλγόριθμος θα μπορούσε να χρησιμοποιηθεί.

7.1 Βελτιώσεις και μελλοντικές επεκτάσεις

Οι βελτιώσεις της παρούσας εργασίας μπορούν να ξεκινήσουν από την βελτίωση του αλγορίθμου κατάτμησης εικόνας, ώστε αυτός να λειτουργεί ικανοποιητικά χωρίς να είναι απαραίτητο οι εικόνες να είναι στατικές. Τέτοιες βελτιώσεις περιλαμβάνουν την χρησιμοποίηση χρωματικής πληροφορίας για κατάτμηση, καθώς επίσης και πληροφορίας βάθους του αντικειμένου, η οποία θα έχει αποκτηθεί από 3D αισθητήρες. Επιπρόσθετα, μπορούν να χρησιμοποιηθούν πιο αποτελεσματικοί αλγόριθμοι, όπως αυτοί που παρουσιάστηκαν στο Κεφάλαιο 2. Η υπολογιστική πολυπλοκότητα μπορεί να είναι μεγαλύτερη, αλλά τα αποτελέσματα θα είναι πιο εύρωστα σε θόρυβο και εξωτερικούς παράγοντες.

Μία πιθανή βελτίωση θα μπορούσε να είναι η χρήση πολυπλοκότερου αλγορίθμου εξαγωγής 3D οπτικής πληροφορίας για την εκτέλεση των πειραμάτων. Με λήψη διαφορετικών εικόνων από διαφορετικούς προσανατολισμούς για το αντικείμενο και χρησιμοποιώντας ανόρθωση εικόνας, αντιστοίχιση σημείων μεταξύ εικόνων και

τριγωνοποίηση, είναι δυνατό να κατασκευαστεί ένας 3Δ χάρτης που θα παρουσιάζει τις 3Δ συντεταγμένες κάθε σημείου του αντικειμένου στο χώρο. Αυτές οι συντεταγμένες μπορούν εύκολα να μετασχηματιστούν ως προς το πλαίσιο της κάμερας, επιτρέποντας έτσι στο ρομπότ τη γνώση των συντεταγμένων κάθε σημείου της επιφάνειας του αντικειμένου.

Στην εργασία αυτή δεν χρησιμοποιήθηκε ο ολικός ακτινικός περιγραφέας ως διάλυμα εκπαίδευσης, παρά μόνο ως μέσο πλοήγησης στο περίγραμμα του σχήματος. Ο ολικός περιγραφέας περιέχει πληροφορίες για τη συμμετρία του αντικειμένου και για την μορφολογία του. Πράγματι, η συνάρτηση του περιγραφέα εάν υποστεί μετατόπιση στον άξονα x και το αποτέλεσμα είναι μία άρτια συνάρτηση, συνεπάγεται την ύπαρξη σημείου συμμετρίας. Μελετώντας λοιπόν την αρτιότητα της συνάρτησης, μπορούν να εξαχθούν σημεία κοντά στο μέσο του αντικειμένου, ασφαλή για εκτέλεση λαβής. Ταυτόχρονα, η κλίση της συνάρτησης υποδηλώνει την ύπαρξη προεξοχών και απότομης μορφολογίας στο σχήμα. Τα σημεία στα οποία η συνάρτηση έχει σταθερή κλίση είναι σημεία που αντιστοιχούν σε ομαλό περίγραμμα. Πιθανά σταθερά σημεία για εκτέλεση λαβής θα μπορούσαν να εκλεγούν εκείνα τα οποία ικανοποιούν τις συνθήκες αρτιότητας και σταθερής κλίσης.

Ο τοπικός ακτινικός περιγραφέας που χρησιμοποιήθηκε στην εργασία, λαμβάνει πληροφορία για το 2Δ περίγραμμα του αντικειμένου και την χρησιμοποιεί τόσο σαν είσοδο σε SVM, όσο και σαν μέσο καθοδήγησης του χεριού του ρομπότ. Αυτή η προσέγγιση μπορεί να επεκταθεί στον 3Δ χώρο, εκπαιδεύοντας το ρομπότ με τη βοήθεια μίας 2Δ συνάρτησης περιγραφέα αντί της 1Δ συνάρτησης που χρησιμοποιήθηκε εδώ. Διαθέτοντας ένα νέφος 3Δ σημείων, και ξεκινώντας από το κεντροειδές τους, μπορεί να χαρτογραφηθεί η επιφάνεια του αντικειμένου με τη βοήθεια μιας 2Δ συνάρτησης $r - \theta$ σε αναλογία με την 1Δ συνάρτηση της παρούσας εργασίας. Το αποτέλεσμα θα είναι μία 2Δ διακριτή συνάρτηση, ή ένας πίνακας που θα περιέχει πληροφορίες για την 3Δ μορφολογία. Ο πίνακας αυτός μπορεί να αναπαρασταθεί με τεχνικές του Κεφαλαίου 2, όπως περιγραφείς Fourier ή ροπές εικόνας, για την μείωση της διαστατικότητας των δεδομένων και την ευκολότερη επεξεργασία τους.

Τέλος, επισημαίνεται η ανάγκη φορητότητας του αλγορίθμου για εκτέλεση σε διαφορετικά ρομποτικά συστήματα. Κάθε διαφορετικό ρομπότ παρέχει μεγάλη ποικιλία από

αισθητήρες και διαφορετικά στοιχεία δράσης, καθιστώντας την εκμετάλλευσή τους απαραίτητη για την μεγιστοποίηση της αποδοτικότητας του αλγορίθμου.

BIBΛΙΟΓΡΑΦΙΑ

Adan. A., Vasquez A., Merchan P., Heradio R. (2011): “Direction kernels: using a simplified 3D model representation for grasping”, *Machine Vision and Applications 24*, Springer-Verlag.

Bohg J., Kragic D. (2009): “Learning grasping points with shape context”, *Robotics and Autonomous Systems 58*, Elsevier.

Denavit J., Hartenberg R. (1964). *Kinematic synthesis of linkages*, McGraw-Hill.

J. Felip., J. Laakonsen, A. Morales, V. Kyrki (2012): “Manipulation primitives: a paradigm for abstraction and execution of grasping and manipulation tasks”, *Robotics and Autonomous Systems 61*, Elsevier.

Flusser J., Suk T., Zitova B. (2009). *Moments and moment invariants in pattern recognition*, Wiley.

Gonzalez R., Woods R. (2008). *Digital Image Processing*, Pearson Education.

Hartley R., Zisserman A. (2003). *Multiple view geometry in computer vision*, Cambridge University Press.

Haykin S. (2009). *Neural networks and machine learning*, Third Edition, Pearson Education.

Hu M. (1962). “Visual pattern recognition by moment invariants”, *IRE Transaction on Information Theory*, vol. 8, pp. 179-187.

Palm R., Iliev B., Bourhane K. (2008). “Recognition of human grasps by time-clustering and fuzzy modeling”, *Robotics and Autonomous Systems 57*, Elsevier.

Platt R. (2006): “Learning and generalizing control-based grasping and manipulation skills”, *University of Massachusetts Amherst*, Phd thesis.

Saedan M., Marcelo H. Ang Jr. (2001): “3D vision based control on an industrial robot”, *International Conference on Robotics and Applications 2001, Florida USA*.

Shapiro L., Stockman G. (2001). *Computer vision*, Prentice Hall.

Siciliano B., Sciavicco L., Villani L., Oriolo G. (2009). *Robotics: Modelling, planning and control*, Springer.

B. Wang, J. Liang, J.W Li, H.G. Kai, H. Liu (2005): “Grasping unknown objects based on 3D model reconstruction”, *International Conference of Advanced Intelligent Mechatronics 2005*.

Zhang Z. (2000). “A flexible new technique for camera calibration”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330-1334.

Αληφραγκής Μ. (2010): “Εικονική ανακατασκευή και ταυτοποίηση αντικειμένων στο χώρο με τεχνικές υπολογιστικής όρασης και τρισδιάστατα γραφικά”, Διδακτορική Διατριβή, Ε.Μ.Π.

Μαραγκός Π. (2005), *Όραση Υπολογιστών*, Εκδόσεις Ε.Μ.Π.

Τζαφέστας Σ. (2003) . *Ρομποτική: Ανάλυση – Έλεγχος – Σχεδιασμός – Προγραμματισμός – Αίσθηση*, Εκδόσεις Ε.Μ.Π.