



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

**Εκπαίδευση Ρομποτικού Βραχίονα στην Οπτική Αναγνώριση
Επιτρεπτών Λαβών Σύνθετων Αντικειμένων σε Επίπεδο μέσω
Επιβλεπόμενης Μάθησης Αυτο-οργανούμενου Χάρτη**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεώργιος Κ. Δαριβιανάκης

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

Αθήνα, Νοέμβριος 2011



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

**Εκπαίδευση Ρομποτικού Βραχίονα στην Οπτική Αναγνώριση
Επιτρεπτών Λαβών Σύνθετων Αντικειμένων σε Επίπεδο μέσω
Επιβλεπόμενης Μάθησης Αυτο-οργανούμενου Χάρτη**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεώργιος Κ. Δαριβιανάκης

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

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3^η Νοεμβρίου 2011

.....
Κ. Τζαφέστας
Επίκουρος Καθηγητής Ε.Μ.Π.

.....
Π. Μαραγκός
Καθηγητής Ε.Μ.Π.

.....
Ν. Μαράτος
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2011

.....
Γεώργιος Κ. Δαριβιανάκης

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

Copyright © Γεώργιος Κ. Δαριβιανάκης, 2011.

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

Περίληψη

Η επιστήμη της Ρομποτικής έχει ως στόχο να δημιουργήσει «σκεπτόμενες» μηχανές, τα ρομπότ, οι οποίες θα μπορούν να αλληλεπιδρούν με το περιβάλλον τους με απώτερο σκοπό να διευκολύνουν τον άνθρωπο στις διάφορες εργασίες που έχει να επιτελέσει. Απαραίτητη προϋπόθεση προκειμένου να συμβεί αυτό είναι τα ρομπότ να μπορούν να χειρίζονται με ιδιαίτερη επιδεξιότητα τα αντικείμενα που βρίσκονται στο περιβάλλον τους. Πεδίο μελέτης της παρούσας διπλωματικής εργασίας αποτελεί η ανάπτυξη ενός ρομποτικού συστήματος το οποίο θα εκπαιδεύεται ώστε να αποκτήσει την δεξιότητα της οπτικής αναγνώρισης των εφικτών λαβών των αντικειμένων. Η εκπαίδευση του ρομποτικού συστήματος υλοποιείται μέσω της αλληλεπίδρασης αυτού με τον άνθρωπο. Ο εκπαιδευτής επιδεικνύει στο ρομπότ τον τρόπο λαβής ορισμένων πρωταρχικών αντικειμένων (επιβλεπόμενη μάθηση) και το ρομποτικό σύστημα συνδυάζοντας αλγορίθμους επεξεργασίας εικόνας και διατάξεων νευρωνικών δικτύων αναπτύσσει την ικανότητα της εύρεσης των επιτρεπτών λαβών σύνθετων αντικειμένων.

Η διπλωματική εργασία μπορεί να χωριστεί σε δυο ουσιώδη μέρη: (α) την θεωρητική ανάλυση και (β) την υλοποίηση του πειράματος. Στο πρώτο μέρος, παρουσιάζεται εκτενώς το θεωρητικό υπόβαθρο στο οποίο στηριχθήκαμε ώστε να επιλύσουμε το πρόβλημά μας. Επικεντρώνασθε κυρίως στην θεωρία της εξαγωγής οπτικών χαρακτηριστικών που σχετίζονται με πληροφορία σχήματος, όπως οι ροπές εικόνας, οι αναλλοίωτες μεταβλητές ροπών και η ανάλυση καμπυλότητας. Παρουσιάζεται η μεθοδολογία της κατάτμησης σχήματος και της στερεοοπτικής όρασης καθώς και βασικές τεχνικές βαθμονόμησης κάμερας. Ιδιαίτερη έμφαση δίνεται στην θεωρία που διέπει τους αυτοοργανούμενους χάρτες, καθώς αποτελούν την διάταξη νευρωνικού δικτύου που χρησιμοποιήθηκε για να εκπαιδευτεί ο ρομποτικός χειριστής. Στο δεύτερο μέρος, αναλύεται η διαδικασία που ακολουθήθηκε για την εκπαίδευση του ρομποτικού χειριστή, όπου αναδεικνύονται τα πλεονεκτήματα της χρήσης αυτοοργανούμενου χάρτη καθώς το σύστημα έχει τη δυνατότητα να εκπαιδευτεί αποτελεσματικά με χρήση μικρού αντιπροσωπευτικού συνόλου εκπαίδευσης. Επίσης, επεξηγείται εκτενώς η μεθοδολογία που εφαρμόστηκε προκειμένου το ρομπότ να είναι σε θέση να εντοπίζει το σύνολο των επιτρεπτών λαβών του αντικειμένου. Τέλος, παρουσιάζονται στιγμιότυπα της υλοποίησης του πειράματος σε περιβάλλον προσομοίωσης Webots, χρησιμοποιώντας τον ρομποτικό βραχίονα Katana 6M180. Τα αποτελέσματα από την προσομοίωση δείχνουν ότι ο ρομποτικός χειριστής ανταποκρίνεται ιδιαιτέρως ικανοποιητικά στο πρόβλημα του εντοπισμού των επιτρεπτών λαβών σύνθετου σχήματος αντικειμένου που είναι τοποθετημένο πάνω σε οριζόντιο τραπέζι.

Λέξεις Κλειδιά: Εύρεση επιτρεπτών λαβών, κατάτμηση σχήματος, στερεοοπτική όραση, αναλλοίωτες μεταβλητές ροπών, αυτοοργανούμενος χάρτης, Webots, Katana.

Abstract

The goal of Robotics science is to create “thinking” machines, that is, robots that will be capable of interacting with their environment with a view to facilitate humans in various tasks. A necessary condition for this to happen is the robot to be able to handle with great skill various objects in its environment. The main goal of this diploma thesis is to develop a robotic system that will be trained to acquire the skill associated to the visual identification of feasible grasp configurations of objects. Training of the robotic system can be performed through human-robot interaction. The human instructor can teach the system how to handle certain primitive objects (in a supervised learning context) and, by combining image processing and neural network learning algorithms, the robotic system can then develop the capability of finding feasible grasp configurations for more complex objects.

This diploma thesis is divided into two main parts: (a) the theoretical analysis and (b) the experimental implementation and assessment. The first part presents in detail the theoretical background used to solve the problem. We focus mainly on the theory of extracting visual features that are related to shape information, like image moments, image moment invariants and curvature. Moreover, stereopsis and camera calibration techniques are employed in this work. Particular emphasis is given to the theory governing self-organizing maps (SOMs), as they constitute the neural network architecture that we used to achieve learning of feasible grasps based on visual features. The second part of this thesis deals with the procedure followed to train the robot manipulator. The advantages of using SOMs are highlighted, enabling the system to achieve good learning performance while using a small representative set of training examples. Moreover, we introduce the methodology used to enable the robot to identify the set of feasible grasp configurations of visible objects. Experimental implementation is performed on a Webots simulation environment, and results are obtained using a robotic arm Katana 6M180. The results presented in this thesis show that the robot manipulator achieves excellent learning performance, responding very well to the problem of identifying feasible grasps configurations of new complex objects mounted on a horizontal table.

Keywords: grasp planning, feasible grasp configurations, stereopsis, image moment invariants, self-organized maps, Webots, Katana

Ευχαριστίες

Πρώτα και κύρια θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Κωνσταντίνο Τζαφέστα που μου έδωσε την ευκαιρία μέσω της ανάθεσης της παρούσας διπλωματικής εργασίας να ασχοληθώ με ένα ιδιαίτερος ενδιαφέρον και σύγχρονο θέμα της ρομποτικής επιστήμης. Επιπλέον του είμαι ευγνώμων για τον χρόνο που μου αφιέρωσε και για τις αρμονικές συνθήκες συνεργασίας που εξασφάλισε κατά την εκπόνηση της εργασίας μου αυτής.

Ακόμα θα ήθελα να ευχαριστήσω την οικογένειά μου, του γονείς μου Κώστα και Γεωργία, που στάθηκαν συμπαραστάτες καθόλη τη διάρκεια των φοιτητικών μου χρόνων και μου παρείχαν κάθε δυνατότητα και άνεση ώστε να αφοσιωθώ απερίσπαστος στην ενασχόληση με τις σπουδές μου αλλά και την αδερφή μου Μαρία για την ηθική στήριξή της τα τελευταία δύο χρόνια της συγκατοίκησής μας.

Τέλος, ένα ευχαριστώ αξίζει στη Βίλη για την γλυκιά της παρουσία δίπλα μου στις εύκολες αλλά και στις δύσκολες στιγμές.

Περιεχόμενα

1	Εισαγωγή.....	15
1.1	Ρομποτική.....	15
1.2	Ρομποτικό σύστημα.....	16
1.3	Αλληλεπίδραση Ανθρώπου – Ρομπότ.....	18
1.4	Εκπαίδευση ρομπότ από τον άνθρωπο.....	18
1.5	Σύνοψη παρούσης διπλωματικής εργασίας.....	20
2	Ροπές Εικόνας και Αναλλοίωτες Μεταβλητές Ροπών.....	25
2.1	Εισαγωγή.....	25
2.2	Κατηγορίες αναλλοίωτων μεταβλητών.....	26
2.3	Ροπές Εικόνας.....	28
2.3.1	Γεωμετρικές και Μιγαδικές Ροπές.....	28
2.4	Αναλλοίωτες Μεταβλητές Ροπών ως προς την μετατόπιση, την περιστροφή και την κλιμάκωση.....	30
2.4.1	Αναλλοίωτες μεταβλητές ως προς την μετατόπιση.....	31
2.4.2	Αναλλοίωτες μεταβλητές ως προς την ομοιόμορφη κλιμάκωση.....	32
2.4.3	Αναλλοίωτες μεταβλητές H_u	33
2.5	Αφινικός Μετασχηματισμός.....	34
2.6	Αναλλοίωτες μεταβλητές ως προς αφινικό μετασχηματισμό (Affine Moment Invariants - AMIs).....	36
2.6.1	AMIs εξαγόμενα από γράφους.....	36
2.7	Αριθμητικά παραδείγματα.....	38
3	Τεχνητά Νευρωνικά Δίκτυα – Αυτοοργανούμενοι Χάρτες (SOM).....	41
3.1	Εισαγωγή.....	41
3.2	Αυτοοργανούμενοι Χάρτες.....	42
3.2.1	Η Διαδικασία Ανταγωνισμού.....	45
3.2.2	Διαδικασία Συνεργασίας.....	46
3.2.3	Διαδικασία Προσαρμογής.....	47
3.2.4	Οι Δύο Φάσεις της Προσαρμοστικής Διαδικασίας: Διάταξη και Σύγκλιση.....	49
3.3	Σύνοψη του Αλγορίθμου SOM.....	52

3.4	Αριθμητικά Παραδείγματα	53
3.5	Ομαδοποίηση Προτύπων Χρησιμοποιώντας SOM.....	56
3.6	Επιβλεπόμενο SOM (Supervised SOM)	57
4	Καμπυλότητα σε Χώρο Κλίμακας (Curvature Scale Space) – Κατάτμηση Σχήματος.....	59
4.1	Εισαγωγή.....	59
4.2	Καμπύλη.....	60
4.3	Ανάλυση Καμπυλότητας Επίπεδων Σχημάτων	61
4.4	Καμπυλότητα σε Χώρο Κλίμακας (Curvature Scale Space)	64
4.5	Κατάτμηση Σχήματος	69
5	Στερεοπτική Όραση και Βαθμονόμηση Κάμερας	75
5.1	Εισαγωγή.....	75
5.2	Κάμερα Οπής	76
5.3	Στερεοπτική Όραση.....	80
5.3.1	Επιπολική Γεωμετρία.....	81
5.3.2	Εύρεση Συντεταγμένων Σημείου - Τριγωνοποίηση	83
5.4	Βαθμονόμηση Κάμερας	86
6	Εκπαίδευση και Προσδιορισμός Παραμέτρων Αυτό-οργανούμενου Χάρτη (SOM).....	91
6.1	Εισαγωγή.....	91
6.2	Πρότυπα Εισόδου Εκπαίδευσης.....	92
6.3	Σύνολο Δεδομένων Ελέγχου	94
6.4	Ποσοστό επιτυχίας νευρωνικού δικτύου.....	95
6.5	Ορισμός κατωφλίων για την εύρεση του νικητή νευρώνα.....	96
6.6	Δοκιμές Εκπαίδευσης Αυτοοργανούμενου Χάρτη	100
6.7	Συμπεράσματα και Αλγόριθμος Εκπαίδευσης.....	105
7	Εντοπισμός επιτρεπτών λαβών τυχαίου σχήματος αντικειμένου.....	107
7.1	Εισαγωγή.....	107
7.2	Περιβάλλον Προσομοίωσης Webots	108
7.3	Ρομποτικός Βραχίονας Katana 6M180	109
7.4	Ορθή Κινηματική Ανάλυση	111
7.4.1	Μέθοδος Denavit - Hartenberg.....	111

7.4.2	Ρομποτικός Βραχίονας Katana 6M180	113
7.5	Ανάστροφη Κινηματική Ανάλυση	115
7.6	Διαδικασία εντοπισμού επιτρεπτών λαβών.....	121
7.7	Υλοποίηση πειράματος	128
8	Συμπεράσματα και Μελλοντικές κατευθύνσεις της έρευνας	131
8.1	Συμπεράσματα	131
8.2	Μελλοντικές κατευθύνσεις της έρευνας	134

Κατάλογος Σχημάτων

Σχήμα 1.1: Μπλοκ διάγραμμα των συστατικών μερών ενός ρομποτικού συστήματος	16
Σχήμα 1.2: Ανοικτή κινηματική αλυσίδα που αποτελείται από δυο περιστροφικές και μια πρισματική άρθρωση.	17
Σχήμα 1.3: Προσανατολισμός της αρπάγης κάθετος προς την επιφάνεια του πάγκου εργα-σίας.	21
Σχήμα 2.1: Δισδιάστατος χώρος με δυο διακριτές τάξεις. Οι αναλλοίωτες μεταβλητές διατηρούν κοντά τις τιμές τους ανεξάρτητα με τις παραμορφώσεις που γίνονται στα σχήματα (Flusser & Suk, 2009).	26
Σχήμα 2.2: Η επιθυμητή συμπεριφορά των ΜΠΚ αναλλοίωτων μεταβλητών ροπής. Σε όλες τις περιπτώσεις, η περιστραμμένη και κλιμακωμένη εικόνα έχει σχεδόν την ίδια τιμή με την αμετάβλητη αρχική εικόνα (Flusser & Suk, 2009).	31
Σχήμα 2.3: Ο αφινικός μετασχηματισμός επεικονίζει ένα τετράγωνο σε ένα παραλληλόγραμμα (Flusser & Suk, 2009).	35
Σχήμα 2.4: Οι χαρακτήρες που χρησιμοποιήθηκαν στο πείραμα. Στη 1 ^η γραμμή περιέχονται τα πρότυπα εκπαίδευσης, ενώ, στις υπόλοιπες άγνωστοι χαρακτήρες (Flusser & Suk, 1994).....	38
Σχήμα 3.1: Δισδιάστατο πλέγμα νευρώνων, με είσοδο τριών διαστάσεων και έξοδο 4x4 διαστάσεων.	43
Σχήμα 3.2: Εκθετική Χωρική Κατανομή προσομοιωμένη από μονοδιάστατο SOM πλέγμα-τος 10 νευρώνων.	54
Σχήμα 3.3: Χωρική Κατανομή Ρόμβου προσομοιωμένη από δισδιάστατο SOM πλέγματος 10x10.....	55
Σχήμα 3.4: Τυπικό παράδειγμα U-Matrix, για 6-διάστατα δεδομένα εκπαίδευσης που χωρίζονται σε δυο κλάσεις μεταξύ τους.	56
Σχήμα 4.1: Αρχική, τυχαία, επίπεδη συνεχής κλειστή καμπύλη	66
Σχήμα 4.2: Αποτύπωση των σημείων εναλλαγής πρόσημου πάνω στις ομαλοποιημένες καμπύλες και στο διάγραμμα του χώρου (p,σ).....	67
Σχήμα 4.3: Προφανείς τρόποι κατάτμησης φιγούρων. Ο (β) είναι ο μοναδικός τρόπος κατάτμησης του (α) ενώ υπάρχουν δυο διαφορετικοί προφανείς τρόποι (δ), (ε) για την κατάτμηση του (γ) (Seyranian & Hoffman, 1999).	69
Σχήμα 4.4: Γεωμετρία σχετιζόμενη με τον "Κανόνα Ελαχίστου".....	70
Σχήμα 4.5: Ο διαχωρισμός του σχήματος ενώνοντας διαδοχικά οριακά σημεία μπορεί να παράγει πολύ περίεργα αποτελέσματα όπως στο (α),(β), ενώ σε κάποιες	

περιπτώσεις όπως η (c) μπορεί να μην έχουμε καν καινούρια υποσχήματα (Seyranian & Hoffman, 1999).....	71
Σχήμα 4.6: Διαφορετικός τρόπος κατάτμησης για τον σταυρό (a), (b) αν και έχουν ίδιο αριθμό οριακών σημείων σε περίπου παρόμοιες θέσεις(Seyranian & Hoffman, 1999).	72
Σχήμα 4.7: Ο κανόνας ελαχίστου μας παρέχει μόνο ένα οριακό σημείο, ενώ το δεύτερο σημείο που πρέπει να ενωθεί έχει μηδενική καμπυλότητα (Seyranian & Hoffman, 1999).....	72
Σχήμα 4.8: Το συντομότερο κόψιμο δεν είναι πάντα το καλύτερο (x w), πρέπει επίσης να διασχίζει και ένα άξονα τοπικής συμμετρίας όπως το κόψιμο (x z) (Seyranian & Hoffman, 1999).....	73
Σχήμα 4.9: Αρχική εικόνα και τα υποσχήματα στα οποία διαχωρίζεται εφαρμόζοντας την μεθοδολογία που αναλύθηκε προηγουμένως.	73
Σχήμα 5.1: Προοπτική προβολή για κάμερα οπής, όπου f το εστιακό βάθος της κάμερας (Μαραγκός, 2005).	76
Σχήμα 5.2: : Στο επίπεδο της εικόνας (image plane) το αντικείμενο μας φαίνεται ανεστραμμένο.	77
Σχήμα 5.3: Προοπτική προβολή αντικειμένου χρησιμοποιώντας εικονικό επίπεδο εικόνας, πριν από την οπή.....	78
Σχήμα 5.4: : Επιπολική Γεωμετρία.....	82
Σχήμα 5.5: Παράλληλα και στοιχισμένα επίπεδα εικόνας για τις δυο κάμερες που χρησιμοποιούμε.	85
Σχήμα 6.1: Πρωτεύον άξονας σχήματος. (Πηγή: Εγχειρίδιο χρήσης Matlab)	93
Σχήμα 6.2: Ορθογώνιο Παραλληλόγραμμο και τετράγωνο. Αν και το ένα σχήμα αποτελεί μια κλιμάκωση του άλλου, ωστόσο οι επιτρεπτές λαβές τους διαφέρουν ...	93
Σχήμα 6.3: Βασικά ή Πρωταρχικά Σχήματα εκπαίδευσης	94
Σχήμα 6.4: Γενικεύσεις πρωταρχικών σχημάτων εκπαίδευσης.....	95
Σχήμα 6.5: Τραπεζίο και οριακές του διατάξεις που είναι το ορθογώνιο παραλληλόγραμμο και το τρίγωνο.....	97
Σχήμα 6.6: Βαθμός επιτυχίας μονοδιάστατου SOM το οποίο έχει εκπαιδευτεί χρησιμοποιώντας μόνο τα τέσσερα πρωταρχικά σχήματα.	101
Σχήμα 6.7: Μονοδιάστατο SOM, χρησιμοποιώντας ως σύνολο εκπαίδευσης μόνο εννέα αντιπροσωπευτικά δείγματα.	102
Σχήμα 6.8: Δισδιάστατο πλέγμα νευρώνων της μορφής [2,x] όπου x ο αριθμός των νευρώνων. Σύνολο εκπαίδευσης αποτελούμενο από 9 εικόνες.	103
Σχήμα 6.9: Δισδιάστατο πλέγμα νευρώνων της μορφής [4,x] όπου x ο αριθμός των νευρώνων. Σύνολο εκπαίδευσης αποτελούμενο από 9 εικόνες.	103

Σχήμα 6.10: Μονοδιάστατο SOM, χρησιμοποιώντας ως σύνολο εκπαίδευσης δεκαέξι εικόνες.....	104
Σχήμα 7.1: Εικονικό τρισδιάστατο περιβάλλον, όπου διακρίνεται ο ρομποτικός βραχίονας Katana 6M180	108
Σχήμα 7.2: Τεχνικά Χαρακτηριστικά του ρομποτικού βραχίονα Katana 6M180, όπως αυτά παρουσιάζονται στο εγχειρίδιο χρήσης της Neuronics AG.	109
Σχήμα 7.3: Οι πέντε βαθμοί ελευθερίας του ρομποτικού βραχίονα Katana 6M180.	110
Σχήμα 7.4: Χώρος εργασίας του ρομποτικού βραχίονα Katana 6M180.	110
Σχήμα 7.5: Μέθοδος Denavit - Hartenberg στην ορθή κινηματική ανάλυση.	111
Σχήμα 7.6: Σχηματική απεικόνιση του Katana 6M180, με καθορισμό των συντεταγμένων των πλαισίων των περιστροφικών αρθρώσεων και του τελικού στοιχείου δράσης.	113
Σχήμα 7.7: Γεωμετρικό ανάλογο για το σύστημα εξισώσεων 7.10 και 7.11.....	117
Σχήμα 7.8: Προσανατολισμός της αρπάγης κάθετος προς την επιφάνεια του πάγκου εργασίας.	120
Σχήμα 7.9: Διαχωρισμούς σύνθετου σχήματος σε υποσχήματα. Η λαβή κατά την διεύθυνση x είναι επιτρεπτή για κάθε υποσχήμα ξεχωριστά, ωστόσο, αδύνατη για το συνολικό σχήμα.	125
Σχήμα 7.10: Μη επιτρεπτή λαβή καθώς υπάρχει τομή μεταξύ της αναπαράστασης της αρπάγης και του ελλειπούς σχήματος	126
Σχήμα 7.11: Η λαβή για τον άξονα y είναι επιτρεπτή καθώς δεν υπάρχει κάποια τομή μεταξύ της αναπαράστασης της αρπάγης και του ελλειπούς σχήματος. Οπότε δεν υφίστανται γεωμετρικοί περιορισμοί που να εμποδίζουν την λαβή από το να πραγματοποιηθεί.	127
Σχήμα 7.12: Λογικό διάγραμμα αλγορίθμου εύρεσης επιτρεπτών λαβών	127
Σχήμα 7.13: Επιτρεπτές λαβές για πρωταρχικά αντικείμενα. Ο κύβος έχει δυο, το τραπέζιο μια και ο κύλινδρος τέσσερις.....	128
Σχήμα 7.14: Επιτρεπτές λαβές για πολύπλοκα αντικείμενα.....	130
Σχήμα 8.1: Διάκριση μεταξύ αντικειμένων που μπορούν να κατατμηθούν και αντικειμένων τα οποία δεν μπορούν με την χρήση του κανόνα ελαχίστου.....	132

Κατάλογος Πινάκων

Πίνακας 2.1: Οι τιμές της πρώτης μεταβλητής H_u την Φ_1 για τους χαρακτήρες κάθε γραμμής. Η πρώτη γραμμή χρησιμοποιείται ως δεδομένα εκπαίδευσης ενώ στις υπόλοιπες παρουσιάζονται τα αποτελέσματα της αναγνώρισης (Flusser & Suk, 1994).	39
Πίνακας 2.2: Οι τιμές της πρώτης μεταβλητής AMI την I_1 για τους χαρακτήρες κάθε γραμμής. Η πρώτη γραμμή χρησιμοποιείται ως δεδομένα εκπαίδευσης ενώ στις υπόλοιπες παρουσιάζονται τα αποτελέσματα της αναγνώρισης (Flusser & Suk, 1994).	40
Πίνακας 6.1: Σύνολο εκπαίδευσης αποτελούμενο μόνο από τρία σχήματα.....	98
Πίνακας 6.2: Διανύσματα βαρών νευρώνων μονοδιάστατου αυτοοργανούμενου χάρτη που έχει εκπαιδευτεί μόνο με τρία σχήματα.....	99
Πίνακας 6.3: Αντιπροσωπευτικά δείγματα των πρωταρχικών σχημάτων και των γενικεύσεων αυτών.....	102
Πίνακας 7.1: Πίνακας παραμέτρων Denavit-Hartenberg για ορθή κινηματική ανάλυση του ρομποτικού βραχίονα Katana 6M180.	114

ΚΕΦΑΛΑΙΟ 1

1 Εισαγωγή

1.1 Ρομποτική

Ρομποτική είναι η επιστήμη που ασχολείται με εκείνες τις μηχανές που μπορούν να αντικαταστήσουν τον άνθρωπο στην εκτέλεση κάποιας εργασίας, η οποία περιλαμβάνει τόσο την λήψη κάποιας απόφασης όσο και την αλληλεπίδραση του ρομποτικού χειριστή με το περιβάλλον στο οποίο βρίσκεται.

Ο όρος *ρομποτική* εισήχθη πρώτη φορά από τον Ρώσο Isaac Asimov, έναν διάσημο συγγραφέα έργων επιστημονικής φαντασίας. Ο Asimov όρισε ως ρομποτική τον τομέα της επιστήμης που έχει αφιερωθεί στην μελέτη των ρομπότ, η λειτουργία των οποίων πρέπει να διέπεται από τους εξής τρεις κανόνες.

- Ένα ρομπότ πρέπει να μην βλάπτει οποιονδήποτε άνθρωπο ή μέσω της απραξίας του να επιτρέπει τον τραυματισμό κάποιου ανθρώπου.
- Το ρομπότ πρέπει να υπακούει στις εντολές που του δίνονται από τους ανθρώπους, εκτός κι αν αυτές οι εντολές παραβαίνουν τον πρώτο κανόνα.
- Το ρομπότ οφείλει να προστατεύει την ύπαρξη του, εφόσον όμως οι ενέργειες που απαιτούνται δεν παραβιάζουν τους παραπάνω δυο κανόνες

Η επιθυμία του ανθρώπου να δημιουργήσει μηχανές στις οποίες θα μπορεί να αναθέτει εργασίες είναι τόσο παλιά που εμφανίζεται ήδη στην αρχαία Ελληνική μυθολογία, σύμφωνα με την οποία ο Ήφαιστος σφυρηλάτησε το γιγάντιο ρομπότ

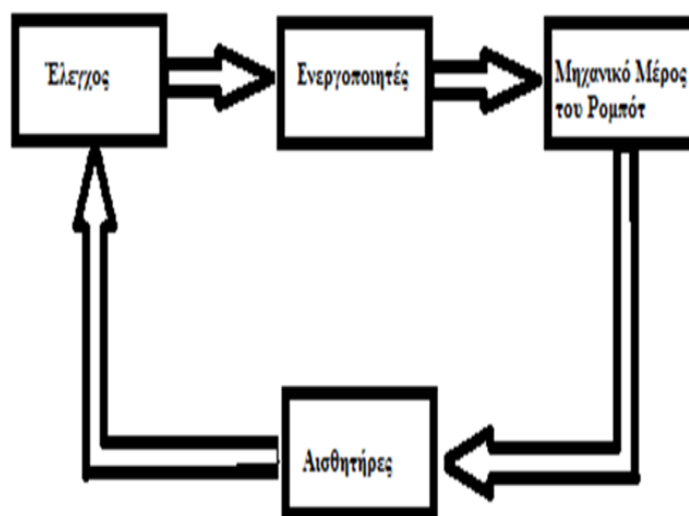
Τάλως και του ανέθεσε την αποστολή να προστατεύει το νησί της Κρήτης από τους εισβολείς.

Στην σύγχρονη εποχή, τη θέση της μυθολογίας έχει πάρει η επιστημονική φαντασία όπου οι άνθρωποι συνεχίζουν να φαντάζονται τα ρομπότ ως ανθρωπόμορφα όντα, που διαθέτουν στοιχειώδη νόηση και έτσι μπορούν να αλληλεπιδρούν με το περιβάλλον τους. Τέτοιου είδους έργα επιστημονικής φαντασίας αποτελούν το Star Wars, Avatar κ.α.

Σύμφωνα λοιπόν με την επιστημονική θεώρηση των παραπάνω σεναρίων επιστημονικής φαντασίας, το ρομπότ αποτελεί μια μηχανή η οποία είναι ικανή να τροποποιεί το περιβάλλον στο οποίο λειτουργεί. Αυτό το επιτυγχάνει με ενέργειες οι οποίες ακολουθούν κάποιους κανόνες, όπως αυτούς που περιέγραψε ο Asimov, ενώ ταυτόχρονα οι ενέργειες του ρομπότ επηρεάζονται άμεσα από τα δεδομένα που λαμβάνει από το περιβάλλον του.

1.2 Ρομποτικό σύστημα

Επομένως εύκολα συμπεραίνεται από τα προαναφερθέντα ότι το ρομπότ στην πραγματικότητα είναι ένα περίπλοκο σύστημα που αποτελείται από πολλά υποσυστήματα, όπως φαίνεται στο μπλοκ διάγραμμα του σχήματος 1.1

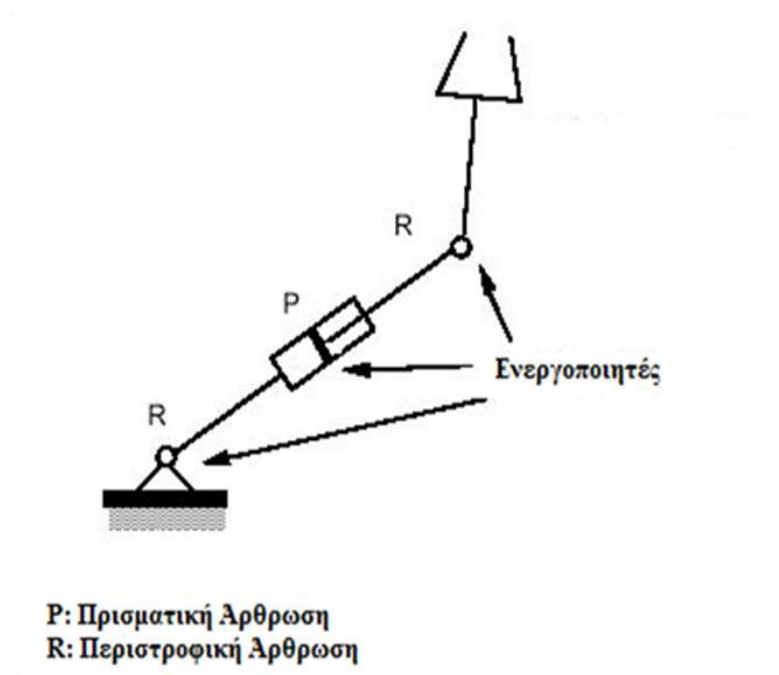


Σχήμα 1.1: Μπλοκ διάγραμμα των συστατικών μερών ενός ρομποτικού συστήματος

Συγκεκριμένα, το μηχανικό μέρος του ρομπότ μπορεί να αποτελείται από βραχίονες, τροχούς, μηχανικά πόδια κ.α. Η κινητικότητα του ρομπότ εξασφαλίζεται μέσω των αρθρώσεων με τις οποίες ενώνονται τα επιμέρους μηχανικά μέρη του ρομπότ που ονομάζονται σύνδεσμοι.

Ο τρόπος διασύνδεσης των αρθρώσεων και των συνδέσμων μας δίνει την κινηματική αλυσίδα του ρομποτικού χειριστή. Μια κινηματική αλυσίδα ονομάζεται ανοικτή όταν υπάρχει μια μοναδική αλληλουχία συνδέσμων που ενώνει τον αρχικό και τον τελικό σύνδεσμο μεταξύ τους. Σε αντίθετη περίπτωση η κινηματική αλυσίδα ονομάζεται κλειστή.

Υπάρχουν δύο είδη αρθρώσεων οι πρισματικές και οι περιστροφικές, όπως φαίνεται και στο σχήμα 1.2. Στην ανοικτή κινηματική αλυσίδα κάθε άρθρωση, είτε πρισματική είτε περιστροφική, δίνει και έναν επιπλέον βαθμό ελευθερίας στη δομή του ρομποτικού χειριστή.



Σχήμα 1.2: Ανοικτή κινηματική αλυσίδα που αποτελείται από δυο περιστροφικές και μια πρισματική άρθρωση.

Την δυνατότητα στο ρομπότ να διεξάγει μια ενέργεια του την παρέχουν οι ενεργοποιητές, οι οποίοι κινούν τα μηχανικά του μέρη. Οι ενεργοποιητές είναι συνήθως σερβοκινητήρες ή άλλες διατάξεις, των οποίων η κίνηση είναι πλήρως ελέγξιμη ανάλογα με τα σήματα εισόδου που τους παρέχονται.

Μέσω των αισθητήρων καθίσταται δυνατή η αλληλεπίδραση του ρομπότ με το περιβάλλον του. Υπάρχουν διάφορες κατηγορίες αισθητήρων όπως αισθητήρες όρασης, κίνησης, δύναμης. Συνήθως, σ' ένα ρομποτικό χειριστή υπάρχουν πολλές αισθητήριες μονάδες κάποιες από αυτές λαμβάνουν δεδομένα για την κατάσταση του ρομπότ, όπως την γωνία περιστροφής κάθε σερβοκινητήρα ενώ κάποιες άλλες λαμβάνουν δεδομένα για την κατάσταση του περιβάλλοντος, όπως οι κάμερες.

Τέλος, μέσω του συστήματος ελέγχου δίνονται οι απαραίτητες εντολές στο ρομποτικό σύστημα ώστε να εκτελέσει την ζητούμενη εργασία. Πιο συγκεκριμένα, το σύστημα ελέγχου, λαμβάνει μέσω των αισθητήρων την κατάσταση στην οποία

βρίσκεται την συγκεκριμένη στιγμή το περιβάλλον καθώς και το ρομπότ και παρέχει εκείνα τα σήματα ελέγχου που οδηγούν το ρομπότ στην εκπλήρωση της εργασίας που του έχει ανατεθεί.

Σύμφωνα λοιπόν με την παραπάνω ανάλυση, το πεδίο της ρομποτικής επιστήμης είναι μια μίξη διάφορων επιστημονικών κλάδων όπως αυτός της μηχανικής, του ελεγχου, των υπολογιστών και της ηλεκτρονικής.

1.3 Αλληλεπίδραση Ανθρώπου – Ρομπότ

Όπως προαναφέρθηκε, σκοπός ενός ρομπότ δεν είναι απλώς να εκτελέσει μια προκαθορισμένη εργασία υπό ελεγχόμενες συνθήκες. Στόχος είναι το ρομπότ να λειτουργεί σε ένα συνεχώς εξελισσόμενο περιβάλλον, στο οποίο θα προσαρμόζεται έτσι ώστε να καταφέρει να υλοποιήσει την εργασία που του έχει ανατεθεί. Παραδείγματος χάρη, εάν σε ένα αυτόνομο κινούμενο ρομπότ ζητηθεί να κινηθεί μεταξύ δύο θέσεων τότε θα πρέπει αυτό να είναι ικανό να αποφεύγει στατικά ή κινούμενα εμπόδια που πιθανώς να υπάρχουν στο ενδιάμεσο των δυο θέσεων.

Οπότε εφόσον στόχος του ρομπότ είναι να λειτουργεί σε ένα συνεχώς μεταβαλλόμενο περιβάλλον οφείλει να αλληλεπιδρά με αυτό. Ιδιαίτερη σημασία έχει η αλληλεπίδραση του ρομπότ με τον άνθρωπο. Ήδη έχουν αναπτυχθεί ρομπότ που μπορούν να αναγνωρίζουν τον χρήστη τους και να παρακολουθούν τις κινήσεις του στο χώρο, να απαντούν σε ερωτήσεις και να κινούνται αυτόνομα εκτελώντας εργασίες που τους έχουν ζητηθεί. Αυτά τα ρομπότ σύντομα θα μπορούν να εκτελούν μια πληθώρα εργασιών που ίσως είναι αδύνατον να κάνει ο άνθρωπος, όπως το να διεισδύουν σε συντρίμια και να εντοπίζουν επιζώντες.

1.4 Εκπαίδευση ρομπότ από τον άνθρωπο

Στα πλαίσια λοιπόν της αλληλεπίδρασης του ανθρώπου με το ρομπότ, περιλαμβάνεται και ο τομέας της εκπαίδευσης. Συγκεκριμένα, το ρομπότ θα πρέπει να είναι ικανό να μαθαίνει μια εργασία που θα του επιδεικνύει ο άνθρωπος. Η επίδειξη της εργασίας αυτής μπορεί να γίνεται είτε σε πραγματικό είτε σε εικονικό περιβάλλον.

Στον πραγματικό κόσμο, ο άνθρωπος έχει την δυνατότητα να επιδεικνύει μια κίνηση σε ένα ρομπότ με δύο τρόπους. Ο πρώτος είναι να το καθοδηγήσει είτε από κοντά αρπάζοντάς το και καθοδηγώντας το ώστε να εκτελέσει την επιθυμητή κίνηση, είτε από μακριά μέσω τηλεχειρισμού. Ο δεύτερος τρόπος είναι να εκτελεί ο άνθρωπος την κίνηση και το ρομπότ να την καταγράφει με δικούς του αισθητήρες όπως μια κάμερα. Στην πρώτη περίπτωση, η αποθήκευση των δεδομένων της κίνησης είναι άμεση καθώς αυτή πραγματοποιείται στον χώρο κίνησης του ρομπότ, ενώ στην δεύτερη περίπτωση τα δεδομένα είναι έμμεσα, καθώς η καταγραφή της γίνεται στο χώρο κίνησης του ανθρώπου και πρέπει να χρησιμοποιηθούν κατάλληλοι αλγόριθμοι

επεξεργασίας εικόνας και αντιστοίχισης με τον χώρο κίνησης του ρομπότ ώστε να εξαχθούν τα δεδομένα εκπαίδευσης.

Ωστόσο, οι δυσκολίες που προκύπτουν στον πραγματικό κόσμο - όπως είναι το κόστος της προμήθειας διαφόρων τύπων ρομποτικών χειριστών, τα λάθη στον χειρισμό που μπορεί ακόμα και να καταστρέψουν τον ρομποτικό χειριστή - έχουν οδηγήσει στην ανάπτυξη λογισμικού εικονικής προσομοίωσης. Με το λογισμικό αυτό οι άνθρωποι έχουν τη δυνατότητα να χειρίζονται απευθείας το ρομπότ και να του επιδεικνύουν κινήσεις καθώς και να πειραματίζονται πολύ γρήγορα με νέες μεθόδους.

Τα δεδομένα εκπαίδευσης λοιπόν προκύπτουν ανάλογα με τον τρόπο με τον οποίο επιδεικνύεται η κίνηση στο ρομπότ. Μπορεί να χρησιμοποιηθεί είτε ολόκληρη η τροχιά είτε ένα μέρος αυτής όπως η τελική θέση. Επειδή μπορεί να μην είναι επιθυμητή η χρήση μεγάλου όγκου δεδομένων, μπορεί να χρησιμοποιηθεί ως δεδομένο εκπαίδευσης μόνο μια γωνία στρέψης κάποιας άρθρωσης. Παραδείγματος χάρη στην παρούσα διπλωματική εργασία ως δεδομένο εκπαίδευσης που παρέχεται από τον χρήστη είναι η γωνία στρέψης της άρθρωσης που χειρίζεται την αρπάγη του ρομποτικού βραχίονα. Τέλος, τα δεδομένα εκπαίδευσης εμπλουτίζονται με επιπλέον στοιχεία που παρέχονται από τις αισθητήριες μονάδες του ρομπότ, όπως η κάμερα και έτσι προκύπτει το τελικό σύνολο δεδομένων εκπαίδευσης.

Είναι αναγκαίο λοιπόν το ρομπότ να ανταποκρίνεται ικανοποιητικά όχι μόνο στις κινήσεις που του έχουν επιδειχθεί αλλά και σε πληθώρα άλλων κινήσεων σε οποιαδήποτε κατάσταση ή περιβάλλον βρεθεί. Αυτό βέβαια, θα πρέπει να γίνει χωρίς να χρειάζεται να δειχθούν πολλές κινήσεις στο ρομπότ καθώς ο χρόνος που απαιτείται από τον άνθρωπο για να εκπαιδεύσει το ρομπότ πρέπει να είναι περιορισμένος.

Κρίνεται λοιπόν απαραίτητο να βρεθεί ένα είδος συνάρτησης ή διαδικασίας αντιστοίχισης (αναλυτική ή μη), στην οποία θα δίνονται ως είσοδος τα δεδομένα εκπαίδευσης και ως έξοδος θα παράγεται μια επιθυμητή συμπεριφορά για το ρομπότ. Το ποιά θα είναι η μέθοδος προσδιορισμού της συνάρτησης αυτής προκύπτει από διάφορους παράγοντες. Εξαρτάται πιο συγκεκριμένα από το αν η είσοδος και η έξοδος είναι συνεχής ή διακριτές τιμές, αν η μέθοδος χρησιμοποιεί δεδομένα για την προσέγγιση της συνάρτησης πριν ή κατά την διάρκεια της κίνησης του ρομπότ, από το αν είναι αδιάφορο ή θεμιτό να κρατάμε τα δεδομένα της επίδειξης καθ' όλη την διάρκεια της εκπαίδευσης ή αν συνεχώς τα δεδομένα αυτά ανανεώνονται. Όπως επισημαίνεται (Hastie et al. 2009), οι στατιστικές μέθοδοι προσέγγισης μιας συνάρτησης εμπίπτουν σε δυο μεγάλες κατηγορίες.

Στην πρώτη κατηγορία εντάσσονται οι *μέθοδοι ταξινόμησης*, οι οποίες κατηγοριοποιούν τα δεδομένα εκπαίδευσης σε διακριτές κλάσεις. Αρχικά, το μεγαλύτερο μέρος των δεδομένων εκπαίδευσης χρησιμοποιείται για τον προσδιορισμό των κλάσεων, ενώ ένα μικρότερο για επαλήθευση. Αφού προσδιοριστούν οι κλάσεις, κάθε μια από τις οποίες αντιστοιχεί σε κάποια ενέργεια του ρομπότ, η αναγνώριση έγκειται απλά στην εύρεση της κλάσης που ανήκει το πρότυπο εισόδου. Από τις πιο γνωστές μεθόδους ταξινόμησης είναι τα Κρυφά

Μαρκοβιανά Μοντέλα (HMM), ο K-means, GMMs (Gaussian Mixture Models), EM (Expectation Maximization) και τα SVM (Support Vector Machines).

Στην δεύτερη κατηγορία εντάσσονται οι μέθοδοι παλινδρόμησης, οι οποίες μέσω επαναληπτικών διαδικασιών προσεγγίζουν τα δεδομένα στο συνεχή χώρο χωρίς να επιχειρούν την ομαδοποίηση τους. Η παλινδρόμηση, επειδή οι συνεχείς τιμές της εξόδου είναι συχνά αποτέλεσμα συνδυασμού διαφόρων δεδομένων εκπαίδευσης, ουσιαστικά εφαρμόζεται σε χαμηλού επιπέδου κινήσεις και όχι σε υψηλού επιπέδου συμπεριφορές. Μια σημαντική διαφορά μεταξύ των μεθόδων παλινδρόμησης είναι αν η προσέγγιση της συνάρτησης αντιστοίχισης πραγματοποιείται κατά την εκτέλεση της κίνησης από το ρομπότ ή πριν από αυτήν. Από τις πιο γνωστές μεθόδους παλινδρόμησης είναι η Οκνηρή Μάθηση (kNN), η Τοπικά Σταθμισμένη Παλινδρόμηση (LWR) και τα Νευρωνικά Δίκτυα (NN).

Στην Οκνηρή Μάθηση, η προσέγγιση της συνάρτησης αντιστοίχισης δεν πραγματοποιείται μέχρις ότου να είναι αναγκαία η αντιστοίχιση κάποιου τωρινού σημείου παρατήρησης. Το μειονέκτημα βέβαια αυτής της μεθόδου είναι ότι πρέπει να κρατούνται συνεχώς στην μνήμη όλα τα δεδομένα εκπαίδευσης.

Στον αντίποδα, βρίσκονται τα Νευρωνικά Δίκτυα όπου η προσέγγιση της συνάρτησης αντιστοίχισης γίνεται πριν από οποιαδήποτε κίνηση του ρομπότ. Πλεονέκτημα της μεθόδου αυτής αποτελεί ότι δεν κρατάει στην μνήμη τα δεδομένα εκπαίδευσης που απαιτούν αρκετό αποθηκευτικό χώρο. Παράλληλα δεν χρειάζεται να υπολογιστούν δεδομένα εκπαίδευσης κατά την διάρκεια εκτέλεσης της κίνησης, ωστόσο η μέθοδος παρουσιάζει μεγάλο κόστος υπολογισμού πριν την εκτέλεση όπου γίνεται και η εκπαίδευση του Νευρωνικού Δικτύου. Μειονέκτημα της χρήσης Νευρωνικών Δικτύων μπορεί να θεωρηθεί το ότι εφόσον η εκπαίδευση δεν είναι επαρκής μπορεί αυτά να «ξεχάσουν» ακόμα και τα δεδομένα με τα οποία έχουν εκπαιδευτεί. Οπότε, αν και θεωρείται μια από τις καλύτερες μεθόδους καθώς είναι ταχύτερη κατά την φάση της εκτέλεσης, απαιτείται προσοχή κατά την φάση της εκπαίδευσης.

1.5 Σύνοψη παρούσας διπλωματικής εργασίας

Προκειμένου το ρομπότ να είναι σε θέση να αλληλεπιδρά με το συνεχώς μεταβαλλόμενο περιβάλλον γύρω του, ώστε να μπορεί να διευκολύνει τους ανθρώπους σε διάφορες καθημερινές εργασίες, πρέπει αυτό να μπορεί να χειρίζεται μια πληθώρα αντικειμένων. Επομένως, ο χειρισμός είναι προαπαιτούμενος για τις περισσότερες ρομποτικές εφαρμογές.

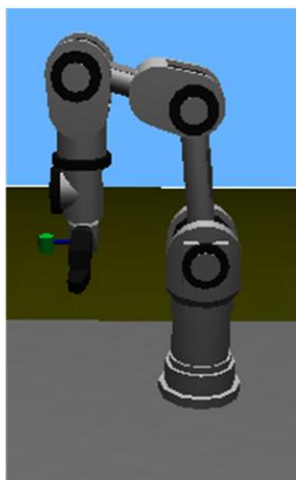
Η μέθοδος με την οποία ένα ρομπότ μπορεί να πιάσει άγνωστα αντικείμενα παραμένει ακόμα ένα άλυτο πρόβλημα. Παρότι οι άνθρωποι αποκτούν αυτή την ικανότητα σχετικά εύκολα από νηπιακή ακόμα ηλικία, ο τρόπος με τον οποίο ο εγκέφαλος εκτελεί την συγκεκριμένη εργασία παραμένει άγνωστος, δυσκολεύοντας έτσι την ανάπτυξη ρομποτικών συστημάτων που μπορούν να μιμηθούν την συγκεκριμένη συμπεριφορά. Συνεπώς, περιοριζόμαστε στην λύση του προβλήματος χρησιμοποιώντας ελεγχόμενες συνθήκες εργασίας.

Η επιλογή μιας κατάλληλης λαβής για ένα δοσμένο αντικείμενο απαιτεί την γνώση της γεωμετρίας τόσο του αντικειμένου όσο και της αρπάγης του ρομποτικού βραχίονα με τον οποίο θα επιχειρηθεί η λαβή. Επιπλέον, απαραίτητη είναι και η γνώση της κινηματικής του βραχίονα ώστε να είναι δυνατός ο έλεγχος του κατά πόσο μια λαβή είναι και επιτεύξιμη. Όλοι αυτοί οι περιορισμοί συνυπάρχουν και πρέπει να συνυπολογίζονται για την ορθή λύση του προβλήματος.

Η γνώση της γεωμετρίας οποιουδήποτε αντικειμένου είναι αδύνατη. Οπότε, επικεντρωνόμαστε στην προσπάθεια απλοποίησης του προβλήματος μας εκμεταλλευόμενοι την ιδέα ότι τα περισσότερα αντικείμενα μπορούν να αποδομηθούν σε απλούστερα για τα οποία μπορεί να υπάρξει γνώση της γεωμετρίας τους (Goldfeder et al., 2007). Αυτά τα απλούστερα (πρωταρχικά) αντικείμενα όταν ενωθούν δίνουν το αρχικό σύνθετο αντικείμενο. Προσδιορίζοντας λοιπόν τις επιτρεπτές λαβές για το σύνολο των πρωταρχικών αντικειμένων και επιλέγοντας αυτές για τις οποίες δεν υφίστανται γεωμετρικοί περιορισμοί, αποκτούμε το σύνολο των επιτρεπτών λαβών του σύνθετου αντικειμένου.

Στην παρούσα διπλωματική εργασία, θα ασχοληθούμε με το πρόβλημα της εύρεσης των επιτρεπτών λαβών τυχαίων αντικειμένων που βρίσκονται τοποθετημένα πάνω σε οριζόντιο τραπέζι. Χρησιμοποιούμε τον ρομποτικό βραχίονα Katana 6M180, στο οποίο έχουμε προσαρμόσει μια κάμερα για να λαμβάνουμε οπτικά δεδομένα.

Καθότι ο ρομποτικός βραχίονας διαθέτει μόνο πέντε βαθμούς ελευθερίας με αποτέλεσμα να μην είναι δυνατή η επίτευξη οποιασδήποτε θέσης και προσανατολισμού στο χώρο, περιορίζουμε τη λύση του προβλήματός μας στην εύρεση εκείνων των επιτρεπτών λαβών του αντικειμένου των οποίων ο άξονας της αρπάγης του ρομποτικού βραχίονα είναι κάθετος στο επίπεδο του πάγκου εργασίας (βλέπε σχήμα 1.3)



Σχήμα 1.3: Προσανατολισμός της αρπάγης κάθετος προς την επιφάνεια του πάγκου εργασίας.

Αφού επομένως μας ενδιαφέρει ο προσδιορισμός των λαβών ως προς μια μόνο όψη του αντικειμένου, χρησιμοποιούμε δισδιάστατα δεδομένα για την υλοποίηση του πειράματος στο εικονικό περιβάλλον Webots. Ωστόσο, κάνουμε μια εξαίρεση στον υπολογισμό της απόστασης που βρίσκεται το αντικείμενο από την κάμερα λήψης όπου αναγκάζομαστε να προβούμε σε τρισδιάστατη ανακατασκευή ενός μόνο σημείου της επιφάνειας του αντικειμένου.

Η παρούσα διπλωματική εργασία, απαρτίζεται από επτά επιπλέον κεφάλαια και η δομή της είναι η παρακάτω

- *Κεφάλαιο 1:* Εισαγωγικές έννοιες γύρω από την επιστήμη της Ρομποτικής καθώς και επιγραμματική αναφορά στα επιμέρους υποσυστήματα από τα οποία απαρτίζεται ένα ρομποτικό σύστημα. Δίνεται έμφαση στην αλληλεπίδραση του ανθρώπου με το ρομπότ και στο πώς αυτό μπορεί να εκπαιδευτεί ώστε να εκτελεί εργασίες που θα διευκολύνουν την ζωή του ανθρώπου.
- *Κεφάλαιο 2:* Παρουσιάζεται αναλυτικά η θεωρία των ροπών εικόνας και η χρησιμότητα που έχουν ως περιγραφείς επίπεδων σχημάτων. Ιδιαίτερη έμφαση δίνεται στις Αναλλοίωτες Μεταβλητές Ροπών (Affine Moment Invariants - AMIs) οι οποίες χρησιμοποιούνται τόσο για την εκπαίδευση του συστήματός μας όσο και για την εύρεση των επιτρεπών λαβών των αντικειμένων. Τέλος παρουσιάζονται πειραματικά δεδομένα για την ευρωστία που παρουσιάζουν τα AMIs σε σχέση με τις ροπές του Hu.
- *Κεφάλαιο 3:* Γίνεται αναφορά στα πλεονεκτήματα χρήσης νευρωνικών δικτύων για την εκπαίδευση του ρομποτικού συστήματος. Παρουσιάζεται ακόμα η θεωρία που διέπει τους αυτοοργανούμενους χάρτες (SOM) αλλά και παραδείγματα εκπαίδευσης χρησιμοποιώντας αυτοοργανούμενους χάρτες
- *Κεφάλαιο 4:* Ορίζεται η έννοια της καμπύλης και αναλύεται η μεθοδολογία που χρησιμοποιείται για τον προσδιορισμό της συνάρτησης καμπυλότητας ενός σχήματος, εξάγοντας την τροχιά του συνόρου του. Επίσης γίνεται διεξοδική αναφορά στην θεωρία που διέπει την κατάτμηση σχήματος και στους τρόπους με τους οποίους μπορεί να αποδομηθεί ένα σύνθετο σχήμα σε απλούστερα.
- *Κεφάλαιο 5:* Αναλύεται ο τρόπος με τον οποίο μπορούμε να εντοπίσουμε τις συντεταγμένες ενός σημείου στο χώρο, χρησιμοποιώντας έννοιες από την στερεοπτική όραση και την επιπολική γεωμετρία. Επιπρόσθετα αναφέρουμε την τεχνική βαθμονόμησης κάμερας που εισήχθη από τον Zhang (2000).

- *Κεφάλαιο 6:* Παρουσιάζεται αναλυτικά η διαδικασία που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου. Περιγράφεται ο τρόπος με τον οποίο παράγονται τα δεδομένα εκπαίδευσης και επαλήθευσης ενώ παράλληλα δίνονται διαγράμματα με τα ποσοστά επιτυχίας του αυτοοργανούμενου χάρτη στα διάφορα σύνολα εκπαίδευσης που επιλέγονται.
- *Κεφάλαιο 7:* Αναλύεται διεξοδικά όλη η διαδικασία που ακολουθήθηκε προκειμένου να εντοπίσουμε τις επιτρεπτές λαβές των αντικειμένων. Αρχικά παρουσιάζεται το εικονικό περιβάλλον προσομοίωσης Webots καθώς και τα τεχνικά χαρακτηριστικά του ρομποτικού βραχίονα Katana 6M180 που χρησιμοποιήθηκε. Τέλος, παρατίθενται στιγμιότυπα της εκτέλεσης του πειράματος στο εικονικό περιβάλλον.
- *Κεφάλαιο 8:* Γίνεται αναφορά στα πλεονεκτήματα και στα μειονεκτήματα της μεθόδου προσδιορισμού επιτρεπτών λαβών ενώ ταυτόχρονα προτείνονται βελτιώσεις και προσθήκες που θα μπορούσαν να γίνουν σε αυτή. Προτείνονται ακόμα ορισμένες κατευθύνσεις έρευνας τις οποίες θα μπορούσε να ακολουθήσει μια μελλοντική εργασία η οποία θα έχει ως βάση την παρούσα διπλωματική εργασία.

Στο τέλος της εργασίας παρατίθεται ολόκληρη η βιβλιογραφία που χρησιμοποιήθηκε για την σύνταξη της παρούσας διπλωματικής εργασίας.

ΚΕΦΑΛΑΙΟ 2

2 Ροπές Εικόνας και Αναλλοίωτες Μεταβλητές Ροπών

2.1 Εισαγωγή

Η αναγνώριση αντικειμένων ή σχημάτων, τα οποία έχουν παραμορφωθεί με διάφορους τρόπους υπήρξε ο στόχος πολλών σύγχρονων ερευνών. Υπάρχουν στην πραγματικότητα τρεις βασικές προσεγγίσεις του προβλήματος, (i) η εξονυχιστική αναζήτηση, (ii) η κανονικοποίηση εικόνας και (iii) η χρήση αναλλοίωτων χαρακτηριστικών.

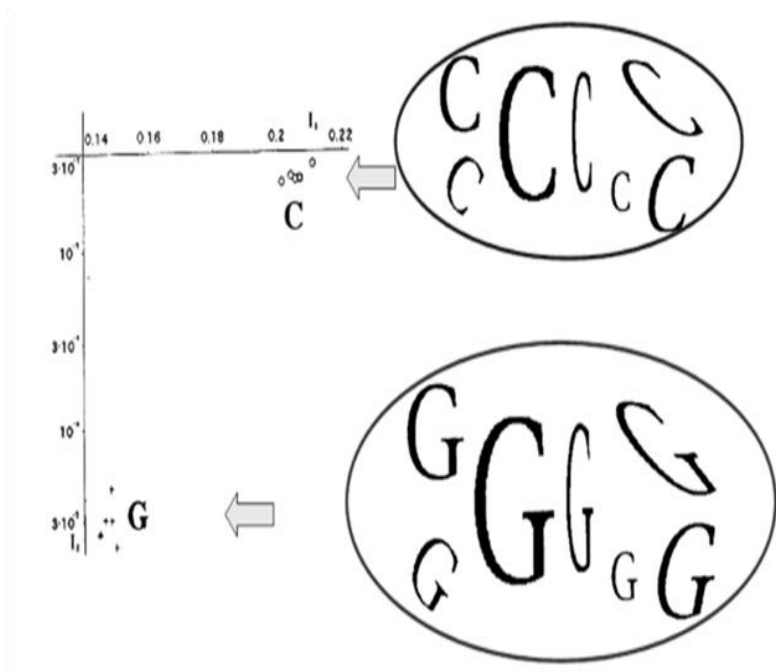
Στην εξονυχιστική αναζήτηση, ψάχνουμε τον παραμετρικό χώρο όλων των πιθανών παραμορφώσεων της εικόνας. Αυτό, σημαίνει ότι, το σύνολο εκπαίδευσης κάθε τάξης δεν αρκεί να περιέχει μόνο ένα αντιπροσωπευτικό δείγμα της τάξης, αλλά και όλες τις παραμορφωμένες εκδοχές του (περιστροφή, μετατόπιση, κλιμάκωση, θόλωμα). Φυσικά μια τέτοια υλοποίηση θα οδηγούσε σε τεράστια χρονική πολυπλοκότητα και είναι πρακτικά ανεφάρμοστη.

Στην μέθοδο της «κανονικοποίησης εικόνας», τα αντικείμενα μετασχηματίζονται ώστε να έχουν ένα συγκεκριμένο προσανατολισμό πριν μπουν στον ταξινομητή. Ωστόσο, αυτή η μέθοδος αν και φαίνεται αποτελεσματική, απαιτεί την λύση δύσκολων ανάστροφων προβλημάτων, τα οποία τις περισσότερες φορές μπορεί να είναι αδύνατον να λυθούν.

Η προσέγγιση που χρησιμοποιεί «αναλλοίωτες μεταβλητές» (Invariants) φαίνεται να είναι η καλύτερη και γι αυτό τον λόγο έχει χρησιμοποιηθεί εκτενώς. Η βασική της ιδέα είναι να περιγράψει το αντικείμενο με ένα σύνολο από βαθμωτές ποσότητες που

ονομάζονται *αναλλοίωτες μεταβλητές (invariants)*, οι οποίες δεν παρουσιάζουν ευαισθησία στις παραμορφώσεις των αντικειμένων ενώ ταυτόχρονα παρέχουν την δυνατότητα να διακρίνουμε αντικείμενα που ανήκουν σε διαφορετικές τάξεις.

Συνήθως, μια μόνο αναλλοίωτη μεταβλητή δεν αρκεί για να διαχωρίσουμε ικανοποιητικά τα σχήματα μας και γι αυτό τον λόγο χρησιμοποιούμε ένα διάνυσμα μεταβλητών μεγέθους n . Έτσι κάθε σχήμα αντιπροσωπεύεται από ένα σημείο στον n -διάστατο χώρο χαρακτηριστικών. (βλέπε σχήμα 2.1).



Σχήμα 2.1: Δισδιάστατος χώρος με δυο διακριτές τάξεις. Οι αναλλοίωτες μεταβλητές διατηρούν κοντά τις τιμές τους ανεξάρτητα με τις παραμορφώσεις που γίνονται στα σχήματα (Flusser & Suk, 2009).

2.2 Κατηγορίες αναλλοίωτων μεταβλητών

Οι υπάρχουσες αναλλοίωτες μεταβλητές, οι οποίες χρησιμοποιούνται προκειμένου να περιγράψουμε επίπεδα σχήματα, μπορούν να κατηγοριοποιηθούν με έναν από τους παρακάτω τρόπους.

- Ανάλογα με τον τύπος της *αμεταβλητότητας* τους. Δηλαδή, αν η τιμή τους παραμένει αναλλοίωτη ως προς την μετατόπιση, περιστροφή ή κλιμάκωση.
- Ανάλογα με τα μαθηματικά εργαλεία που χρησιμοποιούν

1. *Απλοί περιγραφείς σχήματος* - συνεκτικότητα, κυρτότητα κ.α.

2. *Συντελεστές μετασχηματισμού* – υπολογίζονται από τον εκάστοτε μετασχηματισμό της εικόνας – Περιγραφείς Fourier, Περιγραφείς Hadamard.
 3. *Αναλλοίωτες Μεταβλητές Ροπής* – ειδικές συναρτήσεις των ροπών της εικόνας.
 4. *Διαφορίσιμες αναλλοίωτες μεταβλητές* – χρησιμοποιούν τις παραγώγους του συνόρου του αντικειμένου.
- Ανάλογα με το μέρος του αντικειμένου που χρησιμοποιείται προκειμένου να υπολογιστούν οι τιμές των αναλλοίωτων μεταβλητών
 1. *Καθολικά αναλλοίωτες μεταβλητές* – υπολογίζονται στο σύνολο της εικόνας (συμπεριλαμβανομένου και του φόντου εφ' όσον δεν έχει γίνει κατάτμηση). Σε σύγκριση με τις τοπικές αναλλοίωτες μεταβλητές, οι καθολικές είναι πιο ανθεκτικές στον θόρυβο, στη μη ακριβή εξαγωγή του συνόρου του αντικειμένου και άλλους παρόμοιους παράγοντες. Ωστόσο, το κυριότερο μειονέκτημα τους είναι ότι λαμβάνουν υπόψη τους όλα τα συστατικά της εικόνας, οπότε αν σε μια εικόνα υπάρχουν δυο σχήματα και όχι ένα, ακόμα και αν αυτά τα σχήματα είναι όμοια, θα μας δώσουν διαφορετικό αποτέλεσμα αν είχαμε μια εικόνα που περιελάμβανε μόνο ένα σχήμα.
 2. *Τοπικές αναλλοίωτες μεταβλητές* – υπολογίζονται σε μια συγκεκριμένη περιοχή κυρίαρχων σημείων. Οι «διαφορίσιμες αναλλοίωτες μεταβλητές» είναι τυπικοί αντιπρόσωποι αυτής της κατηγορίας. Αφού εξάγουμε το σύνορο του αντικειμένου υπολογίζουμε σε κάθε σημείο του, τις αναλλοίωτες μεταβλητές, ως συναρτήσεις των παραγώγων του συνόρου. Συνεπώς, οι αναλλοίωτες μεταβλητές σε οποιοδήποτε σημείο εξαρτώνται μόνο από μια γειτονιά της καμπύλης του συνόρου, του αντικειμένου. Ωστόσο, παρουσιάζουν ευαισθησία στον θόρυβο και σε μη ακριβή εντοπισμό του συνόρου.
 3. *Ημι-τοπικές αναλλοίωτες μεταβλητές* - προσπαθούν να συνδυάσουν τα πλεονεκτήματα και των δυο παραπάνω κατηγοριών. Αρχικά κατατέμνουν την εικόνα σε βασικά σχήματα και έπειτα χρησιμοποιούν τις καθολικά αναλλοίωτες μεταβλητές προκειμένου να περιγράψουν αυτά τα σχήματα. Αποτελούν τις πιο συχνά χρησιμοποιούμενες αναλλοίωτες μεταβλητές καθώς δεν επηρεάζονται σε μεγάλο βαθμό ούτε από θόρυβο, ούτε από μη επιτυχή εντοπισμό του συνόρου του αντικειμένου, ούτε από άλλα αντικείμενα που ίσως συνυπάρχουν στη εικόνα μας.

Στο παρόν κεφάλαιο ωστόσο επικεντρωνόμαστε στην περιγραφή των αντικειμένων μας χρησιμοποιώντας αναλλοίωτες μεταβλητές ροπών, οι οποίες πρωτοεμφανίστηκαν το 1962, από τον Hu (Hu, 1962). Από τότε, οι αναλλοίωτες μεταβλητές ροπών έχουν γίνει από τους πιο σπουδαίους και συχνά χρησιμοποιούμενους περιγραφείς σχημάτων.

2.3 Ροπές Εικόνας

Οι ροπές είναι μεγέθη που χρησιμοποιούνται προκειμένου να περιγράψουμε τις ιδιότητες και τα ιδιαίτερα χαρακτηριστικά μιας συνάρτησης. Χρησιμοποιούνται στην στατιστική για περιγραφή του σχήματος της συνάρτησης πυκνότητας πιθανότητας και στην κλασική μηχανική προκειμένου να μετρηθεί η κατανομή μάζας ενός στερεού αντικειμένου.

Όπως επισημαίνουν οι Flusser & Suk (2009), οι ροπές μπορούν να θεωρηθούν ως η προβολή μιας συνάρτησης σε μια πολυωνυμική βάση όπως ο μετασχηματισμός Fourier είναι η προβολή μιας συνάρτησης σε μια βάση από αρμονικές συναρτήσεις. Παραθέτουμε κάποιες βασικές έννοιες και συναρτήσεις που θα χρησιμοποιηθούν και περαιτέρω.

Ορισμός1: Με τον όρο συνάρτηση εικόνας (ή εικόνα) εννοούμε οποιαδήποτε τμηματικά συνεχή πραγματική συνάρτηση $f(x, y)$, δύο μεταβλητών, ορισμένη σε ένα συμπαγές σύνολο $D \subset \mathbb{R} \times \mathbb{R}$, με πεπερασμένο μη μηδενικό ολοκλήρωμα.

Ορισμός2: Ροπή M_{pq} μιας εικόνας $f(x, y): D \rightarrow \mathbb{R}$, όπου $D \subset \mathbb{R} \times \mathbb{R}$ και p, q μη αρνητικοί ακέραιοι αριθμοί ορίζεται η ποσότητα

$$M_{pq} = \iint_D p_{pq}(x, y) f(x, y) dx dy \quad (2.1)$$

όπου $p_{00}(x, y), p_{10}(x, y), \dots, p_{kj}(x, y), \dots$ είναι συναρτήσεις πολυωνυμικής βάσης και $r = p + q$ η τάξη της ροπή. Ανάλογα με την πολυωνυμική βάση που χρησιμοποιούμε έχουμε τα παρακάτω είδη ροπών.

2.3.1 Γεωμετρικές και Μιγαδικές Ροπές

Η πιο συνηθισμένη επιλογή βάσης είναι η «βάση δύναμης» $p_{kj}(x, y) = x^k y^j$ από την οποία προκύπτουν οι γεωμετρικές ροπές.

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

Οι γεωμετρικές ροπές χαμηλής τάξης έχουν διαισθητική σημασία - m_{00} είναι η «μάζα» του αντικειμένου (για δυαδικές εικόνες, η m_{00} μας δίνει το εμβαδόν της περιοχής που καταλαμβάνει το αντικείμενο στην εικόνα), m_{10}/m_{00} και m_{01}/m_{00} ορίζουν τις συντεταγμένες του κέντρου μάζας της εικόνας. Οι ροπές δευτέρας τάξης m_{20} και m_{02} περιγράφουν την «κατανομή μάζας» σε σχέση με τους βασικούς άξονες των συντεταγμένων. Στην μηχανική, ονομάζονται ροπές αδράνειας. Η ακτίνα περιστροφής ως προς έναν άξονα, μπορεί επίσης να εκφραστεί σε όρους που περιέχουν ροπές όπως $\sqrt{\frac{m_{20}}{m_{00}}}$ και $\sqrt{\frac{m_{02}}{m_{00}}}$.

Εάν η εικόνα θεωρηθεί ως συνάρτηση πυκνότητας πιθανότητας (pdf) (π.χ. οι τιμές της έχουν κανονικοποιηθεί έτσι ώστε $m_{00} = 1$), τότε m_{10} και m_{01} εκφράζουν τις μέσες τιμές. Σε περίπτωση που έχουμε μηδενικές μέσες τιμές, m_{20} και m_{02} είναι οι διακυμάνσεις της οριζόντιας και κάθετης προβολής, αντίστοιχα. Ενώ, η m_{11} είναι η συμμεταβλητότητα μεταξύ τους.

Η *λέπτυνση* και η *κύρτωση* επίσης μπορούν να περιγραφούν με όρους που χρησιμοποιούν αναλλοίωτες μεταβλητές. Η λέπτυνση της οριζόντιας προβολής ορίζεται ως $m_{30}/\sqrt{m_{20}^3}$ και της κάθετης προβολής ως $m_{03}/\sqrt{m_{20}^3}$. Η λέπτυνση, ως ποσότητα, μετράει την απόκλιση της αντίστοιχης προβολής από την συμμετρία. Εάν η προβολή είναι συμμετρική ως προς το μέσο (π.χ. την αρχή των αξόνων στην συγκεκριμένη περίπτωση), τότε η αντίστοιχη λέπτυνση ισούται με μηδέν. Η κύρτωση μετράει την ένταση των κορυφών στην συνάρτηση πυκνότητας πιθανότητας και ορίζεται ξεχωριστά για κάθε προβολή – η οριζόντια κύρτωση ορίζεται ως m_{40}/m_{20}^2 ενώ η κάθετη κύρτωση ως m_{04}/m_{20}^2 .

Ο χαρακτηρισμός της εικόνας χρησιμοποιώντας τις γεωμετρικές ροπές είναι πλήρης με την εξής έννοια. Οποιαδήποτε συνάρτηση εικόνας, της οποίας οι γεωμετρικές ροπές όλων των τάξεων υπάρχουν και είναι πεπερασμένες μπορεί να ανακατασκευαστεί πλήρως χρησιμοποιώντας μόνο το σύνολο με τις γεωμετρικές ροπές (αυτή η υπόθεση είναι γνωστή ως **θεώρημα μοναδικότητας**).

Αν για πολυωνυμική βάση επιλέξουμε την $p_{kj}(x, y) = (x + iy)^k (x - iy)^j$ τότε παίρνουμε τις «μυγαδικές ροπές» που ορίζονται ως

$$c_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + iy)^p (x - iy)^q f(x, y) dx dy \quad (2.3)$$

Οι γεωμετρικές και οι μυγαδικές ροπές μεταφέρουν την ίδια ποσότητα πληροφορίας. Κάθε μυγαδική ροπή μπορεί να εκφραστεί με όρους γεωμετρικής ροπής όμοιας τάξης

$$c_{pq} = \sum_{k=0}^p \sum_{j=0}^q \binom{p}{k} \binom{q}{j} (-1)^{q-j} \cdot i^{p+q-k-j} \cdot m_{k+j,p+q-k-j} \quad (2.4)$$

και αντίστροφα

$$m_{pq} = \frac{1}{2^{p+q} \cdot i^q} \sum_{k=0}^p \sum_{j=0}^q \binom{p}{k} \binom{q}{j} (-1)^{q-j} \cdot c_{k+j,p+q-k-j} \quad (2.5)$$

2.4 Αναλλοίωτες Μεταβλητές Ροπών ως προς την μετατόπιση, την περιστροφή και την κλιμάκωση

Η μετατόπιση, η περιστροφή και η κλιμάκωση (ΜΠΚ) είναι οι πιο απλοί μετασχηματισμοί χωρικών συντεταγμένων. Ο ΜΠΚ, στην βιβλιογραφία (Teague, 1980) αναφέρεται και ως μετασχηματισμός ομοιότητας, είναι ένας μετασχηματισμός τεσσάρων παραμέτρων για 2Δ σχήματα στο επίπεδο, ο οποίος μπορεί να περιγραφεί ως

$$\mathbf{x}' = sR \cdot \mathbf{x} + \mathbf{t}$$

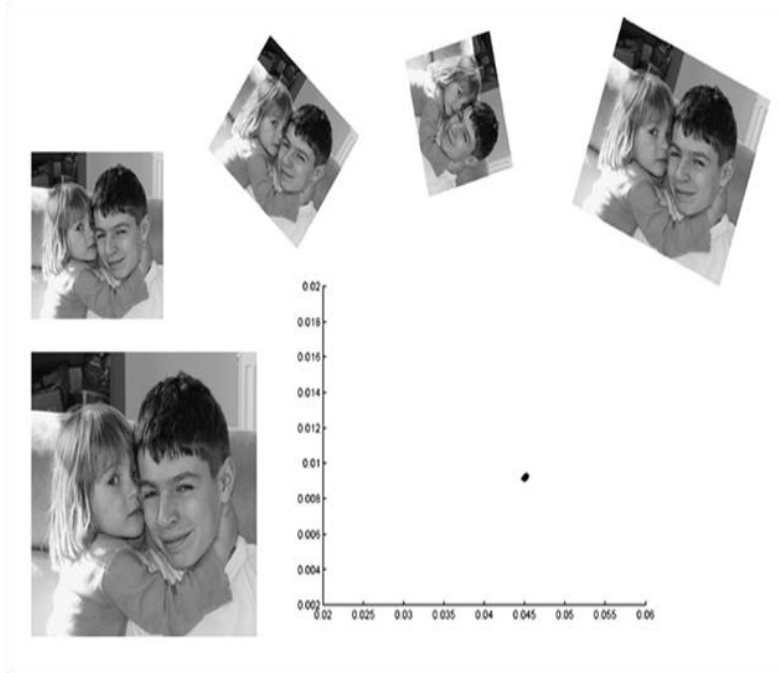
όπου \mathbf{t} είναι το διάνυσμα μετατόπισης, s θετικός συντελεστής κλιμάκωσης (θεωρούμε ότι έχουμε ομοιόμορφη κλιμάκωση μόνο όταν ο συντελεστής κλιμάκωσης είναι ίδιος τόσο για την οριζόντια όσο και για την κάθετη κατεύθυνση) και R ο πίνακας περιστροφής

$$R = \begin{pmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{pmatrix}$$

όπου a η γωνία περιστροφής.

Η αμεταβλητότητα ως προς τον μετασχηματισμό ΜΠΚ είναι απαιτούμενη σχεδόν σε όλες τις πρακτικές εφαρμογές, επειδή το αντικείμενο πρέπει να αναγνωριστεί σωστά, ασχέτως με την θέση, τον προσανατολισμό του και την απόσταση του από την κάμερα (Teague, 1980).

Επιπλέον, το μοντέλο ΜΠΚ παρέχει μια προσέγγιση της πραγματικής παραμόρφωσης της εικόνας όταν η σκηνή είναι επίπεδη και κάθετη στον οπτικό άξονα. Γι αυτό τον λόγο έχει δοθεί πολύ έμφαση στις αναλλοίωτες μεταβλητές ΜΠΚ. Ωστόσο, αν και οι αναλλοίωτες μεταβλητές ως προς την μετατόπιση και την κλιμάκωση είναι πολύ εύκολο να βρεθούν, η εύρεση των αναλλοίωτων μεταβλητών ως προς την περιστροφή είναι πολύ πιο σύνθετη.



Σχήμα 2.2: Η επιθυμητή συμπεριφορά των ΜΠΚ αναλλοίωτων μεταβλητών ροπής. Σε όλες τις περιπτώσεις, η περιστραμμένη και κλιμακωμένη εικόνα έχει σχεδόν την ίδια τιμή με την αμετάβλητη αρχική εικόνα (Flusser & Suk, 2009).

2.4.1 Αναλλοίωτες μεταβλητές ως προς την μετατόπιση

Το αναλλοίωτο ως προς την μετατόπιση μπορεί να επιτευχθεί με δύο τρόπους. Απλά με την μετατόπιση του αντικειμένου ώστε το κέντρο μάζας του να συμπίπτει με την αρχή των αξόνων ή, αντίστροφα, με την μετατόπιση της πολυωνυμικής βάσης στο κέντρο μάζας του αντικειμένου.

Στην περίπτωση των γεωμετρικών αναλλοίωτων μεταβλητών που εξετάζουμε, ένας τέτοιος μετασχηματισμός οδηγεί στις λεγόμενες *κεντρικές ροπές*

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y) dx dy \quad (2.6)$$

όπου

$$x_c = m_{10}/m_{00} \text{ και } y_c = m_{01}/m_{00}$$

είναι οι συντεταγμένες του κέντρου μάζας του αντικειμένου.

Από τον εξίσωση (2.6) προκύπτει άμεσα ότι $\mu_{10} = \mu_{01} = 0$ και $\mu_{00} = m_{00}$. Το αναλλοίωτο ως προς την μετατόπιση για τις κεντρικές ροπές είναι προφανές.

Οι κεντρικές ροπές μπορούν να εκφραστούν με όρους γεωμετρικών ροπών ως εξής

$$\mu_{pq} = \sum_{k=0}^p \sum_{j=0}^q \binom{p}{k} \binom{q}{j} (-1)^{k+j} x_c^k y_c^j m_{p-k, q-j} \quad (2.7)$$

2.4.2 Αναλλοίωτες μεταβλητές ως προς την ομοιόμορφη κλιμάκωση

Το αναλλοίωτο ως προς την κλιμάκωση αποκτιέται με κατάλληλη κανονικοποίηση κάθε ροπής. Κατ' αρχήν, κάθε ροπή μπορεί να χρησιμοποιηθεί ως παράγοντας κανονικοποίησης εφόσον είναι μη μηδενική για όλες τις εικόνες του πειράματος. Από την στιγμή που οι ροπές χαμηλής τάξης είναι πιο ανθεκτικές στον θόρυβο και ευκολότερες να υπολογιστούν, κανονικοποιούμε τις περισσότερες φορές με κατάλληλη δύναμη του μ_{00}

$$v_{pq} = \frac{\mu_{pq}}{\mu_{00}^w} \quad (2.8)$$

όπου

$$w = \frac{p+q}{2} + 1 \quad (2.9)$$

Η ροπή v_{pq} ονομάζεται *κανονικοποιημένη κεντρική ροπή*.

Προκειμένου μια συνάρτηση να θεωρηθεί αναλλοίωτη ως προς την κλιμάκωση πρέπει να ισχύει $h(\lambda x) = \lambda^\Delta h(x)$, όπου Δ η τάξη της κλιμάκωσης. Για ομοιόμορφη κλιμάκωση έχουμε $h(\lambda x) = \lambda h(x)$, ενώ για συναρτήσεις με δυο μεταβλητές η προηγούμενη σχέση γίνεται $h(\lambda x, \lambda y) = \lambda^2 h(x, y)$.

Έστω ότι κάνουμε ομοιόμορφη κλιμάκωση χρησιμοποιώντας τον συντελεστή s τόσο για την κάθετη όσο και για την οριζόντια διεύθυνση. Για την συνάρτηση εικόνας $f(x, y)$ θα έχουμε $f(sx, sy) = s^2 f(x, y)$, οπότε, οι κεντρικές ροπές στις νέες συντεταγμένες θα είναι

$$\begin{aligned} \mu'_{pq} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x' - x'_c)^p (y' - y'_c)^q f'(x', y') dx' dy' \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s^p (x - x_c)^p s^q (y - y_c)^q f(x, y) s^2 dx dy \\ &= s^{p+q+2} \mu_{pq} \end{aligned} \quad (2.10)$$

Έχουμε ότι,

$$\mu'_{00} = s^2 \mu_{00}$$

Συνεπώς, χρησιμοποιώντας τις εξισώσεις (2.8) και (2.9) παίρνουμε

$$v'_{pq} = \frac{\mu'_{pq}}{(\mu'_{00})^w} = \frac{s^{p+q+2} \cdot \mu_{pq}}{(s^2 \mu_{00})^w} = v_{pq}$$

Κάτι που αποδεικνύει την αμεταβλητότητα ως προς την κλιμάκωση των κανονικοποιημένων ροπών.

Ως μειονέκτημα των κανονικοποιημένων ροπών μπορεί να θεωρηθεί ότι η ροπή που χρησιμοποιήθηκε για κανονικοποίηση δεν μπορεί πλέον να χρησιμοποιηθεί για αναγνώριση καθώς η τιμή της μετά την κανονικοποίηση είναι πάντα ένα (στην παραπάνω κανονικοποίηση έχουμε, $v_{00} = 1$). Εάν θέλουμε να κρατήσουμε την τιμή της μηδενικής τάξης ροπής, μπορούμε να κανονικοποιήσουμε με διαφορετικής τάξης ροπή.

2.4.3 Αναλλοίωτες μεταβλητές Hu

Οι αναλλοίωτες μεταβλητές ροπής για την περιστροφή, πρωτοεμφανίστηκαν το 1962 από τον Hu (Hu, 1962), ο οποίος επεξεργαζόμενος τα αποτελέσματα της θεωρίας των αλγεβρικών αναλλοίωτων μεταβλητών κατέληξε στις επτά διάσημες αναλλοίωτες μεταβλητές του ως προς την περιστροφή

$$\phi_1 = m_{20} + m_{02}$$

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

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

$$\phi_4 = (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2$$

$$\phi_5 = (m_{30} - 3m_{12})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) \\ + (3m_{21} - m_{03})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2)$$

$$\phi_6 = (m_{20} - m_{02})((m_{30} + m_{12})^2 - (m_{21} + m_{03})^2) \\ + 4m_{11}(m_{30} + m_{12})(m_{21} + m_{03})$$

$$\phi_7 = (3m_{21} - m_{03})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) \\ - (m_{30} - 3m_{12})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2)$$

Εάν στις παραπάνω σχέσεις αντικαταστήσουμε τις γεωμετρικές ροπές με κεντρικές ή κανονικοποιημένες ροπές, τότε αποκτούμε αναλλοίωτες μεταβλητές όχι μόνο ως προς την περιστροφή αλλά επίσης ως προς την μετατόπιση ή/και την κλιμάκωση.

Όπως επισημαίνουν οι Flusser & Suk (2009) , ο τρόπος με τον οποίο ο Hu εξήγαγε τις αναλλοίωτες μεταβλητές του ήταν ιδιαίτερα περίπλοκος και γι αυτό το λόγο μόνο επτά παρουσιάστηκαν χωρίς να δίνεται κάποια μεθοδολογία για την εξαγωγή αναλλοίωτων μεταβλητών μεγαλύτερης τάξης.

Ωστόσο από την στιγμή που έχουμε τους τύπους, η απόδειξη για την αμεταβλητότητα ως προς την περιστροφή είναι εύκολη. Ας αποδείξουμε την ιδιότητα αυτή για τις ϕ_1 και ϕ_2 χρησιμοποιώντας τις κεντρικές ροπές.

Η δεύτερης τάξης ροπές, μετά την περιστροφή, μπορούν να εκφραστούν ως

$$\mu'_{20} = \cos^2 a \cdot \mu_{20} + \sin^2 a \cdot \mu_{02} - \sin 2a \cdot \mu_{11}$$

$$\mu'_{02} = \sin^2 a \cdot \mu_{20} + \cos^2 a \cdot \mu_{02} + \sin 2a \cdot \mu_{11}$$

$$\mu'_{11} = \frac{1}{2} \sin 2a \cdot (\mu_{20} - \mu_{02}) + \cos 2a \cdot \mu_{11}$$

Οπότε

$$\phi'_1 = \mu'_{20} + \mu'_{02} = (\sin^2 a + \cos^2 a)(\mu_{20} + \mu_{02}) = \phi_1$$

Ομοίως έχουμε και για το ϕ'_2 εφαρμόζοντας τον τύπο, $\cos 2a = \cos^2 a - \sin^2 a$

Μπορεί οι αναλλοίωτες μεταβλητές του Hu να έχουν περιορισμένη ικανότητα αναγνώρισης, αμοιβαία εξάρτηση και να περιορίζονται στην χρήση της δεύτερης και τρίτης τάξης ροπής μόνο, ωστόσο έχουν γίνει κλασσικές και παρόλα τα μειονεκτήματά τους έχουν χρησιμοποιηθεί με επιτυχία σε διάφορους τομείς. Το μεγαλύτερο μειονέκτημα της θεωρίας του Hu είναι ότι δεν παρέχουν την δυνατότητα για περαιτέρω γενίκευση.

2.5 Αφινικός Μετασχηματισμός

Ο αφινικός μετασχηματισμός είναι γενικά ένα γραμμικός μετασχηματισμός των χωρικών συντεταγμένων της εικόνας. Ο αφινικός μετασχηματισμός μπορεί να εκφραστεί ως

$$x' = a_0 + a_1x + a_2y$$

$$y' = b_0 + b_1x + b_2y$$

και σε μορφή πίνακα

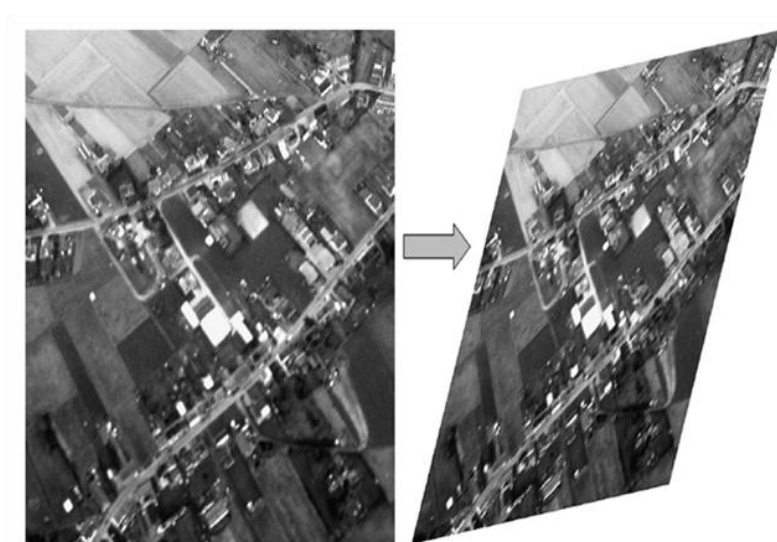
$$x' = Ax + b$$

Όπου

$$A = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \text{ και } b = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix}$$

Ο αφινικός μετασχηματισμός απεικονίζει ένα τετράγωνο σε ένα παραλληλόγραμμο διατηρώντας την συγγραμμικότητα (βλ. Σχήμα 2.3). Προφανώς, ο μετασχηματισμός ΜΠΚ (Μετατόπισης, Περιστροφής, Κλιμάκωσης) είναι απλώς μια ειδική περίπτωση του αφινικού μετασχηματισμού όπου $a_1 = b_2$ και $a_2 = -b_1$. Ορίζουμε ως μήτρα του αφινικού μετασχηματισμού την $J = a_1b_2 - a_2b_1$.

Πρέπει να επισημανθεί ότι ο αφινικός μετασχηματισμός είναι μια ειδική περίπτωση της προοπτικής προβολής μέσω της οποίας απεικονίζονται τα τρισδιάστατα αντικείμενα στην κάμερα. Πιο συγκεκριμένα, αν το αντικείμενο είναι μικρό σε σχέση με την απόσταση της κάμερας από την σκηνή, τότε η προοπτική επίδραση γίνεται αμελητέα και το αφινικό μοντέλο αποτελεί μια ικανοποιητική προσέγγιση του προοπτικού μοντέλου. Γι αυτό τον λόγο ο αφινικός μετασχηματισμός και οι αναλλοίωτες μεταβλητές του είναι τόσο σημαντικοί για την υπολογιστική όραση.



Σχήμα 2.3: Ο αφινικός μετασχηματισμός απεικονίζει ένα τετράγωνο σε ένα παραλληλόγραμμο (Flusser & Suk, 2009).

2.6 Αναλλοίωτες μεταβλητές ως προς αφινικό μετασχηματισμό (Affine Moment Invariants - AMIs)

Για τους λόγους που αναφερθήκαν προηγουμένως, τα AMIs παίζουν σημαντικό ρόλο στην αναγνώριση αντικειμένου ανεξάρτητα με την θέση της κάμερας και έχουν χρησιμοποιηθεί ευρέως.

Χρησιμοποιούνται διάφορες μεθοδολογίες για την εξαγωγή των AMIs. Εδώ θα παρουσιάσουμε τον πιο προφανή και απλό τρόπο, δηλαδή, αυτόν που χρησιμοποιεί την θεωρία γράφων (για περισσότερες πληροφορίες για τον τρόπο εξαγωγής των AMIs, παραπέμπουμε στο βιβλίο των Flusser & Suk (2009)). Η συγκεκριμένη μεθοδολογία παρέχει τα εργαλεία για τον προσδιορισμό οποιασδήποτε τάξης AMIs, ενώ το μεγαλύτερο πλεονέκτημα της είναι ότι επειδή περιέχει γραφική απεικόνιση των AMIs μας επιτρέπει να εντοπίσουμε τυχόν εξαρτήσεις που παρουσιάζονται μεταξύ τους και να τις εξαλείψουμε.

Πάντως, ένα κοινό σημείο όλων των μεθόδων είναι ότι μπορούν να παράγουν όσες αναλλοίωτες μεταβλητές κάποιος επιθυμεί, ωστόσο, μερικές από αυτές είναι εξαρτημένες. Από την στιγμή που οι εξαρτημένες αναλλοίωτες μεταβλητές είναι πρακτικά άχρηστες, συνήθως στις εφαρμογές χρησιμοποιούνται μόνο οι ανεξάρτητες αν και είναι περιορισμένος ο αριθμός τους.

2.6.1 AMIs εξαγόμενα από γράφους

Ας θεωρήσουμε μια εικόνα f και δύο τυχαία σημεία $(x_1, y_1), (x_2, y_2) \in D$. Έστω C_{12} το «εξωτερικό γινόμενο» (cross-product) τους

$$C_{12} = x_1 y_2 - x_2 y_1$$

Εφαρμόζοντας τον αφινικό μετασχηματισμό (θεωρούμε μετασχηματισμούς που δεν περιέχουν μετατόπιση $b = 0$) ισχύει ότι

$$C'_{12} = J \cdot C_{12}$$

Το οποίο σημαίνει ότι η C_{12} αποτελεί μια σχετικά αναλλοίωτη μεταβλητή ως προς τον αφινικό μετασχηματισμό.

Θεωρούμε έναν αριθμό από σημεία και ολοκληρώνουμε τα διαγώνια γινόμενα τους (ή κάποια δύναμη των διαγωνίων γινομένων τους) στο συμπαγές σύνολο της f το D . Αυτά τα ολοκληρώματα μπορούν να εκφραστούν με όρους ροπών και μετά την εξάλειψη των Ιακωβιανών J με κατάλληλη κανονικοποίηση, καταλήγουμε σε αναλλοίωτες μεταβλητές ως προς τον αφινικό μετασχηματισμό.

Πιο συγκεκριμένα, έστω ότι έχουμε r σημεία ($r \geq 2$). Ορίζουμε συνάρτηση I που εξαρτάται από το r και τους μη αρνητικούς ακεραίους n_{kj} ως

$$I(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{k,j=1}^r C_{kj}^{n_{kj}} \cdot \prod_{i=1}^r f(x_i, y_i) dx_i dy_i \quad (2.11)$$

Θεωρούμε ότι $j > k$, καθώς $C_{kj} = -C_{jk}$ και $C_{kk} = 0$.

Εφαρμόζοντας τον αφινικό μετασχηματισμό, η $I(f)$ γίνεται

$$I'(f) = J^w |J|^r \cdot I(f)$$

όπου $w = \sum_{k,j} n_{kj}$ και ονομάζεται βάρος της αναλλοίωτης μεταβλητής και r είναι ο βαθμός της,

Εάν η $I(f)$ κανονικοποιηθεί με τον όρο μ_{00}^{w+r} , αποκτούμε την επιθυμητή αναλλοίωτη μεταβλητή ως προς τον αφινικό μετασχηματισμό

$$\left(\frac{I(f)}{\mu_{00}^{w+r}} \right)' = \left(\frac{I(f)}{\mu_{00}^{w+r}} \right) \quad (2.12)$$

Εάν w περιττός και $J < 0$, τότε η εξίσωση (2.12) έχει έναν επιπλέον παράγοντα -1

Η μέγιστη τάξη των ροπών από τις οποίες η αναλλοίωτη μεταβλητή συντίθεται, αποτελεί και την τάξη της αναλλοίωτης μεταβλητής. Η τάξη αυτή είναι πάντα μικρότερη ή ίση με το βάρος. Ένα επίσης σημαντικό στοιχείο της αναλλοίωτης μεταβλητής είναι η δομή της. Η δομή της αναλλοίωτης μεταβλητής ορίζεται από ένα διάνυσμα ακεραίων $s = (k_2, k_3, \dots, k_s)$, όπου s είναι η τάξη της αναλλοίωτης μεταβλητής και k_j είναι ο συνολικός αριθμός των ροπών τάξης j που περιέχονται σε κάθε όρο της αναλλοίωτης μεταβλητής.

Παρουσιάζουμε την μέθοδο, που δίνεται από τον τύπο (2.11) για δυο απλές αναλλοίωτες μεταβλητές. Αρχικά, έστω $r = 2$ και $w = n_{12} = 2$. Τότε

$$\begin{aligned} I(f) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 y_2 \\ &\quad - x_2 y_1)^2 f(x_1, y_1) f(x_2, y_2) dx_1 dx_2 dy_2 \\ &= 2(m_{20} m_{02} - m_{11}^2) \end{aligned} \quad (2.13)$$

Αν αντικαταστήσουμε στην εξίσωση (2.13) τις γεωμετρικές ροπές με τις αντίστοιχες κεντρικές ροπές και κανονικοποιήσουμε την αναλλοίωτη μεταβλητή με τον

παράγοντα μ_{00}^4 , αποκτούμε μια πλήρως αναλλοίωτη μεταβλητή ως προς τον γενικό αφινικό μετασχηματισμό.

$$I_1 = \frac{1}{\mu_{00}^4} (\mu_{20}\mu_{02} - \mu_{11}^2) \quad (2.14)$$

Αυτή είναι η πιο απλή αναλλοίωτη μεταβλητή ως προς τον αφινικό μετασχηματισμό που περιέχει μόνο ροπές δεύτερης τάξης.

Με παρόμοιο τρόπο, αλλάζοντας ωστόσο κάθε φορά τις παραμέτρους w και r , μπορούμε να εξάγουμε μεγαλύτερης τάξης αναλλοίωτες μεταβλητές ως προς τον αφινικό μετασχηματισμό όπως αναφέρουν οι Flusser & Suk (2009).

Για λόγους απλότητας παραθέτουμε εδώ μόνο τα τέσσερα πρώτα AMIs.

$$I_1 = \frac{1}{\mu_{00}^4} (\mu_{20}\mu_{02} - \mu_{11}^2)$$

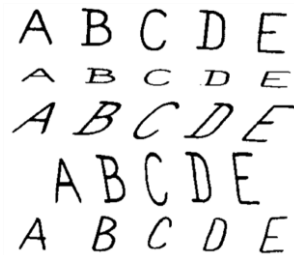
$$I_2 = \frac{1}{\mu_{00}^{10}} (-\mu_{30}^2\mu_{03}^2 + 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} - 4\mu_{30}\mu_{12}^3 - 4\mu_{21}^3\mu_{03} + 3\mu_{21}^2\mu_{12}^2)$$

$$I_3 = \frac{1}{\mu_{00}^7} (\mu_{20}\mu_{21}\mu_{03} - \mu_{20}\mu_{12}^2 - \mu_{11}\mu_{30}\mu_{03} + \mu_{11}\mu_{21}\mu_{12} + \mu_{02}\mu_{30}\mu_{12} - \mu_{02}\mu_{21}^2)$$

$$I_4 = \frac{1}{\mu_{00}^{11}} (-\mu_{20}^3\mu_{03}^2 + 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 3\mu_{20}^2\mu_{02}\mu_{12}^2 - 6\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} - 6\mu_{20}\mu_{11}^2\mu_{12}^2 + 12\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} - 3\mu_{20}\mu_{02}^2\mu_{21}^2 + 2\mu_{11}^3\mu_{30}\mu_{03} + 6\mu_{11}^3\mu_{21}\mu_{12} - 6\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}^2\mu_{02}\mu_{21}^2 + 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} - \mu_{02}^3\mu_{30}^2)$$

2.7 Αριθμητικά παραδείγματα

Παρουσιάζουμε εδώ τα αποτελέσματα του πειράματος που διεξήγαγαν οι Flusser & Suk (1994). Το ακόλουθο πείραμα δείχνει την ισχύ που έχουν τα AMIs στην περιγραφή και αναγνώριση ενός αντικειμένου.



Σχήμα 2.4: Οι χαρακτήρες που χρησιμοποιήθηκαν στο πείραμα. Στη 1^η γραμμή περιέχονται τα πρότυπα εκπαίδευσης, ενώ, στις υπόλοιπες άγνωστοι χαρακτήρες (Flusser & Suk, 1994)

- Δεδομένα.
Χειρόγραφοι κεφαλαίοι χαρακτήρες A, B, C, D και E, χρησιμοποιήθηκαν στο πείραμα (Σχήμα 2.4). Οι χαρακτήρες τις πρώτης γραμμής είναι τα πρότυπα εισόδου (δεδομένα εκπαίδευσης) και όλοι οι υπόλοιποι χαρακτήρες στις επόμενες γραμμές θεωρούνται άγνωστοι. Οπότε σκοπός του πειράματος, είναι να αναγνωριστούν αυτοί οι χαρακτήρες μόνο με ένα τόσο μικρό σύνολο εκπαίδευσης. Όπως φαίνεται στο Σχήμα 2.4, οι χαρακτήρες στις γραμμές δύο έως τέσσερα, έχουν υποστεί αφινικό μετασχηματισμό. Ενώ, στην γραμμή 5 οι χαρακτήρες έχουν υποστεί τυχαίο μη γραμμικό μετασχηματισμό.
- Εξαγωγή Χαρακτηριστικών
Αρχικά, χρησιμοποιήθηκε η μέθοδος της κατωφλιοποίησης για την εξαγωγή των εικόνων. Περαιτέρω, επεξεργασία εικόνας, όπως λέπτυνση, δεν έγινε. Έπειτα, τέσσερις αναλλοίωτες μεταβλητές H_u και τέσσερις μεταβλητές AMIs υπολογίστηκαν για κάθε χαρακτήρα ξεχωριστά.
- Μέθοδος Ταξινόμησης
Η ταξινόμηση των χαρακτήρων έγινε χρησιμοποιώντας τον ταξινομητή ελαχίστης απόστασης στον τετραδιάστατο ευκλείδειο χώρο χαρακτηριστικών. Κάθε χαρακτήρας ταξινομήθηκε δύο φορές, μια χρησιμοποιώντας τις αναλλοίωτες μεταβλητές H_u και άλλη μια χρησιμοποιώντας τα AMIs. Σε περίπτωση που πρέπει να αναγνωριστεί ένα μεγαλύτερο σύνολο χαρακτήρων, τότε μπορεί να χρησιμοποιηθεί ο K-means.
- Αποτελέσματα
Τα πειραματικά αποτελέσματα συνοψίζονται στους δυο παρακάτω πίνακες.

Γραμμή	A	B	C	D	E
1	576	593	1399	793	1119
2	810→D	617→B	1145→E	701→B	880→D
3	741→D	765→D	1918→C	1102→E	1587→C
4	679→D	815→D	2027→C	1099→E	1553→C
5	566→A	567→B	1118→E	814→D	1259→C

Πίνακας 2.1: Οι τιμές της πρώτης μεταβλητής H_u την Φ_1 για τους χαρακτήρες κάθε γραμμής. Η πρώτη γραμμή χρησιμοποιείται ως δεδομένα εκπαίδευσης ενώ στις υπόλοιπες παρουσιάζονται τα αποτελέσματα της αναγνώρισης (Flusser & Suk, 1994).

Όπως είναι εμφανές στο πίνακα 1, η χρησιμοποίηση της πρώτης μεταβλητής H_u για την αναγνώριση, δίνει πολύ χαμηλά αποτελέσματα επιτυχίας.

Γραμμή	A	B	C	D	E
1	82	72	311	115	200
2	82→A	73→B	297→C	105→D	166→E
3	81→A	70→B	287→C	112→D	189→E
4	80→A	72→B	299→C	109→D	176→E
5	74→A	58→B	173→E	116→D	210→E

Πίνακας 2.2: Οι τιμές της πρώτης μεταβλητής AMI την I_1 για τους χαρακτήρες κάθε γραμμής. Η πρώτη γραμμή χρησιμοποιείται ως δεδομένα εκπαίδευσης ενώ στις υπόλοιπες παρουσιάζονται τα αποτελέσματα της αναγνώρισης (Flusser & Suk, 1994).

Όπως μπορούμε να παρατηρήσουμε η τιμή της I_1 για τον ίδιο χαρακτήρα δεν αλλάζει σημαντικά για τις τέσσερις πρώτες γραμμές, όπου οι χαρακτήρες μας έχουν υποστεί αφινικό μετασχηματισμό, ενώ, η απόκλιση που παρουσιάζει για την πέμπτη γραμμή είναι ανεκτή.

Επίσης, πρέπει να επισημάνουμε ότι οι τιμές μεταξύ διαφορετικών χαρακτήρων έχουν μεγάλη απόκλιση, κάτι που μας βοηθάει στο να τους ταξινομήσουμε. Σε αντίθεση, η τιμή της ϕ_1 αλλάζει σημαντικά για τον ίδιο χαρακτήρα από γραμμή σε γραμμή, ενώ για διαφορετικούς χαρακτήρες δεν έχουμε μεγάλη απόκλιση στις μεταξύ τους τιμές με αποτέλεσμα η αναγνώριση να είναι τις περισσότερες φορές εσφαλμένη.

Τα αποτελέσματα των παραπάνω πινάκων δείχνουν ότι οι αναλλοίωτες μεταβλητές H_u δεν μπορούν να χρησιμοποιηθούν ως χαρακτηριστικά για αναγνώριση αντικειμένων που έχουν υποστεί αφινικό μετασχηματισμό. Μόνο το 30% των χαρακτήρων αναγνωρίστηκε ορθώς.

Αντίθετα, οι αναλλοίωτες μεταβλητές ως προς τον αφινικό μετασχηματισμό (AMIs) έχουν επιτυχία πάνω από 95% και το μικρό ποσοστό αποτυχίας οφείλεται στους χαρακτήρες της πέμπτης γραμμής που έχουν υποστεί μη γραμμική παραμόρφωση.

Όπως αναφέρουν οι Flusser & Suk (1994) από διάφορα πειράματα που έχουν διεξαχθεί, όπως αυτό που παρουσιάσαμε παραπάνω, το ποσοστό αναγνώρισης σε περίπτωση που χρησιμοποιήσουμε αναλλοίωτες μεταβλητές H_u δεν κυμαίνεται παραπάνω από 40%, ενώ στην περίπτωση των AMIs έχουν ποσοστό που σε όλες τις περιπτώσεις υπερβαίνει το 90%.

ΚΕΦΑΛΑΙΟ 3

3 Τεχνητά Νευρωνικά Δίκτυα – Αυτοοργανούμενοι Χάρτες (SOM)

3.1 Εισαγωγή

Τα Τεχνητά Νευρωνικά Δίκτυα δημιουργήθηκαν προκειμένου να μπορέσουμε να προσομοιάσουμε την πολύπλοκη λειτουργία του ανθρώπινου εγκεφάλου. Ο ανθρώπινος εγκέφαλος, μπορεί να θεωρηθεί ως ένας εξαιρετικά πολύπλοκος, μη γραμμικός υπολογιστής που εκτελεί παράλληλα πολλαπλές διεργασίες. Οι διεργασίες αυτές μπορεί να είναι ταυτόχρονη επεξεργασία εικόνας, αναγνώριση, μάθηση και αλληλεπίδραση με το περιβάλλον.

Συνεπώς, ένα τεχνητό νευρωνικό δίκτυο είναι μια υπολογιστική μηχανή σχεδιασμένη να εκτελεί μια συγκεκριμένη εργασία ή λειτουργία όπως ακριβώς και ο ανθρώπινος εγκέφαλος. Η δομή του τεχνητού νευρωνικού δικτύου μιμείται κατά το δυνατό εκείνη του βιολογικού νευρωνικού δικτύου, ώστε να εμφανίζει παρόμοιες ιδιότητες. Χρησιμοποιεί, λοιπόν, ένα σύνολο από διασυνδεδεμένα μεταξύ τους υπολογιστικά κύτταρα, που ονομάζονται «νευρώνες». Ο βαθμός αλληλεπίδρασης είναι διαφορετικός για κάθε ζεύγος νευρώνων και καθορίζεται από τα λεγόμενα «συναπτικά βάρη». Η ισχύ των συνδέσεων μεταξύ των νευρώνων, που δίνεται από τα «συναπτικά βάρη», χρησιμοποιείται για την αποθήκευση της γνώσης.

Συγκεκριμένα, καθώς το νευρωνικό δίκτυο αλληλεπιδρά με το περιβάλλον του και μαθαίνει από αυτό, τα συναπτικά βάρη μεταβάλλονται συνεχώς, ενδυναμώνοντας ή αποδυναμώνοντας την ισχύ του κάθε δεσμού. Όλη η εμπειρική γνώση που αποκτά,

επομένως, το νευρωνικό δίκτυο από το περιβάλλον κωδικοποιείται στα συναπτικά βάρη. Αυτά αποτελούν το χαρακτηριστικό εκείνο που δίνει στο δίκτυο την ικανότητα για εξέλιξη και προσαρμογή στο περιβάλλον.

Ακριβώς όπως υπάρχουν διαφορετικοί τρόποι με τους οποίους μαθαίνουν οι άνθρωποι από το περιβάλλον τους, το ίδιο ισχύει και για τα νευρωνικά δίκτυα. Μπορούμε να κατηγοριοποιήσουμε τις διαδικασίες μάθησης μέσω των οποίων λειτουργούν τα νευρωνικά δίκτυα ως εξής: *επιβλεπόμενη μάθηση*, *ενισχυτική μάθηση* και *μη επιβλεπόμενη μάθηση*.

Κατά την *επιβλεπόμενη μάθηση* μας δίνεται ένα σύνολο από ζευγάρια προτύπων εκπαίδευσης (x, y) , $x \in X, y \in Y$ όπου X το σύνολο εισόδου και Y το σύνολο εξόδου. Στόχος της επιβλεπόμενης μάθησης είναι το νευρωνικό δίκτυο να προσομοιάσει την συνάρτηση f που απεικονίζει το σύνολο εισόδου στο σύνολο εξόδου $f: X \rightarrow Y$. Προκειμένου να έχουμε μια ένδειξη κατά πόσο το νευρωνικό μας δίκτυο έχει εκπαιδευτεί επαρκώς, χρησιμοποιούμε κάποιες συναρτήσεις κόστους όπως η συνάρτηση «μέσου τετραγωνικού σφάλματος». Για τη μείωση της τιμής της συνάρτησης κόστους χρησιμοποιούνται διάφοροι αλγόριθμοι όπως της κλίσης (gradient), απότομης καθόδου (steepest descent) κ.α. Τα πολυεπίπεδα Perceptron αποτελούν χαρακτηριστικό δείγμα νευρωνικού δικτύου επιβλεπόμενης μάθησης

Στην *ενισχυτική μάθηση*, τα δεδομένα εισόδου x δεν δίνονται, αλλά παράγονται από την αλληλεπίδραση του συστήματος μας με το περιβάλλον. Για κάθε χρονική στιγμή t , το σύστημα μας εκτελεί μια ενέργεια y_t και το περιβάλλον δίνει απόκριση σ' αυτή την ενέργεια x_t καθώς και ένα στιγμιαίο κόστος c_t , σύμφωνα με κάποια δυναμική απόκριση που είναι συνήθως άγνωστη. Σκοπός, λοιπόν, του νευρωνικού μας είναι να παράγει ενέργειες y_t που ελαχιστοποιούν μια μακροπρόθεσμη ποσότητα όπως το αθροιστικό κόστος. Συνεπώς, στόχος της ενισχυτικής μάθησης δεν είναι να ελαχιστοποιήσει το στιγμιαίο κόστος, αλλά να βρει την αλληλουχία ερεθισμάτων που μας ελαχιστοποιούν το αθροιστικό κόστος.

Τέλος, στην *μη επιβλεπόμενη μάθηση*, ή αυτοοργανούμενη μάθηση, δεν υπάρχει εξωτερικός εκπαιδευτής που να επιβλέπει την διαδικασία μάθησης. Αντί αυτού, υπάρχει ένα ανεξάρτητο από την εργασία μέτρο της ποιότητας της αναπαράστασης που καλείται να μάθει το δίκτυο και οι ελεύθερες παράμετροι του δικτύου βελτιστοποιούνται σε σχέση με αυτό το μέτρο. Οι αυτοοργανούμενοι χάρτες που αναλύονται εκτενώς παρακάτω ανήκουν στην κατηγορία της μη επιβλεπόμενης μάθησης.

3.2 Αυτοοργανούμενοι Χάρτες

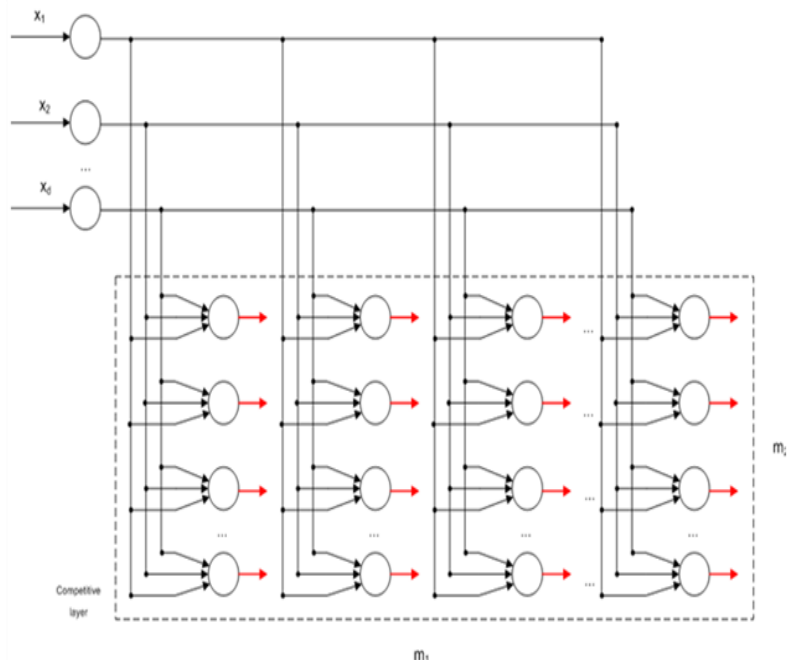
Μια ειδική κατηγορία τεχνητών νευρωνικών δικτύων αποτελούν οι αυτοοργανούμενοι χάρτες (Self-Organizing Maps, SOM). Αυτά τα δίκτυα βασίζονται στην ανταγωνιστική μάθηση. Οι νευρώνες εξόδου του δικτύου ανταγωνίζονται μεταξύ τους για το δικαίωμα ενεργοποίησης, με αποτέλεσμα μόνο ένας νευρώνας

εξόδου να είναι ενεργός ανά πάσα στιγμή. Ο νευρώνας εξόδου που νικά στον ανταγωνισμό (νικητής νευρώνας) απολαμβάνει το καθεστώς που αποκαλείται “winner takes all” (“ο νικητής τα παίρνει όλα”).

Σε έναν αυτοοργανούμενο χάρτη, οι νευρώνες τοποθετούνται στους κόμβους ενός πλέγματος το οποίο είναι συνήθως μονοδιάστατο ή δισδιάστατο. Οι νευρώνες συντονίζονται επιλεκτικά σε διάφορα πρότυπα εισόδου (ερεθίσματα) κατά την πορεία μιας διαδικασίας ανταγωνιστικής μάθησης. Οι θέσεις των νευρώνων που συντονίζονται κατ’ αυτό τον τρόπο, διατάσσονται ο ένας σε σχέση με τον άλλο με τέτοιο τρόπο, έτσι ώστε να δημιουργείται ένα λογικό σύστημα συντεταγμένων για διαφορετικά χαρακτηριστικά εισόδου πάνω στο πλέγμα.

Συνεπώς, ένας αυτοοργανούμενος χάρτης χαρακτηρίζεται από τον σχηματισμό ενός τοπογραφικού χάρτη αποτελούμενου από τα πρότυπα εισόδου, στον οποίο οι χωρικές θέσεις των νευρώνων στο πλέγμα είναι ενδεικτικές των εσωτερικών στατιστικών χαρακτηριστικών που περιέχονται στα πρότυπα εισόδου – σ’ αυτό οφείλεται και ο χαρακτηρισμός τους ως «αυτοοργανούμενοι χάρτες». Ένας αυτοοργανούμενος χάρτης είναι εκ’ φύσεως μη γραμμικός.

Ο κύριος στόχος ενός αυτοοργανούμενου χάρτη (SOM) είναι να μετασχηματίζει ένα πρότυπο εισερχόμενου σήματος, τυχαίας διάστασης, σ’ έναν διακριτό χάρτη μιας ή δυο διαστάσεων και να εκτελεί αυτό το μετασχηματισμό προσαρμοστικά, με κάποιον τοπολογικά διατεταγμένο τρόπο.



Σχήμα 3.1: Δισδιάστατο πλέγμα νευρώνων, με είσοδο τριών διαστάσεων και έξοδο 4x4 διαστάσεων.

Στο σχήμα 3.1 παρουσιάζεται το σχηματικό διάγραμμα ενός δισδιάστατου πλέγματος νευρώνων, το οποίο χρησιμοποιείται ευρέως ως διακριτός χάρτης. Κάθε νευρώνας στο πλέγμα είναι πλήρως συνδεδεμένος με όλους τους πηγαίους κόμβους του επιπέδου εισόδου. Αυτό το δίκτυο αντιπροσωπεύει μια δομή πρόσθιας τροφοδότησης με ένα υπολογιστικό επίπεδο αποτελούμενο από νευρώνες διατεταγμένους σε γραμμές και στήλες. Ένα μονοδιάστατο πλέγμα είναι μια ειδική περίπτωση της διαμόρφωσης που απεικονίζεται στο σχήμα 3.1: σ' αυτήν την ειδική περίπτωση, το υπολογιστικό επίπεδο αποτελείται απλώς από μια στήλη ή γραμμή νευρώνων.

Κάθε πρότυπο εισόδου που παρουσιάζεται στο δίκτυο αποτελείται τυπικά από ένα «σημείο δραστηριότητας». Η θέση και η φύση ενός τέτοιου σημείου συνήθως μεταβάλλεται από το ένα στιγμιότυπο του προτύπου εισόδου στο επόμενο. Συνεπώς, όλοι οι νευρώνες του δικτύου θα πρέπει να εκτίθενται σε επαρκή αριθμό διαφορετικών στιγμιότυπων του προτύπου εισόδου για να διασφαλιστεί ότι η διαδικασία αυτοοργάνωσης θα έχει την ευκαιρία να αναπτυχθεί σωστά.

Ο αλγόριθμος που είναι υπεύθυνος για τον σχηματισμό του αυτοοργανούμενου χάρτη ξεκινά αρχικοποιώντας τα συναπτικά βάρη στο δίκτυο. Αυτό μπορεί να γίνει αναθέτοντας τους μικρές τιμές, επιλεγμένες από μια γεννήτρια τυχαίων αριθμών. Κατ' αυτό τον τρόπο, δεν επιβάλλεται κάποια αρχική «σειρά» στο χάρτη χαρακτηριστικών. Αφού το δίκτυο αρχικοποιηθεί, υπάρχουν τρεις σημαντικές διαδικασίες που εμπλέκονται στο σχηματισμό του αυτοοργανούμενου χάρτη

1. *Ανταγωνισμός*: Για κάθε πρότυπο εισόδου, οι νευρώνες του δικτύου υπολογίζουν τις αντίστοιχες τιμές μιας συνάρτησης διάκρισης. Αυτή η συνάρτηση διάκρισης παρέχει τη βάση για τον ανταγωνισμό μεταξύ των νευρώνων. Ο συγκεκριμένος νευρώνας με τη μεγαλύτερη τιμή στη συνάρτηση διάκρισης δηλώνεται νικητής του ανταγωνισμού
2. *Συνεργασία*: Ο νικητής νευρώνας καθορίζει την χωρική θέση μιας τοπολογικής γειτονιάς διεγερμένων νευρώνων, παρέχοντας έτσι τη βάση για συνεργασία μεταξύ τέτοιων γειτονικών νευρώνων.
3. *Προσαρμογή Συναπτικών Βαρών*: Αυτός ο τελευταίος μηχανισμός επιτρέπει στους διεγερμένους νευρώνες να αυξάνουν τις τιμές της συνάρτησης διάκρισης σε σχέση με το πρότυπο εισόδου μέσω κατάλληλων προσαρμογών που εφαρμόζονται στα συναπτικά βάρη τους. Οι προσαρμογές που γίνονται είναι τέτοιες ώστε η απόκριση του νικητή νευρώνα στην επόμενη εφαρμογή ενός παρόμοιου προτύπου εισόδου να είναι βελτιωμένη.

Οι διαδικασίες του ανταγωνισμού, συνεργασίας και προσαρμογής των συναπτικών βαρών περιγράφονται αναλυτικότερα στις επόμενες ενότητες που ακολουθούν (Haykin, 2009).

3.2.1 Η Διαδικασία Ανταγωνισμού

Έστω m η διάσταση του χώρου εισόδου (δεδομένων). Έστω, επίσης, ότι ένα πρότυπο εισόδου (διάνυσμα) που επιλέγεται τυχαία από το σύνολο εκπαίδευσης και συμβολίζεται ως

$$\mathbf{x} = [x_1, x_2, \dots, x_m]^T$$

Το διάνυσμα συναπτικών βαρών κάθε νευρώνα του δικτύου έχει ίδια διάσταση με το χώρο εισόδου. Έστω ότι το διάνυσμα συναπτικών βαρών του νευρώνα j συμβολίζεται ως

$$\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jm}]^T, \quad j = 1, 2, \dots, l$$

όπου l ο συνολικός αριθμός νευρώνων του δικτύου. Για να βρούμε την βέλτιστη ταύτιση του διανύσματος εισόδου \mathbf{x} με τα διανύσματα συναπτικών βαρών \mathbf{w}_j , συγκρίνουμε τα εσωτερικά γινόμενα $\mathbf{w}_j^T \mathbf{x}$ για $j = 1, 2, \dots, l$ και επιλέγουμε το μεγαλύτερο. Αυτή η μέθοδος υποθέτει ότι εφαρμόζεται το ίδιο κατώφλι σε όλους τους νευρώνες. Το κατώφλι είναι το αρνητικό της πόλωσης. Έτσι, επιλέγοντας το νευρώνα με το μεγαλύτερο εσωτερικό γινόμενο $\mathbf{w}_j^T \mathbf{x}$ ουσιαστικά θα έχουμε καθορίσει τη θέση όπου πρόκειται να κεντραριστεί η τοπολογική γειτονιά των διεγερμένων νευρώνων.

Όπως είναι γνωστό το κριτήριο βέλτιστης ταύτισης, βάσει μεγιστοποίησης του εσωτερικού γινομένου $\mathbf{w}_j^T \mathbf{x}$ είναι μαθηματικώς ισοδύναμο με την ελαχιστοποίηση της ευκλείδειας απόστασης μεταξύ των διανυσμάτων \mathbf{x} και \mathbf{w}_j , υπό τον όρο ότι το \mathbf{w}_j έχει μοναδιαίο μήκος για όλα τα j . Εάν χρησιμοποιήσουμε τον δείκτη $i(\mathbf{x})$ για το νευρώνα που ταιριάζει καλύτερα με το διάνυσμα εισόδου \mathbf{x} , μπορούμε κατόπιν να καθορίσουμε το $i(\mathbf{x})$ εφαρμόζοντας την ακόλουθη συνθήκη, η οποία συνοψίζει και την ουσία της διαδικασίας ανταγωνισμού μεταξύ των νευρώνων

$$i(\mathbf{x}) = \arg(\min_j \|\mathbf{x} - \mathbf{w}_j\|) \quad (3.1)$$

Σύμφωνα με την Εξίσωση (3.1), το $i(\mathbf{x})$ είναι το σημείο εστίασης της προσοχής επειδή θέλουμε να βρούμε την ταυτότητα του νευρώνα i . Ο συγκεκριμένος νευρώνας που ικανοποιεί αυτή την συνθήκη αποκαλείται νευρώνας νικητής για το διάνυσμα εισόδου \mathbf{x} . Η Εξίσωση (3.1) οδηγεί στην ακόλουθη παρατήρηση (Haykin, 2009)

Ένας συνεχής χώρος εισόδου από πρότυπα ενεργοποίησης αντιστοιχίζεται σ' ένα διακριτό χώρο εξόδου νευρώνων μέσω μιας διαδικασίας ανταγωνισμού μεταξύ των νευρώνων του δικτύου

Ανάλογα με την εκάστοτε εφαρμογή, η απόκριση του δικτύου θα μπορούσε να είναι είτε ο δείκτης του νικητή νευρώνα (δηλαδή η θέση του στο πλέγμα) είτε το διάνυσμα συναπτικών βαρών που είναι πλησιέστερο στο διάνυσμα εισόδου, με την Ευκλείδεια έννοια.

3.2.2 Διαδικασία Συνεργασίας

Ο νικητής νευρώνας εντοπίζει το κέντρο μιας τοπολογικής γειτονιάς συνεργαζόμενων νευρώνων. Συγκεκριμένα, ένας νευρώνας που ενεργοποιείται τείνει να διεγείρει τους νευρώνες που βρίσκονται στην άμεση γειτονιά του περισσότερο από εκείνους που βρίσκονται πιο μακριά. Αυτή η παρατήρηση μας οδηγεί στο να ορίσουμε μια τοπολογική γειτονιά γύρω από τον νικητή νευρώνα i και να την κάνουμε να φθίνει ομαλά με την πλευρική απόσταση (Ritter et al., 1992).

Έστω, λοιπόν, ότι το h_{ji} συμβολίζει την τοπολογική γειτονιά που είναι κεντραρισμένη γύρω από το νικητή νευρώνα i και περικλείει ένα σύνολο διεγερμένων (συνεργαζόμενων) νευρώνων, ένα τυπικό δείγμα των οποίων συμβολίζεται ως j . Τότε για την συνάρτηση h_{ji} μπορούμε να υποθέσουμε ότι ικανοποιεί τις παρακάτω δύο απαιτήσεις (Haykin, 2009):

1. Η τοπολογική γειτονιά h_{ji} είναι συμμετρική γύρω από το μέγιστο σημείο που ορίζεται από $d_{ji} = 0$, δηλαδή, αποκτά την μέγιστη τιμή της στο νικητή νευρώνα i για τον οποίο η απόσταση d_{ji} είναι μηδέν
2. Το πλάτος της τοπολογικής γειτονιάς h_{ji} , μειώνεται μονοτονικά με την αύξηση της πλευρικής απόστασης d_{ji} , φθίνοντας στο μηδέν για $d_{ji} \rightarrow \infty$. Αυτή είναι μια αναγκαία συνθήκη για την σύγκλιση.

Μια καλή επιλογή του h_{ji} που ικανοποιεί αυτές τις απαιτήσεις είναι η Γκαουσιανή συνάρτηση

$$h_{j,i(x)} = e^{-\frac{d_{ji}^2}{2\sigma^2}} \quad (3.2)$$

η οποία είναι ανεξάρτητη από την θέση του νικητή νευρώνα i . Η διακύμανση σ είναι το «ενεργό» εύρος της τοπολογικής γειτονιάς και μετρά το βαθμό που οι διεγερμένοι νευρώνες στην κοντινή περιοχή του νευρώνα νικητή συμμετέχουν στη διαδικασία μάθησης. Σε ποιοτικό επίπεδο, η Γκαουσιανή τύπου τοπολογική γειτονιά της Εξίσωσης (3.2) είναι πιο κατάλληλη απ' ό,τι μια ορθογώνια καθώς κάνει τον αλγόριθμο SOM να συγκλίνει πιο γρήγορα ενώ οι κοντινοί νευρώνες στον νικητή δεν λαμβάνουν όλοι το ίδιο ποσοστό στην διαδικασία μάθησης (Erwin et al., 1992).

Για να υπάρξει, λοιπόν, συνεργασία μεταξύ γειτονικών νευρώνων, είναι αναγκαίο η τοπολογική γειτονιά $h_{j,i(x)}$ να εξαρτάται από την πλευρική απόσταση d_{ji} μεταξύ του νικητή i και του διεγερμένου νευρώνα j στο χώρο εξόδου και όχι από κάποιο μέτρο απόστασης στον αρχικό χώρο εισόδου.

Αυτό ακριβώς έχουμε στην Εξίσωση (3.2). Στην περίπτωση του μονοδιάστατου πλέγματος, το d_{ji} είναι ένας ακέραιος $d_{ji} = |j - i|$. Από την άλλη, στην περίπτωση ενός δισδιάστατου πλέγματος ορίζεται ως

$$d_{ji} = \|r_j - r_i\|^2$$

Όπου το διακριτό διάνυσμα r_j ορίζει τη θέση του διεγερμένου νευρώνα j και το r_i ορίζει τη θέση του νικητή νευρώνα i , αμφότερες εκ' των οποίων μετριοούνται στον διακριτό χώρο εξόδου.

Ένα άλλο μοναδικό χαρακτηριστικό του αλγορίθμου SOM είναι ότι το μέγεθος της τοπολογικής γειτονιάς επιτρέπεται να συρρικνώνεται με τον χρόνο. Αυτή η απαίτηση ικανοποιείται κάνοντας το εύρος σ της τοπολογικής συνάρτησης h_{ji} να μειώνεται με τον χρόνο. Μια δημοφιλής επιλογή για την εξάρτηση του σ από το διακριτό χρόνο n είναι η εκθετική μείωση που περιγράφεται ως

$$\sigma(n) = \sigma_0 e^{-\frac{n}{\tau_1}} \quad n = 0, 1, 2, \dots$$

Όπου σ_0 είναι η τιμή του σ κατά την έναρξη του αλγορίθμου SOM και τ_1 είναι μια σταθερά χρόνου. Αντίστοιχα, η συνάρτηση τοπολογικής γειτονιάς λαμβάνει μια δική της, μεταβαλλόμενη μορφή στο χρόνο (Haykin, 2009)

$$h_{j,i(x)}(n) = e^{-\frac{d_{ji}^2}{2(\sigma(n))^2}} \quad (3.3)$$

Έτσι καθώς αυξάνεται ο διακριτός χρόνος n (δηλαδή ο αριθμός των επαναλήψεων), το εύρος $\sigma(n)$ μειώνεται με εκθετικό ρυθμό και η τοπολογική γειτονιά συρρικνώνεται ανάλογα.

Ωστόσο θα πρέπει να επισημάνουμε ότι η συνάρτηση γειτονιάς τελικά θα έχει και πάλι τιμή μονάδας για τον νικητή νευρώνα i , εφόσον η απόσταση d_{ji} για τον νευρώνα j υπολογίζεται στον χώρο του πλέγματος.

3.2.3 Διαδικασία Προσαρμογής

Η διαδικασία προσαρμογής είναι η τελευταία φάση του αυτοοργανούμενου σχηματισμού ενός χάρτη χαρακτηριστικών. Για να είναι αυτοοργανούμενο το δίκτυο,

το διάνυσμα συναπτικών βαρών \mathbf{w}_j του νευρώνα j του δικτύου πρέπει να προσαρμόζεται σε σχέση με το διάνυσμα εισόδου \mathbf{x} .

Όπως αναφέρει ο Haykin (2009), στο αίτημα μάθησης του Hebb, ένα συναπτικό βάρος αυξάνεται με την ταυτόχρονη ύπαρξη προσυναπτικής και μετα-συναπτικής δραστηριότητας. Η χρήση ενός τέτοιου κανόνα είναι ιδιαίτερα κατάλληλη για διαδικασίες συσχετιστικής μάθησης. Ωστόσο, για το είδος της μη επιβλεπόμενης μάθησης η Χεμπιανή υπόθεση, στη βασική της μορφή, δεν είναι ικανοποιητική, για τον ακόλουθο λόγο:

Οι αλλαγές στις συνδέσεις λαμβάνουν χώρα προς μια διεύθυνση μόνο, οδηγώντας τελικά όλα τα συναπτικά βάρη στον κορεσμό. Οπότε η Χεμπιανή υπόθεση πρέπει να τροποποιηθεί ώστε να συμπεριλαμβάνει τον όρο λησμόνησης $g(y_j)\mathbf{w}_j$, όπου \mathbf{w}_j το διάνυσμα συναπτικών βαρών του νευρώνα j , και $g(y_j)$ είναι κάποια βαθμωτή συνάρτηση της απόκρισης y_j .

Η μόνη απαίτηση που επιβάλλεται στην συνάρτηση $g(y_j)$ είναι ο σταθερός όρος στο ανάπτυγμα Taylor της $g(y_j)$ να είναι μηδέν, έτσι ώστε να μπορεί να γραφεί

$$g(y_j) = 0 \quad \text{για } y_j = 0 \quad (3.4)$$

Η σπουδαιότητα αυτής της απαίτησης θα γίνει εμφανής παρακάτω. Δοθείσας μιας τέτοιας συνάρτησης, μπορούμε κατόπιν να εκφράσουμε την αλλαγή στο διάνυσμα βαρών του νευρώνα j του πλέγματος ως εξής

$$\Delta \mathbf{w}_j = \eta y_j \mathbf{x} - g(y_j)\mathbf{w}_j \quad (3.5)$$

όπου η η παράμετρος ρυθμού μάθησης του αλγορίθμου.

Ο πρώτος όρος στη δεξιά πλευρά της Εξίσωσης (3.5) είναι ο Χεμπιανός όρος, ενώ ο δεύτερος είναι ο όρος λησμόνησης. Προκειμένου να ικανοποιηθεί η απαίτηση της εξίσωσης (3.4), επιλέγουμε μια γραμμική συνάρτηση για την $g(y_j)$.

$$g(y_j) = \eta y_j$$

Θεωρώντας επίσης ότι

$$y_j = h_{j,i(x)}$$

αποκτούμε την τελική μορφή της εξίσωσης για την αλλαγή των συναπτικών βαρών των νευρώνων

$$\Delta \mathbf{w}_j = \eta h_{j,i(x)}(\mathbf{x} - \mathbf{w}_j) \quad (3.6)$$

Τέλος, χρησιμοποιώντας φορμαλισμό διακριτού χρόνου, με δεδομένο το διάνυσμα συναπτικών βαρών $\mathbf{w}_j(k)$ του νευρώνα j τη χρονική στιγμή k , ορίζουμε το ενημερωμένο διάνυσμα βαρών $\mathbf{w}_j(k + 1)$ για την χρονική στιγμή $k + 1$ ως εξής

$$\mathbf{w}_j(k + 1) = \mathbf{w}_j(k) + \eta(k) h_{j,i(x)}(k) (\mathbf{x}(k) - \mathbf{w}_j(k)) \quad (3.7)$$

Η Εξίσωση (3.7) εφαρμόζεται σε όλους τους νευρώνες του πλέγματος που βρίσκονται μέσα στην τοπολογική γειτονιά του νικητή νευρώνα i . Συνεπώς, η Εξίσωση (3.7) έχει ως αποτέλεσμα τη μετακίνηση του διανύσματος συναπτικών βαρών \mathbf{w}_i του νικητή νευρώνα i προς το διάνυσμα εισόδου \mathbf{x} . Μετά από επαναλαμβανόμενες παρουσιάσεις των δεδομένων εκπαίδευσης, τα διανύσματα συναπτικών βαρών τείνουν να ακολουθούν την κατανομή των διανυσμάτων εισόδου λόγω της ενημέρωσης που λαμβάνει χώρα στη γειτονιά. Συνεπώς, ο αλγόριθμος οδηγεί σε μια τοπολογική διάταξη (τακτοποίηση) του χάρτη χαρακτηριστικών στο χώρο εισόδου, υπό την έννοια ότι οι νευρώνες που είναι γειτονικοί στο πλέγμα θα τείνουν να έχουν παρόμοια διανύσματα συναπτικών βαρών.

Προϋπόθεση για να υπολογιστεί η εξίσωση (3.7), που μας δίνει τα σηναπτικά βάρη των κόμβων του SOM, είναι να έχει πρότερα υπολογιστεί η συνάρτηση ρυθμού μάθησης $\eta(k)$, η οποία είναι μεταβαλλόμενη με τον χρόνο. Συγκεκριμένα, θα πρέπει να ξεκινά από κάποια αρχική τιμή η_0 και κατόπιν να μειώνεται βαθμιαία με τον χρόνο k . Αυτή η απαίτηση μπορεί να ικανοποιηθεί με τον ακόλουθο ευρετικό κανόνα

$$\eta(n) = \eta_0 \cdot e^{-\frac{k}{\tau_2}} \quad (3.8)$$

όπου τ_2 είναι ακόμα μια σταθερά του αλγορίθμου SOM. Σύμφωνα με αυτό τον δεύτερο ευρετικό κανόνα, η παράμετρος ρυθμού μάθησης φθίνει εκθετικά με τον χρόνο k .

3.2.4 Οι Δύο Φάσεις της Προσαρμοστικής Διαδικασίας: Διάταξη και Σύγκλιση

Αρχικά, η τοπολογική διάταξη του αυτοοργανούμενου χάρτη είναι απροσδιόριστη. Καθώς λοιπόν εκτελείται ο αλγόριθμος της προσαρμοστικής διαδικασίας, ο αυτοοργανούμενος χάρτης οδηγείται σταδιακά σε μια οργανωμένη αναπαράσταση των προτύπων ενεργοποίησης που αντλούνται από τον χώρο εισόδου.

Η διαδικασία προσαρμογής των συναπτικών βαρών του δικτύου, όπως αυτά υπολογίζονται σύμφωνα με την εξίσωση (3.7), μπορεί να γίνει σε δύο διαδοχικές φάσεις: μια φάση διάταξης (τακτοποίησης, αυτοοργάνωσης) και μια φάση σύγκλισης.

Παρακάτω αναλύουμε τις δυο φάσεις από τις οποίες αποτελείται ο αλγόριθμος της προσαρμοστικής διαδικασίας όπως αυτές παρουσιάστηκαν πρώτη φορά από τον Kohonen (1982).

1. Φάση Διάταξης

Κατά την διάρκεια αυτής της πρώτης φάσης της προσαρμοστικής διαδικασίας λαμβάνει χώρα η τοπολογική διάταξη των διανυσμάτων βαρών των νευρώνων του αυτοοργανούμενου χάρτη. Η φάση τοπολογικής διάταξης συνήθως απαιτεί αρκετές επαναλήψεις του αλγορίθμου SOM, ο αριθμός των οποίων καθορίζεται ανάλογα με τα δεδομένα του εκάστοτε προβλήματος.

Παραδείγματος χάριν, αν είναι επιθυμητό να προσεγγιστεί το η χωρική κατανομή μιας εκθετικής συνάρτησης με τη χρήση λίγων δεδομένων εκπαίδευσης, τότε ο αριθμός των επαναλήψεων του αλγορίθμου SOM για την φάση της διάταξης πρέπει να είναι περιορισμένος. Αυτό γίνεται προκειμένου να μην συγκεντρωθούν όλοι οι νευρώνες γύρω από αυτά τα λίγα δεδομένα αλλά να «απλωθούν» και στο χώρο ενδιάμεσα τους ώστε να προσεγγιστεί καλύτερα η εκθετική συνάρτηση. Αντίθετα, εάν το πρόβλημα μας είναι να διαχωρίσουμε τα χρώματα μιας τυχαίας εικόνας και μας δίνονται λίγα δεδομένα που αντιπροσωπεύουν ξεχωριστά χρώματα τότε επιθυμούμε ο αριθμός των επαναλήψεων του αλγορίθμου SOM για την φάση της διάταξης να είναι μεγάλος ώστε οι νευρώνες να συγκεντρωθούν στην εγγύς περιοχή των δεδομένων εκπαίδευσης.

Συνεπώς, μπορεί ο αριθμός των επαναλήψεων του αλγορίθμου SOM να εξαρτάται από το εκάστοτε πρόβλημα ωστόσο θα πρέπει να δοθεί ιδιαίτερη προσοχή στην επιλογή της παραμέτρου ρυθμού μάθησης και της συνάρτησης γειτονιάς, όπως περιγράφεται στη συνέχεια

- Η παράμετρος ρυθμού μάθησης $\eta(k)$ συνήθως ξεκινά με μια μεγάλη τιμή όπως το 0.1 ώστε αρχικά να έχουμε ταχεία σύγκλιση μεταξύ των ενεργοποιημένων νευρώνων και των δειγμάτων ενεργοποίησης. Από το σημείο αυτό και μετά, θα πρέπει να μειώνεται βαθμιαία, αλλά να παραμένει πάνω από μια μικρή τιμή όπως 0.01 (δηλαδή δεν θα πρέπει ποτέ να φτάνει στο μηδέν). Αυτές οι επιθυμητές τιμές ικανοποιούνται με τις ακόλουθες επιλογές στον τύπο της εξίσωσης (3.8)

$$\eta_0 = 0.1$$

$$\tau_2 = \text{αριθμός επαναλήψεων αλγορίθμου SOM}$$

- Η συνάρτηση γειτονιάς $h_{j,i}(k)$ θα πρέπει αρχικά να περιλαμβάνει σχεδόν όλους τους νευρώνες του δικτύου κεντραρισμένους ως προς τον νικητή νευρώνα i και κατόπιν να «συρρικνώνεται» αργά με την πάροδο του χρόνου

Συγκεκριμένα, κατά τη διάρκεια της φάσης τοπολογικής διάταξης, η $h_{j,i(x)}(k)$ θα πρέπει να μπορεί να μειωθεί σε μια μικρή τιμή που θα αντιπροσωπεύει

μόνο λίγους γειτονικούς νευρώνες γύρω από το νικητή νευρώνα, ή στο νικητή νευρώνα μόνο. Υποθέτοντας ότι χρησιμοποιείται ένα δισδιάστατο πλέγμα νευρώνων για το διακριτό χάρτη, μπορούμε να καθορίσουμε το αρχικό μέγεθος σ_0 της συνάρτησης γειτονιάς ίσο με την «ακτίνα» του πλέγματος. Αντίστοιχα, μπορούμε να καθορίσουμε τη σταθερά χρόνου τ_1 στον τύπο της εξίσωσης για τον $\sigma(k)$ ίσο με

$$\tau_1 = \frac{\tau_2}{\log \sigma_0}$$

2. Φάση Σύγκλισης

Αυτή η δεύτερη φάση της προσαρμοστικής διαδικασίας απαιτείται για την εκτέλεση «λεπτών προσαρμογών» στον αυτοοργανούμενο χάρτη έτσι ώστε να παρέχει μια κατά το δυνατόν επακριβή αναπαράσταση του χώρου εισόδου. Επιπλέον, αν και ο αριθμός των επαναλήψεων που απαιτούνται για την σύγκλιση εξαρτάται σε μεγάλο βαθμό από τη διάσταση του χώρου εισόδου. Ένας γενικός κανόνας είναι ο εξής (Kohonen, 1982):

Ο αριθμός επαναλήψεων στη φάση σύγκλισης πρέπει να είναι τουλάχιστον 500-πλάσιος από τον αριθμό των νευρώνων του δικτύου. Έτσι, η φάση σύγκλισης μπορεί να χρειαστεί χιλιάδες επαναλήψεις. Η επιλογή της παραμέτρου ρυθμού μάθησης και της συνάρτησης γειτονιάς θα πρέπει να γίνεται ως εξής:

- Για καλή στατιστική ακρίβεια, η παράμετρος ρυθμού μάθησης $\eta(k)$ θα πρέπει να διατηρείται κατά τη διάρκεια της φάσης σύγκλισης σε μια μικρή τιμή, της τάξης του 0.01. Όπως προαναφέραμε, η $h_{j,i(x)}(n)$ δεν θα πρέπει να επιτρέπεται να μειωθεί στο μηδέν. Διαφορετικά, είναι πιθανό το δίκτυο μας να «παγιδευτεί» σε μια μετασταθή κατάσταση. Μια μετασταθής κατάσταση ανήκει σε μια διαμόρφωση του χάρτη χαρακτηριστικών με κάποια τοπολογική ατέλεια. Η εκθετική μείωση της εξίσωσης (3.8) διασφαλίζει ότι δεν πρόκειται να υπάρξουν μετασταθείς καταστάσεις.
- Η συνάρτηση γειτονιάς $h_{j,i}(k)$ θα πρέπει να περιέχει μόνο τους πλησιέστερους γείτονες ενός νικητή νευρώνα, αριθμός ο οποίος μπορεί τελικά να μειωθεί σε έναν ή μηδέν γειτονικούς νευρώνες.

3.3 Σύνοψη του Αλγορίθμου SOM

Η ουσία του αλγορίθμου SOM του Kohonen (1982) είναι ότι αντικαθιστά τις πολυπλοκότερες και πιο λεπτομερείς ιδιότητες του «Χεμπιανού τύπου» κανόνα και των πλευρικών αλληλεπιδράσεων με έναν απλό γεωμετρικό υπολογισμό. Οι βασικές συνιστώσες και παράμετροι του αλγορίθμου είναι:

- Ένας συνεχής χώρος εισόδου αποτελούμενος από πρότυπα ενεργοποίησης
- Μια τοπολογία του δικτύου με την μορφή ενός πλέγματος νευρώνων, η οποία ορίζει ένα διακριτό χώρο εξόδου
- Μια μεταβαλλόμενη στο χρόνο συνάρτηση γειτονιάς $h_{j,i(x)}(k)$ η οποία ορίζεται γύρω από ένα νικητή νευρώνα $i(x)$
- Μια παράμετρος ρυθμού μάθησης $\eta(k)$ η οποία ξεκινά από μια αρχική τιμή η_0 και κατόπιν μειώνεται βαθμιαία με τον χρόνο k , αλλά δεν φτάνει ποτέ στο μηδέν

Όσον αφορά τη συνάρτηση γειτονιάς και την παράμετρο ρυθμού μάθησης, μπορούμε να χρησιμοποιήσουμε τις εξισώσεις (3.3) και (3.8) αντίστοιχα, για την φάση τοπολογικής διάταξης. Για καλή στατιστική ακρίβεια η $\eta(n)$ θα πρέπει να διατηρείται σε μικρή τιμή κατά τη διάρκεια της σύγκλισης, για σχετικά μεγάλο χρονικό διάστημα. Η συνάρτηση γειτονιάς θα πρέπει να περιέχει μόνο τους πλησιέστερους γείτονες του νικητή νευρώνα στην αρχή της φάσης σύγκλισης και μπορεί τελικά να μειωθεί σε έναν ή μηδέν γειτονικούς νευρώνες.

Τρία είναι τα βασικά βήματα που εμπλέκονται στην εφαρμογή του αλγορίθμου μετά από την αρχικοποίηση: δειγματοληψία, ταίριασμα ομοιότητας και ενημέρωση. Αυτά τα τρία βήματα επαναλαμβάνονται μέχρι να ολοκληρωθεί ο σχηματισμός του αυτοοργανούμενου χάρτη. Ο αλγόριθμος συνοψίζεται ως εξής.

1. *Αρχικοποίηση* : Επέλεξε τυχαίες τιμές για τα αρχικά διανύσματα βαρών $\mathbf{w}_j(0)$. Ο μόνος περιορισμός εδώ είναι ότι το $\mathbf{w}_j(0)$ πρέπει να είναι διαφορετικό για $j = 1, 2, \dots, l$ όπου l ο αριθμός των νευρώνων στο πλέγμα. Μπορεί να είναι επιθυμητό να κρατήσουμε τις τιμές των βαρών μικρές. Ένας άλλος τρόπος αρχικοποίησης του αλγορίθμου είναι να επιλέξουμε τα διανύσματα βαρών από το διαθέσιμο σύνολο διανυσμάτων εισόδου με τυχαίο τρόπο.
2. *Δειγματοληψία* : Πάρε ένα δείγμα \mathbf{x} από το χώρο εισόδου με μια συγκεκριμένη πιθανότητα. Το διάνυσμα \mathbf{x} αντιπροσωπεύει το πρότυπο ενεργοποίησης που εφαρμόζεται στο πλέγμα. Η διάσταση του διανύσματος \mathbf{x} ισούται με m .

3. *Ταίριασμα Ομοιότητας* : Βρες τον «πλέον ταιριαστό» (νικητή) νευρώνα $i(\mathbf{x})$ στο χρονικό βήμα n χρησιμοποιώντας το κριτήριο ελαχίστης απόστασης.

$$i(\mathbf{x}) = \arg(\min_j \|\mathbf{x} - \mathbf{w}_j\|)$$

4. *Ενημέρωση* : Προσάρμοσε το διάνυσμα συναπτικών βαρών όλων των ενεργοποιημένων νευρώνων χρησιμοποιώντας τον τύπο ενημέρωσης

$$\mathbf{w}_j(k+1) = \mathbf{w}_j(k) + \eta(k) h_{j,i(\mathbf{x})}(k) (\mathbf{x}(k) - \mathbf{w}_j(k))$$

5. *Έλεγχος* : Συνέχισε με το βήμα 2 μέχρι να μην παρατηρούνται πλέον ευδιάκριτες αλλαγές στον αυτοοργανούμενο χάρτη.

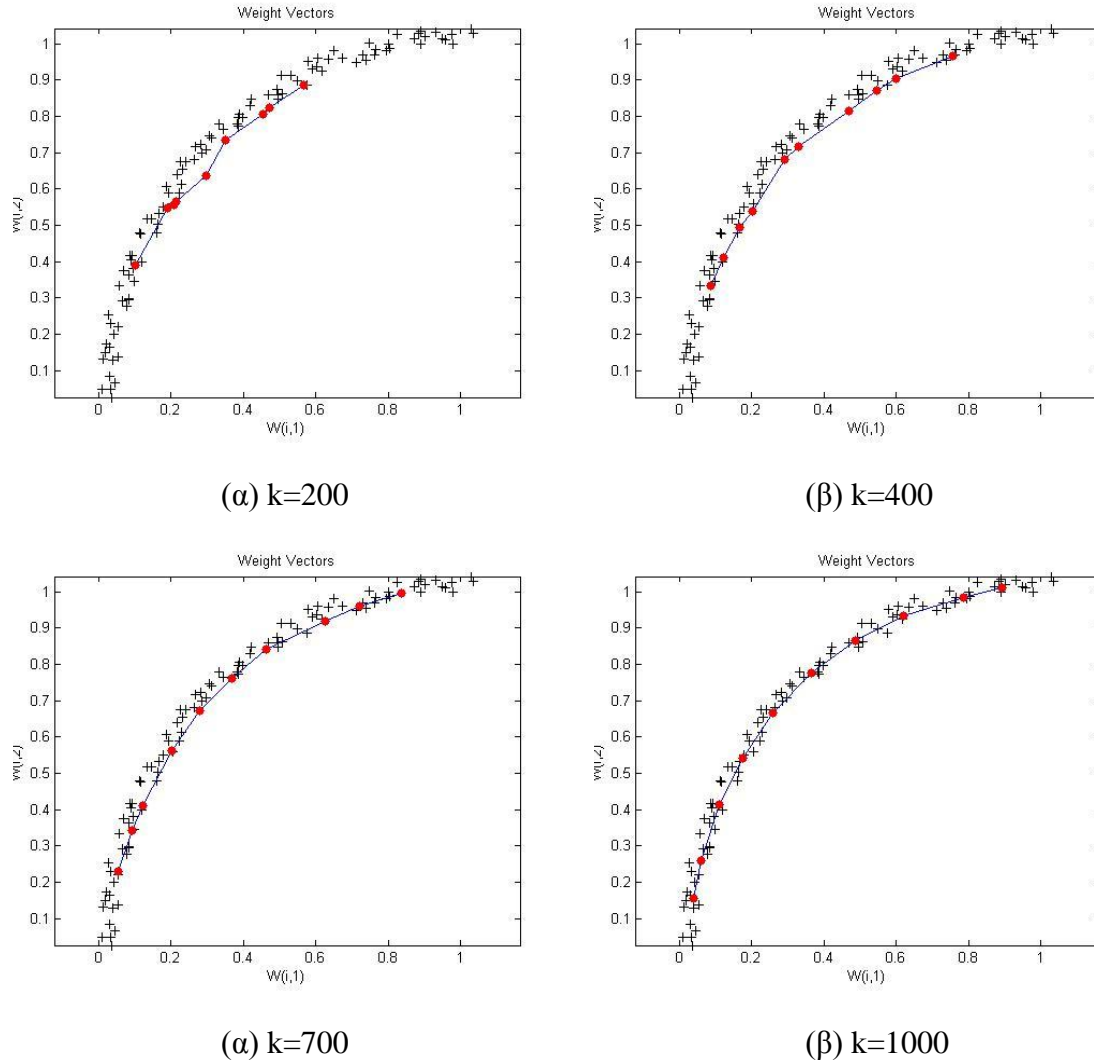
3.4 Αριθμητικά Παραδείγματα

Αφού υλοποιήσαμε ένα SOM σύμφωνα με τα παραπάνω θα επιχειρήσουμε την μελέτη και την ανάλυση των ιδιοτήτων, των δυνατοτήτων και των επιδόσεων του αυτοοργανούμενου χάρτη.

Η χρησιμοποιούμενη μεθοδολογία, όπως θα φανεί και στη συνέχεια, είναι η πειραματική διερεύνηση/αντιμετώπιση χαρακτηριστικών προβλημάτων με στόχο να διαφωτιστούν σημεία που χρήζουν κάποιας εξέτασης και σχολιασμού. Κατά κύριο λόγο χρησιμοποιούμε προβλήματα χαμηλών διαστάσεων, έτσι ώστε να παρέχεται η δυνατότητα (μέσω της οδού της οπτικοποίησης) για εξαγωγή συμπερασμάτων και για βαθύτερη κατανόηση των επιτελούμενων λειτουργιών.

Δημιουργούμε δύο δισδιάστατα σύνολα δεδομένων εκπαίδευσης απλώς με την αποτύπωση σημείων στο χώρο. Η χωρική κατανομή του πρώτου συνόλου εκπαίδευσης που αποτελείται από 100 σημεία στο επίπεδο, προσομοιάζει μια εκθετική συνάρτηση, ενώ του δεύτερου συνόλου εκπαίδευσης που αποτελείται από 600 σημεία στο επίπεδο, προσομοιάζει ένα ρόμβο. Στα παρακάτω σχήματα που παραθέτουμε παρουσιάζουμε πως το SOM, κατά την διάρκεια της εκπαίδευσης του προσεγγίζει αυτές τις δύο χωρικές κατανομές. Για το πρώτο σύνολο εκπαίδευσης, χρησιμοποιούμε μονοδιάστατο πλέγμα με 10 νευρώνες ενώ για το δεύτερο σύνολο εκπαίδευσης χρησιμοποιούμε δισδιάστατο πλέγμα 10×10 νευρώνων.

Έχουμε αρχικοποιήσει τους νευρώνες του SOM ώστε οι αρχικές τιμές τους να βρίσκονται μεταξύ μέγιστης και ελάχιστης τιμής του συνόλου εκπαίδευσης. Όπως είναι εμφανές στο σχήμα 3.2, οι νευρώνες έχουν μια τυχαία αρχική θέση και όσο περνάει ο χρόνος εκπαίδευσης τόσο η τοπολογία τους πλησιάζει την εκθετική. Μάλιστα είναι εντυπωσιακό ότι η τελική θέση των νευρώνων του SOM έχει προσεγγίσει σχεδόν τέλεια την εκθετική συνάρτηση.

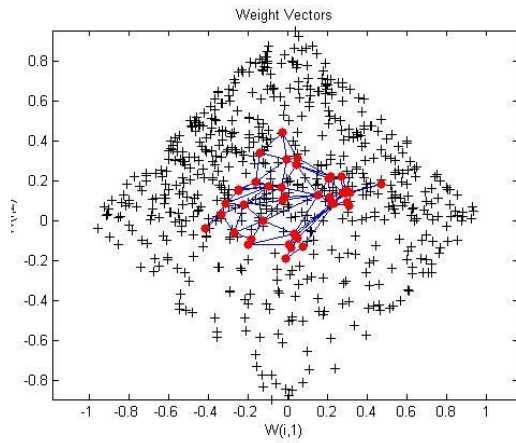


Σχήμα 3.2: Εκθετική Χωρική Κατανομή προσομοιωμένη από μονοδιάστατο SOM πλέγμα-
τος 10 νευρώνων. $\tau_2 = 1000, \eta_0 = 0.1$

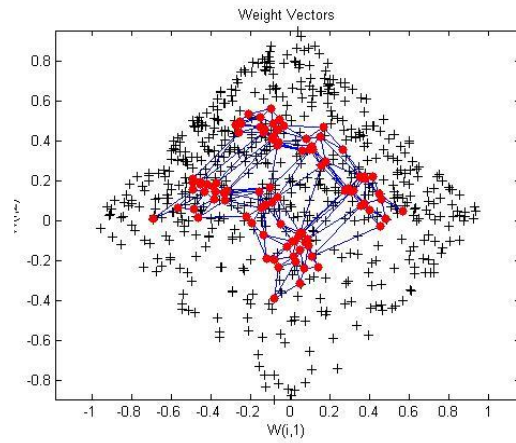
Στο σχήμα 3.3 μπορούμε να παρατηρήσουμε τον τρόπο που εξελίσσεται το δισδιάστατο SOM. Το σύνολο εκπαίδευσης σε αυτή την περίπτωση αποτελείται από σημεία που η χωρική τους τοπολογία προσομοιάζει έναν ρόμβο. Αρχικά, οι νευρώνες του SOM, βρίσκονται κάπου στο κέντρο (μεταξύ μέγιστης και ελάχιστης τιμής του συνόλου εκπαίδευσης) και καθώς οι εποχές εκπαίδευσης περνάνε αρχίζει να προσεγγίζουν την ζητούμενη τοπολογία.

Το γεγονός ότι τελικά το SOM δεν προσεγγίζει τέλεια την χωρική κατανομή, οφείλεται στο ότι τα δείγματα εκπαίδευσης δεν έχουν κατανομηθεί ομοιόμορφα στον χώρο. Για παράδειγμα, παρατηρούμε ότι στο πάνω μέρος υπάρχουν πιο πολλά δείγματα σε σχέση με το κάτω μέρος του κόμβου. Αυτό έχει σαν αποτέλεσμα το συγκεκριμένο πάνω μέρος να έχει μεγαλύτερη ελκτική δύναμη σε σχέση με το κάτω, οπότε να προσελκύει περισσότερους νευρώνες και εν τέλει να προσεγγίζεται

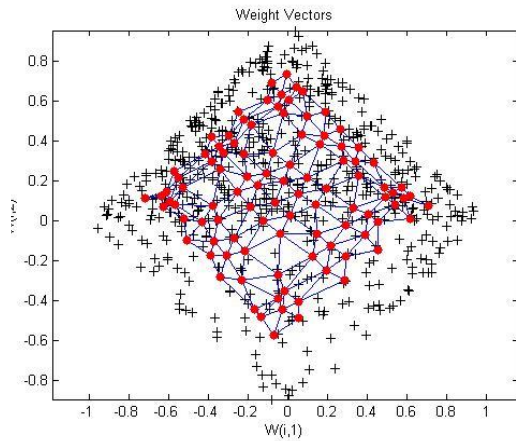
καλύτερα. Το παραπάνω συμπέρασμα είναι εμφανές στα σχήματα 3ε και 3στ καθώς το πλέγμα των νευρώνων είναι πιο αραιό στο κάτω μέρος σε σχέση με το πάνω.



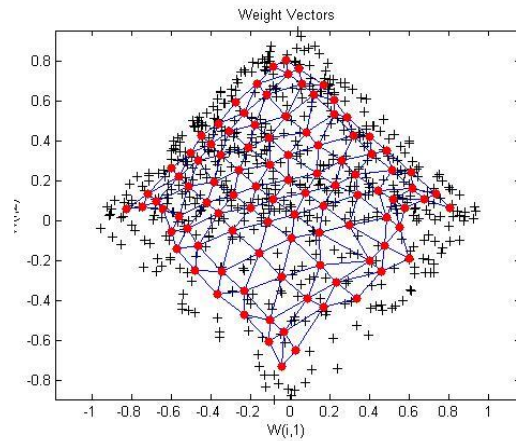
(α) k=200



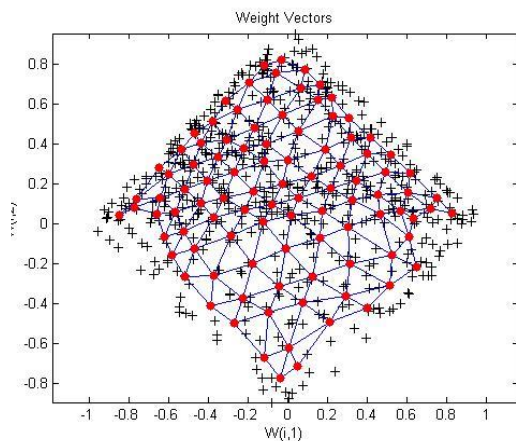
(β) k=400



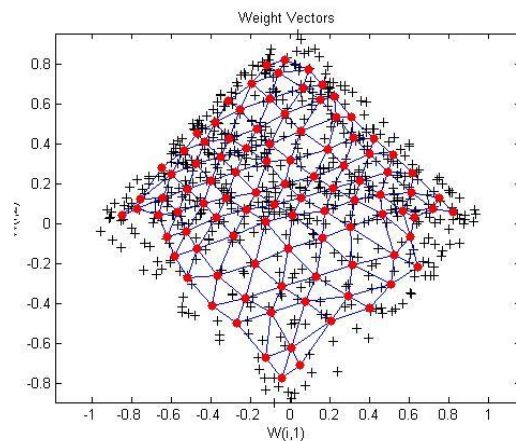
(γ) k=600



(δ) k=800



(ε) k=1000 (Τελική τοπολογία φάσης Διάταξης)



(στ) Τελική τοπολογία φάσης Σύγκλισης

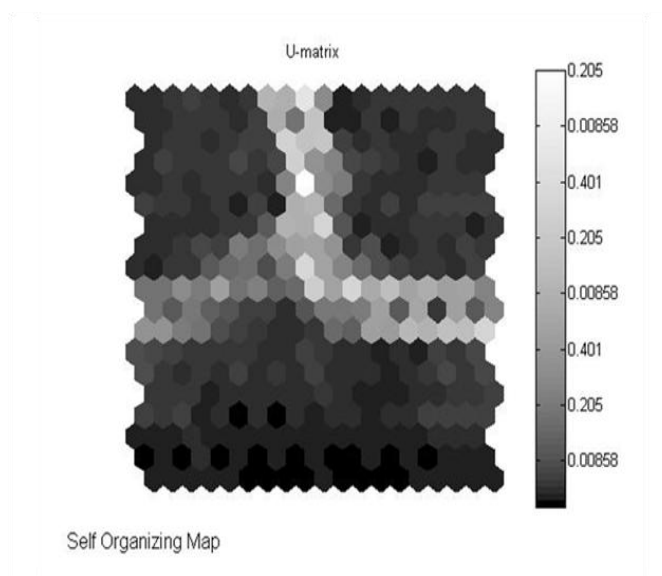
Σχήμα 3.3: Χωρική Κατανομή Ρόμβου προσομοιωμένη από δισδιάστατο SOM πλέγματος 10×10 .

3.5 Ομαδοποίηση Προτύπων Χρησιμοποιώντας SOM

Μία από τις βασικές λειτουργίες που καλείται να επιτελέσει το SOM (ως ένα κλασικό δίκτυο μη επιβλεπόμενης μάθησης) είναι η *ομαδοποίηση προτύπων (clustering)* ή αλλιώς η *μη επιβλεπόμενη ταξινόμηση (unsupervised classification)*. Στόχος του clustering είναι η αντιστοίχιση προτύπων προερχόμενων από σύνολα δεδομένων σε ομάδες (συστάδες) σύμφωνα με ορισμένα κριτήρια ομοιότητας. Ο διαχωρισμός των αρχικών συνόλων σε ομάδες γίνεται στην βάση της μεγιστοποίησης της ομοιότητας μεταξύ των προτύπων της ίδιας ομάδας και στην ελαχιστοποίησης της ομοιότητας μεταξύ προτύπων που ανήκουν σε διαφορετικές ομάδες.

Το SOM, εκτός των δυνατοτήτων που προσφέρει για ομαδοποίηση προτύπων, είναι ένα αποδοτικό εργαλείο για οπτικοποίηση δεδομένων μεγάλων διαστάσεων. Στη βασική του μορφή, ουσιαστικά μετατρέπει τις μη γραμμικές στατιστικές συσχετίσεις προτύπων μεγάλης διάστασης σε απλές γεωμετρικές συσχετίσεις σημείων ενός επιπέδου μικρής διάστασης (το οποίο συνήθως είναι ένα δισδιάστατο πλέγμα νευρώνων/κόμβων). Κατά συνέπεια, καθώς το SOM συμπιέζει την πληροφορία διατηρώντας στις απεικονίσεις τις σημαντικότερες τοπολογικές και μετρικές συσχετίσεις των δεδομένων, μπορεί να υποστηριχθεί ότι παράγει ένα είδος αφαίρεσης. Αυτά τα δύο χαρακτηριστικά, η οπτικοποίηση και η αφαίρεση, σε συνδυασμό με την ομαδοποίηση προτύπων, είναι δυνατό να χρησιμοποιηθούν σε μία σειρά πολύπλοκων και σύνθετων εργασιών.

Μία μεθοδολογία που ενοποιεί τις τρεις αυτές ιδιότητες/δυνατότητες του SOM είναι η γραφική απεικόνιση που καλείται *πίνακας ενοποιημένων αποστάσεων (unified distance matrix ή αλλιώς U-matrix)*. Βάσει αυτής της μεθοδολογίας υπολογίζονται οι αποστάσεις μεταξύ των διανυσμάτων παραμέτρων (βαρών) των γειτονικών νευρώνων και στην συνέχεια αναπαριστώνται ως αποχρώσεις επί ενός επιπέδου. Παράλληλα, η απόχρωση καθενός νευρώνα καθορίζεται από τη μέση τιμή των αποστάσεων του νευρώνα αυτού από όλους τους γειτονικούς σε αυτόν νευρώνες. (Ultsch & Siemon, 1990)



Σχήμα 3.4: Τυπικό παράδειγμα U-Matrix, για 6-διάστατα δεδομένα εκπαίδευσης που χωρίζονται σε δυο κλάσεις μεταξύ τους.

Από το U-Matrix μπορούν να εξαχθούν αρκετά χρήσιμα συμπεράσματα για τα πρότυπα εισόδου και ποιά είναι η χωρική κατανομή τους. Για παράδειγμα από τον U-Matrix του σχήματος 3.4 καταλήγουμε στα παρακάτω συμπεράσματα

- Το πλήθος των ομάδων (Clusters) που μπορούν να διαχωριστούν τα πρότυπα μας είναι δύο (η γραμμή που εμφανίζεται στο κέντρο προς τα αριστερά, δεν μπορεί να θεωρηθεί διαχωριστική καθώς οι γειτονικοί νευρώνες βρίσκονται πολύ κοντά, καθώς η απόχρωσή τους είναι πιο κοντά στο μαύρο παρά στο άσπρο)
- Το μέγεθος κάθε ομάδας μπορεί να φανεί από τον αριθμό των νευρώνων που περιέχει. Συνεπώς, εύκολα αντιλαμβανόμαστε ότι η ομάδα νευρώνων που βρίσκεται πάνω δεξιά στο σχήμα μας περιέχει πολύ μικρότερο αριθμό προτύπων σε σχέση με την άλλη.

Οι διαφοροποιήσεις στα σύνορα οφείλονται στο πόσο κοντά βρίσκονται οι ομάδες μεταξύ τους. Γι αυτό τον λόγο, όταν η απόσταση μεγιστοποιείται έχουμε άσπρο, ενώ όταν βρίσκονται πιο κοντά οι ομάδες έχουμε μαύρο.

3.6 Επιβλεπόμενο SOM (Supervised SOM)

Είναι γενικά αποδεκτό ότι το SOM υλοποιείται ακολουθώντας μία μη επιβλεπόμενη προσέγγιση. Ωστόσο, στην περίπτωση που παρέχεται η επιπρόσθετη πληροφορία των κατηγοριών στις οποίες ανήκουν τα πρότυπα ενός συνόλου δεδομένων, οι δυνατότητες του SOM δύναται να διευρυνθούν πράγμα που έχει άμεσο αντίκτυπο στην περαιτέρω βελτίωση των επιδόσεων του. Η εν λόγω προσέγγιση καλείται *επιβλεπόμενο SOM* καθότι κινείται στα σύνορα μεταξύ επιβλεπόμενης και μη επιβλεπόμενης μάθησης.

Η μεθοδολογία για να μετασχηματιστεί ο αλγόριθμος μη επιβλεπόμενης μάθησης του SOM σε επιβλεπόμενη είναι η εξής (Midenet & Grumbach, 1994):

Το διάνυσμα κάθε προτύπου εισόδου αποτελείται από δύο τμήματα \mathbf{x}_u και \mathbf{x}_s , όπου \mathbf{x}_u είναι το διάνυσμα χαρακτηριστικών του προτύπου και \mathbf{x}_s είναι η κατηγορία στην οποία ανήκει το πρότυπο (βάσει μίας ορθογώνιας κωδικοποίησης των κατηγοριών). Εν συνεχεία τα συνενωμένα διανύσματα προτύπων εισόδου $\mathbf{x} = [\mathbf{x}_u, \mathbf{x}_s]$ χρησιμοποιούνται στην διαδικασία εκπαίδευσης του SOM, γεγονός που αιτιολογεί τον όρο επιβλεπόμενη αφού η πληροφορία των κατηγοριών στις οποίες ανήκουν τα πρότυπα αξιοποιείται αναλόγως.

Αντίθετα, κατά την αναγνώριση - κατηγοριοποίηση ενός αγνώστου κατηγορίας προτύπου μόνο το \mathbf{x}_u τμήμα του διανύσματος αυτού, συγκρίνεται με το αντίστοιχο τμήμα των διανυσμάτων βαρών του SOM. Η δε κατηγορία του αγνώστου προτύπου καθορίζεται από το \mathbf{w}_s τμήμα του διανύσματος βαρών του νευρώνα νικητή.

ΚΕΦΑΛΑΙΟ 4

4 Καμπυλότητα σε Χώρο Κλίμακας (Curvature Scale Space) – Κατάτμηση Σχήματος

4.1 Εισαγωγή

Ένα από τα σημαντικότερα χαρακτηριστικά των επίπεδων σχημάτων είναι η καμπυλότητα. Η συνάρτηση της καμπυλότητας υπολογίζει σε κάθε σημείο κατά μήκος της καμπύλης το μέτρο της μεταβολής του προσανατολισμού της καμπύλης.

Υπάρχουν αρκετοί λόγοι για τους οποίους χρησιμοποιείται η καμπυλότητα ως βάση σε εφαρμογές όρασης υπολογιστών.

- *Βιολογική Όραση*: Από ψυχοφυσικά και ψυχολογικά τεστ που έχουν διεξαχθεί σε ανθρώπους, έχει παρατηρηθεί ότι η ανθρώπινη όραση χρησιμοποιεί τα ακρότατα της συνάρτησης της καμπυλότητας, προκειμένου να τέμνει το περίγραμμα ενός σχήματος, δημιουργώντας μικρότερα τμήματα του περιγράμματος τα οποία έχουν ουσιώδη νόημα για τον ανθρώπινο εγκέφαλο. Για παράδειγμα, ο Attneave (1954) βρήκε μέσω ψυχολογικών πειραμάτων ότι οι άνθρωποι επιλέγουν τα ακρότατα της συνάρτησης της καμπυλότητας προκειμένου να περιγράψουν ένα σχήμα. Ενώ οι Hoffman & Richards (1984) δίνουν στοιχεία από ψυχολογικά τεστ σύμφωνα με τα οποία, η ανθρώπινη όραση χρησιμοποιεί τα σημεία στα οποία έχουμε τοπικά ελάχιστα της συνάρτησης της καμπυλότητας προκειμένου να κατατάμνει μια καμπύλη.

- *Υπολογιστική Όραση*: Τοπικοί υπολογισμοί της καμπυλότητας και της εφαπτομένης είναι αρκετοί ώστε να επανακτήσουμε την τροχιά μιας καμπύλης στην εικόνα. Σημαντικά σημεία, όπως τα ακρότατα και τα σημεία εναλλαγής προσήμου, έχει αποδειχθεί ότι είναι πολύ χρήσιμα στο ταίριασμα και την αναγνώριση σχημάτων.

Στο συγκεκριμένο κεφάλαιο θα αναφερθούμε στις εφαρμογές που έχει η συνάρτηση της καμπυλότητας στα διάφορα προβλήματα υπολογιστικής όρασης. Ξεκινάμε παρουσιάζοντας τον ορισμό της καμπύλης και των σημαντικότερων τύπων της ανάλυσης της καμπυλότητας για επίπεδα σχήματα.

Ιδιαίτερη έμφαση θα δώσουμε στο κομμάτι της “καμπυλότητα σε χώρο κλίμακας” (Curvature Scale Space) όπου βρίσκουμε την καμπυλότητα του επίπεδου σχήματος σε διάφορα επίπεδα λεπτομέρειας. Αυτά τα επίπεδα αποκτώνται μέσω της συνέλιξης της καμπύλης του σχήματος με ένα Γκαουσιανό πυρήνα μεταβλητής διακύμανσης.

Τέλος θα αναφερθούμε στην κατάτμηση σχημάτων (shape decomposition). Η συγκεκριμένη θεωρία μας παρέχει τα εργαλεία ώστε να διαχωρίσουμε ένα αρχικά περίπλοκο επίπεδο σχήμα σε μικρότερα πρωταρχικά σχήματα όπως τετράγωνα, κύκλους, ορθογώνια.

4.2 Καμπύλη

Έστω μια συνεχής διανυσματική συνάρτηση $\vec{C}: [a, b] \rightarrow \mathbb{R}^m$, $m = 2$ ή $m = 3$ όπου $J = [a, b]$ είναι ένα κλειστό πραγματικό διάστημα. Καθώς η μεταβλητή p κινείται στο J , οι τιμές της συνάρτησης $\vec{C}(p)$ ανιχνεύουν το σχήμα της \vec{C} , το οποίο είναι το σύνολο

$$\Gamma = \{ \vec{C}(p) \in \mathbb{R}^m : p \in [a, b] \} \quad (4.1)$$

Αν η \vec{C} είναι συνεχής συνάρτηση, τότε το σχήμα που ανιχνεύεται ονομάζεται *καμπύλη*. Αν $m=2$ τότε θα έχουμε επίπεδη καμπύλη, ενώ αν $m=3$ τότε θα έχουμε μια καμπύλη στον χώρο.

Συνεπώς, από γεωμετρική σκοπιά, η καμπύλη είναι ένα σύνολο από σημεία όπως περιγράφεται στην εξίσωση (4.1) ενώ αναλυτικά, η καμπύλη μπορεί να περιγραφεί ως μια συνεχής απεικόνιση \vec{C} , η οποία ονομάζεται συνεχής τροχιά (Μαραγκός, 2005).

Σε ένα καρτεσιανό σύστημα συντεταγμένων, μια τροχιά αναπαριστάται από το διάνυσμα θέσης της καμπύλης $\vec{C}(p) = (x_1(p), \dots, x_m(p))$ όπου όλα τα $x_i : J \rightarrow \mathbb{R}$ είναι συνεχείς πραγματικές συναρτήσεις. Καθώς η παράμετρος p κινείται στο

διάστημα $[a, b]$, το διάνυσμα θέσης κινείται κατά μήκος της καμπύλης από το αρχικό της σημείο $\vec{C}(a)$ στο τελικό της σημείο $\vec{C}(b)$. Συνεπώς, η τροχιά \vec{C} αποτελεί μια παραμετροποίηση της καμπύλης.

Μια καμπύλη μπορεί να έχει πολλές διαφορετικές παραμετροποιήσεις. Συγκεκριμένα, κάθε συνεχή τροχιά $\vec{Y} : [c, d] \rightarrow \mathbb{R}^m$ ορισμένη ως $\vec{Y}(p) = \vec{C}(h(p))$, όπου η απεικόνιση $h : [c, d] \rightarrow [a, b]$ είναι ένας αυξανόμενος ομοιομορφισμός¹, ανιχνεύει την ίδια καμπύλη όπως η \vec{C} και ονομάζεται ισοδύναμη τροχιά. Η συνάρτηση h ονομάζεται μετασχηματισμός παραμέτρου και δεν επηρεάζει τις σημαντικές γεωμετρικές ιδιότητες της καμπύλης οι οποίες αναφέρθηκαν προηγουμένως.

Συνεπώς μπορούμε να υποθέσουμε ότι όλες οι τροχιές ορίζονται στο μοναδιαίο διάστημα $I = [0, 1]$ καθώς αν μια τροχιά ορίζεται στο διάστημα $[a, b]$ μπορούμε να ορίσουμε μια ισοδύναμη τροχιά χρησιμοποιώντας την συνάρτηση αλλαγής μεταβλητής $h : [0, 1] \rightarrow [a, b]$ όπου $h(p) = a + (b - a)p$.

Τυπικά, μια καμπύλη ορισμένη από μια συνεχή τροχιά \vec{C} είναι το σύνολο όλων των ισοδύναμων τροχιών $\vec{Y}(p) = \vec{C}(h(p))$, όπου h είναι μια αύξουσα, 1-1, συνεχής απεικόνιση του I επί του εαυτού του.

Εάν η τροχιά \vec{C} είναι συνεχώς διαφορίσιμη και οποιαδήποτε πιθανή αλλαγή παραμέτρου h είναι επίσης μια συνεχώς διαφορίσιμη συνάρτηση με θετική παράγωγο στο I , τότε η αντίστοιχη καμπύλη ονομάζεται *ομαλή καμπύλη*.

Μια τροχιά ονομάζεται κλειστή αν $\vec{C}(0) = \vec{C}(1)$. Επειδή στην περίπτωση αυτή το αρχικό σημείο δεν παίζει ρόλο καθώς η καμπύλη είναι κλειστή (δεν έχει αρχή και τέλος), ωστόσο η επιλογή του μπορεί να επηρεάσει το τελικό σύνολο σημείων που θα πάρουμε, χρησιμοποιούμε αντί για το κλειστό διάστημα I τον μοναδιαίο κύκλο S^1 .

Συνεπώς μια κλειστή καμπύλη ορισμένη από μια συνεχή τροχιά του μοναδιαίου κύκλου S^1 στο \mathbb{R}^m είναι το σύνολο όλων των απεικονίσεων $\vec{Y}(p) = \vec{C}(h(p))$, όπου το h είναι μια 1-1, με σταθερό προσανατολισμό, συνεχής απεικόνιση του S^1 επί του εαυτού του.

4.3 Ανάλυση Καμπυλότητας Επίπεδων Σχημάτων

Έστω μια κλειστή επίπεδη ομαλή καμπύλη παραμετροποιημένη ως μια τροχιά $\vec{C} : I \rightarrow \mathbb{R}^2$ τέτοια ώστε το εσωτερικό της να βρίσκεται συνεχώς στα αριστερά καθώς ανιχνεύεται η καμπύλη στην κατεύθυνση που αυξάνει το $p \in I$.

¹ Μια συνάρτηση $h : A \rightarrow B$ ονομάζεται ένας ομοιομορφισμός εάν και μόνο εάν είναι συνεχής, 1-1 και επί, ενώ η αντίστροφη της συνάρτηση h^{-1} είναι συνεχής.

Το διάνυσμα θέσης της καμπύλης δίνεται από τον τύπο

$$\vec{C}(p) = (x(p), y(p))$$

όπου $x(p)$ και $y(p)$ είναι δυο φορές συνεχώς διαφορίσιμες πραγματικές συναρτήσεις στο I .

Το μοναδιαίο εφαπτομενικό διάνυσμα σε κάθε σημείο μπορεί να εκφραστεί ως

$$\vec{T}(p) = \frac{\vec{C}'(p)}{\|\vec{C}'(p)\|} \quad (4.2)$$

όπου ο τόνος ' δηλώνει διαφόριση ως προς p , και

$$u(p) = \|\vec{C}'(p)\| = \sqrt{[x'(p)]^2 + [y'(p)]^2}$$

είναι η ταχύτητα με την οποία το διάνυσμα θέσης ανιχνεύει την καμπύλη.

Εκτός των υπολοίπων παραμετροποιήσεων μίας καμπύλης, υπάρχει μία με ιδιαίτερη σημασία, το μήκος καμπύλου τμήματος ή μήκος τόξου. Θεωρήστε μια καμπύλη η οποία αναπαριστάται από μια τροχιά $\vec{C}(p)$ ορισμένη στο I . Η συνάρτηση μήκους τόξου (ή καμπύλου τμήματος) της είναι η $s(p)$ και ορίζεται σε κάθε $p \in I$ ως το μήκος του τόξου μεταξύ του αρχικού σημείου $A = \vec{C}(0)$ και ενός τυχαίου σημείου $B = \vec{C}(p)$ στην καμπύλη. Γενικά, το μήκος τόξου ορίζεται ως το ανώτατο όριο (supremum) των μηκών όλων των εγγεγραμμένων πολυγωνικών καμπυλών. Το συνολικό μήκος, L , μιας καμπύλης είναι ίσο με $s(1)$.

Αν μια ομαλή καμπύλη $\vec{C}(p)$, έχει πεπερασμένο μήκος L , τότε η παράγωγος του μήκους τόξου υπάρχει και είναι ίση με την συνάρτηση ταχύτητας.

$$s'(p) = \frac{ds(p)}{dp} = u(p)$$

Επιπλέον μπορούμε να υπολογίσουμε το μήκος τόξου οποιουδήποτε καμπύλου τμήματος ολοκληρώνοντας την ταχύτητα

$$s(p) = \int_0^p u(q) dq$$

Οπότε όταν παραμετροποιούμε καμπύλες μέσω του μήκους τόξου, η αντίστοιχη συνάρτηση τροχιάς δεν ορίζεται στο μοναδιαίο διάστημα I αλλά στο διάστημα $[0, L]$ όπου L το μήκος της καμπύλης.

\

Εάν μια ομαλή καμπύλη παραμετροποιηθεί από το μήκος τόξου, τότε το διάνυσμα της ταχύτητας έχει μοναδιαίο μέτρο

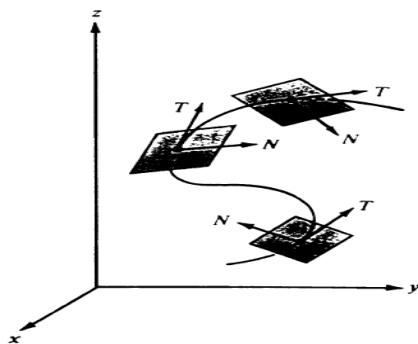
$$\frac{d\vec{C}}{ds} = \frac{d\vec{C}}{dp} \frac{dp}{ds} = \frac{1}{u(p)} \frac{d\vec{C}}{dp} \Rightarrow \left\| \frac{d\vec{C}}{ds} \right\| = \left\| \frac{d\vec{C}}{dp} \right\| \frac{1}{u(p)} = 1$$

Από τα παραπάνω προκύπτει ότι το $\frac{d\vec{C}}{ds}$ είναι το ίδιο με το εφαπτομενικό διάνυσμα \vec{T} σύμφωνα με την εξίσωση (4.2).

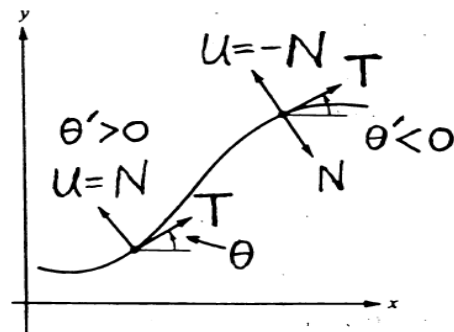
Η παράγωγος του μοναδιαίου εφαπτομενικού διανύσματος είναι κάθετη στο διάνυσμα και κανονικοποιώντας την προκύπτει το μοναδιαίο κάθετο διάνυσμα της καμπύλης

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

Το μοναδιαίο κάθετο διάνυσμα πάντα έχει κατεύθυνση προς το κοίλο μέρος της καμπύλης, όπως φαίνεται και στο παρακάτω σχήμα.



α. Εφαπτομενικό διάνυσμα T και το κάθετο διάνυσμα N που έχει κατεύθυνση προς το κοίλο μέρος της καμπύλης (Μαραγκός, 2005).



β. Γωνία θ που σχηματίζει το εφαπτομενικό διάνυσμα T με τον άξονα των x (Μαραγκός, 2005)

Σε κάθε σημείο της καμπύλης, η γωνία που σχηματίζεται από το εφαπτομενικό διάνυσμα \vec{T} και τον άξονα x ισούται με

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

Η απόλυτη τιμή του ρυθμού μεταβολής της γωνίας, $|\theta'(p)|$ ισούται με την ταχύτητα $\|\vec{T}'(p)\|$.

Αν ο ρυθμός μεταβολή της γωνίας βρεθεί συναρτήσει του μήκους τόξου της καμπύλης, τότε προκύπτει η συνάρτηση της **καμυλότητας**

$$\kappa(p) \triangleq \frac{d\theta}{ds} = \frac{y''x' - x''y'}{[(x')^2 + (y')^2]^{\frac{3}{2}}} \quad (4.3)$$

Εάν γνωρίζουμε το μοναδιαίο εφαπτομενικό διάνυσμα $\vec{T}(s)$ για όλα τα s τότε μπορούμε να προσδιορίσουμε την καμπύλη με κάποια μετατόπιση αφού

$$\vec{C}(s) = \vec{C}(0) + \int_0^s \vec{T}(q) dq$$

Επιπλέον, αν γνωρίζουμε την συνάρτηση της καμυλότητας τότε πάλι μπορούμε να προσδιορίσουμε μονοσήμαντα την καμπύλη με κάποια ευκλείδεια μετατόπιση σύμφωνα και με το παρακάτω θεώρημα.

Θεώρημα. Δυο επίπεδες καμπύλες, που είναι δυο φορές συνεχώς διαφορίσιμες, έχουν την ίδια συνάρτηση καμυλότητας ως προς το μήκος τόξου εάν και μόνο εάν η μια είναι μετασχηματισμός της άλλης ως προς μια ευκλείδεια κίνηση

Συνεπώς η καμυλότητα μιας επίπεδης καμπύλης μας δίνει όλες τις πληροφορίες για το σχήμα της καμπύλης και καμία πληροφορία για την θέση και τον προσανατολισμό της.

4.4 Καμυλότητα σε Χώρο Κλίμακας (Curvature Scale Space)

Έστω μια ομαλή επίπεδη καμπύλη $\vec{C}(p) = (x(p), y(p))$, όπου $p \in I$. Σύμφωνα με την εξίσωση (4.3), ο τύπος για να υπολογίσουμε την συνάρτηση καμυλότητας είναι

$$\kappa(p) = \frac{y''x' - x''y'}{[(x')^2 + (y')^2]^{\frac{3}{2}}}$$

Προκειμένου να υπολογίσουμε την καμυλότητα σε διάφορες κλίμακες λεπτομέρειας, οι παραμετρικές συναρτήσεις $x(p)$ και $y(p)$, όπου $x : I \rightarrow \mathbb{R}$ και $y : I \rightarrow \mathbb{R}$, υπόκεινται σε συνέλιξη με ένα μονοδιάστατο Γκαουσιανό πυρήνα $g(p, \sigma)$ διακόμησης σ . (Marr & Hildreth, 1980)

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

Συνεπώς, η συνάρτηση $X(p, \sigma)$ που προκύπτει από την συνέλιξη του $x(p)$ με τον Γκαουσιανό πυρήνα, ορίζεται ως

$$X(p, \sigma) = x(p) \star g(p, \sigma) = \int_{-\infty}^{\infty} x(u) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(p-u)^2}{2\sigma^2}} du$$

Ομοίως ορίζεται και η συνάρτηση $Y(p, \sigma)$ ως η συνέλιξη του $y(p)$ με τον Γκαουσιανό πυρήνα

$$Y(p, \sigma) = y(p) \star g(p, \sigma) = \int_{-\infty}^{\infty} y(u) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(p-u)^2}{2\sigma^2}} du$$

Οπότε η παραμετρική εξίσωση της καμπύλης που προκύπτει μετά την συνέλιξη έχει την μορφή

$$\vec{C}(p, \sigma) = (X(p, \sigma), Y(p, \sigma))$$

και η συνάρτηση της καμπυλότητας της δίνεται από τον τύπο

$$\kappa(p, \sigma) = \frac{Y''(p, \sigma) \cdot X'(p, \sigma) - X'(p, \sigma) \cdot Y''(p, \sigma)}{[(X(p, \sigma))^2 + (Y(p, \sigma))^2]^{\frac{3}{2}}} \quad (4.4)$$

Προκειμένου να υπολογίσουμε τις παραγώγους των συναρτήσεων $X(p, \sigma)$, $Y(p, \sigma)$ χρησιμοποιούμε τις ιδιότητες της συνέλιξης για την παραγωγή, άρα έχουμε (Mokhtarian & Mackworth, 1984)

$$X'(p, \sigma) = \frac{\partial X(p, \sigma)}{\partial p} = \frac{\partial [x(p) \star g(p, \sigma)]}{\partial p} = x(p) \star \left[\frac{\partial g(p, \sigma)}{\partial p} \right]$$

$$X''(p, \sigma) = \frac{\partial^2 X(p, \sigma)}{\partial p^2} = x(p) \star \left[\frac{\partial^2 g(p, \sigma)}{\partial p^2} \right]$$

Ομοίως

$$Y'(p, \sigma) = \frac{\partial Y(p, \sigma)}{\partial p} = y(p) * \left[\frac{\partial g(p, \sigma)}{\partial p} \right]$$

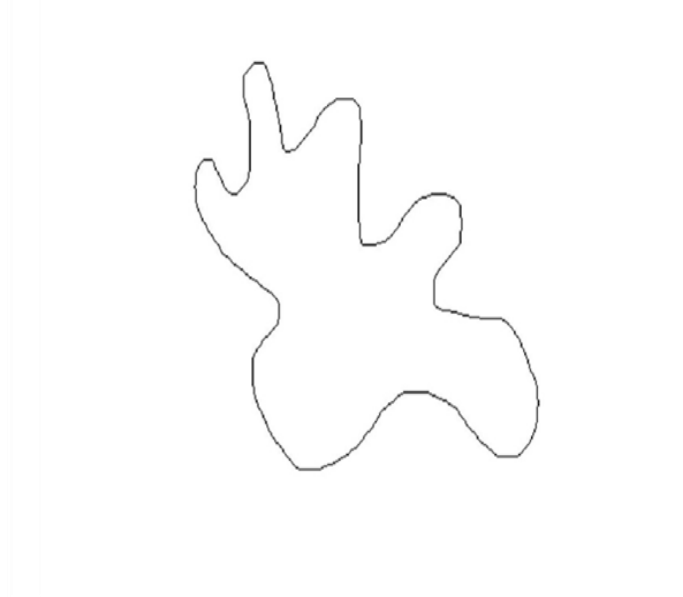
$$Y''(p, \sigma) = \frac{\partial^2 Y(p, \sigma)}{\partial p^2} = y(p) * \left[\frac{\partial^2 g(p, \sigma)}{\partial p^2} \right]$$

Συνεπώς εφαρμόζοντας την εξίσωση (4.4) επαναληπτικά αυξάνοντας κάθε φορά το σ , αποκτούμε την συνάρτηση καμπυλότητας για τα διάφορα επίπεδα ομαλότητας της καμπύλης. Η διαδικασία αυτή ονομάζεται *εξέλιξη* της καμπύλης.

Είναι επιθυμητό να βρούμε για κάθε ένα από αυτά τα επίπεδα τα σημεία εκείνα στα οποία έχουμε αλλαγή προσήμου (zero-crossing) της συνάρτησης της καμπυλότητας (Abbasi et al. 1999).

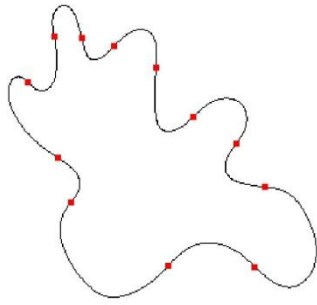
Άρα καθώς η καμπύλη εξελίσσεται μπορούμε να αναπαραστήσουμε τα συγκεκριμένα σημεία στο επίπεδο (p, σ) όπου p η παράμετρος ανίχνευσης της καμπύλης και σ η κλίμακα της. Το διάγραμμα που προκύπτει ονομάζεται «Αποτύπωμα της Καμπυλότητας στο Χώρο Κλίμακας» (*Curvature scale-space fingerprint*).

Στα σχήματα 4.1 και 4.2 που ακολουθούν, δημιουργήσαμε μια τυχαίου σχήματος κλειστή καμπύλη και παρουσιάζουμε την εξέλιξη της καθώς αυξάνεται η διακύμανση σ . Για κάθε επίπεδο ομαλότητας, εντοπίζουμε τα σημεία εναλλαγής προσήμου και τα απεικονίζουμε πάνω στην καμπύλη. Τέλος, φαίνεται το «Αποτύπωμα της Καμπυλότητας στο Χώρο Κλίμακας», όπως αυτό προκύπτει από την απεικόνιση των σημείων εναλλαγής προσήμου στο επίπεδο (p, σ) .

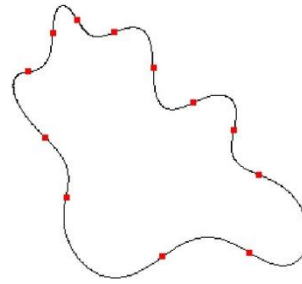


Σχήμα 4.1: Αρχική, τυχαία, επίπεδη συνεχής κλειστή καμπύλη

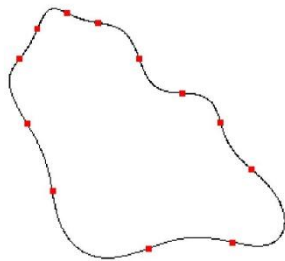
ZeroCrossing Points for $\sigma = 1.49$



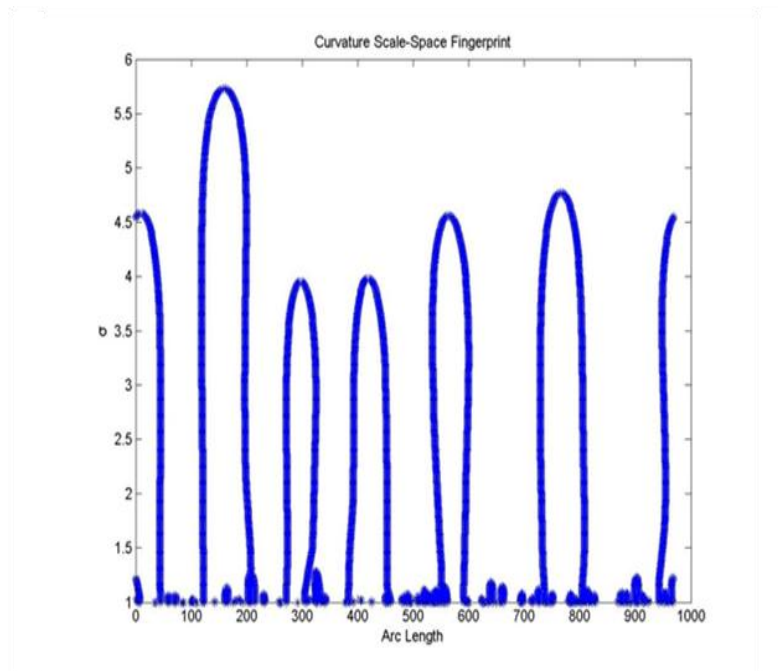
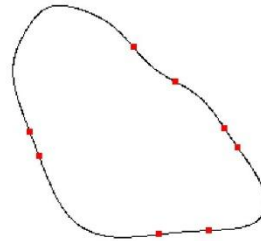
ZeroCrossing Points for $\sigma = 2.49$



ZeroCrossing Points for $\sigma = 3.49$



ZeroCrossing Points for $\sigma = 4.49$



Σχήμα 4.2: Αποτύπωση των σημείων εναλλαγής πρόσημου πάνω στις ομαλοποιημένες καμπύλες και στο διάγραμμα του χώρου (ρ, σ)

Από τα παραπάνω σχήματα παρατηρούμε ότι καθώς το σ αυξάνεται, η καμπύλη γίνεται πιο ομαλή και ο αριθμός των σημείων, στα οποία έχουμε εναλλαγή προσήμου της συνάρτησης της καμπυλότητας, μειώνεται.

Τελικά, όταν το σ θα γίνει αρκετά μεγάλο τότε η καμπύλη μας εκφυλίζεται σε κυρτή ενώ δεν υπάρχουν σημεία στα οποία η συνάρτηση της καμπυλότητας να αλλάζει πρόσημο.

Τα παραπάνω συμπεράσματα είναι εμφανή και στο διάγραμμα που προκύπτει αν αποτυπώσουμε τα σημεία εναλλαγής προσήμου στο επίπεδο (p, σ) . Το διάγραμμα αυτό, ονομάζεται και *CSS εικόνα* της καμπύλης. Αν φέρουμε οριζόντια γραμμή σε κάποιο επίπεδο εξέλιξης της καμπύλης μας, που δίνεται από την παράμετρο σ , τότε η τομή της γραμμής αυτής με το περίγραμμα του διαγράμματος μας δίνει τον αριθμό και την θέση των σημείων εναλλαγής προσήμου στην εκάστοτε ομαλοποιημένη καμπύλη. Για παράδειγμα, μια οριζόντια γραμμή στο $\sigma = 3.5$ μας δείχνει ότι υπάρχουν 12 σημεία εναλλαγής προσήμου.

Όπως επισημαίνουν οι Abbasi et al. (1999), υπάρχουν δύο σημεία εναλλαγής προσήμου καμπυλότητας για κάθε κοίλο ή κυρτό κομμάτι του αρχικού σχήματος μας και καθώς η καμπύλη μας εξελίσσεται (ομαλοποιείται), τα δυο αυτά σημεία πλησιάζουν το ένα το άλλο και δημιουργούν ένα περίγραμμα στην εικόνα CSS του σχήματος. Όταν το αρχικά κοίλο ή κυρτό μέρος της καμπύλης μας, ομαλοποιηθεί, καθώς η καμπύλη εξελίσσεται, τότε τα δύο αυτά σημεία ενώνονται και σχηματίζουν το πλήρες περίγραμμα, όπως φαίνεται στο σχήμα 4.2. Το ύψος αυτού του περιγράμματος, δηλώνει το βάθος και το μέγεθος της αντίστοιχης κυρτότητας ή κοιλότητας που εμφανίζεται σε εκείνο το κομμάτι του σχήματος. Όσο πιο βαθύ και μεγάλο είναι το τμήμα αυτό, τόσο υψηλότερο γίνεται και το περίγραμμα στην εικόνα CSS. Με άλλα λόγια, κάθε τοπικό μέγιστο του περιγράμματος αντιπροσωπεύει ένα κομμάτι του σχήματος.

Η παραπάνω μέθοδος έχει το πλεονέκτημα ότι είναι συμπαγής, καθώς μπορεί να περιγράψει ικανοποιητικά ένα σχήμα μόνο με ένα μικρό αριθμό από τοπικά μέγιστα και την θέση αυτών στην CSS εικόνα.

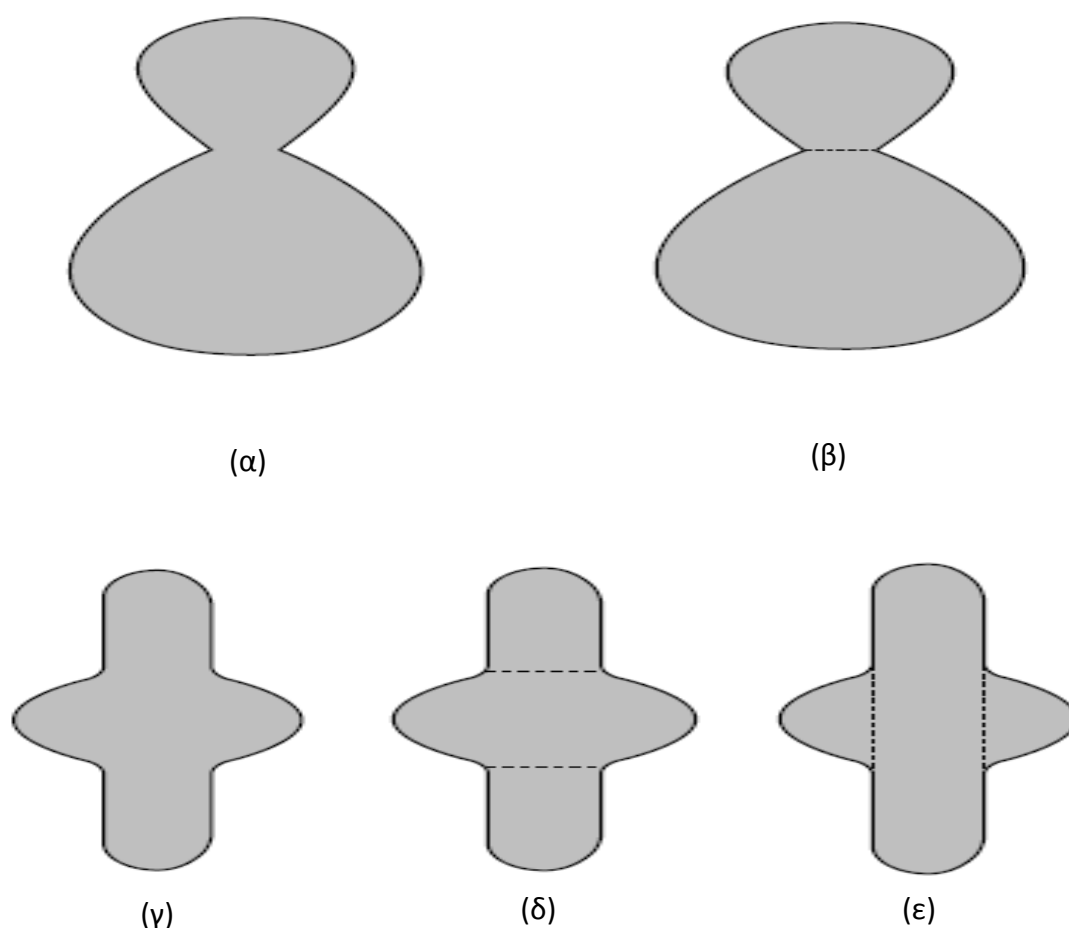
Επίσης, το ύψος των κορυφών των περιγραμμάτων της CSS εικόνας δεν επηρεάζεται από την μετατόπιση ή την αλλαγή του προσανατολισμού του σχήματος μας. Συγκεκριμένα, η μετατόπιση δεν επιφέρει καμιά αλλαγή ενώ η αλλαγή του προσανατολισμού απλώς μεταθέτει δεξιά ή αριστερά τα περιγράμματα, χωρίς να μεταβάλλει το σχήμα τους (Mokhtarian & Mackworth, 1984).

Τέλος αξίζει να επισημάνουμε ότι η μέθοδος μας δεν επηρεάζεται καθόλου από τον θόρυβο, που εκδηλώνεται συνήθως ως μικρός κυματισμός στο σχήμα της καμπύλης, καθώς με την ομαλοποίηση που εφαρμόζουμε στην καμπύλη μας, αυτός εξαφανίζεται πολύ γρήγορα. Συνεπώς, η εικόνα CSS μπορεί να μας παρέχει και ένα μέτρο για το πόσο πρέπει να ομαλοποιήσουμε την καμπύλη μας πριν προβούμε σε περαιτέρω επεξεργασία.

4.5 Κατάτμηση Σχήματος

Η ευκολία με την οποία η ανθρώπινη όραση αναγνωρίζει σχήματα έχει μελετηθεί εκτενώς χωρίς ακόμα να έχει βρεθεί ο ακριβής μηχανισμός που χρησιμοποιείται. Μια υπόθεση σύμφωνα με τους Hoffman & Richards (1984) είναι ότι η ανθρώπινη όραση χρησιμοποιεί κομμάτια συνδυασμένα σε δομικές αναπαραστάσεις. Δηλαδή, αναπαριστά ένα σχήμα με όρους όπως (i) τα σχήματα των συστατικών του μερών και (ii) την χωρική σχέση που έχουν μεταξύ τους τα μέρη αυτά.

Σύμφωνα λοιπόν με αυτή την προσέγγιση, η αναπαράσταση ενός σχήματος γίνεται μέσω της κατάτμησής του σε υποσχήματα και περιγράφοντας την σχέση μεταξύ αυτών των υποσχημάτων. Για παράδειγμα στο Σχήμα 4.3 (α)-(β) η προφανής κατάτμηση οδηγεί στον διαχωρισμό του σε δύο υποσχήματα, ενώ στο Σχήμα 2 (γ)-(ε) οι προφανείς κατατμήσεις του μπορεί να είναι δύο και να οδηγήσουν σε ένα διαφορετικό σύνολο υποσχημάτων.

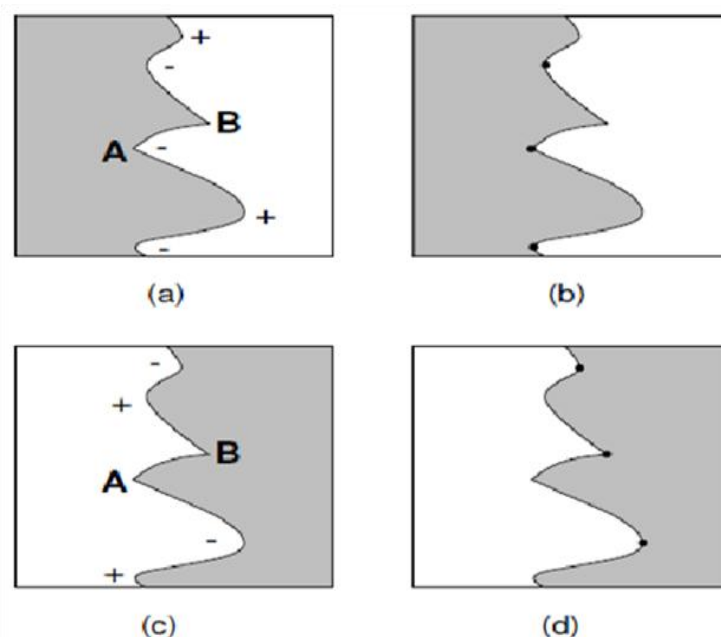


Σχήμα 4.3: Προφανείς τρόποι κατάτμησης φιγούρων. Ο (β) είναι ο μοναδικός τρόπος κατάτμησης του (α) ενώ υπάρχουν δυο διαφορετικοί προφανείς τρόποι (δ), (ε) για την κατάτμηση του (γ) (Seyranian & Hoffman, 1999).

Από τα παραπάνω εγείρεται το ερώτημα, πως η ανθρώπινη όραση κατατέμνει ένα σχήμα σε υποσχήματα. Κάποιοι επιστήμονες, θεωρούν ότι υπάρχει ένα σύνολο από βασικά σχήματα, πρωτότυπα, τα οποία η ανθρώπινη όραση αναζητά σε κάθε σχήμα (Binford,1971). Τέτοια σχήματα στο επίπεδο μπορεί να είναι τετράπλευρα, κύκλοι, ελλείψεις, κ.α. Σύμφωνα με αυτή την θεωρία, λοιπόν, η ανθρώπινη όραση διαχωρίζει ένα σχήμα βρίσκοντας αυτά τα πρωτότυπα στο σχήμα.

Ωστόσο, υπάρχει και η θεωρία ότι η ανθρώπινη όραση χρησιμοποιεί γεωμετρικούς κανόνες με τους οποίους βρίσκει τα όρια μεταξύ των υποσχημάτων ενός σχήματος. Τέτοιου είδους γεωμετρικό κανόνα αποτελεί ο «κανόνας ελαχίστου» που πρότειναν οι Hoffman & Richards (1984). Ο εν λόγω κανόνας θα αναλυθεί περαιτέρω, καθότι είναι ο κανόνας που χρησιμοποιούμε ώστε να κατατμήσουμε τα σχήματα μας.

Κανόνας Ελαχίστου για επίπεδα σχήματα : Χώρισε κάθε σχήμα σε υποσχήματα χρησιμοποιώντας κοίλες ακμές και τα αρνητικά τοπικά ελάχιστα της καμπυλότητας της τροχιάς του συνόρου του σχήματος.

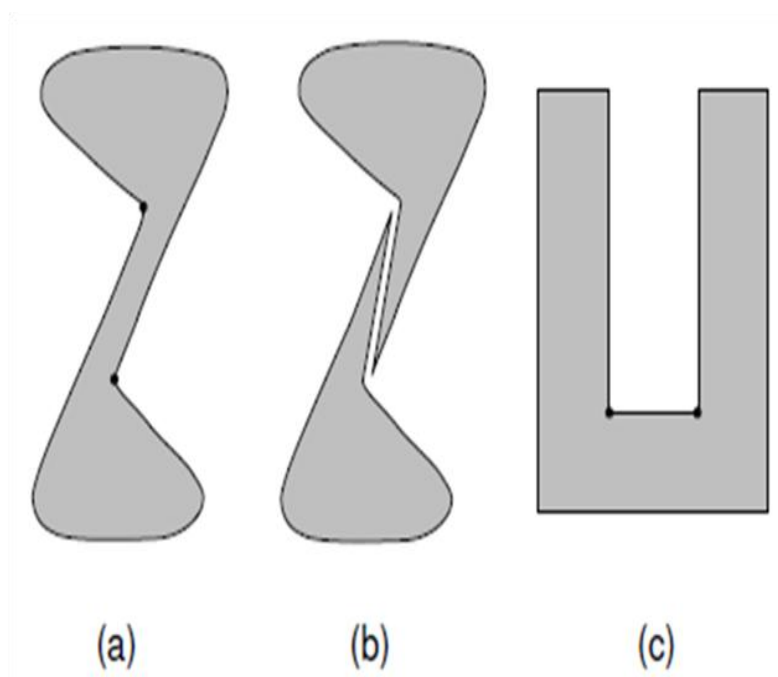


Σχήμα 4.4: Γεωμετρία σχετιζόμενη με τον "Κανόνα Ελαχίστου". Στο (a) παρατηρούμε αυτή την κυματοειδή φιγούρα. Οι περιοχές με θετική και αρνητική καμπυλότητα, φαίνονται με το πρόσημο (+) και (-) αντίστοιχα. Η ακμή με το διακριτικό A είναι κοίλη καθώς έχει κατεύθυνση προς το εσωτερικό του σχήματος ενώ η ακμή με το διακριτικό B είναι κυρτή καθώς έχει κατεύθυνση προς το εξωτερικό του σχήματος. Στο (b) παρουσιάζονται τα σημεία όπου βρίσκονται τα αρνητικά τοπικά ελάχιστα της καμπυλότητας. Αυτά τα σημεία είναι σύμφωνα με τον "Κανόνα Ελαχίστου" τα οριακά σημεία των υποσχημάτων. Στο (c) και (d) φαίνεται το σχήμα ανεστραμμένο και πως εντοπίζονται πάλι τα οριακά σημεία. Πρέπει να τονίσουμε οτι στο (c) η ακμή με το διακριτικό A είναι κυρτή καθώς δείχνει προς το εξωτερικό του σχήματος ενώ η ακμή με το διακριτικό B είναι κοίλη (Seyranian & Hoffman, 1999).

Σύμφωνα με τη ανάλυση που έγινε στο σχήμα 4.4, για το πώς μπορούμε να βρούμε τα οριακά σημεία ο κανόνας ελαχίστου μπορεί να αναδιατυπωθεί ως εξής

Κανόνας Ελαχίστου για επίπεδα σχήματα : Χώρισε κάθε σχήμα σε υποσχήματα χρησιμοποιώντας τα αρνητικά τοπικά ελάχιστα της καμπυλότητας της τροχιάς του συνόρου του σχήματος. Η τροχιά αυτή πρέπει να έχει προκύψει κινούμενοι με τέτοια κατεύθυνση πάνω στο σύνορο του σχήματος, ώστε να έχουμε πάντα το εσωτερικό του στα αριστερά.

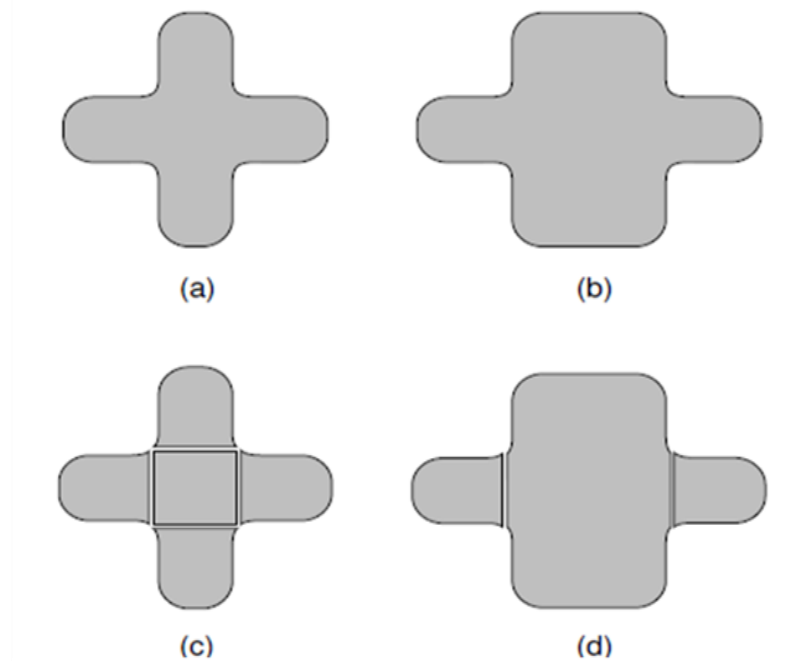
Ο κανόνας ελαχίστου μπορεί να μας παρέχει την ακριβή θέση των οριακών ή κρίσιμων σημείων ωστόσο δεν μας εξηγεί τον τρόπο με τον οποίο αυτά πρέπει να ενωθούν ώστε να σχηματίσουν τα υποσχήματα. Για την ακρίβεια δεν ενώνουμε διαδοχικά οριακά σημεία ώστε να προκύψουν νέα σχήματα, καθώς τότε μπορεί να έχουμε περίεργα αποτελέσματα όπως φαίνεται στο σχήμα 4.4.



Σχήμα 4.5: Άμα κόψουμε το σχήμα μας ενώνοντας διαδοχικά οριακά σημεία μπορεί να προκύψουν πολύ περίεργα αποτελέσματα όπως στο (a),(b), ενώ σε κάποιες περιπτώσεις όπως η (c) μπορεί να μην έχουμε καν καινούρια υπό-σχήματα (Seyranian & Hoffman, 1999).

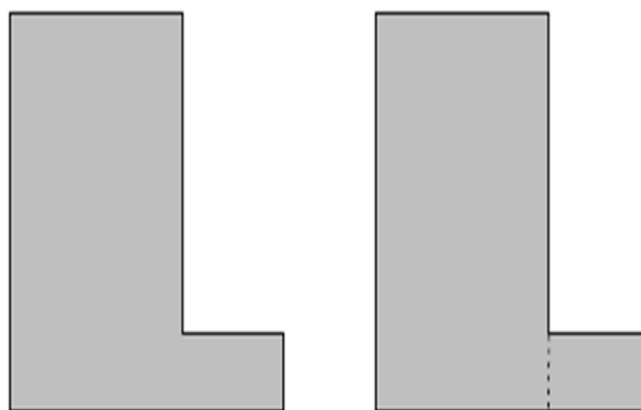
Ωστόσο υπάρχουν και άλλοι περιορισμοί για το πώς θα διαχωρίσουμε ένα σχήμα χρησιμοποιώντας μόνο τα οριακά σημεία που μας παρέχονται από τον κανόνα ελαχίστου. Οι περιορισμοί αυτοί είναι εμφανείς στα σχήματα 4.6, 4.7 που παρουσιάζονται παρακάτω.

Συγκεκριμένα στο σχήμα 4.6, έχουμε δυο σταυρούς όπου και στις δυο περιπτώσεις, ο κανόνας ελαχίστου θα μας δώσει τέσσερα οριακά σημεία. Ωστόσο, ο προφανής τρόπος, με τον οποίο κάποιος άνθρωπος θα διαχώριζε σε υποσχήματα τους δυο σταυρούς, είναι διαφορετικός για κάθε περίπτωση.



Σχήμα 4.6: Διαφορετικός τρόπος κατάτμησης για τον σταυρό (a), (b) αν και έχουν ίδιο αριθμό οριακών σημείων σε περίπου παρόμοιες θέσεις (Seyranian & Hoffman, 1999).

Επιπλέον μειονέκτημα της μεθόδου αποτελεί το γεγονός ότι δεν υποδεικνύει πιο σημείο επιπρόσθετα με το αρνητικό τοπικό ελάχιστο της καμπυλότητας είναι καλό για διαχωρισμό του σχήματος μας, καθώς τέτοια σημεία μπορεί να έχουν και μηδενική τιμή καμπυλότητας, όπως φαίνεται στο σχήμα 4.7.



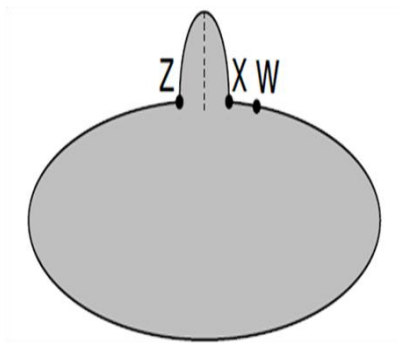
Σχήμα 4.7: Ο κανόνας ελαχίστου μας παρέχει μόνο ένα οριακό σημείο, ενώ το δεύτερο σημείο που πρέπει να ενωθεί έχει μηδενική καμπυλότητα (Seyranian & Hoffman, 1999).

Συνεπώς από τα παραπάνω εύκολα γίνεται κατανοητό ότι δεν υπάρχει συγκεκριμένη μεθοδολογία με την οποία να μπορούμε να πούμε επ' ακριβώς πώς να συνδέσουμε τα οριακά σημεία που μας δίνει η μέθοδος ελαχίστου.

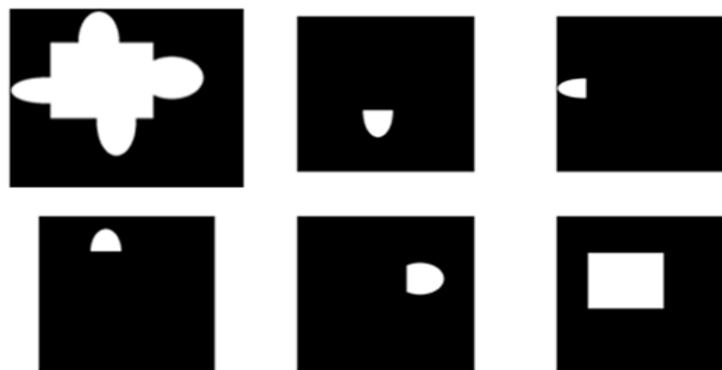
Όπως προτείνεται από τους Seyranian & Hoffman (1999), κατά πόσο ένα κόψιμο θεωρείται καλό έχει σίγουρα να κάνει με την απόσταση των δύο οριακών σημείων που ενώνουμε. Ωστόσο, όπως είδαμε παραπάνω αυτό ως μοναδικό κριτήριο δεν επαρκεί, οπότε σύμφωνα με τους Seyranian & Hoffman (1999) πρέπει το κόψιμο μας να διασχίζει και ένα τοπικό άξονα συμμετρίας.

Με άλλα λόγια, από την στιγμή που ένα κόψιμο διασχίζει ένα άξονα συμμετρίας τότε το συντομότερο κόψιμο είναι και το καλύτερο. Αν δεν διασχίζει ένα άξονα συμμετρίας, τότε απλά το συντομότερο κόψιμο δεν είναι το καλύτερο.

Αυτή η διαπίστωση στηρίζεται στο γεγονός ότι αν ένα κόψιμο αποτυγχάνει να διασχίσει ένα άξονα τοπικής συμμετρίας τότε η περιοχή που αποκόπτεται δεν είναι ικανοποιητικά μεγάλη ώστε να θεωρηθεί παρακλάδι του αρχικού μας σχήματος. Στο σχήμα 4.8, παρουσιάζεται η προαναφερθείσα μέθοδος κατάτμησης. Ενώ στο σχήμα 4.9 παρουσιάζονται και τα υποσχήματα που εξάγονται από την χρήση της παραπάνω μεθόδου.



Σχήμα 4.8: Το συντομότερο κόψιμο δεν είναι πάντα το καλύτερο ($x w$), πρέπει επίσης να διασχίζει και ένα άξονα τοπικής συμμετρίας όπως το κόψιμο ($x z$) (Seyranian & Hoffman, 1999).



Σχήμα 4.9: Αρχική εικόνα και τα υποσχήματα στα οποία διαχωρίζεται εφαρμόζοντας την μεθοδολογία που αναλύθηκε προηγουμένως.

ΚΕΦΑΛΑΙΟ 5

5 Στερεοπτική Όραση και Βαθμονόμηση Κάμερας

5.1 Εισαγωγή

Η όραση επιτρέπει σ' ένα ρομποτικό σύστημα να αποκτά γεωμετρικές και ποσοτικές πληροφορίες για τα αντικείμενα που βρίσκονται στο περιβάλλον του. Οι πληροφορίες αυτές μπορούν να χρησιμοποιηθούν από το σύστημα ελέγχου ώστε να εκτελεστεί μια εργασία από τον ρομποτικό βραχίονα ή ως ανατροφοδότηση στο σύστημα μας για τον έλεγχο κάποιας εργασίας.

Ως παράδειγμα θεωρείστε ένα ρομποτικό βραχίονα στον οποίο έχουμε προσαρμόσει μια κάμερα και του αναθέτουμε την εργασία να αρπάξει ένα αντικείμενο χρησιμοποιώντας μια αρπάγη. Μέσω της κάμερας το ρομπότ αποκτά πληροφορίες για το σχήμα, την θέση και τον προσανατολισμό του αντικειμένου. Έτσι το σύστημα ελέγχου μπορεί να σχεδιάσει μια τροχιά που να οδηγεί την αρπάγη στην σωστή θέση και προσανατολισμό ώστε να εκτελεστεί η εργασία μας.

Το βασικό αισθητήριο όργανο για την υπολογιστική όραση είναι η *κάμερα*. Η λειτουργία μιας κάμερας, μοντελοποιείται συνήθως ως μια αντιστοίχιση ανάμεσα στον τρισδιάστατο κόσμο (χώρο αντικειμένων) και στη δισδιάστατη εικόνα. Η ανάλυση των περισσοτέρων μοντέλων κάμερας γίνεται με την βοήθεια εργαλείων προβολικής γεωμετρίας. Έτσι, όπως θα δούμε και στη συνέχεια, γεωμετρικές οντότητες της κάμερας όπως το κέντρο προβολής και το επίπεδο της εικόνας, μπορούν να υπολογιστούν πολύ εύκολα από την αλγεβρική τους αναπαράσταση σε

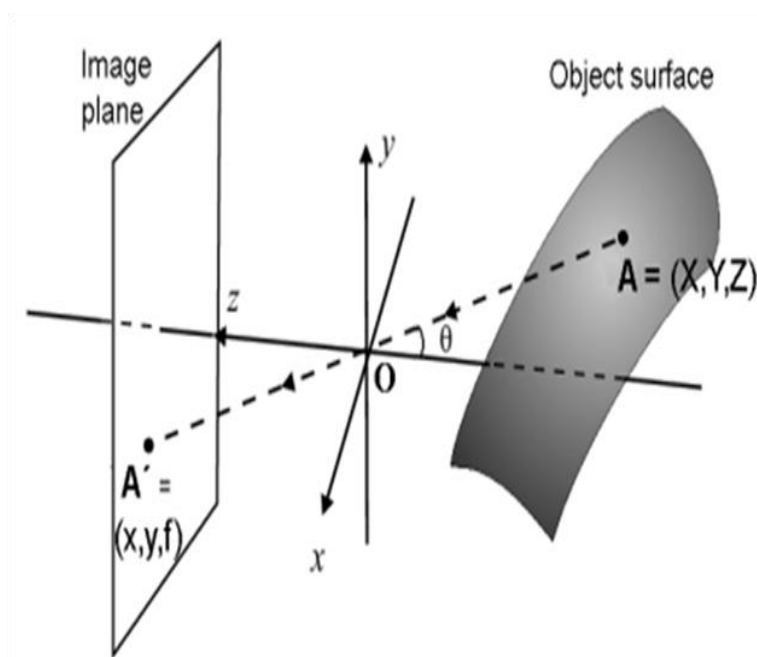
πίνακες. Στο παρών κεφάλαιο θα ασχοληθούμε με το απλούστερο μοντέλο κάμερας, το μοντέλο οπής (pinhole model).

Αφού, λοιπόν αναλύσουμε τα βασικά μαθηματικά και αλγεβρικά μοντέλα που διέπουν την κάμερα οπής, θα δώσουμε έμφαση στις έννοιες της στερεοπτικής όρασης και πως αυτή μέσω της επιπολικής γεωμετρίας και της μεθόδου της τριγωνοποίησης, μας επιτρέπει να εξάγουμε χρήσιμη πληροφορία για τις συντεταγμένες του αντικειμένου μας στο χώρο. Ως δεδομένα για να βρούμε τις συντεταγμένες του αντικειμένου μας στο χώρο, έχουμε μόνο δυο λήψεις του από δυο κάμερες που βρίσκονται σε διαφορετική θέση η μια ως προς την άλλη.

Τέλος θα αναφερθούμε και στο πρόβλημα της βαθμονόμησης κάμερας. Η βαθμονόμηση της κάμερας αποτελεί απαραίτητο βήμα στη 3Δ υπολογιστική όραση προκειμένου να εξαχθεί μετρική πληροφορία από 2Δ εικόνες. Μέσω της βαθμονόμησης, υπολογίζουμε τόσο τις εσωτερικές όσο και τις εξωτερικές παραμέτρους της κάμερας μας. Ο βασικός αλγόριθμος στον οποίο θα δοθεί έμφαση είναι αυτός που παρουσίασε ο Zhang (2000) και χρησιμοποιείται ευρέως στην πράξη.

5.2 Κάμερα Οπής

Ο σκοπός μιας κάμερας ως αισθητήρας όρασης είναι να μετράει την ένταση του φωτός που ανακλάται από κάποιο σώμα. Γι αυτό τον λόγο, χρησιμοποιείται ένα *φωτοευαίσθητο στοιχείο* ή *εικονοστοιχείο (pixel)* το οποίο έχει την ιδιότητα να μετασχηματίζει την ενέργεια του φωτός σε ηλεκτρική ενέργεια. Οι πιο γνωστοί τύποι αισθητήρων είναι οι CCD και CMOS που βασίζονται στο φωτοηλεκτρικό φαινόμενο των ημιαγωγών.



Σχήμα 5.1: Προοπτική προβολή για κάμερα οπής, όπου f το εστιακό βάθος της κάμερας (Μαραγκός, 2005).

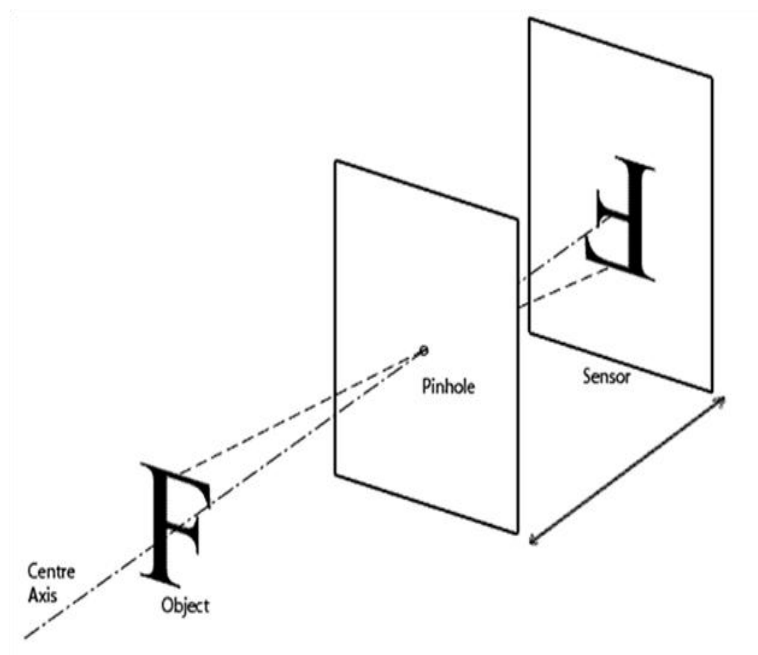
Αρχικά θεωρούμε κεντρική προβολή των σημείων του χώρου στο επίπεδο. Το κέντρο της προβολής ονομάζεται και κέντρο της κάμερας ή οπτικό κέντρο. Η κάθετη ευθεία στο επίπεδο της εικόνας που διέρχεται από το κέντρο της κάμερας, ονομάζεται κύριος ή πρωτεύον άξονας. Επίσης το σημείο τομής του κύριου άξονα με το επίπεδο της εικόνας ονομάζεται κύριο ή πρωτεύον σημείο (βλέπε Σχήμα 5.1).

Η ακτίνα φωτός που εκπέμπεται από ένα σημείο $A = (X^c, Y^c, Z^c)$ του 3Δ αντικειμένου, ορισμένο ως προς το οπτικό κέντρο, αφού διέλθει μέσω της οπής της κάμερας, τέμνει το επίπεδο εικόνας στο σημείο $A' = (x, y, f)$ όπου f το εστιακό βάθος της κάμερας. Οπότε από τους τύπους της προοπτικής προβολής έχουμε ότι

$$x = -f \frac{X^c}{Z^c}$$

$$y = -f \frac{Y^c}{Z^c}$$

Το πρόσημο μείον στις παραπάνω εξισώσεις της προοπτικής προβολής οφείλεται στο γεγονός ότι η εικόνα του 3Δ αντικειμένου φαίνεται ανεστραμμένη στο επίπεδο της εικόνας (Σχήμα 5.2).



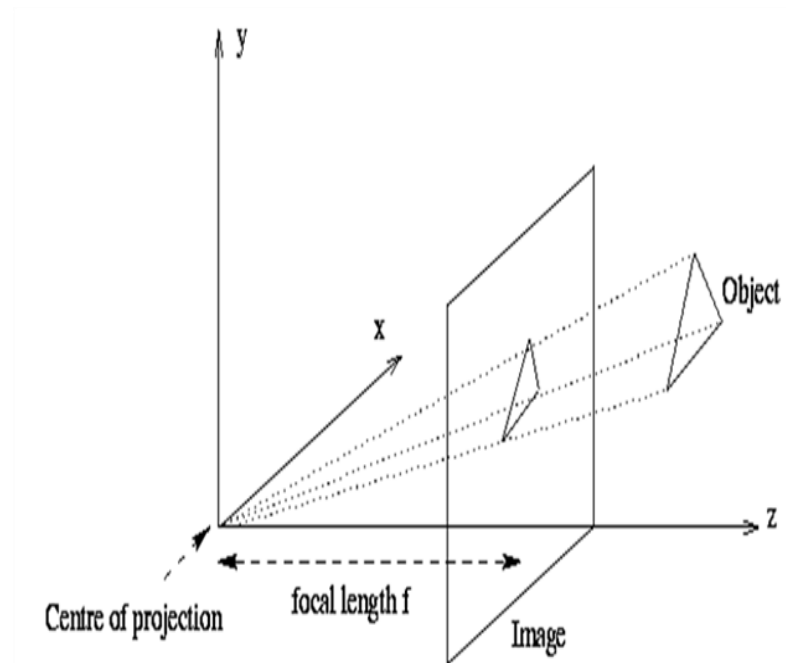
Σχήμα 5.2: : Στο επίπεδο της εικόνας (image plane) το αντικείμενο μας φαίνεται ανεστραμμένο.

Αυτό το φαινόμενο, μπορεί να αποφευχθεί, για λόγους ευκολίας υπολογισμών, θεωρώντας ένα εικονικό επίπεδο εικόνας τοποθετημένο πριν από την οπή σε απόσταση $Z = -f$ (βλέπε Σχήμα 5.3). Με αυτό τον τρόπο, το σχήμα δεν απεικονίζεται στο επίπεδο της εικόνας αναστραμμένο και η περαιτέρω επεξεργασία του απλουστεύεται.

Στην περίπτωση αυτή οι εξισώσεις της προοπτική προβολής που παραθέσαμε παραπάνω μετασχηματίζονται ως

$$x = f \frac{X^c}{Z^c} \quad (5.1)$$

$$y = f \frac{Y^c}{Z^c} \quad (5.2)$$



Σχήμα 5.3: Προοπτική προβολή αντικειμένου χρησιμοποιώντας εικονικό επίπεδο εικόνας, πριν από την οπή.

Όπως προαναφέραμε, στις κάμερες η αισθητήρια μονάδα είναι το εικονοστοιχείο (pixel). Συνεπώς, στις εφαρμογές, πρέπει να μετατρέπονται οι συντεταγμένες του επιπέδου της εικόνας (x, y) σε συντεταγμένες εικονοστοιχείων (X_I, Y_I) .

Ωστόσο, το μοντέλο της κάμερας οπής υποθέτει ότι οι εικονικές συντεταγμένες είναι ευκλείδειες συντεταγμένες με ίδια κλίμακα και στις δυο αξονικές κατευθύνσεις. Στην περίπτωση χρήσης CCD κάμερας υπάρχει η πιθανότητα μη τετραγωνικών εικονοστοιχείων. Αυτό έχει ως αποτέλεσμα την εισαγωγή άνισων παραγόντων κλίμακας σε κάθε κατεύθυνση.

Πιο συγκεκριμένα, αν ο αριθμός εικονοστοιχείων ανά μονάδα απόστασης μετρημένη σε εικονικές συντεταγμένες είναι a_x και a_y κατά μήκος των κατευθύν-

σεων x, y αντίστοιχα, τότε ο μετασχηματισμός των εξισώσεων της προοπτικής προβολής (5.1), (5.2) για το σύστημα συντεταγμένων των εικονοστοιχείων δίνεται από τις παρακάτω εξισώσεις (Shapiro & Stockman, 2001).

$$X_I = f \frac{a_x X^c}{Z^c} + X_0 \quad (5.3)$$

$$Y_I = f \frac{a_y Y^c}{Z^c} + Y_0 \quad (5.4)$$

Όπου X_0 και Y_0 δηλώνουν την μετατόπιση του κεντρικού εικονοστοιχείου του επιπέδου της εικόνας ως προς τον οπτικό άξονα.

Ο παραπάνω μη γραμμικός μετασχηματισμός μπορεί να γραφεί σε γραμμική μορφή, θεωρώντας την ομοιογενή αναπαράσταση του σημείου (x_I, y_I, z_I) μέσω των σχέσεων

$$X_I = \frac{x_I}{\lambda}$$

$$Y_I = \frac{y_I}{\lambda}$$

Όπου $\lambda > 0$. Συνεπώς, οι εξισώσεις (5.3) και (5.4) γράφονται στην μορφή (Siciliano et al., 2009)

$$\begin{bmatrix} x_I \\ y_I \\ z_I \end{bmatrix} = \lambda \begin{bmatrix} X_I \\ Y_I \\ 1 \end{bmatrix} = \mathbf{\Omega} \mathbf{\Pi} \begin{bmatrix} X^c \\ Y^c \\ Z^c \\ 1 \end{bmatrix} \quad (5.5)$$

Όπου

$$\mathbf{\Omega} = \begin{bmatrix} f a_x & 0 & X_0 \\ 0 & f a_y & Y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{\Pi} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Θεωρώντας ότι το οπτικό κέντρο της κάμερας μας, βρίσκεται στο σημείο $O_c = (x_c, y_c, z_c)$, ως προς τις συντεταγμένες βάσης, και \mathbf{T}_b^c η μήτρα μετασχηματισμού των συντεταγμένων βάσης ως προς το πλαίσιο της κάμερας, τότε η συνολική μήτρα μετασχηματισμού που μας δίνει τις συντεταγμένες στο πλαίσιο της κάμερας ενός σημείου στο χώρο εκφρασμένο ως προς το πλαίσιο βάσης είναι

$$\mathbf{\Xi} = \mathbf{\Omega} \cdot \mathbf{\Pi} \cdot \mathbf{T}_b^c \quad (5.6)$$

Η μήτρα της εξίσωσης (5.6), ονομάζεται και *μήτρα βαθμονόμησης κάμερας*.

Πρέπει να επισημάνουμε ότι η παραπάνω μήτρα περιέχει τόσο *εσωτερικές παραμέτρους* της κάμερας - όπως τους όρους (a_x, a_y, X_0, Y_0, f) της μήτρας $\mathbf{\Omega}$, οι οποίες εσωτερικές παράμετροι εξαρτώνται από τους αισθητήρες και τους φακούς της κάμερας - όσο και *εξωτερικούς παραμέτρους* όπως η μήτρα \mathbf{T}_c^b που εξαρτάται από την σχετική θέση και προσανατολισμό της κάμερας ως προς το πλαίσιο βάσης.

Πολλές τεχνικές βαθμονόμησης κάμερας έχουν αναπτυχθεί προκειμένου να μπορέσουμε να προσδιορίσουμε τόσο τις εξωτερικές όσο και τις εσωτερικές παραμέτρους της κάμερας. Στόχος είναι να προσδιορίσουμε τον μετασχηματισμό μεταξύ του καρτεσιανού επιπέδου της βάσης και του πλαισίου της κάμερας όσο ακριβέστερα μπορούμε. Στην ενότητα, «Βαθμονόμηση Κάμερας», αναλύουμε εκτενώς την μέθοδο που πρότεινε ο Zhang (2000).

Στην περίπτωση που οι εσωτερικές παράμετροι της κάμερας είναι γνωστοί τότε για υπολογιστική ευκολία, μας βολεύει να αναφερόμαστε στις κανονικοποιημένες συντεταγμένες (X, Y) , ορισμένες ως

$$\lambda \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \mathbf{\Pi} \begin{bmatrix} X^c \\ Y^c \\ Z^c \\ 1 \end{bmatrix} \quad (5.7)$$

Οι παραπάνω συντεταγμένες συμπίπτουν με αυτές των εξισώσεων (5.1) και (5.2) στην περίπτωση που έχουμε $f = 1$.

Έχοντας υπολογίσει τις κανονικοποιημένες συντεταγμένες είναι εύκολο να παράγουμε τις συντεταγμένες των εικονοστοιχείων με τον παρακάτω μετασχηματισμό που προκύπτει άμεσα από την εξίσωση (5.5)

$$\begin{bmatrix} X_I \\ Y_I \\ 1 \end{bmatrix} = \mathbf{\Omega} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (5.8)$$

5.3 Στερεοπτική Όραση

Η δισδιάστατη εικόνα που μας δίνει η κάμερα δεν παρέχει από μόνη της κάποια πληροφορία για το βάθος του αντικειμένου, που ορίζεται ως η απόσταση του παρατηρούμενου αντικειμένου σε σχέση με την κάμερα.

Στην περίπτωση που έχουμε διαθέσιμη μόνο μια λήψη από την κάμερα, πρέπει να γνωρίζουμε το γεωμετρικό μοντέλο του αντικειμένου, προκειμένου να αποκτήσουμε

κάποια πληροφορία για το βάθος του. Ωστόσο, η γνώση του γεωμετρικού μοντέλου τυχαίου σχήματος αντικειμένου είναι αδύνατη οπότε πρέπει να αναζητηθούν διαφορετικοί μέθοδοι προσδιορισμού του βάθους.

Μια τέτοια μέθοδος είναι αυτή κατά την οποία το βάθος του αντικειμένου υπολογίζεται χρησιμοποιώντας μόνο δύο λήψεις της ίδιας σκηνής, από διαφορετική οπτική γωνία κάθε φορά. Με τον όρο «σκηνή» εννοούμε το αντικείμενο μαζί με το φόντο από το οποίο περιβάλλεται. Οι δύο λήψεις μπορούν να προέρχονται είτε από δυο διαφορετικές κάμερες είτε από μια κινούμενη κάμερα. Η παραπάνω μέθοδος στην βιβλιογραφία αναφέρεται ως *Στερεοπτική Όραση* (Hartley & Zisserman, 2003).

Προκειμένου να εφαρμοστεί η μέθοδος της στερεοπτικής όρασης για τον εντοπισμό του βάθους του αντικειμένου πρέπει προηγουμένως να έχουν λυθεί δύο ουσιώδη προβλήματα. Το πρώτο, που αναφέρεται ως «πρόβλημα αντιστοίχισης», έγκειται στον εντοπισμό των όμοιων σημείων που υπάρχουν στις δύο εικόνες. Με τον όρο «όμοια σημεία» αναφερόμαστε στα σημεία εκείνα τα οποία αντιστοιχούν στην προβολή του ίδιου σημείου στα επίπεδα εικόνες των δυο καμερών.

Όπως είναι φανερό, το πρόβλημα αντιστοίχισης είναι πολύπλοκο και η λύση του στηρίζεται στην εύρεση γεωμετρικών περιορισμών που υπάρχουν ανάμεσα στις δύο εικόνες καθώς και στον εντοπισμό λεπτομερειών της σκηνής που εμφανίζονται το ίδιο και στις δυο λήψεις.

Το δεύτερο ουσιώδες πρόβλημα είναι ο υπολογισμός της σχετικής θέσης και του προσανατολισμού της κάμερας ως προς τη παρατηρούμενη σκηνή. Εκτός, λοιπόν, από την ύπαρξη των δυο λήψεων είναι απαραίτητο να γνωρίζουμε και την οπτική γωνία της κάμερας όταν αυτές οι δύο λήψεις αποκτήθηκαν. Το συγκεκριμένο πρόβλημα είναι εξίσου δύσκολο με το προηγούμενο ωστόσο η λύση του δεν είναι απαραίτητη στις περισσότερες εφαρμογές καθώς η θέση και ο προσανατολισμός της κάμερας είναι γνωστός εκ των προτέρων. Για παράδειγμα, όταν η κάμερα είναι προσαρμοσμένη σε ένα ρομποτικό βραχίονα τότε μπορεί να ελεγχθεί η κίνηση του βραχίονα ώστε να πάρουμε μια δεύτερη λήψη, ενώ όταν χρησιμοποιούμε δυο κάμερες στο χώρο μπορούμε από πριν να γνωρίζουμε την θέση τους.

5.3.1 Επιπολική Γεωμετρία

Έστω ότι διαθέτουμε δύο κάμερες. Θεωρούμε τα πλαίσια αναφοράς 1 και 2, αντίστοιχα για κάθε κάμερα. Επιπλέον, έστω $\mathbf{o}_{1,2}^1$ το διάνυσμα θέσης και \mathbf{R}_2^1 ο πίνακας περιστροφής, του πλαισίου 2 ως προς το πλαίσιο 1. Τότε, οι συντεταγμένες ενός σημείου P, ως προς το πλαίσιο 1, γνωρίζοντα τις συντεταγμένες του ως προς το πλαίσιο 2, θα δίνεται από την εξίσωση

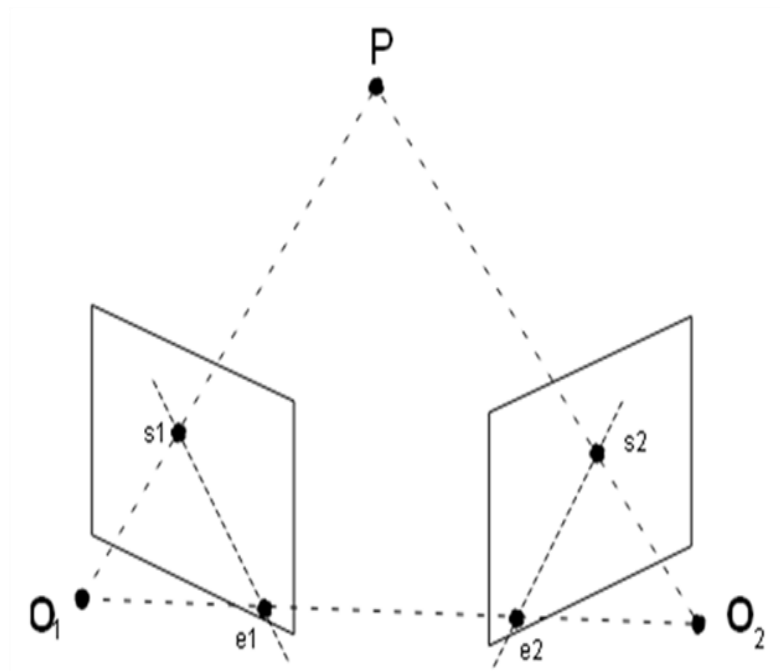
$$\mathbf{p}^1 = \mathbf{o}_{1,2}^1 + \mathbf{R}_2^1 \cdot \mathbf{p}^2 \quad (5.9)$$

Έστω \mathbf{s}_1 και \mathbf{s}_2 οι συντεταγμένες των προβολών του σημείου P , στα δυο επίπεδα εικόνας των καμερών, αντίστοιχα (βλέπε σχήμα 5.4). Έστω $\tilde{\mathbf{s}}_1$ και $\tilde{\mathbf{p}}_i$ τα επαυξημένα (ομογενοποιημένα) διανύσματα. Από την εξίσωση (5.7) έχουμε ότι

$$\lambda_i \cdot \tilde{\mathbf{s}}_i = \mathbf{\Pi} \cdot \tilde{\mathbf{p}}_i = \mathbf{p}^i \quad i = 1,2 \quad (5.10)$$

Αντικαθιστώντας στην εξίσωση (5.9) την (5.10) παίρνουμε

$$\lambda_1 \cdot \tilde{\mathbf{s}}_1 = \mathbf{o}_{1,2}^1 + \lambda_2 \cdot \mathbf{R}_2^1 \cdot \tilde{\mathbf{s}}_2 \quad (5.11)$$



Σχήμα 5.4: : Επιπολική Γεωμετρία

Πολλαπλασιάζοντας κάθε μέρος της εξίσωσης (5.11) με τον αντισυμμετρικό πίνακα $S(\mathbf{o}_{1,2}^1)$ όπου

$$\mathbf{o}_{1,2}^1 = \begin{bmatrix} o_x \\ o_y \\ o_z \end{bmatrix} \text{ και } S(\mathbf{o}_{1,2}^1) = \begin{bmatrix} 0 & -o_z & o_y \\ o_z & 0 & -o_x \\ -o_y & o_x & 0 \end{bmatrix}$$

έχουμε ότι

$$\lambda_1 \cdot S(\mathbf{o}_{1,2}^1) \cdot \tilde{\mathbf{s}}_1 = S(\mathbf{o}_{1,2}^1) \cdot \mathbf{o}_{1,2}^1 + \lambda_2 \cdot S(\mathbf{o}_{1,2}^1) \cdot \mathbf{R}_2^1 \cdot \tilde{\mathbf{s}}_2 \Leftrightarrow$$

$$\lambda_1 \cdot S(\mathbf{o}_{1,2}^1) \cdot \tilde{\mathbf{s}}_1 = \lambda_2 \cdot S(\mathbf{o}_{1,2}^1) \cdot \mathbf{R}_2^1 \cdot \tilde{\mathbf{s}}_2$$

πολλαπλασιάζοντας την παραπάνω εξίσωση με $\widetilde{\mathbf{s}}_1^T$ παίρνουμε ότι

$$\begin{aligned}\lambda_1 \widetilde{\mathbf{s}}_1^T \cdot S(\mathbf{o}_{1,2}^1) \cdot \widetilde{\mathbf{s}}_1 &= \lambda_2 \widetilde{\mathbf{s}}_1^T \cdot S(\mathbf{o}_{1,2}^1) \cdot \mathbf{R}_2^1 \cdot \widetilde{\mathbf{s}}_2 \Leftrightarrow \\ \lambda_2 \widetilde{\mathbf{s}}_1^T \cdot S(\mathbf{o}_{1,2}^1) \cdot \mathbf{R}_2^1 \cdot \widetilde{\mathbf{s}}_2 &= 0\end{aligned}$$

Η οποία πρέπει να ισχύει για οποιοδήποτε $\lambda_2 > 0$.

Συνεπώς η παραπάνω εξίσωση είναι ισοδύναμη με την εξίσωση *επιπολικού περιορισμού* (Hartley & Zisserman, 2003).

$$\widetilde{\mathbf{s}}_1^T \cdot \mathbf{E} \cdot \widetilde{\mathbf{s}}_2 = 0 \quad (5.12)$$

Όπου $\mathbf{E} = S(\mathbf{o}_{1,2}^1) \cdot \mathbf{R}_2^1$ είναι ένας (3×3) πίνακας, γνωστός και ως *ουσιώδης μήτρα*. Η εξίσωση (5.12) εκφράζει σε αναλυτική μορφή τον γεωμετρικό περιορισμό που υπάρχει μεταξύ των προβολών του ίδιου σημείου στα επίπεδα εικόνας των δύο καμερών.

Η ερμηνεία του γεωμετρικού περιορισμού της εξίσωσης (5.12), η οποία μπορεί να γίνει κατανοητή από την εποπτεία του σχήματος 5.4, είναι ότι τα διανύσματα \mathbf{s}_1 , $\mathbf{R}_2^1 \cdot \mathbf{s}_2$, και $\mathbf{o}_{1,2}^1$, που είναι όλα εκφρασμένα ως προς το πλαίσιο 1, πρέπει να ανήκουν στο ίδιο επίπεδο που ονομάζεται *επιπολικό επίπεδο*.

Μέσω του επιπολικού επιπέδου, μπορούμε να απλουστεύσουμε το πρόβλημα της αντιστοίχισης όμοιων σημείων μεταξύ των δύο λήψεων της κάμερα που ορίσαμε προηγουμένως. Πιο συγκεκριμένα, άμα γνωρίζουμε το διάνυσμα \mathbf{s}_1 και το διάνυσμα θέσης $\mathbf{o}_{1,2}^1$ τότε το αντίστοιχο σημείο της προβολής του P στο πλαίσιο 2 θα βρίσκεται πάνω στο επίπεδο που ορίζουν τα δύο αυτά διανύσματα, το επιπολικό επίπεδο. Συνεπώς, εφόσον έχουμε τομή δύο επιπέδων, επίπεδο εικόνας δεύτερης κάμερας και επιπολικό επίπεδο, το αντίστοιχο σημείο της προβολής του P στο πλαίσιο 2, θα βρίσκεται στην ευθεία που ορίζεται από την τομή των δύο επιπέδων (βλέπε σχήμα 5.4). Η ευθεία αυτή ονομάζεται *επιπολική ευθεία* και όπως μπορεί να παρατηρηθεί στο σχήμα 5.4, όντως το σημείο s_2 βρίσκεται πάνω στην ευθεία αυτή (Hartley & Zisserman, 2003)

Με αυτόν τον τρόπο απλουστεύεται σημαντικά το πρόβλημα αναζήτησης όμοιων σημείων καθώς πλέον η αναζήτηση επικεντρώνεται σε μια γραμμή και όχι σε ολόκληρο το επίπεδο εικόνας της κάμερας.

5.3.2 Εύρεση Συντεταγμένων Σημείου - Τριγωνοποίηση

Το πρόβλημα της εύρεσης του βάθους του αντικειμένου από την κάμερα παρατήρησης είναι αντίστοιχο με τον προσδιορισμό των συντεταγμένων κάποιου σημείου της επιφάνειας του αντικειμένου ως προς το πλαίσιο της κάμερας. Για την

επίλυση του προβλήματος της εύρεσης των συντεταγμένων χρησιμοποιούμε την γεωμετρική μέθοδο της *τριγωνοποίησης* (Siciliano et al., 2009).

Στόχος είναι να υπολογιστούν οι συντεταγμένες του σημείου $p = [p_x, p_y, p_z]^T$, ορισμένο ως προς το πλαίσιο βάσης έχοντας ως δεδομένα μόνο τις κανονικοποιημένες συντεταγμένες $s_1 = [X_1, Y_1]^T$ και $s_2 = [X_2, Y_2]^T$ των προβολών του P στα επίπεδα εικόνας των δύο καμερών.

Υποθέτουμε, για απλότητα στους υπολογισμούς, ότι το πλαίσιο βάσης συμπίπτει με το πλαίσιο 1 της κάμερας. Τότε ισχύουν τα παρακάτω.

$$\mathbf{p}^1 = \mathbf{p}, \quad \mathbf{o}_{1,2}^1 = \mathbf{o}, \quad \mathbf{R}_2^1 = \mathbf{R}$$

Από τις εξισώσεις (5.10) και (5.11) παράγονται οι παρακάτω εξισώσεις

$$\mathbf{p} = \lambda_1 \cdot \widetilde{\mathbf{s}}_1 \quad (5.13)$$

$$\mathbf{p} = \mathbf{o} + \lambda_2 \cdot \mathbf{R} \cdot \widetilde{\mathbf{s}}_2 \quad (5.14)$$

όπου η εξίσωση (5.13) δίνει την παραμετρική εξίσωση της οπτικής ακτίνας που περνάει από το O_1 και το σημείο των συντεταγμένων s_1 , ενώ η εξίσωση (5.14) δίνει την παραμετρική εξίσωση της οπτικής ακτίνας που περνάει από το O_2 και το σημείο των συντεταγμένων s_2 . Οι δυο παραμετρικές εξισώσεις είναι εκφρασμένες ως προς το πλαίσιο βάσης, που έχει θεωρηθεί ότι συμπίπτει με το πλαίσιο 1 της κάμερας.

Συνεπώς, οι συντεταγμένες του σημείου P δίνονται από την τομή των δύο οπτικών ακτινών και μπορούν να υπολογιστούν λύνοντας το σύστημα των εξισώσεων (5.13) και (5.14) ως προς την παράμετρο \mathbf{p} .

Από την εξίσωση (5.13) προκύπτει το παρακάτω σύστημα

$$\left. \begin{array}{l} p_x = \lambda_1 X_1 \\ p_y = \lambda_1 Y_1 \\ p_z = \lambda_1 \end{array} \right\} = \left. \begin{array}{l} p_x - X_1 p_z = 0 \\ p_y - Y_1 p_z = 0 \end{array} \right\} \Leftrightarrow \begin{bmatrix} 1 & 0 & -X_1 \\ 0 & 1 & -Y_1 \end{bmatrix} \mathbf{p} = 0 \quad (5.15)$$

Με παρόμοιο τρόπο, αφού πρώτα πολλαπλασιαστούν και τα δύο μέλη της εξίσωσης (5.14) με τον πίνακα \mathbf{R}^T προκύπτει το παρακάτω σύστημα

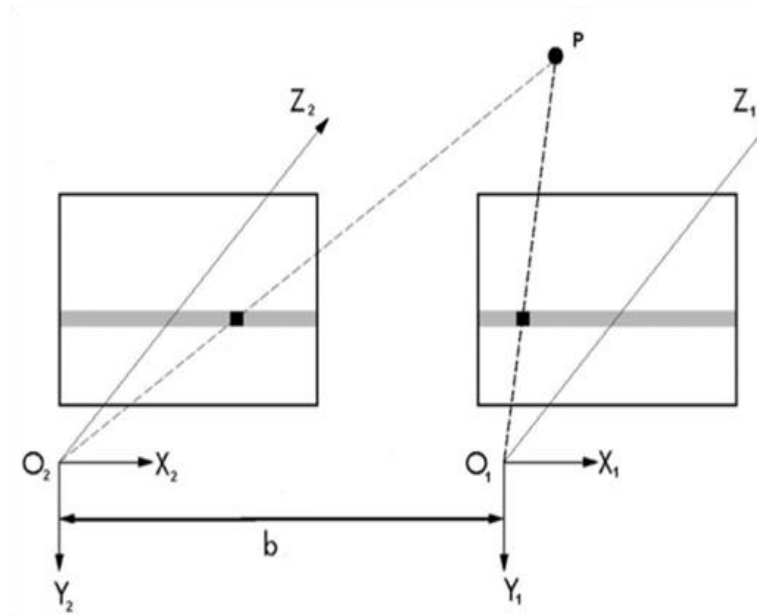
$$\begin{bmatrix} r_1^T - X_2 r_3^T \\ r_2^T - Y_2 r_3^T \end{bmatrix} \mathbf{p} = \begin{bmatrix} o_x - o_z X_2 \\ o_y - o_z Y_2 \end{bmatrix} \quad (5.16)$$

Όπου $\mathbf{R} = [r_1 \ r_2 \ r_3]$ και $\mathbf{R}^T \mathbf{o} = [o_x \ o_y \ o_z]^T$

Οι εξισώσεις (5.15) και (5.16) ορίζουν ένα σύστημα τεσσάρων εξισώσεων και τριών αγνώστων, το οποίο είναι γραμμικό ως προς το \mathbf{p} . Απ' αυτές τις εξισώσεις μόνο τρεις είναι ανεξάρτητες, στην ιδανική περίπτωση που οι δύο οπτικές ακτίνες τέμνονται στο σημείο P .

Στην πράξη ωστόσο, επειδή στο σύστημα υπεισέρχεται θόρυβος, όλες οι παραπάνω εξισώσεις είναι ανεξάρτητες μεταξύ τους και το σύστημα είναι αδύνατο. Γι αυτό τον λόγο πρέπει να χρησιμοποιηθούν αλγόριθμοι που βασίζονται σε μεθόδους ελαχίστων τετραγώνων προκειμένου να αποκτηθεί μια προσεγγιστική λύση.

Ωστόσο, ο υπολογισμός του \mathbf{p} μπορεί να απλοποιηθεί πολύ στην περίπτωση που οι δύο κάμερες έχουν παράλληλα και στοιχισμένα επίπεδα όπως φαίνεται στο σχήμα 5.5.



Σχήμα 5.5: Παράλληλα και στοιχισμένα επίπεδα εικόνας για τις δυο κάμερες που χρησιμοποιούμε.

Στην περίπτωση αυτή, που είναι ιδιαίτερα συχνή στην πράξη, έχουμε $\mathbf{R} = \mathbf{I}$ και $\mathbf{R}^T \mathbf{o} = [b \ 0 \ 0]^T$, όπου $b > 0$ η απόσταση μεταξύ των δύο πλαισίων όπως φαίνεται και στο σχήμα 5.5. Συνεπώς, η λύση του συστήματος των εξισώσεων (5.15) και (5.16) είναι (Siciliano et al., 2009)

$$p_x = \frac{X_1 b}{X_1 - X_2} \quad (5.17)$$

$$p_y = \frac{Y_1 b}{X_1 - X_2} = \frac{Y_2 b}{X_1 - X_2} \quad (5.18)$$

$$p_z = \frac{b}{X_1 - X_2} \quad (5.19)$$

Αμα δίνονται ως είσοδος συντεταγμένες εικονοστοιχείων (pixel) αντί για κανονικοποιημένες συντεταγμένες εφαρμόζεται ο παρακάτω μετασχηματισμό και οι σχέσεις (5.17)-(5.19)

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \mathbf{\Omega}^{-1} \begin{bmatrix} X_I \\ Y_I \\ 1 \end{bmatrix}$$

5.4 Βαθμονόμηση Κάμερας

Ένα σημαντικό πρόβλημα για τις ρομποτικές εφαρμογές που χρησιμοποιούν κάμερα, είναι αυτό της βαθμονόμησης. Με την βαθμονόμηση προσπαθούμε να προσδιορίσουμε τόσο τις εσωτερικές όσο και τις εξωτερικές παραμέτρους της κάμερας. Άρα επικεντρωνόμαστε στον προσδιορισμό της μήτρας βαθμονόμησης της εξίσωσης 5.6.

Αρχικά, προσδιορίζουμε της εσωτερικές παραμέτρους, που δίνονται από τον πίνακα $\mathbf{\Omega}$. Προκειμένου να το επιτύχουμε αυτό, χρησιμοποιούμε την τεχνική βαθμονόμησης που πρότεινε ο Zhang (2000), σύμφωνα με την οποία η κάμερα χρειάζεται να παρατηρήσει ένα επίπεδο αντικείμενο βαθμονόμησης σε μερικούς διαφορετικούς προσανατολισμούς (τουλάχιστον δύο). Είτε η κάμερα είτε το αντικείμενο μπορούν ελεύθερα να κινηθούν χωρίς να πρέπει να είναι γνωστή αυτή η κίνηση.

Θεωρούμε μια κάμερα και θέτουμε το σημείο O_c ως το οπτικό κέντρο της. Έστω O_o η αρχή των αξόνων του πλαισίου αναφοράς του επίπεδου αντικειμένου βαθμονόμησης. Έστω, n σημεία του αντικειμένου βαθμονόμησης με

$$r_{o,i}^o = \begin{bmatrix} X_i^o \\ Y_i^o \\ Z_i^o \end{bmatrix}, \quad i = 1, \dots, n$$

τα διανύσματα θέσης των σημείων αυτών ως προς το πλαίσιο αναφοράς του αντικειμένου. Για απλοποίηση των υπολογισμών θεωρούμε ότι το αντικείμενο βαθμονόμησης βρίσκεται στο επίπεδο $Z_i^o = 0$.

Από την εξίσωση 5.5 έχουμε για τις συντεταγμένες εικονοστοιχείων $\tilde{z}_i = [X_{Ii} \ Y_{Ii} \ 1]^T$, της προβολής των σημείων του αντικειμένου στο επίπεδο εικόνας της κάμερας ότι

$$\lambda_i \begin{bmatrix} X_{li} \\ Y_{li} \\ 1 \end{bmatrix} = \mathbf{\Omega} \mathbf{\Pi} \begin{bmatrix} X_i^c \\ Y_i^c \\ 0 \\ 1 \end{bmatrix} = \mathbf{\Omega} \mathbf{\Pi} \mathbf{T}_o^c \begin{bmatrix} X_i^o \\ Y_i^o \\ 0 \\ 1 \end{bmatrix} = \mathbf{\Omega} [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{o}_{c,o}^c] \begin{bmatrix} X_i^o \\ Y_i^o \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} X_i^o \\ Y_i^o \\ 1 \end{bmatrix}$$

Όπου \mathbf{T}_o^c η μήτρα μετασχηματισμού, \mathbf{r}_1 και \mathbf{r}_2 η πρώτη και η δεύτερη στήλη του πίνακα περιστροφής \mathbf{R}_o^c . Ο πίνακας $\mathbf{H} = \mathbf{\Omega} [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{o}_{c,o}^c]$ ονομάζεται πίνακας επιπέδου ομοιογραφήματος και στόχος μας είναι να τον προσδιορίσουμε, γνωρίζοντας μόνο τις συντεταγμένες των n σημείων του αντικειμένου, ως προς το πλαίσιο αναφοράς O_o και τις προβολές αυτών των σημείων στο επίπεδο της εικόνας.

Αν η παραπάνω σχέση πολλαπλασιαστεί με τον αντισυμμετρικό πίνακα $\mathbf{S}(\tilde{\mathbf{c}}_i)$ δίνει την παρακάτω εξίσωση

$$\mathbf{S}(\tilde{\mathbf{c}}_i) \mathbf{H} \cdot [X_i^o \quad Y_i^o \quad 1] = \mathbf{0} \quad (5.20)$$

Η εξίσωση (20) είναι συμμετρική ως προς \mathbf{H} , οπότε μπορεί να γραφεί στην παρακάτω μορφή.

$$\mathbf{A}_i(\mathbf{c}_i) \cdot \mathbf{h} = \mathbf{0} \quad (5.21)$$

Όπου ο \mathbf{h} είναι ένας (9×1) διάνυσμα στήλης και προκύπτει άμα βάλουμε τις στήλες του $\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3]$ την μια δίπλα στην άλλη.

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{bmatrix}$$

Ενώ ο πίνακας $\mathbf{A}_i(\mathbf{c}_i)$ είναι της μορφής

$$\mathbf{A}_i(\mathbf{c}_i) = [\mathbf{S}(\tilde{\mathbf{c}}_i) X_i^o \quad \mathbf{S}(\tilde{\mathbf{c}}_i) Y_i^o \quad \mathbf{S}(\tilde{\mathbf{c}}_i)]$$

Από την στιγμή που ο βαθμός του αντισυμμετρικού πίνακα $\mathbf{S}(\tilde{\mathbf{c}}_i)$ είναι το πολύ 2, συνεπάγεται ότι και ο βαθμός του πίνακα $\mathbf{A}_i(\mathbf{c}_i)$ θα είναι επίσης το πολύ 2. Συγκεκριμένα, ορίζοντας το διάνυσμα $\overline{\mathbf{M}}_i = [X_i^o \quad Y_i^o \quad 1]^T$ και εκτελώντας τις πράξεις για τον $\mathbf{A}_i(\mathbf{c}_i)$, η τελική μορφή του πίνακα θα είναι η ακόλουθη

$$\mathbf{A}_i(\mathbf{c}_i) = \begin{bmatrix} \overline{\mathbf{M}}_i^T & 0^T & -X_{li} \cdot \overline{\mathbf{M}}_i^T \\ 0^T & \overline{\mathbf{M}}_i^T & -Y_{li} \cdot \overline{\mathbf{M}}_i^T \end{bmatrix}$$

Η διάσταση του $\mathbf{A}_i(\mathbf{c}_i)$ είναι ένας πίνακας (2×9) . Αν δοθούν $n > 4$ σημεία, τότε έχουμε n εξισώσεις της μορφής (5.21), οι οποίες μπορούν να γραφούν ως εξίσωση πινάκων $L \cdot \mathbf{h} = \mathbf{0}$, όπου $L : 2n \times 9$ πίνακας.

Οπότε, σύμφωνα με τον Zhang (2000), το παραπάνω σύστημα είναι υπερπλήρες και μια λύση του, υποκείμενη στον περιορισμό $\|\mathbf{h}\| = 1$, αποτελεί το ιδιοδιάνυσμα του $L^T L$ που αντιστοιχεί στην μικρότερη ιδιοτιμή. Στον πίνακα L κάποια στοιχεία είναι σταθερά, κάποια σε συντεταγμένες εικονοστοιχείων, κάποια σε παγκόσμιες συντεταγμένες και κάποια πολλαπλασιασμός και των δύο. Αυτό το γεγονός κάνει τον L μη επαρκώς αριθμητικά ορισμένο.

Συνεπώς, είναι σκόπιμο πριν την επίλυση του συστήματος να γίνει μια ομαλοποίηση των δεδομένων. Η συνηθέστερη ομαλοποίηση είναι η εξής: Τα σημεία της εικόνας μεταφέρονται έτσι ώστε το κέντρο μάζας να είναι στο 0 και κλιμακούμενα έτσι ώστε η RMS απόσταση από την αρχή των αξόνων να είναι ίση με $\sqrt{2}$. Αν T , ο πίνακας ομαλοποίησης τότε ισχύει: $\mathbf{c}_i^{norm} = T \cdot \mathbf{c}_i$

Το είδος της ομαλοποίησης που πρέπει να επιβληθεί στα 2Δ σημεία του χώρου αποτελεί ένα κάπως πιο δύσκολο πρόβλημα. Στην περίπτωση που η μεταβολή του βάθους είναι σχετικά ομαλή επιλέγουμε αντίστοιχη ομαλοποίηση με πριν. Έτσι λοιπόν σε αυτήν την περίπτωση το κέντρο μάζας των σημείων μεταφέρεται στην αρχή των αξόνων και οι συντεταγμένες τους κλιμακώνονται έτσι ώστε η RMS απόσταση από την αρχή να είναι ίση με $\sqrt{3}$. Αν U , ο πίνακας ομαλοποίησης τότε ισχύει: $\mathbf{M}_i^{norm} = U \cdot \mathbf{M}_i$

Η γραμμική εκτίμηση του ομοιογραφήματος αποτελεί την αρχικοποίηση του επαναληπτικού αλγορίθμου Levenberg-Marquardt (Madsen & Nielsen, 2004) που ελαχιστοποιεί το γεωμετρικό λάθος του $\hat{\mathbf{H}}$, όπου $\hat{\mathbf{H}}$, το ομοιογράφημα που έχει προέλθει με την χρήση ομαλοποιημένων συντεταγμένων. Προκειμένου να αποκτηθεί το πραγματικό ομοιο-γράφημα \mathbf{H} πρέπει να γίνει ο παρακάτω μετασχηματισμός

$$\mathbf{H} = \mathbf{T}^{-1} \cdot \hat{\mathbf{H}} \cdot \mathbf{U}$$

Η λύση που παίρνουμε από το ιδιοδιάνυσμα του $L^T L$ που αντιστοιχεί στην μικρότερη ιδιοτιμή, είναι της μορφής $\zeta \mathbf{h}$, όπου ζ κάποιος συντελεστής κλιμάκωσης. Άρα, η λύση για τον \mathbf{H} θα είναι της μορφής $\zeta \mathbf{H}$.

Συνεπώς θα ισχύει ότι

$$\zeta[\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3] = \Omega[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{o}_{c,o}^e]$$

Όπου, όπως έχουμε προαναφέρει h_i δηλώνει την i στήλη του πίνακα \mathbf{H} . Χρησιμοποιώντας τη γνώση ότι τα \mathbf{r}_1 και \mathbf{r}_2 είναι ορθοκανονικά, παίρνουμε τους παρακάτω περιορισμούς (Zhang, 2000)

$$\mathbf{h}_1^T \Omega^{-T} \Omega^{-1} \mathbf{h}_2 = 0 \quad (5.22)$$

$$\mathbf{h}_1^T \Omega^{-T} \Omega^{-1} \mathbf{h}_1 = \mathbf{h}_2^T \Omega^{-T} \Omega^{-1} \mathbf{h}_2 \quad (5.23)$$

Οι σχέσεις (5.22) και (5.23) αποτελούν δύο βασικούς περιορισμούς στις εσωτερικές παραμέτρους, δοσμένου ενός ομοιογραφήματος. Επειδή ένα ομοιογράφημα έχει 8 βαθμούς ελευθερίας και 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 δεύτερη θέση είναι παράλληλο σε σχέση με την πρώτη θέση, τότε το δεύτερο ομοιογράφημα δεν παρέχει επιπλέον γεωμετρικούς περιορισμούς για την επίλυση του προβλήματος και μιλάμε για εκφυλισμένη περίπτωση. Οπότε σύμφωνα με την μεθοδολογία που πρότεινε ο Zhang (2000), πρέπει να δοθεί ιδιαίτερη προσοχή ώστε οι λήψεις που λαμβάνουμε για το αντικείμενο βαθμονόμησης να μην είναι παράλληλες.

Οι σχέσεις (5.22) και (5.23) είναι γραμμικές και μπορούν να γραφούν στη μορφή

$$\mathbf{A}'\mathbf{b} = \mathbf{0} \quad (5.24)$$

Στην παραπάνω εξίσωση, ο \mathbf{A}' είναι ένας (2×6) πίνακας με συντελεστές που εξαρτάται από τα $\mathbf{h}_1, \mathbf{h}_2$, ενώ ο $\mathbf{b} = [b_{11} \ b_{12} \ b_{22} \ b_{13} \ b_{23} \ b_{33}]$ είναι ο πίνακας του οποίου τα στοιχεία b_{ij} δίνονται από τον συμμετρικό πίνακα

$$\mathbf{B} = \mathbf{\Omega}^{-T}\mathbf{\Omega}^{-1} = \begin{bmatrix} 1/a_x^2 & 0 & -X_0/a_x^2 \\ 0 & 1/a_y^2 & -Y_0/a_y^2 \\ -X_0/a_x^2 & -Y_0/a_y^2 & 1 + X_0^2/a_x^2 + Y_0^2/a_y^2 \end{bmatrix} \quad (5.25)$$

Αν διαθέτουμε k εικόνες του επιπέδου βαθμονόμησης και επαναλάβουμε όλη την διαδικασία που περιγράψαμε μέχρι τώρα k φορές, τότε αποκτούμε $2k$ εξισώσεις της μορφής (5.25). Αυτές οι εξισώσεις, στην περίπτωση που έχουμε $k \geq 3$, έχουν μοναδική λύση $\gamma\mathbf{b}$ όπου γ συντελεστής κλιμάκωσης. Από την πίνακα $\gamma\mathbf{B}$ που προκύπτει προσδιορίζουμε τις εσωτερικές παραμέτρους της κάμερας μας ως εξής (Siciliano et al., 2009).

$$X_0 = -\frac{b'_{13}}{b'_{11}} \quad (5.26)$$

$$Y_0 = -\frac{b'_{23}}{b'_{22}} \quad (5.27)$$

$$fa_x = \sqrt{\frac{\xi}{b'_{11}}} \quad (5.28)$$

$$fa_y = \sqrt{\frac{\xi}{b'_{22}}} \quad (5.29)$$

Όπου έχουμε ότι

$$b'_{ij} = \gamma b_{ij}$$

$$\xi = b'_{13} + b'_{23} + b'_{33}$$

Έχοντας προσδιορίσει τις εσωτερικές παραμέτρους της κάμερας μας, μπορούμε εύκολα να προσδιορίσουμε και τις εξωτερικές ως εξής

$$r_1 = \lambda \Omega^{-1} h_1 \quad (5.30)$$

$$r_2 = \lambda \Omega^{-1} h_2 \quad (5.31)$$

$$r_3 = r_1 \times r_2 \quad (5.32)$$

$$o_{0,c}^c = \lambda \Omega^{-1} h_3 \quad (5.33)$$

Συνεπώς με τις εξισώσεις (5.30)-(5.33) προσδιορίζουμε την μήτρα μετασχηματισμού T_0^c .

Άρα προκειμένου να προσδιορίσουμε τις εξωτερικές παραμέτρους της κάμερας μας που βρίσκονται στην μήτρα μετασχηματισμού T_c^b , αρκεί να εφαρμόσουμε την παρακάτω εξίσωση,

$$T_c^b = T_0^b (T_0^c)^{-1} \quad (5.34)$$

ΚΕΦΑΛΑΙΟ 6

6 Εκπαίδευση και Προσδιορισμός Παραμέτρων Αυτοοργανούμενου Χάρτη (SOM)

6.1 Εισαγωγή

Η διαδικασία που ακολουθείται για την εύρεση των επιτρεπτών λαβών ενός σύνθετου σχήματος αντικειμένου μπορεί να χωρισθεί σε επιμέρους βήματα. Η κατάτμηση σχήματος μέσω της οποίας το σύνθετο σχήμα αποδομείται σε πρωταρχικά, αποτελεί το αρχικό στάδιο επεξεργασίας. Ακολουθεί η διαδικασία εύρεσης των συντεταγμένων του κέντρου μάζας κάθε πρωταρχικού σχήματος ως προς το πλαίσιο της κάμερας. Έτσι γνωρίζουμε την ακριβή θέση στην οποία πρέπει να βρεθεί η αρπάγη του ρομποτικού βραχίονα προκειμένου να πιάσει το αντικείμενο. Ωστόσο πρέπει να προσδιοριστούν και οι επιτρεπτές λαβές, δηλαδή, να βρεθούν οι διάφοροι προσανατολισμοί που μπορεί να πάρει η αρπάγη ώστε να πιάσει το αντικείμενο.

Συνεπώς, θεωρείται σκόπιμη η χρήση κάποιας διάταξης νευρωνικού δικτύου, ώστε ανάλογα με το σχήμα που θα δίνεται ως είσοδος στο σύστημα, να διαφοροποιείται και η έξοδος δηλαδή το σύνολο των επιτρεπτών λαβών του αντικειμένου. Το σχήμα που θα δίνεται ως είσοδος θα εξάγεται μέσω της διαδικασίας της κατάτμησης που περιγράψαμε αναλυτικά σε προηγούμενο κεφάλαιο και είτε θα είναι κάποιο από τα βασικά σχήματα είτε κάποια παραμόρφωση αυτών. Ως *βασικά ή πρωταρχικά σχήματα* ορίζουμε το ορθογώνιο παραλληλόγραμμο, τον κύκλο, το οκτάγωνο, το τραπέζιο και το τρίγωνο.

Διακρίνονται επομένως δυο περιπτώσεις: είτε (α) το σχήμα που δίνεται ως είσοδος είναι κάποιο από τα πρωταρχικά σχήματα για τα οποία έχει ήδη εκπαιδευτεί το νευρωνικό δίκτυο, παρέχοντας άρα ως έξοδο τις αντίστοιχες επιτρεπτές λαβές, είτε (β) το σχήμα εισόδου είναι κάποιο άγνωστο σχήμα για το οποίο το δίκτυο πρέπει να μην δίνει καμιά επιτρεπτή λαβή. Η διάταξη νευρωνικού δικτύου που θα χρησιμοποιηθεί, είναι ο αυτοοργανούμενος χάρτης (SOM). Όπως έχει αναφερθεί στο κεφάλαιο 3, ένας αυτοοργανούμενος χάρτης ανήκει στην κατηγορία νευρωνικών δικτύων μη επιβλεπόμενης μάθησης. Ωστόσο, χρησιμοποιώντας τη μεθοδολογία που περιγράφεται στην παραγράφου 3.6 (επιβλεπόμενο SOM), μπορεί να χρησιμοποιηθεί και για επιβλεπόμενη μάθηση. Για περισσότερες πληροφορίες για την θεωρία των αυτοοργανούμενων χαρτών (SOM) μπορεί κάποιος να ανατρέξει στο κεφάλαιο 3.

Παρακάτω, παρουσιάζεται τόσο η διαδικασία που ακολουθήθηκε προκειμένου να εκπαιδευτεί το νευρωνικό δίκτυο όσο και τα κριτήρια που εφαρμόστηκαν για την επιλογή της τοπολογίας και του ελάχιστου αριθμού ώστε η εκπαίδευση του αυτοοργανούμενου χάρτη να είναι επαρκής.

6.2 Πρότυπα Εισόδου Εκπαίδευσης

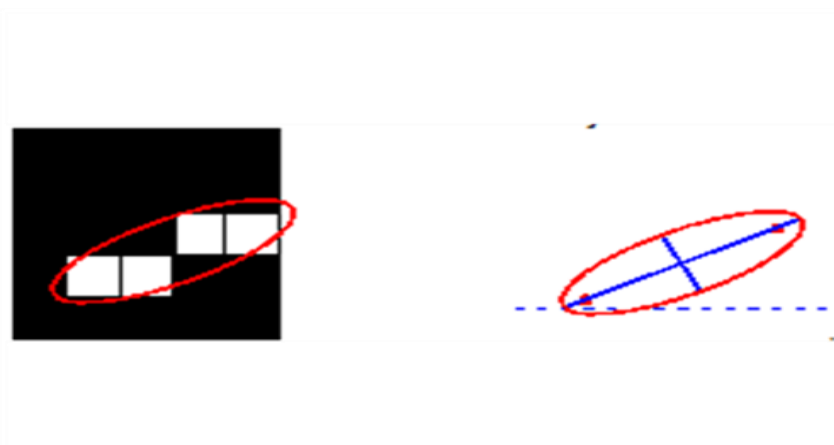
Ως δεδομένα εκπαίδευσης δίνονται δισδιάστατες εικόνες μαζί με ένα διάνυσμα τεσσάρων στοιχείων για κάθε εικόνα, που ορίζει τις επιτρεπτές λαβές για το σχήμα που απεικονίζεται. Με την χρήση αλγορίθμων επεξεργασίας εικόνας, εξάγουμε κρίσιμα χαρακτηριστικά του σχήματος της εικόνας, όπως αναλλοίωτες μεταβλητές, εκκεντρότητα και έτσι δημιουργούμε το *διάνυσμα χαρακτηριστικών*. Έπειτα, ενώνουμε το διάνυσμα χαρακτηριστικών με το διάνυσμα επιτρεπτών λαβών που δίνει ο εκπαιδευτής για κάθε σχήμα και έτσι αποκτούμε το τελικό διάνυσμα εκπαίδευσης που δίνεται ως είσοδος στο νευρωνικό μας δίκτυο (SOM) και ονομάζεται *πρότυπο εισόδου*.

Το διάνυσμα επιτρεπτών λαβών αποτελείται από τέσσερα στοιχεία που δηλώνουν τις επιτρεπτές γωνίες περιστροφής στις οποίες μπορεί να βρεθεί η αρπάγη του ρομποτικού βραχίονα ώστε να πιάσει το αντικείμενο. Έχουμε διακριτοποιήσει αυτές τις γωνίες σε 0, 45, 90 και 135 μοίρες από τον πρωτεύοντα άξονα του σχήματος. Ως *πρωτεύοντα άξονα* ορίζουμε τον μεγαλύτερο άξονα της περιγεγραμμένης έλλειψης του σχήματος ο οποίος ορίζει και τον προσανατολισμό του (βλέπε σχήμα 6.1)

Συνεπώς, αν το διάνυσμα επιτρεπτών λαβών έχει, για παράδειγμα, την μορφή $[0 \ 1 \ 1 \ 0]$, τότε για επιτρεπτή λαβή του αντικειμένου, πρέπει η ρομποτική αρπάγη να έχει περιστραφεί είτε 45 μοίρες, είτε 90 μοίρες ως προς τον πρωτεύοντα άξονα (άξονας προσανατολισμού). Εφόσον τα τέσσερα στοιχεία του διανύσματος επιτρεπτών λαβών δίνονται από τον εκπαιδευτή, δικαιολογείται και ο όρος «επιβλεπόμενη μάθηση» για το SOM.

Όπως προαναφέραμε, τα στοιχεία του διανύσματος χαρακτηριστικών εξάγονται από την δισδιάστατη εικόνα. Πιο συγκεκριμένα, τα δύο πρώτα είναι αναλλοίωτες

μεταβλητές ως προς τον αφινικό μετασχηματισμό (Affine Moment Invariants – AMIs) και χρησιμοποιούνται για να διαχωριστούν τα διάφορα σχήματα μεταξύ τους, χωρίς ωστόσο να επηρεάζεται αυτός ο διαχωρισμός από κάποιον αφινικό μετασχηματισμό που μπορεί να έχει υποστεί το σχήμα όπως μετατόπιση, κλιμάκωση ή περιστροφή. Τα υπόλοιπα δύο χαρακτηριστικά του διανύσματος εκπαίδευσης είναι ο λόγος του μεγαλύτερου προς τον μικρότερο άξονα της περιγεγραμμένης στο σχήμα έλλειψης (βλέπε σχήμα 6.1) και η εκκεντρότητα. Ο λόγος των αξόνων και η εκκεντρότητα χρησιμοποιούνται ώστε να διακρίνουμε ιδιόμορφες διατάξεις όμοιων σχημάτων.



Σχήμα 6.1: Πρωτεύον άξονας σχήματος. (Πηγή: Εγχειρίδιο χρήσης Matlab)

Για παράδειγμα, έστω ότι διαθέτουμε δύο εικόνες εισόδου (βλέπε Σχήμα 6.2). Στην πρώτη εικόνα απεικονίζεται ένα ορθογώνιο παραλληλόγραμμο που έχει μήκος πολύ μεγαλύτερο συγκριτικά με το πλάτος του, ενώ στην δεύτερη ένα τετράγωνο.



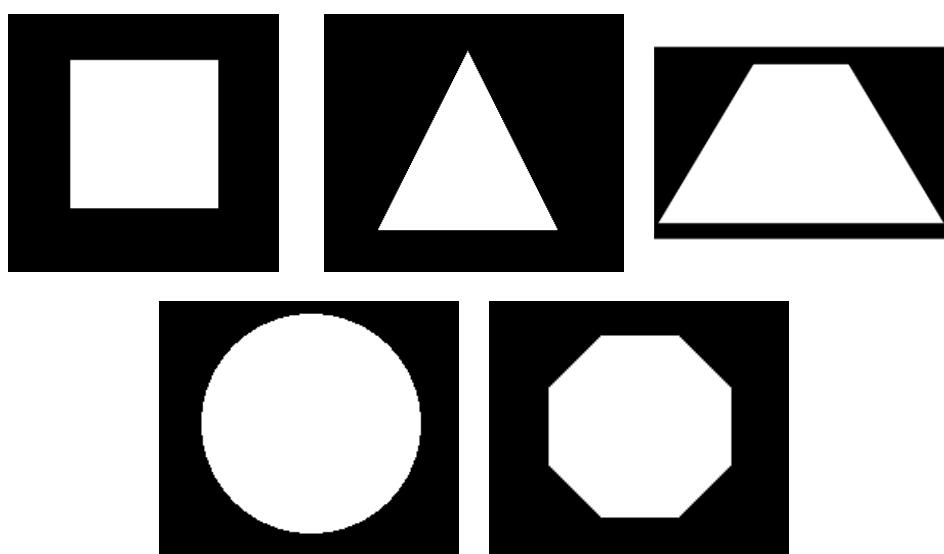
Σχήμα 6.2: Ορθογώνιο Παραλληλόγραμμο και τετράγωνο. Αν και το ένα σχήμα αποτελεί μια κλιμάκωση του άλλου, ωστόσο οι επιτρεπτές λαβές τους διαφέρουν

Οι αναλλοίωτες μεταβλητές AMIs θα έχουν την ίδια τιμή και για τις δύο εικόνες καθώς η μια αποτελεί κλιμάκωση της άλλης. Ωστόσο οι επιτρεπτές λαβές πρέπει να είναι διαφορετικές καθώς στην περίπτωση που έχουμε το τετράγωνο ως είσοδο, οι επιτρεπτές λαβές μπορούν να είναι τόσο η κάθετη όσο και η οριζόντια, ενώ στην περίπτωση που έχουμε το ορθογώνιο παραλληλόγραμμο για λόγους ευστάθειας της διάταξης ορίζουμε ως επιτρεπτή λαβή να είναι μόνο αυτή που γίνεται κατά μήκος του μικρότερου άξονα του αντικειμένου. Συνεπώς, χρησιμοποιούμε την *εκκεντρότητα* - που αποτελεί ένα μέτρο κατά πόσο το σχήμα μας αποκλίνει από το να είναι κυκλικό - και τον *λόγο του μεγαλύτερου προς τον μικρότερο άξονα της περιγεγραμμένης στο σχήμα έλλειψης* προκειμένου να διακρίνουμε περιπτώσεις όπως αυτές.

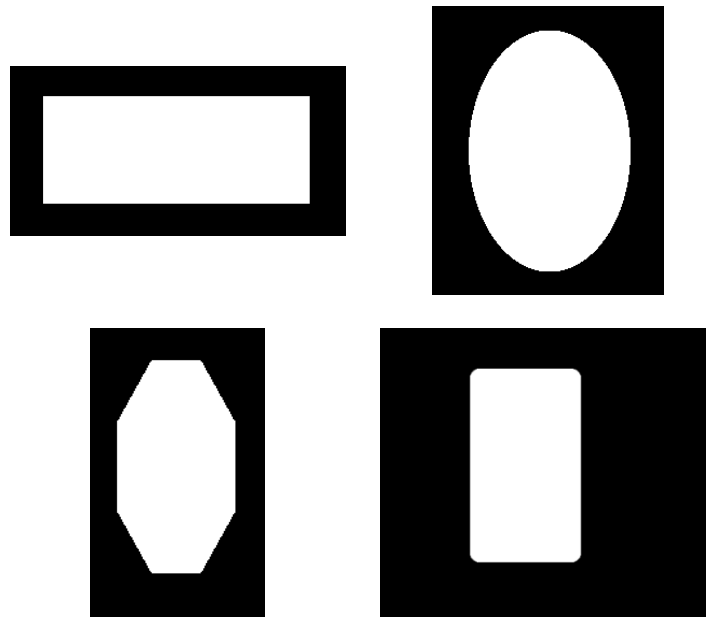
6.3 Σύνολο Δεδομένων Ελέγχου

Προκειμένου να ορίσουμε ένα μέτρο του κατά πόσο το νευρωνικό δίκτυο έχει εκπαιδευτεί επαρκώς, δημιουργούμε ένα σύνολο *δεδομένων ελέγχου* (*Test Set*) και εξετάζουμε τον βαθμό επιτυχίας του νευρωνικού δικτύου πάνω σ' αυτό το σύνολο δεδομένων.

Για να δημιουργήσουμε το σύνολο δεδομένων ελέγχου, υλοποιήσαμε πρόγραμμα στην γλώσσα προγραμματισμού OpenGL μέσω του οποίου αυτοματοποιήσαμε την διαδικασία σχεδίασης τόσο πρωταρχικών σχημάτων (βλέπε σχήμα 6.3) όσο και γενικεύσεων αυτών (βλέπε σχήμα 6.4). Για κάθε σχήμα που παράγαμε υπολογίζαμε το διάνυσμα χαρακτηριστικών του και δίναμε ως είσοδο το διάνυσμα επιτρεπτών λαβών του. Με αυτό τον τρόπο δημιουργήσαμε ένα σύνολο δεδομένων ελέγχου το οποίο περιείχε τετρακόσια ξεχωριστά σχήματα, αριθμό ικανοποιητικό για να θεωρηθεί αξιόπιστο.



Σχήμα 6.3: Βασικά ή Πρωταρχικά Σχήματα εκπαίδευσης



Σχήμα 6.4: Γενικεύσεις πρωταρχικών σχημάτων εκπαίδευσης

Με τον όρο γενικεύσεις σχημάτων, εννοούμε επίπεδα σχήματα τα οποία προκύπτουν από τα πρωταρχικά έχοντας υποστεί κάποιου είδους μετασχηματισμό ή αλλοίωση (βλέπε σχήμα 6.4). Για παράδειγμα μια γενίκευση ενός τετραγώνου θα μπορούσε να είναι ένα ορθογώνιο παραλληλόγραμμο, περιστραμμένο, το οποίο έχει υποστεί και κάποιου είδους ομαλοποίηση ώστε οι γωνίες του να μην είναι τόσο απότομες.

Ο λόγος που ενσωματώνουμε τέτοιου είδους παραμορφώσεις στο σύνολο ελέγχου είναι επειδή θέλουμε το σύστημα μας να ανταποκρίνεται ικανοποιητικά όχι μόνο σε ένα περιορισμένο αριθμό σχημάτων, αλλά σε πολλά και διαφορετικά μεταξύ τους σχήματα, δηλαδή, θέλουμε το σύστημα μας να έχει την ιδιότητα της *γενίκευσης*. Άλλωστε, αυτός είναι και ένας από του λόγους που χρησιμοποιούμε νευρωνικά δίκτυα καθώς είναι επιθυμητό με ένα περιορισμένο σύνολο δειγμάτων εκπαίδευσης το σύστημα να ανταποκρίνεται, ικανοποιητικά, σ' ένα υπερσύνολο στο οποίο θα υπάρχει και η τυχαία γενική περίπτωση. Σε αντίθετη περίπτωση, θα χρησιμοποιούσαμε μια τεράστια βάση δεδομένων όπου θα περιέχονται όλες οι δυνατές περιπτώσεις. Κάτι το οποίο είναι πρακτικά αδύνατο, καθώς δεν μπορούμε να καλύψουμε όλες τις δυνατές περιπτώσεις, ενώ ακόμα κι αν μπορούσαμε, η υπολογιστική ισχύς που θα χρειαζόμασταν θα ήταν μεγάλη.

6.4 Ποσοστό επιτυχίας νευρωνικού δικτύου

Ο εκπαιδευτής λοιπόν, όπως έχουμε προαναφέρει, παρέχει το σύνολο εκπαίδευσης του νευρωνικού δικτύου, το οποίο σύνολο αποτελείται από δύο μέρη: (α) τα σχήματα εκπαίδευσης και (β) το διάνυσμα επιτρεπτών λαβών για κάθε σχήμα εκπαί-

δευσης. Από τα σχήματα εκπαίδευσης χρησιμοποιώντας κατάλληλους αλγορίθμους επεξεργασίας εικόνας εξάγουμε τα αντίστοιχα διανύσματα χαρακτηριστικών τα οποία ενωνόμενα με τα διανύσματα επιτρεπτών λαβών δίνουν το τελικό σύνολο προτύπων εισόδου με το οποίο εκπαιδεύεται το νευρωνικό δίκτυο. Αφότου έχει εκπαιδευτεί το νευρωνικό δίκτυο χρησιμοποιούμε το σύνολο δεδομένων ελέγχου προκειμένου να διαπιστώσουμε εάν το νευρωνικό δίκτυο με το συγκεκριμένο σύνολο εκπαίδευσεως παρουσιάζει υψηλά ποσοστά επιτυχίας. Η διαδικασία που ακολουθείται για τον προσδιορισμό του ποσοστού επιτυχίας ενός εκπαιδευμένου νευρωνικού δικτύου περιγράφεται εκτενώς παρακάτω.

Συγκεκριμένα, γνωρίζοντας το διάνυσμα χαρακτηριστικών (δυο AMIs, λόγος αξόνων και εκκεντρότητα) κάθε δείγματος του συνόλου ελέγχου εντοπίζουμε ποιος νευρώνας του εκπαιδευμένου αυτοοργανούμενου χάρτη (SOM) ενεργοποιείται. Ως κριτήριο χρησιμοποιούμε την ελάχιστη απόσταση μεταξύ του διανύσματος χαρακτηριστικών και των τεσσάρων πρώτων στοιχείων του διανύσματος βαρών του κάθε νευρώνα.

Εφόσον βρούμε τον νικητή νευρώνα ελέγχουμε εάν οι τελευταίες τέσσερις διαστάσεις του διανύσματος βαρών του συμπίπτουν με τα στοιχεία του διανύσματος επιτρεπτών λαβών του δείγματος του συνόλου ελέγχου. Εφόσον αυτό συμβαίνει έχουμε επιτυχία για το συγκεκριμένο δείγμα του συνόλου ελέγχου, σε αντίθετη περίπτωση έχουμε αποτυχία. Οπότε το ποσοστό επιτυχίας προκύπτει από τον παρακάτω τύπο

$$\text{ποσοστό επιτυχίας} = \frac{\text{επιτυχώς αναγνωρισμένα δείγματα του συνόλου δεδομένων ελέγχου}}{\text{συνολικός αριθμός δεδομένων ελέγχου}}$$

6.5 Ορισμός κατώφλιων για την εύρεση του νικητή νευρώνα

Η διαδικασία της εύρεσης του νικητή νευρώνα μπορεί να φαντάζει απλή όπως περιγράφηκε παραπάνω, καθώς στηρίζεται στην επιλογή του νευρώνα που «απέχει» τη μικρότερη απόσταση από το διάνυσμα χαρακτηριστικών του προς αναγνώριση σχήματος. Ωστόσο, την παραπάνω διαδικασία πρέπει να την διέπουν κάποια κριτήρια. Συγκεκριμένα, πρέπει να ορίσουμε ένα κατώφλι για το ανώτατο όριο της απόστασης που μπορεί να έχει το διάνυσμα χαρακτηριστικών του προς αναγνώριση σχήματος από τον νικητή νευρώνα. Επιπλέον, εφόσον ικανοποιείται ο πρώτος περιορισμός, οπότε έχουμε νικητή νευρώνα, προκειμένου να προσδιορίσουμε το διάνυσμα επιτρεπτών λαβών πρέπει να ορίσουμε ένα δεύτερο κατώφλι πάνω από το οποίο μια λαβή θα θεωρείται ορθή. Παρακάτω, προσδιορίζουμε επακριβώς αυτά τα δύο κατώφλια.

Το πρώτο κατώφλι που χρησιμοποιούμε είναι για την αναγνώριση ενός καινούριου αντικειμένου που δεν έχει εισαχθεί στην φάση της εκπαίδευσης. Πιο συγκεκριμένα, από την στιγμή που ως κριτήριο για τον εντοπισμό του νευρώνα νικητή είναι μόνο η απόσταση, αυτό έχει ως αποτέλεσμα, αν δοθεί ως είσοδος στο σύστημα μας ένα καινούριο σχήμα που δεν υπήρχε στην φάση της εκπαίδευσης, αυτό να αναγνωριστεί λανθασμένα ακόμα κι αν η απόσταση του από τον κοντινότερο

νευρώνα είναι τόσο μεγάλη που να μην δικαιολογεί ότι βρίσκεται στην ίδια κλάση με τον νευρώνα νικητή. Γι αυτό τον λόγο εισάγουμε το συγκεκριμένο κατώφλι, το οποίο μας επιτρέπει να ορίσουμε έναν νευρώνα ως νικητή, μόνο αν η απόσταση του από το πρότυπο εισόδου είναι μικρότερη από το κατώφλι. Έτσι, εισάγεται στο σύστημα η έννοια την «αναγνώρισης νέου αντικειμένου» (new object decision), καθώς σε περίπτωση που εισαχθεί αντικείμενο για το οποίο το σύστημα δεν έχει εκπαιδευτεί τότε δεν επιχειρείται ο ορισμός επιτρεπτών λαβών.

Για να προσδιορίσουμε το συγκεκριμένο όριο αρκεί να υπολογίσουμε το εύρος της κλάσης που απαρτίζεται από σχήματα τραπεζίου. Ο λόγος που επιλέγουμε την συγκεκριμένη κλάση για τον υπολογισμό του κατωφλίου είναι επειδή μπορούν να προσδιοριστούν επακριβώς οι οριακές τιμές της οπότε να βρεθεί και το εύρος της. Όπως είναι γνωστό, όταν η μικρή βάση του τραπεζίου έχει μηδενικό μήκος τότε το τραπέζιο εκφυλίζεται σε τρίγωνο, ενώ όταν οι δύο βάσεις του τραπεζίου έχουν το ίδιο μήκος τότε το τραπέζιο εκφυλίζεται σε ορθογώνιο παραλληλόγραμμο (βλέπε σχήμα 6.5).



Σχήμα 6.5: Τραπέζιο και οριακές του διατάξεις που είναι το ορθογώνιο παραλληλόγραμμο και το τρίγωνο

Κρατώντας λοιπόν το μήκος της μεγάλης βάσης του τραπεζίου σταθερό και μεταβάλλοντας το μήκος της μικρής βάσης, μπορούμε να προσδιορίσουμε με συνεχή τρόπο τον χώρο που καταλαμβάνει η κλάση των σχημάτων τραπεζίου. Ως οριακές καταστάσεις έχουμε το τρίγωνο και το ορθογώνιο παραλληλόγραμμο που ανήκουν σε διαφορετικές κλάσεις. Εξάγουμε λοιπόν τα διανύσματα χαρακτηριστικών για τις οριακές καταστάσεις της κλάσης και υπολογίζουμε την μεταξύ τους απόσταση που μας δίνει το εύρος της κλάσης. Λαμβάνοντας υπόψη κάποιου είδους ανοχή για το πότε ένα σχήμα θεωρείται τραπέζιο (το μήκος της μικρής πλευράς να είναι ανάμεσα στο 10% και το 90% του μήκους της μεγάλης) υπολογίζουμε το τελικό εύρος της κλάσης. Θέτουμε, λοιπόν, ως τιμή κατωφλίου το μισό του εύρους της κλάσης θεωρώντας πως αν κάποιος νευρώνας βρεθεί στο κέντρο της κλάσης τότε οποιοδήποτε σχήμα τραπεζίου θα μπορεί να αναγνωρισθεί από αυτόν.

Το διάνυσμα χαρακτηριστικών όπως έχει προαναφερθεί αποτελείται από δύο αναλλοίωτες μεταβλητές AMIs, τον λόγο του μεγαλύτερου προς τον μικρότερο άξονα της περιγεγραμμένης έλλειψης στο σχήμα και την εκκεντρότητα είναι αναλλοίωτο ως

προς την περιστροφή, την κλιμάκωση και την μετατόπιση του σχήματος. Συνεπώς η τιμή που προκύπτει για το κατώφλι, είναι ανεξάρτητη από την επιλογή που έχει γίνει για το μήκος της μεγάλης πλευράς του τραpezίου στο συγκεκριμένο πείραμα και μπορεί να θεωρηθεί γενική. Εκτελώντας λοιπόν το προαναφερθέν πείραμα υπολογίζουμε την τιμή του κατωφλίου ίση με 8.

Το δεύτερο κατώφλι έχει να κάνει με το όριο πάνω από το οποίο μια λαβή θεωρείται ότι αναγνωρίζεται. Όπως είναι γνωστό, το διάνυσμα επιτρεπτών λαβών, για κάθε σχήμα εκπαίδευσης, που παρέχει ο εκπαιδευτής, περιέχει δεδομένα που είναι διακριτά. Δηλαδή αν μια λαβή είναι επιτρεπτή τότε το αντίστοιχο στοιχείο του διανύσματος επιτρεπτών λαβών είναι 1 ενώ όταν δεν είναι επιτρεπτή τότε είναι 0. Ωστόσο, κατά την αρχικοποίηση του διανύσματος βαρών του κάθε νευρώνα του αυτοοργανούμενου χάρτη οι τιμές των τελευταίων τεσσάρων στοιχείων του που δηλώνουν και τις επιτρεπτές λαβές κυμαίνονται στο διάστημα $[0, 1]$. Έτσι, κατά την διάρκεια της εκπαίδευσης και με την πάροδο των προτύπων εισόδου τα βάρη τελικά αρχίζουν να συγκλίνουν είτε προς το 0 είτε προς το 1 ανάλογα με το ποια πρότυπα εισόδου θα έλκουν τον εκάστοτε νευρώνα.

Προκειμένου να προσδιορίσουμε την τιμή του δεύτερου κατωφλίου θα χρησιμοποιήσουμε ξανά την κλάση των σχημάτων τραpezίου. Αυτή την φορά θα εκπαιδεύσουμε ένα μονοδιάστατο αυτοοργανούμενο χάρτη (SOM) αποτελούμενο από 20 νευρώνες. Το σύνολο εκπαίδευσης αποτελείται από τρία σχήματα, το ορθογώνιο παραλληλόγραμμα, το τραπέζιο και το τρίγωνο. Το τρίγωνο δεν έχει καμία επιτρεπτή λαβή οπότε το διάνυσμα επιτρεπτών λαβών του είναι της μορφής $[0\ 0\ 0\ 0]$, ενώ το ορθογώνιο παραλληλόγραμμα και το τραπέζιο έχουν διάνυσμα επιτρεπτών λαβών $[0\ 0\ 1\ 0]$ καθώς και στα δύο σχήματα η μόνη επιτρεπτή λαβή είναι αυτή που σχηματίζει γωνία με τον πρωτεύοντα άξονα του κάθε σχήματος ίση με 90° . Στον πίνακα 6.1, παρουσιάζεται το σύνολο εκπαίδευσης

	AMI	AMI	Ratio	Eccent	0°	45°	90°	135°
Τρίγωνο	36,55	57,72	1,56	0,77	0,00	0,00	0,00	0,00
Τραπέζιο	32,42	40,95	1,40	0,70	0,00	0,00	1,00	0,00
Ορθ. Παραλ.	27,41	26,38	1,80	0,83	0,00	0,00	1,00	0,00

Πίνακας 6.1: Σύνολο εκπαίδευσης αποτελούμενο μόνο από τρία σχήματα

Στον πίνακα 6.2 που ακολουθεί παρουσιάζονται τα διανύσματα βαρών των νευρώνων του αυτοοργανούμενου χάρτη μετά το πέρας της εκπαίδευσης. Από την εποπτεία του πίνακα 6.2 μπορούν να εξαχθούν δύο σημαντικά συμπεράσματα:

- Εάν δύο σχήματα έχουν το ίδιο διάνυσμα επιτρεπτών λαβών (τραπέζιο και ορθογώνιο παραλληλόγραμμα), οι νευρώνες που βρίσκονται στο ενδιάμεσο τους έχουν επίσης το ίδιο διάνυσμα επιτρεπτών λαβών (νευρώνες 1 έως 11). Το συμπέρασμα αυτό ήταν αναμενόμενο καθώς τα πρότυπα εισόδου που

έλκουν τους συγκεκριμένους νευρώνες έχουν το ίδιο διάνυσμα επιτρεπτών λαβών, με αποτέλεσμα τα βάρη των νευρώνων να συγκλίνουν σε κάθε επανάληψη του αλγορίθμου εκπαίδευσης όλο και πιο κοντά στο επιθυμητό διάνυσμα.

- Εάν δύο σχήματα έχουν διαφορετικό διάνυσμα επιτρεπτών λαβών ως προς μια διάσταση (τραπέζιο και τρίγωνο) τότε ο αριθμός των νευρώνων που έχουν τιμή στη διάσταση αυτή μικρότερη από 0.5 είναι ίσος με τον αριθμό των νευρώνων που έχουν τιμή μεγαλύτερη από 0.5 (νευρώνες 11 έως 20). Συνεπώς η τιμή 0.5 μπορεί να θεωρηθεί ως το διαχωριστικό επίπεδο μεταξύ των δυο κλάσεων. Εκτός αυτού παρατηρούμε ότι οι νευρώνες «απλώνονται» στο διανυσματικό χώρο με ομοιόμορφο τρόπο ώστε να καλύψουν ικανοποιητικά τον χώρο που βρίσκεται ενδιάμεσα των δειγμάτων εκπαίδευσης

Αρ. Νευρ.	AMI	AMI	Ratio	Eccent	0°	45°	90°	135°
1	27,41499	26,38014	1,8	0,831479	0	0	1	0
2	27,50473	26,64142	1,792784	0,829096	0	0	0,999999	0
3	27,59816	26,91345	1,78527	0,826614	0	0	0,999999	0
4	27,87639	27,72353	1,762893	0,819224	0	0	1	0
5	28,4039	29,2594	1,72047	0,805214	0	0	0,999998	0
6	29,80326	33,33372	1,60793	0,768047	0	0	0,999996	0
7	31,31188	37,72626	1,486614	0,727982	0	0	0,999972	0
8	31,89369	39,42086	1,439886	0,712551	0	0	0,999844	0
9	32,20806	40,33814	1,41481	0,704276	0	0	0,999426	0
10	32,32934	40,70059	1,406033	0,701408	0	0	0,997451	0
11	32,41961	40,95137	1,397513	0,698555	0	0	1	0
12	32,53186	41,41087	1,402334	0,700576	0	0	0,97197	0
13	32,70204	42,09826	1,408672	0,703289	0	0	0,931438	0
14	32,99071	43,2684	1,41986	0,708048	0	0	0,861804	0
15	33,73873	46,30277	1,449084	0,720466	0	0	0,680889	0
16	35,10808	51,85796	1,502618	0,743211	0	0	0,349629	0
17	35,91324	55,1243	1,534093	0,756584	0	0	0,154857	0
18	36,22934	56,40661	1,546445	0,761832	0	0	0,0784	0
19	36,41557	57,16205	1,55372	0,764923	0	0	0,03336	0
20	36,5535	57,72157	1,559109	0,767213	0	0	1,72E-08	0

Πίνακας 6.2: Διανύσματα βαρών νευρώνων μονοδιάστατου αυτοοργανούμενου χάρτη που έχει εκπαιδευτεί μόνο με τρία σχήματα

Επομένως, ορίζουμε ως τιμή για το δεύτερο κατώφλι το 0.5 με το σκεπτικό ότι εφόσον τα διανύσματα εκπαίδευσης έχουν επιλεγεί με τέτοιο τρόπο ώστε το διαχωριστικό επίπεδο να βρίσκεται στο μέσο τότε όσοι νευρώνες έχουν τιμή

μεγαλύτερη από 0.5 θα ανήκουν στη μια κλάση ενώ όσοι έχουν τιμή μικρότερη από 0.5 στην άλλη.

Όπως μπορεί να παρατηρηθεί από τον πίνακα 6.2, όσο μεγαλύτερος είναι ο αριθμός των νευρώνων τόσο αυτοί καλύπτουν καλύτερα τον διανυσματικό χώρο εισόδου. Οπότε, το πρώτο κατώφλι για την αναγνώριση νέου αντικειμένου είναι άμεσα εξαρτώμενο από τον αριθμό των νευρώνων. Για παράδειγμα, σε περίπτωση που υπάρχει μόνο ένας νευρώνας για να καλυφθεί ολόκληρη η κλάση των σχημάτων τραapeζίου τότε το πρώτο κατώφλι έχει τιμή 8. Αν όμως χρησιμοποιήσουμε περισσότερους νευρώνες τότε καλύπτουμε την κλάση με μικρότερη τιμή καθώς οι νευρώνες απλώνονται στον διανυσματικό χώρο. Οπότε ένα νέο σχήμα, όπως το τρίγωνο που προηγουμένως είχε απόσταση 9 από τον κεντρικό νευρώνα της κλάσης και κατατασσόταν ως άγνωστο, σε περίπτωση που χρησιμοποιηθούν περισσότεροι νευρώνες να έχει απόσταση 3 και να αναγνωρισθεί εσφαλμένα. Συνεπώς ο λόγος ύπαρξης του κατωφλίου αναγνώρισης νέου αντικειμένου είναι για αποτροπή εσφαλμένης αναγνώρισης περιπτώσεων όπου στην είσοδο δίνονται σχήματα τα οποία διαφέρουν κατά πολύ από τα σχήματα για τα οποία έχει εκπαιδευτεί το νευρωνικό δίκτυο άρα η απόσταση τους δεν επηρεάζεται σημαντικά από την αύξηση των νευρώνων.

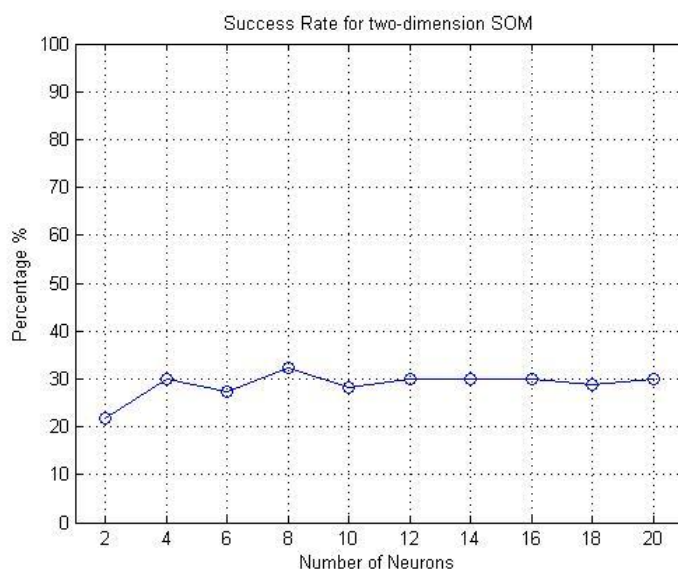
6.6 Δοκιμές Εκπαίδευσης Αυτοοργανούμενου Χάρτη

Το σύνολο ελέγχου μας παρέχει ένα μέτρο για το πόσο ικανοποιητικά εκπαιδευμένο είναι το σύστημα μας. Έτσι μπορούμε να βρούμε τον συνδυασμό - δειγμάτων εκπαίδευσης και τοπολογίας του αυτοοργανούμενου χάρτη - για τον οποίο το νευρωνικό δίκτυο παρουσιάζει υψηλό ποσοστό επιτυχίας.

Αρχικά, εκπαιδεύουμε το νευρωνικό δίκτυο χρησιμοποιώντας τον ελάχιστο αριθμό δειγμάτων, δηλαδή μόνο τα πρωταρχικά σχήματα που απεικονίζονται στο σχήμα 6.3 (τετράγωνο, κύκλος, οκτάγωνο, τραπέζιο και τρίγωνο). Συνεπώς, το σύνολο εκπαίδευσης αποτελείται μόνο από πέντε εικόνες. Χρησιμοποιούμε μονοδιάστατο SOM και κάθε φορά αλλάζουμε τον αριθμό των νευρώνων του. Για κάθε αλλαγή του αριθμού των νευρώνων, εκπαιδεύουμε ξανά το νευρωνικό μας και έπειτα βρίσκουμε τον βαθμό επιτυχίας της νέας διάταξης, χρησιμοποιώντας το σύνολο ελέγχου.

Στο σχήμα 6.6 που ακολουθεί, μπορούμε να δούμε τα αποτελέσματα της εφαρμογής της παραπάνω διαδικασίας. Όπως παρατηρούμε η μέγιστη τιμή του ποσοστού επιτυχίας είναι κοντά στο 30%. Ο λόγος για τον οποίο έχουμε ένα τόσο χαμηλό ποσοστό επιτυχίας είναι επειδή το σύνολο εκπαίδευσης δεν περιέχει τις γενικεύσεις των πρωταρχικών σχημάτων. Ακόμα, λοιπόν, και αν ο αυτοοργανούμενος χάρτης μας επιτύχει στο να κατανέμει στην σωστή κλάση τα σχήματα που δίνονται, αποτυγχάνει στο να αναγνωρίσει το σωστό σύνολο επιτρεπτών λαβών. Αυτό συμβαίνει επειδή το σύνολο επιτρεπτών λαβών διαφοροποιείται, ακόμα και για σχήματα που το ένα

αποτελεί απλά μια κλιμάκωση του άλλου (βλέπε τετράγωνο με ορθογώνιο παραλληλόγραμμο)



Σχήμα 6.6: Βαθμός επιτυχίας μονοδιάστατου SOM το οποίο έχει εκπαιδευτεί χρησιμοποιώντας μόνο τα τέσσερα πρωταρχικά σχήματα.

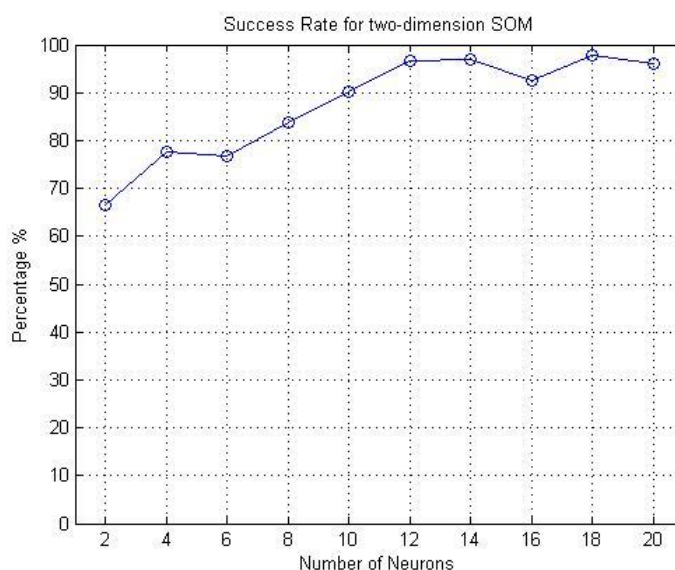
Επαναλαμβάνουμε το παραπάνω πείραμα, χρησιμοποιώντας πάλι μονοδιάστατο SOM. Ωστόσο, αυτή την φορά το σύνολο εκπαίδευσης, θα περιέχει αντιπροσωπευτικά δείγματα τόσο των πρωταρχικών σχημάτων όσο και των γενικεύσεων αυτών. Συγκεκριμένα, ορίζουμε ως αντιπροσωπευτικά εκείνα τα σχήματα η επιλογή των οποίων ως δειγμάτων εκπαίδευσης μας δίνει σωστό διαχωρισμό των διαφορών κλάσεων. Για παράδειγμα, έχουμε επιλέξει ότι ένα ορθογώνιο παραλληλόγραμμο που ο λόγος των πλευρών του είναι μικρότερος από 1.5 μπορεί να πιαστεί και από τις δύο πλευρές όπως το τετράγωνο, ενώ σε αντίθετη περίπτωση μπορεί να πιαστεί μόνο από την μικρότερη πλευρά. Συνεπώς προκειμένου να έχουμε σωστό διαχωρισμό επιλέγουμε ως αντιπροσωπευτικά δείγματα ένα τετράγωνο και ένα ορθογώνιο παραλληλόγραμμο με λόγο πλευρών 2. Αντίστοιχα, θεωρούμε ότι ένα τραπέζιο που η μικρή πλευρά του έχει μήκος μεγαλύτερο από 90% του μήκους της μεγάλης μπορεί να πιαστεί όπως το τετράγωνο, ενώ αν η μικρή πλευρά έχει μήκος μικρότερο από το 10% της μεγάλης δεν μπορεί να πιαστεί καθόλου. Σε κάθε άλλη περίπτωση το τραπέζιο έχει μόνο μια επιτρεπτή λαβή. Οπότε ως αντιπροσωπευτικά δείγματα για την κλάση των σχημάτων τραapeζίου είναι ένα τετράγωνο, ένα τραπέζιο με τη μικρή βάση στο 80% της μεγάλης, ένα τραπέζιο με την μικρή βάση στο 20% της μεγάλης και ένα τρίγωνο. Όπως προαναφέραμε στην παράγραφο 6.5 εφόσον τα δυο τραπέζια που χρησιμοποιούμε ως αντιπροσωπευτικά δείγματα έχουν το ίδιο διάνυσμα επιτρεπτών λαβών και οι νευρώνες που θα βρίσκονται ενδιάμεσα τους θα έχουν επίσης το ίδιο διάνυσμα επιτρεπτών λαβών, οπότε καλύπτεται ολόκληρη η κλάση

σχημάτων τραπεζίου που η μικρή βάση έχει μήκος από 20% έως 80% της μεγάλης. Με παρόμοιο σκεπτικό ορίζουμε τα αντιπροσωπευτικά διανύσματα για τον κύκλο και το οκτάγωνο, οπότε το συνολικό σύνολο δεδομένων εκπαίδευσης δίνεται στον πίνακα 6.3.

Αρ. Νευρ.	AMI	AMI	Ratio	Eccent	0°	45°	90°	135°
1	36,5535	57,72157	1,559109	0,767213	0	0	0	0
2	35,3699	52,14855	1,443953	0,721376	0	0	1	0
3	27,47836	26,54003	1,703933	0,809676	0	0	1	0
4	27,41009	26,36743	1	0	1	0	1	0
5	27,41214	26,37275	2	0,866025	0	0	1	0
6	25,00012	18,75028	1	0,00064	1	1	1	1
7	24,99994	18,74987	2,988999	0,864423	0	0	1	0
8	25,112	19,0055	1	0	1	1	1	1
9	25,11841	19,02017	2,015427	0,868224	0	0	1	0

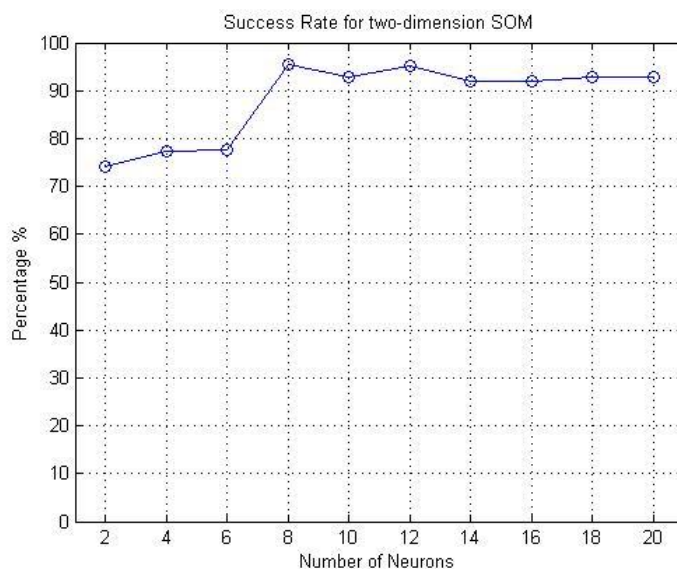
Πίνακας 6.3: Αντιπροσωπευτικά δείγματα των πρωταρχικών σχημάτων και των γενικεύσεων αυτών

Στο σχήμα 6.7 που ακολουθεί, μπορούμε να επιβεβαιώσουμε πλήρως τις προηγούμενες παρατηρήσεις μας. Χρησιμοποιώντας μόνο εννέα εικόνες καλύπτουμε σχεδόν όλες τις περιπτώσεις καθώς το ποσοστό επιτυχίας μας ξεπερνάει το 96%. Το φαινόμενο που περιγράψαμε προηγουμένως με την αύξηση των νευρώνων, που έχει ως αποτέλεσμα να καλύπτουμε επαρκέστερα το χώρο μεταξύ των δειγμάτων εκπαίδευσης φαίνεται χαρακτηριστικά στο σχήμα 6.7, καθώς το ποσοστό επιτυχίας του νευρωνικού μας συνεχώς αυξάνεται μέχρι και τους 14 νευρώνες.

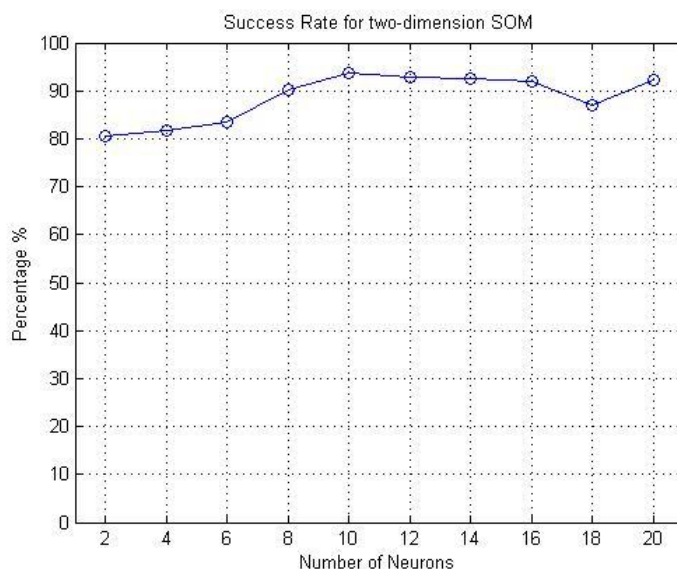


Σχήμα 6.7: Μονοδιάστατο SOM, χρησιμοποιώντας ως σύνολο εκπαίδευσης μόνο εννέα αντιπροσωπευτικά δείγματα.

Επαναλαμβάνουμε το παραπάνω πείραμα, χρησιμοποιώντας αυτή τη φορά ως τοπολογία δισδιάστατο SOM. Στο σχήμα 6.8 που ακολουθεί έχουμε χρησιμοποιήσει τοπολογία για τον αυτοοργανούμενο χάρτη της μορφής $[2, x]$ όπου x ο αριθμός των νευρώνων, ενώ στο σχήμα 6.9 χρησιμοποιούμε τοπολογία της μορφής $[4, x]$.



Σχήμα 6.8: Δισδιάστατο πλέγμα νευρώνων της μορφής $[2,x]$ όπου x ο αριθμός των νευρώνων. Σύνολο εκπαίδευσης αποτελούμενο από 9 εικόνες.

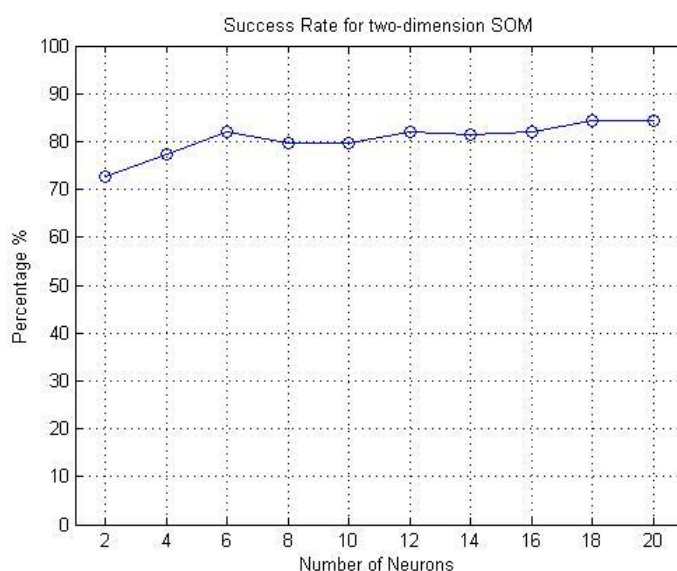


Σχήμα 6.9: Δισδιάστατο πλέγμα νευρώνων της μορφής $[4,x]$ όπου x ο αριθμός των νευρώνων. Σύνολο εκπαίδευσης αποτελούμενο από 9 εικόνες.

Αυτό που παρατηρούμε τόσο στο σχήμα 6.8 όσο και στο σχήμα 6.9, είναι ότι με το δισδιάστατο πλέγμα νευρώνων έχουμε πολύ υψηλά ποσοστά επιτυχίας. Πιο συγκεκριμένα, για δισδιάστατο πλέγμα της μορφής $[2, x]$, έχουμε ποσοστό επιτυχίας που ξεπερνάει το 96% για την διάταξη $[2,8]$. Ωστόσο, καθώς αυξάνεται πάρα πολύ ο αριθμός των νευρώνων, και η τοπολογία γίνεται πιο πυκνή, το ποσοστό επιτυχίας αρχίζει να πέφτει αν και πάλι διατηρείται σε υψηλά επίπεδα. Η διαπίστωση αυτή, μπορεί να εξαχθεί και εποπτικά από τα σχήματα 6.8 και 6.9, όπου το ποσοστό επιτυχίας, αρχικά αυξάνει μέχρι να φτάσει στην μέγιστη τιμή και έπειτα αρχίζει να ταλαντώνεται, μέχρις ότου ισορροπήσει για πολύ μεγάλο αριθμό νευρώνων.

Σύμφωνα με τα παραπάνω, λοιπόν, εξάγεται το συμπέρασμα ότι για την ικανοποιητική εκπαίδευση (ποσοστό επιτυχίας $> 90\%$) του αυτοοργανούμενου χάρτη μας, χρειαζόμαστε ελάχιστο αριθμό δειγμάτων εκπαίδευσης, τα οποία ωστόσο, πρέπει να είναι χαρακτηριστικοί αντιπρόσωποι της κλάσης τους. Αυτό, είναι εμφανές στα σχήματα 6.7, 6.8 και 6.9 όπου πετυχαίνουμε ποσοστά επιτυχίας κοντά στο 97%, χρησιμοποιώντας ως σύνολο εκπαίδευσης μόνο εννέα εικόνες.

Η σπουδαιότητα της επιλογής αντιπροσωπευτικών δειγμάτων για την εκπαίδευση του αυτοοργανούμενου χάρτη φαίνεται στα αποτελέσματα του παρακάτω πειράματος. Συγκεκριμένα, χρησιμοποιήσαμε μονοδιάστατο SOM το οποίο εκπαίδευσάμε με 20 συνολικά εικόνες επιλεγμένες τυχαία από ένα σύνολο με πρωταρχικά σχήματα και γενικεύσεις αυτών. Στο σύνολο εκπαίδευσης περιέχονται εικόνες για όλες τις δυνατές περιπτώσεις που μπορούν να προκύψουν ωστόσο δεν είναι προσεκτικά επιλεγμένες ώστε να θεωρούνται αντιπροσωπευτικές.



Σχήμα 6.10: Μονοδιάστατο SOM, χρησιμοποιώντας ως σύνολο εκπαίδευσης δεκαέξι εικόνες.

Όπως παρατηρούμε στο σχήμα 6.10 το ποσοστό επιτυχίας είναι σαφώς μειωμένο σε σχέση με την περίπτωση που χρησιμοποιούμε σύνολο εκπαίδευσης αποτελούμενο μόνο από εννέα σχήματα. Ο λόγος που συμβαίνει αυτό είναι επειδή τα δείγματα εκπαίδευσης δεν είναι αντιπροσωπευτικά. Συνεπώς δεν καθορίζονται σωστά οι διαχωριστικές επιφάνειες μεταξύ των διαφορετικών περιπτώσεων που μπορεί να προκύψουν. Οπότε όσα δεδομένα του συνόλου ελέγχου βρίσκονται κοντά στα όρια τα οποία έχουμε θέσει για αλλαγή του διανύσματος επιτρεπτών λαβών, όπως συμβαίνει στο τετράγωνο και το ορθογώνιο παραλληλόγραμμο, αποτυγχάνουν να αναγνωριστούν σωστά.

Σε όλα τα παραπάνω σχήματα 6.7 – 6.10 παρατηρούμε ότι η μορφή της διακύμανσης του ποσοστού επιτυχίας σε συνάρτηση με τον αριθμό των νευρώνων, είναι αύξουσα μέχρις ότου να φτάσει στο μέγιστό της για πρώτη φορά. Για οποιαδήποτε περαιτέρω αύξηση του αριθμού των νευρώνων, παρατηρούμε ότι το ποσοστό επιτυχίας αρχίζει να εμφανίζει τυχαίες διακυμάνσεις μικρής κλίμακας κοντά στην μέγιστη τιμή. Συνεπώς, αφού το σύστημα μας επιτύχει το μέγιστο ποσοστό επιτυχίας για συγκεκριμένο αριθμό νευρώνων, η οποιαδήποτε αύξηση του αριθμού των νευρώνων οδηγεί σε κορεσμό, χωρίς περαιτέρω βελτίωση των αποτελεσμάτων.

6.7 Συμπεράσματα και Αλγόριθμος Εκπαίδευσης

Από την παραπάνω πειραματική διερεύνηση, καταλήγουμε στα εξής συμπεράσματα για την διαδικασία που πρέπει να ακολουθηθεί ώστε να εκπαιδύσουμε τον αυτοοργανούμενο χάρτη του συστήματος αναγνώρισης επιτρεπτών ρομποτικών λαβών.

- Θέλουμε οι νευρώνες του αυτοοργανούμενου χάρτη να κατανέμονται με τέτοιο τρόπο στον χώρο καταστάσεων (οπτικών χαρακτηριστικών και επιτρεπτών λαβών) του συστήματος, ώστε να καλύπτουν και τις περιοχές που βρίσκονται ανάμεσα στις κλάσεις, επιτρέποντας τον επιτυχή εντοπισμό και σε γενικές περιπτώσεις.

Για τον λόγο αυτό, επιλέγουμε, η φάση της διάταξης να διαρκεί λίγες εποχές, περίπου 100, ενώ η φάση της σύγκλισης πολλές περισσότερες, περίπου 1000. Η συνάρτηση γειτονιάς μειώνεται εκθετικά κατά την φάση της διάταξης, ενώ η τιμή που έχει κατά την φάση της σύγκλισης είναι μικρή, περίπου 0.8, ώστε να επιλέγεται κυρίως ο νευρώνας νικητής και όχι οι γειτονικοί του.

- Η εκπαίδευση του συστήματος μας μπορεί να γίνει με τον ελάχιστο αριθμό δειγμάτων εκπαίδευσης, εφόσον τα δείγματα αυτά είναι αντιπροσωπευτικά όλων των δυνατών διατάξεων που μπορούν να προκύψουν. Με τον όρο διαφορετικές διατάξεις εννοούμε διαφορετικά σχήματα αλλά και διαφορετικά διανύσματα επιτρεπτών λαβών.

- Οι τιμές των κατώφλιων που χρησιμοποιούνται για την αναγνώριση, προσδιορίζονται σύμφωνα με το πείραμα που διεξαγάγαμε με τα σχήματα τραπεζίου και χρησιμοποιούνται ακολούθως γενικά σε όλα τα πειράματα, ανεξαρτήτως των δεδομένων εκπαίδευσης. Συγκεκριμένα, το κατώφλι που αναγνωρίζουμε τότε η λαβή είναι ορθή έχει τιμή 0.5, ενώ το κατώφλι για την αναγνώριση νέου αντικειμένου έχει τιμή 8.
- Εφόσον έχουμε ορίσει το σύνολο εκπαίδευσης, εκπαιδεύουμε τον αυτοοργανούμενο χάρτη (SOM), χρησιμοποιώντας είτε μονοδιάστατη είτε δισδιάστατη τοπολογία διάταξης των νευρώνων. Αυξάνουμε κάθε φορά τον αριθμό των νευρώνων που χρησιμοποιούμε μέχρις ότου το ποσοστό επιτυχίας να αρχίσει να διακυμαίνεται, οπότε βρισκόμαστε σε κορεσμό. Η τοπολογία για την οποία έχουμε το μέγιστο πρώτη φορά, αποτελεί και την βέλτιστη.

Εφαρμόζοντας λοιπόν τον παραπάνω αλγόριθμο για την συγκεκριμένη περίπτωση, αν επιλέξουμε να χρησιμοποιήσουμε μονοδιάστατο πλέγμα για τον αυτοοργανούμενο χάρτη, τότε η βέλτιστη τοπολογία είναι αυτή που χρησιμοποιεί 14 νευρώνες. Αντίθετα, αν χρησιμοποιήσουμε δισδιάστατο πλέγμα τότε η βέλτιστη τοπολογία είναι η [2,8]. Τόσο με την χρήση μονοδιάστατου όσο και δισδιάστατου αυτοοργανούμενου χάρτη το βέλτιστο ποσοστό επιτυχίας υπερβαίνει το 96%.

ΚΕΦΑΛΑΙΟ 7

7 Εντοπισμός επιτρεπτών λαβών τυχαίου σχήματος αντικειμένου

7.1 Εισαγωγή

Προτού οποιοδήποτε επιστημονικό πείραμα εκτελεστεί στον πραγματικό κόσμο, πρέπει να διέλθει από συγκεκριμένα στάδια ανάλυσης. Ο αριθμός των σταδίων αυτών, διαφοροποιείται ανάλογα με το είδος και την πολυπλοκότητα του πειράματος που θα εκτελεστεί, ωστόσο, μπορούν να κατηγοριοποιηθούν σε τρία κύρια μέρη.

- Το πρώτο μέρος, αποτελείται από την θεωρητική ανάλυση του προβλήματος. Στο μέρος αυτό παρουσιάζονται οι βασικές θεωρητικές σχέσεις στις οποίες βασιζόμαστε προκειμένου να υλοποιήσουμε το πείραμά μας.
- Το δεύτερο μέρος, είναι η υλοποίηση του πειράματος σε κάποιο περιβάλλον προσομοίωσης το οποίο θα μας παρέχει μια προσέγγιση για την συμπεριφορά του υπό εξέταση συστήματος στον πραγματικό κόσμο.
- Το τρίτο και σημαντικότερο μέρος, είναι η διεξαγωγή του πειράματος στον πραγματικό κόσμο.

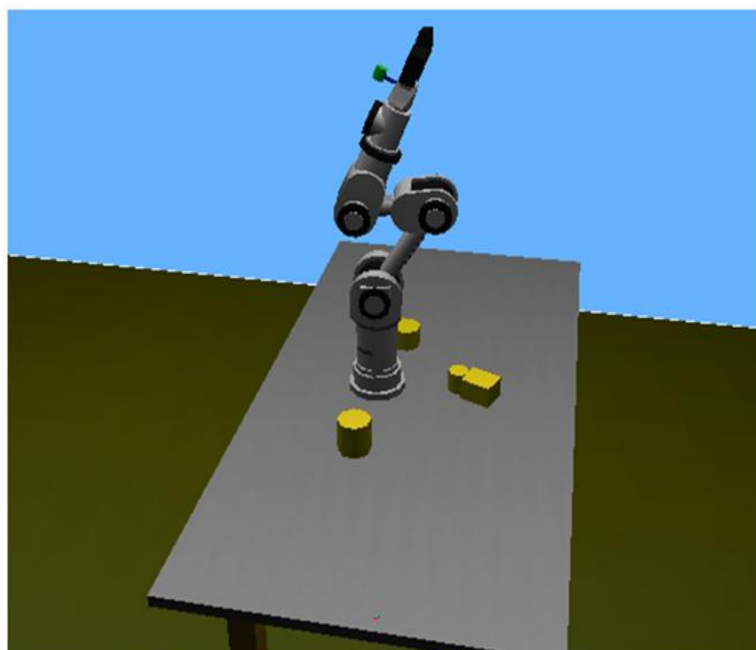
Στο παρόν κεφάλαιο, θα ασχοληθούμε με το δεύτερο μέρος της πειραματικής διαδικασίας και συγκεκριμένα με την υλοποίηση του πειράματος μας στο εικονικό περιβάλλον Webots. Για τον σκοπό αυτό θα χρησιμοποιηθεί ο ρομποτικός βραχίονας Katana 6M180 που αναλύεται εκτενώς παρακάτω.

Το πείραμα που διεξάγουμε έχει ως στόχο να εντοπίσει τις σωστές λαβές με τις οποίες μπορεί η αρπάγη ενός ρομποτικού βραχίονα να πιάσει ένα τυχαίου σχήματος αντικείμενο που έχει τοποθετηθεί πάνω σε οριζόντιο τραπέζι. Η υλοποίηση του πειράματος προϋποθέτει διάφορα στάδια επεξεργασίας όπως αυτά της κατάτμηση σχήματος, εύρεση βάθους κ.α. που θα αναφερθούν επιγραμματικώς παρακάτω καθώς η θεωρητική τους βάση έχει αναλυθεί εκτενώς σε προηγούμενα κεφάλαια.

Τέλος, θα παρουσιάσουμε αρκετά στιγμιότυπα της εκτέλεσης του κώδικα μας στο εικονικό περιβάλλον για διάφορα τυχαίου σχήματος αντικείμενα ώστε να γίνει εμφανής η ορθή λειτουργία του συστήματος μας.

7.2 Περιβάλλον Προσομοίωσης Webots

Το Webots είναι ένα πακέτο λογισμικού προσομοίωσης ρομπότ που αναπτύχθηκε από την εταιρεία Cyberbotics. Προσφέρει ένα γρήγορο περιβάλλον διαμόρφωσης πρωτοτύπων, το οποίο επιτρέπει στο χρήστη να δημιουργήσει τρισδιάστατους εικονικούς κόσμους και τα αντικείμενα που περιέχονται σ' αυτούς, με ιδιαίτερη ευκολία. Ο χρήστης μπορεί να προσθέσει ενεργά αντικείμενα όπως ρομποτικούς βραχίονες και κινούμενα ρομπότ. Σ' αυτά τα ενεργά αντικείμενα μπορούν να προσαρτηθούν επιπλέον συσκευές όπως κάμερες καθώς και διαφόρων τύπων αισθητήρες όπως κίνησης, απόστασης, δύναμης κ.α.



Σχήμα 7.1: Εικονικό τρισδιάστατο περιβάλλον, όπου διακρίνεται ο ρομποτικός βραχίονας Katana 6M180

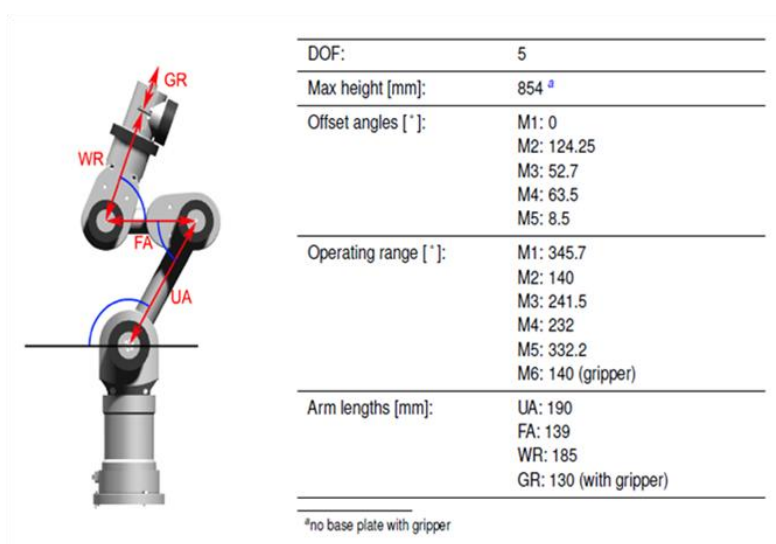
Συγκεκριμένα, με την ήδη έτοιμη βιβλιοθήκη που παρέχεται από το Webots καταφέραμε να προσομοιάσουμε πιστά την λειτουργία του ρομποτικού βραχίονα Katana 6M180 καθώς και το περιβάλλον που θα χρησιμοποιηθεί για την εκτέλεση του πειράματος (βλέπε σχήμα 7.1).

Η κίνηση των ρομποτικού βραχίονα γίνεται ακολουθώντας τους νόμους που διέπουν την κινηματική και την δυναμική των αντικειμένων καθώς μπορούμε να εισάγουμε στα αντικείμενα παραμέτρους όπως η μάζα, ο συντελεστής τριβής, ροπές αδράνειας κ.α.. Συνεπώς, η κίνηση του ρομποτικού βραχίονα προσεγγίζει σε ικανοποιητικό βαθμό την αντίστοιχη του πραγματικού κόσμου. Αυτός είναι ο κυριότερος λόγος που έχει καταστήσει το Webots ευρέως διαδεδομένο καθώς είναι τόσο ρεαλιστικό που μπορείς να ελέγξεις πραγματικά ρομπότ μέσω του προσομοιωτή.

Επιπλέον πλεονεκτήματα του εικονικού περιβάλλοντος Webots είναι ότι υποστηρίζει διάφορες γλώσσες προγραμματισμού όπως MATLAB, C++, Java, Python, ενώ ταυτόχρονα παρέχει διεπαφές με αρκετά πραγματικά ρομπότ, καθιστώντας τον κώδικα μας άμεσα εφαρμόσιμο στην πράξη.

7.3 Ρομποτικός Βραχίονας Katana 6M180

Ο ρομποτικός βραχίονας Katana 6M180 έχει δημιουργηθεί από την ελβετική εταιρεία Neuronics AG. Περιλαμβάνει πέντε περιστροφικές αρθρώσεις και την αρπάγη ως τελικό στοιχείο δράσης, η οποία εκτελεί κινήσεις ανοίγματος και κλεισίματος.



Σχήμα 7.2: Τεχνικά Χαρακτηριστικά του ρομποτικού βραχίονα Katana 6M180, όπως αυτά παρουσιάζονται στο εγχειρίδιο χρήσης της Neuronics AG.

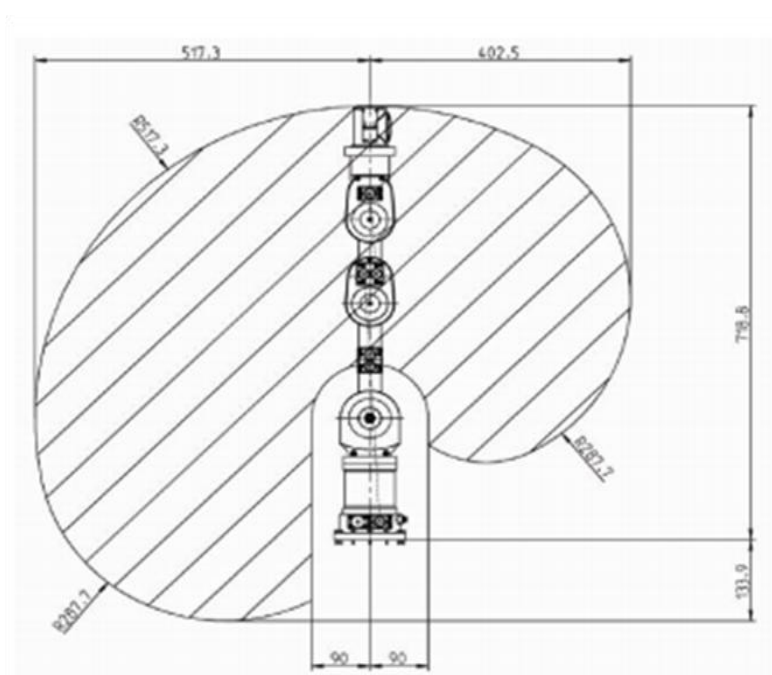
Στο σχήμα 7.2 που προηγείται, παρουσιάζονται τα τεχνικά χαρακτηριστικά του ρομπότ, όπως αυτά δίνονται από το εγχειρίδιο χρήσης που παρέχει η Neuronics AG. Συγκεκριμένα, δίνονται οι προεπιλεγμένες (offset) γωνίες των σερβοκινητήρων και το περιστροφικό τους εύρος, καθώς και τα μήκη κάθε βραχίονα.

Όπως είναι εμφανές, ο ρομποτικός βραχίονας Katana 6M180 διαθέτει πέντε βαθμούς ελευθερίας, οι οποίοι φαίνονται στο σχήμα 7.3 που ακολουθεί.



Σχήμα 7.3: Οι πέντε βαθμοί ελευθερίας του ρομποτικού βραχίονα Katana 6M180

Χώρος εργασίας (workspace) ονομάζεται η περιοχή που ορίζεται από το κέντρο του τελικού στοιχείου δράσης, όταν οι αρθρώσεις του ρομποτικού χειριστή μπορούν να εκτελέσουν όλες τις επιτρεπτές κινήσεις. Οπότε, ο χώρος εργασίας του Katana 6M180 όπως δίνεται στο εγχειρίδιο χρήσης, φαίνεται στο σχήμα 7.4



Σχήμα 7.4: Χώρος εργασίας του ρομποτικού βραχίονα Katana 6M180.

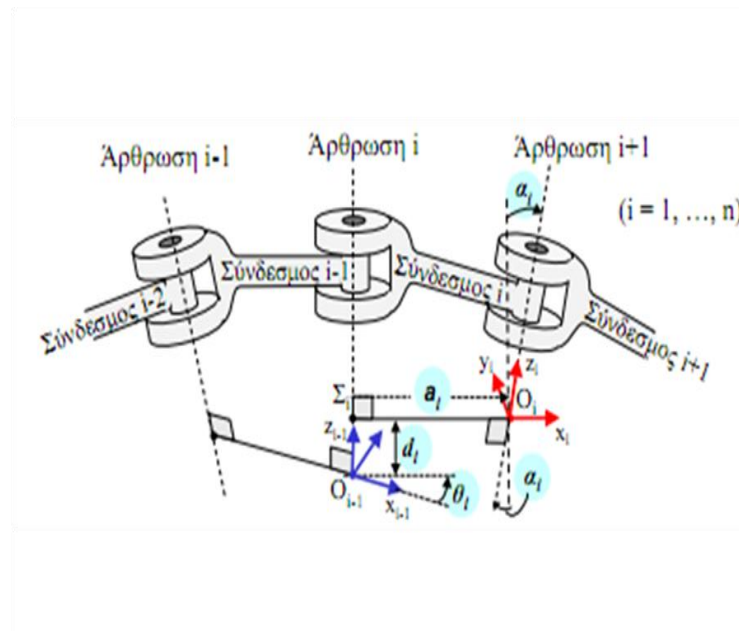
Όπως είναι γνωστό, αν ο ρομποτικός βραχίονας έχει λιγότερους από έξι βαθμούς ελευθερίας δεν μπορεί να πάρει οποιαδήποτε θέση και προσανατολισμό στο χώρο. Οπότε θέτουμε περιορισμούς για το περιβάλλον εργασίας, όπως, ότι τα αντικείμενα μας θα πρέπει να βρίσκονται σε θέσεις που μπορούν να επιτευχθούν από τον ρομποτικό χειριστή έχοντας συγκεκριμένο προσανατολισμό που επιθυμούμε.

Παρακάτω παρουσιάζουμε αναλυτικά την ορθή και ανάστροφη κινηματική ανάλυση για τον ρομποτικό χειριστή Katana 6M180, θέτοντας ωστόσο κάποιους περιορισμούς για την ανάστροφη κινηματική ανάλυση, καθώς διαθέτουμε μόνο πέντε βαθμούς ελευθερίας.

7.4 Ορθή Κινηματική Ανάλυση

7.4.1 Μέθοδος Denavit - Hartenberg

Για την ορθή κινηματική ανάλυση του ρομποτικού βραχίονα, μπορούμε να ακολουθήσουμε την μέθοδο Denavit – Hartenberg (D-H). Η συστηματική αυτή μέθοδος μας βοηθάει στο να ορίσουμε την σχετική θέση και προσανατολισμό μεταξύ δυο διαδοχικών συνδέσμων.



Σχήμα 7.5: Μέθοδος Denavit - Hartenberg στην ορθή κινηματική ανάλυση.

Βασικό στοιχείο της μεθόδου, αποτελεί ο σωστός ορισμός των πλαισίων κάθε άρθρωσης. Προκειμένου να γίνει αυτό, πρέπει να τηρηθούν οι παρακάτω τέσσερις κανόνες (Denavit & Hartenberg, 1964).

1. Επιλέγουμε τον άξονα z_i κατά μήκος του άξονα της άρθρωσης $i + 1$

2. Τοποθετούμε το κέντρο O_i στην τομή του άξονα z_i με την κοινή κάθετο των αξόνων z_{i-1} και z_i .
3. Επιλέγουμε τον άξονα x_i κατά μήκος της κοινής καθέτου των αξόνων z_{i-1} και z_i , με φορά από την άρθρωση i στην άρθρωση $i + 1$.
4. Επιλέγουμε τον άξονα y_i , ώστε να έχουμε δεξιόστροφο σύστημα συντεταγμένων για το πλαίσιο i .

Έχοντας ορίσει τα πλαίσια αναφοράς κάθε άρθρωσης, όπως φαίνεται στο σχήμα 7.5, απομένει πλέον να προσδιορίσουμε τις τέσσερις παραμέτρους της μεθόδου που είναι οι $\mathbf{d}_i, \theta_i, \mathbf{a}_i, \alpha_i$. Οι παράμετροι αυτοί ορίζονται ως εξής:

- \mathbf{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-1} και του άξονα x_i , γύρω από τον άξονα z_{i-1} κατά την δεξιόστροφη φορά.
- \mathbf{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 του άξονα z_{i-1} και του άξονα z_i γύρω από τον άξονα x_i κατά την δεξιόστροφη φορά.

Έχοντας ορίσει τα πλαίσια των αρθρώσεων και τις παραμέτρους τις μεθόδους D-H, ακολουθούμε τα παρακάτω βήματα, προκειμένου να ορίσουμε την σχετική θέση και προσανατολισμό 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 \mathbf{d}_i του πλαισίου $i - 1$ κατά μήκος του άξονα z_{i-1} .

Βήμα 3: Μετατόπιση \mathbf{a}_i (μήκος της κοινής καθέτου) κατά τον άξονα x_i

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

Από το Βήμα 1 και Βήμα 2, παίρνουμε την μήτρα μετασχηματισμού

$$T_{\Sigma_i}^{i-1} = \begin{bmatrix} \cos \theta_i & -\sin \theta_i & 0 & 0 \\ \sin \theta_i & \cos \theta_i & 0 & 0 \\ 0 & 0 & 1 & \mathbf{d}_i \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Ενώ εφαρμόζοντας το Βήμα 3 και Βήμα 4 έχουμε την μήτρα μετασχηματισμού

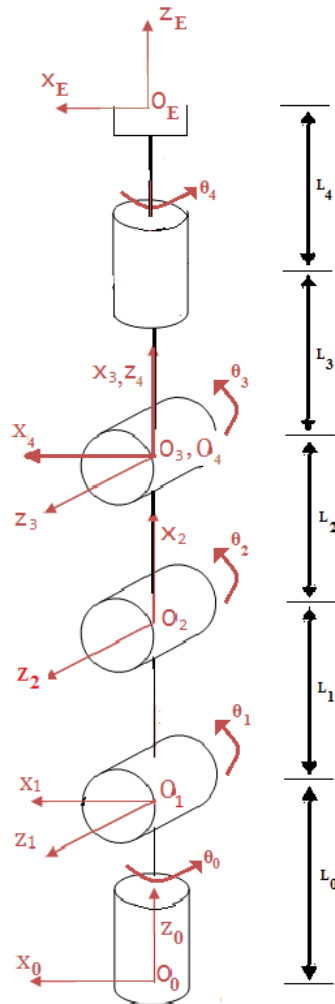
$$T_i^{\Sigma_i} = \begin{bmatrix} 1 & 0 & 0 & \mathbf{a}_i \\ 0 & \cos \alpha_i & -\sin \alpha_i & 0 \\ 0 & \sin \alpha_i & \cos \alpha_i & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Οπότε, συνολικά, η μήτρα μετασχηματισμού, που μας δίνει τη σχετική θέση και προσανατολισμό του πλαισίου i ως προς το πλαίσιο $i - 1$, είναι η ακόλουθη:

$$T_i^{i-1} = T_{\Sigma_i}^{i-1} \cdot T_i^{\Sigma_i} = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \cos a_i & \sin\theta_i \sin a_i & a_i \cos\theta_i \\ \sin\theta_i & \cos\theta_i \cos a_i & -\cos\theta_i \sin a_i & a_i \sin\theta_i \\ 0 & \sin a_i & \cos a_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.1)$$

7.4.2 Ρομποτικός Βραχίονας Katana 6M180

Για το στιγμιότυπο του ρομποτικού χειριστή που φαίνεται στο σχήμα 7.6, εφαρμόζουμε τη μέθοδο D-H για ορθή κινηματική ανάλυση.



Σχήμα 7.6: Σχηματική απεικόνιση του Katana 6M180, με καθορισμό των συντεταγμένων των πλαισίων των περιστροφικών αρθρώσεων και του τελικού στοιχείου δράσης.

Οι παράμετροι της μεθόδου D-H , για κάθε πλαίσιο αναφοράς δίνονται στον πίνακα που ακολουθεί

Πλαίσιο	θ_i	d_i	a_i	α_i
1	θ_0	L_0	0	$-\frac{\pi}{2}$
2	$-\frac{\pi}{2} + \theta_1$	0	L_1	0
3	θ_2	0	L_2	0
4	$\frac{\pi}{2} + \theta_3$	0	0	$\frac{\pi}{2}$
E	θ_4	$L_3 + L_4$	0	0

Πίνακας 7.1: Πίνακας παραμέτρων Denavit-Hartenberg για ορθή κινηματική ανάλυση του ρομποτικού βραχίονα Katana 6M180.

Συνεπώς, έχοντας τις παραμέτρους D-H για κάθε πλαίσιο εφαρμόζουμε την εξίσωση (7.1), προκειμένου να προσδιορίσουμε τους πίνακες μετασχηματισμού T_i^{i-1} .

$$T_1^0 = \begin{bmatrix} \cos \theta_0 & 0 & -\sin \theta_0 & 0 \\ \sin \theta_0 & 0 & \cos \theta_0 & 0 \\ 0 & -1 & 0 & L_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.2)$$

$$T_2^1 = \begin{bmatrix} \sin \theta_1 & \cos \theta_1 & 0 & L_1 \sin \theta_1 \\ -\cos \theta_1 & \sin \theta_1 & 0 & -L_1 \cos \theta_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.3)$$

$$T_3^2 = \begin{bmatrix} \cos \theta_2 & -\sin \theta_2 & 0 & L_2 \cos \theta_2 \\ \sin \theta_2 & \cos \theta_2 & 0 & L_2 \sin \theta_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.4)$$

$$T_4^3 = \begin{bmatrix} -\sin \theta_3 & 0 & \cos \theta_3 & 0 \\ \cos \theta_3 & 0 & \sin \theta_3 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.5)$$

$$T_E^4 = \begin{bmatrix} \cos \theta_4 & -\sin \theta_4 & 0 & 0 \\ \sin \theta_4 & \cos \theta_4 & 0 & 0 \\ 0 & 0 & 1 & L_3 + L_4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.6)$$

Έχοντας τους πίνακες μετασχηματισμού T_i^{i-1} από τις εξισώσεις 7.2 – 7.6, μπορούμε να υπολογίσουμε την συνολική μήτρα, εφαρμόζοντας την ακόλουθη εξίσωση

$$T_E^0 = T_1^0 \cdot T_2^1 \cdot T_3^2 \cdot T_4^3 \cdot T_E^4 \quad (7.7)$$

Η εξίσωση 7.7 μας δίνει την θέση και τον προσανατολισμό του τελικού στοιχείου δράσης ως προς το πλαίσιο της βάσης.

Προκειμένου να υπολογίσουμε αναλυτικά τον πίνακα που προκύπτει από την εξίσωση 7.7 χρησιμοποιήσαμε το προγραμματιστικό περιβάλλον MATLAB.

Οπότε η συνολική μήτρα μετασχηματισμού που δίνεται από την εξίσωση 7.7 είναι η εξής:

$$T_E^0 = \begin{bmatrix} c_{123}c_0c_4 - s_0s_4 & -c_4s_0 - c_{123}c_0s_4 & s_{123}c_0 & c_0(L_2s_{12} + (L_3 + L_4)s_{123} + L_1s_1) \\ c_0s_4 + c_{123}c_4s_0 & c_0c_4 - c_{123}s_0s_4 & s_{123}s_0 & s_0(L_2s_{12} + (L_3 + L_4)s_{123} + L_1s_1) \\ -s_{123}c_4 & s_{123}s_4 & c_{123} & L_0 + L_2c_{12} + (L_3 + L_4)c_{123} + L_1c_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.8)$$

όπου

$$c_{xyz} = \cos(\theta_x + \theta_y + \theta_z)$$

$$c_{xy} = \cos(\theta_x + \theta_y)$$

$$c_x = \cos(\theta_x)$$

$$s_{xyz} = \sin(\theta_x + \theta_y + \theta_z)$$

$$s_{xy} = \sin(\theta_x + \theta_y)$$

$$s_x = \sin(\theta_x)$$

7.5 Ανάστροφη Κινηματική Ανάλυση

Η επίλυση του προβλήματος της ανάστροφης κινηματικής ανάλυσης έγκειται στον εύρεση των τιμών των μεταβλητών $\theta_0, \theta_1, \theta_2, \theta_3, \theta_4$, έχοντας ως δεδομένη την θέση και τον προσανατολισμό της αρπάγης του ρομπότ.

Πραγματοποιώντας την επίλυση της ορθής κινηματικής ανάλυσης, προκύπτει μοναδική λύση για την θέση και τον προσανατολισμό της αρπάγης, που δίνεται από την εξίσωση 7.8. Όσον αφορά στην επίλυση της ανάστροφης κινηματικής ανάλυσης,

δεν είναι βέβαιο αν υπάρχει λύση, αλλά ακόμη κι αν υπάρχει, δεν είναι βέβαιο ότι αυτή είναι η μοναδική.

Για την επίλυση του προβλήματος της ανάστροφης κινηματικής ανάλυσης, θεωρούμε ότι μας δίνεται η θέση και ο προσανατολισμός του τελικού στελέχους του ρομποτικού βραχίονα, δηλαδή της αρπάγης του ρομπότ. Ο πίνακας που θα μας δίνεται θα είναι της μορφής

$$T_E^0 = \begin{bmatrix} x_{E,x} & y_{E,x} & z_{E,x} & p_{E,x} \\ x_{E,y} & y_{E,y} & z_{E,y} & p_{E,y} \\ x_{E,z} & y_{E,z} & z_{E,z} & p_{E,z} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Μέσω αυτού θα πρέπει να βρούμε τις παραμέτρους $\theta_0, \theta_1, \theta_2, \theta_3, \theta_4$ έτσι ώστε ο παραπάνω πίνακας να ισούται με τον τελικό πίνακα που προέκυψε από την επίλυση του προβλήματος της ορθής κινηματικής ανάλυσης.

Χρησιμοποιώντας λοιπόν τον πίνακα που δίνεται από την εξίσωση 7.8 εξάγονται οι παρακάτω σχέσεις για τις μεταβλητές των περιστροφικών αρθρώσεων.

- Αν $p_{E,x} \neq 0$

$$\tan \theta_0 = \frac{s_0}{c_0} = \frac{p_{E,y}}{p_{E,x}} \Rightarrow \theta_0 = \tan^{-1} \left(\frac{p_{E,y}}{p_{E,x}} \right)$$

- Αν $p_{E,x} = 0$

$$c_0 = 0 \Rightarrow \begin{cases} \theta_0 = \pi/2 \\ \theta_0 = -\pi/2 \end{cases}$$

Άρα συνολικά για την θ_0 έχουμε

$$\theta_0 = \begin{cases} \tan^{-1} \left(\frac{p_{E,y}}{p_{E,x}} \right) & , \text{αν } p_{E,x} \neq 0 \\ \pm \frac{\pi}{2} & , \text{αν } p_{E,x} = 0 \end{cases} \quad (7.9)$$

Έχοντας προσδιορίσει την γωνία θ_0 , μπορούμε πλέον να υπολογίσουμε και τις υπόλοιπες γωνίες από τις παρακάτω εξισώσεις

$$c_{123} = z_{E,z} \Rightarrow \theta_1 + \theta_2 + \theta_3 = \pm \cos^{-1}(z_{E,z}) \quad (7.10)$$

επίσης

$$L_2 s_{12} + (L_3 + L_4) s_{123} + L_1 s_1 = \begin{cases} \frac{p_{E,x}}{c_0} & , p_{E,x} \neq 0 \\ \frac{p_{E,y}}{s_0} & , p_{E,x} = 0 \end{cases} = a \quad , \text{όπου } a \text{ σταθερά}$$

Αφού μέσω της εξίσωσης 7.10 προσδιορίζεται η γωνία $\theta_{o\lambda} = \theta_1 + \theta_2 + \theta_3$, η παραπάνω σχέση μπορεί να γραφεί

$$L_1 s_1 + L_2 s_{12} = a - (L_3 + L_4) s_{123} = b \quad (7.11)$$

όπου b σταθερά

Επιπλέον, έχουμε ότι

$$p_{E,z} = L_0 + L_2 c_{12} + (L_3 + L_4) c_{123} + L_1 c_1 \Leftrightarrow$$

$$L_1 c_1 + L_2 c_{12} = p_{E,z} - L_0 - (L_3 + L_4) c_{123} = d \quad (7.12)$$

όπου d σταθερά

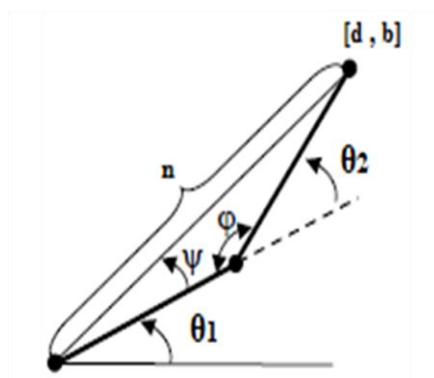
Οπότε λύνοντας το σύστημα που δημιουργείται από τις εξισώσεις 7.10, 7.11 και 7.12 υπολογίζουμε τις μεταβλητές $\theta_1, \theta_2, \theta_3$. Συγκεκριμένα για τη λύση του συστήματος έχουμε

$$\begin{aligned} b^2 + d^2 &= (L_1 s_1 + L_2 s_{12})^2 + (L_1 c_1 + L_2 c_{12})^2 \\ &= L_1^2 s_1^2 + L_2^2 s_{12}^2 + 2L_1 L_2 s_1 s_{12} + L_1^2 c_1^2 + L_2^2 c_{12}^2 + 2L_1 L_2 c_1 c_{12} \\ &= L_1^2 + L_2^2 + 2L_1 L_2 (c_1 c_{12} + s_1 s_{12}) = L_1^2 + L_2^2 + 2L_1 L_2 c_2 \Leftrightarrow \end{aligned}$$

$$c_2 = \frac{b^2 + d^2 - L_1^2 - L_2^2}{2L_1 L_2} \Leftrightarrow$$

$$\theta_2 = \pm \cos^{-1} \left(\frac{b^2 + d^2 - L_1^2 - L_2^2}{2L_1 L_2} \right) \quad (7.13)$$

Προκειμένου να υπολογίσουμε το θ_1 , θα χρησιμοποιήσουμε το γεωμετρικό ανάλογο των εξισώσεων 7.11 και 7.12 το οποίο είναι ένας επίπεδος ρομποτικός χειριστής δύο αρθρώσεων, όπως φαίνεται στο σχήμα 7.7



Σχήμα 7.7: Γεωμετρικό ανάλογο για το σύστημα εξισώσεων 7.10 και 7.11.

Οπότε η γεωμετρική λύση για το θ_1 θα προκύψει ως εξής

- Από νόμο ημιτόνων στο τρίγωνο

$$\frac{\sin \phi}{n} = \frac{\sin \psi}{L_2} \Rightarrow L_2 \sin(\pi - \theta_2) = n \sin \psi \Rightarrow \psi = \sin^{-1} \left(\frac{L_2 s_2}{n} \right)$$

$$\text{όπου } n = \sqrt{b^2 + d^2}$$

Επίσης

$$\tan(\theta_1 + \psi) = \frac{b}{d} \Rightarrow \theta_1 = \tan^{-1} \left(\frac{b}{d} \right) - \psi$$

$$\theta_1 = \tan^{-1} \left(\frac{b}{d} \right) - \sin^{-1} \left(\frac{L_2 s_2}{\sqrt{b^2 + d^2}} \right) \quad (7.14)$$

Μέσω της εξίσωσης 7.10 προσδιορίζουμε και την μεταβλητή θ_3 ως

$$\theta_3 = \pm \cos^{-1} z_{E,z} - \theta_1 - \theta_2 \quad (7.15)$$

Τέλος για τον προσδιορισμό της γωνίας θ_4 , έχουμε

- Αν $s_{123} \neq 0$

$$s_4 = \frac{y_{E,z}}{s_{123}} \Rightarrow \theta_4 = \sin^{-1} \left(\frac{y_{E,z}}{s_{123}} \right)$$

- Αν $s_{123} = 0$ και $c_{123} = 1$

$$c_0 c_4 - s_0 s_4 = x_{E,x} \Leftrightarrow$$

$$\cos(\theta_0 + \theta_4) = x_{E,x} \Leftrightarrow$$

$$\theta_4 = \pm \cos^{-1} x_{E,x} - \theta_0$$

- Αν $s_{123} = 0$ και $c_{123} = -1$

$$-c_0 c_4 - s_0 s_4 = x_{E,x} \Leftrightarrow$$

$$-\cos(\theta_4 - \theta_0) = x_{E,x} \Leftrightarrow$$

$$\theta_4 = \pm \cos^{-1} x_{E,x} + \theta_0$$

Οπότε για την μεταβλητή θ_4 έχουμε συνολικά

$$\theta_4 = \begin{cases} \sin^{-1} \left(\frac{y_{E,z}}{s_{123}} \right) & , \text{αν } s_{123} \neq 0 \\ \pm \cos^{-1} x_{E,x} - \theta_0 & , \text{αν } s_{123} = 0 \text{ και } c_{123} = 1 \\ \pm \cos^{-1} x_{E,x} + \theta_0 & , \text{αν } s_{123} = 0 \text{ και } c_{123} = -1 \end{cases} \quad (7.16)$$

Οι εξισώσεις 7.9, 7.13, 7.14, 7.15, 7.16 μας δίνουν όλες τις λύσεις του ανάστροφου κινηματικού προβλήματος. Ωστόσο, μόνο ένα υποσύνολο αυτών είναι αποδεκτές. Συγκεκριμένα, χρησιμοποιήσαμε μερικές από τις εξισώσεις του πίνακα της ορθής κινηματικής ανάλυσης προκειμένου να εξάγουμε τις μεταβλητές των περιστροφικών αρθρώσεων. Ωστόσο, προκειμένου η λύση μας να είναι ορθή, πρέπει να υπάρχει πλήρης αντιστοιχία μεταξύ του δοσμένου πίνακα και του πίνακα που προκύπτει από την ορθή κινηματική ανάλυση. Αυτός ο έλεγχος είναι απαραίτητος καθώς ο ρομποτικός βραχίονας Katana 6M180 έχει πέντε βαθμούς ελευθερίας, οπότε δεν μπορεί να επιτύχει οποιαδήποτε θέση και προσανατολισμό στον χώρο εργασίας. Συνεπώς, μπορεί το ανάστροφο κινηματικό πρόβλημα να μας δώσει λύση καθώς χρησιμοποιούμε ένα πεπερασμένο σύνολο εξισώσεων για την λύση του, ωστόσο, αυτή η λύση να είναι εκτός του επιτρεπτού χώρου εργασίας.

Επιπλέον, πρέπει να επισημάνουμε ότι οι περιστροφικές αρθρώσεις έχουν ένα συγκεκριμένο εύρος εντός του οποίου μπορούν να κινηθούν όπως περιγράφεται στα τεχνικά χαρακτηριστικά του ρομποτικού βραχίονα katana 6M180. Ειδικότερα, τα όρια για τις μεταβλητές των περιστροφικών αρθρώσεων έχουν ως εξής:

$$-172.5^\circ \leq \theta_0 \leq 173^\circ$$

$$-107^\circ \leq \theta_1 \leq 32.9^\circ$$

$$-127^\circ \leq \theta_2 \leq 114.8^\circ$$

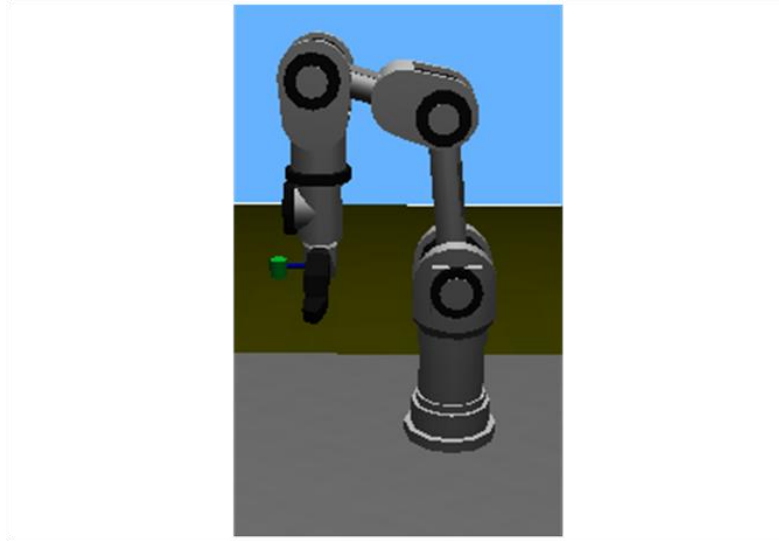
$$-116^\circ \leq \theta_3 \leq 116^\circ$$

$$-160^\circ \leq \theta_4 \leq 171.5^\circ$$

Οπότε, οι μεταβλητές των περιστροφικών αρθρώσεων που θα προκύψουν από την λύση του ανάστροφου κινηματικού προβλήματος θα πρέπει να βρίσκονται και εντός αυτών των ορίων.

Όπως προκύπτει άμεσα από τα παραπάνω, η εύρεση μιας τουλάχιστον επιτρεπτής λύσης για το ανάστροφο κινηματικό πρόβλημα απαιτεί πληθώρα ελέγχων συνθηκών. Ωστόσο, στο πρόβλημα που έχουμε να επιλύσουμε το οποίο είναι η εύρεση των βέλτιστων λαβών για αντικείμενα τυχαίας θέσης και προσανατολισμού πάνω σε επίπεδο τραπέζι, γνωρίζουμε εξ' αρχής ότι ο άξονας z_E της αρπάγης έχει διεύθυνση κάθετη προς την επιφάνεια του πάγκου εργασίας και φορά που δείχνει προς αυτόν, $z_E = [0 \ 0 \ -1]^T$. Οπότε, επιλύουμε το ανάστροφο κινηματικό πρόβλημα για την διάταξη που απεικονίζεται στο σχήμα 7.7. Αφού έχουμε συγκεκριμένη διάταξη για το ρομπότ το σύνολο δυνατών λύσεων που προκύπτει είναι σαφώς μικρότερο απ' ότι στην γενική περίπτωση. Συνεπώς, η συνολική μήτρα μετασχηματισμού, που δίνεται από την εξίσωση 7.8, στην περίπτωση αυτή, γίνεται

$$T_E^0 = \begin{bmatrix} -\cos(\theta_4 - \theta_0) & \sin(\theta_4 - \theta_0) & 0 & c_0(L_2s_{12} + L_1s_1) \\ \sin(\theta_4 - \theta_0) & \cos(\theta_4 - \theta_0) & 0 & s_0(L_2s_{12} + L_1s_1) \\ 0 & 0 & -1 & L_0 - (L_3 + L_4) + L_1c_1 + L_2c_{12} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.1)$$



Σχήμα 7.8: Προσανατολισμός της αρπάγης κάθετος προς την επιφάνεια του πάγκου εργασίας.

Οπότε, οι εξισώσεις που μας δίνουν τις μεταβλητές των περιστροφικών αρθρώσεων και είναι οι 7.9, 7.13, 7.14, 7.15, 7.16 μετασχηματίζονται ως ακολούθως

$$\theta_{ολ} = \theta_1 + \theta_2 + \theta_3 = \pi$$

Ισχύει πάντα ότι $p_{E,x} \neq 0$ καθώς δεν έχει νόημα μια τέτοια θέση για το τελικό στοιχείο δράσης, οπότε

$$\theta_0 = \tan^{-1} \left(\frac{p_{E,y}}{p_{E,x}} \right) \quad (7.17)$$

$$\theta_2 = \pm \cos^{-1} \left(\frac{b^2 + d^2 - L_1^2 - L_2^2}{2L_1L_2} \right) \quad (7.18)$$

$$\theta_1 = \tan^{-1} \left(\frac{b}{d} \right) - \sin^{-1} \left(\frac{L_2 s_2}{\sqrt{b^2 + d^2}} \right) \quad (7.19)$$

$$\theta_3 = \pi - \theta_1 - \theta_2 \quad (7.20)$$

$$\theta_4 = \pm \cos^{-1} x_{E,x} + \theta_0 \quad (7.21)$$

Οπότε από το σύνολο των λύσεων που προκύπτει για την συγκεκριμένη περίπτωση, επιλέγουμε αυτές που είναι επιτρεπτές ελέγχοντας μόνο αν ικανοποιούν τους περιορισμούς για το εύρος κίνησης των περιστροφικών αρθρώσεων.

7.6 Διαδικασία εντοπισμού επιτρεπτών λαβών

Για την υλοποίηση του συγκεκριμένου πειράματος χρησιμοποιήθηκε ο ρομποτικός βραχίονας Katana 6M180 σε περιβάλλον προσομοίωσης Webots. Στο έτοιμο πρότυπο του ρομποτικού βραχίονα το οποίο παρέχεται από το Webots, προσαρμόσαμε μια κάμερα κοντά στο τελικό στοιχείο δράσης, όπως φαίνεται στο σχήμα 7.8. Τόσο οι εσωτερικές, όσο και οι εξωτερικές παράμετροι της κάμερας θεωρούνται κατά την πειραματική αυτή διερεύνηση γνωστές οπότε δεν χρειάστηκε να εφαρμόσουμε κάποια τεχνική βαθμονόμησης για τον προσδιορισμό τους.

Η θέση της κάμερας είναι πάνω στον άξονα του τελικού στοιχείου δράσης, οπότε η μήτρα μετασχηματισμού που μας δίνει την θέση και τον προσανατολισμό αυτής ως προς την αρπάγη, είναι σταθερή και ανεξάρτητη των μεταβλητών περιστροφής των αρθρώσεων του ρομποτικού βραχίονα. Οπότε γνωρίζοντας την σταθερή μήτρα μετασχηματισμού T_c^E , η μήτρα μετασχηματισμού της κάμερα ως προς το πλαίσιο βάσης δίνεται από την παρακάτω εξίσωση.

$$T_c^0 = T_E^0 \cdot T_c^E \quad (7.22)$$

Η λύση για το ανάστροφο κινηματικό πρόβλημα της κάμερα, λοιπόν, είναι ακριβώς η ίδια με αυτή του ανάστροφου κινηματικού προβλήματος ως προς την αρπάγη καθώς ο πίνακας μετασχηματισμού T_c^E είναι ανεξάρτητος των μεταβλητών περιστροφής των αρθρώσεων του ρομποτικού βραχίονα. Οπότε από τις εξισώσεις 7.9, 7.13, 7.14, 7.15, 7.16 αλλάζουμε μόνο τις σταθερές b, d , ώστε να είναι συμβατές με την θέση της κάμερας. Στην περίπτωση του προβλήματος που εξετάζουμε χρησιμοποιούμε τις εξισώσεις 7.17-7.21 ελέγχοντας πάντα αν η λύση μας ικανοποιεί τους περιορισμούς για τις γωνίες των μεταβλητών των περιστροφικών αρθρώσεων.

Γνωρίζοντας, λοιπόν, τη λύση του ανάστροφου κινηματικού προβλήματος της κάμερας, μπορούμε πλέον να κάνουμε παράλληλη κίνηση του ρομποτικού βραχίονα ώστε να πάρουμε δυο διαφορετικές λήψεις του αντικείμενο για το οποίο θέλουμε να βρούμε τις επιτρεπτές λαβές. Έτσι χρησιμοποιώντας τις εξισώσεις 5.17-5.19 της παραγράφου 5.3.2, μπορούμε να εντοπίσουμε την θέση του κέντρου μάζας του αντικείμενου ως προς το πλαίσιο της κάμερας. Οι εξισώσεις 5.17-5.19 παρουσιάζονται και εδώ για ευκολία

$$p_x = \frac{X_1 b}{X_1 - X_2} \quad (7.23)$$

$$p_y = \frac{Y_1 b}{X_1 - X_2} = \frac{Y_2 b}{X_1 - X_2} \quad (7.24)$$

$$p_z = \frac{b}{X_1 - X_2} \quad (7.25)$$

όπου $p = [p_x, p_y, p_z]^T$ οι συντεταγμένες του κέντρου μάζας ως προς το πλαίσιο της κάμερας, $c_i = [X_i, Y_i, 1]^T$ οι κανονικοποιημένες συντεταγμένες του κέντρου μάζας ως προς το πλαίσιο της κάμερας και b η απόσταση μεταξύ των δύο παράλληλων λήψεων.

Εφόσον μας δίνονται ως δεδομένα συντεταγμένες εικονοστοιχείων αντί για κανονικοποιημένες συντεταγμένες, πραγματοποιούμε τον παρακάτω μετασχηματισμό και έπειτα εφαρμόζουμε τις σχέσεις (7.23)-(7.25)

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \Omega^{-1} \begin{bmatrix} X_I \\ Y_I \\ 1 \end{bmatrix}$$

Προϋπόθεση για την εκτέλεση του αλγορίθμου εύρεσης επιτρεπτών λαβών για τυχαίου σχήματος αντικείμενο που έχει τοποθετηθεί πάνω σε επίπεδη επιφάνεια, είναι να έχει προηγηθεί η εκπαίδευση του νευρωνικού δικτύου. Όπως, αναφέραμε εκτενώς στο κεφάλαιο 6, για την εκπαίδευση του συστήματος χρησιμοποιείται αυτο-οργανούμενος χάρτης (SOM) και η εκπαίδευση γίνεται με το ελάχιστο δυνατό σύνολο σχημάτων. Ο αλγόριθμος για την βέλτιστη εκπαίδευση του νευρωνικού δικτύου, έχει παρουσιαστεί εκτενώς στο κεφάλαιο 6 και συγκεκριμένα στην παράγραφο 6.6.

Αφού περιγράψαμε τις προϋποθέσεις για την εκτέλεση του αλγορίθμου εύρεσης των επιτρεπτών λαβών για τυχαίου σχήματος αντικείμενο που έχει τοποθετηθεί πάνω σε επίπεδη επιφάνεια, παρουσιάζουμε τα βήματα από τα οποία αποτελείται ο αλγόριθμός μας

- **Βήμα 1: Κατάτμηση εικόνας**

Η απεικόνιση ενός τρισδιάστατου αντικειμένου από την κάμερα που χρησιμοποιείται στο περιβάλλον προσομοίωσης, είναι μια έγχρωμη δισδιάστατη εικόνα. Ωστόσο, προκειμένου να εξαχθούν χρήσιμα χαρακτηριστικά για το σχήμα και τον προσανατολισμό του αντικειμένου, με χρήση κατάλληλων αλγορίθμων επεξεργασίας εικόνας, η έγχρωμη εικόνα μετατρέπεται σε ασπρόμαυρη. Η έγχρωμη εικόνα αρχικά μετατρέπεται σε γκριζα, που σημαίνει ότι οι διάφορες αποχρώσεις στον χώρο RGB μέσω ενός μετασχηματισμού παίρνουν κάποια τιμή στο σύνολο $[0, 256]$ όπου η τιμή 0 αντιστοιχεί στο μαύρο ενώ η τιμή 256 στο άσπρο. Έπειτα, προκειμένου να μετατρέψουμε την γκριζα εικόνα σε ασπρόμαυρη χρησιμοποιούμε

ένα κατώφλι ανάμεσα στο 0 και το 256. Όσα εικονοστοιχεία έχουν τιμή φωτεινότητας μεγαλύτερη από το κατώφλι παίρνουν την τιμή 1 ενώ τα υπόλοιπα την τιμή 0. Το σύνολο των εικονοστοιχείων που έχουν τιμή 1 ορίζουν και το σχήμα μας στην ασπρόμαυρη εικόνα, ενώ τα υπόλοιπα ορίζουν το φόντο.

Ο αλγόριθμος κατάτμησης εικόνας που περιγράφηκε παραπάνω μπορεί να χρησιμοποιηθεί μόνο σε περιπτώσεις όπου το χρώμα τόσο του αντικειμένου όσο και του φόντου είναι γνωστά. Σε περίπτωση που δεν υπάρχει τέτοιου είδους γνώση εκ' των προτέρων πρέπει να εφαρμοστούν πιο πολύπλοκοι και αξιόπιστοι αλγόριθμοι κατάτμησης εικόνας (Μαραγκός, 2005) (Shapiro & Stockman, 2001).

- **Βήμα 2: Κατάτμηση σχήματος**

Η μέθοδος για την κατάτμηση ενός σύνθετου σχήματος σε υποσχήματα γίνεται χρησιμοποιώντας τον κανόνα ελαχίστου. Σύμφωνα με τον κανόνα αυτόν κάθε σχήμα χωρίζεται σε υποσχήματα χρησιμοποιώντας τα αρνητικά τοπικά ελάχιστα της καμπυλότητας της τροχιάς του συνόρου του σχήματος, τα οποία ονομάζονται κρίσιμα σημεία. Η τροχιά αυτή πρέπει να έχει προκύψει κινούμενοι με τέτοια κατεύθυνση πάνω στο σύνορο του σχήματος, ώστε να έχουμε πάντα το εσωτερικό του στα αριστερά. Ο τρόπος που ενώνουμε αυτά τα κρίσιμα σημεία είναι τέτοιος ώστε η τομή που θα δημιουργηθεί να διασχίζει κάποιον τοπικό άξονα συμμετρίας του σχήματος. Αναλυτική παρουσίαση της μεθόδου έχει γίνει στο κεφάλαιο 4.

Με την κατάτμηση σχημάτων καταφέρνουμε να διασπάσουμε ένα πολύπλοκο σχήμα σε απλούστερα. Αυτά τα απλούστερα σχήματα, συνήθως, μπορούν να προσεγγιστούν από τα πρωταρχικά ή κάποια γενίκευση αυτών. Ωστόσο, υπάρχει η περίπτωση κατά την οποία το αρχικό μας σχήμα δεν χρειάζεται κατάτμηση καθώς μπορεί να είναι κάποιο πρωταρχικό σχήμα ή γενίκευση αυτού. Κριτήριο, λοιπόν, για το αν κάποιο σχήμα χρειάζεται διαχωρισμό αποτελεί η κυρτότητα. Αν ένα σχήμα είναι κυρτό τότε δεν χρειάζεται διαχωρισμό καθώς δεν υπάρχουν ούτε εσοχές ούτε εξογκώματα σ' αυτό. Στο κυρτό σχήμα, λοιπόν, δεν υπάρχουν αρνητικά τοπικά ελάχιστα της καμπυλότητας της τροχιάς του συνόρου του σχήματος.

Συνεπώς, κριτήριο για τον διαχωρισμό ενός σχήματος σε απλούστερα αποτελεί η ύπαρξη ή όχι κρίσιμων σημείων. Οπότε, στην περίπτωση που το σχήμα μας είναι κυρτό, το βήμα 2 θα μας δώσει ένα μοναδικό υποσχήμα που θα ταυτίζεται με το αρχικό σχήμα.

- **Βήμα 3: Εύρεση συντεταγμένων κέντρου μάζας και προσανατολισμού**

Επαναλαμβάνουμε τα βήματα 1 και 2, για μια δεύτερη λήψη του αντικειμένου με την κάμερα. Οι δύο λήψεις πρέπει να έχουν παράλληλα και στοιχισμένα επίπεδα. Οπότε αφού εντοπίσουμε τα ζευγάρια των όμοιων υποσχημάτων εντοπίζουμε τις συντεταγμένες σε εικονοστοιχεία των κέντρων μάζας τους. Οι συντεταγμένες του κέντρου μάζας για διδιάστατη εικόνα δίνονται από τους τύπους

$$centroid_{x,i} = \frac{\sum_{j=1}^N X_{j,i}}{N}, \quad centroid_{y,i} = \frac{\sum_{j=1}^N Y_{j,i}}{N}$$

Όπου $(X_{j,i}, Y_{j,i})$ οι συντεταγμένες των ενεργών εικονοστοιχείων της εικόνας i . Έπειτα εφαρμόζουμε τις σχέσεις 7.23 – 7.25 για $c_i = [centroid_{x,i}, centroid_{y,i}, 1]^T$ και προσδιορίζουμε τις συντεταγμένες του κέντρου μάζας κάθε εικόνας ως προς το πλαίσιο της κάμερας, άρα και ως προς το πλαίσιο της βάσης καθώς ο πίνακας T_c^0 έχει ήδη προσδιοριστεί.

Για τον προσδιορισμό του προσανατολισμού κάθε υποσχήματος, χρησιμοποιούμε την ήδη υλοποιημένη συνάρτηση `regionprops` του προγραμματιστικού περιβάλλοντος MATLAB. Η συνάρτηση επιστρέφει την γωνία μεταξύ του πρωτεύοντα άξονα και του x άξονα του επιπέδου της εικόνας. Ως πρωτεύον άξονας ορίζεται ο μεγαλύτερος άξονας της περιγεγραμμένης έλλειψης στο υποσχήμα.

- **Βήμα 4: Προσδιορισμός συνόλου λαβών**

Με το βήμα 3, διαχωρίζουμε ένα σχήμα σε υποσχήματα. Εάν δεν χρειάζεται διαχωρισμός τότε υπάρχει ένα μοναδικό υποσχήμα που συμπίπτει με το αρχικό σχήμα εισόδου. Για κάθε υποσχήμα πρέπει να βρεθεί το διάνυσμα των επιτρεπτών λαβών του. Για να επιτευχθεί αυτό ακολουθείται η παρακάτω διαδικασία.

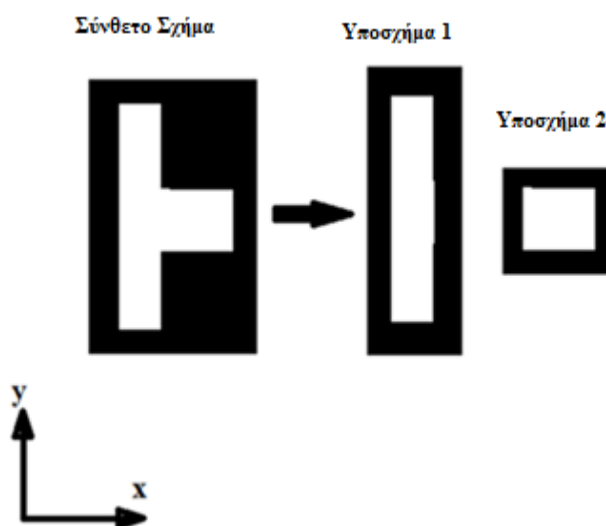
Αρχικά, εξάγουμε το διάνυσμα χαρακτηριστικών που αποτελείται από δυο AMIs, τον λόγο του μεγαλύτερου προς τον μικρότερο άξονα του σχήματος και την εκκεντρότητα. Βρίσκουμε την ευκλείδεια απόσταση του διανύσματος χαρακτηριστικών από κάθε νευρώνα του αυτοοργανούμενου χάρτη μας, συγκρίνοντας το διάνυσμα χαρακτηριστικών με τα πρώτα τέσσερα στοιχεία του διανύσματος βαρών των νευρώνων. Ορίζουμε ως νικητή νευρώνα, αυτόν που απέχει την μικρότερη απόσταση από το διάνυσμα χαρακτηριστικών, εφ' όσον αυτή η απόσταση είναι μικρότερη από ένα κατώφλι. Η τιμή του κατωφλίου είναι 8 και προσδιορίζεται επ' ακριβώς στην παράγραφο 6.5 του κεφαλαίου 6.

Έπειτα, προσδιορίζουμε το διάνυσμα επιτρεπτών λαβών για κάθε υποσχήμα με την χρήση ενός κατωφλίου και των τεσσάρων τελευταίων στοιχείων του διανύσματος βαρών των νευρώνων. Η τιμή του κατωφλίου ορίζεται 0.5 και προσδιορίζεται επ' ακριβώς επίσης στην παράγραφο 6.5 του κεφαλαίου 6. Για καθένα από τα τελευταία τέσσερα στοιχεία του διανύσματος βαρών των νευρώνων ελέγχουμε αν η τιμή του είναι μεγαλύτερη από το κατώφλι και σε περίπτωση που ισχύει θέτουμε τιμή 1 αλλιώς θέτουμε τιμή 0 στο διάνυσμα επιτρεπτών λαβών. Η τιμή 1 δείχνει ότι ο συγκεκριμένος προσανατολισμός της ρομποτικής λαβής είναι επιτρεπτός. Όπως έχουμε προαναφέρει στο κεφάλαιο 6, κάθε στοιχείο του διανύσματος επιτρεπτών λαβών δηλώνει μια γωνία περιστροφής ως προς τον πρωτεύοντα άξονα του αντικειμένου κινούμενοι αντιωρολογιακά γύρω από το κέντρο μάζας του. Οι γωνίες αυτές είναι 0, 45, 90 και 135 μοίρες. Ο πρωτεύοντας άξονας του αντικειμένου είναι αυτός που δηλώνει τον προσανατολισμό του. Για παράδειγμα, αν το διάνυσμα επιτρεπτών

λαβών είναι της μορφής $[0\ 1\ 0\ 1]$ σημαίνει ότι επιτρεπτές λαβές είναι αυτές για τις οποίες η αρπάγη του ρομπότ έχει περιστραφεί 45 ή 135 μοίρες ως προς τον πρωτεύοντα άξονα του αντικειμένου με φορά αντιωρολογιακή. Η περιστροφή γίνεται γύρω από το κέντρο μάζας του υποσχήματος.

- **Βήμα 5: Προσδιορισμός επιτρεπτών λαβών συνολικού σχήματος**

Στην περίπτωση που το αρχικό σχήμα δεν χρειάζεται διαχωρισμό τότε το βήμα 5 είναι περιττό καθώς τότε το σύνολο επιτρεπτών λαβών του υποσχήματος συμπίπτει με αυτό του συνολικού σχήματος. Ωστόσο στην περίπτωση που το σχήμα χρειάζεται διαχωρισμό σε υποσχήματα τότε πρέπει να ελέγξουμε κατά πόσο οι επιτρεπτές λαβές για κάθε υποσχήμα μπορούν να είναι επιτρεπτές λαβές και για το σύνθετο σχήμα το οποίο προκύπτει ως η ένωση όλων των επιμέρους υποσχημάτων (βλέπε σχήμα 7.9).



Σχήμα 7.9: Διαχωρισμούς σύνθετου σχήματος σε υποσχήματα. Η λαβή κατά την διεύθυνση x είναι επιτρεπτή για κάθε υποσχήμα ξεχωριστά, ωστόσο, αδύνατη για το συνολικό σχήμα.

Για να ελέγξουμε λοιπόν ποιες λαβές είναι επιτρεπτές στο αρχικό σύνθετο σχήμα ακολουθήσαμε την εξής διαδικασία. Αρχικά, γνωρίζοντας τις διαστάσεις της αρπάγης του ρομποτικού βραχίονα καθώς και το βάθος στο οποίο βρίσκεται το αντικείμενο, υπολογίζουμε το χώρο που θα καταλάμβανε η αρπάγη στην εικόνα αν βρισκόταν στο ίδιο βάθος με το αντικείμενο αυτό. Έπειτα από την εικόνα αφαιρούμε μόνο το υποσχήμα του οποίου τις επιτρεπτές λαβές εξετάζουμε δημιουργώντας με αυτό τον τρόπο το *ελλίπες σχήμα*. Για το υποσχήμα που αφαιρέσαμε γνωρίζουμε τον ακριβή προσανατολισμό για κάθε μια από τις επιτρεπτές λαβές του (Βήμα 4) καθώς και τις συντεταγμένες του κέντρου μάζας του πάνω στην εικόνα. Έτσι μπορούμε να

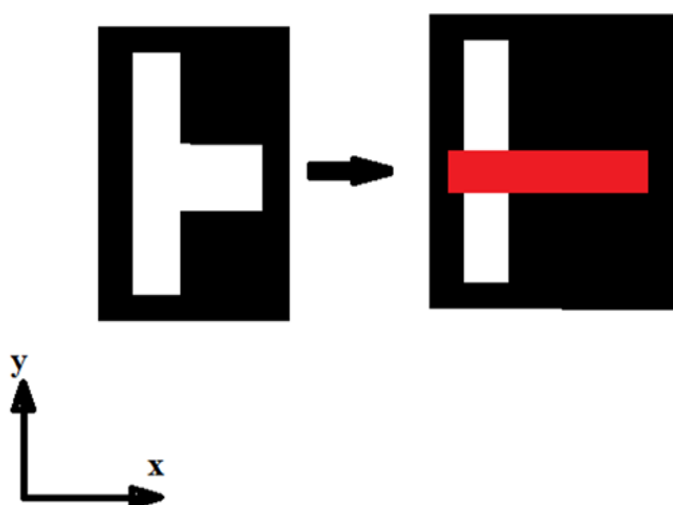
αναπαράστησουμε το σχήμα της αρπάγης στην εικόνα για κάθε επιτρεπτή λαβή του υποσχήματος (βλέπε σχήματα 7.10 και 7.11)

Εφόσον η αναπαράσταση της αρπάγης στην εικόνα για κάποια επιτρεπτή λαβή του υποσχήματος τέμνει το ελλειπές σχήμα τότε η συγκεκριμένη λαβή δεν είναι επιτρεπτή για το αρχικό σύνθετο σχήμα. Ο λόγος που συμβαίνει αυτό είναι ότι σε περίπτωση που η αρπάγη επιχειρούσε να πιάσει το σύνθετο αντικείμενο με την συγκεκριμένη λαβή τότε αυτή θα προσέκρουε σε κάποιο μέρος του αντικειμένου. Εφαρμόζοντας, λοιπόν, την παραπάνω μέθοδο για το σύνολο των υποσχημάτων προσδιορίζουμε το τελικό σύνολο επιτρεπτών λαβών του αρχικού σύνθετου σχήματος.

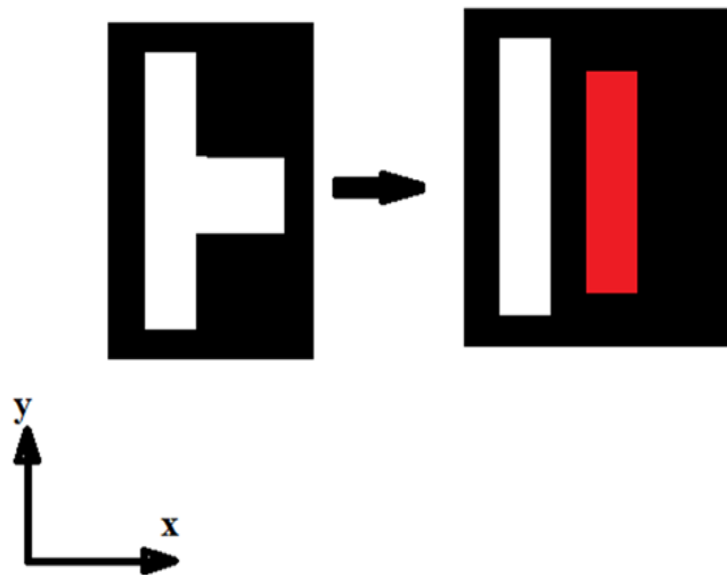
Η εφαρμογή της μεθόδου είναι εμφανής στα σχήματα 7.10 και 7.11 που ακολουθούν όπου εξετάζουμε κατά πόσον οι επιτρεπτές λαβές του υποσχήματος 2 (βλέπε σχήμα 7.9) είναι επιτρεπτές και στο αρχικό σύνθετο σχήμα. Το υποσχήμα 2 έχει συνολικά δύο επιτρεπτές λαβές, μια στη διεύθυνση του άξονα x και μια στη διεύθυνση του άξονα y . Στο σχήμα 7.10, παρατηρούμε ότι η λαβή που έχει τη διεύθυνση του άξονα x δεν είναι επιτρεπτή στο σύνθετο σχήμα καθώς η αναπαράσταση της αρπάγης για την συγκεκριμένη λαβή γύρω από το κέντρο μάζας του υποσχήματος 2, τέμνει το ελλειπές σχήμα. Αντίθετα, στο σχήμα 7.11, η λαβή είναι επιτρεπτή τόσο για το υποσχήμα, όσο και για το συνολικό σχήμα.

Επαναλαμβάνοντας την παραπάνω διαδικασία και για το υποσχήμα 1 προσδιορίζουμε το τελικό σύνολο επιτρεπτών λαβών του αρχικού σύνθετου σχήματος. Συγκεκριμένα, το αρχικό σύνθετο σχήμα έχει μόνο δυο επιτρεπτές λαβές που έχουν την ίδια διεύθυνση y αλλά διαφορετική θέση καθώς η μια γίνεται γύρω από το κέντρο μάζας του υποσχήματος 1 ενώ η άλλη γύρω από το κέντρο μάζας του υποσχήματος 2.

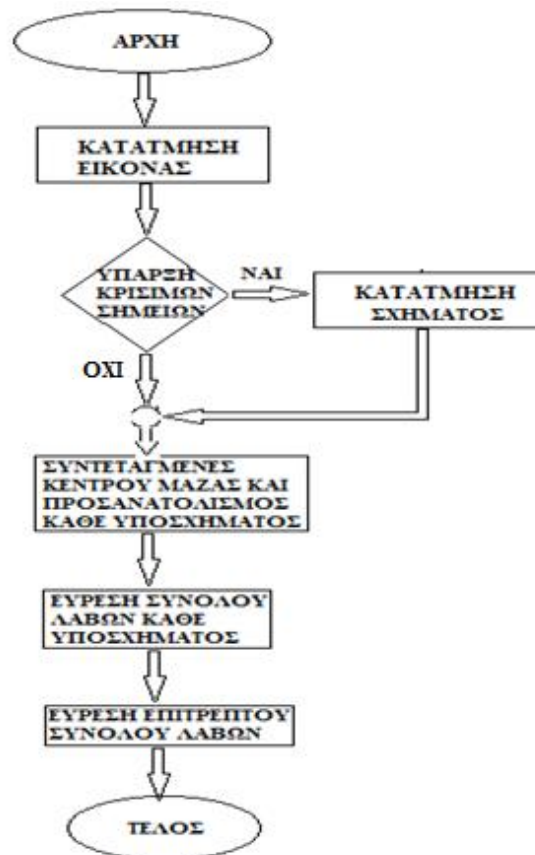
Στο σχήμα 7.12 παρουσιάζουμε το λογικό διάγραμμα του αλγορίθμου εύρεσης επιτρεπτών λαβών



Σχήμα 7.10: Μη επιτρεπτή λαβή καθώς υπάρχει τομή μεταξύ της αναπαράστασης της αρπάγης και του ελλειπούς σχήματος



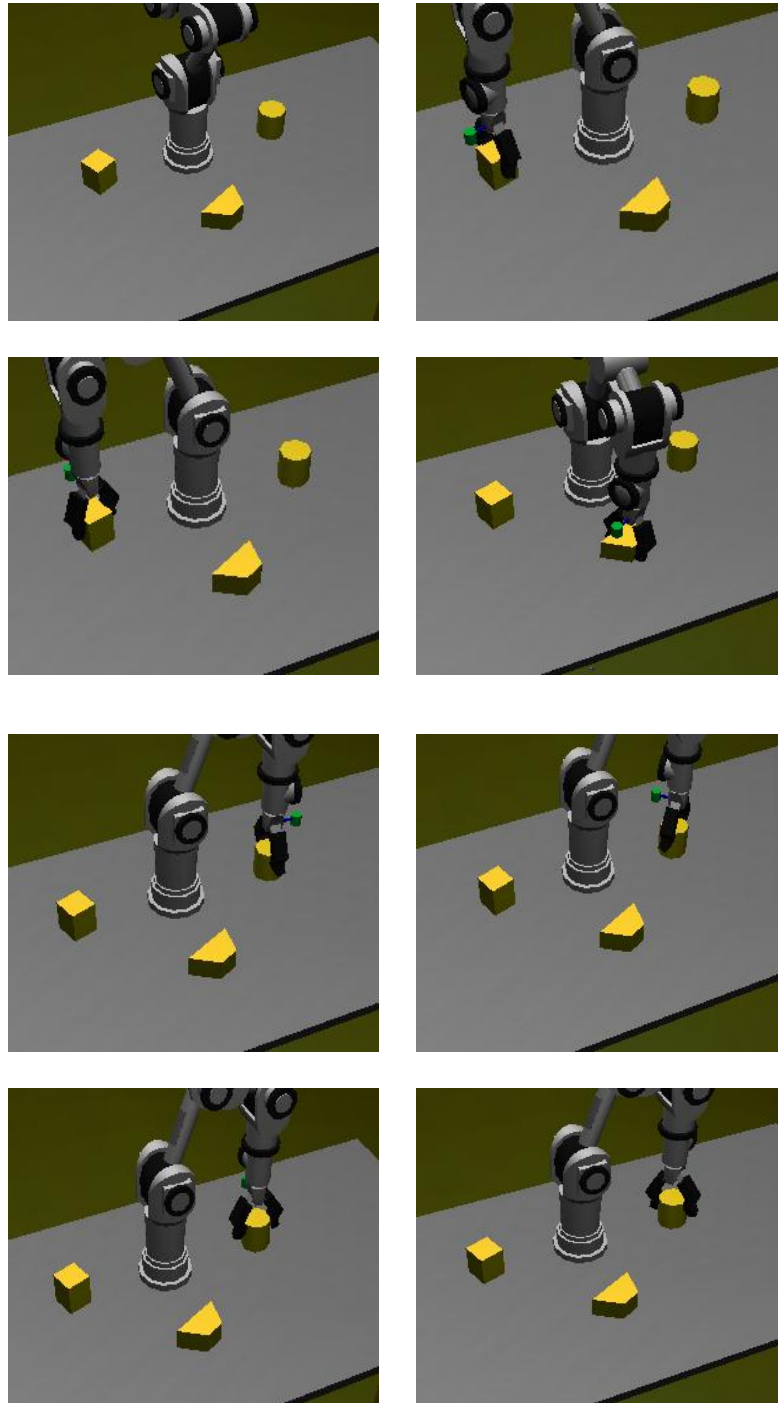
Σχήμα 7.11: Η λαβή για τον άξονα y είναι επιτρεπτή καθώς δεν υπάρχει κάποια τομή μεταξύ της αναπαράστασης της αρπάγης και του ελλιπούς σχήματος. Οπότε δεν υφίστανται γεωμετρικοί περιορισμοί που να εμποδίζουν την λαβή από το να πραγματοποιηθεί.



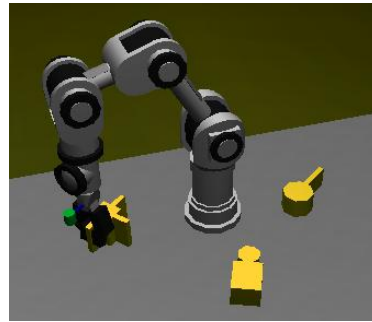
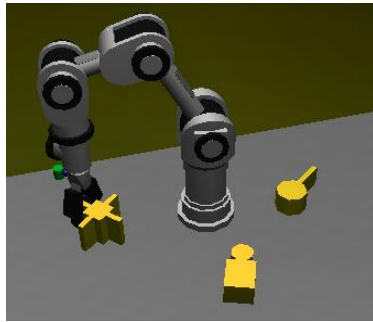
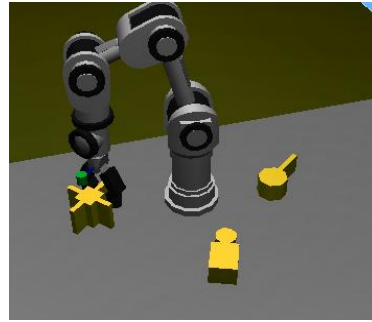
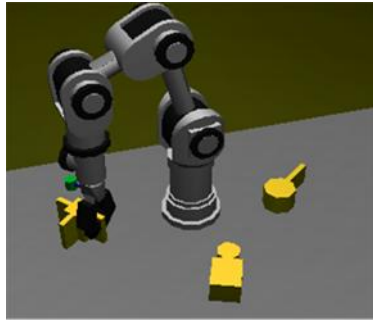
Σχήμα 7.12: Λογικό διάγραμμα αλγορίθμου εύρεσης επιτρεπτών λαβών

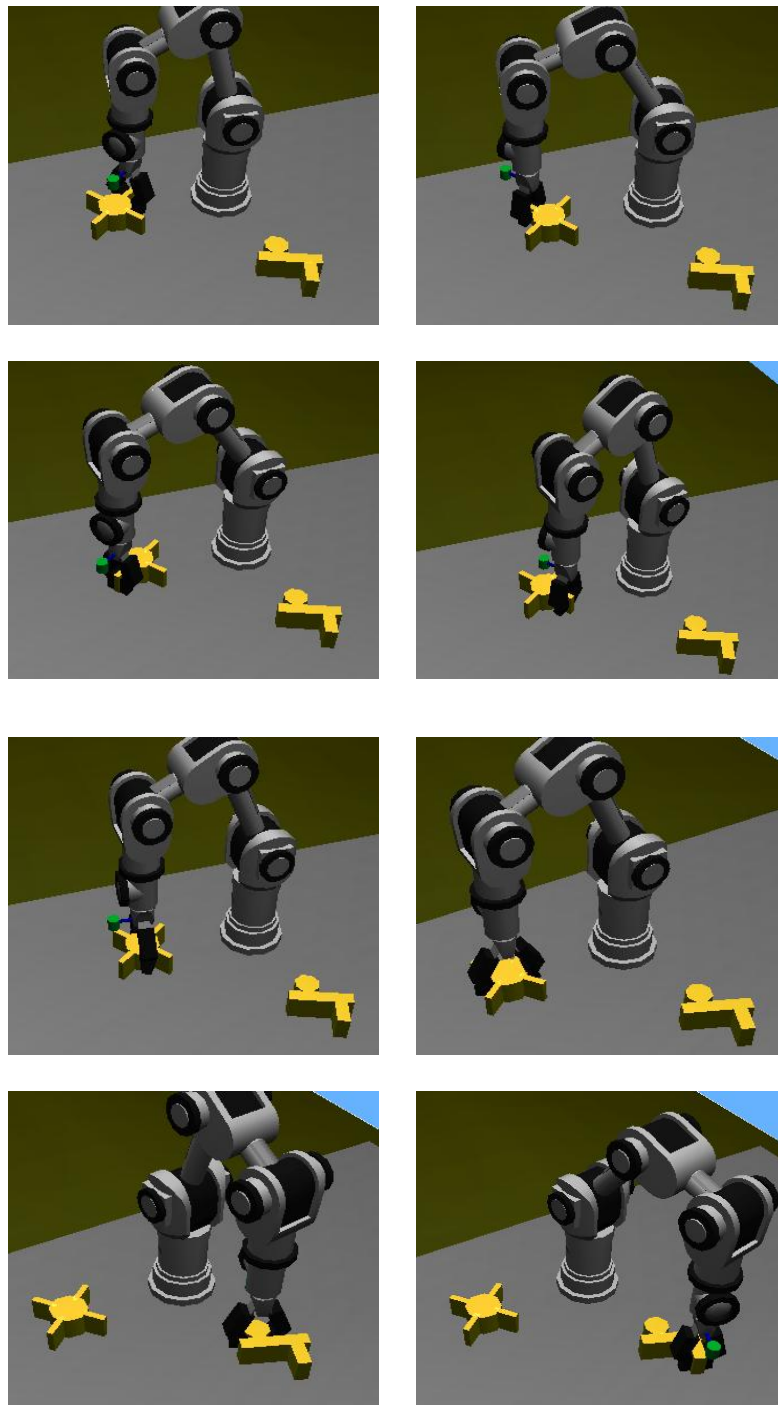
7.7 Υλοποίηση πειράματος

Στην παράγραφο αυτή παρουσιάζουμε στιγμιότυπα εκτέλεσης του αλγορίθμου που περιγράψαμε εκτενώς στην παράγραφο 7.6. Στα σχήματα που ακολουθούν φαίνονται οι επιτρεπτές λαβές για αντικείμενα που είναι τοποθετημένα πάνω σε τραπέζι και έχουν τυχαία θέση και προσανατολισμό.



Σχήμα 7.13: Επιτρεπτές λαβές για πρωταρχικά αντικείμενα. Ο κύβος έχει δυο, το τραπέζιο μια και ο κύλινδρος τέσσερις.





Σχήμα 7.14: Επιτρεπτές λαβές για πολύπλοκα αντικείμενα.

Στα σχήματα 7.13 και 7.14, παρουσιάσαμε ορισμένα παραδείγματα εκτέλεση του πειράματος μας, από τα οποία είναι εμφανές ότι ο αλγόριθμος μας εντοπίζει όλες τις επιτρεπτές λαβές των αντικειμένων.

ΚΕΦΑΛΑΙΟ 8

8 Συμπεράσματα και Μελλοντικές κατευθύνσεις της έρευνας

8.1 Συμπεράσματα

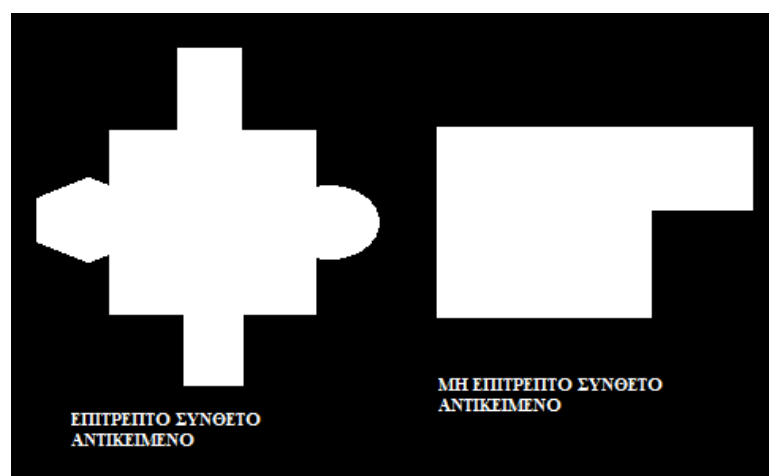
Στη παρούσα διπλωματική εργασία αναπτύχθηκε ένα ρομποτικό σύστημα που είναι ικανό να αναγνωρίζει το σύνολο των επιτρεπτών λαβών των αντικειμένων χρησιμοποιώντας μόνο οπτική πληροφορία. Η οπτική πληροφορία προέρχεται από μια κάμερα που έχει προσαρμοστεί πάνω στον ρομποτικό βραχίονα. Γίνεται χρήση αλγορίθμων επεξεργασίας εικόνας για την εξαγωγή χρήσιμων χαρακτηριστικών των δισδιάστατων σχημάτων που προκύπτουν ως απεικόνιση στην κάμερα των τρισδιάστατων αντικειμένων του χώρου. Η πληροφορία που εξάγεται δίνεται ως είσοδος σε ένα νευρωνικό δίκτυο το οποίο παρέχει στην έξοδό του το διάλυμα επιτρεπτών λαβών για το συγκεκριμένο αντικείμενο. Έπειτα με χρήση κατάλληλων αλγορίθμων στερεοοπτικής όρασης εντοπίζεται η θέση του αντικειμένου στον τρισδιάστατο χώρο. Τέλος ελέγχεται κατά πόσο οι επιτρεπτές λαβές του αντικειμένου ικανοποιούν συγκεκριμένους περιορισμούς που θέτονται από την γεωμετρία του σχήματος και όταν προσδιοριστούν οι τελικές επιτρεπτές λαβές, επιχειρείται η αρπαγή του αντικειμένου από τον ρομποτικό χειριστή.

Η εικόνα που προέρχεται από την κάμερα είναι στον χρωματικό χώρο RGB και πρέπει να μετατραπεί σε ασπρόμαυρη προκειμένου να επεξεργαστεί. Η τεχνική κατάτμησης που χρησιμοποιείται είναι απλή και εφαρμόσιμη όταν υπάρχει γνώση του χρώματος τόσο του αντικειμένου όσο και του φόντου. Στην γενική περίπτωση πρέπει

να αναζητηθούν πιο πολύπλοκες μέθοδοι κατάτμησης εικόνας που προσφέρουν μεγαλύτερη αξιοπιστία ως προς το αποτέλεσμα. Πρέπει να επισημανθεί ότι δε δίνεται έμφαση στο θέμα της κατάτμησης εικόνας καθώς ξεφεύγει από τους στόχους και τα πλαίσια της εργασίας αυτής.

Το ρομποτικό οπτικό σύστημα, προκειμένου να έχει πρακτική εφαρμογή, δηλαδή να μπορεί να εντοπίζει αντικείμενα στο χώρο και να οδηγεί την εκτέλεση εργασιών χειρισμού που προϋποθέτουν επιτυχή λήψη των αντικειμένων μέσω της ρομποτικής αρπάγης, πρέπει οπωσδήποτε να παρουσιάζει ανεξαρτησία ως προς την περιστροφή, την κλιμάκωση και τη μετατόπιση των αντικειμένων στο χώρο. Για τον λόγο αυτό τα οπτικά χαρακτηριστικά που εξάγονται από τις δισδιάστατες εικόνες εισόδου πρέπει να είναι αναλλοίωτα ως προς τις τρεις αυτές παραμορφώσεις. Συγκεκριμένα, χρησιμοποιούνται οι αναλλοίωτες μεταβλητές ως προς τον αφινικό μετασχηματισμό (Affine Moment Invariants – AMIs) που παρουσιάζουν ιδιαίτερη ευρωστία ως προς τις παραμορφώσεις των σχημάτων ενώ ταυτόχρονα μπορούν να διαχωρίσουν τα διαφορετικά σχήματα μεταξύ τους. Επίσης, εισάγεται ο λόγος του μεγαλύτερου προς τον μικρότερο άξονα της περιγεγραμμένης στο σχήμα έλλειψης καθώς και η εκκεντρότητα προκειμένου να διαχωριστούν ιδιόμορφες διατάξεις μεταξύ όμοιων σχημάτων.

Βασική ιδέα που χρησιμοποιείται ώστε να καταστεί δυνατή η απόκριση του συστήματος σε σύνθετα αντικείμενα είναι η κατάτμηση σχήματος. Θεωρούμε ότι ένα σύνθετο σχήμα μπορεί να αποδομηθεί σε απλούστερα για τα οποία μπορούν να αναγνωριστούν ευκολότερα οι επιτρεπτές λαβές τους και μέσω της διαδικασίας της ανασύνθεσης να προκύψουν οι επιτρεπτές λαβές του αρχικού σύνθετου σχήματος. Εφαρμόζουμε την τεχνική της «συντομότερης διάτμησης» για τα οριακά σημεία που προκύπτουν από τον κανόνα ελαχίστου. Ωστόσο, η τεχνική αυτή παρουσιάζει μειονεκτήματα καθώς δεν μας παρέχει πάντα όλα τα οριακά σημεία, όπως αναφέρεται στο κεφάλαιο 4.5. Έτσι περιορίζουμε την υλοποίηση του πειράματός μας μόνο σε αντικείμενα τα οποία έχουν εσοχές και προεξοχές που δεν συμπίπτουν με κάποια πλευρά του αντικειμένου (βλέπε σχήμα 8.1).



Σχήμα 8.1: Διάκριση μεταξύ αντικειμένων που μπορούν να κατατμηθούν και αντικειμένων τα οποία δεν μπορούν με την χρήση του κανόνα ελαχίστου.

Συνεπώς, το νευρωνικό δίκτυο καλείται να αναγνωρίσει τις λαβές απλών σχημάτων που προέρχονται από τη διαδικασία της κατάτμησης σχήματος. Οπότε εκπαιδεύουμε ένα αυτοοργανούμενο χάρτη σε πρωταρχικά σχήματα και γενικεύσεις αυτών. Ο τρόπος επιλογής των δειγμάτων εκπαίδευσης έχει πολύ σημαντικό ρόλο καθώς καθορίζει τις διαχωριστικές επιφάνειες μεταξύ των διαφόρων διατάξεων επιτρεπτών λαβών που έχουμε ορίσει. Για παράδειγμα, έχουν διαφορετικό διάνυσμα επιτρεπτών λαβών το τετράγωνο με το ορθογώνιο παραλληλόγραμμο αν και το ένα αποτελεί μια κλιμάκωση του άλλου. Το νευρωνικό δίκτυο λοιπόν μπορεί να εκπαιδευτεί με ελάχιστο αριθμό δειγμάτων αρκεί αυτά να είναι αντιπροσωπευτικά. Επιπλέον, χρησιμοποιούμε δυο κατώφλια: (α) ένα για την αποτροπή ορισμού λαβών σε μη εκπαιδευμένα σχήματα τα οποία διαφέρουν αρκετά από τα σχήματα για τα οποία έχει εκπαιδευτεί. Τέτοιο σχήμα αποτελεί το μη επιτρεπτό σύνθετο αντικείμενο του σχήματος 8.1. Η τιμή για το πρώτο κατώφλι είναι 8. (β) ένα κατώφλι για το πότε μια λαβή αναγνωρίζεται. Η τιμή του δεύτερου κατωφλίου είναι 0.5 εφόσον έχουμε ορίσει τα δείγματα εκπαίδευσης ώστε να είναι αντιπροσωπευτικά. Σε αντίθετη περίπτωση όπου τα δείγματα εκπαίδευσης έχουν ορισθεί με τυχαίο τρόπο πρέπει να θέσουμε πιο στενά όρια αναγνώρισης όπως τιμή κατωφλίου 0.8 ώστε να είμαστε σίγουροι ότι δεν θα έχουμε εσφαλμένη αναγνώριση. Ωστόσο το ποσοστό επιτυχίας θα είναι μικρό.

Εφόσον έχουμε εντοπίσει τις επιτρεπτές λαβές για τα υποσχήματα, προσδιορίζουμε τις επιτρεπτές λαβές του σύνθετου αντικειμένου με χρήση της μεθόδου της αναπαράστασης της αρπάγης, που περιγράφεται στην παράγραφο 7.6. Η μέθοδος αυτή μας παρέχει ικανοποιητικά αποτελέσματα ωστόσο δεν είναι τόσο εύρωστη. Συγκεκριμένα, σε περίπτωση που το βάθος του αντικειμένου δεν έχει βρεθεί σωστά καθώς είτε ο ρομποτικός βραχίονας δε μπορεί να εκτελέσει εντελώς παράλληλη κίνηση είτε ο θόρυβος από το περιβάλλον έχει επίπτωση στον προσδιορισμό των όμοιων σημείων της μεθόδου τριγωνοποίησης, τότε η αναπαράσταση της αρπάγης στο επίπεδο της εικόνας θα είναι εσφαλμένη. Έτσι υπάρχει η περίπτωση να αποκλείεται κάποια επιτρεπτή λαβή χωρίς να υπάρχουν γεωμετρικοί περιορισμοί που να το επιβάλλουν. Τέτοιου είδους προβλήματα δεν υπάρχουν στο περιβάλλον προσομοίωσης Webots που χρησιμοποιούμε άρα η μέθοδος της αναπαράστασης της αρπάγης λειτουργεί ιδιαίτερα ικανοποιητικά.

Γενικότερα, μπορεί να εξαχθεί το συμπέρασμα ότι η διαδικασία που προτείνουμε στην παρούσα διπλωματική εργασία για τον εντοπισμό των επιτρεπτών λαβών των αντικειμένων είναι αρκετά εύρωστη και παρέχει αξιόπιστα αποτελέσματα στο περιβάλλον προσομοίωσης Webots. Από πειράματα που διεξήχθησαν για αντικείμενα που είναι τοποθετημένα πάνω σε οριζόντιο τραπέζι τα αποτελέσματα είναι ιδιαίτερα ικανοποιητικά καθώς αναγνωρίστηκαν ορθώς όλες οι λαβές στην πλειονότητα των παραδειγμάτων. Η εκπαίδευση του ρομποτικού συστήματος είναι απλή και η αλληλεπίδραση του ρομπότ με τον άνθρωπο είναι ελάχιστη καθώς ο τελευταίος χρειάζεται πολύ λίγο χρόνο για να εκπαιδεύσει το ρομπότ ώστε αυτό να ανταποκρίνεται επαρκώς και σε σύνθετα αντικείμενα.

8.2 Μελλοντικές κατευθύνσεις της έρευνας

Επεκτάσεις και πιθανές βελτιώσεις της παρούσας διπλωματικής εργασίας περιλαμβάνουν, καταρχήν, τη χρήση κάποιας καλύτερης μεθόδου κατάτμησης εικόνας καθώς και τη βελτίωση της μεθόδου που χρησιμοποιείται για την κατάτμηση σχήματος ώστε να μπορούν να κατατμηθούν πιο πολύπλοκα σχήματα. Επιπλέον αντί της χρήσης μιας κάμερας πάνω στην αρπάγη θα μπορούσαν να χρησιμοποιηθούν δυο στατικές κάμερες. Με αυτόν τον τρόπο δεν θα εξαρτάται η εύρεση του βάθους από το κατά πόσο ο ρομποτικός βραχίονας μπορεί η όχι να εκτελέσει παράλληλη κίνηση.

Επίσης με την χρήση δυο στατικών καμερών θα μπορεί να δοθεί και ανάδραση στο σύστημα ώστε να επαναπροσδιορίζει συνεχώς τη θέση του. Το βήμα αυτό είναι σημαντικό σε περίπτωση υλοποίησης του πειράματος στον πραγματικό κόσμο καθώς ακόμα κι αν η οπτική πληροφορία δίνει σωστά αποτελέσματα ως προς τη θέση στην οποία πρέπει να κινηθεί ο ρομποτικός βραχίονας, η χρήση σερβοκινητήρων εισάγει συσσωρευμένο σφάλμα στο σύστημα. Η διαδικασία εύρεσης των συντεταγμένων λοιπόν του αντικειμένου θα πρέπει να ανανεώνεται ανά τακτά χρονικά διαστήματα.

Επιπλέον, καλό θα ήταν να χρησιμοποιηθούν για την εκπαίδευση και διαφορετικές διατάξεις νευρωνικών δικτύων ώστε να συγκριθούν τα πειραματικά αποτελέσματα με τα αντίστοιχα που προκύπτουν από τη χρήση αυτοοργανούμενων χαρτών. Για παράδειγμα, τα ασαφή συστήματα (fuzzy system) θα ήταν μια καλή επιλογή για την εκπαίδευση του συστήματος καθώς παρέχουν μια ασαφή διαμέριση του χώρου εισόδου με αποτέλεσμα να λαμβάνονται υπόψη και τυχόν ανοχές που θέλουμε να θέσουμε.

Η μέθοδος που παρουσιάζεται στην παρούσα διπλωματική εργασία αναγνωρίζει τις επιτρεπτές λαβές των αντικειμένων που είναι τοποθετημένα πάνω σε οριζόντιο τραπέζι περιορίζοντας μ' αυτό τον τρόπο σημαντικά το πεδίο δράσης του ρομπότ. Σε μια μελλοντική εργασία θα μπορούσε να επιχειρηθεί η λαβή αντικειμένων τα οποία έχουν τυχαία θέση και προσανατολισμό στο χώρο. Στην περίπτωση αυτή δεν πρέπει να αναγνωρίζονται οι επιτρεπτές λαβές μόνο ως προς μια κατεύθυνση αλλά ως προς διάφορες κατευθύνσεις και προσανατολισμούς. Για τον λόγο αυτό επιβεβλημένο θα ήταν να χρησιμοποιηθεί εξαρχής τρισδιάστατη πληροφορία, όπως ένα νέφος σημείων για την περιγραφή της γεωμετρίας του σχήματος και την εξαγωγή των οπτικών χαρακτηριστικών.

Σε περίπτωση χρήσης τρισδιάστατης πληροφορίας προκειμένου τα σύνθετα αντικείμενα να αποδομηθούν σε απλούστερα θα μπορούσαν να χρησιμοποιηθούν τα πολύεδρα ως βασικά αντικείμενα. Η εκπαίδευση επομένως του ρομποτικού συστήματος να γίνεται με βασικά πολύεδρα όπως ο κύβος, ο κύλινδρος κ.α. .

Τέλος, εκτός από τρισδιάστατη πληροφορία καλό θα ήταν να χρησιμοποιηθεί και ένας ρομποτικός βραχίονας που να έχει τουλάχιστον έξι βαθμούς ελευθερίας ώστε να είναι δυνατή η επίτευξη οποιασδήποτε θέσης και προσανατολισμού στο τρισδιάστατο χώρο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Abbasi S., Mokhtarian F., Kittler J. (1999). “Curvature scale space image in shape similarity retrieval”, *Multimedia Systems*, vol. 7, no. 6, pp. 467-476.

Attneave F. (1954). “Some informational aspects of visual perception”, *Psychological Review*, vol. 61, no. 3, pp. 183-193.

Binford T. (1971). “Visual perception by computer”, In: *Proceedings of IEEE Conf. Systems and Controls*, Miami.

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

Erwin E, Obermayer K., Schulten K. (1992). “Self-organizing maps: Stationary states, metastability and convergence rate”, *Biological Cybernetics*, vol. 67, pp. 35-45.

Flusser J., Suk T. (1994). “Affine moment invariants: A new tool for character recognition”, *Pattern Recognition Letters*, vol. 15, pp. 433-436.

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

Goldfeder C., Allen P., Lackner C., Pelosof R. (2007). “Grasp planning via decomposition trees”, In: *Proceedings of IEEE Int. Conference on Robotics and Automation, Rome*.

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

Hastie T., Tibshirani R, Friedman J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*, Second Edition, Springer.

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

Hoffman D., Richards W. (1984). “Parts of recognition”, *Cognition*, vol. 18, pp. 65-96.

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

Kohonen T. (1982). “Self-organized formation of topologically correct feature maps”, *Biological Cybernetics*, vol. 43, pp. 59-69.

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

Madsen K., Nielsen H. (2004). *Methods for non-linear least squares problems*, Informatics and Mathematical Modelling, Technical University of Denmark, DTU.

Marr D., Hildreth E. (1980). “Theory of edge detection”, In: *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 207, pp. 187-217.

Midenet S., Grumbach A. (1994). “Learning associations by self-organization: the LASSO model”, *Neurocomputing*, vol. 6, pp. 343 – 361.

Mokhtarian F., Mackworth A. (1984). *Scale-based descriptions of planar curves*, Technical Report, University of British Columbia, Vancouver, BC, Canada.

Ritter H., Martinetz T., Schulten K. (1992). *Neural computation and self-organizing maps: An introduction*, Addison-Wesley Longman Publishing Co.

Seyranian G., Hoffman D. (1999). “Parsing silhouettes: The short-cut rule”, *Perception and Psychophysics*, vol. 61, pp. 636-660.

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

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

Teague M. (1980). “Image analysis via the general theory of moments”, *Optical Society of America*, vol. 70, pp. 920-930.

Ultsch H., Siemon H. (1990). “Kohonen's self-organizing feature maps for exploratory data analysis”, In: *Proceedings of International Neural Networks Conference (INNC)*, pp. 305-308.

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