



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

Προσαρμοζόμενος Έλεγχος Αυτοκινούμενου Ρομπότ Διαφορικής
Οδήγησης Για Την Παρακολούθηση Εξωτερικών Γεωμετρικών
Χαρακτηριστικών Με Χρήση Τεχνικών Ασαφούς Συλλογιστικής
Και Ενισχυτικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΩΝΣΤΑΝΤΙΝΟΣ Ε. ΔΟΓΕΑΣ

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

Αθήνα, Σεπτέμβριος 2014



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

Προσαρμοζόμενος Έλεγχος Αυτοκινούμενου Ρομπότ Διαφορικής
Οδήγησης Για Την Παρακολούθηση Εξωτερικών Γεωμετρικών
Χαρακτηριστικών Με Χρήση Τεχνικών Ασαφούς Συλλογιστικής
Και Ενισχυτικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΩΝΣΤΑΝΤΙΝΟΣ Ε. ΔΟΓΕΑΣ

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

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12η Σεπτεμβρίου 2014

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

.....
Γεώργιος Π. Παπαβασιλόπουλος
Καθηγητής Ε.Μ.Π.

.....
Σπυρίδων Γ. Τζαφέστας
Ομότιμος Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2014

.....

Κωνσταντίνος Ε. Δογέας

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

©Κωνσταντίνος Ε. Δογέας, 2014

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Η πλοήγηση αυτόνομων ρομπότ σε άγνωστο περιβάλλον, αποτελεί ένα πολυδιάστατο πρόβλημα, που κατατάσσεται στη μηχανική συστημάτων λόγω των διαφορετικών επιστημονικών πεδίων που απαιτεί. Έναν ιδιαίτερο τομέα έρευνας, αποτελεί η Μηχανική Μάθηση, μια και αποτελεί το μέσο για τη διαχείριση δεδομένων και τη δημιουργία αυτόνομων συμπεριφορών. Στη διπλωματική αυτή εργασία, εξετάζεται, η κίνηση ρομπότ διαφορικής οδήγησης σε άγνωστο περιβάλλον, εφαρμόζοντας Ασαφή Ενισχυτική Μάθηση, για την περίπτωση ακολούθησης τοίχου. Γίνεται κινηματική ανάλυση για τα ρομπότ διαφορικής οδήγησης, παρουσιάζεται η μέθοδος της Γενικής Αντίληψης και το θεωρητικό υπόβαθρο για την Ασαφή Λογική, την Ενισχυτική Μάθηση και την Ασαφή Ενισχυτική Μάθηση. Στο κομμάτι υλοποίησης, παρουσιάζεται το ρομπότ Khepera 3, στο οποίο εφαρμόστηκε η Ασαφής Ενισχυτική Μάθηση και μελετάται η συμπεριφορά των αισθητήρων του. Μελετάται επίσης η συμπεριφορά των αισθητήρων ενός πραγματικού ρομπότ Khepera 2. Ακολουθεί υλοποίηση Ασαφούς Ελέγχου και αριθμητικά αποτελέσματα, τόσο για την περίπτωση πραγματικού ρομπότ Khepera 2 όσο και για το ρομπότ Khepera 3 σε περιβάλλον εξομοίωσης. Τέλος υλοποιείται Ασαφής Ενισχυτική Μάθηση και εφαρμόζεται στο ρομπότ Khepera 3, σε περιβάλλον εξομοίωσης, όπου και παρουσιάζονται τα αριθμητικά αποτελέσματα και τα τελικά συμπεράσματα.

Λέξεις Κλειδιά: Ακολούθηση Τοίχου, Ασαφής Λογική, Ενισχυτική Μάθηση, Ασαφής Ενισχυτική Μάθηση, Ρομπότ Διαφορικής Οδήγησης, Γενική Αντίληψη, Αυτόνομο Ρομπότ, Q-Μάθηση

Abstract

Navigation of autonomous robots in unknown environments is a multidimensional problem, that can be classified as a systems engineering problem, due to the different scientific fields it requires. A specific field of research is Machine Learning, it being the medium for data management and creation of autonomous behaviors. This Diploma Thesis examines the movement of differential drive robots, in unknown environments, by applying the Fuzzy Reinforcement Learning method, for the wall following case. After a kinematic analysis of differential drive robots, the General Perception concept is presented, followed by the theoretical background for Fuzzy Logic, Reinforcement Learning and Fuzzy Reinforcement Learning. In the implementation part, a brief presentation of the Khepera 3 differential drive robot, is made, since it was the platform on which Fuzzy Reinforcement Learning was applied. Sensor behavior is presented for both the Khepera 3 and a real Khepera 2. An implementation of Fuzzy Control is made for a real Khepera 2 robot as well as a simulated Khepera 3 robot and results are presented. An implementation of Fuzzy Reinforcement Learning for a simulated Khepera 3 robot is made, with results and conclusions.

Keywords: Wall Following, Fuzzy Logic, Reinforcement Learning, Fuzzy Reinforcement Learning, Differential Drive Robot, General Perception, Autonomous Robot, Q-Learning

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή μου, Κωνσταντίνο Σ. Τζαφέστα, Επίκουρο Καθηγητή Ε.Μ.Π., για τη δυνατότητα που μου έδωσε, να εκπονήσω τη διπλωματική αυτή εργασία, σε μια δύσκολη για μένα περίοδο. Σας ευχαριστώ και σαν επιστήμονα και σαν άνθρωπο.

Τον πατέρα μου Νώντα, για τα εφόδια που μου έδωσε και που δεν πρόλαβε να δει να αποφέρουν καρπούς.

Τη μητέρα μου Κλειώ, την αδερφή μου Μαριάννα και τους Χρήστο και Χρυσή Φλώρου, που μου στάθηκαν και μου στέκονται, ανεξαρτήτως συνθηκών.

Τέλος, ευχαριστώ τη Μαρία, για τη στήριξη και την αγάπη της.

Περιεχόμενα

1	Εισαγωγή	17
1.1	Ρομπότ	17
1.2	Μηχανική Μάθηση	18
1.3	Σκοπός της Διπλωματικής Εργασίας	19
1.4	Οργάνωση Κειμένου	20
2	Κινηματική Ανάλυση	21
2.1	Κινηματική Ανάλυση Ρομπότ Διαφορικής Οδήγησης	21
2.2	Ορθή Κινηματική Ανάλυση Ρομπότ Διαφορικής Οδήγησης	23
2.3	Ανάστροφη Κινηματική Ανάλυση Ρομπότ Διαφορικής Οδήγησης	23
2.4	Οδομετρία Ρομπότ Διαφορικής Οδήγησης	25
3	Σύνθεση Δεδομένων Αισθητήρων	27
3.1	Κυριότερες Μέθοδοι Σύνθεσης Αισθητήρων	27
3.2	Η Μέθοδος της Γενικής Αντίληψης	28
4	Ασαφής Λογική	30
4.1	Ασαφή Συνολα	30
4.2	Ασαφείς Μεταβλητές	31
4.3	Ασαφείς Τελεστές	32
4.4	Ασαφείς Κανόνες και Ασαφής Λογική	34
4.5	Αποασαφοποίηση	35
4.5.1	Μέθοδος Μέσου Μεγίστων (MOM)	35
4.5.2	Μέθοδος Κέντρου Βαρύτητας (COG)	35
4.5.3	Μέθοδος Ύψους (HM)	35
4.6	Ασαφές Συμπερασματικό Σύστημα	36
4.7	Ασαφές Συμπερασματικό Σύστημα Takagi-Sugeno (TS-FIS)	36
4.8	Σύγκριση (M-FIS) και (TS-FIS)	38
5	Ενισχυτική Μάθηση (Reinforcement Learning)	40
5.1	Γενικά για την Ενισχυτική Μάθηση	40
5.2	Η συνάρτηση ανατροφοδότησης	41
5.3	Μαρκοβιανές Διαδικασίες Λήψης Αποφάσεων	42
5.4	Μαθηματικό Πλαίσιο	42
5.4.1	Η συνάρτηση αξίας	42
5.4.2	Επίλυση της Εξίσωσης Bellman	44
5.5	Χρονικές διαφορές (Temporal Differences)	45
5.5.1	TD(0)	45
5.5.2	Ίχνος Επιλεξιμότητας	46
5.5.3	TD(λ)	47
5.5.4	Μάθηση Q (Q-Learning)	47
5.6	Εξερεύνηση/Εκμετάλλευση	48

6	Ασαφής Λογική με Ενισχυτική Μάθηση	50
6.1	Μορφή του Ασαφούς Συμπερασματικού Συστήματος	50
6.2	Αναπαράσταση της συνάρτησης Q από ένα Ασαφές Συμπερασματικό Σύστημα	51
6.3	Συνεχείς χώροι δράσεων και καταστάσεων	52
6.4	Εξαγωγή Γνώσης	52
6.4.1	Επιλογή Συμπερασμάτων	53
6.4.2	Αλγόριθμος	54
7	Υλικό και Λογισμικό Υλοποίησης	56
7.1	Το Ρομπότ Khepera 3	56
7.1.1	Γενικά για το ρομπότ	56
7.1.2	Σύστημα Κίνησης	57
7.1.3	Σύστημα Αισθητήρων	58
7.1.4	Σύστημα Τροφοδοσίας και Σύνδεση με Υπολογιστή	60
7.2	Λογισμικό V-REP	60
8	Βαθμονόμηση Αισθητήρων Πραγματικού Ρομπότ	62
8.1	Δεδομένα κατασκευαστή	62
8.2	Δεδομένα πραγματικού ρομπότ	64
8.3	Αξιολόγηση των Αισθητήρων	65
9	Υλοποίηση Ασαφούς Λογικής	67
9.1	Είσοδοι	67
9.2	Έξοδοι	68
9.3	Βάση Κανόνων	68
9.4	Αριθμητικά Αποτελέσματα Πραγματικού Ρομπότ	69
9.5	Αριθμητικά Αποτελέσματα του Ρομπότ σε Περιβάλλον Εξομοίωσης	72
9.5.1	Αποτελέσματα για Ευθεία Κίνηση	72
9.5.2	Αποτελέσματα για Τετράγωνη Αρένα	74
9.5.3	Αποτελέσματα σε Αρένα με Διαφορετικούς Τύπους Γωνιών	77
10	Υλοποίηση Ασαφούς Λογικής με Ενισχυτική Μάθηση	81
10.1	Βάση Κανόνων	81
10.2	Ανταμοιβές	82
10.3	Δράσεις	82
10.4	Επιλογή Δράσης από τις Διαθέσιμες	83
10.5	Αλγόριθμος Q-μάθησης	84
10.6	Αριθμητικά Αποτελέσματα του Ρομπότ σε Περιβάλλον Εξομοίωσης	85
10.6.1	Αποτελέσματα για Ευθεία Κίνηση	86
10.6.2	Αποτελέσματα για Τετράγωνη Αρένα	92
10.6.3	Αποτελέσματα σε Αρένα με Διαφορετικούς Τύπους Γωνιών	96
11	Επίλογος	100
11.1	Συμπεράσματα	100
11.2	Μελλοντικές Επεκτάσεις	101

Κατάλογος σχημάτων

2.1	Όχημα διαφορικής οδήγησης	22
2.2	Κίνηση του ρομπότ γύρω από το <i>ICC</i>	24
2.3	Γεωμετρία ρομπότ διαφορικής οδήγησης	25
3.1	Γενική Αντίληψη p	29
4.1	Τραπεζοειδής συνάρτηση συμμετοχής	31
4.2	Τριγωνοειδής συνάρτηση συμμετοχής	31
4.3	Αρχιτεκτονική Ασαφούς Συμπερασματικού Συστήματος Takagi-Sugeno (TS-FIS)	38
5.1	Μοντελοποίηση δράσης-ανάδρασης, μεταξύ του πράκτορα και του περιβάλλοντος, στην ενισχυόμενη μάθηση.(Το D αναπαριστά χρονική καθυστέρηση) .	41
5.2	Αρχιτεκτονική Επαναλήψεων επί της αξίας	45
6.1	Σχέση μεταξύ των $x, A(x), \hat{Q}(x, A(x))$	52
7.1	Το Ρομπότ Khepera 3 [10]	56
7.2	Pulse Width Modulation - PWM [10]	57
7.3	Διάταξη των αισθητήρων υπερύθρων του Khepera 3 [10]	59
7.4	Διάταξη των αισθητήρων υπερήχων του Khepera 3 [10]	59
7.5	Το περιβάλλον εξομοίωσης V-REP, με αρένα κίνησης και ρομπότ Khepera 3 .	61
8.1	Μετρήσεις περιβαλλοντικού φωτός χρησιμοποιώντας πηγή ισχύος 50W [27] .	63
8.2	Μετρήσεις απόστασης των αισθητήρων σε λευκό χαρτί [11]	63
8.3	Μετρήσεις απόστασης των αισθητήρων σε διαφορετικές επιφάνειες [26] . . .	64
8.4	Αρένα που κατασκευάστηκε	64
8.5	Μετρήσεις Αισθητήρων του πραγματικού ρομπότ	65
9.1	Γωνία Αντίληψης α	68
9.2	Αντίληψη p	68
9.3	Γωνία Οδήγησης	69
9.4	Γωνία και απόσταση σε σχέση με τον τοίχο σε ευθεία κίνηση, υπό Ασαφή Λογική για πραγματικό ρομπότ	70
9.5	Γωνία και απόσταση σε σχέση με τον τοίχο κατά την εκτέλεση κλειστής αριστερής γωνίας, υπό Ασαφή Λογική για πραγματικό ρομπότ	71
9.6	Γωνία και απόσταση σε σχέση με τον τοίχο κατά την εκτέλεση αμβλείας αριστερής γωνίας, υπό Ασαφή Λογική για πραγματικό ρομπότ	71
9.7	Αντίληψη κατά την εκτέλεση ευθείας κίνησης, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	72
9.8	Γωνία Αντίληψης κατά την εκτέλεση ευθείας κίνησης, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	73

9.9	Σφάλμα από την επιθυμητή απόσταση κατα την εκτέλεση ευθείας κίνησης, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	73
9.10	Φορές που το ρομπότ κατάφερε να βρεθεί στην επιθυμητή απόσταση και λαμβάνει ανταμοιβή κατα την εκτέλεση ευθείας κίνησης, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	74
9.11	Αντίληψη κατα την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	74
9.12	Γωνία Αντίληψης κατα την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	75
9.13	Σφάλμα από την επιθυμητή απόσταση κατα την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	75
9.14	Φορές που το ρομπότ κατάφερε να βρεθεί στην επιθυμητή απόσταση και λαμβάνει ανταμοιβή κατα την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	76
9.15	Μονοπάτι για τετράγωνη αρένα, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	76
9.16	Αντίληψη κατα την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	77
9.17	Γωνία Αντίληψης κατα την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	78
9.18	Σφάλμα από την επιθυμητή απόσταση κατα την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	78
9.19	Φορές που το ρομπότ κατάφερε να βρεθεί στην επιθυμητή απόσταση και λαμβάνει ανταμοιβή κατα την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	79
9.20	Μονοπάτι κατα την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης	79
10.1	Κεντρικές τιμές Συνάρτησης Συμμετοχής Εξόδου, γύρω από τις οποίες κυμαίνονται οι πιθανές δράσεις	82
10.2	Αντίληψη κατά την εκτέλεση ευθείας κίνησης 100, 250, 500 εποχών, υπό Ασαφή Ενισχυτική Μάθηση	87
10.3	Γωνία Αντίληψης κατά την εκτέλεση ευθείας κίνησης 100, 250, 500 εποχών, υπό Ασαφή Ενισχυτική Μάθηση	88
10.4	Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση ευθείας κίνησης 100, 250, 500 εποχών, υπό Ασαφή Ενισχυτική Μάθηση	89
10.5	Λήψη ανταμοιβών κατά την εκτέλεση ευθείας κίνησης 100, 250, 500 εποχών, υπό Ασαφή Ενισχυτική Μάθηση	90
10.6	Λήψη ανταμοιβών κατά την εκτέλεση ευθείας κίνησης 100 εποχών, υπό Ασαφή Ενισχυτική Μάθηση σε σύγκριση με ευθεία κίνηση υπό Ασαφή Έλεγχο	91
10.7	Μονοπάτι για τετράγωνη αρένα, υπό Ασαφή Ενισχυτική Μάθηση	92
10.8	Αντίληψη κατά την εκτέλεση τετράγωνης αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1500 εποχές	93
10.9	Γωνία αντίληψης για τον τοίχο κατά την εκτέλεση τετράγωνης αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1500 εποχές	93
10.10	Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση τετράγωνης αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1500 εποχές	94
10.11	Λήψη ανταμοιβών κατά την εκτέλεση τετράγωνης αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1500 εποχές	94
10.12	Λήψη ανταμοιβών κατά την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Ενισχυτική Μάθηση σε σύγκριση με εκτέλεση τετράγωνης αρένας υπό Ασαφή Έλεγχο	95

10.13	Απόσταση από τον τοίχο κατά την εκτέλεση της αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές	96
10.14	Γωνία Αντίληψης κατά την εκτέλεση της αρένας διαφορετικών γωνιών υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές	97
10.15	Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση της αρένας διαφορετικών γωνιών υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές	97
10.16	Λήψη ανταμοιβών κατά την εκτέλεση της αρένας διαφορετικών γωνιών υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές	98
10.17	Λήψη ανταμοιβών κατά την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Ενισχυτική Μάθηση σε σύγκριση τον Ασαφή Έλεγχο	98
10.18	Μονοπάτι που ακολούθησε το ρομπότ στον εξομοιωτή κατα την εκτέλεση της αρένας διαφορετικών γωνιών υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές .	99

Κατάλογος πινάκων

4.1 Πίνακας αληθείας ασαφών λογικών τελεστών για δυαδικά στοιχεία	34
9.1 Βάση κανόνων	69
10.1 Τροποποιημένη Βάση κανόνων	82
10.2 Παράμετροι μάθησης της υλοποίησης	85
10.3 Αρχική συνάρτηση συνάρτηση συμμετοχής εξόδου	85
10.4 Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου για ευθεία κίνηση 100 εποχών	86
10.5 Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου για ευθεία κίνηση 250 εποχών	86
10.6 Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου για ευθεία κίνηση 500 εποχών	86
10.7 Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου, τετράγωνη αρένα, 1500 εποχές	92
10.8 Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου, αρένας διαφορετικών γωνιών, 1000 εποχές	96

Κεφάλαιο 1

Εισαγωγή

1.1 Ρομπότ

Από τα αυτόματα του Έρωνα του Αλεξανδρέως μέχρι τα σύγχρονα ρομπότ εξερεύνησης πλανητών, η κατασκευή μηχανών που εκτελούν αυτόματα εργασίες αποτέλεσε πεδίο έρευνας για την ανθρωπότητα, στα πλαίσια της ανάγκης για εξασφάλιση καλύτερων συνθηκών ζωής, βελτιστοποιώντας την χρήση ανθρώπινων πόρων κατα την επίτευξή τους.

Ο όρος ρομπότ (robot) εισήχθη από τον Τσέχο συγγραφέα Karel Čapek στο θεατρικό του έργο "R.U.R. (Rossum's Universal Robots)" το 1920. Οι ρίζες της λέξης βρίσκονται στην λέξη "rabota" της Παλαιάς Βουλγαρικής Γλώσσας, που συναντάται σε πολλές Σλαβικές γλώσσες και σημαίνει εργασία.

Όσον αφορά στον ορισμό της λέξης ρομπότ, δεν υπάρχει κοινά αποδεκτός ορισμός. Ο εγκυκλοπαιδικός ορισμός που δίνεται από τη Βρετανική Εγκυκλοπαίδεια είναι : " Οποιαδήποτε αυτόματα λειτουργούσα μηχανή, που αντικαθιστά την ανθρώπινη προσπάθεια, ακόμα και αν δεν προσομοιάζει στα ανθρώπινα όντα στην εμφάνιση, ή δεν εκτελεί λειτουργίες με ανθρώπινο τρόπο". Παρόλαυτά οι προσεγγίσεις στο τι ακριβώς συνιστά ένα ρομπότ ποικίλουν, γεγονός που ίσως να εξηγείται από το ότι σαν αντικείμενο είναι διαρκώς εξελισσόμενο.

Ο όρος *Ρομποτική* (Robotics), χρησιμοποιείται για να περιγράψει το πεδίο επιστημονικής μελέτης των ρομπότ και εισήχθη από τον συγγραφέα επιστημονικής φαντασίας Dr. Isaac Asimov, στην ιστορία του, "Liar" το 1941.

Στη σύγχρονη εποχή, τα ρομπότ, χρησιμοποιούνται σε πολλούς διαφορετικούς τομείς, από τη διασκέδαση, την ιατρική, τη βιομηχανία και τη διαστημική έρευνα, μέχρι τις πολεμικές επιχειρήσεις.

Μια κατηγοριοποίηση των ρομπότ θα μπορούσε να είναι η παρακάτω (χωρίς να είναι η μοναδική) :

- *Αυτόνομα Ρομπότ* (Autonomous Robots) : Λειτουργούν ανεξάρτητα, χωρίς ανθρώπινη παρέμβαση. Είναι Αυτόνομα, γιατί, είναι ευθύνη του ρομπότ να αναγνωρίσει αλλαγές στο περιβάλλον του και να προσαρμοστεί σε αυτό.
- *Προ-προγραμματισμένα Ρομπότ* (Pre-programmed Robots) : Λειτουργούν σε απλά, ελεγχόμενα περιβάλλοντα και δεν απαιτούν ιδιαίτερα ευφυή συστήματα ελέγχου.
- *Τηλεκατευθυνόμενα Ρομπότ* (Teleoperated Robots) : Ελέγχονται από απόσταση, μέσω κάποιας διεπαφής (interface), από ανθρώπους.

- *Αυξητικά ρομπότ (Augmenting Robots)* : Συνδέονται απευθείας στο σώμα του ανθρώπινου χειριστή μέσω κάποιων χειριστηρίων ή ακόμα και σε απευθείας επαφή με το ανθρώπινο σώμα. Οι κινήσεις του ανθρώπινου σώματος (ή ακόμα και η εγκεφαλική δραστηριότητα) ελέγχουν κάποιο ρομποτικό τελεστή.

Η εξέλιξη της ρομποτικής επιστήμης υπήρξε και προβλέπεται, να συνεχίσει να είναι, ραγδαία. Μαζί με τα πλεονεκτήματά της, φέρνει στο προσκήνιο το διάλογο πάνω στα πιθανά προβλήματα που μπορεί να δημιουργήσει η διάδοσή της.

Ηθικά ζητήματα (που είχε θέσει ο Dr. Isaac Asimov με τους Τρεις Νόμους της Ρομποτικής), όπως και κοινωνικά ζητήματα καθώς εργατικά χέρια αντικαθίστανται από ρομπότ και τα προβλήματα ανεργίας που αυτό μπορεί να δημιουργήσει.

1.2 Μηχανική Μάθηση

Καθώς η έρευνα στην τεχνολογία υπολογιστών προχωρά, η δυνατότητα αποθήκευσης και επεξεργασίας δεδομένων αυξάνεται. Δεδομένα προερχόμενα από την ανθρώπινη εμπειρία, στατιστικά δεδομένα, ή και δεδομένα μετρήσεων, συγκεντρώνονται, αποθηκεύονται και είναι προσβάσιμα από όλο και μεγαλύτερες αποστάσεις, με τη βοήθεια των δικτύων υπολογιστών.

Τι σημαίνουν αυτά τα δεδομένα; Αν ήταν γνωστές οι σχέσεις που τα διέπουν, δε θα χρειαζόταν καμία ανάλυση τους, καμία επεξεργασία τους, απλά θα επιλέγαμε τον τρόπο πρόσβασής τους. Επειδή όμως οι σχέσεις και οι υποκείμενες διεργασίες μας είναι άγνωστες, συλλέγουμε τα δεδομένα και ελπίζουμε να εξάγουμε απαντήσεις από αυτά.

Η πίστη πως μπορούμε να εξάγουμε απαντήσεις από τα δεδομένα, οφείλεται, σε ένα από τα πλέον σημαντικά χαρακτηριστικά της επιστημονικής μεθόδου, το γεγονός πως θεωρούμε ότι πίσω από τα δεδομένα, υπάρχει μια διαδικασία που μπορεί να τα περιγράψει. Αν και η διαδικασία μας είναι άγνωστη, γνωρίζουμε πως η διαδικασία δεν είναι τελείως τυχαία. Πρέπει να εντοπίσουμε ποιά πρότυπα χαρακτηριστικά παρουσιάζει αυτή.

Ακόμα και αν δεν μπορέσουμε να ταυτοποιήσουμε την διαδικασία πλήρως, μπορούμε να κατασκευάσουμε μια προσέγγιση της. Ίσως η προσέγγιση να μην εξηγήει πλήρως την υποκείμενη διαδικασία των δεδομένων, αλλά μπορεί να μας εξηγήσει κάποια από τα δεδομένα και μας επιτρέπει να προχωρήσουμε στη χρήση τους για την εφαρμογή που τα χρειαζόμαστε.

Η εφαρμογή Μηχανικής Μάθησης (Machine Learning) σε βάσεις δεδομένων, ονομάζεται *Εξόρυξη Δεδομένων (Data Mining)*.

Η εξόρυξη δεδομένων βρίσκει εφαρμογές σε πολλούς τομείς της σύγχρονης κοινωνικοοικονομικής ζωής. Εμπόριο, διαφήμιση, χρηματιστήριο, ιατρική, φυσική, τηλεπικοινωνίες, ρομποτική αποτελούν μερικούς μόνο από τους τομείς, στους οποίους η εξόρυξη δεδομένων χρησιμοποιείται. Παρά τη διαφορετική φύση των τομέων, η διαχείριση των δεδομένων παραμένει κατα βάση ίδια : Ανάλυση των δεδομένων, ώστε να βρεθεί συσχέτιση μεταξύ τους και εξαγωγή ενός μοντέλου.

Η Μηχανική Μάθηση, δεν είναι μόνο ένα πρόβλημα βάσεων δεδομένων. Είναι και κομμάτι της Τεχνητής Νοημοσύνης (Artificial Intelligence). Για να είναι ένα σύστημα νοήμον, πρέπει σε ένα μεταβαλλόμενο περιβάλλον να παρουσιάζει τη δυνατότητα της μάθησης. Αν

το σύστημα μπορεί να μάθει και να προσαρμοστεί σε αλλαγές, ο σχεδιαστής του δεν χρειάζεται να προβλέπει και να παρέχει λύσεις για όλες τις πιθανές καταστάσεις συστήματος.

Ως Μηχανική Μάθηση ορίζεται ο προγραμματισμός υπολογιστών για τη βελτιστοποίηση κριτηρίων απόδοσης, χρησιμοποιώντας ένα δείγμα δεδομένων ή προηγούμενη εμπειρία. Το μοντέλο μπορεί να είναι προβλεπτικό, εκτιμώντας μελλοντικές συμπεριφορές, περιγραφικό, προσπαθώντας να εξάγει γνώση από τα δεδομένα, ή και τα δύο.

Κύριο εργαλείο της Μηχανικής Μάθησης είναι η στατιστική θεωρία, έτσι ώστε να κατασκευαστούν μαθηματικά συμπερασματικά μοντέλα χρησιμοποιώντας ένα δείγμα δεδομένων.

1.3 Σκοπός της Διπλωματικής Εργασίας

Σκοπός της διπλωματικής αυτής εργασίας, είναι η υλοποίηση ενός ευφυούς αλγορίθμου κίνησης, για κινούμενο, σε άγνωστο περιβάλλον, ρομπότ διαφορικής οδήγησης, για την περίπτωση ακολούθησης τοίχου σε συγκεκριμένη απόσταση.

Χρησιμοποιώντας την έννοια της *Γενικής Αντίληψης* [8], για τη σύνθεση των δεδομένων των αισθητήρων, υλοποιήθηκε μια αναπαράσταση του περιβάλλοντος και μία βάση κανόνων, για κάθε βήμα κίνησης του ρομπότ, χρησιμοποιώντας την Ασαφή Λογική.

Με βάση την Ασαφή αναπαράσταση του περιβάλλοντος, πραγματοποιήθηκε ασαφής έλεγχος πάνω σε πραγματικό ρομπότ Khepera 2, για το πρόβλημα ακολούθησης τοίχου σε συγκεκριμένη απόσταση, σε άγνωστο περιβάλλον. Το ίδιο έγινε και σε περιβάλλον εξομοίωσης V-REP, της Coppelia Robotics για το ρομπότ Khepera 3.

Έπειτα, έχοντας την Ασαφή αναπαράσταση του περιβάλλοντος, χρησιμοποιήθηκε Μηχανική Μάθηση, τύπου Ενισχυτικής Μάθησης και πιο συγκεκριμένα Q-μάθησης, για τη βελτίωση της εξόδου του συστήματος. Υλοποίηση της Ασαφούς Ενισχυτικής Μάθησης έγινε στο περιβάλλον εξομοίωσης V-REP, της Coppelia Robotics για το ρομπότ Khepera 3.

Το πρόβλημα της κίνησης σε άγνωστο περιβάλλον, αποτελεί ένα από τα σημαντικότερα προβλήματα του τομέα των κινητών ρομπότ. Η Ασαφής Λογική προσφέρει γλωσσικές περιγραφές και αντίληψη καταστάσεων, ανάλογες της ανθρώπινης. Η ανθρώπινη εμπειρία χρησιμοποιείται στην Ασαφή λογική, τόσο για την αναπαράσταση του κόσμου, όσο και για τις κατάλληλες δράσεις που θα πρέπει να κάνει το ρομπότ.

Παρόλαυτά, το κομμάτι δράσεων της Ασαφούς λογικής, ορίζεται από τον άνθρωπο, βασισμένο σε εμπειρικά δεδομένα. Η συμπεριφορά του συστήματος μπορεί να μην είναι βέλτιστη. Με την εισαγωγή της Ενισχυτικής Μάθησης εξερευνώνται δράσεις κοντά στην προτεινόμενη από τον άνθρωπο δράση. Η Ενισχυτική Μάθηση, με τη σύγκλιση που εξασφαλίζει, προσφέρει τη βέλτιστη δράση για μια κατάσταση.

Αντίστοιχα, η Ενισχυτική Μάθηση εξερευνώντας πιθανές δράσεις, μπορεί να εξετάζει δράσεις οι οποίες με βάση την ανθρώπινη εμπειρία δεν οδηγούν στο στόχο. Η Ασαφής Λογική αποκλείει τέτοιες δράσεις, εξοικονομώντας υπολογιστικό χρόνο.

1.4 Οργάνωση Κειμένου

Στα πρώτα κεφάλαια γίνεται παρουσίαση του θεωρητικού υποβάθρου που χρησιμοποιήθηκε για την υλοποίηση της διπλωματικής αυτής εργασίας.

Στο Κεφάλαιο 2 γίνεται Κινηματική Ανάλυση για τα ρομπότ διαφορικής οδήγησης. Συνεχίζοντας, στο Κεφάλαιο 3 γίνεται παρουσίαση της έννοιας της Γενικής Αντίληψης για τη σύνθεση δεδομένων αισθητήρων ρομπότ. Στα κεφάλαια 4, 5 και 6 παρουσιάζονται διαδοχικά οι βασικές αρχές για την Ασαφή Λογική, την Ενισχυτική Μάθηση καθώς και για την Ασαφή Ενισχυτική Μάθηση.

Στα υπόλοιπα κεφάλαια παρουσιάζεται το πρακτικό κομμάτι της διπλωματικής εργασίας. Στο κεφάλαιο 7 παρουσιάζεται το ρομπότ Khepera 3, στο οποίο έγινε η υλοποίηση. Διάταξη αισθητήρων, χαρακτηριστικά κινητήρων, διαστάσεις και οδομετρία του ρομπότ. Στο κεφάλαιο 8 αναλύονται τα χαρακτηριστικά των αισθητήρων, μετρήσεις που έγιναν σε αυτούς καθώς και οι περιορισμοί τους.

Συνεχίζοντας, στο κεφάλαιο 9, δίνονται οι λεπτομέρειες της υλοποίησης του Ασαφούς Ελέγχου, ενώ στο κεφάλαιο 10, υλοποιείται ο έλεγχος με Ασαφή Ενισχυτική Μάθηση και παρουσιάζονται τα αριθμητικά αποτελέσματα. Η διπλωματική εργασία κλείνει με το κεφάλαιο συμπερασμάτων και πιθανές μελλοντικές προεκτάσεις με βάση τον έλεγχο με Ασαφή Ενισχυτική Μάθηση.

Κεφάλαιο 2

Κινηματική Ανάλυση

2.1 Κινηματική Ανάλυση Ρομπότ Διαφορικής Οδήγησης

Όχημα διαφορικής οδήγησης, ονομάζεται κάθε όχημα με δύο τροχούς, μη κατευθυντήριους, στον ίδιο άξονα, όπου η κίνηση του καθενός από αυτούς είναι ανεξάρτητη και η φορά κίνησής των δύναται να είναι και η ωρολογιακή και η αντίστροφη αυτής.

Η διαφορική οδήγηση είναι αρκετά διαδεδομένη στα ρομπότ λόγω της απλότητας στην υλοποίησή της αλλά και λόγω της απλότητας στην κινηματική της ανάλυση.

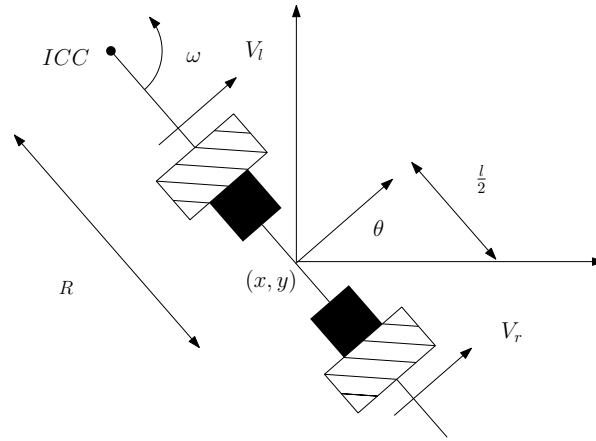
Από τη στιγμή που οι τροχοί των ρομπότ διαφορικής οδήγησης δεν είναι κατευθυντήριοι, η κατευθυντικότητα τους είναι $\delta_s = 0$ ενώ η κινητικότητά τους είναι $\delta_m = 2$ μια και η κίνησή τους γίνεται μόνο στην επιφάνεια που ορίζεται από τους άξονες x, y . Οι βαθμοί ελευθερίας dof_{dd} των ρομπότ διαφορικής οδήγησης είναι άρα :

$$dof_{dd} = \delta_m + \delta_s = 2 + 0 = 2 \quad (2.1)$$

Θα πρέπει να τονιστεί εδώ πως αν και η κίνηση γίνεται στην επιφάνεια που ορίζεται από τους άξονες x, y , παρόλαυτά, το διάνυσμα κίνησης του ρομπότ δεν μπορεί να είναι παράλληλο με τον άξονα που ορίζουν οι 2 τροχοί.

Ανά πάσα στιγμή ορίζεται το Στιγμιαίο Κέντρο Τροχιάς (ICC – Instantaneous Center of Curvature) το οποίο για τα οχήματα διαφορικής οδήγησης βρίσκεται πάντα πάνω στον άξονα που ορίζουν οι 2 τροχοί αλλά όχι απαραίτητα ανάμεσα σε αυτούς. Το ICC αποτελεί τον άξονα γύρω από τον οποίο περιστρέφεται το ρομπότ.

Όπου :



Σχήμα 2.1: Όχημα διαφορικής οδήγησης

ICC : το Στιγμιαίο Κέντρο Τροχιάς του οχήματος

ω : η γωνιακή ταχύτητα με την οποία περιστρέφεται το ρομπότ γύρω από το ICC

R : η απόσταση μεταξύ του ICC και του μέσου της απόστασης μεταξύ των δυο τροχών

l : η απόσταση μεταξύ των δύο τροχών

V_l, V_r : οι ταχύτητες του αριστερού και δεξιού τροχού αντίστοιχα

(x, y) : η θέση του ρομπότ στο σύστημα αξόνων αναφοράς

θ : η γωνία κατα την οποία έχει στραφεί το ρομπότ σε σχέση με το σύστημα αξόνων αναφοράς

Επειδή κάθε στιγμή ο ρυθμός περιστροφής ω και των δυο τροχών, ως προς το ICC είναι ο ίδιος, έχουμε :

$$\omega \left(R + \frac{l}{2} \right) = V_r \quad (2.2)$$

$$\omega \left(R - \frac{l}{2} \right) = V_l \quad (2.3)$$

Αν επιλύσουμε ως R προς και ω αντίστοιχα, προκύπτουν οι ακόλουθες σχέσεις :

$$R = \frac{l}{2} \frac{V_l + V_r}{V_r - V_l} \quad \omega = \frac{V_r - V_l}{l} \quad (2.4)$$

Με βάση αυτές τις σχέσεις, αναφορά πρέπει να γίνει για τις ακόλουθες περιπτώσεις :

- Άν $V_l = V_r$ τότε έχουμε κίνηση προς τα μπροστά. Η απόσταση R γίνεται άπειρη ενώ η γωνιακή ταχύτητα ω μηδενίζεται και συνεπώς δεν έχουμε περιστροφή με το ρομπότ να κινείται σε ευθεία γραμμή.
- Άν $V_l = -V_r$ τότε η απόσταση R μηδενίζεται και τότε το ρομπότ περιστρέφεται γύρω από τον εαυτό του (γύρω από το μέσο του άξονα που ορίζουν οι δύο τροχοί).
- Άν $V_l = 0$ ή $V_r = 0$ τότε έχουμε περιστροφή γύρω από τον αριστερό ή δεξιό τροχό αντίστοιχα.

Με παραμέτρους ελέγχου τις ταχύτητες V_l, V_r μπορούμε να κάνουμε το ρομπότ να κινηθεί προς κάποια επιθυμητή θέση και διεύθυνση.

2.2 Ορθή Κινηματική Ανάλυση Ρομπότ Διαφορικής Οδήγησης

Χρησιμοποιώντας τις σχέσεις (2.4) που δίνουν την απόσταση R και τη γωνιακή ταχύτητα, προκύπτει η θέση του Στιγμαίου Κέντρου Τροχιάς του ρομπότ ICC ως προς τις ταχύτητες V_l, V_r , στο σύστημα αξόνων αναφοράς :

$$ICC = [x - R\sin(\theta), y + R\cos(\theta)] \quad (2.5)$$

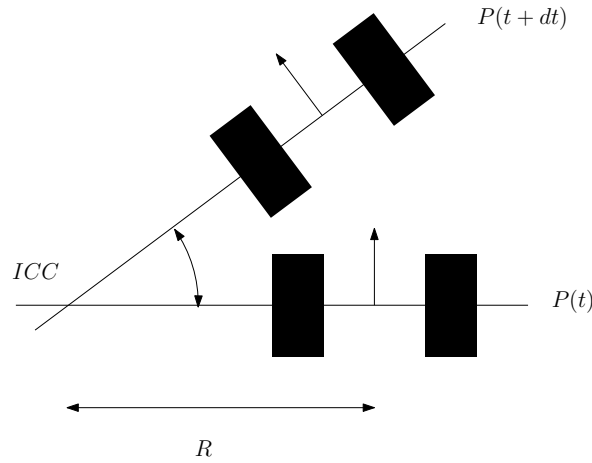
Αν τη στιγμή t το ρομπότ βρίσκεται στη θέση (x, y) και έχει στραφεί κατα γωνία θ σε σχέση με το σύστημα αξόνων αναφοράς, τότε τη στιγμή $t + \delta t$ θα έχει περιστραφεί σε απόσταση R από το ICC με γωνιακή ταχύτητα ω και θα βρίσκεται στη θέση (x', y') ενώ θα έχει στραφεί κατα γωνία θ' σε σχέση με το σύστημα αξόνων αναφοράς ώστε :

$$\begin{pmatrix} x' \\ y' \\ \theta' \end{pmatrix} = \begin{pmatrix} \cos(\omega\delta t) & -\sin(\omega\delta t) & 0 \\ \sin(\omega\delta t) & \cos(\omega\delta t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x - ICC_x \\ y - ICC_y \\ \theta \end{pmatrix} + \begin{pmatrix} ICC_x \\ ICC_y \\ \omega\delta t \end{pmatrix} \quad (2.6)$$

2.3 Ανάστροφη Κινηματική Ανάλυση Ρομπότ Διαφορικής Οδήγησης

Η θέση ενός ρομπότ, ικανού να κινηθεί στην κατεύθυνση Θ_t με ταχύτητα $V(t)$ είναι :

$$\begin{aligned} x(t) &= \int_0^t V(t) \cos[\theta(t)] dt \\ y(t) &= \int_0^t V(t) \sin[\theta(t)] dt \\ \Theta(t) &= \int_0^t \omega(t) dt \end{aligned} \quad (2.7)$$



Σχήμα 2.2: Κίνηση του ρομπότ γύρω από το ICC

Και στην περίπτωση του ρομπότ διαφορικής οδήγησης οι παραπάνω εξισώσεις γίνονται :

$$\begin{aligned} x(t) &= \frac{1}{2} \int_0^t [u_r(t) + u_l(t)] \cos[\theta(t)] dt \\ y(t) &= \frac{1}{2} \int_0^t [u_r(t) + u_l(t)] \sin[\theta(t)] dt \\ \Theta(t) &= \frac{1}{l} \int_0^t [u_r(t) - u_l(t)] dt \end{aligned} \quad (2.8)$$

Στην ανάστροφη κινηματική ανάλυση ενός ρομπότ διαφορικής οδήγησης, αναζητάται τέτοιος έλεγχος ώστε να προκύψει μια συγκεκριμένη θέση και κατεύθυνση (x, y, θ) . Προκύπτει όμως το εξής πρόβλημα : Τα ρομπότ διαφορικής οδήγησης είναι μη ολονομικά, μια και δεν έχουν τη δυνατότητα κίνησης στη διεύθυνση του άξονα των τροχών τους. Συνεπώς δεν μπορούμε να απαιτήσουμε συγκεκριμένη θέση και κατεύθυνση, δίδοντας τις ταχύτητες των τροχών. Πρέπει να οριστεί και η ακολουθούμενη τροχιά.

Το πρόβλημα, παρακάμπτεται, ακολουθώντας στρατηγική δύο κινήσεων. Το ρομπότ κινείται σε ευθεία γραμμή έως ότου φτάσει στην επιθυμητή θέση (x, y) και έπειτα πραγματοποιεί επι τόπου περιστροφή, για να φτάσει την επιθυμητή γωνία θ . οι δυο ανεξάρτητες κινήσεις είναι :

Αν $u_l = u_r = u$ προκύπτει ευθεία κίνηση του ρομπότ και οι εξισώσεις κίνησης γίνονται :

$$\begin{bmatrix} x' \\ y' \\ \theta' \end{bmatrix} = \begin{bmatrix} x + u \sin(\theta) \delta t \\ y + u \cos(\theta) \delta t \\ \theta \end{bmatrix} \quad (2.9)$$

Αν $u_l = -u_r = u$ προκύπτει κίνηση του ρομπότ γύρω από τον άξονά του και οι εξισώσεις κίνησης γίνονται :

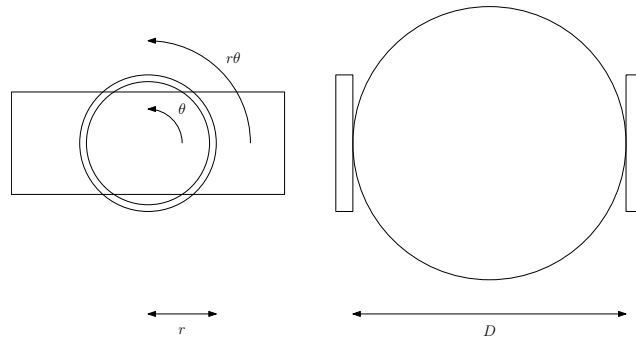
$$\begin{bmatrix} x' \\ y' \\ \theta' \end{bmatrix} = \begin{bmatrix} x \\ y \\ \frac{\theta + 2u\delta t}{l} \end{bmatrix} \quad (2.10)$$

2.4 Οδομετρία Ρομπότ Διαφορικής Οδήγησης

Οι ταχύτητες αριστερού και δεξιού τροχού V_l, V_r που χρησιμοποιήθηκαν παραπάνω, είναι γραμμικές ταχύτητες. Ο έλεγχος των τροχών, γίνεται, ορίζοντας μια γωνιακή ταχύτητα V_{wheel} , για κάθε τροχό, σε ακτίνια ανα δευτερόλεπτο (rad/sec).

Με δεδομένη τη V_{wheel} , πρέπει να βρεθεί η γραμμική ταχύτητα που προκύπτει για τη κίνηση του τροχού. Έστω r_{wheel} η ακτίνα των τροχών του ρομπότ, D_{wheel} το μήκος του άξονα μεταξύ των τροχών και V_{wheel} η γωνιακή ταχύτητα του τροχού σε ακτίνια ανα δευτερόλεπτο.

Σε χρόνο t ο κάθε τροχός θα έχει διανύσει απόσταση :



Σχήμα 2.3: Γεωμετρία ρομπότ διαφορικής οδήγησης

$$D_{wheel} = r_{wheel}V_{wheel}t \quad (2.11)$$

Μπορούμε συνεπώς να εξάγουμε το χρόνο t που απαιτείται, αν οι τροχοί κινούνται προς την ίδια κατεύθυνση με την ίδια ταχύτητα, για ευθεία μετακίνηση του ρομπότ σε κάποια επιθυμητή απόσταση.

Η περίμετρος του ρομπότ είναι :

$$C = \pi D_{wheel} \quad (2.12)$$

Αν οι τροχοί κινούνται προς αντίθετη κατεύθυνση, και το μέτρο της ταχύτητάς τους είναι το ίδιο, το ICC συμπίπτει με το κέντρο του άξονα των τροχών και το ρομπότ περιστρέφεται γύρω από το κέντρο του άξονά του.

2.4. ΟΔΟΜΕΤΡΙΑ ΡΟΜΠΟΤ ΔΙΑΦΟΡΙΚΗΣ ΟΔΗΓΗΣΗΣ

Σε αυτή την περίπτωση για να πραγματοποιηθεί περιστροφή του ρομπότ κατά γωνία ϕ , απαιτείται χρόνος t :

$$t = \frac{\phi C}{2\pi V_{wheel} r_{wheel}} \quad (2.13)$$

Κεφάλαιο 3

Σύνθεση Δεδομένων Αισθητήρων

Ο συνδυασμός δεδομένων, από διαφορετικές αισθητηριακές πηγές, με τέτοιο τρόπο έτσι ώστε η εξαγόμενη πληροφορία να είναι "καλύτερη" σε σχέση με το αν γινόταν διαχείριση για κάθε αισθητηριακή πηγή ξεχωριστά, ονομάζεται *Σύνθεση Αισθητήρων* (Sensor Fusion).

Σε αυτό το κεφάλαιο παρουσιάζονται οι κυριότεροι αλγόριθμοι Σύνθεσης Αισθητήρων καθώς και η έννοια της *Γενικής Αντίληψης* που χρησιμοποιήθηκε στη διπλωματική εργασία για την Σύνθεση Αισθητήρων.

3.1 Κυριότερες Μέθοδοι Σύνθεσης Αισθητήρων

Οι αισθητήρες, ανάλογα με την ποιότητά τους, επηρεάζονται λιγότερο ή περισσότερο, από προβλήματα θορύβου, που επηρεάζουν την αξιοπιστία των μετρήσεων.

Πιθανοτικά μοντέλα είναι τα ως επι το πλείστον χρησιμοποιούμενα για το φιλτράρισμα θορύβου αισθητηριακών διατάξεων, με ανάλογο κόστος σε υπολογιστικό χρόνο για την εξαγωγή του μοντέλου θορύβου, με ότι αυτό συνεπάγεται για συστήματα πραγματικού χρόνου.

Οι κυριότερες μέθοδοι σύνθεσης δεδομένων αισθητήρων στη βιβλιογραφία, είναι οι ακόλουθες :

- *Κεντρικό Οριακό Θεώρημα* (Central Limit Theorem) : Στη θεωρία πιθανοτήτων, σύμφωνα με το κεντρικό οριακό θεώρημα, κάτω από συγκεκριμένες συνθήκες, ο αριθμητικός μέσος ενός αρκετά μεγάλου αριθμού δειγμάτων, ανεξάρτητων τυχαίων μεταβλητών, με καλά ορισμένες αναμενόμενες τιμές και καλά ορισμένη διακύμανση, θα είναι περίπου κανονικά κατανομημένος.
- *Φίλτρο Kalman* (Kalman Filter) : Είναι ένας αλγόριθμος που χρησιμοποιεί μια σειρά μετρήσεων, που παρατηρούνται με την εξέλιξη του χρόνου και επηρεάζονται από θόρυβο και εκτιμά τις άγνωστες μεταβλητές θορύβου, για τη σειρά μετρήσεων, που τείνουν να είναι πιο ακριβείς, σε σχέση με μια και μόνη μέτρηση. Το φίλτρο Kalman, λειτουργεί αναδρομικά.
- *Δίκτυο Bayes* (Bayesian Network) : Είναι ένα πιθανοτικό γραφικό μοντέλο που αναπαριστά ένα σύνολο τυχαίων μεταβλητών και των υπο όρους εξαρτήσεων τους μέσω ενός κατευθυνόμενου ακυκλικού γράφου.
- *Θεωρία Dempster-Shafer* (Dempster-Shafer Theory) : Επιτρέπει το συνδυασμό στοιχείων, από διαφορετικές πηγές και φτάνει σε ένα βαθμό πεποίθησης (που αντιστοιχεί σε μια συνάρτηση πεποίθησης), για όλα τα διαθέσιμα στοιχεία.

3.2 Η Μέθοδος της Γενικής Αντίληψης

Η μέθοδος της Γενικής Αντίληψης, όπως περιγράφεται από τους Reinhard Braunstingl και Pedro Sanz και Jose Manuel Ezkerra [8], δεν προσπαθεί να μοντελοποιήσει το περιβάλλον. Αντίθετα, επιχειρεί την κατασκευή ενός γενικού μοντέλου του χώρου, από τις μετρήσεις, αναπαριστώντας το με ένα διάνυσμα που ονομάζεται Διάνυσμα Γενικής Αντίληψης.

Σε αντίθεση με τους περίπλοκους υπολογισμούς των πιθανοτικών μοντέλων και με δεδομένο πως οι αισθητήρες στο πρόβλημα της ακολούθησης τοίχου έχουν πολλές διαφορετικές παραμέτρους θορύβου (επιφάνεια τοίχου, γωνία αντικειμένου, ατέλειες στην τοποθέτηση των αισθητήρων, θόρυβος λόγω κατασκευαστικής μεθόδου των αισθητήρων), η μέθοδος της Γενικής Αντίληψης παίρνει τα δεδομένα ως έχουν, αγνοώντας τον θόρυβο και τα συνδυάζει, προσπαθώντας να αποκτήσει μια αξιόπιστη αναπαράσταση του χώρου, σε πραγματικό χρόνο (χρόνος ενημέρωσης των αισθητήρων).

Σε κάθε αισθητήρα υπερύθρων i του ρομπότ, αντιστοιχείται ένα διάνυσμα αντίληψης p_i . Η διεύθυνσή του αντιστοιχεί στη διεύθυνση του άξονα του αισθητήρα και το μέτρο του είναι συνάρτηση της μετρούμενης απόστασης από αυτόν τον αισθητήρα :

$$p_i = \frac{d_{max} - d_i}{d_{max} - d_{min}} \quad (3.1)$$

Όπου d_{min} και d_{max} είναι η ελάχιστη και μέγιστη απόσταση, αντίστοιχα, για τις οποίες ένα αντικείμενο μπορεί να ανιχνευθεί αξιόπιστα.

Το p_i έχει τιμές μεταξύ 0 και 1 :

$$p_i = \begin{cases} 0 & \text{αν } d_i > d_{max} \\ 1 & \text{αν } d_i < d_{min} \end{cases} \quad (3.2)$$

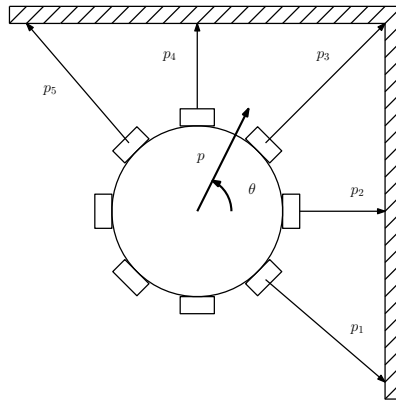
Το διάνυσμα αντίληψης, ομοιάζει στα ιστογράμματα διανυσματικών πεδίων, μα είναι συνδεδεμένο με τον αισθητήρα, αντί για ένα κελί ενός πλέγματος. Το συνολικό διάνυσμα αντίληψης, p αποτελείται από όλα τα επιμέρους διανύσματα αντίληψης των αισθητήρων p_i , η διεύθυνσή του ισούται με το άθροισμα των διευθύνσεων των αισθητήρων και το μέτρο του ισούται με τη μέγιστη αισθητηριακή αντίληψη :

$$p = p_{i,max} \frac{\sum p_i}{|\sum p_i|} \quad (3.3)$$

Στο σχήμα αναπαρίσταται η λειτουργία της γενικής αντίληψης στην περίπτωση ακολούθησης τοίχου :

Η μέθοδος της γενικής αντίληψης, προσφέρεται για γλωσσική περιγραφή καταστάσεων στις οποίες μπορεί να βρεθεί ένα ρομπότ και συνεργεί εξαιρετικά με την ασαφή λογική. Ο συνδυασμός προσομοιάζει στον τρόπο διαχείρισης του προβλήματος ακολούθησης τοίχου από τον άνθρωπο.

Στο σχήμα αναπαρίσταται η λειτουργία της γενικής αντίληψης στην περίπτωση ακολούθησης τοίχου :



Σχήμα 3.1: Γενική Αντίληψη p

Η μέθοδος της γενικής αντίληψης, προσφέρεται για γλωσσική περιγραφή καταστάσεων στις οποίες μπορεί να βρεθεί ένα ρομπότ και συνεργεί εξαιρετικά με την ασαφή λογική. Ο συνδυασμός προσομοιάζει στον τρόπο διαχείρισης του προβλήματος ακολούθησης τοίχου από τον άνθρωπο.

Κεφάλαιο 4

Ασαφής Λογική

Η ασαφής λογική αποτελεί πλέον, μια καθιερωμένη μέθοδο ελέγχου διεργασιών και συστημάτων όχι μόνο σε ερευνητικό αλλά και σε εφαρμοσμένο επίπεδο. Η περιγραφή της έγινε για πρώτη φορά συγχρόνως από τον *Lotfi Zadeh* [32] και από τον *Dieter Klaua* [18], και βασίζεται στη θεωρία των ασαφών συνόλων, με βάση την οποία, η περιγραφή συνόλων μπορεί να γίνει με πιο ασαφή κριτήρια συγκριτικά με την κλασική θεωρία συνόλων.

Η διάδοσή της οφείλεται στην απλότητά της, στις δυνατότητες που παρέχει για μοντελοποίηση μη γραμμικών συναρτήσεων, αλλά και στο γεγονός πως οι κανόνες της μπορούν να περιλαμβάνουν πληροφορίες προερχόμενες από την ανθρώπινη εμπειρία και μάλιστα, η περιγραφή των κανόνων αυτών, προσομοιάζει στην ανθρώπινη φυσική γλώσσα.

Στο κεφάλαιο αυτό περιγράφονται οι βασικές αρχές της ασαφούς λογικής και των ασαφών συστημάτων, των συστημάτων δηλαδή που ελέγχονται με βάση αυτή.

4.1 Ασαφή Συνολα

Στην κλασική θεωρία συνόλων, ένα στοιχείο είτε ανήκει είτε δεν ανήκει σε ένα σύνολο. Τα σύνολα πρέπει να είναι πολύ συγκεκριμένα, ώστε να περιγράφουν όσο το δυνατόν καλύτερα, τα στοιχεία που καλούνται να περιλάβουν. Το κατά πόσο ένα στοιχείο αποτελεί μέρος ενός συνόλου, στηρίζεται σε μια καθαρά δυαδική λογική, αποκλείοντας μερική επαλήθευση ενός συνόλου από ένα στοιχείο.

Η θεωρία ασαφών συνόλων, εισάγει στα στοιχεία μια περιγραφή για το κατά πόσο αυτά αποτελούν μέρη ενός συνόλου. Το σύνολο δεν είναι ανάγκη να περιγράφει πλήρως το στοιχείο που περικλείει. Το στοιχείο πλέον περιγράφει, σε πιο βαθμό το ίδιο ικανοποιεί τις συνθήκες του συνόλου. Δίνεται λοιπόν η δυνατότητα για ορισμό συνόλων τα οποία δεν έχουν τα σαφή όρια που απαιτεί η κλασική θεωρία συνόλων και από αυτό ακριβώς το γεγονός προκύπτει και η ονομασία.

Αν X είναι το σύνολο όλων των πιθανών εισόδων για ένα σύστημα και x καθένα από τα στοιχεία αυτού, τότε ορίζεται ως ασαφές σύνολο FS με βάση το X , το σύνολο των στοιχείων $(x, \mu_{FS}(x))$ τέτοιο ώστε :

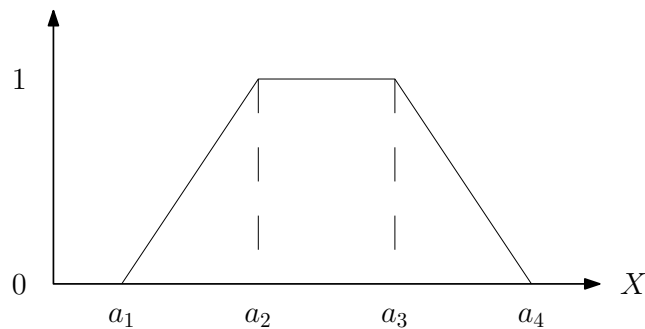
$$FS = \{x, \mu_{FS}(x) | x \in X, 0 \leq \mu_{FS}(x) \leq 1\} \quad (4.1)$$

Όπου $\mu_{FS}(x)$ μια τιμή που ονομάζεται τιμή συμμετοχής του x στο σύνολο FS και που περιγράφει το κατά πόσο το στοιχείο x ανήκει στο ασαφές σύνολο. Η τιμή αυτή ορίζεται από τη συνάρτηση μ_{FS} , που ονομάζεται συναρτηση συμμετοχής και ουσιαστικά αντιστοιχίζει κάθε στοιχείο του συνόλου X σε μια τιμή ανάμεσα στο 0 και το 1.

Η μορφή της συνάρτησης μ_{FS} ποικίλει. Στα ασαφή συστήματα συναντάται συχνότερα η τραπεζοειδής μορφή συνάρτησης συμμετοχής και η υποπερίπτωσή της, της τριγωνικής, ενώ απαντώνται επίσης η γκαουσιανή μορφή, η γενικευμένη καμπανοειδής μορφή και η πολυω-
νυμική μορφή.

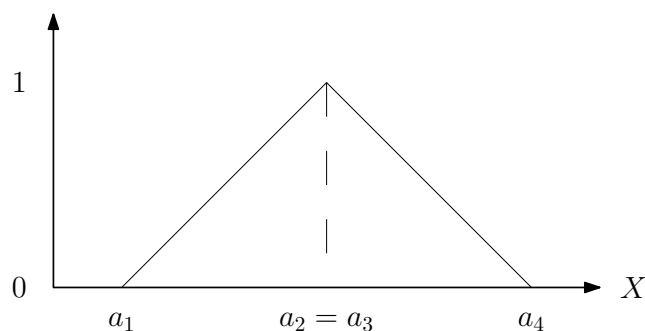
Η τραπεζοειδής μορφή, περιγράφεται από τον τύπο :

$$\mu_{FS}(x) = \begin{cases} \max(0, \frac{\alpha_4 - x}{\alpha_4 - \alpha_3}), & x > \alpha_3 \\ \max(0, \frac{x - \alpha_1}{\alpha_2 - \alpha_1}), & x < \alpha_2 \\ 1 & \text{διαφορετικά} \end{cases} \quad (4.2)$$



Σχήμα 4.1: Τραπεζοειδής συνάρτηση συμμετοχής

Είναι προφανές πως για την περίπτωση όπου $\alpha_2 = \alpha_3$ η τραπεζοειδής μορφή μετατρέπεται σε τριγωνική.



Σχήμα 4.2: Τριγωνοειδής συνάρτηση συμμετοχής

4.2 Ασαφείς Μεταβλητές

Οι ασαφείς μεταβλητές, αποτελούν ένα τρόπο περιγραφής συνόλων με γλωσσικούς όρους, παρόμοιους με αυτούς που χρησιμοποιούνται στην ανθρώπινη ομιλία. Λόγω αυτού του γεγονότος, αναφέρονται στη βιβλιογραφία και σαν γλωσσικές μεταβλητές.

Ασαφής μεταβλητή X_f ονομάζεται κάθε ασαφής αναπαράσταση του συνόλου X με γλωσσικούς όρους προσδίδοντας μια ποιοτική περιγραφή στα στοιχεία. Για την ασαφή μεταβλητή X_f μπορούν να οριστούν πολλαπλοί γλωσσικοί όροι που αντιστοιχούν σε ασαφή υποσύνολα του X .

Αν L_i είναι οι n το πλήθος ($i = 1, 2, \dots, n$) γλωσσικοί όροι μιας ασαφούς μεταβλητής X_f τότε κάθε γλωσσικός όρος αντιστοιχεί σε ένα ασαφές υποσύνολο και όλοι μαζί περιγράφουν το ασαφές σύνολο X_f :

$$L_i = \{x, \mu_{L_i}(x) \mid x \in X, 0 \leq \mu_{L_i}(x) \leq 1\} \mid 1 \leq i \leq n \quad (4.3)$$

$$X_f = L_i \mid 1 \leq i \leq n \quad (4.4)$$

Οι γλωσσικοί όροι αποτελούν ουσιαστικά μια μέθοδο κατάτμησης του συνόλου X η οποία ονομάζεται ασαφής κατάτμηση. Τα είδη ασαφών κατατμήσεων είναι δύο και η διάκρισή τους εξαρτάται από το αν, για τις συναρτήσεις συμμετοχής που αντιστοιχούν σε κάθε ένα από τα στοιχεία, ισχύει :

$$\sum_{i=1}^n \mu_{L_i}(x) = 1, \quad \forall x \in X \quad (4.5)$$

- Η Κανονική ή αλλιώς αδύναμη κατάτμηση : Σε αυτή την περίπτωση, δεν υπάρχει περιορισμός ως προς το πως θα γίνει η κατάτμηση του συνόλου των στοιχείων. Οι συναρτήσεις συμμετοχής που περιγράφουν τους γλωσσικούς όρους δεν είναι απαραίτητο να καλύπτουν όλο το πλήθος των στοιχείων του συνόλου και η Σχέση (4.5) δεν είναι απαραίτητο να ισχύει, το άθροισμα μπορεί να είναι μικρότερο της μονάδας.
- Η Ισχυρή κατάτμηση : Σε αυτή τη περίπτωση το σύνολο των στοιχείων πρέπει να περιγράφεται από κάποιο ή κάποιους γλωσσικούς όρους. Η Σχέση (4.5) πρέπει να ισχύει.

Η ισχυρή κατάτμηση εγγυάται πως όλα τα στοιχεία του συνόλου περιγράφονται από κάποια ή κάποιες συναρτήσεις συμμετοχής με αποτέλεσμα την πιο εύκολη διαχείρισή τους για την υλοποίηση του επιθυμητού ελέγχου. Στις περισσότερες πρακτικές εφαρμογές, προτιμάται η ισχυρή κατάτμηση.

4.3 Ασαφείς Τελεστές

Η ασαφής λογική αποτελεί ουσιαστικά μια επέκταση της *Boolean* λογικής με τη διαφορά πως αντί για τις δυαδικές τιμές 0 και 1, χρησιμοποιεί και τις ενδιάμεσες τιμές αυτών για την περιγραφή στοιχείων.

Για την τέλεση λογικών ασαφών πράξεων θα πρέπει συνεπώς να εκφραστεί ένα σύνολο ασαφών λογικών τελεστών, βασισμένο στους τελεστές *OXI*, *KAI*, *Ή* της άλγεβρας *Boole* τέτοιο ώστε να μπορεί να διαχειριστεί τις ασαφείς μεταβλητές. Οι ασαφείς αυτοί λογικοί

τελεστές, θα αποτελούν ένα υπερσύνολο των κλασικών τελεστών.

Για δύο ασαφή σύνολα A και B , όπου :

$$A = \{x, \mu_A(x) \mid x \in X, 0 \leq \mu_A(x) \leq 1\} \quad (4.6)$$

$$B = \{x, \mu_B(x) \mid x \in X, 0 \leq \mu_B(x) \leq 1\} \quad (4.7)$$

ορίζονται οι ακόλουθοι ασαφείς τελεστές, κατα παρόμοιο τρόπο με την λογική Boolean :

- Τελεστής *OXI* (ασαφές συμπλήρωμα συνόλου, η μορφή είναι ίδια και για το σύνολο B) :

$$OXI \ A = A' = \{x, \mu_{A'}(x) \mid x \in X, \mu_{A'}(x) = 1 - \mu_A(x)\} \quad (4.8)$$

$$OXI \ B = B' = \{x, \mu_{B'}(x) \mid x \in X, \mu_{B'}(x) = 1 - \mu_B(x)\} \quad (4.9)$$

- Τελεστής *KAI* (Ασαφής επαλήθευση και των δύο ασαφών συνόλων) :

$$A \ KAI \ B = A \cap B = T_{\text{νόρμα}}(\mu_A(x), \mu_B(x)) \quad (4.10)$$

- Τελεστής *Ή* (Ασαφής επαλήθευση ενός από τα δύο ή και των δύο ασαφών συνόλων) :

$$A \ Ή \ B = A \cup B = S_{\text{νόρμα}}(\mu_A(x), \mu_B(x)) \quad (4.11)$$

Όπου $S_{\text{νόρμα}}(\dots, \dots)$ είναι ένας κατάλληλος τελεστής δυαδικής αντιστοίχισης, που πρέπει να ικανοποιεί ένα σύνολο συνθηκών που περιγράφεται παρακάτω και το οποίο διαφέρει από το σύνολο συνθηκών της $T_{\text{νόρμα}}(\dots, \dots)$.

Ως $T_{\text{νόρμα}}(\dots, \dots)$ ορίζεται κάθε δυαδικός τελεστής αντιστοίχισης ο οποίος ικανοποιεί :

- Τα όρια τιμών : $T(0, 0) = 0, T(a, 1) = T(1, a) = a$
- Τη μονοτονία : $T(a, b) \leq T(c, d)$ αν $a \leq c$ και $b \leq d$
- Την αντιμεταθετικότητα : $T(a, b) = T(b, a)$
- Την προσεταιριστικότητα : $T(a, T(b, c)) = T(T(a, b), c)$

Ως $S_{\text{νόρμα}}(\dots, \dots)$ ορίζεται κάθε δυαδικός τελεστής αντιστοίχισης ο οποίος ικανοποιεί :

- Τα όρια τιμών : $S(1, 1) = 1, S(a, 0) = S(0, a) = a$
- Τη μονοτονία : $S(a, b) \leq S(c, d)$ αν $a \leq c$ και $b \leq d$
- Την αντιμεταθετικότητα : $S(a, b) = S(b, a)$
- Την προσεταιριστικότητα : $S(a, S(b, c)) = S(S(a, b), c)$

a	b	$\neg b$	$a \wedge b$	$a \vee b$	$a \downarrow b$	$a \oplus b$	$a \implies b$	$a \uparrow b$
0	0	1	0	0	1	0	1	1
0	1	0	0	1	0	1	1	1
1	0		0	1	0	1	0	1
1	1		1	1	0	0	1	0

Πίνακας 4.1: Πίνακας αληθείας ασαφών λογικών τελεστών για δυαδικά στοιχεία

Οι πλέον συνηθισμένες $T_{\nu\omicron\rho\mu\alpha}(\dots, \dots)$ και $S_{\nu\omicron\rho\mu\alpha}(\dots, \dots)$ στα ασαφή συστήματα είναι οι συναρτήσεις ελαχίστου και μεγίστου αντίστοιχα και ήταν οι συναρτήσεις που επιλέχθηκαν και στη διπλωματική αυτή εργασία. Αυτό δε σημαίνει σε καμία περίπτωση βέβαια, πως αποτελούν και τις μοναδικές, αλλά η απλότητά τους και ο εύκολος υπολογισμός τους τις καθιστά ιδιαίτερα δημοφιλείς.

Στον Πίνακα 4.1 φαίνονται οι πίνακες αληθείας για τους ασαφείς λογικούς τελεστές, στην περίπτωση δυαδικών στοιχείων. Οι πίνακες αυτοί είναι ίδιοι με τους πίνακες αληθείας, της κλασικής άλγεβρας Boole.

4.4 Ασαφείς Κανόνες και Ασαφής Λογική

Οι ασαφείς κανόνες αποτελούν το κύριο και πλέον χαρακτηριστικό γνώρισμα της ασαφούς λογικής.

Έστω ένας κανόνας, όπου προϋπόθεση για να ισχύει το B αποτελεί το να ισχύει το A :

$$AN\ A\ TOTE\ B$$

Με βάση την παραδοσιακή λογική, προϋπόθεση για να συμβαίνει το B είναι να επαληθεύεται το A (Modus Ponens). Στην ασαφή λογική η επαλήθευση αυτού του κανόνα, μπορεί να είναι και μερική, το B δηλαδή να ισχύει, με όποιο βαθμό αλήθειας ισχύει το A .

Ας υποθέσουμε τώρα πως το x είναι A' και πως ο κανόνας είναι AN το x είναι A $TOTE$ το y είναι B (Modus Tollens). Αν υποθέσουμε πως υπάρχει ένα ασαφές σύνολο B' ορισμένο προσεγγιστικά στο σύνολο B τότε, με την προϋπόθεση πως το ασαφές σύνολο A' είναι ορισμένο προσεγγιστικά στο σύνολο A , τότε μπορεί να εξαχθεί το συμπέρασμα πως το " y είναι B' ". Ορίζεται λοιπόν η ασαφής λογική, ως :

$$(x \text{ είναι } A' \text{ και } A \rightarrow B) \rightarrow (B \text{ είναι } B') \text{ τέτοιο ώστε :}$$

$$\mu_{B'}(y) = \max_x \min[\mu_{A'}(x), \mu_R(x, y)] \tag{4.12}$$

Όπου $\mu_R(x, y)$ είναι η ασαφής συνάρτηση συμμετοχής συνεπαγωγής (\rightarrow).

Η εξίσωση 4.12 μπορεί να επαναδιατυπωθεί σε πιο γενική μορφή :

$$B' = A' \circ R = A' \circ (A \rightarrow B) \tag{4.13}$$

Με βάση τα παραπάνω, για την εξαγωγή της ασαφούς συνάρτησης συμμετοχής συσχέτισης, θα πρέπει αρχικά να βρεθούν τα κοινά σημεία της με την $\mu_{A'}(x)$ (με τη χρήση του τελεστή ελαχίστου) και έπειτα να υπολογιστεί η προβολή της $\mu_{B'}(x)$ στο επίπεδο z-y.

Για τον υπολογισμό της ασαφούς συνάρτησης συμμετοχής συσχέτισης $\mu_R(x, y)$, υποθέτεται πως η συσχέτιση είναι τύπου, αιτίου αιτιατού μεταξύ του A και του B και συνεπώς μπορεί να χρησιμοποιηθεί η $T_{\text{νόρμα}}$ για τον υπολογισμό του $\mu_R(x, y)$:

$$\mu_R(x, y) = T_{\text{νόρμα}}(\mu_A(x), \mu_B(y)) \quad (4.14)$$

Συνεπώς, η διαδικασία ερμηνείας ενός ασαφούς (AN-TOTE) κανόνα, συνοψίζεται σε 3 βήματα :

1. Ασαφοποίηση, που σημαίνει την αντιστοίχιση των ενεργοποιημένων μεταβλητών εισόδου στην προϋπόθεση ισχύος του κανόνα, σύμφωνα με το βαθμό συμμετοχής τους
2. Εφαρμογή των ασαφών τελεστών στην προϋπόθεση ισχύος του κανόνα, ώστε να επιλυθούν όλες οι ασαφοποιημένες τιμές σε μία τιμή που ονομάζεται βαθμός επαλήθευσης του κανόνα, εφαρμόζοντας τους τελεστές $T_{\text{νόρμα}}$ και $S_{\text{νόρμα}}$
3. Εφαρμογή της $T_{\text{νόρμα}}$ στο βαθμό επαλήθευσης του κανόνα και το συμπέρασμα του κανόνα.

4.5 Αποασαφοποίηση

Αποασαφοποίηση είναι η διαδικασία ποσοτικοποίησης του αποτελέσματος ενός ασαφούς συστήματος. Υπάρχουν τρεις διαδεδομένες μέθοδοι, για τη μετατροπή από τη γλωσσική μεταβλητή που μας δίνει ένα ασαφές σύστημα σαν έξοδο, σε μια τιμή.

4.5.1 Μέθοδος Μέσου Μεγίστων (MOM)

Η μέθοδος του Μέσου Μεγίστων, υπολογίζει τον μέσο όρο των ασαφών συμπερασμάτων με τις μέγιστες τιμές αλήθειας, αγνοώντας το σχήμα της συνάρτησης συμμετοχής εξόδου.

Αυτό το χαρακτηριστικό της είναι και το μειονέκτημά της, μια και για συναρτήσεις συμμετοχής διαφορετικού σχήματος, μπορεί να δώσει το ίδιο αποτέλεσμα, αν αυτές έχουν τις ίδιες μέγιστες τιμές αλήθειας.

4.5.2 Μέθοδος Κέντρου Βαρύτητας (COG)

Η μέθοδος του Κέντρου Βαρύτητας, αποτελεί την πλέον δημοφιλή τεχνική αποασαφοποίησης και προσομοιάζει στην τεχνική εύρεσης του κέντρου βαρύτητας στη φυσική. Ο σταθμισμένος μέσος της συνάρτησης συμμετοχής, υπολογίζεται με βάση το εμβαδό που περικλείει η συνάρτηση συμμετοχής εξόδου.

4.5.3 Μέθοδος Ύψους (HM)

Η μέθοδος Ύψους εφαρμόζεται μόνο στην περίπτωση, όπου η συνάρτηση συμμετοχής εξόδου είναι τομή συμμετρικών συναρτήσεων. Ολοκληρώνεται σε δύο βήματα. Πρώτα η συνάρτηση συμμετοχής εξόδου, μετατρέπεται σε τιμή χρησιμοποιώντας τη μέθοδο του Κέντρου

Βαρύτητας και έπειτα εφαρμόζεται σταθμισμένος μέσος για τις τιμές αληθείας των ενεργοποιημένων κανόνων. Πλεονέκτημα αυτής της μεθόδου είναι η απλότητά της και χρησιμοποιείται κατά κόρον σε υλοποιήσεις νευρο-ασαφών μοντέλων για μείωση των απαιτούμενων υπολογισμών.

4.6 Ασαφές Συμπερασματικό Σύστημα

Τα ασαφή συμπερασματικά συστήματα αποτελούνται από ένα σύνολο κανόνων (υπό-σύνολο αυτών των κανόνων, αποτελούν ένα σύνολο (AN-TOTE) κανόνων) που προσπαθούν να μιμηθούν την ανθρώπινη εμπειρική διαδικασία λήψης αποφάσεων.

Ός Ασαφές Συμπερασματικό Σύστημα (Fuzzy Inference System), ορίζεται ένα σύστημα, για την εξαγωγή συμπερασμάτων, χρησιμοποιώντας ασαφείς εισόδους στις οποίες εφαρμόζονται ασαφείς κανόνες. Ασαφές συμπέρασμα ονομάζεται η διαδικασία χαρτογράφησης της συνεπαγωγής μιας οιασδήποτε εισόδου σε μια έξοδο, χρησιμοποιώντας την ασαφή λογική. Η χαρτογράφηση αυτή δίνει μια βάση, στην οποία μπορεί να στηριχθεί η λήψη αποφάσεων.

Υπάρχουν δύο τύποι ασαφών συμπερασματικών συστημάτων που μπορούν να υλοποιηθούν : τα τύπου-Mamdani(M-FIS) [20] και τα τύπου-Takagi-Sugeno(TS-FIS) [30]. Η διαφορά τους έγκειται στον τρόπο εξαγωγής των εξόδων. Για τις ανάγκες της διπλωματικής αυτής εργασίας χρησιμοποιήθηκε ένα TS-FIS, λόγω της δυνατότητας που παρέχει για αναπαράσταση των συμπερασμάτων σαν συναρτήσεις των εισόδων, σε αντίθεση με τον τύπο Mamdani που προϋποθέτει γλωσσικό ορισμό των εξόδων.

4.7 Ασαφές Συμπερασματικό Σύστημα Takagi-Sugeno (TS-FIS)

Οι πλέον συνήθεις τύποι TS-FIS είναι οι μηδενικής και πρώτης τάξεως. Στη μορφή μηδενικής τάξης, οι συναρτήσεις συμμετοχής εξόδου είναι μονήρεις αιχμές. ενώ στη μορφή πρώτης τάξης, οι συναρτήσεις συμμετοχής εξόδου, είναι ένας γραμμικός συνδυασμός των μεταβλητών εισόδου.

Η διαφορά των δύο μορφών έγκειται στο συμπερασματικό κομμάτι των κανόνων. Δεδομένων των υπολογιστικών απαιτήσεων της μορφής πρώτης τάξης αλλά και της πολυπλοκότητας που εισάγει στο όλο σύστημα, για τις ανάγκες της παρούσας εργασίας, επιλέχθηκε η μορφή μηδενικής τάξης.

Οι κανόνες ενός ασαφούς συμπερασματικού συστήματος, αποτελούνται από ένα αυθαίρετα επιλεγμένο αριθμό κανόνων (ικανό να ανταποκριθεί στις ανάγκες τις εκάστοτε περίπτωσης), διαφορετικών μορφών, που αποτελούνται από συνδυασμούς των τελεστών ΚΑΙ και Η.

Έστω πως ένα ασαφές συμπερασματικό σύστημα έχει ένα σύνολο από N κανόνες και πως το διάνυσμα εισόδου έχει n στοιχεία. Τότε ο κάθε κανόνας μπορεί να γραφεί ως :

Διάνυσμα εισόδου : $X = (x_1, \dots, x_n)$

Κανόνας i : AN x_1 είναι L_1^i ΚΑΙ x_2 είναι L_2^i ΚΑΙ ... ΚΑΙ x_n είναι L_n^i TOTE y είναι o^i

Όπου $L_j^i, j = 1, \dots, n, i = 1, \dots, N$ είναι η γλωσσική περιγραφή της μεταβλητής εισόδου x_j η οποία συμμετέχει στον i κανόνα. ενώ o^i είναι η σταθερή μονήρης έξοδος του i κανόνα.

Ο αριθμός των κανόνων θα είναι :

$$N = \prod_{i=1}^n N_{L_i} \quad (4.15)$$

Με $N_{L_i}, i = 1, \dots, n$ να είναι ο αριθμός των ασαφών συνόλων (γλωσσικές περιγραφές) της μεταβλητής εισόδου x_i .

Οι κανόνες, των οποίων οι προϋποθέσεις έχουν μη μηδενικές τιμές συμμετοχής, κατά την εφαρμογή της εισόδου X , ονομάζονται ενεργοποιημένοι κανόνες, το δε σύνολο των ενεργοποιημένων κανόνων αναφέρεται ως $A(X)$. Το μέγιστο πλήθος ενεργοποιημένων κανόνων $A(X)$, για κάποια είσοδο X , στην περίπτωση ισχυρής ασαφούς κατάτμησης είναι 2^n .

Μετά την ασαφοποίηση του διανύσματος εισόδου και τον υπολογισμό των συναρτήσεων συμμετοχής, ακολουθεί ο υπολογισμός των τιμών αληθείας των κανόνων a_{R_i} με εφαρμογή του τελεστή $T_{\text{νόρμα}}$, οπότε προκύπτει :

$$a_{R_i}(X) = \prod_{j=1}^n \mu_{L_j^i}(x_j), \quad \forall R_i \in A(X) \quad (4.16)$$

Με $\mu_{L_j^i}$ να είναι η συνάρτηση συμμετοχής που αντιστοιχεί στο L_j^i .

Οι έξοδοι πλήρους ενεργοποίησης των κανόνων, μπορούν να αναπαρασταθούν σαν ασαφή σύνολα τα οποία περιέχουν μια μόνο τιμή του διαστήματος τιμών της μεταβλητής εξόδου, στη μορφή μοναδιαίας συνάρτησης :

$$\mu_{o^i}(y) \begin{cases} 1 & y = o^i \\ 0 & \text{διαφορετικά} \end{cases} \quad (4.17)$$

Στη φάση εξαγωγής συμπερασμάτων, πολλαπλασιάζοντας τις εξόδους πλήρους ενεργοποίησης των κανόνων με τις τιμές αληθείας των ενεργοποιημένων κανόνων, έχουμε συναρτήσεις ίσες με τις τιμές αληθείας για τους ενεργοποιημένους κανόνες και μηδενικές για τους μη ενεργοποιημένους.

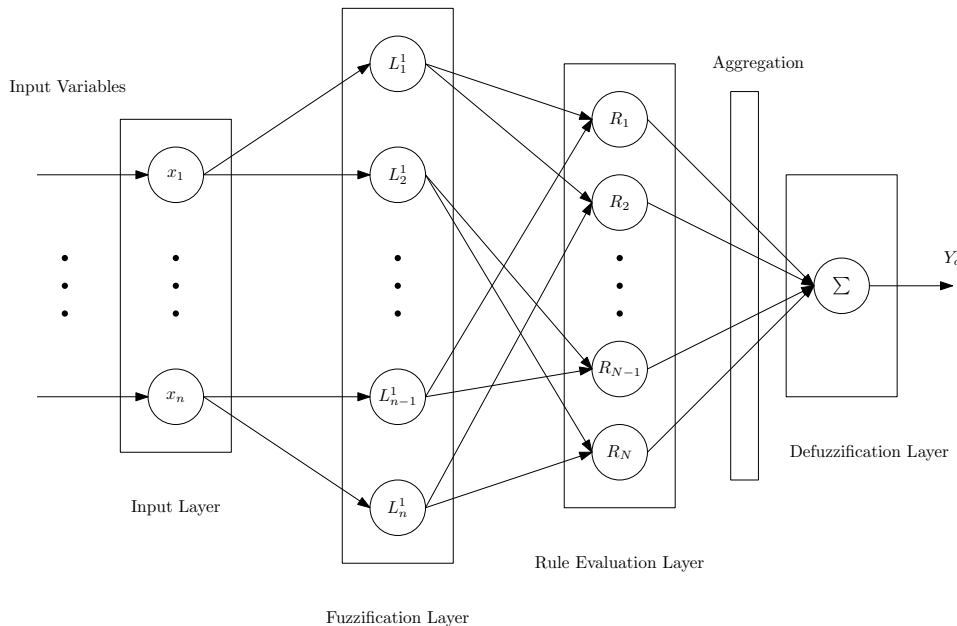
Μπορούμε πλέον, έχοντας υπολογίσει το βάρος όλων των ενεργοποιημένων κανόνων, να υπολογίσουμε την έξοδο του ασαφούς συμπερασματικού συστήματος $Y_o(X)$ που θα είναι ίση με το σταθμισμένο μέσο όρο των εξόδων των ενεργοποιημένων κανόνων (με βάση τη μέθοδο Κέντρου Βαρύτητας για την αποασαφοποίηση) και ονομάζεται "Μοντέλο Πολλαπλών Εισόδων Μιας Εξόδου":

$$Y_o(X) = \frac{\sum_{\forall R_i \in A(X)} a_{R_i}(X) o^i}{\sum_{\forall R_i \in A(X)} a_{R_i}} \quad (4.18)$$

4.8. ΣΥΓΚΡΙΣΗ (M-FIS) ΚΑΙ (TS-FIS)

Συνοψίζοντας, η λειτουργία ενός Ασαφούς Συμπερασματικού Συστήματος Takagi-Sugeno πραγματοποιείται σε 5 βήματα :

1. Ασαφοποίηση : Ασαφοποίηση των μεταβλητών εισόδου, δηλαδή αναζήτηση του σε ποιό βαθμό οι εισοδοί ανήκουν σε κάποιο από τα ορισμένα ασαφή σύνολα.
2. Αξιολόγηση των κανόνων : Εφαρμογή των ασαφών τελεστών (KAI και \acute{H}) στις προϋποθέσεις των κανόνων, χρησιμοποιώντας τις $T_{\nu\acute{o}\rho\mu\alpha}$ και $S_{\nu\acute{o}\rho\mu\alpha}$ αντίστοιχα.
3. Συμπέρασμα των κανόνων : Χρησιμοποιώντας την $T_{\nu\acute{o}\rho\mu\alpha}$ γίνεται η σύνδεση προϋπόθεσης και συμπεράσματος για τον κάθε κανόνα. Για μηδενικής τάξης σύστημα, αυτό απλώς σημαίνει τον πολλαπλασιασμό του βαθμού αληθείας του κάθε κανόνα με τη μονάδα.
4. Συνδυασμός των συμπερασμάτων όλων των κανόνων : Είναι η διαδικασία ενοποίησης των συμπερασμάτων όλων των ενεργοποιημένων κανόνων.
5. Αποασαφοποίηση : Λήψη μιας αντιπροσωπευτικής εξόδου για το σύστημα, με χρήση της μεθόδου Κέντρου Βαρύτητας, για όλους τους κανόνες.



Σχήμα 4.3: Αρχιτεκτονική Ασαφούς Συμπερασματικού Συστήματος Takagi-Sugeno (TS-FIS)

4.8 Σύγκριση (M-FIS) και (TS-FIS)

Η κύρια διαφορά μεταξύ των Mamdani(M-FIS) και Takagi-Sugeno(TS-FIS) είναι πως οι συναρτήσεις συμμετοχής εξόδου, είναι σταθερές ή γραμμικές για τα Takagi-Sugeno(TS-FIS) ενώ για το Mamdani(M-FIS) είναι ασαφή σύνολα. Το τελευταίο απαιτεί υπολογιστική ισχύ για το συνδυασμό των συμπερασμάτων των κανόνων, με τη χρήση της $S_{\nu\acute{o}\rho\mu\alpha}$. Επιπλέον, μια και κάθε έξοδος θα είναι ένα ασαφές σύνολο, αντί για μια τιμή, απαιτούνται περίπλοκοι υπολογισμοί για την αποασαφοποίηση, αντί για την απλότητα του σταθμισμένου μέσου που χρησιμοποιεί το Takagi-Sugeno(TS-FIS).

Με δεδομένη την περιπλοκότητα των υπολογισμών που απαιτεί το Mamdani(M-FIS), επικράτησε στις πρακτικές εφαρμογές (όπου απαιτείται ταχύτητα υπολογισμών ενώ παράλληλα ο χώρος καταστάσεων είναι μεγάλος) η χρήση του Takagi-Sugeno(TS-FIS).

Κεφάλαιο 5

Ενισχυτική Μάθηση (Reinforcement Learning)

5.1 Γενικά για την Ενισχυτική Μάθηση

Η Ενισχυτική Μάθηση, αφορά μια σειρά προβλημάτων, στην οποία ένας πράκτορας (agent) εξελίσσεται αναλύοντας συνέπειες των πράξεων του, χάρη σε ένα απλό βαθμωτό σήμα (την ανταμοιβή) που λαμβάνει από το περιβάλλον. Οι ρίζες της βρίσκονται στην συμπεριφορική ψυχολογία, πρωτοπόρος της οποίας υπήρξε ο *Dr Ivan Pavlov* [25], αλλά λόγω της γενικής φύσης της, έγινε αντικείμενο μελέτης τόσο στο χώρο του αυτομάτου ελέγχου όσο και στον χώρο της επιστήμης υπολογιστών. Ιδιαίτερο ενδιαφέρον παρουσιάζει το γεγονός πως η εισαγωγή της Ενισχυτικής Μάθησης στο χώρο της τεχνητής νοημοσύνης, έγινε από το *Dr Richard S. Sutton* [29], ο οποίος κατέχει πτυχίο ψυχολογίας και εκπόνησε τη διδακτορική του διατριβή πάνω στην επιστήμη των υπολογιστών.

Τα δύο βασικά χαρακτηριστικά της Ενισχυτικής Μάθησης είναι τα εξής :

- Ο πράκτορας αλληλεπιδρά με το περιβάλλον και το ζευγάρι "πράκτορας και περιβάλλον" αποτελούν ένα δυναμικό σύστημα.
- Το σήμα ενίσχυσης του πράκτορα, είναι είτε ανταμοιβή είτε τιμωρία και με βάση αυτό μεταβάλλει τη συμπεριφορά του.

Στην Επιβλεπόμενη Μάθηση, γνωστή και ως "μάθηση με αυθεντία", το σύστημα μάθησης γνωρίζει ανα πάσα στιγμή την απόκλιση του από το επιθυμητό, μια και για κάθε είσοδο, είναι γνωστή η επιθυμητή έξοδος και άρα η διαφορά ανάμεσα σε επιθυμητή και πραγματική έξοδο, χρησιμοποιείται για τη μεταβολή των παραμέτρων ελέγχου.

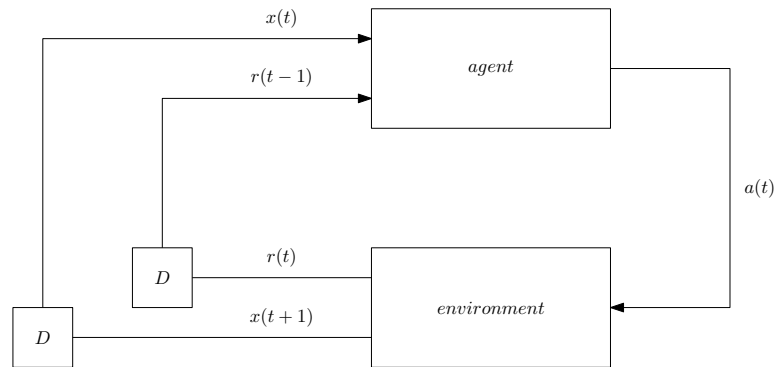
Αντίθετα, στην Ενισχυτική Μάθηση, γνωστή και ως "μάθηση με κριτική", το λαμβανόμενο σήμα ενίσχυσης (το οποίο μπορεί να είναι θετικό, αρνητικό ή ουδέτερο) απλά υποδεικνύει τι πρέπει να κάνει ο πράκτορας, χωρίς να επιβάλλει το πως. Ο πράκτορας χρησιμοποιεί το σήμα για να αποφασίσει μια πολιτική που θα τον οδηγήσει σε ένα μακροπρόθεσμο στόχο.

Μια ακόμη σημαντική διαφορά, ανάμεσα στην Ενισχυτική και την Επιβλεπόμενη Μάθηση είναι πως η πρώτη είναι πραγματικού χρόνου, μια και οι δράσεις του πράκτορα μεταβάλλουν την κατάσταση του περιβάλλοντος. Για να επιτευχθεί ο στόχος, ο πράκτορας συνδυάζει διαφορετικές δράσεις ακολουθώντας μια πολιτική, προσπαθώντας να επιλέξει την πολιτική που θα μεγιστοποιήσει τις μελλοντικές ανταμοιβές.

Η ακολουθούμενη διαδικασία είναι η εξής :

1. Τη χρονική στιγμή t , ο πράκτορας βρίσκεται στην κατάσταση $x(t)$
2. Ο πράκτορας επιλέγει μια από τις διαθέσιμες δράσεις $a(t)$ για τη συγκεκριμένη κατάσταση
3. Η επιλεγμένη δράση συνεπάγεται :
 - το πέρασμα σε μια νέα κατάσταση $x(t+1)$
 - την ανταμοιβή ενίσχυσης $r(t)$
4. $t \leftarrow t+1$
5. Επιστροφή στο 2, ή τερματισμός, αν η νέα κατάσταση είναι ο στόχος

Έστω πως X είναι το σύνολο των καταστάσεων και A το σύνολο των δράσεων. Η ανταμοιβή ενίσχυσης $r(t)$ είναι η συνέπεια της δράσης $a(t)$ που επιλέχθηκε στην κατάσταση $x(t)$. Η συνάρτηση ανταμοιβής ενίσχυσης, είναι η εφαρμογή του γινομένου χώρου $X \times A$ στο R ($r : X \times A \rightarrow R$). Σε πρώτη φάση τα X και A θεωρούνται διακριτά.



Σχήμα 5.1: Μοντελοποίηση δράσης-ανάδρασης, μεταξύ του πράκτορα και του περιβάλλοντος, στην ενισχυόμενη μάθηση. (Το D αναπαριστά χρονική καθυστέρηση)

5.2 Η συνάρτηση ανατροφοδότησης

Σκοπός του πράκτορα, είναι να μεγιστοποιήσει τη σώρευση μελλοντικών ανταμοιβών. Η συνάρτηση ανατροφοδότησης $R(t)$ είναι μια μακροπρόθεσμη μέτρηση ανταμοιβών.

Υπάρχουν τρεις διαφορετικοί τρόποι έκφρασης της ανατροφοδότησης στη στοχαστικά ενισχυτική μάθηση, παίρνοντας το άθροισμα των παρακάτω εξισώσεων :

- Το μοντέλο πεπερασμένου ορίζοντα
 Σε αυτή τη περίπτωση, ονομάζεται ορίζοντας, ένας πεπερασμένος αριθμός βημάτων στο μέλλον. Υπάρχει μια κατάσταση τερματισμού και η αλληλουχία δράσεων μεταξύ της αρχικής κατάστασης και της κατάστασης τερματισμού, ονομάζεται περίοδος. Αν K είναι ο αριθμός των βημάτων ώστε να επέλθει τερματισμός (Το K μπορεί να είναι προκαθορισμένο, μπορεί και όχι), τότε :

$$R(t) = r_t + r_{t+1} + \dots + r_{t+K-1} \quad (5.1)$$

- Το μοντέλο μειούμενης ανταμοιβής

Το κριτήριο αυτό χρησιμοποιείται όταν η αλληλουχία δράσεων είναι άπειρη. Εισάγετε ένα συντελεστής μείωσης $\gamma, 0 \leq \gamma \leq 1$ οπότε :

$$R(t) = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \quad (5.2)$$

Η τιμή του γ επιτρέπει τη διαμόρφωση της περιόδου για την οποία ο πράκτορας λαμβάνει υπόψη τις ανταμοιβές. Αν το $\gamma = 0$, ο πράκτορας συμπεριφέρεται "μυοπικά" και λαμβάνει υπόψη μόνο την άμεση ανταμοιβή. όσο το γ πλησιάζει τη μονάδα, τόσο ο πράκτορας λαμβάνει υπόψη και τις μελλοντικές ανταμοιβές.

- Το μοντέλο μέσης ανταμοιβής

Είναι ένα κριτήριο που χρησιμοποιείται στην περίπτωση του R-Learning και εξετάζει τον μέσο όρο των μελλοντικών ανταμοιβών :

$$R(t) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n r_{t+k} \quad (5.3)$$

5.3 Μαρκοβιανές Διαδικασίες Λήψης Αποφάσεων

Η διάδραση μεταξύ του πράκτορα και του περιβάλλοντος του, μπορεί να μοντελοποιηθεί σαν Μαρκοβιανό Πρόβλημα Απόφασης που επιλύεται με μεθόδους Δυναμικού Προγραμματισμού. Θεωρούμε μια στοχαστική διαδικασία X , της οποίας οι πιθανές τιμές, ανήκουν σε ένα πεπερασμένο αριθμό καταστάσεων, $X = \{x_0, x_1, \dots, x_k\}$. Σε ένα Μαρκοβιανό Πρόβλημα Απόφασης, οι μεταβάσεις από την κατάσταση i στην κατάσταση j εξαρτώνται μόνο από τις επιτρεπόμενες δράσεις στην κατάσταση i . Η πιθανότητα μετάβασης είναι :

$$p_{ij}(\alpha) = Pr(X_{t+1} = x_j | X_t = x_i, \alpha(t) = \alpha) \quad (5.4)$$

όπου $\alpha \in A_i$

Σκοπός ενός Μαρκοβιανού Προβλήματος Απόφασης, είναι, για κάθε αρχική κατάσταση, να βρεθεί μια βέλτιστη πολιτική, τέτοια ώστε να αποτελεί μια αλληλουχία δράσεων που μεγιστοποιεί την ανταμοιβή.

5.4 Μαθηματικό Πλαίσιο

5.4.1 Η συνάρτηση αξίας

Σε κάθε κατάσταση x_t , ο πράκτορας επιλέγει μια δράση, σύμφωνα με μια πολιτική π :

$$\alpha_t = \pi(x_t) \quad (5.5)$$

Η εκτίμηση μιας πολιτικής γίνεται με βάση τις μελλοντικές ανταμοιβές, συνυπολογίζοντας το συντελεστή μείωσης. Έτσι η αξία της κατάστασης x ακολουθώντας την πολιτική π δίνεται από τον τύπο :

$$V^\pi(x) = E\left\{\sum_{t=0}^{\infty} \gamma^t r(x_t, \pi(x_t)) \mid x_0 = x\right\} \forall x \in X \quad (5.6)$$

όπου το $E\{r(x_t, \pi(x_t))\}$ αναπαριστά την αναμενόμενη ενίσχυση, όταν ακολουθείται η προτεινόμενη δράση από την πολιτική π στην κατάσταση x_t . Για απλοποίηση, αν $R(x, \alpha) = E\{r(x, \alpha)\}$, η $V^\pi(x)$ γράφεται :

$$V^\pi(x) = R(x, \pi(x)) + \gamma \sum_{y \in X} p_{xy}(\pi(x)) V^\pi(y) \quad (5.7)$$

με $p_{xy}(\pi(x))$ να είναι η πιθανότητα της μετάβασης από την κατάσταση x στην κατάσταση y ενώ εφαρμόζεται η δράση $\pi(x)$.

Η συνάρτηση $x \rightarrow V^\pi(x)$ καθορίζει τη συνάρτηση αξίας, υπό την πολιτική π . Η συνάρτηση αυτή δύναται να υπολογιστεί, αν είναι γνωστή η αρχική κατάσταση καθώς και όλες οι πιθανότητες μετάβασης, οπότε και επιλύεται το γραμμικό σύστημα εξισώσεων. Επίσης υπάρχει μια βέλτιστη πολιτική π^* που ορίζεται ως :

$$V^{\pi^*}(x) \geq V^\pi(x), \forall x \in X, \forall \pi \quad (5.8)$$

Η βέλτιστη αξία μιας κατάστασης είναι (για λόγους απλότητας συμβολίζεται $V^* = V^{\pi^*}$:

$$V^*(x) = \max_{\pi} E\left\{\sum_{t=0}^{\infty} \gamma^t r_t\right\} \quad (5.9)$$

Η οποία, στην περίπτωση πεπερασμένου ορίζοντα οδηγεί στην Εξίσωση Βελτιστοποίησης του Bellman (γνωστή και ως εξίσωση του Δυναμικού Προγραμματισμού) :

$$V^*(x) = \max_{\alpha \in A_x} \{R(x, \alpha) + \gamma \sum_y p_{xy}(\alpha) V^*(y)\}, \forall x \in X \quad (5.10)$$

Και η βέλτιστη πολιτική, εξάγεται από το V^* :

$$\pi^*(x) = \max_{\alpha \in A_x} \{R(x, \alpha) + \gamma \sum_y p_{xy}(a) V^*(y)\} \quad (5.11)$$

5.4.2 Επίλυση της Εξίσωσης Bellman

Υπαρχουν δυο δημοφιλείς μέθοδοι υπολογισμού της βέλτιστης πολιτικής στους οποίους βασίζονται δυο διαφορετικοί αλγόριθμοι Ενισχυόμενης Μάθησης :

- Επαναλήψεις επί της πολιτικής, με απευθείας διαχείριση των πολιτικών.
- Επαναλήψεις επί των αξιών, με αναζήτηση της βέλτιστης συνάρτησης αξιών.

Επαναλήψεις επί της πολιτικής

Η μέθοδος αυτή υπολογισμού, ξεκινά από οποιαδήποτε αρχική πολιτική και τη βελτιώνει μέσω συνεχών επαναλήψεων. Ο αλγόριθμος που ακολουθείται είναι :

Algorithm 1 Αλγόριθμος Επανάληψης επί της πολιτικής

```

επιλογή μιας αρχικής πολιτικής ,  $\hat{\pi}$ 
while η πολιτική συνεχίζει να μεταβάλλεται do
     $\pi = \hat{\pi}$ 
    υπολογισμός  $V^\pi$ 
    βελτίωση της πολιτικής, για κάθε κατάσταση :
     $\hat{\pi}(x) = \arg \max_a \{R(x, \alpha) + \gamma \sum_y p_{xy}(a) V^\pi(y)\}$ 
end while

```

Η αξία της πολιτικής, προκύπτει από την επίλυση του γραμμικού συστήματος εξισώσεων. Τότε, για κάθε κατάσταση, παρατηρείται το κατα πόσο είναι δυνατό να βελτιωθεί η πολιτική, μεταβάλλοντας την πρώτη δράση που λαμβάνεται σε αυτή την κατάσταση.

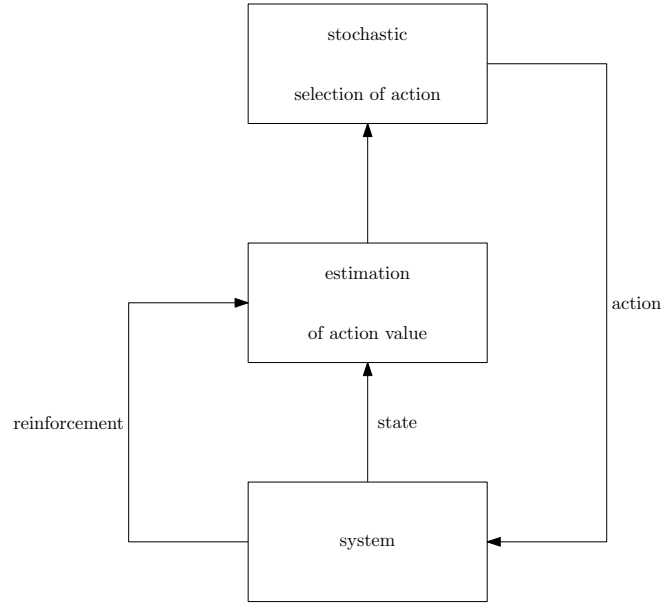
Η ταχύτητα υπολογισμού των επαναλήψεων παραμένει γρήγορη, όσο ο χώρος δράσεων παραμένει μικρός ή όταν μικρός αριθμός επαναλήψεων απαιτείται για τη λήψη ικανοποιητικού αποτελέσματος. Παρόλαυτά η μέθοδος έχει μεγάλες υπολογιστικές απαιτήσεις, και αποφεύγεται στην περίπτωση χώρων με μεγάλο αριθμό πιθανών δράσεων.

Επαναλήψεις επί της αξίας

Ζητούμενο σε αυτή τη μέθοδο δεν είναι ο καθορισμός της πολιτικής, αλλά η εύρεση της δράσης που αποδίδει τη μέγιστη αξία για την κάθε κατάσταση. Αν $Q^\pi(x, \alpha)$ είναι η αξία μιας δράσης α όταν ακολουθείται στην κατάσταση x όταν ακολουθείται η πολιτική π , τότε :

$$Q^\pi(x, \alpha) = R(x, \alpha) + \gamma \sum_y p_{xy}(a) V^\pi(y) \quad (5.12)$$

Η συνάρτηση $Q^* = Q^{\pi^*}$ αντιστοιχεί στη βέλτιστη πολιτική και η εξίσωση βελτιστοποίησης του Bellman γίνεται :



Σχήμα 5.2: Αρχιτεκτονική Επαναλήψεων επί της αξίας

$$V^*(x) = \max_{\alpha \in A_x} Q^*(x, \alpha) \quad (5.13)$$

Για την εκτίμηση του $V^*(\cdot)$, γίνεται υπόθεση πεπερασμένου οριζοντα, εξετάζοντας διαδοχικά τα στάδια επαναλήψεων, από το πρώτο ως το τελευταίο. Αν $V_n(\cdot)$ είναι η εκτίμηση της συνάρτησης αξίας στο n -οστό στάδιο, ο αλγόριθμος επαναλήψεων είναι :

Algorithm 2 Αλγόριθμος Επανάληψης επί της αξίας

επιλογή μιας αρχικής συνάρτησης αξίας, $V_0(x)$, $\forall x \in X$
while $0 \leq n \leq n_{final}$ **do**
 for $x \in X$ **do**
 υπολογισμός $Q_{n+1}(x, \alpha) = R(x, \alpha) + \gamma \sum_y p_{xy}(a) V_n(y)$, $\forall \alpha \in A$
 $V_{n+1}(x) = \max_{\alpha \in A_x} Q_{n+1}(x, \alpha)$
 end for
end while

Η αρχιτεκτονική του αλγορίθμου επαναλήψεων επί της αξίας, φαίνεται στο σχήμα 5.2.

5.5 Χρονικές διαφορές (Temporal Differences)

5.5.1 TD(0)

Θεωρούμε την περίπτωση όπου ο πράκτορας δεν γνωρίζει το περιβάλλον του. Η άμεση ενίσχυση για κάθε ζευγάρι κατάστασης-δράσης είναι άγνωστη, όπως επίσης άγνωστες είναι οι πιθανότητες μετάβασης από κατάσταση σε κατάσταση για κάθε δράση.

Υπάρχουν δύο δυνατές προσεγγίσεις :

- Εξαγωγή της συνάρτησης αξίας μέσω μοντέλου. Η μέθοδος αυτή χαρακτηρίζεται σαν έμμεση μέθοδος.

- Εκτίμηση της συνάρτησης αξίας εμπειρικά. Η μέθοδος αυτή χαρακτηρίζεται σαν άμεση μέθοδος.

Στην παρούσα εργασία χρησιμοποιήθηκε η εμπειρική εκτίμηση της συνάρτησης αξίας. Καθώς η αξία κάθε κατάστασης είναι άγνωστη, χρησιμοποιούνται διαδοχικές αξιολογήσεις της βέλτιστης συνάρτησης αξίας V^* .

Τη χρονική στιγμή t , η συνάρτηση $\hat{V}_t(x_t)$ αναπαριστά την αξιολόγηση της $V^*(x_t)$. Η δράση a_t κάνει τη μετάβαση από τη x_t στη x_{t+1} και η άμεση ενίσχυση είναι r_t . Η νέα αξιολόγηση της αξίας θα είναι με βάση την 5.7:

$$r_t + \gamma \hat{V}_t(x_{t+1}) \quad (5.14)$$

Αυτή η αξιολόγηση είναι μοναδική για την κατάσταση x_{t+1} .

Η χρονική διαφορά μεταξύ των δυο διαδοχικών αξιολογήσεων της αξίας της x_{t+1} είναι $r_t + \gamma \hat{V}_t(x_{t+1}) - \hat{V}_t(x_t)$. Η διαφορά αυτή χρησιμοποιείται για την ενημέρωση της αξίας :

$$\hat{V}_{t+1}(x_t) = \hat{V}_t(x_t) + \beta \{r_t + \gamma \hat{V}_t(x_{t+1}) - \hat{V}_t(x_t)\} \quad (5.15)$$

Όπου β είναι το κέρδος μάθησης, $0 \leq \beta \leq 1$. Αν αυτό το κέρδος μάθησης μειώνεται σταδιακά, μετά από κάθε βήμα, καθώς θα πλησιάζει προς το μηδέν, οι αξιολογήσεις της αξίας θα συγκλίνουν προς τη βέλτιστη συνάρτηση αξίας.

5.5.2 Ίχνος Επιλεξιμότητας

Η τεχνική TD(0) , λαμβάνει υπόψη μόνο την παρούσα και την επόμενη κατάσταση. Μπορεί να γίνει επέκταση της αξιολόγησης για όλες τις καταστάσεις, σύμφωνα με ένα βαθμό επιλεξιμότητας μιας κατάστασης στο παρελθόν. Το ίχνος αυτό επιλεξιμότητας μπορεί να οριστεί ως :

$$e_t(x) = \begin{cases} 1 & \text{αν } x = x_t \\ \gamma \lambda e_{t-1}(x) & \text{διαφορετικά} \end{cases} \quad (5.16)$$

Όπου λ μια μειούμενη παράμετρος από τιμές κοντά στη μονάδα, προς τιμές κοντά στο μηδέν, που αναπαριστά την εξασθένιση του ίχνους προηγούμενης επίσκεψης σε μια κατάσταση.

Το ίχνος επιλεξιμότητας απομνημονεύει τη συχνότητα επίσκεψης μιας κατάστασης. Μειώνεται εκθετικά με το χρόνο, εκτός και αν επανενεργοποιηθεί από νέα επίσκεψη. Η αξία κάθε κατάστασης γίνεται :

$$\hat{V}_{t+1}(x_t) = \hat{V}_t(x_t) + \beta \{r_t + \gamma \hat{V}_t(x_{t+1}) - \hat{V}_t(x_t)\} e_t(x) \quad (5.17)$$

5.5.3 TD(λ)

Η μέθοδος χρονικής διαφοράς, χρησιμοποιείται από όλους τους αλγόριθμους της ενισχυτικής μάθησης άλλοτε με $\lambda = 0$ και άλλοτε με $\lambda > 0$. Ο πλήρης αλγόριθμος φαίνεται παρακάτω :

Algorithm 3 Αλγόριθμος TD(λ)

```

 $t = 0, V_t(x) = 0, e_t(x) = 0 \forall x \in X$ 
while δεν έχει επέλθει σύγκλιση do
    παρατήρησε τη μετάβαση  $x_t \rightarrow x_{t+1}$ 
    υπολογισμός  $e_t(x)$ 
    υπολογισμός  $\hat{V}(x), \forall x$ 
     $t \leftarrow t + 1$ 
end while

```

5.5.4 Μάθηση Q (Q-Learning)

Η μάθηση Q πρωτοαναφέρθηκε από τον *Chris Watkins* το 1989 [31]. Αποτελεί ίσως τον πιο δημοφιλή αλγόριθμο Ενισχυτικής Μάθησης, κυρίως λόγω της απλότητάς του.

Μάθηση Q(0)

Είναι η περίπτωση μάθησης Q που βασίζεται στις χρονικές διαφορές τύπου TD(0), λαμβάνοντας υπόψη, μόνο το επόμενο βήμα. Ο πράκτορας παρατηρεί την παρούσα κατάσταση, x_t και πραγματοποιεί μια δράση a_t , σύμφωνα με την ανταμοιβή που λαμβάνει. Έπειτα, ενημερώνει την αξία της δράσης, λαμβάνοντας υπόψη την άμεση ενίσχυση r_t αλλά και την εκτιμώμενη αξία της νεας κατάστασης $V_t(x_{t+1})$:

$$V_t(x_{t+1}) = \max_{b \in A_{t+1}} Q(x_{t+1}, b) \quad (5.18)$$

ενημέρωση αξίας δράσης :

$$Q(x_t, \alpha_t) \leftarrow Q(x_t, \alpha_t) + \beta \{r_t + \gamma V_t(x_{t+1}) - Q(x_t, \alpha_t)\} \quad (5.19)$$

Όπου β είναι μια μαθησιακή παράμετρος τέτοια ώστε $\beta \rightarrow 0$ καθώς $t \rightarrow \infty$. Η εξίσωση, μπορεί να ξαναγραφεί :

$$Q(x_t, \alpha_t) \leftarrow (1 - \beta)Q(x_t, \alpha_t) + \beta \{r_t + \gamma V_t(x_{t+1})\} \quad (5.20)$$

Αυτή η ενημέρωση, αντιστοιχεί στο κέντρο βάρους των παλαιών και νεότερων αξιολογήσεων, με βάρος β .

Η μάθηση Q παρουσιάζει τα ακόλουθα χαρακτηριστικά :

- Οι εκτιμήσεις των αξιών του Q , είναι ανεξάρτητες της πολιτικής που ακολουθείται από τον πράκτορα. Ο τελευταίος μπορεί να ακολουθήσει οποιαδήποτε πολιτική, συνεχίζοντας πάντα να κατασκευάζει σωστές εκτιμήσεις της αξίας των δράσεων.
- Οι αξίες του Q είναι εκμεταλλεύσιμες πολύ πριν τη σύγκλιση, η οποία πολλές φορές καθυστερεί.
- Υπάρχει σύγκλιση προς τη βέλτιστη πολιτική.

Ιδιαίτερη προσοχή πρέπει να δωθεί στο να εξερευνηθούν όλες οι πιθανές δράσεις σε κάθε κατάσταση. Αν η δράση με τη μεγαλύτερη τιμή Q , επιλεγεί πολύ γρήγορα, μπορεί να οδηγήσει σε τοπικά ελάχιστα :

$$\alpha = \arg \max_{b \in A_x} Q(x, b) \quad (5.21)$$

Τέλος πρέπει να σημειωθεί πως ενίσχυση διάφορη του μηδενός, μπορεί να σημάνει το τέλος μιας περιόδου. Σε αυτή την περίπτωση δεν υπάρχει επόμενη κατάσταση και ο πράκτορας ξεκινά μια νέα ακολουθία εκπαίδευσης. Η εξίσωση ενημέρωσης σε αυτή τη περίπτωση είναι :

$$Q(x_t, \alpha_t) \leftarrow (1 - \beta)Q(x_t, \alpha_t) + \beta r_t \quad (5.22)$$

Μάθηση $Q(\lambda)$

Η μάθηση $Q(\lambda)$ αποτελεί τη γενικότερη περίπτωση, η οποία χρησιμοποιεί το ίχνος επιλεξιμότητας. Τα ίχνη επιλεξιμότητας ορίζονται στον χώρο καταστάσεων $X \times A$ και η 5.16 γίνεται :

$$e_t(x, \alpha) = \begin{cases} 1 & \text{αν } x = x_t \text{ και } \alpha = \alpha_t \\ \gamma \lambda e_{t-1}(x, \alpha) & \text{διαφορετικά} \end{cases} \quad (5.23)$$

Αν και η σύγκλιση της $Q(\lambda)$ προς τη βέλτιστη Q^* δεν εξασφαλίζεται για $\lambda > 0$, η εκπαίδευση του πράκτορα είναι γρηγορότερη. Ο αλγόριθμος για την μάθηση $Q(\lambda)$ φαίνεται στον αλγόριθμο 4.

5.6 Εξερεύνηση/Εκμετάλλευση

Μετά τη σύγκλιση της Q -μάθησης, η βέλτιστη πολιτική εκτελείται, επιλέγοντας τη δράση, που σε κάθε κατάσταση, μεγιστοποιεί τη συνάρτηση Q :

$$\alpha = \arg \max_{b \in A_x} Q^*(x, b) \quad (5.24)$$

Algorithm 4 Αλγόριθμος Q(λ)

```

 $t = 0, Q(x, \alpha) = 0, e_t(x, \alpha) = 0, \forall (x, \alpha) \in X \times A$ 
while δεν έχει επέλθει σύγκλιση ή λ πολύ κοντά στο 0 do
  παρατήρησε τη μετάβαση  $x_t \rightarrow x_{t+1}$ 
   $\epsilon_{tQ} = r_t + \gamma V_t x_{t+1} - Q(x_t, \alpha_t)$ 
   $\epsilon_{tV} = r_t + \gamma V_t x_{t+1} - V_t(x_t)$ 
  υπολογισμός  $e_t(x, \alpha)$ 
  υπολογισμός  $Q(x, \alpha), \forall (x, \alpha) \in X \times A$ 
   $Q(x_t, \alpha_t) \leftarrow Q(x_t, \alpha_t) + \beta \epsilon_{tQ} e_t(x_t, \alpha_t)$ 
   $Q(x, \alpha) \leftarrow Q(x, \alpha) + \beta \epsilon_{tV} e_t(x, \alpha), \forall (x, \alpha) \neq (x_t, \alpha_t)$ 
   $t \leftarrow t + 1$ 
end while

```

Η πολιτική αυτή ονομάζεται *άπληστη*. Ωστόσο, στην έναρξη της εκπαίδευσης, οι τιμές του $Q(x, \alpha)$ δεν έχουν κάποιο νόημα. Η εφαρμογή της άπληστης πολιτικής πολύ γρήγορα, οδηγεί σε τοπικά ελάχιστα. Για να εξαχθεί μια χρήσιμη εκτίμηση του Q, είναι απαραίτητο να εξερευνηθεί και να εκτιμηθεί το σύνολο των δυνατών δράσεων, για όλες τις καταστάσεις. Η φάση αυτή ονομάζεται φάση *εξερεύνησης*, ενώ η φάση στο τέλος της εκπαίδευσης, ονομάζεται φάση *εκμετάλλευσης*.

Το δίλημμα Εξερεύνησης/Εκμετάλλευσης, σε κάθε κατάσταση, για τον πράκτορα, εκφράζεται ως :

- μια δράση για την οποία, η αναμενόμενη ανταμοιβή είναι καλής ποιότητας
- μια δράση, της οποίας η ποιότητα τη στιγμή της επιλογής δεν είναι τόσο καλή, αλλά που η επιλογή της θα μπορούσε να οδηγήσει σε ίσως καλές ανεξερεύνητες καταστάσεις.

Η απάντηση σε αυτό το δίλημμα, είναι να εξερευνηθεί επαρκώς ο χώρος δράσεων, και να δοκιμαστούν όλες οι πιθανές δράσεις, για όλες τις καταστάσεις, αλλά κάτι τέτοιο είναι χρονοβόρο.

Ζητούμενο λοιπόν είναι να βρεθεί ένας συμβιβασμός για οικονομία υπολογιστικού χρόνου. Τρεις δημοφιλείς μέθοδοι Εξερεύνησης και Εκμετάλλευσης είναι οι ακόλουθες :

- Ψευδοστοχαστική Μέθοδος : Η δράση με την μεγαλύτερη αξία, έχει μια πιθανότητα P να επιλεγεί, διαφορετικά, μια δράση επιλέγεται τυχαία.
- Ψευδοεξαντλητική Μέθοδος : Η δράση με την μεγαλύτερη αξία, έχει μια πιθανότητα P να επιλεγεί, διαφορετικά, η δράση που έχει επιλεγεί λιγότερο στο παρελθόν, επιλέγεται.
- Κατανομή Boltzmann : Η δράση a επιλέγεται με την πιθανότητα :

$$P(a|x) = \frac{\exp(\frac{1}{T}Q(x, a))}{\sum_b \exp(\frac{1}{T}Q(x, b))} \quad (5.25)$$

Η Ψευδοστοχαστική και η Ψευδοεξαντλητική μέθοδοι, δίνουν σε γενικές γραμμές, καλύτερα αποτελέσματα.

Κεφάλαιο 6

Ασαφής Λογική με Ενισχυτική Μάθηση

Η ενισχυτική μάθηση έχει χρησιμοποιηθεί για τη βελτιστοποίηση Ασαφών Συμπερασματικών συστημάτων με δυο διαφορετικές προσεγγίσεις :

- Μέσω επαναλήψεων επί της πολιτικής.
- Μέσω επαναλήψεων επί της τιμής, που αποτελεί γενική περίπτωση της Q-μάθησης.

Για συστήματα συνεχών χώρων καταστάσεων, το Ασαφές Συμπερασματικό Σύστημα, εξυπηρετεί, τόσο στην αναπαράσταση των τιμών Q όσο και στον υπολογισμό της εφαρμοζόμενης δράσης. Η χρήση Ασαφών Συμπερασματικών συστημάτων έχει τα ακόλουθα πλεονεκτήματα :

- Η εύκολη αναπαράσταση των τιμών Q , λόγω της δυνατότητας των Ασαφών Συμπερασματικών Συστημάτων για προσέγγιση τιμών.
- Η δυνατότητα διαχείρισης συνεχών χώρων καταστάσεων και δράσεων.
- Η ενσωμάτωση γνώσης που είναι διαθέσιμη για το σύστημα από πριν, και η ευκολία κατανόησης της εξαγόμενης γνώσης , μετά την φάση εκπαίδευσης του συστήματος.

Η υλοποίηση Ασαφούς Λογικής με Ενισχυτική Μάθηση, της διπλωματικής αυτής εργασίας, βασίστηκε στην προσέγγιση του *Pierre Yves Glorennec* [15].

6.1 Μορφή του Ασαφούς Συμπερασματικού Συστήματος

Θεωρούνται μόνο μηδενικής τάξης Takagi-Sugeno ασαφή συμπερασματικά συστήματα, που χρησιμοποιούνται συχνά για μοντελοποίηση και έλεγχο πραγματικών συστημάτων. Υπενθυμίζεται πως η μορφή των κανόνων είναι :

$$\text{AN } x_1 \text{ είναι } A_1^i \text{ ΚΑΙ } \dots \text{ ΚΑΙ } x_n \text{ είναι } A_n^i \text{ τότε } y = c_i$$

Όπου A_i^j είναι ένα ασαφές σύνολο στον χώρο των x_j και c_i είναι ένας πραγματικός αριθμός, για $j = 1$ έως n και $i = 1$ έως N , με n και N να είναι οι διαστάσεις του χώρου εισόδων και ο αριθμός των κανόνων αντίστοιχα.

Για λόγους απλοποίησης τοποθετούμε τους ακόλουθους περιορισμούς :

- Η μορφή των ασαφών συνόλων είναι ορισμένη από το χρήστη *a priori*.

- Η χρησιμοποιούμενη κατάτμηση είναι ισχυρή για τις μεταβλητές εισόδου :

$$(\forall x) \left(\sum_i \alpha_i(x) = 1 \right) \quad (6.1)$$

Όπου α_i ο βαθμός αλήθειας του κανόνα i .

- Στο τμήμα παραδοχών των κανόνων, ο χρησιμοποιούμενος τελεστής είναι ο τελεστής ΚΑΙ.

Αν $x = (x_1, \dots, x_n)^t$, η χαρτογράφηση εισόδου-εξόδου θα είναι $x \leftarrow y(x)$ και η έξοδος για ένα μηδενικής τάξης σύστημα Takagi-Sugeno θα είναι :

1. Υπολογισμός των βαθμών συμμετοχής στα διάφορα ασαφή σύνολα $\mu_{A_j^i}(x_j)$ για $i = 1$ έως N και $j = 1$ έως n
2. Υπολογισμός του βαθμού αλήθειας για κάθε κανόνα, με δεδομένο το διάνυσμα εισόδου : $\alpha_i(x) = \prod_j \mu_{A_j^i}(x_j)$
3. Υπολογισμός της εξόδου :

$$y(x) = \frac{\sum_i \alpha_i(x) c_i}{\sum_i \alpha_i(x)} = \sum_i \alpha_i(x) \times c_i \quad (6.2)$$

6.2 Αναπαράσταση της συνάρτησης Q από ένα Ασαφές Συμπερασματικό Σύστημα

Αρχικά θεωρούμε μόνο μια δράση α . Θέλουμε να υπολογίσουμε την συνδεόμενη αξία Q με τη συνάρτηση :

$$x \rightarrow y = \hat{Q}(x, a) = FIS(x), x \in X \quad (6.3)$$

Η συνάρτηση ορίζεται από τους ασαφείς κανόνες :

$$\begin{aligned} \text{AN } x \text{ είναι } S_1 \text{ τότε } y &= c_1 \\ &\dots \\ \text{AN } x \text{ είναι } S_N \text{ τότε } y &= c_N \end{aligned}$$

Για διάνυσμα εισόδου x_t , η αξιολόγηση τιμής της δράσης α δίνεται από την εξίσωση :

$$\hat{Q}(x_t, a) = FIS(x_t) = \sum_i \alpha_i(x_t) \times c_i \quad (6.4)$$

Αν θέσουμε $c_i = q[S_i, \alpha]$ εισάγουμε την έννοια της ποιότητας q για τη δράση α στο ασαφές συμπερασματικό σύστημα και έτσι έχουμε :

$$\hat{Q}(x_t, a) = FIS(x_t) = \sum_i \alpha_i(x_t) \times q[S_i, \alpha] \quad (6.5)$$

6.3 Συνεχείς χώροι δράσεων και καταστάσεων

Μπορεί πλέον να χρησιμοποιηθεί το ασαφές συμπερασματικό σύστημα, για τον υπολογισμό τόσο της δράσης, όσο και της σχετιζόμενης ποιότητας Q με αυτή. Οι κανόνες μπορούν πλέον να γραφούν, για $i = 1$ έως N :

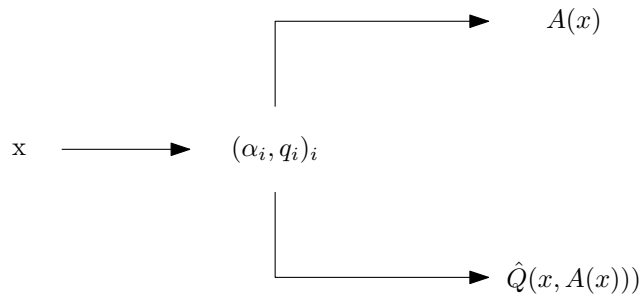
ΑΝ x είναι S_i τότε $y = \alpha_i$ με q_i

οπότε για διάνυσμα εισόδου x μπορούν να βρεθούν η δράση $A(x)$ που αντιστοιχεί σε αυτό, όπως και η σχετιζόμενη τιμή $Q(x, A(x))$:

$$A(x) = \sum_i^N \alpha_i(x) \times \alpha_i \quad (6.6)$$

$$\hat{Q}(x, A(x)) = \sum_i^N \alpha_i(x) \times q_i \quad (6.7)$$

Στο παρακάτω σχήμα οπτικοποιείται η σχέση μεταξύ των $x, A(x), \hat{Q}(x, A(x))$:



Σχήμα 6.1: Σχέση μεταξύ των $x, A(x), \hat{Q}(x, A(x))$

6.4 Εξαγωγή Γνώσης

Η εξαγωγή γνώσης αποτελεί γενικότερη περίπτωση της διακριτής Q -μάθησης. Καθώς συνδέονται διαφορετικές δράσεις σε μια κατάσταση, δημιουργούνται και ανταγωνιζόμενα συμπεράσματα για κάθε κανόνα. Κάθε συμπέρασμα έχει τη δική του q -τιμή που ενημερώνεται με κάθε επανάληψη. Η διαδικασία μάθησης συνιστάται στο να βρεθεί το καλύτερο σύνολο κανόνων, αυτό που θα βελτιστοποιήσει τις μελλοντικές ενισχύσεις. Η αρχική βάση κανόνων αποτελείται από N κανόνες :

ΑΝ x είναι S_i τότε :

$$\begin{aligned}
 & y = \alpha[i, 1] \text{ με } q[i, 1] = 0 \\
 \text{Ή } & y = \alpha[i, 2] \text{ με } q[i, 2] = 0 \\
 & \vdots \\
 \text{Ή } & y = \alpha[i, J] \text{ με } q[i, J] = 0
 \end{aligned}$$

όπου $(\alpha[i, j])_{j=1}^J$ είναι πιθανές λύσεις, των οποίων η ποιότητα αρχικοποιείται στο μηδέν.

Για κάθε κανόνα i , έστω $k[i] \in \{1, J\}$ ο δείκτης της δράσης που επιλέχθηκε με βάση την ποιότητά της. Η αξία της δράσης και η ποιότητά της γίνονται :

$$A(x) = \sum_i^N \alpha_i(x) \times \alpha[i, k[i]] \quad (6.8)$$

$$\hat{Q}(x, A(x)) = \sum_i^N \alpha_i(x) \times q[i, k[i]] \quad (6.9)$$

6.4.1 Επιλογή Συμπερασμάτων

Η υλοποίηση της ασαφούς Q-μάθησης συνεπάγεται πολλαπλά ανταγωνιζόμενα συμπεράσματα ανα κανόνα. Σε αντίθεση με την επιβλεπόμενη μάθηση, τα συμπεράσματα δεν τροποποιούνται. Αναλύονται κατα τη φάση της εξερεύνησης και οι q-τιμές τους ενημερώνονται. Στην φάση εκμετάλλευσης, η άπληστη πολιτική λαμβάνει υπόψη μόνο το καλύτερο συμπέρασμα ανάμεσα στα διαθέσιμα. Είναι σημαντική συνεπώς η αρχική επιλογή συμπερασμάτων σύμφωνα με την a priori γνώση του συστήματος :

- Άγνωστο σύστημα:

Είναι η πιο δυσμενής περίπτωση. Αν y_{min} και y_{max} είναι το κάτω και άνω όριο του συμπεράσματος, τα J συμπεράσματα θα είναι ίσα κατανεμημένα στο διάστημα $[y_{min}, y_{max}]$.

- Μερικώς γνωστό σύστημα:

Η γνώση μπορεί να αναπαρασταθεί ποιοτικά, με γλωσσικές μεταβλητές, όπως "Λίγο", "Πολύ" κ.ο.κ.. Σε αυτή την περίπτωση τα συμπεράσματα πρέπει να δίνουν διαφορετικές ποσοτικές περιγραφές των όρων "Λίγο", "Πολύ", κ.τ.λ.

- Ακριβής γνώση του συστήματος:

Μερικοί κανόνες μπορεί να είναι επακριβώς γνωστοί. Σε αυτή την περίπτωση δε χρειάζεται φάση εξερεύνησης. Οι κανόνες χρησιμοποιούνται απευθείας στη φάση εκμετάλλευσης.

Οι παραπάνω περιπτώσεις περιγράφουν ουσιαστικά και τα στάδια της διαδικασίας μάθησης. Ο πράκτορας ξεκινάει μη έχοντας γνώση του περιβάλλοντος, μετά από μερικές επαναλήψεις αποκτά μερική γνώση του και προς το τέλος των επαναλήψεων συγκλίνει προς την ακριβή γνώση του περιβάλλοντός του.

Είναι πολύ σημαντικό να τονισθεί πως οι αρχικές q-τιμές πρέπει να είναι μηδενικές, ώστε να μην αλλοιωθεί η εκπαίδευση.

6.4.2 Αλγόριθμος

Τη χρονική στιγμή t , επιλέγεται η δράση $A(x_t)$ και η νέα κατάσταση είναι x_{t+1} . Η αξία αυτής της κατάστασης, ανταποκρίνεται στην ποιότητα της καλύτερης δυνατής διαθέσιμης δράσης. Αν $q[i, \max[i]]$ είναι η μέγιστη q -τιμή για τον κανόνα i , η αξία της κατάστασης είναι :

$$V_t(x_{t+1}) = \sum_i^N \alpha_i(x_{t+1}) \times q[i, \max[i]] \quad (6.10)$$

Το σφάλμα χρονικής διαφοράς είναι :

$$\Delta Q = \beta \{r_t + \gamma V_t(x_{t+1}) - \hat{Q}(x_t, A(x_t))\} \quad (6.11)$$

και η μεταβολή της ποιότητας των κανόνων γίνεται :

$$\Delta q[i, j] = \begin{cases} \epsilon \times \Delta Q \times \alpha_i(x_t) & j = k[i] \\ 0 & \text{διαφορετικά} \end{cases} \quad (6.12)$$

Ίχνος Επιλεξιμότητας

Το ίχνος επιλεξιμότητας αποτελεί, όπως έχει αναφερθεί, ένα τρόπο σύνδεσης του παρόντος ζεύγους δράσης-κατάστασης, με προηγούμενες επισκέψεις στο ζεύγος αυτό. Όσο πιο πρόσφατη είναι η τελευταία επίσκεψη, τόσο περισσότερο λαμβάνεται υπόψη η ποιότητα του αντίστοιχου κανόνα.

Αν $e[i, j]$ είναι η επιλεξιμότητα του συμπεράσματος j του κανόνα i , αυτή ενημερώνεται μετά την επιλογή συμπερασμάτων από την πολιτική εξερεύνησης-εκμετάλλευσης :

$$e[i, j] = \begin{cases} \alpha_i(x_t) & j = k[i] \\ \gamma \lambda e[i, j] & \text{διαφορετικά} \end{cases} \quad (6.13)$$

Η μεταβολή της ποιότητας των κανόνων, γίνεται :

$$\Delta Q = \epsilon \times \Delta Q \times e[i, j] \quad (6.14)$$

Πριν τη παρουσίαση του αλγορίθμου, πρέπει να γίνουν δυο παρατηρήσεις :

- Η ενημέρωση των q -τιμών για συμπεράσματα που χρησιμοποιήθηκαν στην κατάσταση x_t λαμβάνει χώρα στην κατάσταση x_{t+1} .
- Απαιτούνται οι ακόλουθες σταθερές
 - ο συντελεστής μείωσης γ , $0 \leq \gamma \leq 1$
 - το κέρδος μάθησης β , που χρησιμοποιείται για την ενημέρωση του $\hat{Q}(x_t, A(x_t))$

- το κέρδος μάθησης ϵ , που χρησιμοποιείται για την ενημέρωση των q-τιμών
- η παράμετρος εξασθένησης ίχνους επιλεξιμότητας λ
- η παράμετρος εξερεύνησης, δηλαδή η πιθανότητα να επιλεγεί η δράση με τη μεγαλύτερη ποιότητα

Ο αλγόριθμος εξαγωγής γνώσης είναι :

Algorithm 5 Αλγόριθμος Q-μάθησης

1. $t = 0$, ίχνος επιλεξιμότητας $e[i, j] = 0$, Πίνακας ποιότητας $Q = 0$, παρατήρηση x_t
 2. Για κάθε κανόνα i , υπολογισμός του βαθμού αληθείας των κανόνων $\alpha_i(x_t)$ για την κατάσταση x_t
 3. Για κάθε κανόνα i , επιλογή μιας από τις j διαθέσιμες δράσεις που συνδέονται με τον κανόνα i , χρησιμοποιώντας e-greedy αλγόριθμο επιλογής.
 4. Υπολογισμός του πίνακα δράσεων $A(x_t)$ και της ποιότητας $Q(x_t, A(x_t))$ αυτών των δράσεων
 5. Υπολογισμός της επιλεξιμότητας των συμπερασμάτων, $e[i, j] = \alpha_i(x_t)$ αν j είναι η παρούσα δράση, $e[i, j] = \gamma \lambda e[i, j]$ διαφορετικά
 6. Πραγματοποίηση της δράσης $A(x_t)$ και παρατήρηση της κατάστασης x_{t+1}
 7. Έλεγχος Κατάστασης. Λήψη της ανταμοιβής r_t
 8. Για κάθε κανόνα i , υπολογισμός του βαθμού αληθείας των κανόνων $\alpha_i(x_{t+1})$ για την κατάσταση x_{t+1}
 9. Υπολογισμός της βέλτιστης πολιτικής $V_t(x_{t+1})$ και της μεταβολής ποιότητας $\Delta Q(V_t(x_{t+1}), A(x_t))$ αυτών των δράσεων
 10. Ενημέρωση του σφάλματος χρονικής διαφοράς και των παραμέτρων ποιότητας $\Delta Q[i, j]$ και $q[i, j] \leftarrow q[i, j] + \Delta q[i, j]$
 11. $t = t + 1$ και εκτέλεση του βήματος 3
-

Κεφάλαιο 7

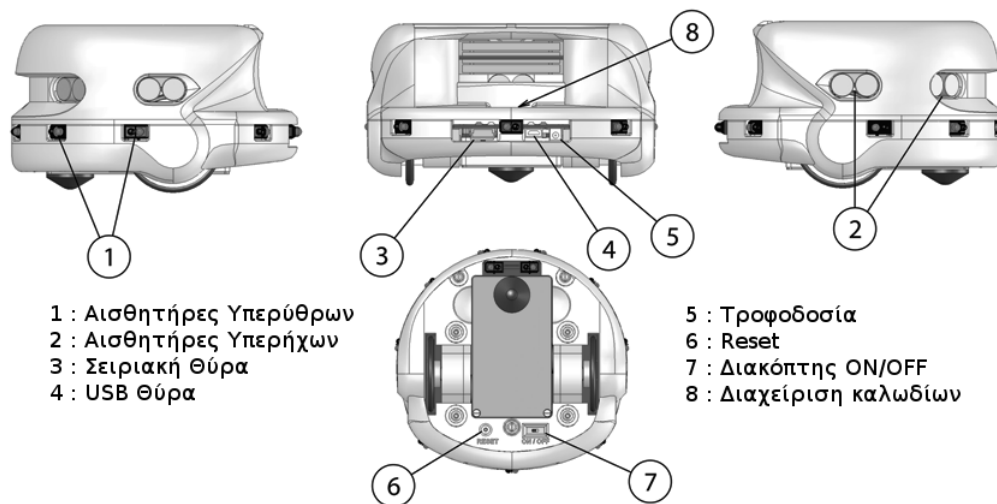
Υλικό και Λογισμικό Υλοποίησης

7.1 Το Ρομπότ Khepera 3

7.1.1 Γενικά για το ρομπότ

Ο έλεγχος με ασαφή λογική και ενισχυτική μάθηση, εφαρμόστηκε πάνω στο ρομπότ διαφορικής οδήγησης Khepera 3. Μερικές από τις δοκιμές αισθητήρων, καθώς και ο απλός έλεγχος με ασαφή λογική, εφαρμόστηκε στο ρομπότ Khepera 2.

Σχεδιασμένα αρχικά, από το Εργαστήριο Μικροεπεξεργαστών και Διεπαφών του Πολυτεχνείου της Λωζάνης (EPFL), διατίθενται στην αγορά από την εταιρία K-TEAM και αποτελούν πλατφόρμες για εφαρμογή αλγορίθμων ρομποτικής, σε πανεπιστημιακό περιβάλλον, έχοντας συγχρόνως, χαμηλό κόστος απόκτησης.



Σχήμα 7.1: Το Ρομπότ Khepera 3 [10]

Ακολουθεί, στο κεφάλαιο αυτό, μια παρουσίαση των βασικότερων συστημάτων του ρομπότ καθώς και των παρεχομένων δυνατοτήτων διασύνδεσης με υπολογιστή.

7.1.2 Σύστημα Κίνησης

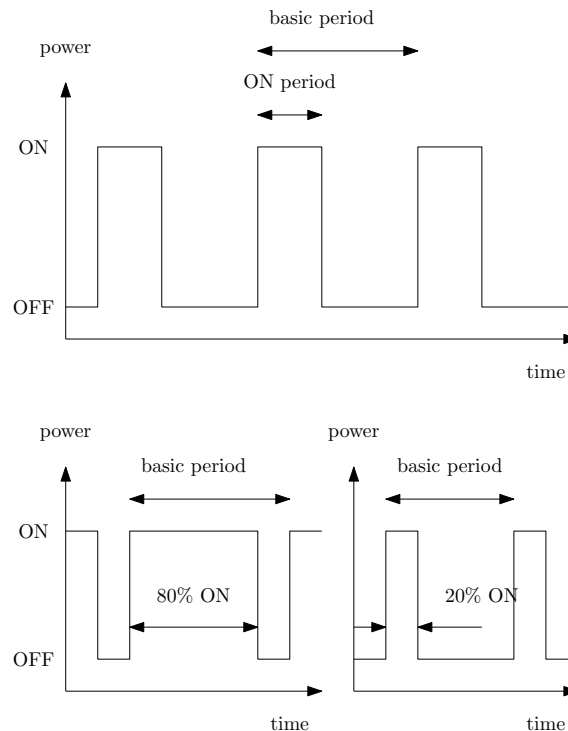
Κινητήρες και Έλεγχος Κινητήρων

Κάθε τροχός κινείται από ένα DC κινητήρα, συνδεδεμένο με τον αντίστοιχο τροχό, μέσω μείωσης 43.2 : 1. Ο ίδιος ο κινητήρας έχει μείωση 27 : 1, ενώ το κιβώτιο έχει μείωση 1.6 : 1. Ο κινητήρας έχει το δικό του βαθμονομημένο κωδικοποιητή, τοποθετημένο στον άξονά του, ο οποίος δίνει 16 παλμούς ανα πλήρη περιστροφή. Η διαθέσιμη ανάλυση είναι λοιπόν 691.2 ανα περιστροφή του τροχού, τιμή που αντιστοιχεί σε 54 παλμούς ανα 10mm κίνησης του ρομπότ (η διάμετρος του τροχού είναι 41mm, οπότε σε κάθε περιστροφή, το ρομπότ διανύει 128.8mm).

Οι κινητήρες οδηγούνται από δύο ανεξάρτητους ελεγκτές, υλοποιημένους σε κύκλωμα PIC18F4431. Το κύκλωμα έχει άμεσο έλεγχο της τροφοδοσίας του κινητήρα και μπορεί να αναγνώσει τους παλμούς του βαθμωτού κωδικοποιητή.

Κάθε ελεγκτής λειτουργεί σαν διακόπτης για τον αντίστοιχο κινητήρα, με συγκεκριμένη συχνότητα και σε συγκεκριμένες στιγμές. Έτσι ο κινητήρας ελέγχεται με βάση τη μέση τροφοδοσία, η οποία μεταβάλλεται αλλάζοντας την περίοδο για την οποία ο κινητήρας είναι ON. Αυτό σημαίνει πως μόνο ο λόγος μεταξύ των περιόδων ON και OFF μεταβάλλεται. Αυτή η μέθοδος ονομάζεται διαμόρφωση πλάτους παλμού (Pulse Width Modulation – PWM). Ορίζεται ως κύκλος λειτουργίας (duty cycle) το πηλίκο του χρόνου κατά τον οποίο ο κινητήρας τροφοδοτείται, προς τη συνολική διάρκεια της περιόδου του παλμού.

Οι ελεγκτές των κινητήρων μπορούν να πραγματοποιήσουν έλεγχο ταχύτητας και θέσης για τους κινητήρες, θέτοντας τη σωστή τιμή διαμόρφωσης πλάτους παλμού, σύμφωνα με την πραγματική ταχύτητα και θέση που παρέχεται από τους βαθμονομημένους κωδικοποιητές.



Σχήμα 7.2: Pulse Width Modulation - PWM [10]

Οι DC κινητήρες μπορούν να ελεγχθούν και μέσω PID ελεγκτή. Κάθε όρος του ελεγκτή (Αναλογικός, Ολοκλήρωσης, Παραγώγισης) συνδέεται με μια σταθερά, K_p , K_i , K_d , που

7.1. ΤΟ ΡΟΜΠΟΤ ΚΗΕΡΕΡΑ 3

αποτελεί το βάρος για τον κάθε όρο.

Η μέγιστη ταχύτητα που μπορεί να επιτευχθεί από το ρομπότ είναι $196\text{mm}/\text{sec}$ και η ελάχιστη $9\text{mm}/\text{sec}$.

Οδομετρία του Ρομπότ Khepera 3

Η ακτίνα τροχών του Khepera 3 είναι $r_{wheel} = 21\text{mm}$ και η απόσταση μεταξύ των τροχών είναι $D = 88.41\text{mm}$. Σύμφωνα με τα όσα ειπώθηκαν στο κεφάλαιο για τα ρομπότ διαφορετικής οδήγησης, αν V_{wheel} είναι το μέτρο της ταχύτητας των τροχών και αυτοί κινούνται σε αντίθετη φορά, τότε για να πραγματοποιηθεί περιστροφή του ρομπότ κατα γωνία ϕ (σε ακτίνια), η ταχύτητα θα πρέπει να εφαρμοστεί στους τροχούς για χρόνο :

$$t = \frac{\phi \pi D}{2\pi V_{wheel} r_{wheel}} \quad (7.1)$$

Αντίστοιχα, για ευθεία κίνηση, αν S είναι η απόσταση στην οποία θέλουμε να κινηθεί το ρομπότ :

$$t = \frac{D}{V_{wheel} r_{wheel}} \quad (7.2)$$

7.1.3 Σύστημα Αισθητήρων

Αισθητήρες Υπερύθρων

Το σύστημα αισθητήρων υπερύθρων, αποτελείται από έντεκα πομποδέκτες υπερύθρου φωτός, TCRT5000 της Vishay Telefunken. Οι τιμές που επιστρέφουν είναι των 12 bits . Η θέση τους στο ρομπότ είναι τέτοια, ώστε οι έξι από αυτούς να καλύπτουν το πρόσθιο ημικύκλιο του ρομπότ και οι άλλοι τρεις να καλύπτουν το οπίσθιο. Δύο αισθητήρες χρησιμοποιούνται για παρακολούθηση γραμμής.

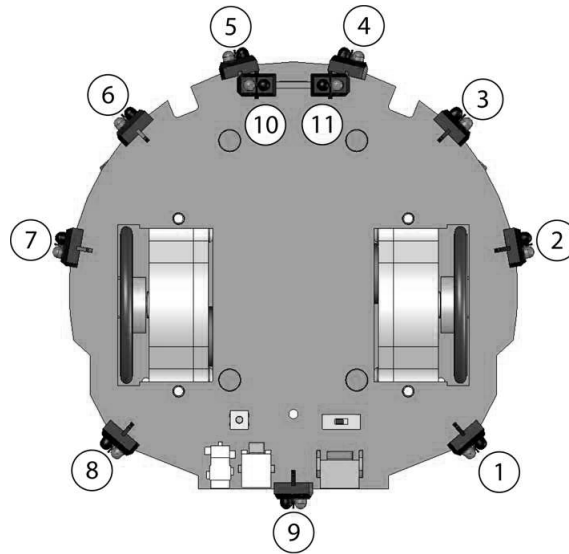
Η αρίθμηση των αισθητήρων, καθώς και η ανάγνωσή τους, γίνεται με αντίθετη ωρολογιακή φορά, κοιτώντας το ρομπότ από τη βάση του :

Οι αισθητήρες, έχουν 2 καταστάσεις λειτουργίας :

- Στη πρώτη κατάσταση λειτουργίας, οι αισθητήρες μετρούν το υπέρυθρο φως στο περιβάλλον λειτουργίας του ρομπότ, χωρίς εκπομπή φωτός από τους ίδιους.

Αυτή η κατάσταση λειτουργίας μπορεί να χρησιμοποιηθεί τόσο για κίνηση του ρομπότ με χρήση φάρων υπερύθρου φωτός σαν οδηγούς, όσο και για τον περιορισμό του θορύβου που υπεισέρχεται στις μετρήσεις των, λόγω του φωτισμού (φυσικού ή τεχνητού) που υπάρχει στο χώρο λειτουργίας.

Ο χρόνος που μεσολαβεί μεταξύ των αναγνώσεων δύο διαδοχικών αισθητήρων είναι στα 3 ms και ο χρήστης έχει καινούρια δεδομένα για ολόκληρη την ομάδα των αισθητήρων κάθε 33ms . Ανά πάσα στιγμή η διαθέσιμη μέτρηση είναι η προηγούμενη μέτρηση που έλαβε χώρα.



Σχήμα 7.3: Διάταξη των αισθητήρων υπεριώθρων του Khepera 3 [10]

- Στη δεύτερη κατάσταση λειτουργίας, οι αισθητήρες μετρούν το φώς που αντανακλάται από αντικείμενα που βρίσκονται στο χώρο.

Οι αισθητήρες εκπέμπουν υπέρυθρο φώς και επιστρέφουν σαν μέτρηση, τη διαφορά ανάμεσα στην τιμή υπεριώθρου φωτός που επέστρεψε λόγω ανάκλασης και την τιμή υπεριώθρου φωτός του περιβάλλοντος λειτουργίας.

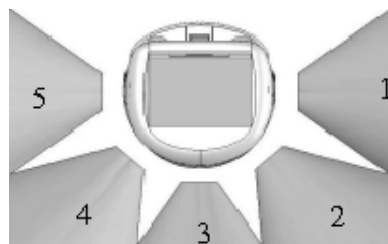
Σε ότι αφορά το χρόνο μεταξύ των αναγνώσεων των αισθητήρων, ισχύει ότι ισχύει και για την κατάσταση λειτουργίας μέτρησης υπέρυθρου φωτός στο χώρο.

Αισθητήρες Υπερήχων

Το ρομπότ είναι εφοδιασμένο με 5 αισθητήρες υπερήχων της Midas Component με πομποδέκτη τον 400ST100 και δέκτη τον 400SR100.

Οι αισθητήρες υπερήχων ανιχνεύουν αντικείμενα σε αποστάσεις από 20cm μέχρι 4m. Κάθε μέτρηση επιστρέφει τον αριθμό των ανιχνεύσεων με τη μέθοδο της ανάκλασης, την απόσταση στην οποία συνέβη η ανάκλαση σε cm, το πλάτος του ανακλούμενου κύματος και το χρόνο που αυτό ελήφθη. Όσον αφορά τον λευκό θόρυβο, αυτός ανιχνεύεται από τον δέκτη και αφαιρείται από το πλάτος του κύματος που λαμβάνει ο πομποδέκτης.

Οι θέσεις των αισθητήρων υπερήχων, σε κάτοψη, φαίνονται στο παρακάτω σχήμα :



Σχήμα 7.4: Διάταξη των αισθητήρων υπερήχων του Khepera 3 [10]

7.1.4 Σύστημα Τροφοδοσίας και Σύνδεση με Υπολογιστή

Το ρομπότ είναι εφοδιασμένο με 2 μπαταρίες Li-Ion, που παρέχουν 1400mAh στα 7.4V.

Όταν χρησιμοποιείται το βασικό ρομπότ (χωρίς κάποια από τις διαθέσιμες επεκτάσεις), αυτό μεταφράζεται σε σχεδόν 4 ώρες λειτουργίας, χωρίς να χρειαστεί επαναφόρτιση, σύμφωνα με τον κατασκευαστή. Όπως με όλες τις μπαταρίες, η απόδοσή τους μειώνεται με την πάροδο των χρόνων.

Πέρα από την αυτόνομη λειτουργία με τη χρήση μπαταριών, υπάρχει και η δυνατότητα λειτουργίας, με σύνδεση σε εξωτερική πηγή ρεύματος. Για αυτό το λόγο μαζί με το ρομπότ παρέχεται μετασχηματιστής AC/DC.

Η σύνδεση μεταξύ Khepera 2 και υπολογιστή, γίνεται μέσω σειριακής θύρας RS232, σύνδεσης USB ή ασύρματα, μέσω σύνδεσης Bluetooth.

Η σύνδεση RS232 υστερεί της σύνδεσης USB, ειδικά αν αναλογιστεί κανείς τη δυνατότητα της τελευταίας για ταυτόχρονη τροφοδοσία με ρεύμα της συνδεδεμένης συσκευής αλλά και τη μεγαλύτερη ταχύτητα μεταφοράς δεδομένων που παρέχει. Παρόλαυτα οι σειριακές θύρες με την απλότητα που τις διακρίνει, αποτελούν μια ικανοποιητική λύση για ένα εκπαιδευτικό ρομπότ.

Λογισμικό

Ο παρεχόμενος, από τον κατασκευαστή, οπτικός δίσκος, προσφέρει προγράμματα, περιβάλλοντα εργασίας και βιβλιοθήκες, συμβατά (κατά περίπτωση) με τα λειτουργικά συστήματα MAC OS, Windows, Linux, Unix, Solaris που αποτελούν και τα πλέον διαδεδομένα στον χώρο των υπολογιστών.

Για τις ανάγκες της διπλωματικής αυτής εργασίας, χρησιμοποιήθηκαν οι παρεχόμενες βιβλιοθήκες για την επικοινωνία ρομπότ και υπολογιστή, σε προγραμματιστικό περιβάλλον Matlab. Οι συγκεκριμένες βιβλιοθήκες περιείχαν αρχεία κλειστού κώδικα (επέκταση .dll) που είναι συμβατά μόνο με Windows και συνεπώς, η ανάπτυξη έγινε σε αυτό το λειτουργικό σύστημα.

Επεκτάσεις

Στο επάνω μέρος του Khepera 3 βρίσκονται υποδοχές σύνδεσης επεκτάσεων.

Αυτές είναι διαθέσιμες από την εταιρεία (μεταξύ των επιλογών βρίσκονται, διαφορετικές κάμερες, ένας βραχίονας, επεξεργαστής, μετρητής απόστασης laser), ενώ υπάρχει και η δυνατότητα ανάπτυξης καινούριων από τρίτους, ανατρέχοντας στα τεχνικά φύλλα που περιγράφουν τις υποδοχές σύνδεσης.

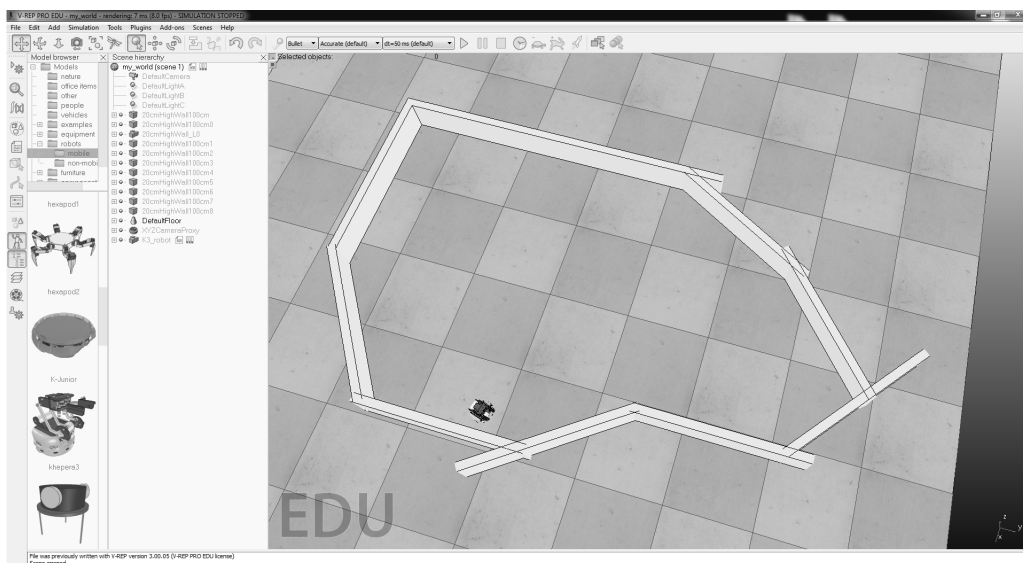
7.2 Λογισμικό V-REP

Το περιβάλλον εξομίωσης ρομπότ V-REP της Coppelia Robotics αποτελεί μια πλατφόρμα για την υλοποίηση ελέγχου σε διαφορετικά μοντέλα ρομπότ. Διαθέτει βιβλιοθήκες για επικοινωνία με τις περισσότερες δημοφιλείς γλώσσες προγραμματισμού, όπως C/C++,

Python, Java, Lua, Matlab, Octave και Rubi. Χρησιμοποιείται, μεταξύ άλλων, για την ανάπτυξη αλγορίθμων σε εξομοίωση εργοστασιακών αυτοματισμών, προτυποποίηση και επαλήθευση ερευνητικών ρομποτικών προγραμμάτων και εξ' αποστάσεως παρακολούθηση πραγματικών ρομπότ.

Διαθέτει βιβλιοθήκη μοντέλων για πολλά από τα δημοφιλή εκπαιδευτικά και βιομηχανικά ρομπότ, εργαλεία κινηματικής ανάλυσης, μοντέλα αισθητήρων, ενώ εξομοιώνει φυσικά δυναμικά συστήματα υποστηρίζοντας τρεις διαφορετικές μηχανές : τη Bullet physics library, την Open Dynamics Engine και τη Vortex Dynamics.

Η επιλογή του έγινε με γνώμονα όχι μόνο την επάρκειά του, αλλά και με βάση το γεγονός, πώς, σε αντίθεση με άλλα εμπορικά περιβάλλοντα εξομοίωσης, διατίθεται δωρεάν για ερευνητικούς, μη εμπορικούς σκοπούς, χωρίς περιορισμούς στο χρόνο χρήσης του και χωρίς κλειδωμένες λειτουργίες.



Σχήμα 7.5: Το περιβάλλον εξομοίωσης V-REP, με αρένα κίνησης και ρομπότ Khepera 3

Κεφάλαιο 8

Βαθμονόμηση Αισθητήρων Πραγματικού Ρομπότ

Σε αυτό το κεφάλαιο μελετώνται τα χαρακτηριστικά των αισθητήρων του πραγματικού ρομπότ. Τα πειραματικά δεδομένα αισθητήρων, πάρθηκαν πάνω στο ρομπότ Khepera 2, το οποίο σε γενικές γραμμές έχει τα χαρακτηριστικά του Khepera 3.

Η κυριότερη διαφορά που οδήγησε στη χρήση ρομπότ Khepera 3 σε περιβάλλον εξομοίωσης, είναι το γεγονός πως οι αισθητήρες υπέρυθρων του Khepera 3 παρέχουν τη δυνατότητα ανίχνευσης αντικειμένων σε μεγαλύτερες αποστάσεις και με μεγαλύτερη ακρίβεια. Για την υλοποίηση της Ασαφούς Λογικής με Ενισχυτική Μάθηση και τη λήψη ουσιαστικών αποτελεσμάτων, αυτό ήταν απαραίτητο.

Παρόλαυτα η αρχή λειτουργίας των αισθητήρων παραμένει αμετάβλητη μεταξύ των δύο τύπων και η παρουσίασή της είναι επιβεβλημένη. Επιπλέον η υλοποίηση της απλής Ασαφούς Λογικής υλοποιήθηκε με επιτυχία τόσο στο πραγματικό ρομπότ Khepera 2, όσο και στο ρομπότ Khepera 3 σε περιβάλλον εξομοίωσης.

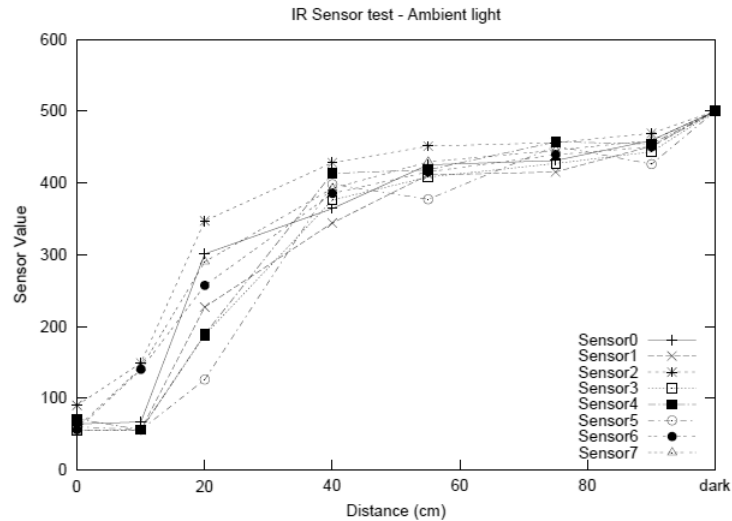
Πρέπει να τονιστεί εδώ, πως λόγω της θορυβώδους φύσης τους, οι υπέρυθροι αισθητήρες χρησιμοποιούνται συνήθως σαν αισθητήρες εγγύτητας και όχι για μέτρηση απόστασης. Σαν αισθητήρες εγγύτητας περιγράφονται και στα εγχειρίδια των δύο ρομπότ.

8.1 Δεδομένα κατασκευαστή

Το Khepera 2 είναι εφοδιασμένο με 8 αισθητήρες υπέρυθρων TCRT1000 της Vishay Telefunken. Σύμφωνα με το εγχειρίδιο του κατασκευαστή οι αισθητήρες παρουσιάζουν μεγάλη ευαισθησία στο φωτισμό του περιβάλλοντος και πιο συγκεκριμένα σε εκπομπές πηγών φωτός κοντά στο υπέρυθρο φάσμα. Ουσιαστικά οι αισθητήρες πρέπει να παραμείνουν μακριά, τόσο από απευθείας φυσικό ηλιακό φως, όσο και από τεχνητό φωτισμό εσωτερικών χώρων.

Το γράφημα 8.1 δείχνει τις μετρήσεις του κατασκευαστή, για το μετρούμενο από τους αισθητήρες, φως περιβάλλοντος, για πηγή φωτός ισχύος 50W τοποθετημένη σε διάφορες αποστάσεις από τον αισθητήρα.

Οι επιστρεφόμενες, από τους αισθητήρες, τιμές, αυξάνονται, καθώς η ένταση του φωτός μειώνεται. Η μέγιστη τιμή επιτυγχάνεται σε συνθήκες σχεδόν σκότους. Σύμφωνα με τον κατασκευαστή οι ιδανικές συνθήκες φωτισμού είναι υπο ομοιόμορφο, σταθερό και καθαρό

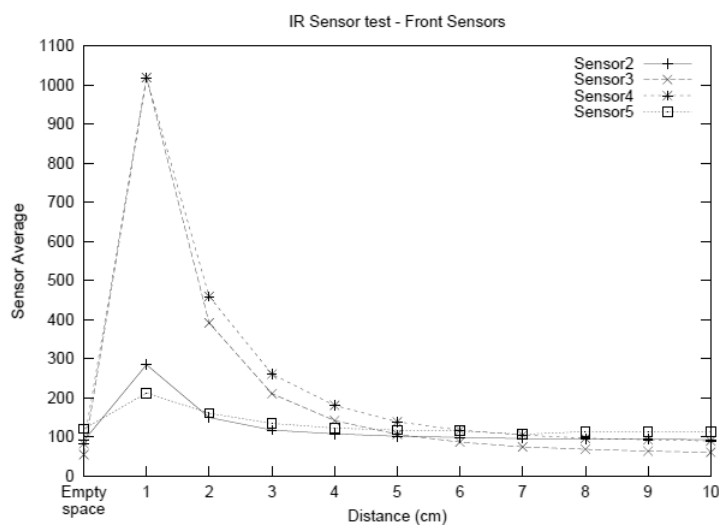


Σχήμα 8.1: Μετρήσεις περιβαλλοντικού φωτός χρησιμοποιώντας πηγή ισχύος 50W [27]

φως, μικρής έντασης.

Για τη μέτρηση απόστασης, οι αισθητήρες μετρούν το επιστρεφόμενο, λόγω ανάκλασης σε επιφάνειες, φως. Το χρώμα, το υλικό και η επιφάνεια των αντικειμένων επηρεάζουν την ανταπόκριση των αισθητήρων, ενώ σε συνδυασμό με τον θόρυβο (λόγω φωτός) από το περιβάλλον, οι μετρήσεις που έχουν παρθεί από τον κατασκευαστή δεν μπορούν να θεωρηθούν σαν αναφορά, παρά μόνο σαν γενικές πληροφορίες ανταπόκρισης των αισθητήρων, για τη μέθοδο βαθμονόμησης.

Στις μετρήσεις απόστασης του κατασκευαστή, χρησιμοποιήθηκε λευκό χαρτί διαστάσεων $5 \times 5 \text{ cm}$ ώστε να επιτευχθεί ικανοποιητική ανάκλαση. Η επιφάνεια τοποθετήθηκε παράλληλα προς τον αισθητήρα και το γράφημα των αποτελεσμάτων φαίνεται στο σχήμα 8.2.

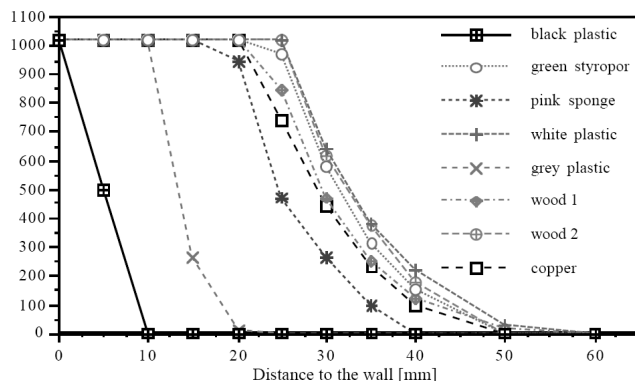


Σχήμα 8.2: Μετρήσεις απόστασης των αισθητήρων σε λευκό χαρτί [11]

Ο κατασκευαστής παρέχει επίσης δεδομένα μέτρησης απόστασης των αισθητήρων, για διαφορετικές επιφάνειες. Επιφάνειες που παρουσιάζουν ανακλαστικά χαρακτηριστικά (λευ-

8.2. ΔΕΔΟΜΕΝΑ ΠΡΑΓΜΑΤΙΚΟΥ ΡΟΜΠΟΤ

κές επιφάνειες, κατά προτίμηση γυαλιστερές), επιτρέπουν μετρήσεις σε μεγαλύτερες αποστάσεις. Τα δεδομένα του κατασκευαστή, φαίνονται στο σχήμα 8.3.



Σχήμα 8.3: Μετρήσεις απόστασης των αισθητήρων σε διαφορετικές επιφάνειες [26]

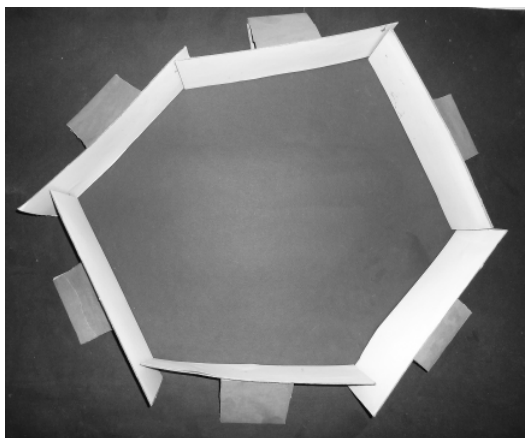
8.2 Δεδομένα πραγματικού ρομπότ

Όπως έχει ήδη φανεί από τα δεδομένα του κατασκευαστή, οι αισθητήρες υπερύθρων παρουσιάζουν ιδιαίτερη ευαισθησία στο θόρυβο και στο είδος της επιφάνειας που χρησιμοποιείται για την ανίχνευση.

Για τη βαθμονόμηση των αισθητήρων στο ρομπότ, λαμβάνοντας υπόψη τις παραπάνω ευαισθησίες, κατασκευάστηκε αρένα για την υλοποίηση του αλγορίθμου ασαφούς λογικής:

- Για το δάπεδο επιλέχθηκε χαρτί μαύρου μιάτ χρώματος έτσι ώστε να μειωθούν στο ελάχιστο οι αντανακλάσεις από εκεί.
- Για τα τμήματα τοίχου επιλέχθηκε λευκό, ανακλαστικό χαρτί εκτύπωσης φωτογραφιών, της Kodak ώστε να είναι όσο το δυνατόν καλύτερη η ανίχνευση από τους αισθητήρες.

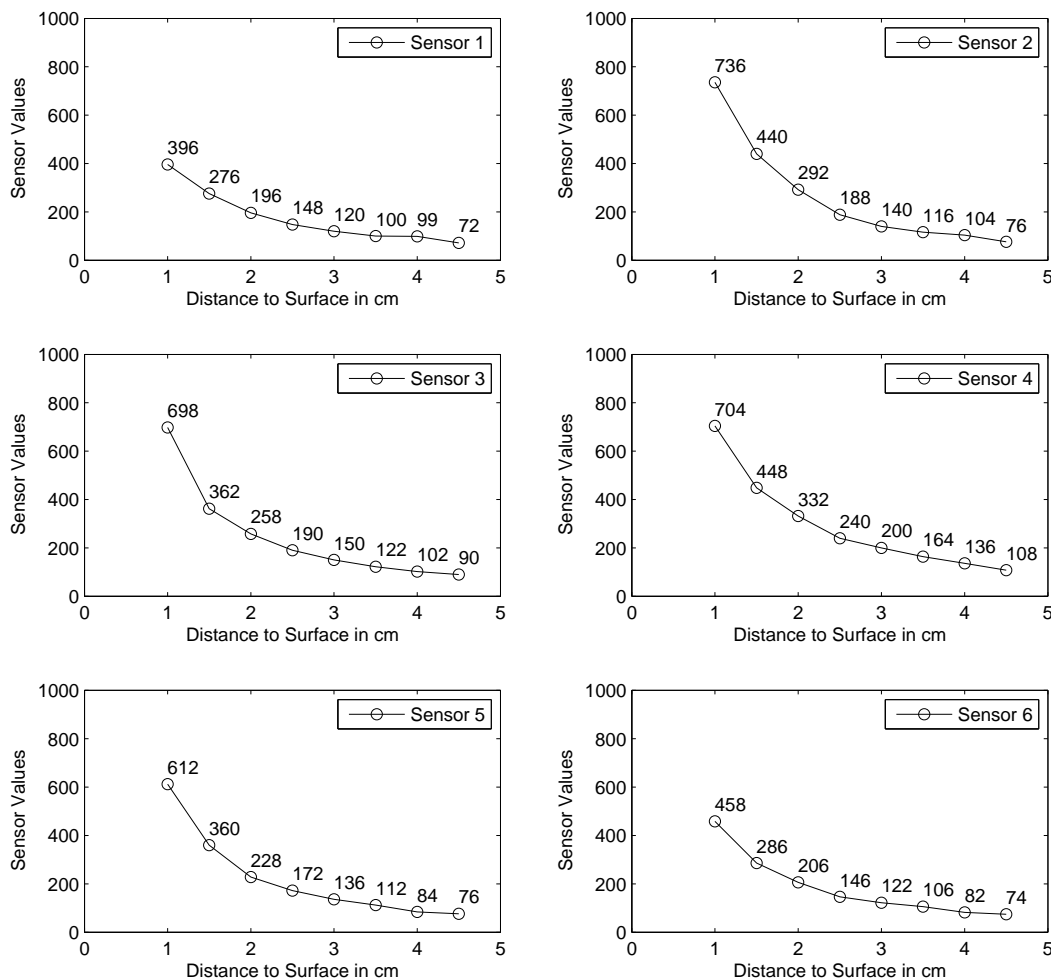
Η αρένα φαίνεται στο σχήμα 8.4.



Σχήμα 8.4: Αρένα που κατασκευάστηκε

Επειδή οι αισθητήρες του ρομπότ έχουν διαφορές μεταξύ τους (κάποιοι δεν ήταν τοποθετημένοι επακριβώς κάθετα), έπρεπε να ληφθούν μετρήσεις για κάθε έναν ξεχωριστά. Μετρήσεις έγιναν στο διάστημα 1 έως 4.5cm ανά 0.5cm. Για κάθε απόσταση ελήφθησαν πολλαπλές μετρήσεις λόγω θορύβου και τα έπειτα βρέθηκε ο μέσος όρος των μετρήσεων.

Ακολουθούν τα γραφήματα των τιμών του κάθε αισθητήρα, σε σχέση με την απόσταση (6 αισθητήρες μια και αυτοί χρησιμοποιήθηκαν για την ασαφή λογική) :



Σχήμα 8.5: Μετρήσεις Αισθητήρων του πραγματικού ρομπότ

8.3 Αξιολόγηση των Αισθητήρων

Η μέγιστη απόσταση ανίχνευσης των αισθητήρων του Khepera 2, δεν δίνει τη δυνατότητα ικανοποιητικής κατάτμησης του διαστήματος, για την υλοποίηση της Ασαφούς Λογικής και ακόμα περισσότερο για την υλοποίηση της Ασαφούς Λογικής με Ενισχυτική Μάθηση. Αυτό είναι προφανές στο γράφημα 8.5. Οι αισθητήρες παρουσιάζουν τεράστιες διαφορές μεταξύ τους και ειδικότερα, όσο μεγαλώνει η απόσταση, μικραίνει και η διακριτική τους ικανότητα.

8.3. ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΑΙΣΘΗΤΗΡΩΝ

Με δεδομένη την ανακρίβεια των υπερύθρων αισθητήρων πρέπει να χρησιμοποιηθεί κάποια μέθοδος διαχείρισης των δεδομένων τους. Αν και στη μέτρηση απόστασης, τα δεδομένα μπορούν να χρησιμοποιηθούν εύκολα, παρά τις ανακρίβειες, η γωνία ανίχνευσης τοίχου από έναν αισθητήρα δεν είναι αξιόπιστη, λόγω πιθανής απόκλισης της γωνίας τοποθέτησης του αισθητήρα αλλά και λόγω της επιφάνειας και πιθανής κλίσης του ανιχνευόμενου αντικειμένου.

Κεφάλαιο 9

Υλοποίηση Ασαφούς Λογικής

Ο ασαφής ελεγκτής σχεδιάστηκε πάνω στις παραδοχές πως δεν υπάρχουν κινητοί τοίχοι ή εμπόδια και πως δεν υπάρχουν τυφλά σημεία για τους αισθητήρες. Ο τοίχος είναι πάντα αντιληπτός από τουλάχιστον έναν από τους αισθητήρες, ανεξάρτητα από τον προσανατολισμό του ρομπότ.

9.1 Είσοδοι

Ο ασαφής ελεγκτής αποτελείται από μια βάση κανόνων που συνθέτεται από 20 κανόνες. Οι κανόνες αυτοί αναπαριστούν οδηγίες για το ρομπότ, όσον αφορά τη συμπεριφορά του υπό συγκεκριμένες συνθήκες. Οι μεταβλητές εισόδου είναι η γωνία α μεταξύ του διανύσματος ταχύτητας του ρομπότ και του διανύσματος γενικής αντίληψης και το μέτρο p του διανύσματος γενικής αντίληψης.

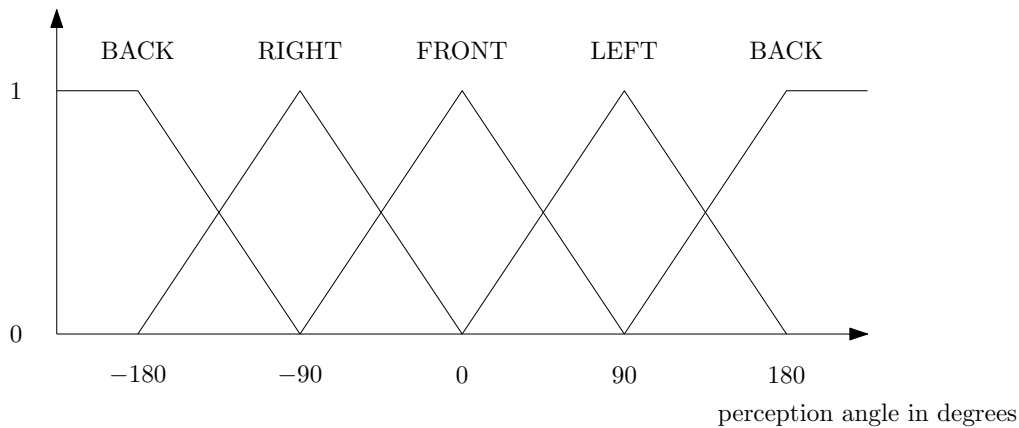
Το ρομπότ εκτιμά μια κατάσταση, χρησιμοποιώντας το διάνυσμα της γενικής αντίληψης. Ονομάζουμε :

- Η γωνία α μεταξύ του διανύσματος ταχύτητας του ρομπότ και του διανύσματος γενικής αντίληψης ονομάζεται "Γωνία Αντίληψης"
- Το μέτρο p του διανύσματος γενικής αντίληψης ονομάζεται "Αντίληψη"

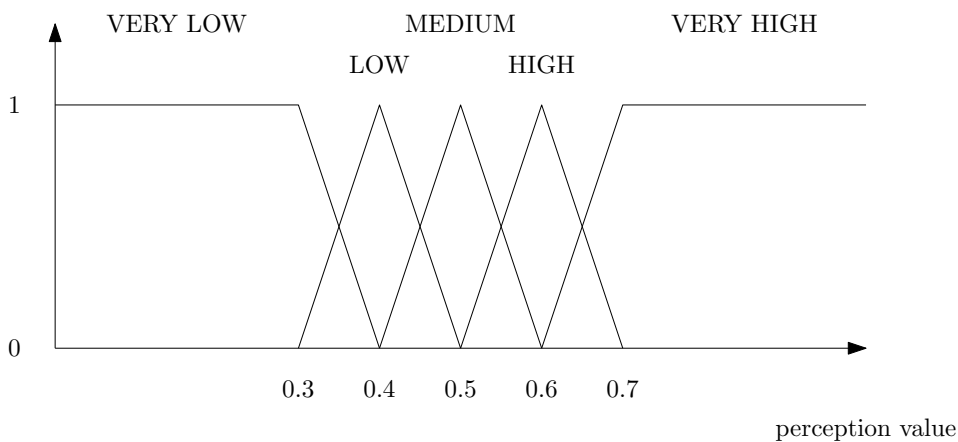
Οι καταστάσεις περιγράφονται γλωσσικά. Για παράδειγμα, αν η γωνία αντίληψης είναι 45° , μια πιθανή γλωσσική περιγραφή είναι "ΑΡΙΣΤΕΡΑ ΚΑΙ ΜΠΡΟΣΤΑ". Στις 60° , η γωνία αντίληψης εξακολουθεί να είναι "ΑΡΙΣΤΕΡΑ ΚΑΙ ΜΠΡΟΣΤΑ", αλλά σε μικρότερο βαθμό. Ο βαθμός στον οποίο κάθε κατάσταση περιγράφεται από τη γλωσσική περιγραφή "ΑΡΙΣΤΕΡΑ ΚΑΙ ΜΠΡΟΣΤΑ" περιγράφεται από τη συνάρτηση συμμετοχής.

Η γωνία αντίληψης υλοποιήθηκε με τέσσερις γλωσσικές περιγραφές που υπερκαλύπτονται. Κυμαίνονται μεταξύ της περιγραφής "ΠΙΣΩ" και της περιγραφής "ΜΠΡΟΣΤΑ". Οι συναρτήσεις συμμετοχής, φαίνονται στο σχήμα 9.1 και η κατάτμηση είναι ισχυρή. Να τονιστεί εδώ πως η περιγραφή "ΠΙΣΩ", αν και εμφανίζεται δύο φορές αριστερά και δεξιά, αποτελεί μια και μόνη περιγραφή.

Η αντίληψη υλοποιήθηκε με πέντε γλωσσικές περιγραφές που υπερκαλύπτονται. Κυμαίνονται μεταξύ της περιγραφής "ΧΑΜΗΛΗ" και της περιγραφής "ΥΨΗΛΗ". Οι συναρτήσεις συμμετοχής, φαίνονται στο σχήμα 9.2 και η κατάτμηση είναι επίσης ισχυρή.



Σχήμα 9.1: Γωνία Αντίληψης α



Σχήμα 9.2: Αντίληψη p

9.2 Έξοδοι

Οι μεταβλητές εξόδου της βάσης κανόνων είναι επίσης ασαφείς μεταβλητές. Η γωνία οδήγησης, επιλέγεται ανάλογα με την είσοδο και έχει πέντε γλωσσικές μεταβλητές. Για παράδειγμα, αν η αντίληψη τοίχου είναι "ΥΨΗΛΗ" και ο τοίχος βρίσκεται "ΑΡΙΣΤΕΡΑ" τότε επιθυμούμε το ρομπότ να εκτελέσει "ΓΡΗΓΟΡΗ ΔΕΞΙΑ" στροφή.

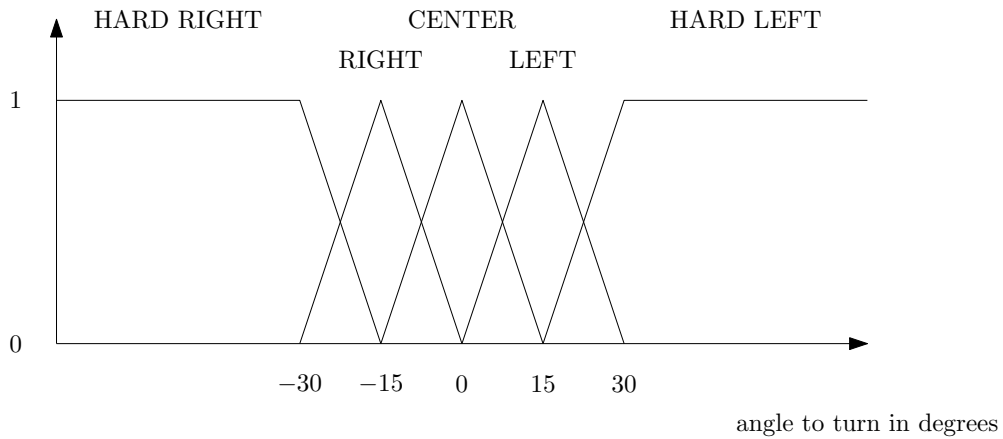
Η γωνία οδήγησης υλοποιήθηκε με 5 γλωσσικές περιγραφές, οι οποίες κυμαίνονται μεταξύ της περιγραφής "ΓΡΗΓΟΡΗ ΔΕΞΙΑ" και της περιγραφής "ΓΡΗΓΟΡΗ ΑΡΙΣΤΕΡΗ". Οι συναρτήσεις συμμετοχής για τη γωνία οδήγησης, φαίνονται στο σχήμα 9.3 και η κατάτμηση είναι ισχυρή.

9.3 Βάση Κανόνων

Η βάση κανόνων αποτελείται, όπως ήδη αναφέρθηκε, από 20 κανόνες. Οι κανόνες είναι της μορφής :

$$AN \langle \text{προυπόθεση} \rangle \text{ TOTE } \langle \text{συμπέρασμα} \rangle$$

Σε αντίθεση με τη Boolean λογική, η προϋπόθεση μπορεί να ικανοποιείται μερικώς. Το συμπέρασμα αυτού του κανόνα θα ικανοποιείται μερικώς επίσης. Ο τελεστής ΚΑΙ συνδυάζει



Σχήμα 9.3: Γωνία Οδήγησης

προυποθέσεις του κανόνα μέσω της νόρμας \min και ο τελεστής H συνδυάζει προϋποθέσεις του κανόνα μέσω της νόρμας \max .

Στην περίπτωση ενεργοποίησης πολλών κανόνων συγχρόνως, πρέπει να βρεθεί ένας έλεγχος που να συνδυάζει τις αντιφατικές εξόδους των κανόνων. Στην υλοποίηση μας χρησιμοποιήθηκε η μέθοδος της συσχετιστικής άθροισης με κεντροειδή αποασαφοποίησηση.

Οι κανόνες που χρησιμοποιήθηκαν έχουν τη γωνία αντίληψης και την αντίληψη σαν προϋποθέσεις και τη γωνία οδήγησης σαν συμπέρασμα. Παραδείγματος χάριν :

ΑΝ γωνία αντίληψης "ΑΡΙΣΤΕΡΑ" ΚΑΙ αντίληψη "ΥΨΗΛΗ"
ΤΟΤΕ γωνία οδήγησης "ΓΡΗΓΟΡΗ ΔΕΞΙΑ"

Στον πίνακα 9.1, φαίνεται η πλήρης βάση κανόνων για δεξιά ακολουθήση τοίχου.

	Πολύ Χαμηλή	Χαμηλή	Μέση	Υψηλή	Πολύ Υψηλή
Πίσω	<i>HR</i>	<i>R</i>	<i>R</i>	<i>C</i>	<i>C</i>
Δεξιά	<i>HR</i>	<i>R</i>	<i>C</i>	<i>L</i>	<i>HL</i>
Μπροστά	<i>C</i>	<i>C</i>	<i>L</i>	<i>HL</i>	<i>HL</i>
Αριστερά	<i>HR</i>	<i>HR</i>	<i>R</i>	<i>R</i>	<i>R</i>

Πίνακας 9.1: Βάση κανόνων

9.4 Αριθμητικά Αποτελέσματα Πραγματικού Ρομπότ

Ο ασαφής ελεγκτής υλοποιήθηκε σε πραγματικό ρομπότ Khepera 2. Το ρομπότ έχει 8 αισθητήρες αλλά οι 2 οπίσθιοι είναι τυφλοί ως προς τα πλαινά του ρομπότ, οπότε χρησιμοποιήθηκαν μόνο οι 6 πρόσθιοι. Ο έλεγχος γίνεται για τη γωνία στροφής του ρομπότ. Το ρομπότ εκτελεί ευθεία κίνηση για ορισμένη απόσταση (στην υλοποίηση 1cm), σταματά, ελέγχει την κατάσταση στην οποία βρίσκεται, μέσω του αλγορίθμου, επιλέγεται και εκτελείται η στροφή και συνεχίζεται η ευθεία κίνηση.

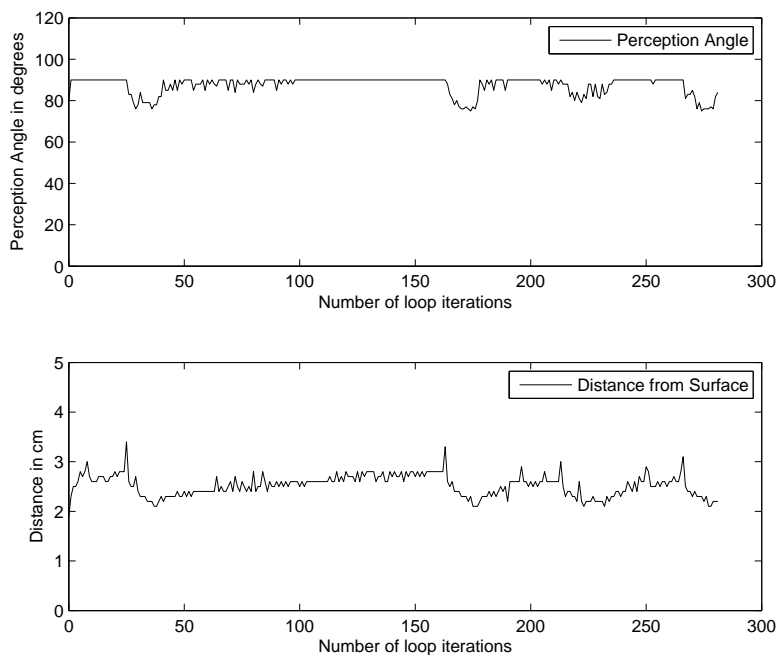
Το ρομπότ μπορούσε να ακολουθήσει τον τοίχο της αρένας, αλλά λόγω της κακής ανάλυσης και θορυβώδους φύσης των αισθητήρων, παρουσίαζε μικρό μόνιμο σφάλμα από την

9.4. ΑΡΙΘΜΗΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΑΓΜΑΤΙΚΟΥ ΡΟΜΠΟΤ

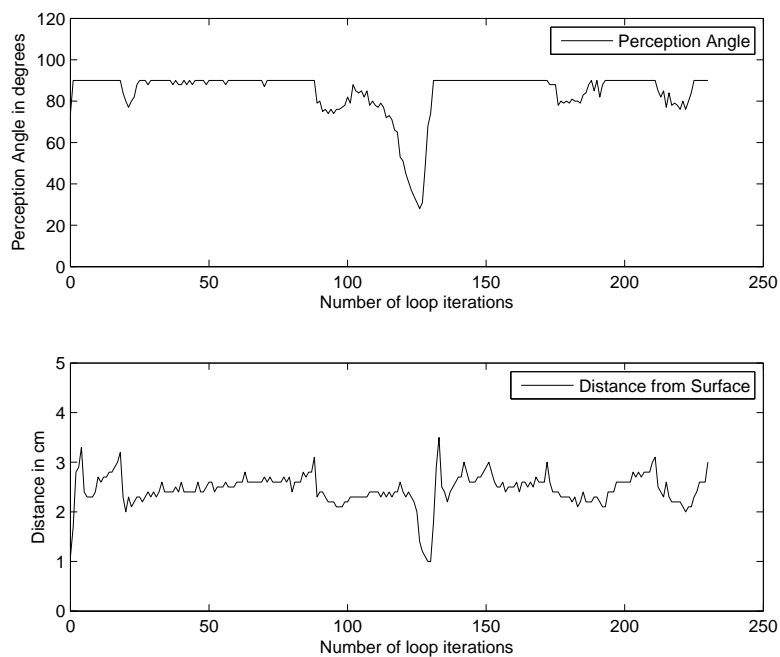
επιθυμητή απόσταση. Παρόλαυτα ο ασαφής έλεγχος λειτούργησε ικανοποιητικά. Η υλοποίηση έγινε για αριστερή ακολούθηση τοίχου.

Ο ελεγκτής υλοποιήθηκε σε περιβάλλον Matlab και η επικοινωνία με το ρομπότ έγινε μέσω των παρεχομένων από τον κατασκευαστή βιβλιοθηκών.

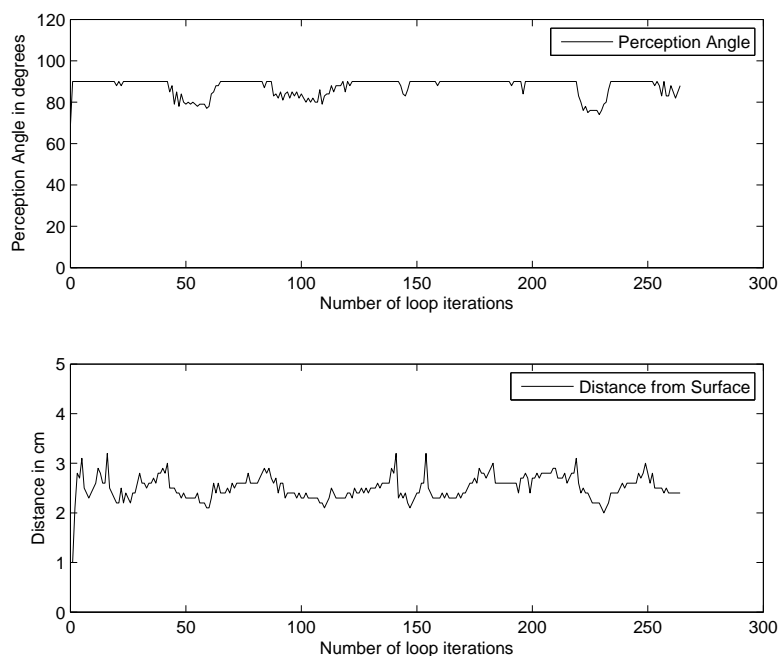
Στα σχήματα 9.4, 9.5, 9.6 φαίνεται η εξέλιξη της απόστασης και της γωνίας του ρομπότ σε σχέση με τον τοίχο, καθώς το ρομπότ εκτελεί διαφορετικούς τύπους γωνιών, ως προς τον αριθμό επαναλήψεων του αλγορίθμου ελέγχου. Από τα σχήματα φαίνεται πως ο Ασαφής Έλεγχος προσπαθεί να διατηρήσει την απόσταση του ρομπότ από τον τοίχο και ανταποκρίνεται στις αλλαγές αντίληψης για το περιβάλλον, χωρίς όμως να μπορεί να φτάσει στην απόσταση-στόχο. Ειδικά στην περίπτωση της ευθείας κίνησης του σχήματος 9.4, η επίδραση του θορύβου, είναι εμφανής με την απόκλιση να φτάνει το 1cm. Η γωνία αντίληψης παρουσιάζει καλύτερη συμπεριφορά, μια και σε αντίθεση με την απόσταση, όπου λαμβάνεται η τιμή του αισθητήρα με την πιο κοντινή ένδειξη, η γωνία εξάγεται συνθέτοντας τις ενδείξεις, όλων των αισθητήρων που ανιχνεύουν κάτι, με αποτέλεσμα μεγαλύτερη αξιοπιστία μέτρησης.



Σχήμα 9.4: Γωνία και απόσταση σε σχέση με τον τοίχο σε ευθεία κίνηση, υπό Ασαφή Λογική για πραγματικό ρομπότ



Σχήμα 9.5: Γωνία και απόσταση σε σχέση με τον τοίχο κατά την εκτέλεση κλειστής αριστερής γωνίας, υπό Ασαφή Λογική για πραγματικό ρομπότ



Σχήμα 9.6: Γωνία και απόσταση σε σχέση με τον τοίχο κατά την εκτέλεση αμβλείας αριστερής γωνίας, υπό Ασαφή Λογική για πραγματικό ρομπότ

9.5 Αριθμητικά Αποτελέσματα του Ρομπότ σε Περιβάλλον Εξομοίωσης

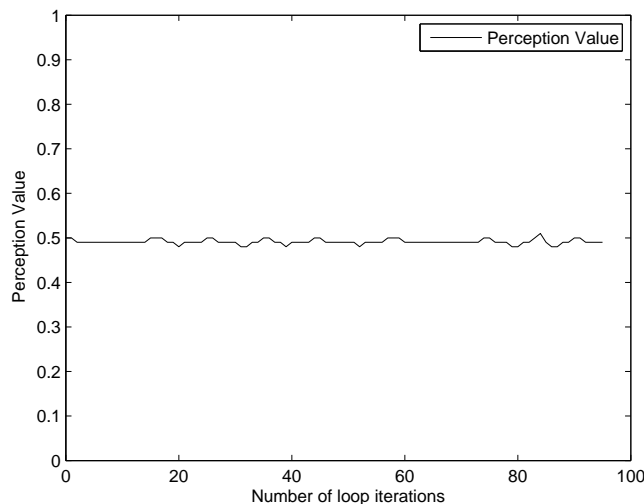
Ο Ασαφής Έλεγχος υλοποιήθηκε σε περιβάλλον εξομοίωσης V-REP για το ρομπότ Khepera 3. Οι καλύτεροι αισθητήρες του ρομπότ Khepera 3 δίνουν τη δυνατότητα υλοποίησης της Ασαφούς Ενισχυτικής Μάθησης στη συνέχεια. Ενώ για το Khepera 2 η μέγιστη απόσταση ανίχνευσης είναι στα 5cm, οι αισθητήρες του Khepera 3, επιτρέπουν ανίχνευση επιφανειών μέχρι τα 25cm.

Ο έλεγχος γίνεται για τη γωνία στροφής του ρομπότ. Το ρομπότ εκτελεί ευθεία κίνηση για ορισμένη απόσταση (στην υλοποίηση 1cm), σταματά, ελέγχει την κατάσταση στην οποία βρίσκεται, μέσω του αλγορίθμου, επιλέγεται και εκτελείται η στροφή και συνεχίζεται η ευθεία κίνηση. Ο Ασαφής Ελεγκτής δοκιμάστηκε σε αρένες με διαφορετικούς τύπους γωνιών, σε ευθεία κίνηση, σε τετράγωνη αρένα καθώς και σε αρένα με διαφορετικούς τύπους γωνιών. Παρουσιάζεται η γωνία στην οποία αντιλαμβάνεται τον τοίχο, η απόσταση από τον τοίχο, το σφάλμα από την επιθυμητή απόσταση, καθώς και το μονοπάτι τροχιάς που το ρομπότ ακολούθησε κατά την εκτέλεση.

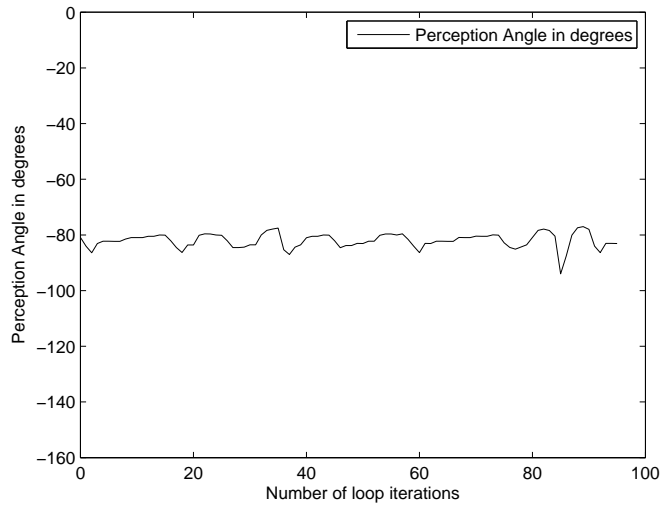
9.5.1 Αποτελέσματα για Ευθεία Κίνηση

Αρχικά δοκιμάστηκε η συμπεριφορά του Ασαφούς Ελέγχου για ευθεία κίνηση. Το βήμα εξομοίωσης ήταν 50ms.

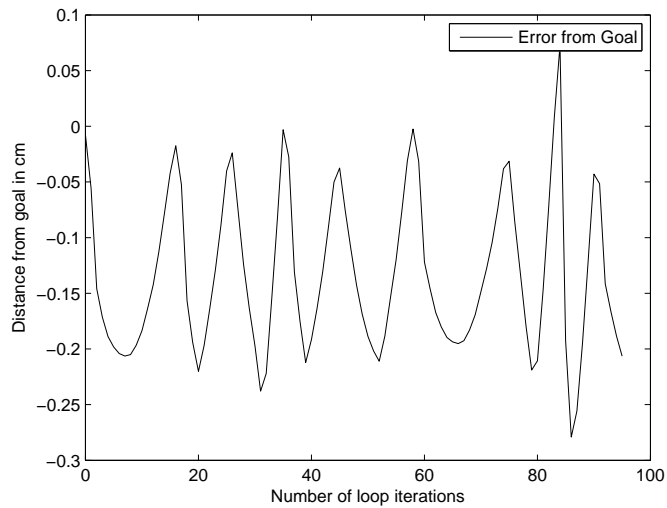
Στα σχήματα 9.7, 9.8, 9.9, φαίνεται η τιμή της Αντίληψης για τον τοίχο, η Γωνία Αντίληψης, το σφάλμα από την επιθυμητή απόσταση. Στο σχήμα 9.10, φαίνονται οι φορές που το ρομπότ κατάφερε να βρεθεί στην επιθυμητή απόσταση και λαμβάνει ανταμοιβή (η οποία εξηγείται και παρουσιάζεται στην υλοποίηση της Ασαφούς Ενισχυτικής Μάθησης).



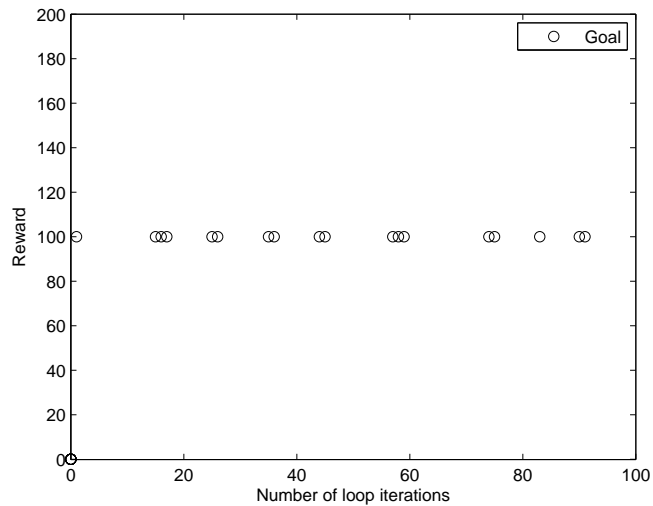
Σχήμα 9.7: Αντίληψη κατά την εκτέλεση ευθείας κίνησης, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης



Σχήμα 9.8: Γωνία Αντίληψης κατά την εκτέλεση ευθείας κίνησης, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης



Σχήμα 9.9: Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση ευθείας κίνησης, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης

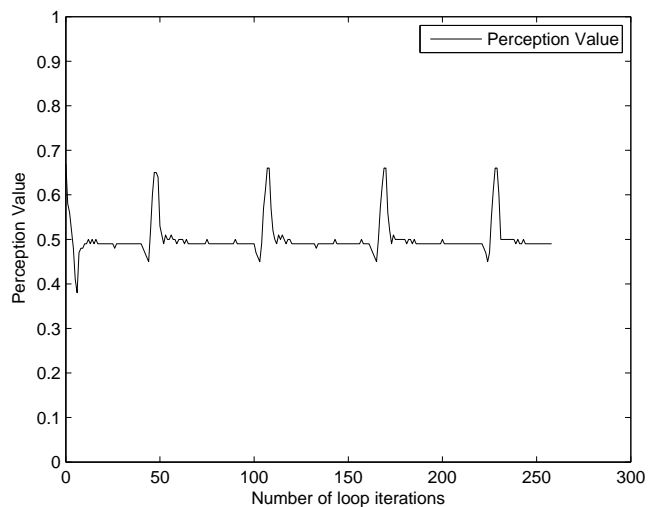


Σχήμα 9.10: Φορές που το ρομπότ κατάφερε να βρεθεί στην επιθυμητή απόσταση και λαμβάνει ανταμοιβή κατά την εκτέλεση ευθείας κίνησης, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης

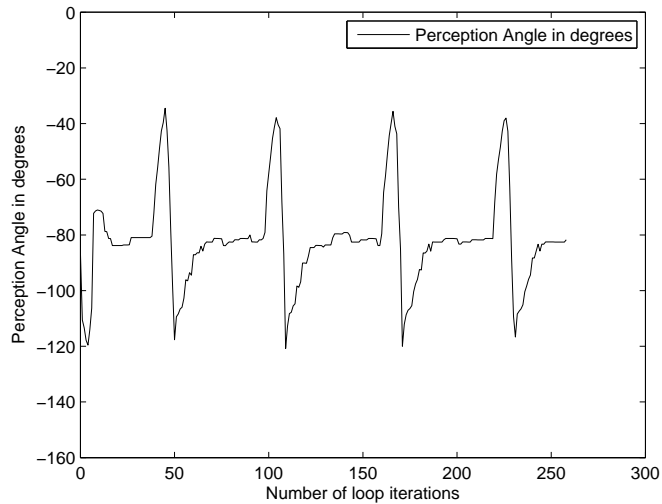
Από τα αποτελέσματα στον εξομοιωτή, φαίνεται πως ο Ασαφής Ελεγκτής ακολουθεί με επιτυχία τον τοίχο, επιτυγχάνοντας κάποιες φορές το στόχο. Η απόσταση, μένει κοντά στον στόχο με απόκλιση που φτάνει μέχρι τα $0.27cm$ και η γωνία αντίληψης έχει μικρές διακυμάνσεις.

9.5.2 Αποτελέσματα για Τετράγωνη Αρένα

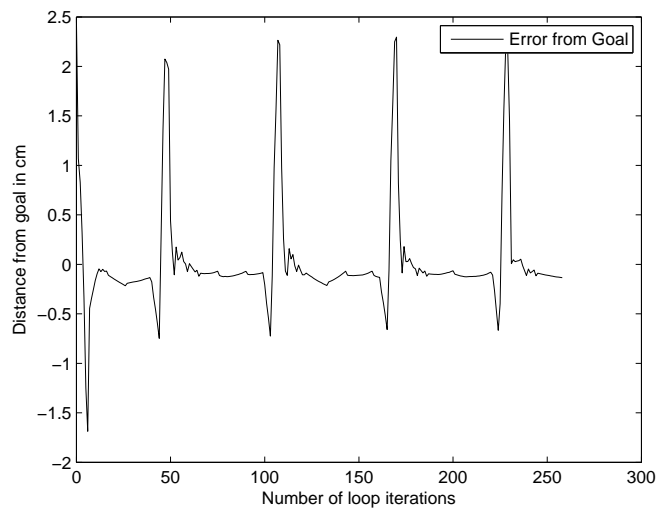
Εξετάζεται η περίπτωση κίνησης του Khepera 3 σε τετράγωνη αρένα. Η τετράγωνη αρένα δίνει τη δυνατότητα να παρατηρηθεί η συμπεριφορά του ρομπότ στη διόρθωση της απόστασης κατά την έξοδο από γωνία.



Σχήμα 9.11: Αντίληψη κατά την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης

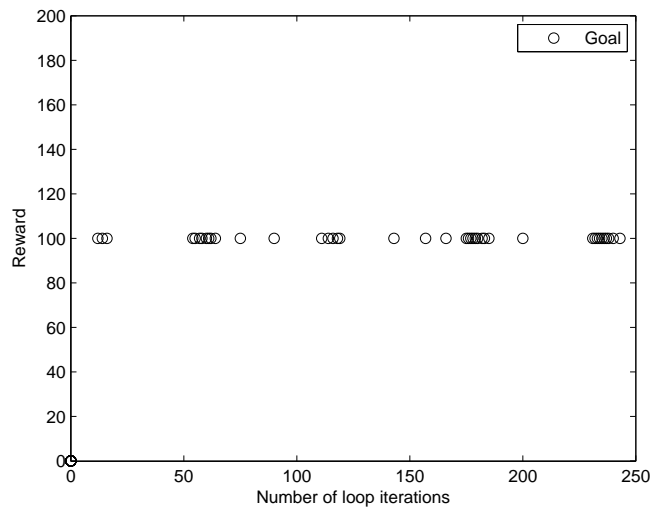


Σχήμα 9.12: Γωνία Αντίληψης κατά την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης

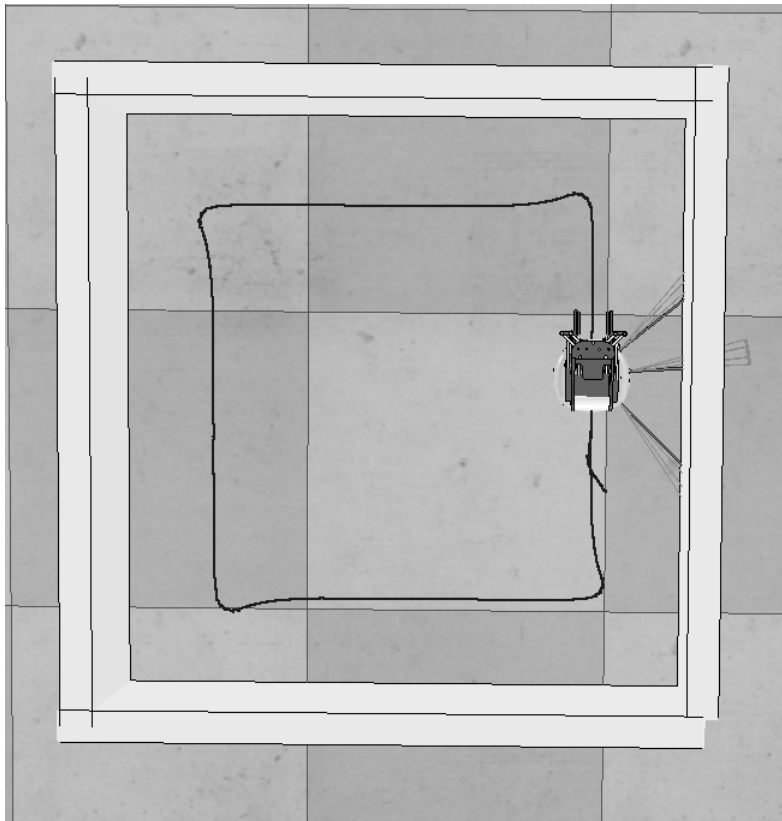


Σχήμα 9.13: Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης

9.5. ΑΡΙΘΜΗΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΟΥ ΡΟΜΠΟΤ ΣΕ ΠΕΡΙΒΑΛΛΟΝ ΕΞΟΜΟΙΩΣΗΣ



Σχήμα 9.14: Φορές που το ρομπότ κατάφερε να βρεθεί στην επιθυμητή απόσταση και λαμβάνει ανταμοιβή κατά την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης



Σχήμα 9.15: Μονοπάτι για τετράγωνη αρένα, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης

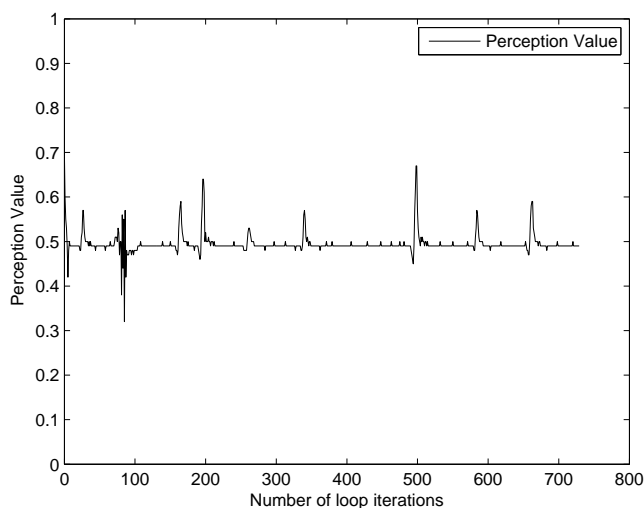
Στο σχήμα 9.15 φαίνεται το μονοπάτι που ακολούθησε το ρομπότ, για την περίπτωση τετράγωνης αρένας. Στο σχήμα 9.13, φαίνεται το σφάλμα από την επιθυμητή απόσταση, ενώ στα σχήματα 9.11, 9.12 φαίνονται αντίστοιχα η εξέλιξη της απόστασης του ρομπότ από τον τοίχο, με την πάροδο του χρόνου, καθώς και η γωνία αντίληψης. Στο σχήμα 9.14, φαίνονται οι φορές που το ρομπότ κατάφερε να βρεθεί στην επιθυμητή απόσταση.

Το ρομπότ διαχειρίζεται τις γωνίες ικανοποιητικά και μετά την έξοδο επιστρέφει κοντά στην απόσταση-στόχο. Όπως και στην περίπτωση της ευθείας κίνησης, υπάρχει ένα σφάλμα κατά την εκτέλεση των τμημάτων ευθείας, το οποίο οφείλεται στο γεγονός ότι, δεν γίνεται να επιλεγεί, με βάση την ανθρώπινη εμπειρία, η ιδανική συνάρτηση συμμετοχής εξόδου.

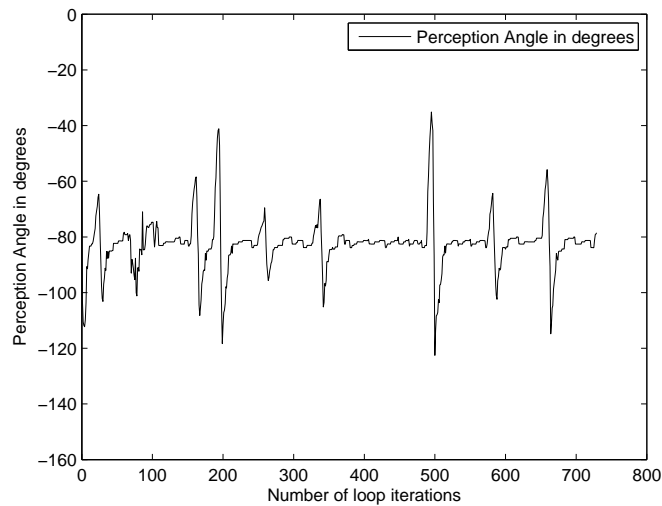
9.5.3 Αποτελέσματα σε Αρένα με Διαφορετικούς Τύπους Γωνιών

Κλείνοντας, παρουσιάζεται η συμπεριφορά του Ασαφούς Ελέγχου, καθώς εκτελείται αρένα με διαφορετικά ήδη γωνιών. Τα αριθμητικά αποτελέσματα αυτής της περίπτωσης, καταδεικνύουν τη δυνατότητα του αλγορίθμου να διαχειριστεί μη επαναλαμβανόμενες καταστάσεις, πράγμα σύνηθες σε κίνηση σε άγνωστο περιβάλλον.

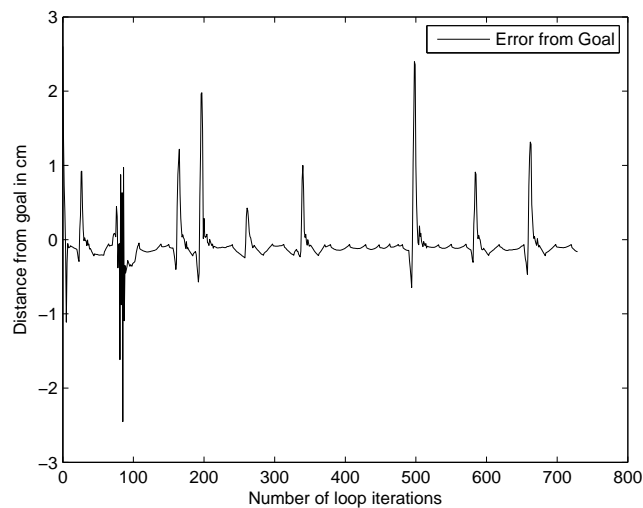
Ακολουθούν το μονοπάτι τροχιάς που ακολούθησε το ρομπότ, στο σχήμα 9.20, η εξέλιξη της Αντίληψης, της Γωνίας Αντίληψης και σφάλματος στα σχήματα 9.16, 9.17 και 9.18 αντίστοιχα, καθώς και ο αριθμός των περιπτώσεων που το ρομπότ κατάφερε να επιτύχει την απόσταση-στόχο στο σχήμα 9.19. Στο μονοπάτι τροχιάς υπάρχει αρίθμηση στη σειρά με την οποία εκτελέστηκαν οι στροφές για καλύτερη κατανόηση των γραφημάτων.



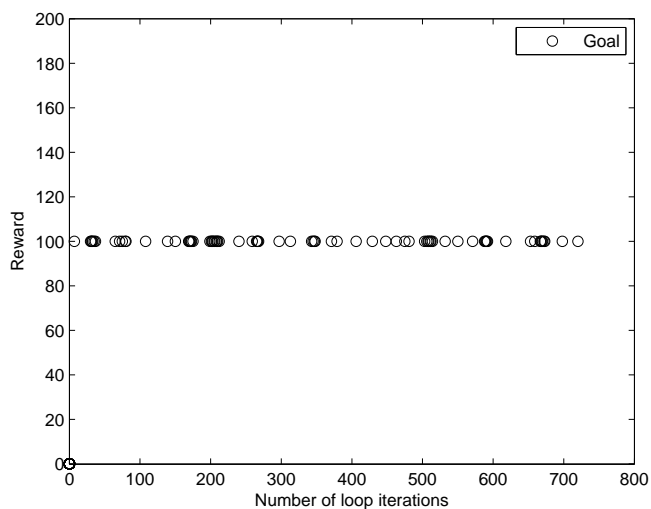
Σχήμα 9.16: Αντίληψη κατά την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης



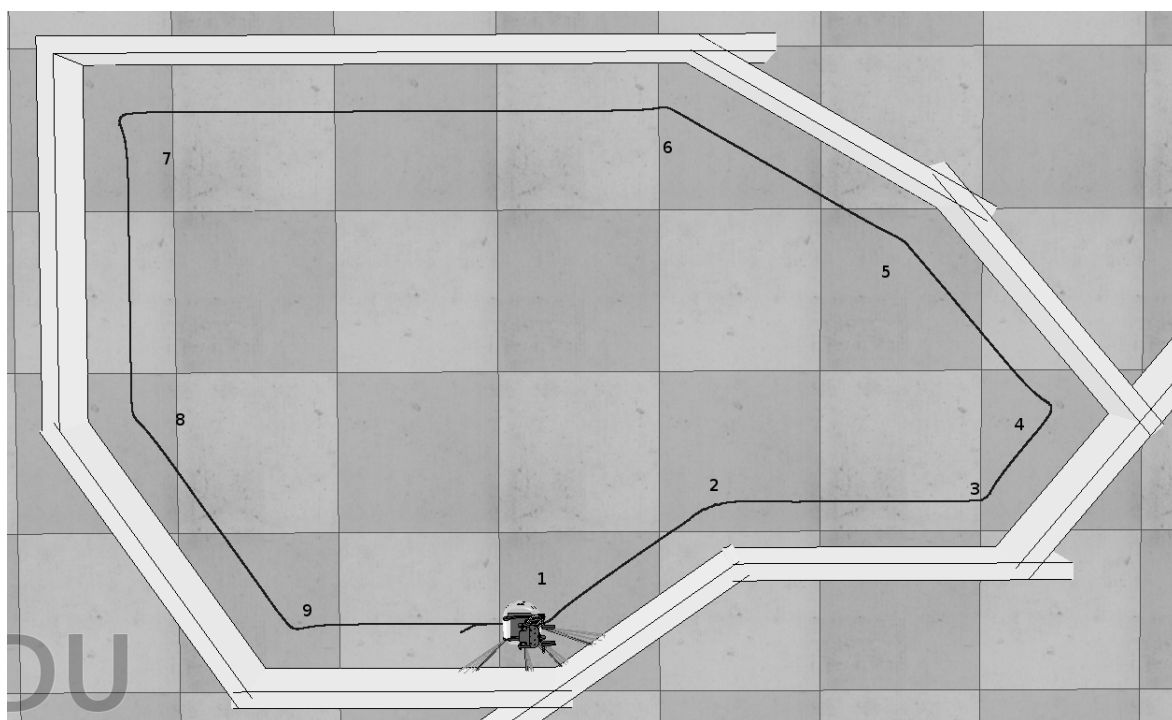
Σχήμα 9.17: Γωνία Αντίληψης κατά την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης



Σχήμα 9.18: Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης



Σχήμα 9.19: Φορές που το ρομπότ κατάφερε να βρεθεί στην επιθυμητή απόσταση και λαμβάνει ανταμοιβή κατά την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης



Σχήμα 9.20: Μονοπάτι κατά την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Λογική σε περιβάλλον εξομοίωσης

9.5. ΑΡΙΘΜΗΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΟΥ ΡΟΜΠΟΤ ΣΕ ΠΕΡΙΒΑΛΛΟΝ ΕΞΟΜΟΙΩΣΗΣ

Στα κομμάτια ευθείας κίνησης, το ρομπότ ακολουθεί με μικρές διορθώσεις όπου χρειάζεται. Στη γωνία 2, αργεί στην εκτέλεση μόνο και μόνο λόγω των θέσεων των αισθητήρων. Η γενική αντίληψη αλλάζει γρήγορα για τη γωνία 2, κατά την εκτέλεση και το ρομπότ προσπαθεί να ανταποκριθεί. Μετά από μερικά βήματα, συνεχίζει κανονικά.

Λόγω της μορφής της συνάρτησης συμμετοχής εξόδου, η απόσταση-στόχος επιτυγχάνεται κάποιες φορές, αλλά παραμένει το σφάλμα που παρατηρήθηκε και στις προηγούμενες αρένες. Η εφαρμογή της Ενισχυτικής Μάθησης στον Ασαφή Έλεγχο, στο επόμενο κεφάλαιο, έρχεται να βελτιώσει ακριβώς αυτό το σφάλμα, συμπληρώνοντας την ανθρώπινη εμπειρία που εισάγεται, μέσω της συνάρτησης συμμετοχής εξόδου.

Κεφάλαιο 10

Υλοποίηση Ασαφούς Λογικής με Ενισχυτική Μάθηση

Έχοντας ήδη περιγράψει την μορφή των ασαφών μεταβλητών εισόδου και τη βάση των ασαφών κανόνων, σε αυτό το κεφάλαιο υλοποιείται η Ενισχυτική Μάθηση και πιο συγκεκριμένα η Q-Μάθηση.

10.1 Βάση Κανόνων

Η βάση κανόνων που χρησιμοποιήθηκε για ασαφή έλεγχο αλλάζει, ώστε το συμπερασματικό κομμάτι της να εμπεριέχει διαφορετικές δράσεις προς επιλογή, από την Ενισχυτική Μάθηση.

Για την υλοποίηση χρησιμοποιήθηκαν 21 πιθανές δράσεις ανα κανόνα (σύνολο 20 κανόνων), οπότε για τον ελεγκτή, υπάρχουν συνολικά 420 δυνατές δράσεις. Στο σχήμα 10.1 φαίνεται η συνάρτηση συμμετοχής εξόδου για τις κεντρικές τιμές που επιλέχθηκαν, ενώ στον πίνακα 10.1, φαίνεται η τροποποιημένη βάση κανόνων για δεξιά ακολουθία τοίχου .

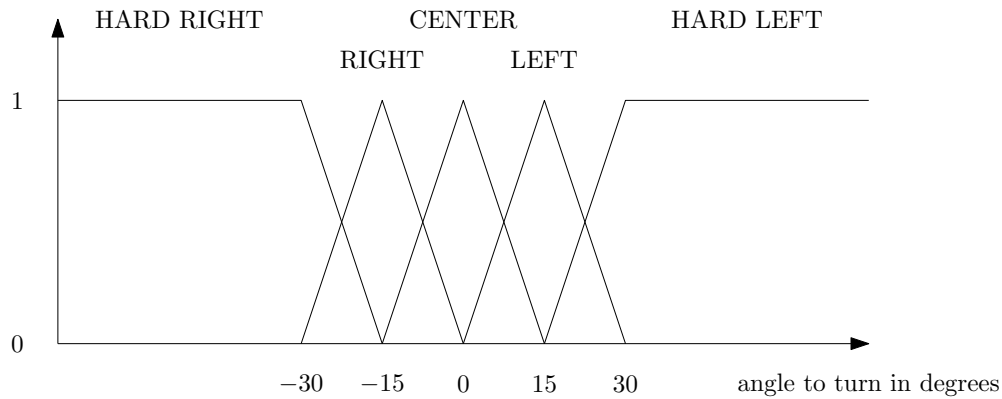
Η κάθε δράση, έχει ένα συνδεδεμένο δείκτη ποιότητας, ο οποίος θα ενημερώνεται κατά τη διάρκεια της φάσεως εξερεύνησης του αλγορίθμου. Οι δείκτες ποιότητας αρχικοποιούνται με μηδενικές τιμές.

Η μορφή των κανόνων αλλάζει και πλέον για κάθε κατάσταση του ρομπότ, υπάρχουν διαφορετικές πιθανές δράσεις με τις αντίστοιχες συνδεδεμένες ποιότητές τους.

Για παράδειγμα :

ΑΝ γωνία αντίληψης "ΑΡΙΣΤΕΡΑ ΜΠΡΟΣΤΑ" ΚΑΙ αντίληψη "ΥΨΗΛΗ" ΤΟΤΕ :

- "ΓΡΗΓΟΡΗ ΔΕΞΙΑ ΤΥΠΟΥ 1" με ποιότητα q_1
- "ΓΡΗΓΟΡΗ ΔΕΞΙΑ ΤΥΠΟΥ 2" με ποιότητα q_2
- "ΓΡΗΓΟΡΗ ΔΕΞΙΑ ΤΥΠΟΥ 3" με ποιότητα q_3



Σχήμα 10.1: Κεντρικές τιμές Συνάρτησης Συμμετοχής Εξόδου, γύρω από τις οποίες κυμαίνονται οι πιθανές δράσεις

	Πολύ Χαμηλή	Χαμηλή	Μέση	Υψηλή	Πολύ Υψηλή
Πίσω	HR_1^{21}	R_1^{21}	R_1^{21}	C_1^{21}	C_1^{21}
Δεξιά	HR_1^{21}	R_1^{21}	C_1^{21}	L_1^{21}	HL_1^{21}
Μπροστά	C_1^{21}	C_1^{21}	L_1^{21}	HL_1^{21}	HL_1^{21}
Αριστερά	HR_1^{21}	HR_1^{21}	R_1^{21}	R_1^{21}	R_1^{21}

Πίνακας 10.1: Τροποποιημένη Βάση κανόνων

Οι πιθανές δράσεις για τον κάθε κανόνα, δεν είναι τυχαίες, όπως για την περίπτωση της χωρίς γνώση Ενισχυτικής Μάθησης. Αντίθετα, οι πιθανές δράσεις, αποτελούν υποσύνολα της συνάρτησης συμμετοχής εξόδου του ασαφούς ελεγκτή.

Τέλος, ο ελεγκτής δεν χάνει χρόνο εξερευνώντας δράσεις οι οποίες γνωρίζουμε πως δεν οδηγούν στο στόχο, αλλά βελτιώνει τις πιθανές λύσεις του ασαφούς ελεγκτή.

10.2 Ανταμοιβές

Με βάση την έννοια της Γενικής Αντίληψης, θέλουμε το ρομπότ να κινείται στο μέσο της απόστασης d_{min} και d_{max} αξιόπιστων μετρήσεων των αισθητήρων.

Για την υλοποίηση στο ρομπότ Khepera 3, $d_{min} = 8cm$ και $d_{max} = 22cm$, οπότε το ρομπότ ανταμείβεται όταν το μέτρο της Γενικής Αντίληψης είναι ίσο με 0.5, στην περίπτωση μας τα 15cm. Για όλες τις άλλες περιπτώσεις, η ανταμοιβή είναι μηδενική.

Το συγκεκριμένο σχήμα συνάρτησης ανταμοιβής, επιλέχθηκε λόγω της μικρής ανάλυσης των αισθητήρων του Khepera 3. Αν δεν υπήρχε αυτός ο περιορισμός και ήταν δυνατή η κατάτμηση τόσο της αντίληψης καταστάσεων, όσο και του χώρου δράσεων σε μεγαλύτερα διαστήματα, θα είχε προτιμηθεί να δίνονται μικρότερες ανταμοιβές, για ενδιάμεσες καταστάσεις, που οδηγούν στο στόχο.

10.3 Δράσεις

Για την υλοποίηση, με δεδομένο πως προϋπήρχε το σύστημα ασαφούς ελέγχου, επιλέχθηκε οι δράσεις να εξερευνούν τιμές γύρω από τις ενδεικνυόμενες από τη συνάρτηση

συμμετοχής εξόδου του ασαφούς ελέγχου (Σχήμα 10.1). Αφού γίνει η επιλογή δράσεων, γίνεται η αποασαφοποίηση και η γωνία ελέγχου εφαρμόζεται στο ρομπότ.

Αν για παράδειγμα, η συνάρτηση συμμετοχής εξόδου, σε μια δεδομένη κατάσταση, δίνει γωνία στροφής 30° , η ενισχυτική μάθηση έχει να επιλέξει ανάμεσα σε 21 διαφορετικές γωνίες γύρω από την τιμή 30° . Οι 21 αυτές δράσεις επιλέχθηκαν, γύρω από τις τιμές εξόδου του ασαφούς ελεγκτή, σε διαστήματα 1° .

Το πλεονέκτημα αυτής της προσέγγισης είναι πως χρησιμοποιείται η ανθρώπινη εμπειρία, που συνοφίζεται στη συνάρτηση συμμετοχής εξόδου του ασαφούς ελεγκτή και βελτιώνεται εξερευνώντας δράσεις κοντά στην προτεινόμενη με τη χρήση της ενισχυτικής μάθησης. Ουσιαστικά εξαλείφονται τα προβλήματα των δυο μεθόδων, όταν χρησιμοποιούνται η κάθε μια ξεχωριστά.

Με το τέλος της εξερεύνησης το σύστημα λειτουργεί υπό Ασαφή Έλεγχο, με τη βελτιωμένη συνάρτηση συμμετοχής εξόδου που ανακαλύφθηκε.

10.4 Επιλογή Δράσης από τις Διαθέσιμες

Σε κάθε κατάσταση, επιλέγεται μια δράση, χρησιμοποιώντας έναν αλγόριθμο Εξερεύνησης / Εκμετάλλευσης. Ο αλγόριθμος επιλογής πρέπει να διατηρεί μια ισορροπία ανάμεσα στην εξερεύνηση και την εκμετάλλευση. Αν επιλεγούν δράσεις που δίνουν άμεσα ανταμοιβή, ίσως το σύστημα οδηγηθεί σε τοπικά ελάχιστα. Αν επιλεγούν δράσεις τυχαία, ο ρυθμός μάθησης του συστήματος θα είναι πολύ αργός και με δεδομένο πως η μάθηση γίνεται σε απευθείας σύνδεση και σε πραγματικό χρόνο, η εξαγόμενη γνώση θα αργεί να αποκτηθεί και μετά από κάποια βήματα εξομοίωσης, ο πράκτορας δεν θα τα λαμβάνει τόσο πολύ υπόψη, λόγω του συντελεστή μείωσης του ίχνους επιλεξιμότητας.

Για την υλοποίηση επιλέχθηκε ένας *e-greedy* αλγόριθμος επιλογής δράσεων, ο οποίος διατηρεί μια καλή ισορροπία ανάμεσα στην εξερεύνηση και στην εκμετάλλευση. Ο αλγόριθμος έχει ως εξής :

Algorithm 6 E-greedy Αλγόριθμος Επιλογής Δράσης

```

r τυχαίος αριθμός ( $0 < r < 1$ )
 $e_{t=0} = 0.99$ 
if  $r < e$  then
    επιλογή δράσης τυχαία
else
    επιλογή της δράσης με την υψηλότερη ανταμοιβή
end if
 $e_{t+1} = 0.9999e_t$ 

```

Ο αλγόριθμος ξεκινώντας, επιλέγει δράσεις τυχαία, χτίζοντας τη βάση δεδομένων του για την ποιότητα της κάθε δράσης. Με την πάροδο του χρόνου, εξερευνεί όλο και λιγότερο, χρησιμοποιώντας τη γνώση που έχει συγκεντρώσει, με τη βοήθεια των ανταμοιβών.

10.5 Αλγόριθμος Q-μάθησης

Ο αλγόριθμος, είναι αυτός που περιγράφηκε στο θεωρητικό κεφάλαιο της Ασαφούς Ενισχυτικής Μάθησης. Ξαναπαρουσιάζεται εδώ, προσαρμοσμένος στους χώρους καταστάσεων και στις δράσεις του προβλήματος ακολουθήθησης τοίχου. Για την υλοποίηση έχει ως εξής :

Algorithm 7 Αλγόριθμος Q-μάθησης για ακολουθήθηση τοίχου

1. $t = 0$, ίχνος επιλεξιμότητας $e[i, j] = 0$, Πίνακας ποιότητας $Q = 0$, παρατήρηση περιβάλλοντος x_t
 2. Για κάθε κανόνα i , υπολογισμός του βαθμού αληθείας των κανόνων $\alpha_i(x_t)$ για την κατάσταση x_t
 3. Για κάθε κανόνα i , επιλογή μιας από τις j διαθέσιμες δράσεις (στροφές) που συνδέονται με τον κανόνα i , $a[i, j]$, χρησιμοποιώντας e-greedy αλγόριθμο επιλογής.
 4. Υπολογισμός του πίνακα δράσεων $A(x_t)$ και της ποιότητας $Q(x_t, A(x_t))$ αυτών των δράσεων
 5. Υπολογισμός της επιλεξιμότητας των συμπερασμάτων, $e[i, j] = \alpha_i(x_t)$ αν j είναι η παρούσα δράση, $e[i, j] = \gamma \lambda e[i, j]$ διαφορετικά
 6. Πραγματοποίηση της δράσης (στροφής) $A(x_t)$ και παρατήρηση της κατάστασης x_{t+1}
 7. Έλεγχος της απόστασης από τον τοίχο. Λήψη της ανταμοιβής r_t
 8. Για κάθε κανόνα i , υπολογισμός του βαθμού αληθείας των κανόνων $\alpha_i(x_{t+1})$ για την κατάσταση x_{t+1}
 9. Υπολογισμός της βέλτιστης πολιτικής $V_t(x_{t+1})$ και της μεταβολής ποιότητας $\Delta Q(V_t(x_{t+1}), A(x_t))$ αυτών των δράσεων
 10. Ενημέρωση του σφάλματος χρονικής διαφοράς και των παραμέτρων ποιότητας $\Delta Q[i, j]$ και $q[i, j] \leftarrow q[i, j] + \Delta q[i, j]$
 11. $t = t + 1$ και εκτέλεση του βήματος 3
-

10.6 Αριθμητικά Αποτελέσματα του Ρομπότ σε Περιβάλλον Εξομοίωσης

Όπως και στην περίπτωση του Ασαφούς Ελέγχου, εξετάστηκαν διάφορες περιπτώσεις γωνιών, όπως και η ευθεία κίνηση για το ρομπότ Khepera 3 σε περιβάλλον εξομοίωσης VREP.

Ο έλεγχος γίνεται για τη γωνία στροφής του ρομπότ. Το ρομπότ εκτελεί ευθεία κίνηση για ορισμένη απόσταση (στην υλοποίηση 1cm), σταματά, ελέγχει την κατάσταση στην οποία βρίσκεται, μέσω του αλγορίθμου, επιλέγεται και εκτελείται η στροφή και συνεχίζεται η ευθεία κίνηση.

Οι παράμετροι μάθησης που χρησιμοποιήθηκαν, φαίνονται στον πίνακα 10.2.

γ	λ	β	ϵ	$e - greed$
0.99	0.5	0.99	0.99	0.99

Πίνακας 10.2: Παράμετροι μάθησης της υλοποίησης

Όπου γ ο συντελεστής μείωσης μελλοντικών ανταμοιβών, λ η εξασθένιση του ίχνους επιλεξιμότητας, β το κέρδος μάθησης συντελεστή ποιότητας της παρούσας κατάστασης, ϵ το κέρδος μάθησης για την ενημέρωση των q-τιμών, $e - greed$ η παράμετρος εξερεύνησης. Το γ επιλέχθηκε κοντά στη μονάδα, ώστε ο πράκτορας να λαμβάνει υπόψη μελλοντικές ανταμοιβές. Το λ είναι 0.5 ώστε οι προηγούμενες επισκέψεις σε μια κατάσταση να λαμβάνονται υπόψη, χωρίς όμως να θεωρούνται ακριβείς. Τα κέρδη μάθησης καθώς και η παράμετρος εξερεύνησης μειώνονται σταδιακά στο 0 αναλόγως με το πόσες εποχές μάθησης θέλουμε, πριν την ολοκλήρωση της φάσης εξερεύνησης. Αν N οι εποχές μάθησης, η μείωση σε κάθε εποχή είναι $\frac{1}{N}$.

Επειδή η κίνηση γίνεται σε άγνωστο περιβάλλον, η φάση εξερεύνησης πρέπει να είναι αρκετά μεγάλη. Ο ελεγκτής κατασκευάζει μια βάση εμπειρικών δεδομένων η οποία χρησιμοποιείται με βάση των αλγόριθμο επιλογής δράσεων. Ακόμα και αν το ρομπότ δεν συναντήσει μια κατάσταση, κατά τη φάση της εξερεύνησης, η αρχική συνάρτηση συμμετοχής εξόδου του Ασαφούς Ελέγχου, εγγυάται την αντιμετώπισή της, αν και η λύση δεν είναι βέλτιστη.

Η ασαφοποίηση των μεταβλητών εισόδου παραμένει ίδια με την περίπτωση του Ασαφούς Ελέγχου. Αυτό που αλλάζει είναι η συνάρτηση συμμετοχής εξόδου. Στην υλοποίηση, σαν αρχική συνάρτηση συμμετοχής εξόδου, χρησιμοποιήθηκε αυτή που φαίνεται στον πίνακα 10.3.

	Πολύ Χαμηλή	Χαμηλή	Μέση	Υψηλή	Πολύ Υψηλή
Πίσω	-30	-15	-15	0	0
Δεξιά	-30	-15	0	15	30
Μπροστά	0	0	15	30	30
Αριστερά	-30	-30	-15	-15	-15

Πίνακας 10.3: Αρχική συνάρτηση συμμετοχής εξόδου

10.6.1 Αποτελέσματα για Ευθεία Κίνηση

Αρχικά δοκιμάστηκε η συμπεριφορά της Ασαφούς Ενισχυτικής Μάθησης για ευθεία κίνηση. Δοκιμές έγιναν για εξερεύνηση 100, 250 και 300 εποχών. Το βήμα εξομοίωσης ήταν 50ms.

Στα σχήματα 10.2, 10.3, 10.4, 10.5 φαίνεται η απόσταση από τον τοίχο, η Γωνία Αντίληψης, το σφάλμα από την επιθυμητή απόσταση, καθώς και η λήψη ανταμοιβών με την πάροδο του χρόνου. Στους πίνακες 10.4, 10.5, 10.6 φαίνονται οι ανακαλυφθείσες συναρτήσεις συμμετοχής.

Παρατηρείται ότι, μετά την ολοκλήρωση της φάσης εξερεύνησης, το ρομπότ διατηρεί την επιθυμητή απόσταση, με ένα μικρό σφάλμα της τάξης του 1mm. Το σφάλμα αυτό οφείλεται τόσο στο βήμα εξομοίωσης που περιορίζει τη μικρότερη δυνατή γωνία που μπορεί να εκτελέσει το ρομπότ σε ένα βήμα, όσο και σε δυναμικά φαινόμενα οδομετρίας.

Όσον αφορά τις ανταμοιβές, παρατηρείται η αύξηση της συχνότητας λήψης τους, καθώς η μάθηση προχωρά και σταθεροποιείται μόλις η εξερεύνηση ολοκληρωθεί. Καθώς αυξάνεται ο αριθμός των εποχών (και άρα ο χρόνος μάθησης), τόσο καλύτερη και η συμπεριφορά του ρομπότ.

	Πολύ Χαμηλή	Χαμηλή	Μέση	Υψηλή	Πολύ Υψηλή
Πίσω	-40	-16	-19	-10	-10
Δεξιά	-40	-21	0	14	20
Μπροστά	-10	-10	14	38	20
Αριστερά	-40	-40	-25	-25	-25

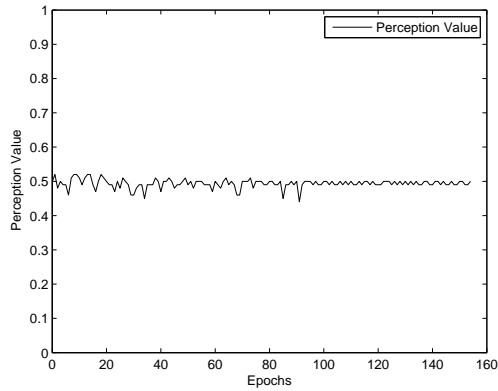
Πίνακας 10.4: Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου για ευθεία κίνηση 100 εποχών

	Πολύ Χαμηλή	Χαμηλή	Μέση	Υψηλή	Πολύ Υψηλή
Πίσω	-40	-15	-24	-10	-10
Δεξιά	-40	-23	-2	7	20
Μπροστά	-10	0	24	29	20
Αριστερά	-40	-40	-25	-25	-25

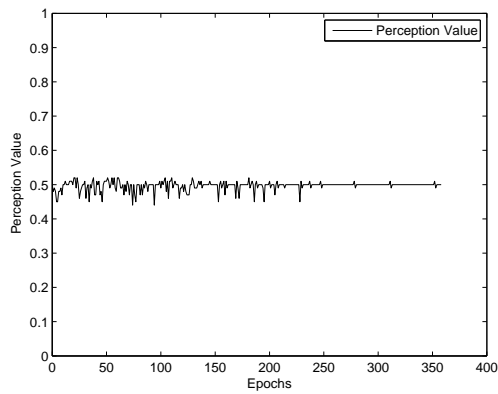
Πίνακας 10.5: Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου για ευθεία κίνηση 250 εποχών

	Πολύ Χαμηλή	Χαμηλή	Μέση	Υψηλή	Πολύ Υψηλή
Πίσω	-40	-21	-8	-10	-10
Δεξιά	-40	-24	-2	15	20
Μπροστά	-10	4	24	25	20
Αριστερά	-40	-40	-25	-25	-25

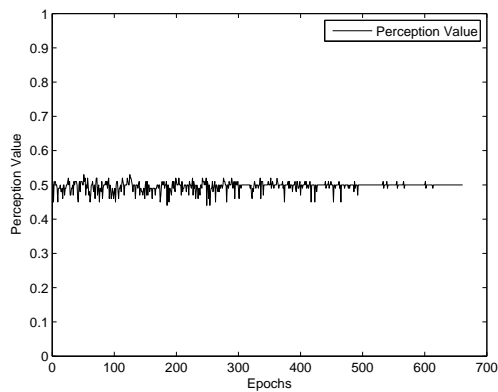
Πίνακας 10.6: Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου για ευθεία κίνηση 500 εποχών



(α') 100 εποχές

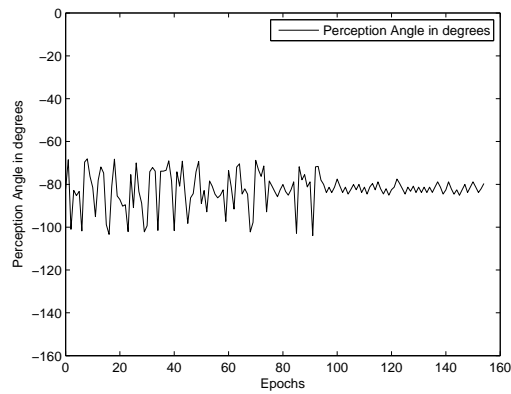


(β') 250 εποχές

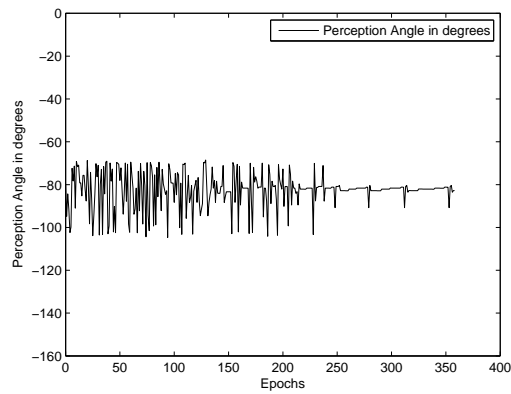


(γ') 500 εποχές

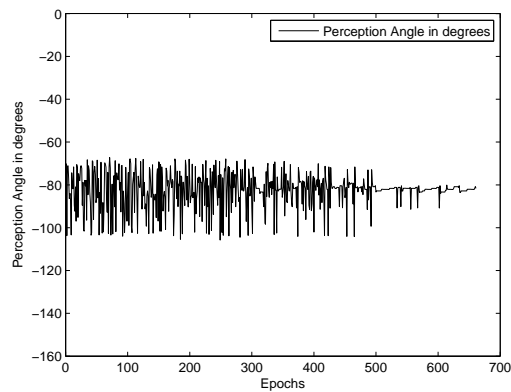
Σχήμα 10.2: Αντίληψη κατά την εκτέλεση ευθείας κίνησης 100, 250, 500 εποχών, υπό Ασαφή Ενισχυτική Μάθηση



(α') 100 εποχές

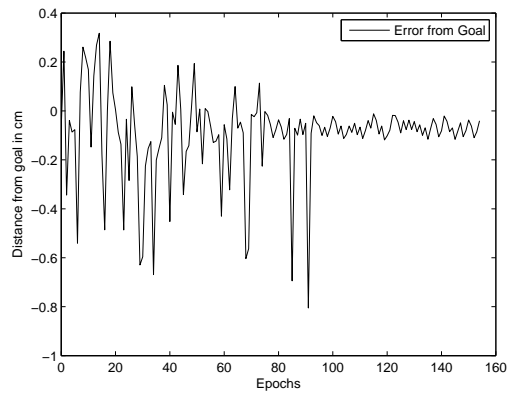


(β') 250 εποχές

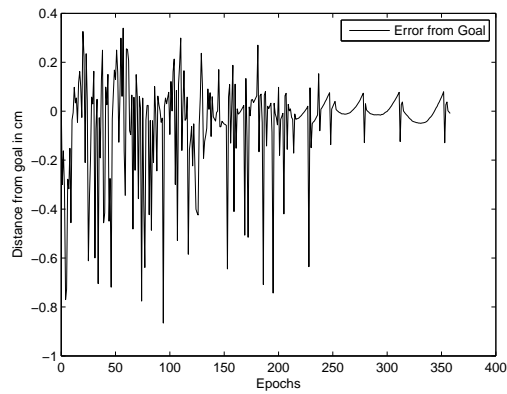


(γ') 500 εποχές

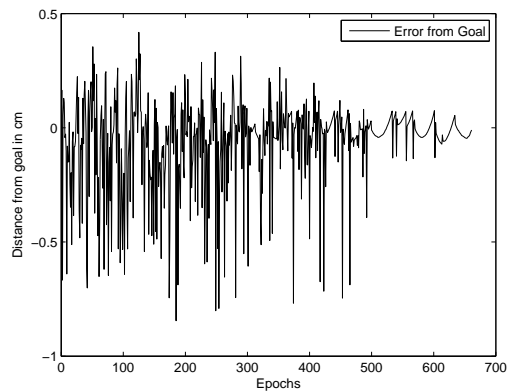
Σχήμα 10.3: Γωνία Αντίληψης κατά την εκτέλεση ευθείας κίνησης 100, 250, 500 εποχών, υπό Ασαφή Ενισχυτική Μάθηση



(α') 100 εποχές



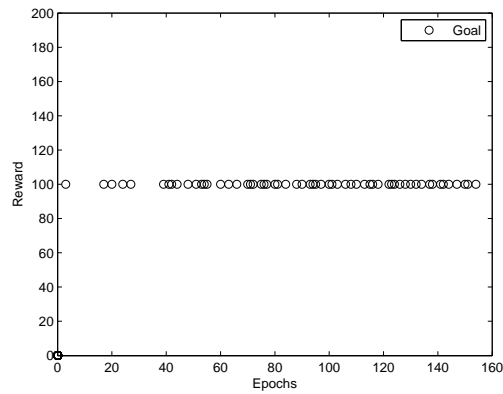
(β') 250 εποχές



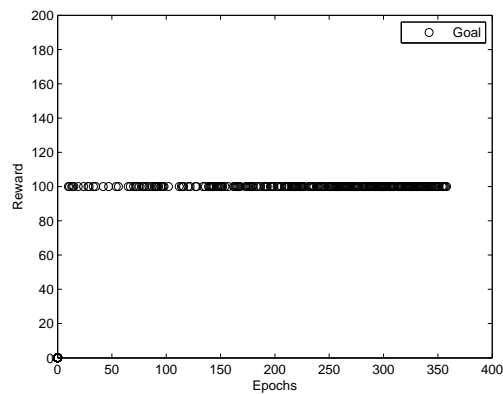
(γ') 500 εποχές

Σχήμα 10.4: Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση ευθείας κίνησης 100, 250, 500 εποχών, υπό Ασαφή Ενισχυτική Μάθηση

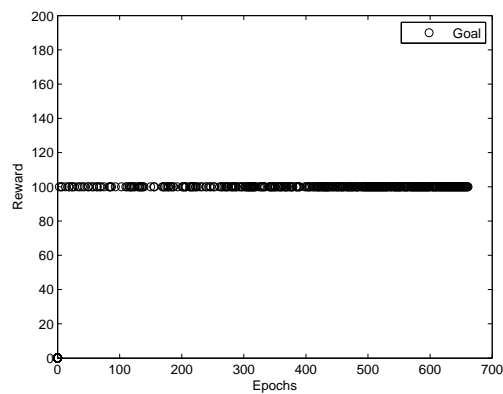
10.6. ΑΡΙΘΜΗΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΟΥ ΡΟΜΠΟΤ ΣΕ ΠΕΡΙΒΑΛΛΟΝ ΕΞΟΜΟΙΩΣΗΣ



(α') 100 εποχές



(β') 250 εποχές

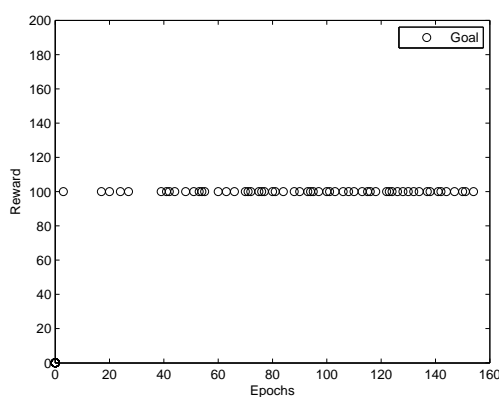


(γ') 500 εποχές

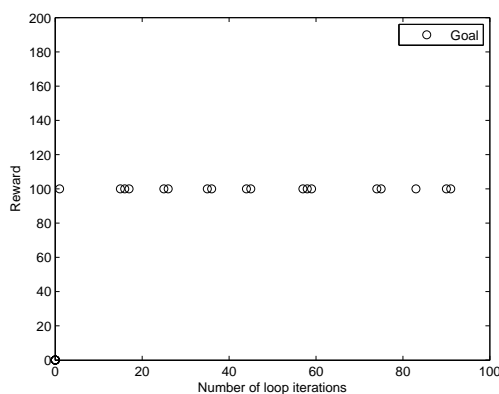
Σχήμα 10.5: Λήψη ανταμοιβών κατά την εκτέλεση ευθείας κίνησης 100, 250, 500 εποχών, υπό Ασαφή Ενισχυτική Μάθηση

Η διάρκεια κάθε εποχής, είναι ίση, με το χρόνο που απαιτείται για την εκτέλεση μιας επανάληψης του αλγορίθμου, ξεκινώντας από το διάβασμα των αισθητήρων, εφαρμόζοντας τον έλεγχο και εκτελώντας την επιλεγμένη δράση. Μετά την ολοκλήρωση της φάσης Μάθησης, ο έλεγχος που εφαρμόζεται, είναι Ασαφής, χρησιμοποιώντας την ανακαλυφθείσα συνάρτηση συμμετοχής.

Σε σύγκριση με τον απλό Ασαφή Έλεγχο, ήδη από την περίπτωση των 100 εποχών, φαίνεται σημαντική βελτίωση στον αριθμό των επισκέψεων στην απόσταση-στόχο. Στις 250 εποχές ο αριθμός των επισκέψεων πληθαίνει, ενώ υπάρχουν και μεγάλα κομμάτια ευθείας κίνησης πάνω στην απόσταση στόχο. Η αύξηση των εποχών στις 500, δεν επιφέρει ιδιαίτερες αλλαγές στη συμπεριφορά του ρομπότ, γεγονός που σημαίνει, ότι ο αριθμός των εποχών μάθησης ήταν μεγαλύτερος από τον απαιτούμενο για την εκμάθηση της ευθείας κίνησης. Ενδιαφέρον παρουσιάζει το γεγονός, πως η συμπεριφορά της Ασαφούς Ενισχυτικής Μάθησης είναι καλύτερη, πριν ακόμα αυτή ολοκληρωθεί. Μια σύγκριση του αριθμού επισκέψεων, στην απόσταση-στόχο, φαίνεται στο σχήμα 10.6.



(α') Ασαφής Ενισχυτική Μάθηση



(β') Ασαφής Έλεγχος

Σχήμα 10.6: Λήψη ανταμοιβών κατά την εκτέλεση ευθείας κίνησης 100 εποχών, υπό Ασαφή Ενισχυτική Μάθηση σε σύγκριση με ευθεία κίνηση υπό Ασαφή Έλεγχο

Οι ανακαλυφθείσες συναρτήσεις συμμετοχής, παρουσιάζουν ομοιότητες, για καταστάσεις που δεν συναντώνται κατά την ευθεία κίνηση σε δεξιά ακολουθία τοίχου. Αυτό είναι αναμενόμενο μια και οι καταστάσεις αυτές δεν εξερευνήθηκαν και η μόνη γνώση είναι αυτή που προέρχεται από τη συνάρτηση συμμετοχής εξόδου του απλού ασαφούς ελέγχου. Η ύπαρξή τους, αν και δεν είναι βέλτιστη, εγγυάται τη δράση από το ρομπότ με χρήση

απλού ασαφούς ελέγχου, εφ'όσον αυτές οι καταστάσεις απαντηθούν μετά την ολοκλήρωση της μάθησης.

10.6.2 Αποτελέσματα για Τετράγωνη Αρένα

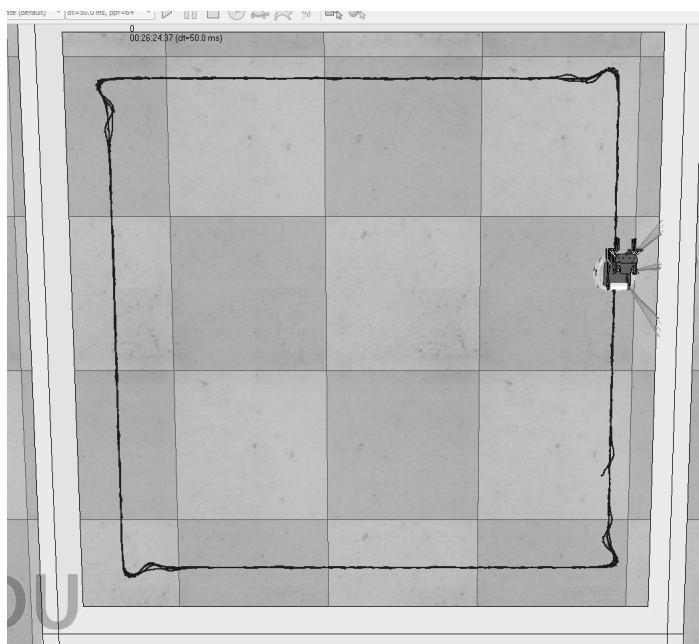
Εξετάζεται η περίπτωση κίνησης του Khepera 3 σε τετράγωνη αρένα. Η τετράγωνη αρένα δίνει τη δυνατότητα να παρατηρηθεί η συμπεριφορά του ρομπότ στη διόρθωση της απόστασης κατά την έξοδο από γωνία, ενώ μπορούν να παρατηρηθούν και οι διαφορετικές τροχιές που ακολούθησε το ρομπότ πριν και μετά την ολοκλήρωση της μάθησης.

Στο σχήμα 10.7 φαίνεται το μονοπάτι που ακολούθησε το ρομπότ, για την περίπτωση τετράγωνης αρένας. Φαίνονται οι διαφορετικές τροχιές που ακολούθησε το ρομπότ εξερευνώντας διαφορετικές λύσεις σε διαφορετικές εκτελέσεις της αρένας.

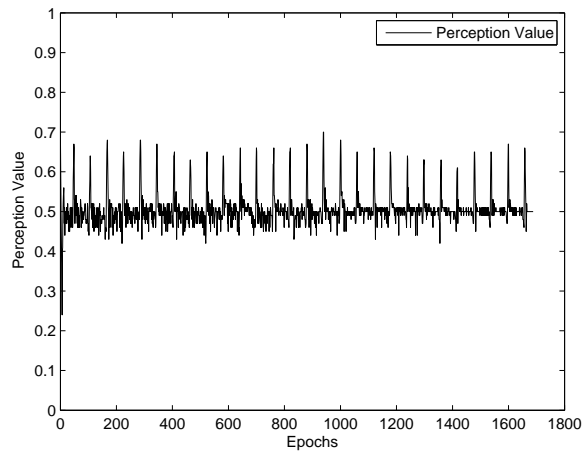
Στο σχήμα 10.10, φαίνεται το σφάλμα από την επιθυμητή απόσταση, ενώ στα σχήματα 10.8, 10.9 φαίνονται αντίστοιχα η εξέλιξη της απόστασης του ρομπότ από τον τοίχο, με την πάροδο του χρόνου, καθώς και η γωνία αντίληψης. Στο σχήμα 10.11, φαίνεται η πρόοδος στη λήψη ανταμοιβών. Τέλος, στον πίνακα 10.7 φαίνεται η ανακαλυφθείσα συνάρτηση συμμετοχής.

	Πολύ Χαμηλή	Χαμηλή	Μέση	Υψηλή	Πολύ Υψηλή
Πίσω	-20	-25	-24	7	4
Δεξιά	-40	-19	-2	24	37
Μπροστά	-10	-6	18	20	20
Αριστερά	-40	-40	-25	-25	-25

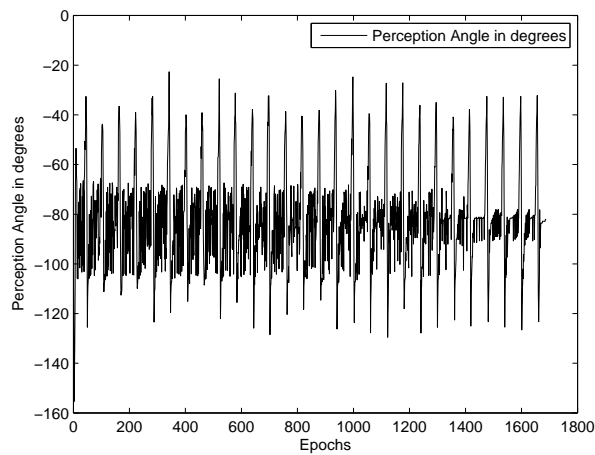
Πίνακας 10.7: Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου, τετράγωνη αρένα, 1500 εποχές



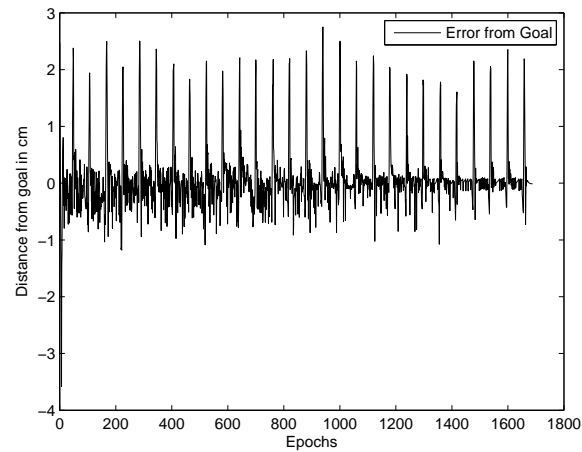
Σχήμα 10.7: Μονοπάτι για τετράγωνη αρένα, υπό Ασαφή Ενισχυτική Μάθηση



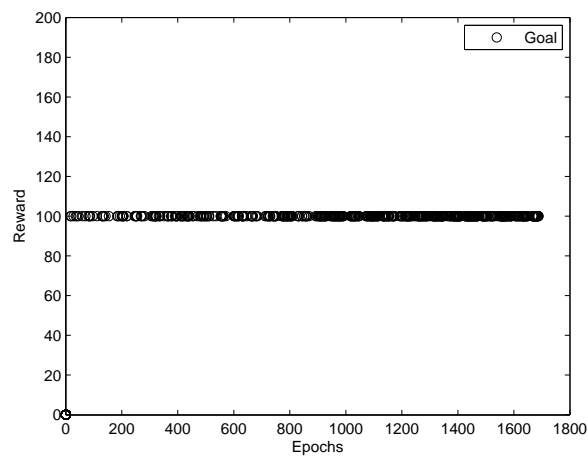
Σχήμα 10.8: Αντίληψη κατά την εκτέλεση τετράγωνης αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1500 εποχές



Σχήμα 10.9: Γωνία αντίληψης για τον τοίχο κατά την εκτέλεση τετράγωνης αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1500 εποχές



Σχήμα 10.10: Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση τετράγωνης αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1500 εποχές

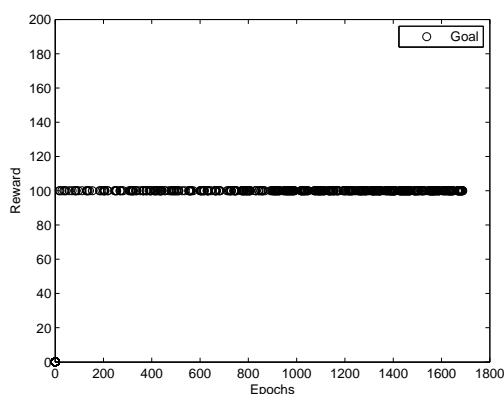


Σχήμα 10.11: Λήψη ανταμοιβών κατά την εκτέλεση τετράγωνης αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1500 εποχές

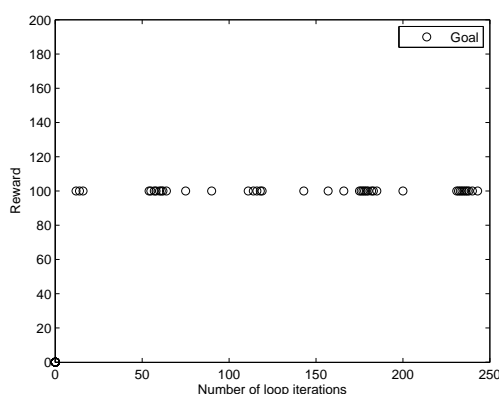
Το ρομπότ για την εκτέλεση τετράγωνης αρένας, επιστρέφει στην επιθυμητή απόσταση, μετά την εκτέλεση της γωνίας, με το μικρό σφάλμα $1mm$, που μπορεί να οφείλεται, στο βήμα εξομοίωσης (οι αισθητήρες του πραγματικού ρομπότ, ανανεώνουν τα δεδομένα τους σε μικρότερο χρόνο $20ms$ από το ελάχιστο επιτρεπόμενο βήμα εξομοίωσης $50ms$), σε σφάλμα οδομετρίας (καθώς οι τροχοί αλλάζουν ταχύτητα στην εκτέλεση στροφών), ή ακόμα και λόγω θορύβου στο μοντέλο εξομοίωσης των αισθητήρων. Στο μονοπάτι που ακολουθήθηκε φαίνονται οι διαφορετικές τροχιές κατά την έξοδο από τη γωνία και παρατηρείται μείωση στην υπέρβαση της επιθυμητής απόστασης κατά την έξοδο.

Το σφάλμα απόστασης κατά την εκτέλεση της γωνίας (οι αιχμές που παρατηρούνται) θα μπορούσε να μειωθεί, χρησιμοποιώντας μεγαλύτερες γωνίες στην αρχική συνάρτηση συμμετοχής εξόδου. Κάτι τέτοιο όμως, κατά τη φάση της αποασαφοποίησης θα προκαλούσε μεγαλύτερες γωνίες στην έξοδο για ευθεία κίνηση, από ενδείξεις των αισθητήρων, μπροστά η πίσω και το σφάλμα ευθείας κίνησης θα ήταν μεγαλύτερο. Αν επιλεγόταν μεγαλύτερο πλήθος δυνατών δράσεων, οι χρόνοι μάθησης θα γίνονταν πολύ μεγαλύτεροι.

Σε σύγκριση με τον Ασαφή Έλεγχο, στο σχήμα 10.17, φαίνεται πως οι επισκέψεις στην επιθυμητή απόσταση-στόχο, είναι πολύ περισσότερες. Όπως και στην ευθεία κίνηση, οι επιτεύξεις στόχου, από τα πρώτα στάδια της μάθησης είναι πιο συχνές και το πλήθος τους αυξάνεται, με τις εποχές εκπαίδευσης.



(α') Ασαφής Ενισχυτική Μάθηση



(β') Ασαφής Έλεγχος

Σχήμα 10.12: Λήψη ανταμοιβών κατά την εκτέλεση τετράγωνης αρένας, υπό Ασαφή Ενισχυτική Μάθηση σε σύγκριση με εκτέλεση τετράγωνης αρένας υπό Ασαφή Έλεγχο

10.6.3 Αποτελέσματα σε Αρένα με Διαφορετικούς Τύπους Γωνιών

Κλείνοντας, παρουσιάζεται η συμπεριφορά της Ασαφούς Ενισχυτικής Μάθησης, καθώς εκτελεί την αρένα με διαφορετικά ήδη γωνιών, που χρησιμοποιήθηκε και στην περίπτωση του Ασαφούς Ελέγχου.

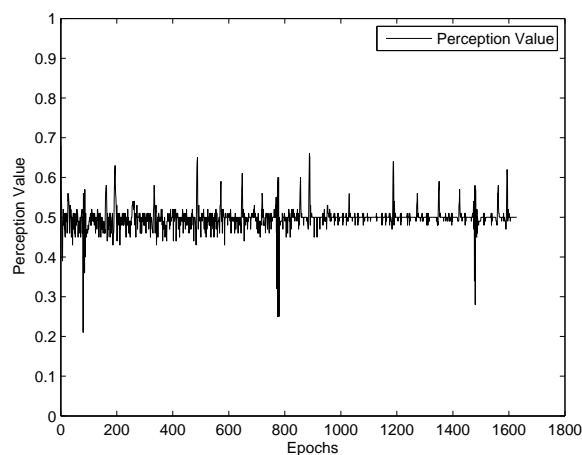
Μετά την ολοκλήρωση της μάθησης, η απόσταση από τον τοίχο μετά την έξοδο από γωνία, είναι η επιθυμητή με το μικρό σφάλμα που υπήρχε και στις προηγούμενες περιπτώσεις. Η διάρκεια μάθησης ήταν 1000 εποχές. Το ρομπότ αντιμετώπισε όλους τους τύπους γωνιών ικανοποιητικά και στο μονοπάτι εκτέλεσης φαίνονται οι διαφορετικές τροχιές.

Η ανακαλυφθείσα συνάρτηση συμμετοχής, λόγω της ποικιλίας διαφορετικών γωνιών, παρουσιάζει αρκετές διαφορές, στις αριθμητικές τιμές εξόδου, σε σχέση με την ευθεία κίνηση και την κίνηση σε τετράγωνη αρένα. Αυτό οφείλεται στο ότι, το ρομπότ συνάντησε καταστάσεις που δεν υπήρχαν στα προηγούμενα παραδείγματα, όπως αμβλείες γωνίες. Οι καινούργιες αυτές καταστάσεις, εξερευνήθηκαν στη φάση μάθησης και συνεπώς μεταβλήθηκε η αρχική συνάρτηση συμμετοχής εξόδου. Η επίτευξη της επιθυμητής απόστασης, όπως και στις προηγούμενες περιπτώσεις, είναι κατα πολύ συχνότερη, σε σχέση με τον Ασαφή Έλεγχο.

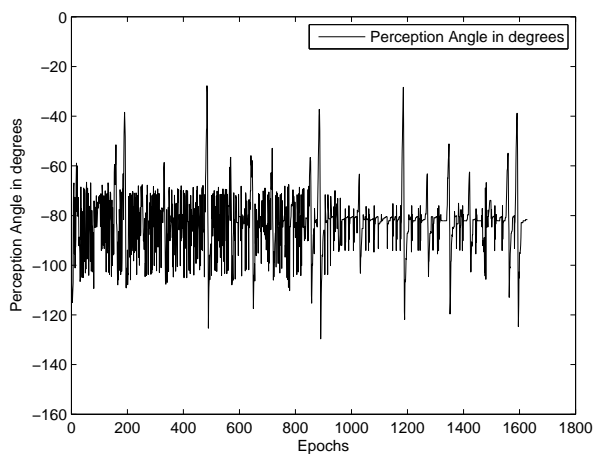
Ακολουθεί η ανακαλυφθείσα συνάρτηση συμμετοχής, το μονοπάτι που ακολουθήθηκε, καθώς και τα γραφήματα απόστασης, γωνίας αντίληψης, σφάλματος και λήψης ανταμοιβών για την περίπτωση αρένας διαφορετικών τύπων γωνιών.

	Πολύ Χαμηλή	Χαμηλή	Μέση	Υψηλή	Πολύ Υψηλή
Πίσω	-23	-15	-15	5	0
Δεξιά	-37	-25	-1	21	40
Μπροστά	-10	1	14	33	29
Αριστερά	-40	-40	-25	-25	-25

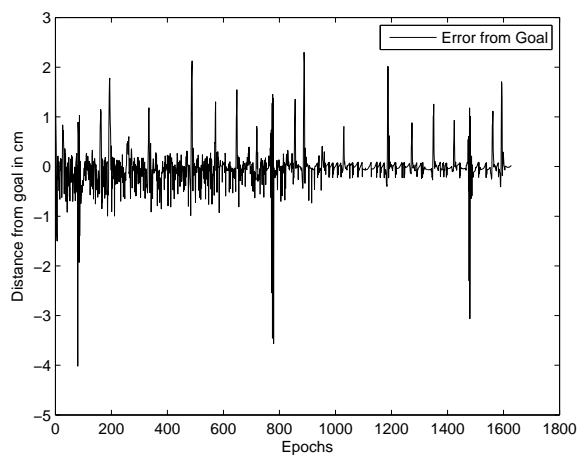
Πίνακας 10.8: Ανακαλυφθείσα συνάρτηση συμμετοχής εξόδου, αρένας διαφορετικών γωνιών, 1000 εποχές



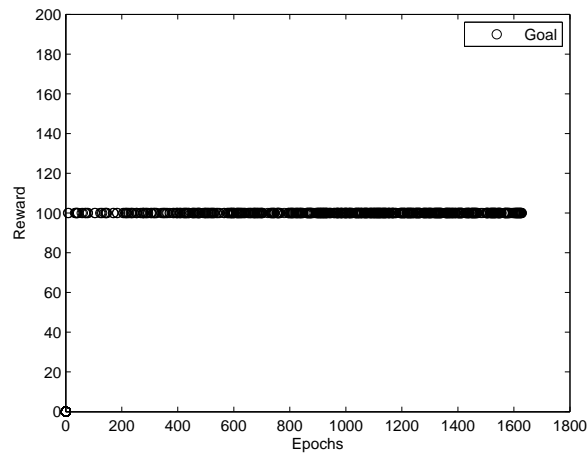
Σχήμα 10.13: Απόσταση από τον τοίχο κατά την εκτέλεση της αρένας υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές



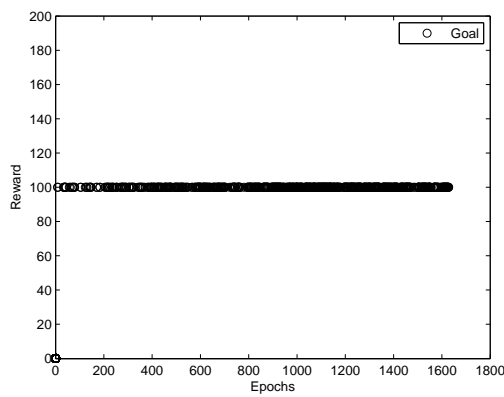
Σχήμα 10.14: Γωνία Αντίληψης κατά την εκτέλεση της αρένας διαφορετικών γωνιών υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές



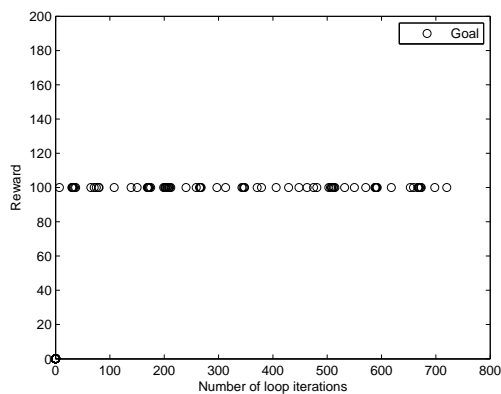
Σχήμα 10.15: Σφάλμα από την επιθυμητή απόσταση κατά την εκτέλεση της αρένας διαφορετικών γωνιών υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές



Σχήμα 10.16: Λήψη ανταμοιβών κατά την εκτέλεση της αρένας διαφορετικών γωνιών υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές



(α') Ασαφής Ενισχυτική Μάθηση



(β') Ασαφής Έλεγχος

Σχήμα 10.17: Λήψη ανταμοιβών κατά την εκτέλεση της αρένας διαφορετικών γωνιών, υπό Ασαφή Ενισχυτική Μάθηση σε σύγκριση τον Ασαφή Έλεγχο



Σχήμα 10.18: Μονοπάτι που ακολούθησε το ρομπότ στον εξομοιωτή κατά την εκτέλεση της αρένας διαφορειακών γωνιών υπό Ασαφή Ενισχυτική Μάθηση, 1000 εποχές

Κεφάλαιο 11

Επίλογος

11.1 Συμπεράσματα

Στη διπλωματική αυτή εργασία υλοποιήθηκε έλεγχος ρομπότ διαφορικής οδήγησης, κινούμενου σε άγνωστο περιβάλλον, για την περίπτωση ακολούθησης τοίχου. Χρησιμοποιήθηκε η μέθοδος της Γενικής Αντίληψης για τη διαχείριση δεδομένων αισθητήρων, υλοποιήθηκε Ασαφής Ελεγκτής για το ρομπότ διαφορικής οδήγησης Khepera 3, σε περιβάλλον εξομοίωσης καθώς και Ασαφής ελεγκτής για πραγματικό ρομπότ Khepera 2. Τέλος υλοποιήθηκε αλγόριθμος Ενισχυτικής Μάθησης, τύπου Q-μάθησης για τη βελτίωση του Ασαφούς Ελεγκτή. Ο τελευταίος δοκιμάστηκε σε περιβάλλον εξομοίωσης για το ρομπότ Khepera 3.

Τα Ασαφή Συμπερασματικά Συστήματα εξακολουθούν να αποτελούν μια εξαιρετική μέθοδο επίλυσης προβλημάτων. Χωρίς μεγάλες υπολογιστικές απαιτήσεις, με αναπαράσταση μεταβλητών που προσομοιάζει στην ανθρώπινη και με τη δυνατότητα ενσωμάτωσης προηγούμενης εμπειρίας στο σύστημα, έχουν καταφέρει να παραμείνουν στο προσκήνιο της επιστημονικής έρευνας, σχεδόν 50 χρόνια μετά την πρώτη περιγραφή της Ασαφούς Λογικής.

Η Ενισχυτική Μάθηση αποτελεί μια μέθοδο Μηχανικής Μάθησης, που δίνει τη δυνατότητα εξερεύνησης λύσεων και εύρεσης βέλτιστων πολιτικών, σε πραγματικό χρόνο και χωρίς επίβλεψη. Η έννοια των ανταμοιβών, μέσω των οποίων δίδεται η ενίσχυση, παρέχει εξαιρετική συνέργεια με τα Ασαφή Συμπερασματικά Συστήματα.

Με τη χρήση Ασαφούς Ενισχυτικής Μάθησης, το ρομπότ εξερευνά το περιβάλλον και συσσωρεύει εμπειρία. Παλαιότερη εμπειρία χρησιμοποιείται, με μειωμένη όμως εμπιστοσύνη, μέσω του συντελεστή μείωσης. Μετά το τέλος της φάσης εξερεύνησης, χρησιμοποιείται η συνάρτηση συμμετοχής εξόδου που ανακαλύφθηκε. Ακόμα και αν το ρομπότ δεν συνάντησε κάποια κατάσταση στη φάση εξερεύνησης, η αρχική συνάρτηση συμμετοχής εγγυάται αντιμετώπιση αυτής της κατάστασης, απλά με μη βέλτιστο τρόπο.

Στα αριθμητικά αποτελέσματα, φαίνεται πως ο αλγόριθμος Ενισχυτικής Μάθησης βελτιώνει τον Ασαφή Έλεγχο, χρησιμοποιώντας την συσσωρευμένη εμπειρία και φτάνει στην απόσταση στόχο με πολύ μικρό σφάλμα. Αν αναλογιστεί κανείς το γεγονός, ότι η εφαρμογή απλού Ασαφούς Ελέγχου περιέχει ένα σημαντικό χρονικό κόστος, όπου η ανθρώπινη εμπειρία μεταφράζεται σε συναρτήσεις συμμετοχής εξόδου και αυτές βελτιώνονται μέσα από διαδικασίες δοκιμής και λάθους, το χρονικό κόστος εφαρμογής της φάσης Μάθησης, της Ασαφούς Ενισχυτικής Μάθησης, για κίνηση στο ίδιο περιβάλλον, είναι μικρό.

11.2 Μελλοντικές Επεκτάσεις

Η υλοποίηση του ελέγχου δεν είναι αποκλειστική για ρομπότ διαφορικής οδήγησης. Μπορεί να χρησιμοποιηθεί σε οποιοδήποτε ρομπότ, με κατάλληλο ορισμό των Ασαφών Μεταβλητών και φυσικά για το αντίστοιχο κινηματικό μοντέλο. Το γεγονός πως ο έλεγχος είναι ουσιαστικά πολλαπλής πλατφόρμας, δίνει μια πληθώρα εφαρμογών σε πιο πολύπλοκα ρομπότ. Επιπλέον η συνάρτηση συμμετοχής εξόδου που δίνεται αρχικά, αν είναι καλά ορισμένη, προστατεύει το ρομπότ από εξερεύνηση καταστάσεων που μπορεί να είναι καταστρεπτικές (σύγκρουση με τον τοίχο).

Η μέθοδος της Ασαφούς Ενισχυτικής Μάθησης γενικότερα, προσφέρεται για την επίλυση προβλημάτων απόκτησης ρομποτικών δεξιοτήτων, που αφορούν αλληλεπίδραση με τον άνθρωπο, για τη βελτιστοποίηση της εκτέλεσης συνεργατικών και υποβοηθητικών λειτουργιών, που απαιτούν συνδυασμένο έλεγχο κινήσεων και χειρισμών. Ο Ασαφής Έλεγχος, προσφέρει τα γλωσσικά εργαλεία για ενσωμάτωση της ανθρώπινης εμπειρίας, ενώ η Ενισχυτική Μάθηση έρχεται να τη βελτιώσει.

Βιβλιογραφία

- [1] Σπυρίδων Γ. Τζαφέστας. *Εισαγωγή στην Τεχνητή Νοημοσύνη και τα Έμπειρα Συστήματα Τεύχος Α*. Σπυρίδων Γ. Τζαφέστας, Αθήνα, 2nd edition, 1996.
- [2] Σπυρίδων Γ. Τζαφέστας. *Ρομποτική Τόμος I Ανάλυση και Έλεγχος*. Σπυρίδων Γ. Τζαφέστας, Αθήνα, 1998.
- [3] Κωνσταντίνος Σ. Τζαφέστας. *Ρομποτική II: Ευφυή και Επιδέξια Ρομποτικά Συστήματα*, Συμπληρωματικές Σημειώσεις Μαθήματος «Ρομποτική II», 2006.
- [4] Ιωάννης Ν. Καρύγιαννης. Συμβολή Στην Ανάπτυξη Πολυπρακτορικής Αρχιτεκτονικής Αναπτυξιακού Ρομποτικού Ελέγχου Στη Βάση Ασαφούς Ενισχυτικής Μάθησης: Εφαρμογή Στον Επιδέξιο Ρομποτικό Χειρισμό. <http://artemis-new.cslab.ece.ntua.gr:8080/jspui/handle/123456789/6564>, 2013.
- [5] Θεόδωρος Ρεκατσίνας. Ανάπτυξη μιας Προσαρμοστικής Πολιτικής Αντικατάστασης Αρχείων με χρήση Ενισχυτικής Μάθησης. <http://artemis-new.cslab.ece.ntua.gr:8080/jspui/handle/123456789/5250>, 2010.
- [6] Σωτήριος Αποστολόπουλος. Ανάπτυξη Κατανεμημένου Πολυπρακτορικού Συστήματος Από Κοινού Ενισχυτικής Μάθησης Για Την Αυτόνομη Ρύθμιση Του Βηματισμού Σε Τετράποδο Ρομπότ. <http://artemis-new.cslab.ece.ntua.gr:8080/jspui/handle/123456789/5788>, 2012.
- [7] Ethem Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2nd edition, 2010.
- [8] Reinhard Brauningl, Pedro Sanz, and Jose Manuel Ezkerra. Fuzzy logic wall following of a mobile robot based on the concept of general perception. In *Proceedings of the 7th International Conference on Advanced Robotics*, pages 367–376, Sant Feliu De Guixols (Spain), 1995.
- [9] Gregory Dudek and Michael Jenkin. *Computational Principles of Mobile Robotics*. Cambridge University Press, New York, NY, USA, 2000.
- [10] F.Lamercy and J.Tharin for K-Team S.A. *Khepera III User Manual*. K-Team S.A., 2013.
- [11] Pierre Bureau for K-Team S.A. *Khepera 2 IR Sensors Report*. K-Team S.A., 2002.
- [12] Pierre Bureau for K-Team S.A. *Khepera 2 Programming Manual*. K-Team S.A., 2002.
- [13] Marc Andreas Freese. *V-REP User Manual*. Coppelia Robotics, 2013.
- [14] P.-Y. Glorennec and L. Jouffe. Fuzzy q-learning. In *Fuzzy Systems, 1997., Proceedings of the Sixth IEEE International Conference on*, volume 2, pages 659–662 vol.2, Jul 1997.
- [15] Pierre Yves Glorennec. Reinforcement learning: an overview. In *European Sym. on Intelligent Techniques*, 2000.

-
- [16] Vishay Semiconductor GmbH. *Application of Optical Reflex Sensors TCRT1000, TCRT5000, CNY70 Data Sheet*. Vishay Semiconductor GmbH, 1999.
- [17] Vishay Semiconductor GmbH. *TCRT1000/TCRT1010 Reflective Optical Sensor with Transistor Output Data Sheet*. Vishay Semiconductor GmbH, 1999.
- [18] Siegfried Gottwald. An early approach toward graded identity and graded membership in set theory. *Fuzzy Sets and Systems*, 161(18):2369 – 2379, 2010.
- [19] J.N. Karigiannis, T.I. Rekatsinas, and C.S. Tzafestas. Fuzzy rule based neuro-dynamic programming for mobile robot skill acquisition on the basis of a nested multi-agent architecture. In *Robotics and Biomimetics (ROBIO), 2010 IEEE International Conference on*, pages 312–319, Dec 2010.
- [20] E.H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1):1 – 13, 1975.
- [21] M. Mucientes, R. Alcalá, J. Alcalá-Fdez, and J. Casillas. Learning weighted linguistic rules for mobile robotics. In *Proceedings of the 11th IFSA World Congress*, pages 1279–1285, Beijing (China), 2005.
- [22] M. Mucientes, D. L. Moreno, A. Bugarín, and S. Barro. Evolutionary learning of a fuzzy controller for wall-following behavior in mobile robotics. *Soft Computing*, 10(10):881–889, 2006.
- [23] Alireza Ferdowsizadeh Naeeni. Advanced multi-agent fuzzy reinforcement learning. Master’s thesis, Dalarna University, Sweden, 2004.
- [24] S. Nefti, M. Oussalah, K. Djouani, and J. Pontnau. Intelligent adaptive mobile robot navigation. *Journal of Intelligent and Robotic Systems*, 30(4):311–329, 2001.
- [25] Iwan Petrowitsch Pawlow, Walter B. Cannon, and W. Horsley Gantt. *Lectures on conditioned reflexes. Twenty-five years of objective study of the higher nervous activity "behaviour" of animals*. New York, 1928.
- [26] K-Team S.A. *Khepera User Manual*. K-Team S.A., 1999.
- [27] K-Team S.A. *Khepera 2 User Manual*. K-Team S.A., 2002.
- [28] Roland Siegwart and Illah R. Nourbakhsh. *Introduction to Autonomous Mobile Robots*. Bradford Company, Scituate, MA, USA, 2004.
- [29] Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.
- [30] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-15(1):116–132, Jan 1985.
- [31] Christopher John Cornish Hellaby Watkins. *Learning from Delayed Rewards*. PhD thesis, King’s College, Cambridge, UK, May 1989.
- [32] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.