

طراحی یک کنترل کننده فازی تطبیقی برای ناوبری ربات های متحرک با استفاده از یادگیری تقویتی

دانشگاه امیرکبیر

دانشکده مهندسی کامپیوتر

محمد رضا میدی

شراره پابویه

(۱) مقدمه

چکیده

ناوبری برای ربات های متحرک عبارت است از حرکت به طرف یک هدف مشخص درحالیکه ربات از بخورد با موانع پرهیز می نماید. در یک محیط دینامیک که دائمًا در حال تغییر است، استفاده از روش های سراسری غیر ممکن یا بسیار پرهیز است، چرا که در این روش ها مدل ریاضی و یا نقشه کامل محیط مورد نیاز می باشد. اما تهیه مدل ریاضی و یا نقشه کامل محیط کاری پیچیده است و با تغییر محیط ناگزیر از به روز در آوردن مدل ریاضی و یا نقشه آن خواهیم شد. باین ترتیب روش های سراسری برای حل چنین مسائلی غیر عملی می باشند. پس به ناچار باید از روش های محلی استفاده نمود.

در این روش های محلی ربات در محیط گردش می کند و اطلاعات موردنیاز خود را با یک حسگرها یش به صورت محلی جمع آوری می نماید. ربات میتواند براساس این اطلاعات درک شده، در حالت های مختلف محیط تصمیم های مناسب اتخاذ نموده، عمل درستی را انجام دهد. به عبارت دیگر حسگرها ربات از طریق یک پردازنده ساده مانند شبکه های عصبی یا سیستم های فازی وغیره، مستقیماً به عملکردهای ربات متصل می شوند و اطلاعات خام در هر لحظه از زمان به ایجاد عملیات موتوری منجر می گردند.

سیستم های استنتاج فازی، به خاطر معماری ساده و هزینه محاسباتی پایین، در زمینه رباتیک توجه زیادی را به خود جلب نموده اند. یک سیستم استنتاج فازی به تنهایی برای یک محیط ساده که قوانین حاکم بر آن رابه آسانی می توان استخراج نمود کافی است [۱]. ولی برای محیط های پیچیده تر استخراج و تنظیم قوانین فازی لازم می باشد. برخی از محققان برای بدست آوردن قوانین فازی از یک شبکه عصبی که وزن های آن نموده اند [۲]، [۳]، [۴].

در این مقاله مراحل طراحی یک کنترل کننده فازی تطبیقی، جهت رفتار ناوبری متوسط دو ربات متحرک در یک محیط ناشناخته شرح داده شده است. هر ربات قادر است با یک حسگرها خود را با محیط را درک کند و با یک کنترل کننده خود تصمیم مناسبی اتخاذ و در محیط اجرانماید. ورودی کنترل کننده مقادیر حسگرها ربات و خروجی آن سرعت چرخ های چپ و راست ربات هستند. جهت تعیین پارامترهای توابع عضویت فازی ورودی و خروجی، با توجه به شرایط محیط در هر قدم زمانی از اتوماتون یادگیر استفاده می شایم. ما جهت یافتن قوانین فازی برای ناوبری هر ربات از الگوریتم یادگیری Q استفاده می کنیم، اما برای همکاری ربات ها از الگوریتم یادگیری Q باتابع ارزش توزیع شده استفاده می شود. ربات در محیط گردش می کند و در حالت های مختلف عمل های مجاز مختلف را می آزماید و مقادیر Q را بروز رساند. این مقادیر Q جهت یافتن قوانین فازی مناسب در هر قدم زمانی به کار می روند. هر رفتار به طور جداگانه پیاده رسانی شده و سپس روشی جدید جهت ترکیب رفتارهای معرفی می شایم. در انتها با کمک آزمایش های مختلف کارآیی راهبرد ارائه شده را نشان می دهیم.

الگوریتم مابه دلیل عدم استفاده از رهبردهایی نظری شبکه های عصبی یا الگوریتم های تکاملی که در اغلب تحقیقات مورداستفاده قرار گرفته اند، ساده و روش میباشد و چون نیازی به مشتق کمی و یا سایر عملیات پیچیده ریاضی ندارد کم هزینه و سریع می باشد. ربات با یک این کنترل کننده ساده میتواند مسیری نرم، هموار، کوتاه و حتی مناسب ترازویش های مذکور را انتخاب کرده و به سمت هدف خود پیش روید. در این کنترل کننده ترکیب رفتارهای سادگی، با روشی قابل درک و با حداقل هزینه ممکن عملی میباشد. به کار گرفتن اتوماتون یادگیری باعث می شود که کنترل کننده خود را با تغییرات محیط تطبیق نموده و شرایط محیط عملکرد آن کاهش نمی یابد. نکته مهم این است که تصمیم گیری بالاطلاعات بسیار ناچیز در مورد طبیعت محیط صورت می گیرد و ربات بدون داشتن هرگونه پیش فرضی به تنهایی میتواند نحوه تصمیم گیری، انتخاب و اجرای عمل مناسب را پیاموزد.

کیپرا از دو لایه متناظر با در برد اصلی

CPU تشکیل شده است: برد حسی حرکتی و برد لد

Pivot و عقب است، این پیکربندی برای مقابله با موئی

جهلو و مقابله با موئی پیچیده بسیار مناسب است چرا که باعث می

هندسی پیچیده بسیار مکان در جا بگردد.

شود ریات بتواند بدون تغییر مکان در جا بگردد.

کیپرا دارای نوعی حرکت است که بر عملکردهای جانبی

و متقارن استوار بوده در اغلب جانوران دیده می

شود، در ضمن ساختار مکانیکی و سیمایتیک سیمایتیکی

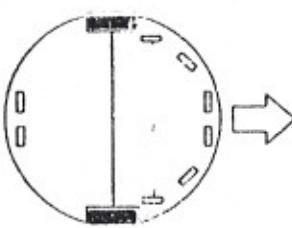
منطبق بر مدل سیمایتیکی از مدل های سیمایتیکی

مداد است. سیستم سنسوری ریات پایه بسیار ساده

و از هشت سنسور مسافت پایه مادون قرمز تشکیل

شده است که بر محیط ریات پخش شده اند، این

پیکربندی سنسوری در شکل (۲) نشان داده شده

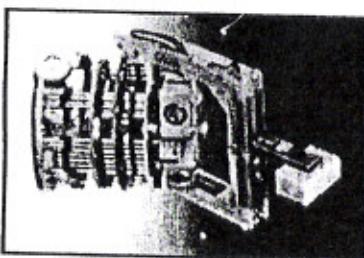


شکل (۲) پیکربندی سنسورهای مسافت پایه

این ترکیب بندی باعث حساسیت پیشتر در
این سمت ریات می کرد که می توان آنرا بعنوان جلوی
ریات شناسایی کرد. سنسورها فاصله از لبه فرسنده
و لبه گیرنده نورمادون قرمز تشکیل شده است. آنها می
توانند حضور اشیاء را با فرستادن و اندازه گرفتن نور
منعکس شده ریایی کنند. از این سنسورها همچنین
می توان بعنوان سنسورهای نورمادون قرمز غیرفعال
استفاده کرد.

کیپرا یک ریات استوانه ای به قطر ۵۰mm و
با ارتفاع متغیر است. این ریات مینیاتوری توسعه
میکرو انفرادی تکنیک F.Mondada ، A.Guignard ، E.Franzi
در سوئیس، طراحی و توسعه یافته و الگون توسعه
شرکت K-Team کنسلیت فدرال لوزان واقع
کیپرا به آنماشگر اجراه می دهد که یک محیط بزرگ را
در یک سطح محدود محقق کند (شکل ۱).

۳) ریات کیپرا

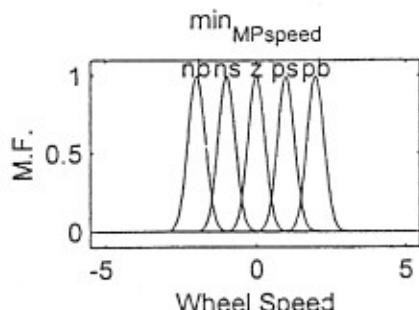


شکل ۱) ریات کیپرا

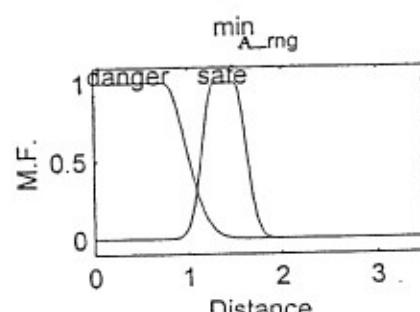
کیپرا مجهز به پاتریهای اسست که آنرا تا حدت
زمان مشخصی خودکفا نگاه می دارند. برد U يك
برد ازنه موتورولا همراه با EEPROM و RAM را در
خود جای داده است. یک مبدل AD اجراه کسب
سیگنال های آنالوگ و اصله از برد حسی - حرکتی را
می دهد. یک خط سریال برای اتصال ریات به کامپیوتر
و نیز تنظیمه اندزی در ریات تعییه کشته است. ارتباط
الکتریکی بین برد U يك و برد حسی - حرکتی توسط
اتصالاتی که پایه مکانیکی نیز هستند، برقرار می گردد.
این اتصالات بین برد U يك توسعه را برای اضافه کردن
ماجل های دیگر به ریات فراهم می کنند.

Kubota و Fukuda قانون مکافه ای جهت بدست اوردن تالی های قوانین
فازی استفاده نموده اند [۵]. در برخی از تحقیقات
الگریتم های زنتیکی و تکامل مستقیما برای مشخص
کردن و یا بهینه نوردن قوانین مازی به کار رفته اند [۶]
در این تحقیق ماقولین فازی را با استفاده از
یادگیری Q در هر قدم زمانی استخراج می نماییم.

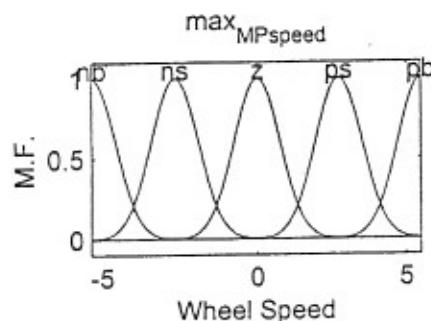
در این مقاله ابتدا به معرفی ریات کیپرا که در
پردازنده سپس در بخش ۲ ساختار کنترل کننده را
شرس می تهییم. اتماتون تصادفی پادگیر و گلریشم
یادگیر به ترتیب در پخش های ۱ و ۰ مفصل توضیح
داده خواهند شد. پخش ۱ به رفتار همکاری اختصاص
داده شده است. در پخش ۰ به بردسی عدکر کنترل
کننده فازی تطبیقی طراحی شده می پردازم و تابع
آزمایش ها را بیان می نماییم. پخش آخر شامل تابع
پیشنهادات می باشد و به دنبال آن لیست تعدادی از
مراجع مورد استفاده در این تحقیق ملاحظه می
فرمایید.



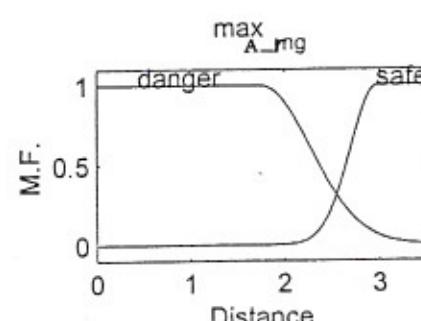
تابع عضویت فازی خروجی بازای
MP-Speed
حداکثر



تابع عضویت فازی ورودی بازای
A-mg
حداکثر



شکل ۴- تابع عضویت فازی خروجی بازای
MP-Speed
حداکثر



شکل ۲- تابع عضویت فازی ورودی بازای
A-mg
حداکثر

نزدیک ترین مانع را در جهت خود اندازه گیری می نماید. این مقدار باید نشان داده می شود. برای هر یک از متغیرهای ورودی و خروجی تعدادی متغیر زبانی و تابع عضویت فازی تعریف می شود. تابع عضویت فازی گاوی شکل انتخاب شده است. پارامترهای این تابع برای هر متغیر زبانی ورودی بر حسب بر دستوجه حسگرهای ریات (A_{mg}) و برای هر متغیر زبانی خروجی بر حسب حداکثر سرعت مجاز ریات (MP_Speed) در هر قدم زمانی بصورت تطبیقی انتخاب می شوند. این کار با کمک اتوماتون تصادفی یا دیگر خطی انجام می شود.

به شکل ۲ توجه فرمایید. این شکل تابع عضویت فازی ورودی را بازای حداکثر و حداکثر بر دستوجه نشان می دهد. شکل ۴ تابع عضویت فازی خروجی را بازای حداکثر و حداکثر (MP_Speed) یا حداکثر سرعت مجاز ریات نشان می دهد. در این شکل z (zero), n (negative), s (small), b (big), p (positive) می باشند.

مادر آزمایشات خود فرض نموده ایم که ربات حداکثر با سرعتی برابر با قطر خود حرکت می نماید و بیشترین فاصله ای که هر حسگر ریات در مقابل خود تشخیص میدهد، براساس میزان پراکندگی موانع در محیط بین ۱۹۲ تا ۸۲ میلی متر متفاوت است. در شبیه سازی ها تمام مقادیر با ۵۵ میلی متر که قطر ریات می باشد مقياس شده اند، و ریات تنها از حسگرهای جلو و طرفین خود برای ناوبری استفاده می نماید.

۳) ساختار کنترل کننده فازی

برای کنترل ربات از یک کنترل کننده فازی تطبیقی استفاده می نماییم. این کنترل کننده سه ورودی از سنسورها دریافت می نماید که به ترتیب، فاصله تا نزدیک ترین مانع سمت چپ، جلو و سمت راست ریات می باشند. کنترل کننده دو خروجی دارد که سرعت های چرخ چپ و چرخ راست می باشند. ریات در محیط گردش می کند، و در هر قدم زمانی هر سنسور ریات مانند فاصله ریات تا

کند و بر عکس وقتی تعداد موانع اطراف ربات کم باشد لازم است برد توجه ربات و همچنین سرعت حرکت آن افزایش یابند. باین ترتیب وقتی تراکم موانع زیاد است، ربات به آنها نزدیک شده و از میانشان عبور می نماید. در چنین حالتی چون خطر برخورد زیادتر است حرکت ربات آهسته تر صورت می گیرد. بر عکس وقتی تراکم موانع در اطراف ربات کم است، ربات بیشتر به جستجوی نواحی بدون مانع می پردازد و با توجه به کمتر بودن خطر برخورد سریع تر حرکت می نماید.

فرض کنید A_{mg} برد توجه و حداقل سرعت مجاز ربات باشند برای تنظیم این دو مقدار از اتوماتون یادگیر استفاده می نماییم. ورودی اتوماتون β یا مقدار تنبیه می باشد. این مقدار با کمکتابع F که آرگومان های آن از سنسورها گرفته می شود، مشخص می گردد. هر آرگومان یک مقدار با ینری است. این مقدار با ینری مشخص می کند که آیا سنسور مربوطه در حوزه دید خود مانع را تشخیص داده است یا نه. اتوماتون مورد استفاده دارای دو عمل مختلف می باشد. یک عمل Increase جهت افزایش مقادیر A_{mg} و MP_Speed و دیگری Decrease جهت کاهش این مقادیر می باشد. هر عمل یک مقدار احتمال مربوط به خوددارد که با p نشان داده می شود. این مقدار احتمال در هر قدم زمانی با توجه به عمل انتخاب شده و تنبیه دریافت شده به کمک رابطه زیر بروز درآورده می شود:

$$IF \alpha(n) = \alpha_i, \quad (2)$$

WHEN $\beta=0$

$$\begin{aligned} p_j(n+1) &= (1-a) \cdot p_j(n), \text{for all } j \neq i \\ p_i(n+1) &= p_i(n) + a \cdot [1 - p_i(n)] \end{aligned}$$

WHEN $\beta=1$

$$\begin{aligned} p_j(n+1) &= b / (r-1) + (1-b) \cdot p_j(n), \text{for all } j \neq i \\ p_i(n+1) &= (1-b) \cdot p_i(n) \end{aligned}$$

احتمال هر دو عمل در آغاز کار با هم مساوی و هر کدام برابر $5/4$ است. انتخاب شده اند. پارامترهای a و b بطور تجربی مساوی با $4/5$ در نظر گرفته شده اند. n قدم زمانی و r تعداد عمل های اتوماتون می باشند.

قوانين فازی به شکل if-then بیان می شوند. مثلاً قانون نام را می توان به کمک رابطه (۱) نشان داد. این قوانین در هر قدم زمانی با کمک الگوریتم یادگیری Q مشخص می شوند. در این رابطه در قسمت (۵) جزئیات بیشتری بیان شده است.

IF d_1 is ILV_{i,1} and d_2 is ILV_{i,2} and d_3 is ILV_{i,3}
THEN v_1 is OLV_{i,1} and v_2 is OLV_{i,2} (۱)

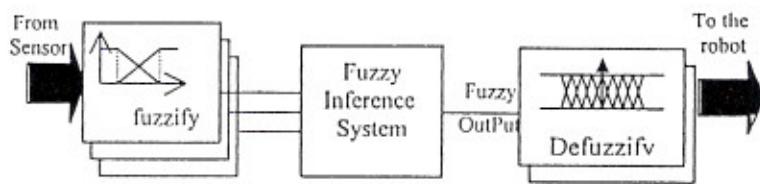
OLV و ILV Input Linguistic Variable و Output Linguistic Variable میباشند.

سیستم استنتاج فازی مدانی است که بر اساس عملکردهای استنتاج دفاری Center_of_Area عمل می کند. روش دفاری کردن کننده را نشان می دهد. برای درک معناری کلی کنترل کننده را نشان می دهد. این روش روشن تر نحوه ارتباط کنترل کننده فازی، اتوماتون یادگیر و الگوریتم یادگیری Q به شکل ۶ توجه نمایید.

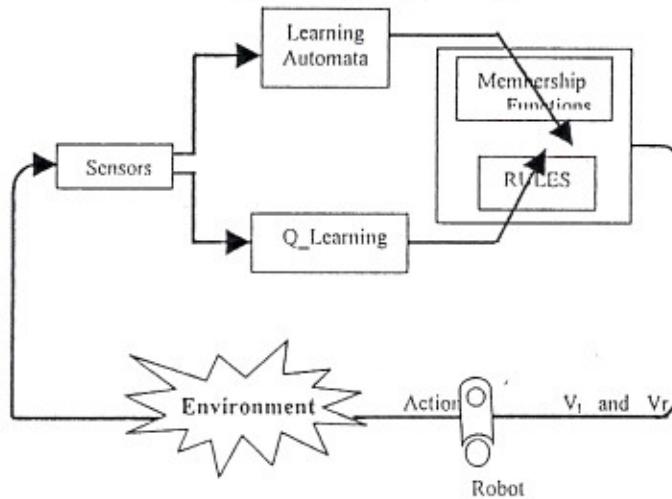
۴) شرح اتوماتون یادگیر

یک اتوماتون یک ماشین یامکانیزم کنترلی است، با این قابلیت که می تواند در این تغییرات محیط تطبیق دهد. یک اتوماتون بدون داشتن هرگونه اطلاعات در مورد عمل پیشنهادی سعی بر یافتن جواب مناسب برای مساله دارد. در آغاز مقدار احتمال برای تمام عمل هایکسان است. عملی به تصادف انتخاب می گردد. سپس پاسخ محیط به این عمل خاص مشاهده می شود. مقدار احتمال این عمل بر اساس این پاسخ محیط بروز درآورده می شود. این مراحل مجدداتکرار می شوند. یک اتوماتون که به شکل فوق سعی در بهبود عملکرد خود داشته باشد اتوماتون یادگیر گفته می شود [۸].

طراحی رفتارهای مختلف برای یک ربات متحرک در یک محیط دینامیک (دائمًا در حال تغییر) کاملاً وابسته به شرایط محیط می باشد. در طراحی رفتار پرهیز از مانع، کنترل برد توجه ربات و نیز سرعت حرکت آن با توجه به تعداد موانع در عملکرد سیستم کنترل و بروز رفتار دلخواه بسیار موثر است. هر گاه تعداد موانع در اطراف ربات زیاد باشد، لازم است ربات بیشتر روی نواحی نزدیک تر به خود تمرکز داشته باشد و با سرعت آهسته تر حرکت



شکل ۵) معماری کنترل کننده فازی



شکل ۶) نحوه ارتباط کنترل کننده فازی، اوتوماتون تصادفی یادگیر و یادگیری Q

عملی که در یک شرایط خاص منجر به شکست شود، احتمال آن با کمک رابطه زیر به دست می‌آید:

شانس کمتری برای انتخاب خواهد داشت و بر عکس،
به این ترتیب با توجه به تراکم موانع در اطراف ربات در
هر قدم زمانی می‌توان Δ_{rng} و Δ_{Speed} مناسب را
یافت. از Δ_{rng} جهت مشخص کردن پارامترهای توابع
عضویت فازی ورودی و از Δ_{Speed} جهت مشخص
کردن پارامترهای توابع عضویت فازی خروجی
استفاده می‌شود (شکل های ۲ و ۴).

مقدار تنبیه β با کمک رابطه زیر به دست می‌آید:

$$IF \alpha(n) = 'Increase' \quad (3)$$

$$IF(ActiveSensorNum \geq 1/3.TotalSensorNum)$$

$$\beta = 1$$

$$ELSE IF \alpha(n) = 'Decrease'$$

$$IF(ActiveSensorNum < 1/3.TotalSensorNum)$$

$$\beta = 1$$

$$ELSE$$

$$\beta = 0$$

۵) شرح الگوریتم یادگیر

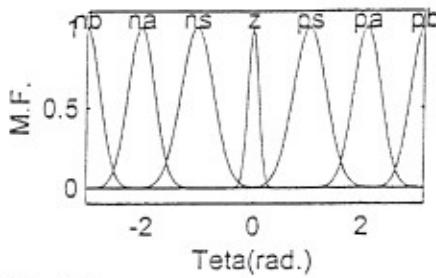
هدف الگوریتم یادگیر یافتن نگاشتی مناسب از مجموعه حالت‌های محیط به مجموعه عملیات مجاز و ممکن برای ربات می‌باشد. گفتیم که در شبیه سازی از ربات کیپرا استفاده نموده ایم، این ربات ۸ سنسور دارد که در آزمایشات تنها سه سنسور آن، جهت تشخیص موانع در جلو، سمت راست و سمت چپ ربات به کار می‌روند. هر سنسور فاصله تا نزدیک ترین مانع به ربات را در یک جهت خاص اندازه گیری می‌نماید. مقدار اندازه گیری شده با توجه به توابع عضویت فازی

$ActiveSensorNumber$ تعداد سنسورهایی است که حداقل یک مانع در حوزه دید خود تشخیص داده اند و $Total Sensor Number$ تعداد کل سنسورهای ربات می‌باشد. در هر قدم زمانی اوتوماتون با توجه به احتمال هر عمل و باروش رویوت ویل، عملی راجه اجراء انتخاب می‌نماید. این عمل بررسی بر دستوجه وحداکثر سرعت مجاز تاثیرگذاشته، مقدار آنها را به ترتیب به اندازه ۰/۰۵ و ۰/۰۵ در شبیه سازی ها معادل ۰/۰۵ و ۰/۰۷۵ میلی مترا در دنیای واقعی تغییر می‌دهد.

می شود. بدون اینکه در کارآیی الگوریتم اثربال ملاحظه ای داشته باشد.

۶ رفتار هدف یابی

ناوبری در حقیقت ترکیب دو رفتار پرهیز از مانع و هدف یابی است. برای هدف یابی نیز از روش مشابه پرهیز از مانع یعنی سیستم فازی و یادگیری Q استفاده می نماییم. برای مشخص کردن حالت های ورودی از زاویه میان ریات و هدف استفاده می شود. این زاویه در فاصله $[pi - pi]$ محاسبه می گردد. چرخش خلاف جهت عقریه های ساعت با مقدار مثبت و چرخش در جهت حرکت عقریه های ساعت با مقدار منفی نشان داده می شود (جهت مثلثات استاندارد). زاویه میان ریات و هدف می تواند هر یک از مقادیر فازی شکل ۷ را اختیار نماید.



شکل ۷) توابع عضویت فازی ورودی برای رفتار هدف یابی نگاشت میان مجموعه حالت و مجموعه اعمال مجاز توسط یادگیری Q انجام می گیرد. ابتدا مقدار Q برای تمام زوج های حالت و عمل مجاز یک در نظر گرفته می شود. یک الگوریتم انتخاب، در ۲۰ درصد از موقعیت بطور تصادفی (exploration) و در ۸۰ درصد از موقعیت با توجه به مقدار Q، عملی را جهت انجام در هر یک از حالات ممکن انتخاب می کند. (مرحله انتخاب قوانین فازی) ریات این عمل را در محیط اجرا می کند و بازای آن پاداش یا تنبیه دریافت می دارد. که برای بروز در آوردن مقادیر Q(s,a) با کمک رابطه (۵) استفاده می شود. γ هردو مساوی $\alpha/9$. انتخاب شده اند. در رفتار هدف یابی میزان پاداش مناسب با کاهش زاویه میان جهت حرکت ریات و هدف تعریف می شود. در صورتی که این زاویه کوچکتر از $8/\pi$ باشد پاداش زیانتری دریافت می گردد.

وروی برای هر سنسور می تواند در یکی از دو ناحیه safe یا dangerous واقع شود. بنابر این روابط نسبت به موضع در ۸ وضعیت مختلف ممکن است قرار گیرد. حال باید مشخص کرد که در هر وضعیت چه عملی مناسب تر است.

الگوریتم یادگیری استفاده شده Q-Learning می باشد. برای حالتی که هر سه طرف ریات در ناحیه خطرنانک تشخیص داده شده اند، عمل مناسب چرخش درجا در نظر گرفته شده است. برای سایر حالات الگوریتم یادگیر عمل مناسب را تعیین می نماید.

در ابتدا مقدار Q برای تمام زوج های حالت و عمل مجاز یک در نظر گرفته می شود. یک الگوریتم انتخاب، در ۲۰ درصد از موقعیت بطور تصادفی (exploration) و در ۸۰ درصد از موقعیت با توجه به مقدار Q، عملی را جهت انجام در هر یک از حالات ممکن استفاده می گردد. این عمل را در محیط اجرا می کند و بازای آن پاداش یا تنبیه دریافت می دارد، که برای بروز در آوردن مقادیر Q(s,a) استفاده می گردد.

برای رفتار پرهیز از مانع مقدار پاداش براساس میزان افزایش نزدیک ترین فاصله اندازه گیری شده تامانع توسط هرسگر ریات، تعیین میگردد. هرسگر جلو اهمیت بیشتری دارد، بنابراین به آن وزن بزرگتری داده می شود. و اگر در چهت این هرسگر فاصله تامانع افزایش یابد، پاداش بیشتری دریافت خواهد شد.

$$\begin{aligned} \text{delta_d}_i &= \text{new_d}_i - \text{prev_d}_i, i = 1, 2, 3 \\ r &= \sum_{i=1}^3 w_i \cdot \text{delta_d}_i \\ w &= \{1, 10, 1\} \end{aligned} \quad (4)$$

میزان افزایش نزدیک ترین فاصله delta_d_i اندازه گیری شده تامانع توسط هرسگر ریات می باشد و وزن W اهمیت آن هرسگر است. برای بروز در آوردن مقادیر Q(s,a) از رابطه زیر استفاده می گردد. α و γ هر دو برابر $9/40$ انتخاب شده اند.

$$\begin{aligned} Q(s, a) &= Q(s, a) + \alpha[r + \gamma V(y) - Q(s, a)] \quad \alpha \in (0, 1] \\ V(y) &= \max_a Q(x, a) \end{aligned}$$

الگوریتم شرح داده شده با دقت یک رقم اعشار پس از ۴۸۰ قدم زمانی با تعداد ۴ برشورد همگرا می شود. اما اگر از قسمت اعشار صرف نظر کنیم با تنها ۲۷۰ قدم زمانی و ۲۰ برشورد الگوریتم همگرا

1)		$\theta = -\pi, \theta = \pi$	5)		$\theta = 0$
2)		$-\pi < \theta < -\pi/2$	6)		$0 < \theta < \pi/2$
3)		$\theta = -\pi/2$	7)		$\theta = \pi/2$
4)		$-\pi/2 < \theta < 0$	8)		$\pi/2 < \theta < \pi$

شکل ۸) نموده هایی از حالات در ریات همکار نسبت به هم

یک الگوریتم انتخاب در ۲۰ درصد از موقعیت بطور تصادفی (exploration) و در ۸۰ درصد از موارد با توجه به مقادیر Q عملی را جهت انجام در هر یک از حالات ممکن انتخاب می کند. (مرحله انتخاب قوانین فازی) ریات این عمل را در محیط اجرا می کند و بازی آن پاداش یا تنبیه دریافت می دارد. که برای بروز در آوردن مقادیر Q استفاده می شود.

در اینجا چون از الگوریتم یادگیری توزیع شده استفاده می شود علاوه بر پاداش و حالت بعدی خود ریات، پاداش و حالت بعدی ریات دیگر نیز در بروز در آوردن مقادیر Q مورد استفاده قرار می گیرند.

برای مشخص کردن مقدار تنبیه یا پاداش از دو معیار استفاده می شود. یکی اینکه فاصله ریات تا ریات دیگری که در جلو خود حس کرده کاهش یا افزایش یابد، که متناسب با هیزان این کاهش یا افزایش، تنبیه یا پاداش دریافت می دارد. دیگری اینکه دو ریات با یکدیگر برخورد کنند که در این صورت یک تنبیه ثابت مساوی ۱۰- در نظر گرفته می شود.

۸) ترکیب رفتارهای همکاری، پرهیز از مانع و هدف یابی

با زمینه بحث ترکیب کردن رفتارهای پرهیز از مانع و هدف یابی مطرح است. اما این بار وضعیت دو ریات نسبت به هم نیز در تصمیم گیری برای انتخاب عمل مناسب بسیار مهم است. برای حل مساله از راه حلی شبیه به ناوبری استفاده می نماییم. به این ترتیب که همکاری و پرهیز از مانع از اولویت بالاتری نسبت به هدف یابی برخوردارند. حال اگر در یک قدم زمانی خاص بتوان عملی یافته که برای هرسه منظور مناسب باشد، آن عمل را انتخاب کرده انجام می دهیم. در غیر

۷) همکاری

در اینجا نیز از روش مشابه قبل بعضی سیستم فازی و یادگیری Q استفاده می نماییم با این تفاوت که نگاشتمیان حالات محیط و اعمال مجاز توسط الگوریتم یادگیری Q توزیع شده انجام می گیرد.

برای مشخص کردن حالت های ورودی از دو معیار استفاده می نماییم . یکی همسایگی است که مشخص می کند آیا سنسور جلو ریات، وجود ریات دیگر را حس کرده است یا نه. اگر حس نکرده باشد ریات می تواند از همان الگوریتم ناوبری ساده استفاده نماید و نیازی به همکاری نیست . اما اگر این سنسور وجود ریات دیگر را تشخیص داده باشد ، وضعیت حساس تری دارد و باید به گونه ای متفاوت عمل کند. چرا که ریات دوم خود قادر به تصمیم گیری، حرکت و تغییر شرایط محیط است. از طرفی این ریات دوم هم در محیط وظایفی دارد که اگر نتواند به آنها عمل کند عملکرد کل سیستم کاهش می یابد. معیار دوم زاویه میان راستای حرکت ریات و ریاتی که در جلوی خود تشخیص داده است. می باشد.

الگوریتم یادگیری Q توزیع شده برای بروز در آوردن مقادیر $Q(s,a)$ از رابطه زیر استفاده می نماید:

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha(R(s,a) + \gamma \sum_j f(j) V_j(s'))$$

$$V_j(s') = \max_a Q_j(s',a)$$

وقتی تعداد نودها (ریات ها) دو باشند و زن های $f(1,1)=f(2,2)=1$ باشند رابطه (مثلاً برای ریات ۱) به شکل زیر تبدیل می گردد. γ, α برابر $9/6$. انتخاب میشوند.

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha(R(s,a) + \gamma(V_1(s') + V_2(s')))$$

$$V(s') = \max_a Q(s',a)$$

ترکیب رفتارهای مختلف موضوعی مفصل و
گستردۀ تحت عنوان Behavior Blending و Behavior Fusion می باشد . ما برای ترکیب رفتارها از روشی بسیار ساده و بسیار کار آمد استفاده نموده ایم، به این ترتیب که به طور ثابت رفتار پرهیز از مانع از اولیت بالاتری نسبت هدف یابی برخوردار است . حال اگر در یک قدم زمانی خاص می توان عملی مناسب برای هر دو رفتار انتخاب نمود، آن عمل را انتخاب کرده و انجام می دهیم، در غیر این صورت عملی را انجام می دهیم که برای رفتار پرهیز از مانع مناسب باشد. این کار با یک ماکریزیم گیری ساده انجام می شود.

$$Q = \min(\text{avoid}Q(s_j, a_i), t \text{ target}Q(t_n, a_i)) \quad (9)$$

$$\text{nav}Q(s_j, a_i) = \begin{cases} Q & \text{if } Q \geq \tau \\ \text{avoid}Q(s_j, a_i) & \text{otherwise} \end{cases}$$

مقادیر Q به دست آمده برای رفتار پرهیز از مانع target Q به دست آمده برای رفتار هدف یابی و $\text{nav}Q$ حاصل برای رفتارناوبری می باشند. با مشخص شدن مقادیر Q برای رفتارناوبری، قوانین فازی مناسب در هر قدم زمانی تعیین شده، رفتارناوبری با موقوفیت انجام می پذیرد.

جدول ۱ و شکل ۹ نشانگر اهمیت به کارگیری اتوماتون یادگیر برای تطبیقی کردن کنترل کننده می باشند. در این جدول D_{average} خطر ماکریزیم، میانگین خطرو Length Moving طول مسیر می باشد. خطر ماکریزیم طبق رابطه زیر محاسبه می گردد:

$$D = \sum_{j=1}^n \left(1.0 - \frac{x_j(t)}{A - \text{rng}(t)} \right) \quad (10)$$

$$D_{\text{max}} = \text{Max}_i(D)$$

که در آن D مقدار خطر در لحظه t ، $x_j(t)$ فاصله اندازه گیری شده تا نزدیک ترین مانع به ربات، توسط سنسور زام و n تعداد سنسورهای ربات می باشند. اقدم زمانی و D_{max} ماکریزیم خطر است. به عبارتی ماکریزیم خطر، بیشترین D اندازه گیری شده طی مسیر، در قدم های زمانی مختلف می باشد.

این صورت عملی را انجام می دهیم که برای دو منظور همکاری و پرهیز از مانع مناسب باشد.

این کار با کمکتابع nav2rob-choose انجام شده است. در هر قدم زمانی ابتدا وضعیت ربات نسبت به هدف (gstate) و نیز وضعیت ربات نسبت به دیگر مشخص می گردد $Cstate$ ، $gstate$ هستند که بیشترین شباهت را به حالت جاری دارند . (حالات جاری در این حالت ها بزرگترین درجه عضویت را دارد).

$$Q1 = \min(Q(gstate, a_n), Q(ObstacleAv idState_{**}, a_n))$$

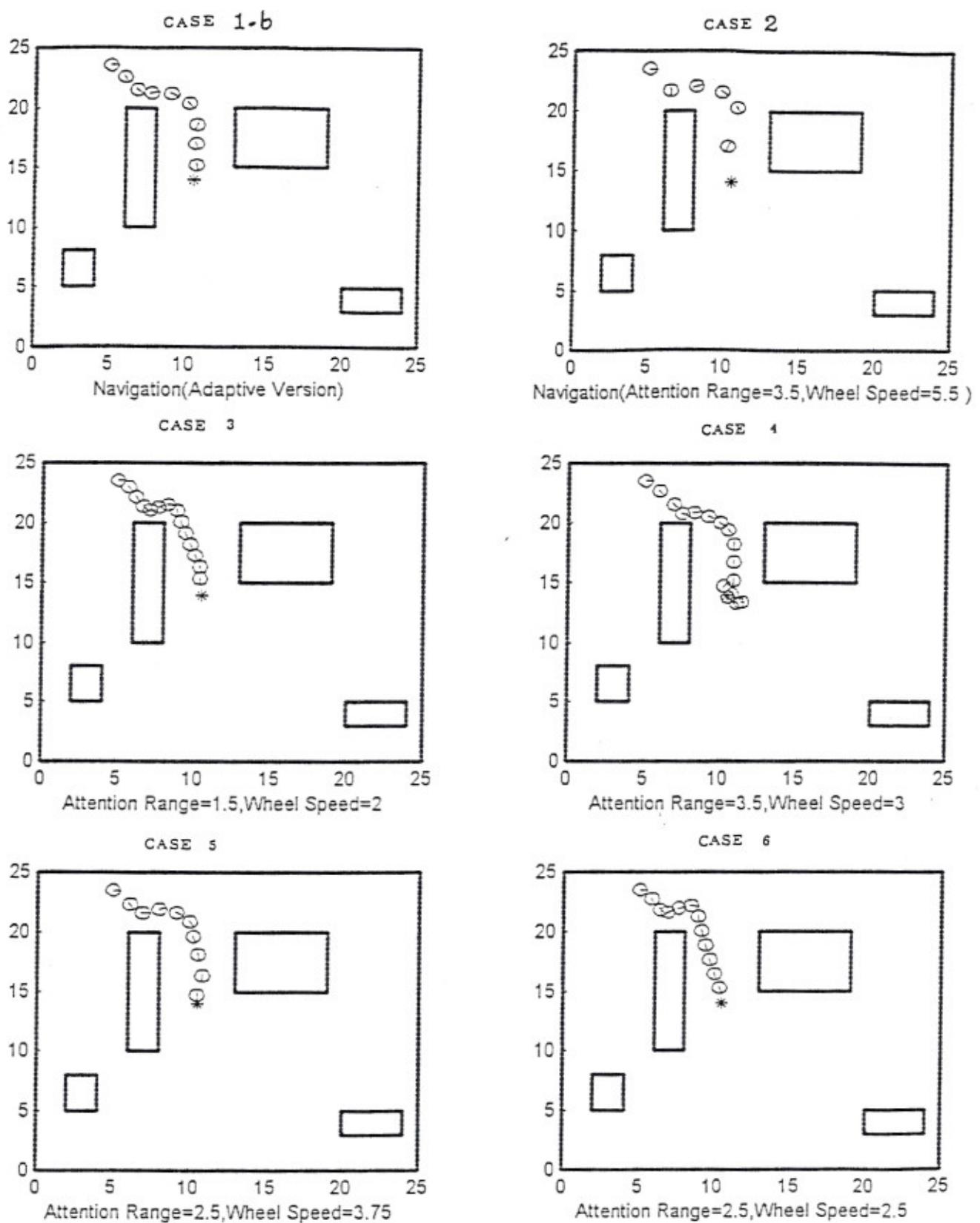
$$Q2 = \min(Q(gstate, a_n), Q(ObstacleAv idState_{**}, a_n))$$

$$Q3 = \min(Temp1, Temp2) \quad \text{for } n = 1, \dots, 25, m = 1, \dots, 8$$

در اینجا نیز یک الگوریتم انتخاب در ۲۰ درصد از موقعیت به طور تصادفی (exploration) و در ۸۰ درصد از موارد با توجه به ماکریزیم $Q3$ یا $Q2$ فازی را انتخاب می نماید. باین ترتیب که اگر ماکریزیم $Q3$ از یک مقدار آستانه بیشتر باشد، عمل مناسب برای هر سه رفتار را ارائه می دهد . در غیر این صورت ماکریزیم $Q2$ بررسی می شود و جهت انتخاب عمل مناسب استفاده می گردد.

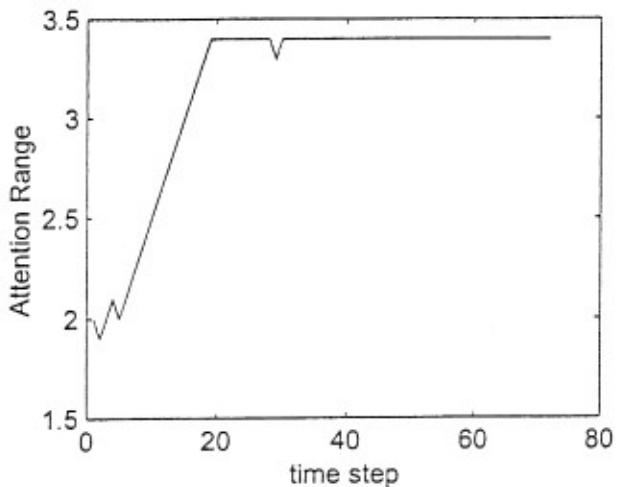
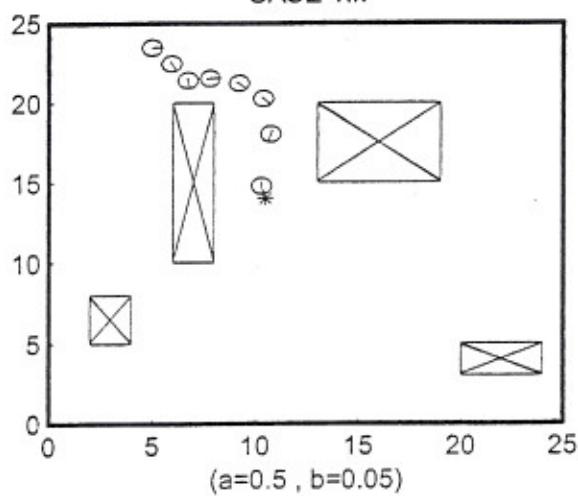
۹) بررسی عملکرد کنترل کننده طراحی شده

ناوبنی عبارت است از حرکت ربات در یک محیط به سمت هدف در حالیکه از مانع پرهیز می کند. برای حل مساله ناوبری دو روش کلی موجود است. روش اول اینکه هر دو رفتار پرهیز از مانع و هدف یابی را به ترتیب کنترل کنند. این روش بزرگتر شدن فضای جستجو و کندتر شدن همگرایی می باشد. روش دوم (که ما از این روش استفاده می کنیم) این است که هر یک از دو رفتار پرهیز از مانع و هدف یابی بطور جداگانه بررسی و یادگرفته شود و سپس با ترکیب این دو رفتار، ناوبری تحقق یابد. مثلاً اگر هدف در سمت راست ربات باشد و مانع در جلو آن، رفتار پرهیز از مانع ربات را از جلو رفتن منع می کند و رفتار هدف یابی آن را به سمت راست هدایت می کند، به این ترتیب ربات نهایتاً به راست می رود و ناوبری انجام می پذیرد. اما ترکیب این دو رفتار همیشه به این سادگی نیست مثلاً حالتی را در نظر آورید که مانع و هدف هر دو در مقابل ربات قرار دارند و مانع میان ربات و هدف می باشد.

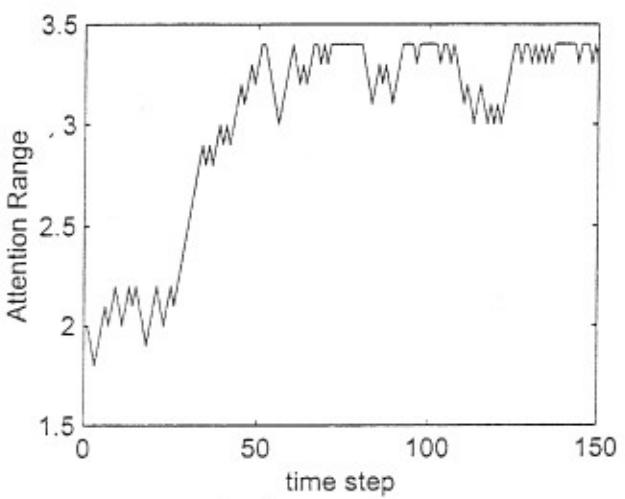
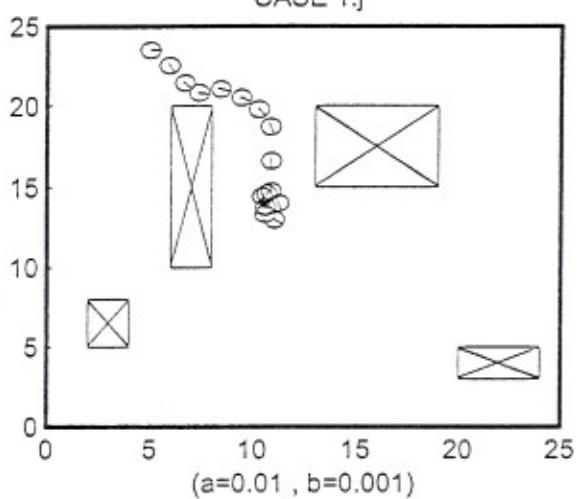


شکل ۹) اهمیت اتوماترنس تصاویر پارکیر در شرط عملکرد کنترل کننده

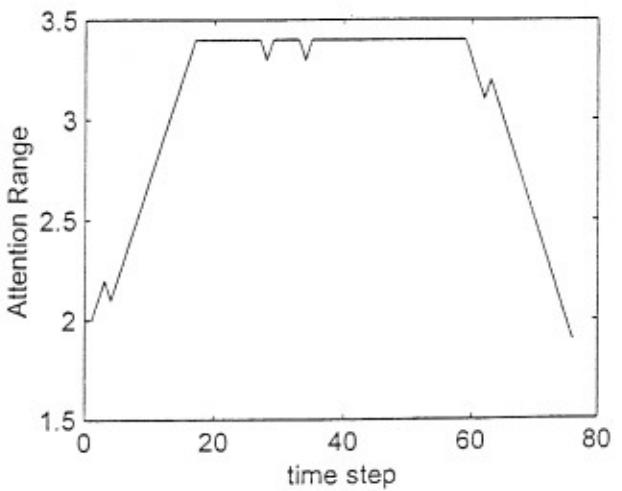
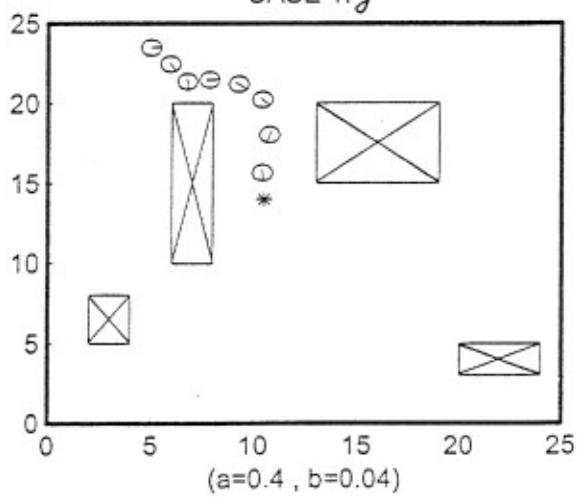
CASE 1.k



CASE 1.j



CASE 1.g



CASE	A-rng	Speed	S.L.A.param a	S.L.A.param b	D _{average}	MovingLength	D _m	TimeStep
1.a	Adaptive	Adaptive	0.2	0.2	0.1088	13.0527	0.8234	83
1.b	Adaptive	Adaptive	0.4	0.4	0.1095	12.6224	0.8678	78
1.r	Adaptive	Adaptive	0.5	0.5	0.1179	13.5473	0.8235	84
1.c	Adaptive	Adaptive	0.6	0.6	0.1211	13.0256	0.8235	83
1.d	Adaptive	Adaptive	0.8	0.8	0.1006	17.6195	0.8235	108
1.e	Adaptive	Adaptive	0.1	0.01	0.1343	13.1003	0.8705	78
1.f	Adaptive	Adaptive	0.2	0.02	0.1334	20.8869	0.8398	116
1.g	Adaptive	Adaptive	0.4	0.04	0.1250	13.1157	0.8235	76
1.k	Adaptive	Adaptive	0.5	0.05	0.1434	13.0741	0.8235	72
1.h	Adaptive	Adaptive	0.6	0.06	0.1354	12.6825	0.8235	71
1.i	Adaptive	Adaptive	0.8	0.08	0.1021	14.6554	0.8235	95
1.j	Adaptive	Adaptive	0.01	0.001	0.1254	21.5008	0.8833	*175
1.l	Adaptive	Adaptive	0.02	0.002	0.1208	13.0323	0.8235	103
1.m	Adaptive	Adaptive	0.04	0.004	0.1414	17.5586	0.8235	106
1.r	Adaptive	Adaptive	0.05	0.005	0.1449	12.8454	0.8423	78
1.n	Adaptive	Adaptive	0.06	0.006	0.1409	12.9513	0.8235	77
1.o	Adaptive	Adaptive	0.08	0.008	0.1427	13.1618	0.8435	80
2	3.5	5.5	None	None	0.1755	13.8931	0.8571	59
3	1.5	2	None	None	0.0539	12.1369	0.6340	139
4	3.5	3	None	None	0.1466	16.4693	0.9445	*178
5	2.5	3.75	None	None	0.0978	12.9748	0.6	92
6	2.5	2.5	None	None	0.1060	12.6233	0.7	118

*: can't find goal in 150 time steps.

Table 1.The obtained results of the various cases examined in fig. 9

هدف دو نیز مربوط به ربات دو بوده و با علامت * نشان داده شده است. ربات ها را پس از هر ۱۰ قدم زمانی در محیط نمایش داده ایم.

در شکل ۱۲ دو ربات، در حالت ۵، از مجموعه

حالت های تعریف شده در شکل (۸) قرار داده شده اند. یعنی ۱ phi و ۲ (زاویه میان راستای هر ربات با افق در جهت مثبت استاندارد) هر دو صفر هستند. همان طور که دیده می شود، در این حالت ربات ۲ راه ربات ۱ را سد نموده است. ولی ربات ۲ تراحمی برای ربات ۱ ندارد. ربات ۲ ابتدا بسیار آهسته حرکت می نماید و در عین حال سعی دارد ربات ۲ را از پایین دور بزند چرا که در بالا با مانع مواجه است (شکل ۸-۱۲).

رفته رفتہ ۲ به سمت هدف خود می رود و از مسیر ۱ خارج می شود. با کنار رفتن ۲ از سر راه ۱ هر ربات

مسیر خود را به سمت هدفش طی می کند. شکل ۸-۱۲-f

تفییرات برد توجه را برای هریک از ربات ها نمایش می

دهد. وقتی که شرایط موقعیت ربات نسبت به موانع

ثابت باشد، برد توجه هم ثابت می شود. بر عکس، با

تفییرات مداوم محیط و میزان پراکندگی موانع،

تفییرات برد توجه هم سریع ترمی شود.

در شکل ۱۲ دو ربات در حالت ۸، از مجموعه

حالت های تعریف شده در شکل ۸ قرار داده شده اند.

یعنی ۱ phi برابر صفر و ۲ pi/۶ می باشد.

در این حالت هر دو ربات راه یکدیگر را سد نموده اند.

بنابر این چون هر ربات برای عبور به سمت

هدف خود با ربات دیگر مواجه می شود حرکت ربات

ها به آهستگی انجام می شود. (شکل ۸-۱۳-۵)، پس از

عبور ۲ مسیر حرکت برای ۱ باز و برای ۲ نسبتاً باز

می شود. پس ۱ با سرعت بیشتر به سمت هدف خود

پیش می رود. ولی ۲ که هنوز مسیر خود را کاملاً باز

نمی یابد آهسته تر حرکت می نماید (شکل ۸-۱۳-۵) در

آخر با خارج شدن ۲ از مسیر ۱، هر دو ربات به سمت

هدف خود پیش می روند (شکل ۸-۱۳-۵).

ربات ۲ پس از ۵۹۰ قدم زمانی و ربات ۱

پس از ۳۹۰ قدم زمانی به هدف خود می رسد. هر قدم

زمانی ۱/۰ ثانیه می باشد. طی این مسیر خطر ماکزیم

برای ۲ مقدار ۱/۵ و برای ۱ برابر ۸/۴ می باشد.

میانگین خطر نیز برای ۲ مساوی ۱۱/۰ و برای ۱ برابر

۰/۰۷ می باشد. شکل (۸-۱۳) تغییرات برد توجه را

برای ربات ها نمایش می دهد.

این میانگین با کمک رابطه زیر محاسبه می شود:

$$D_{average} = \frac{1}{P_{time}} \cdot \frac{1}{n} \cdot \sum_{i=1}^{P_{time}} \sum_{j=1}^n \left(1.0 - \frac{x_j(t)}{A - rng(t)} \right) \quad (11)$$

P_{time} کل قدم های زمانی لازم، از لحظه

شروع حرکت تا لحظه رسیدن به هدف می باشد و

$A - rng(t)$ برد توجه بر قدم زمانی است. بادقت

در شکل ۹ و توجه به جدول ۱ می توان به نتایج

زیردست یافت:

۱_ در صورت عدم استفاده از اتوماتون

یادگیر (کنترل کننده غیر تطبیقی) یانتخاب نامناسب

پارامترهای آن ، مسیر ناهموار خواهد شد

(CASE ۱_g و ۱_b)، یا میزان خطر طی مسیر

افزایش خواهد یافت. ۱_g یا ۱_b (CASE ۱_k)، یا

رسیدن به هدف با تأخیر انجام خواهد شد و ربات دیرتر

به هدف خود خواهد رسید. ۱_g یا ۱_b (CASE ۱_g).

۲_ اگرچه می توان بازای بعضی مقادیر

خاص برای MP_Speed یا A_mg به نتایج خوبی دست

یافت (CASE ۵)، اما اولاً این نتایج به خوبی حالت

طبیقی نخواهد بود و ثانیاً یافتن آنها عملی پرهزینه

و زمانبند خواهد بود. باین ترتیب اهمیت اتوماتون

یادگیر که ای تواند در هر قدم زمانی مقدار مناسب این

متغیرها را بطور اتوماتیک محاسبه کند آشکار ترمی

گردد.

شکل (۱۰) نمونه ای از رفتار ناوبری یک ربات

متحرك در یک محیط خاص را نشان می دهد. ربات پس

از ۳۲۹ قدم زمانی (هر قدم زمانی ۱/۰ ثانیه می باشد)

، مسیری به طول ۲/۳ را طی می نماید. ماکزیم خطر

۱/۲۸ و خطر میانگین ۱۲/۰ اندازه گیری شده اند.

ربات برای رسیدن به هدف خود مسیری مناسب ، کوتاه

و هموار انتخاب می نماید و به خوبی به هدف خود می

رسد. شکل ۱۱ مثال دیگی را نشان می دهد. به نمودار

هر شکل توجه کنید. این نمودار تغییرات برد توجه

ربات را در حین حرکت نشان می دهد. این تغییرات

حاصل عملکرد اتوماتون یادگیر می باشد.

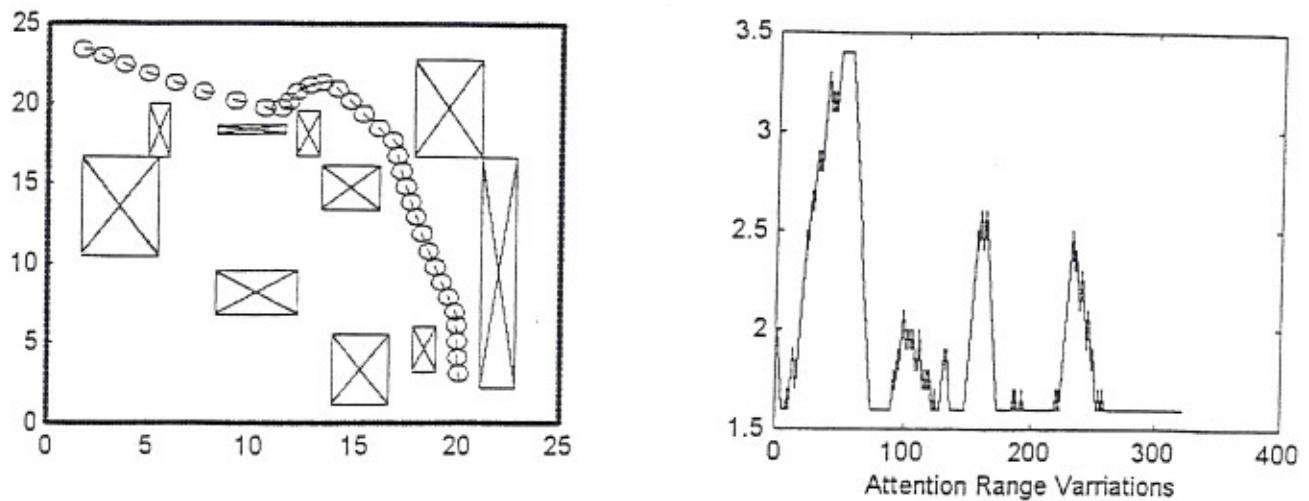
شکل های (۱۲) و (۱۲) نحوه عملکرد دو ربات

در یک محیط را نشان می دهند. در این شکل های

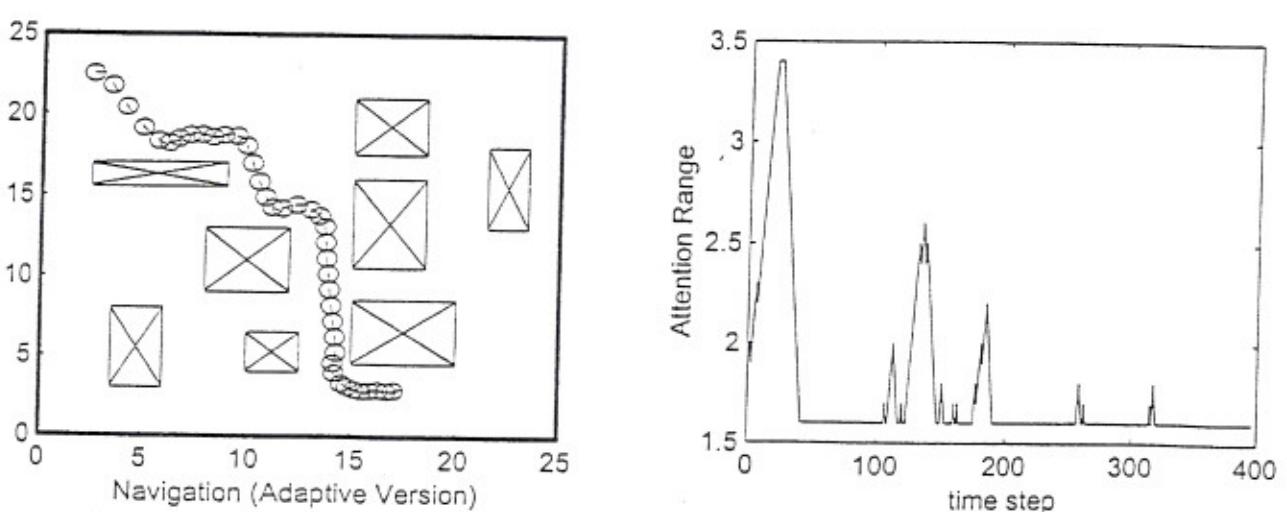
نماینده ربات (robot)، g نماینده هدف (goal) و نقطه

شروع (start) می باشند. هدف یک مربوط به ربات یک

می باشد و با علامت نشان داده شده است.



شکل ۱۰) تاوبری یک ربات متحرک در یک محیط نمونه با استفاده از کنترل کننده فازی تطبیقی

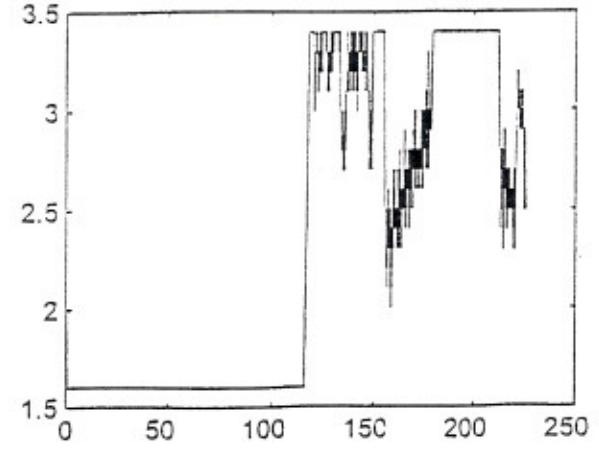
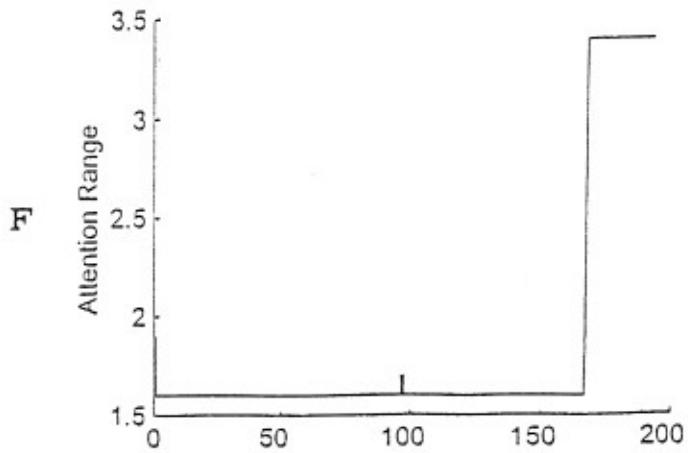
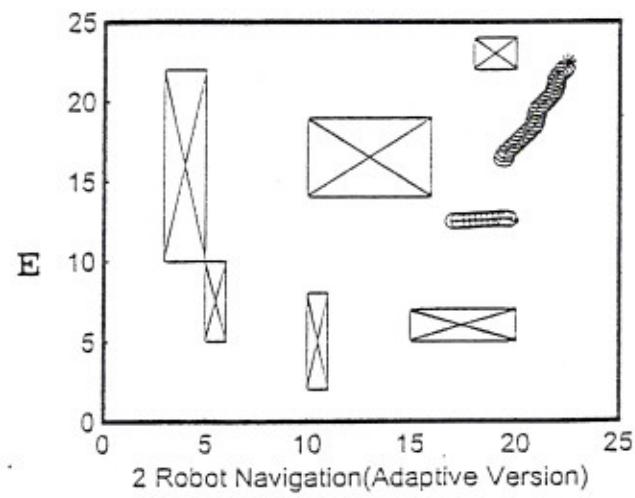
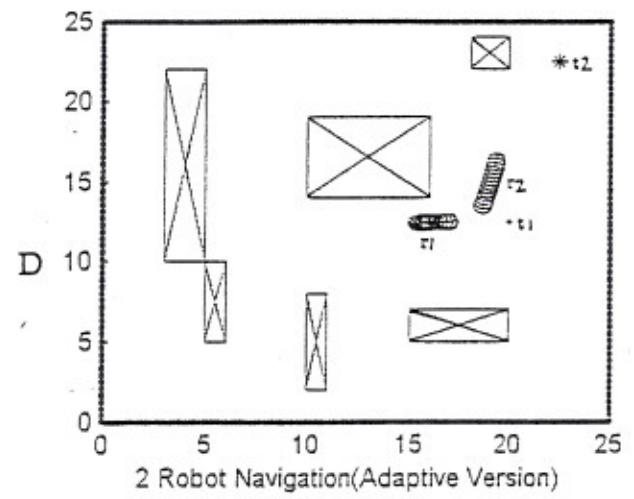
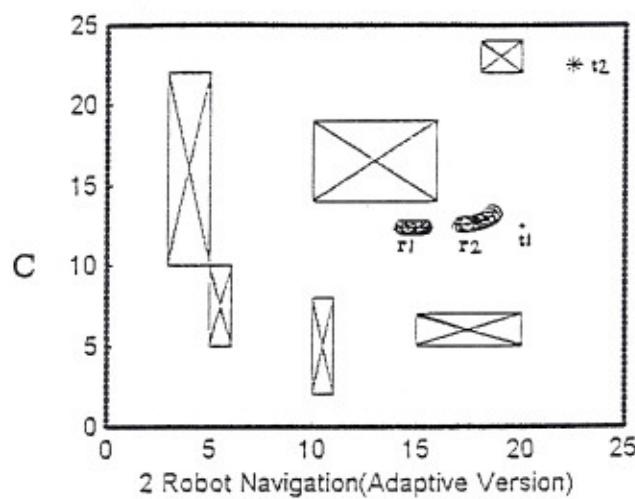
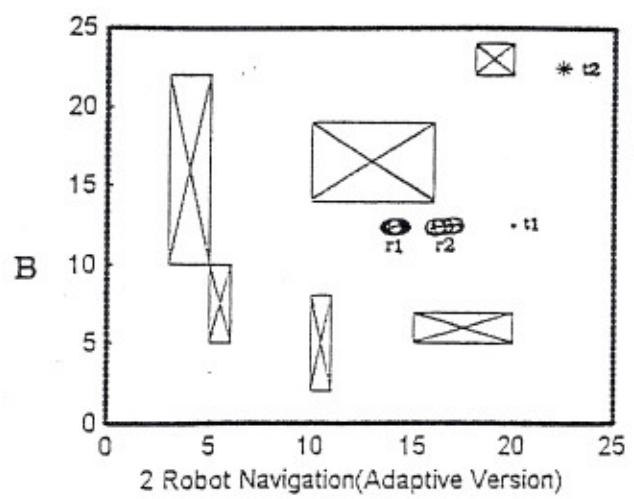
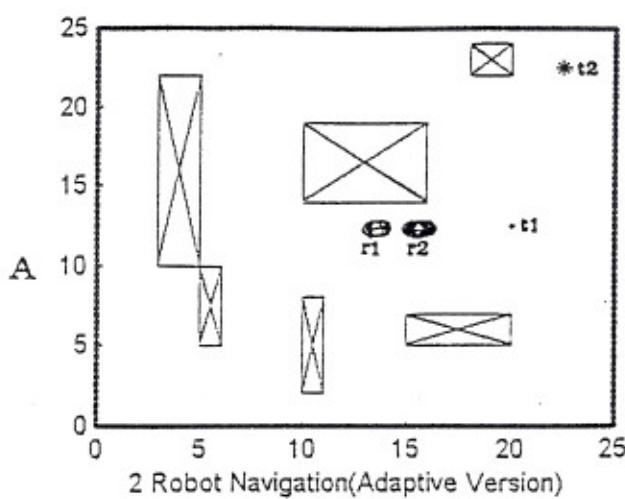


شکل ۱۱) تاوبری یک ربات متحرک در یک محیط نمونه با استفاده از کنترل کننده فازی تطبیقی

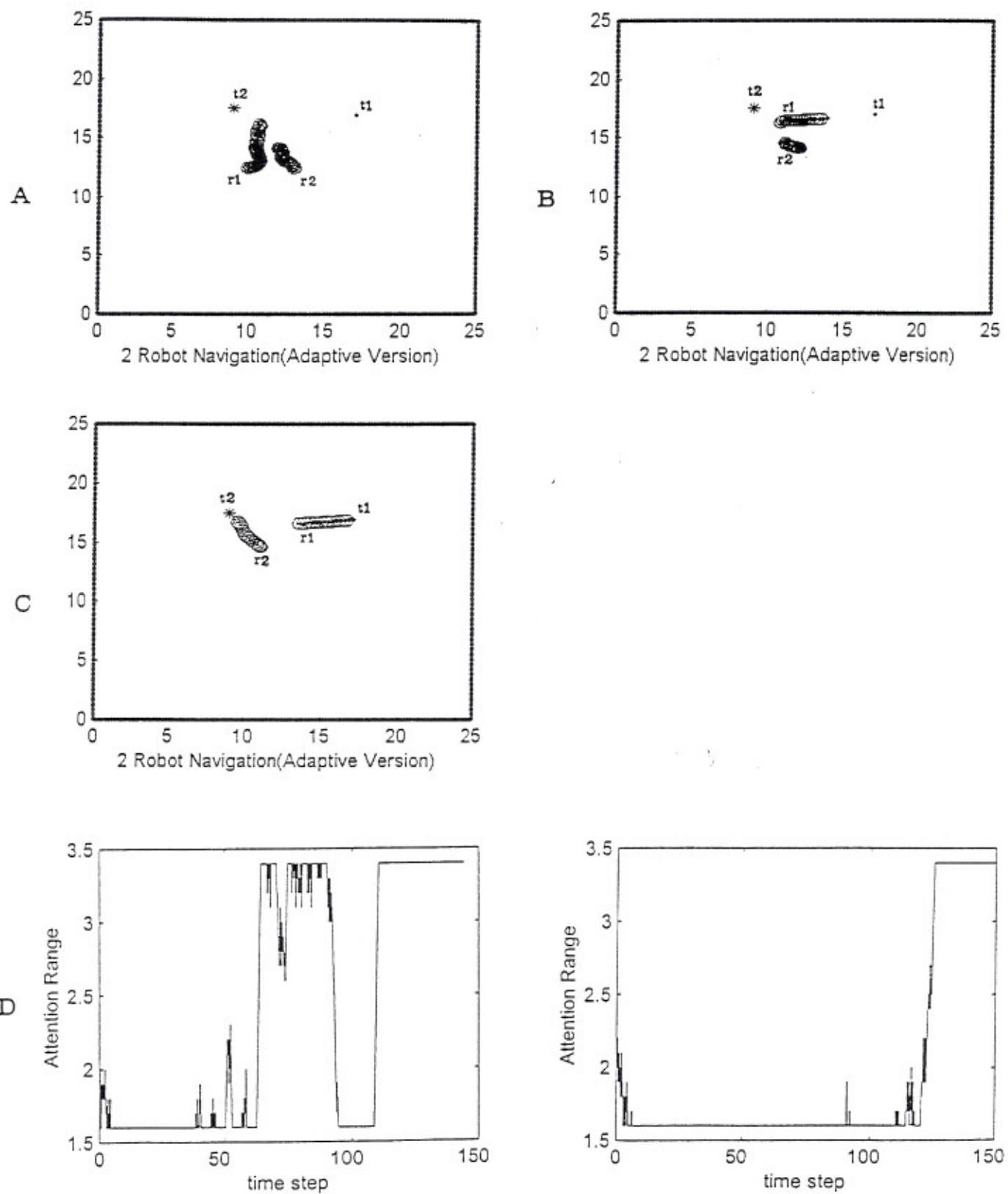
Figure	A-rng	Speed	$D_{average}$	MovingLength	D_m	TimeStep
10	Adaptive	Adaptive	0.13	33.3	0.98	339
11	Adaptive	Adaptive	0.12	32.02	0.79	443

Table) Obtained Results

جدول ۲) نتایج اندازه گیری شده



شکل ۱۲) تابعیت دو ربات متحرک در یک محیط نمونه با استفاده از کنترل کننده فازی تطبیقی



شکل ۱۲) تابعیت دو ربات متحرک در یک محیط ثابت با استفاده از کنترل کننده فازی تطبیقی

۹) نتایج و پیشنهادات

در یک محیط دینامیک که دائماً در حال تغییر است، استفاده از روش های سراسری غیرممکن یا بسیار پرهزینه است، چرا که در این روش ها مدل ریاضی و یا نقشه کامل محیط مورد نیاز می باشد. اما تهیه مدل ریاضی و یا نقشه کامل محیط کاری پیچیده است و با تغییر محیط ناگزیر از به روز نه آوردن مدل ریاضی و یا نقشه آن خواهیم شد. باین ترتیب روش های سراسری برای حل چنین مسائلی غیر عملی می باشند. پس به ناچار باید از روش های محلی استفاده نمود. در روش های محلی اطلاعات خام بصورت محلی در هر لحظه از زمان به ایجاد عملیات موتوری منجر می گردد. تهیه این اطلاعات بر عهده حسگرهای ریات متحرک می باشد.

سیستم های فازی ساده، به راحتی قابل درک و کم هزینه می باشند. این سیستم ها در موارد ساده ای که قوانین فازی حاکم بر محیط را در دست داریم، به تنهایی جهت حل مساله قابل استفاده اند؛ در محیط های پیچیده تر استخراج و تنظیم قوانین فازی، جهت عملکرد مناسب و قابل قبول سیستم لازم می باشند.

در این تحقیق یک کنترل کننده فازی تطبیقی را معرفی نمودیم. در این کنترل کننده توابع عضویت فازی ورودی و خروجی، گاویسی شکل انتخاب شده اند، اما پارامترهای آنها با توجه به شرایط محیط در هر قدم زمانی تعیین و تنظیم میگردد (کنترل کننده تطبیقی). تنظیم پارامترهای توابع عضویت با کمک اتوماتون یادگیر انجام می گیرد.

آزمایشات نشان می دهد که اگر این پارامترها را در کل مسیر حرکت ثابت اختیار کنیم (کنترل کننده غیر تطبیقی)، مسیر حاصل ناهموارتر، طولانی تر، پر خطر تر و یا توان با تأخیر بیشتر خواهد شد. البته می توان مقادیر ثابتی برای این پارامترها یافت، که منجر به پاسخ های قابل قبولی می گردد، اما این پاسخ ها به خوبی حالت تطبیقی نخواهند بود. یافتن این مقادیر نیز کار دشواری می باشد و هزینه محاسباتی و زمان زیادی لازم دارد. اتوماتون یادگیر در تعیین مقادیر مناسبی برای این پارامترها در شرایط مختلف محیط بسیار مطلوب عمل می نماید.

برای یافتن قوانین فازی سیستم در هر قدم زمانی، همانطور که قبل اشاره شد، مستقیماً از

یادگیری تقویتی استفاده می نماییم. الگوریتم مورد استفاده، الگوریتم یادگیری Q می باشد. برای هر زوج حالت - عمل یک مقدار ارزش Q تعیین می گردد که نشان می دهد هر عمل در هر حالت چقدر مناسب یا نا مناسب است. جهت رفتار همکاری از یادگیری Q باتابع ارزش توزع شده استفاده می شود.

الگوریتم مابه دلیل عدم استفاده از رهبردهایی نظری شبکه های عصبی یا الگوریتم های تکاملی که در اغلب تحقیقات مورد استفاده قرار گرفته اند، ساده و روشن میباشد و چون نیازی به مشتق گیری و یا سایر عملیات پیچیده ریاضی ندارد کم هزینه و سریع می باشد. ربات با کمک این کنترل کننده ساده میتواند مسیری نرم، هموار، کوتاه و حتی مناسب تر روش های مذکور را انتخاب کرده و به سمت هدف خود پیش روید. این کنترل کننده ترکیب رفتارهای سادگی، بارو شی قابل درک و با حداقل هزینه ممکن عملی میباشد. به کار گرفتن اتوماتون یادگیری باعث می شود که کنترل کننده خود را با تغییرات محیط تطبیق دهد و در ترتیبی باعوض شدن شرایط محیط عملکرد آن کاهش نمی یابد. نکته مهم این است که تصمیم گیری با اطلاعات بسیار ناچیز در مردم طبیعت محیط صورت می گیرد. ربات بدون داشتن هرگونه پیش فرضی به تنهایی میتواند نحوه تصمیم گیری، انتخاب اجرای عمل مناسب را بیاموزد.

در ادامه این پژوهه می توان تحقیقات متنوع و مختلفی را انجام داد. به عنوان نمونه می توان با تغییر شکل و یا تعداد توابع عضویت فازی ورودی و خروجی، تاثیر این تغییرات در نحوه عملکرد سیستم را بررسی نمود و به روابط و یا تابعیتی در این زمینه دست یافت. می توان بجای سیستم استنتاج فازی مدانی از سیستم های استنتاج دیگری استفاده نمود و تاثیر این کار را بررسی کرد. می توان روش های دفازی سازی مختلف را آزمود. و یا این رفتارها را با تعداد سنسرهای بیشتر یا کمتر بررسی کرد. ما در الگوریتم یادگیری تقویتی خود، برای مشخص کردن پاداش یا تنبیه معیارهایی خاص در نظر گرفتیم و روابطی مشخص قرار دادیم. می توان برای هر رفتار از روابطی متفاوت استفاده کردو الگوریتم را بررسی نمود و در صورت امکان به یک چارچوب ریاضی برای تعیین این روابط دست یافت.

- [V] Dong-Wook Lee, Kwee-Bo Sim,"An evolution of cellular automata neural systems using DNA coding method",*IEEE International Fuzzy Systems Conference Proceedings*, pp. 117_122,vol. 1,1999.
- [A] Cem Ünsal,"Intelligent Navigation of Autonomous Vehicles in an Automated Highway System:Learning Methods and Interacting Vehicles Approach", Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of
the requirements for the degree of Doctor of Philoshopy in Electrical Engineering, Blacksburg, Virginia, January 1, 1997.
<http://www.cs.cmu.edu/~unsal/diss/cemsdiss.html>
- [A] Leslie Pack Kaelbling & Michael L. Littman,"Reinforcement Learning: A Survey",
<http://www.cs.brown.edu/people/lpk/rl-survey/rl-survey.html>
- [A] <http://diwww.epfl.ch/lami/robots/k-amily/khepera.html>

برای رفتار همکاری می توان ظرفات های زیادتری قائل شد. مثلاً بجای ثوبتی بودن توقف یک ربات، جهت عبور ربات دیگر، هر ربات پس از سنجش موقعیت خود و ربات دیگر تصمیم به حرکت یا توقف بگیرد. می توان به تحقیقاتی آماری جهت بررسی مساله پرداخت و به نتایج ارزشمندی دست یافت. یا برای تطبیقی کردن کنترل کننده از روشهای اتوماتون یا دگیر استفاده نمود. یا با تغییر پارامترهای اتوماتون و یا یادگیری تقویتی به نتایجی جدید دست یافت.

مراجع

- [V] Kai-Tai Song, Jen-Chau Tai,"Fuzzy Navigation Of A Mobile Robot", *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*,pp. 621_627,vol 1,1992.
- [T] Hee Rak Beom, Hyung Suck Cho,"A sensor-based navigation for a mobile robot using fuzzylogic and reinforcement learning",*IEEE Transactions on Systems, Man and Cybernetics*,pp. 464_477,vol 25, March 1995.
- [T] Beom, H.R., Koh, K.C., Cho, H.S., "Behavioral control in mobile robot navigation using fuzzy decision making approach".
- [E] *Robotic Systems and the Real IROS '94. Proceedings of the IEEE/RSJ/GI International Conference on intelligent Robots and Systems*,pp.1938_1945,vol.3,1994.
- [E] Yung, N.H.C., Ye, C., "Self-learning fuzzy navigation of mobile vehicle", *3rd International Conference on Signal Processing*,pp. 1465_1468 vol.2,1996.
- [T] Fukuda, T., Kubota, N., "An intelligent robotic system based on a fuzzy approach", *Proceedings of the IEEE*,pp. 1448_1470,vol 87,Sept. 1999.