Q-öğrenme algoritması ile mobil robotların yol planlaması
Path planning of mobile robots with Q-learning

Halil Çetin
Elektrik-Elektronik Mühendisliği Bölümü
Selçuk Üniversitesi, Mühendislik Fakültesi
Konya, Türkiye
halil_c@hotmail.com

Akif Durdu
Elektrik-Elektronik Mühendisliği Bölümü
Selçuk Üniversitesi, Mühendislik Fakültesi
Konya, Türkiye
durdu.1@selcuk.edu.tr

***Özetçe*—Gelişimine hızla devam eden robotik sistemler, günlük hayatımızın içinde daha çok yerini almaktadır. Mobil hareket edebilen robotlar, bulundukları ortamlarda hareket edebilecekleri alanların haritalarını çıkarabilmekte ve çıkardıkları bu haritaların içinde belirlenen hedeflere en kısa yoldan giderek minimum sürede ulaşabilmektedirler. Bu çalışmada mobil robotlar ile elde edildiği varsayılan haritalarda, hedefi belli olan bir noktaya Q-öğrenme algoritması kullanarak en kısa zamanda gidilebilecek yol planlaması yapılmaktadır. Q-öğrenme algoritması, bir takviyeli (reinforcement) öğrenme türü olup bulunduğu ortamı algılayan ve kendi başına kararlar alabilen bir sistemin, hedefine ulaşabilmesinde doğru kararlar almayı nasıl öğrenebileceğini gösterir. Bir mobil robotun, birkaç örnek haritada ki hedeflerin farklı noktalarda konumlandırıldığı durumlarda, doğru yolları Q-öğrenme algoritması çalıştırılarak bulabildiği gösterilmiştir.**

Anahtar Kelimeler — Q-öğrenme; eş zamanlı konumlama ve haritalama; yol belirleme.

***Abstract*—Robotic systems which rapidly continue its development are increasingly used in our daily life. Mobile robots both draw the map where they may move in their environment and reach to the determined target in the shortest time by going on the shortest way in their prepared map. In this paper, Q-learning-based path planning algorithm is presented to find a target in the maps which are obtained by mobile robots. Q-learning is a kind of reinforcement learning algorithm that detects its environment and shows a system which makes decisions itself that how it can learn to make true decisions about reaching its target. The fact that a mobile robot truly finds targets that are located on different points in a few sample maps by processing our proposed Q-learning-based path planning algorithm is shown at the end of the paper.**

Keywords — Q-learning; simultaneously localization and mapping; path planning.

I. GİRİŞ

Günümüzde, haritası veya ortamı bilinen mobil robot uygulamalarında, belirli bir hedefe ulaşabilmek için yol planlaması yapmak ve robotları bu yolda hareket ettirmek oldukça önemli bir konudur. Özellikle savunma sanayisinde, insan hayatının tehlikede olduğu ortamların keşif edilmesinde robot teknolojisi etkili bir şekilde kullanılarak minimum insan kaybı olmasına çalışılmaktadır. Eş zamanlı konumlama ve haritalama yöntemiyle keşfi yapılan tehlikesi yüksek bir ortamda bir bombanın (veya bir hedefin) imha edilmesi, yine robotlar aracığı ile en kısa yolun belirlenmesi sayesinde gerçekleştirilebilir. Yol planlaması ile ilgili verilebilecek güzel örneklerden bazıları, robot süpürgeler veya robot refakatçilerdir. Robotların günlük hayatımızda daha yaygın kullanılması, çalıştırılan algoritmaların hızlı olmasına ve gerçek zamanlı olarak uygulanabilmesine bağlıdır.

R. Valencia ve arkadaşları optimum navigasyon stratejisini tasarlayan bir metot sunmaya çalışmışlardır [1]. Standart özellik tabanlı eş zamanlı konumlama ve haritalama işlemlerinden sonuç elde eden olasılıklı inanç ağlarının direkt olarak yörüngeleri planlanamadığı için bunun yerine pozisyon tabanlı Eş Zamanlı Konumlama ve Haritalama grafiklerini kullanmayı uygun görmüşlerdir. Önerilen yöntemin inanç uzayında tanımlanmış olması ve harita referans noktasından bağımsız olarak, yalnızca pozisyonlar arasındaki bağıl bilgiyi kodlaması avantaj olarak verilmiştir. Ancak harita üzerinde hedef nokta belirtilmediğinde robotun bu noktayı arayıp bulması problemi çözülmemiş ve sonraki çalışmalarda çözümleneceği belirtilmiştir.

E. Guevara ve arkadaşları, eş zamanlı konumlama ve haritalama için tekrarlayan bir yapıya sahip Yapay Sinir Ağlarını (YSA) ve yol bulma problemi için Q-öğrenme algoritmalarını kullanılmışlardır [2]. Burada harita önceden verilmediği için bu iki yapının birleşiminde olasılık hesaplamaları kullanılmıştır. Böylece simülasyon ortamındaki bir mobil robotun, bilinmeyen bir ortamda hem ortamı keşfedip hem de yol belirlemesi simüle edilerek güçlü bir navigasyon sistemi geliştirilmeye çalışılmıştır.

978-1-4799-4874-1/14/$31.00 ©2014 IEEE

C. Li ve arkadaşları, Q-Öğrenme ile YSA yapılarını birleştirerek daha doğru ve daha hızlı sonuç elde edilmeye çalıştılar. Bu iki algoritma kullanılarak algılama alanı ile hareket alanı arasındaki harita ilişkisi daha doğru kurulabilmektedir. Çok boyutlu ileri beslemeli YSA sayesinde yol belirleme denetleyicisi oluşturulmuş ve yol belirleme görevi beş sınıflı bir sınıflandırma problemine indirgenmiştir. Q-Öğrenme, YSA için eğitim örneklerini toplamada kullanılmıştır. Böylece YSA'nın yavaş öğrenme dezavantajı giderilmeye çalışılmıştır. Bir başka çalışmada ise, Q-öğrenme algoritmasının gereksiz işlemler yaparak zaman kaybına sebep olmasını engellemek amacıyla, algoritma tabanında küçük değişikliklere gidilmiştir [3]. Örneğin bir hücrenin Q değerinin her işlem döngüsünde değişmesini önlemek için mantıksal bir değişkenle kontrol edilmesi sağlamış ve bu değişken izin verdiği sürece döngüdeki işlemlere tabi tutulması, aksi takdirde döngü içerisindeki işlemlere dâhil edilmemesi sağlanmıştır. Bu sayede işlem süresi azaltılarak daha verimli bir çalışma yapılmıştır.

Eş zamanlı konumlama ve haritalama konusunda yapılan çalışmaların çoğunluğunda YSA ve/veya Q-öğrenme algoritmalarının kullanıldığı, bu algoritmalar değişse bile konumlama ve haritalama tanımlarının değişmediği, özellikle yol belirleme/planlama sürecinin önemli bir yer tuttuğu görülmektedir. Q-öğrenme algoritması, bir takviyeli (reinforcement) öğrenme metodu olup bulunduğu ortamı algılayan ve kendi başına kararlar alabilen bir sistemin, hedefine ulaşabilmesinde doğru kararlar almayı kendi kendine nasıl öğrenebileceğini gösterir. Bu çalışmada mobil robotlar ile elde edildiği varsayılan haritalarda, hedefi belli olan bir noktaya Q-öğrenme algoritması kullanarak en kısa zamanda gidilebilecek yol planlaması yapılmaktadır.

Bu bildirinin ilerleyen kısımlarında öncelikli olarak uygulanan yöntemden bahsedilecek, daha sonra -önerilen yöntem bölümünde- bizim çalışmamızda Q-öğrenme yönteminin nasıl çalıştırıldığı anlatılacaktır. Dördüncü ve beşinci kısımlarda yol belirlemenin, uygulama örnekleriyle birlikte nasıl yapıldığı verilecektir. Son bölümde ise elde edilen sonuçların tartışılması ve gelecekte yapılacak iş planlaması açıklanacaktır.

II. YÖNTEM

A. Eş zamanlı Konumlama Ve Haritalama

Uygun seçilmiş yapay zekâ uygulamalarıyla donatılmış robotlar, yeterli bilgiye sahip olduklarında konumlama veya haritalama yapabilirler. Konumlama; bir harita üzerinde belirli bir noktanın koordinatlarının belirlenmesi ya da konum bilgisi olarak kullanılan başka bir yöntem yardımıyla noktanın yerinin belirlenmesi olarak tanımlanabilir. Haritalama; bir noktadan başlanarak, belirli bir alanın sonuna kadar o alanda bulunan nesne ve yüzeylerin iki boyutlu veya üç boyutlu olarak şekillendirilmesi, yerlerinin belirlenmesi şeklinde tanımlanabilir.

Hata oranı yok sayılabilecek kadar düşük sensörlere sahip hareketli bir robot, ortamda kendi konumunu bildiği zaman, sensörleriyle ortam hakkında bilgi toplayıp ortamın haritasını oluşturabilir. Benzer şekilde ortamın haritasını bilen bir robot, ortamdaki konumunu sensörleri yardımıyla tespit edebilir. Yani yapay zekâ programları, ortamın haritasını çıkarmak için konum bilgisine, konum bilgisini çıkarmak için ortamın harita bilgisine ihtiyaç duyarlar [4]. Ancak çoğu zaman hatasız bir konum bilgisine ya da hatasız bir harita bilgisine erişmek mümkün olmayabilir. Bu nedenle program belirli aralıklarla ya da belirli kriterler meydana geldiğinde anlık olarak konumlama ya da haritalama yapabilmelidir. Bu işleme Eş Zamanlı Konumlama ve Haritalama denir.

Eş Zamanlı Konumlama ve Haritalama, tam ve çevrimiçi eş zamanlı konumlama ve haritalama olarak ikiye ayrılır. Tam eş zamanlı konumlama ve haritalamada belirli bir süre hareket eden robotun hareketi boyunca bulunduğu konumlar ile harita tespit edilirken çevrimiçi eş zamanlı konumlama ve haritalamada ise robotun son konumu ve harita tespit edilir [4].

B. Q-Öğrenme

Takviyeli öğrenme, deneme yanılma üzerine kurulu bir öğrenme tekniğidir. Bağımsız mobil robotların performanslarına uygun değerlendirmeler sağlayarak beklenmeyen sonuçların öğrenilmesinde kullanılabildiği gibi robotların çevrimiçi öğrenmelerini gerçekleştirmede de kullanılabilmektedir. Çevrimiçi öğrenme ve adaptasyon, gereken deneyim için yeterli miktarda örneği toplamada her robotun öğrenme algoritması işlemleri için arzu edilen özelliklerdir. Karmaşık bir ortamda, gerçek ortamla etkileşim yoluyla çevrimiçi öğrenme gereklidir. Bu çalışmada, yol planlama davranışı üzerine durulmaktadır. Q-öğrenme, sadeliği ve çevrimiçi öğrenmeye müsait olması nedeniyle robotik uygulamalarda sıkça tercih edilen takviyeli öğrenme algoritmalarından biridir [3].

Q-öğrenmede, bir ajanın tüm olası durumları ve verilen durumda tüm olası hareketleri deterministtik olarak bilinir. Diğer bir deyişle, verilen bir A ajanı için, $S_0,S_1,S_2,\ldots,S_n$ olmak üzere n adet olası durum, her bir durumda $a_0,a_1,a_2,\ldots,a_m$ olmak üzere m adet olası hareket vardır. Belirli bir durum-hareket çiftinde, ajanın aldığı bu çifte özel ödül, anlık ödül olarak isimlendirilir. Örneğin $r(S_i, a_j)$, S_i durumunda bir a_j hareketi yaparak A ajanının aldığı ödülü belirtir. Bir ajan, bulunduğu durumdan/konumdan bir sonrakini seçme işlemini belirli bir kuralı takip ederek yapar. Bu kural, bir sonraki durumundan, durumların sonraki geçişlerinden ajanın alabileceği toplam ödülleri maksimuma çıkartmaya çalışır. Örneğin S_i durumunda bulunan bir ajan sonraki en iyi durumu seçmeyi bekliyor olsun. a_j hareketi nedeniyle S_i durumunun Q-değeri (1) eşitliğinde verilmiştir [5].

$$Q(S_i, a_j) = r(S_i, a_j) + \gamma \max_{a'} Q(\delta(S_i, a_j), a') \quad (1)$$

Burada γ, öğrenme katsayısıdır. Genelde 0 ile 1 arasında seçilen bu değer, 0 olursa hiçbir öğrenmenin yapılamayacağı veya Q-değerlerinin ödül dışında bir değer alamayacağı görülmektedir.

$\delta(S_i, a_j)$, S_i durumunda a_j hareket seçimi nedeniyle elde edilen sonraki durumu belirtmektedir. Sonraki durum S_k olsun. $Q(\delta(S_i,a_j), a') = Q(S_k, a')$. Sonuç olarak $Q(S_i, a_j)$'yi maksimize eden (a')'nin seçimi ilginç bir problemdir. Yukarıdaki Q-öğrenme için temel engel, bir S_k durumundaki tüm olası (a') hareketleri için Q-değerlerini bilmektir. Her seferinde, sonraki durumlardan en uygununu belirlemede, belli bir durumdaki tüm olası hareketlerin Q-değerlerini almak için bellekten veri çekmesi gerekmektedir. Bu yüzden sonraki durumu seçmek için fazla zaman harcanmaktadır [5].

Q-öğrenme algoritmasının durdurma kriteri, Q-durum tablosundaki değerlerin değişim miktarlarının belirli bir değerin altında kalmasıdır. Değişim miktarının azalması, tablonun kararlı hale gelmesi anlamına gelir ve bu tablo baz alınarak yol planlama yapılabilmektedir. Bu çalışmada durdurma kriteri olarak iterasyon sayısı ele alınmıştır ve belirtilen iterasyon miktarlarının ne kadar sürede sonuca ulaştıklarını incelenmiştir. Başlangıç noktasından hedefe ulaşmak için yapılan işlemlerin tamamı, bir iterasyon olarak ele alınmıştır.

III. YOL PLANLAMA

Yol belirleme, bir ortam üzerinde, belirli bir noktadan hedef noktasına ulaşmak için, iki nokta arasındaki uygun yolları bulmaya denir. Bu yollar içinden problemin çözümüne uygun olarak en kısa ya da en uzun yolun seçilmesi, en az manevraya sahip ya da en az karmaşık ortamlardan geçen yolun seçilmesi, tam tersi olarak en çok dönüşe sahip ve karmaşık yolların seçilmesi tercihe bağlıdır.

Q-öğrenme algoritmasının işletilmesinden elde edilen Q-durum tablosu(matrisi), robot için yol belirlemede ve karar vermede kullanacağı koşullu harita gibidir. R-ödül\ceza tablosu algoritma işletilirken kullanılacak ve temel olarak referans alınacak tablo, yani diğer bir değişle harita iken; yol belirleme aşamasında tüm işlemleri bitirmiş robot işlemcisi artık harita üzerinde karar verme mekanizmasını, Q-durum tablosunu kullanarak çalıştırır.

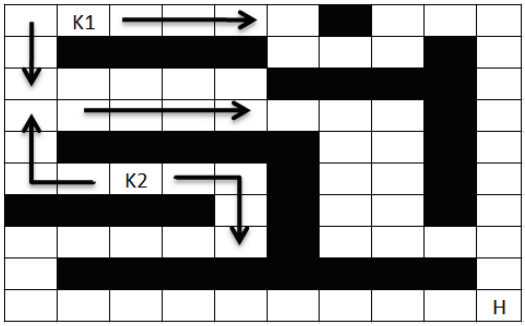
Q tablosu üzerindeki değerleri belirleyen algoritma, robot hangi noktada olursa olsun hedefe ulaşılabilecek en uzak noktaya en düşük değerin gelmesini sağlar. Belirli bir iterasyon sonunda durum tablosunda ödülün bulunduğu noktaya göre tüm değerler belirlenir ve çok değişmeyen, neredeyse sabit bir hal alır.

Robot hangi noktadan başlarsa başlasın, etrafında ceza değerine sahip olmayan, yani gitmekte serbest olduğu tüm durumların bir listesini alır, bu liste içerisinde en büyük Q değerine sahip olan durumu seçer. Eğer iki veya daha fazla durum aynı değere sahipse, bunlardan herhangi birini seçer. Çünkü aynı değere sahip olan durumların hedefe\ödüle olan uzaklığı aynıdır.

Bu mantıkla Q-öğrenme algoritması, serbest bir yüzeyde yol belirleme için kullanılabileceği gibi, labirent gibi çok engelli bir ortamda da kullanılabilir. Ayrıca ortama bağlı olarak hedefe ulaşan yolu birden fazla olan labirentler içinde en kısa yolu seçme olanağı vardır.

IV. DENEY VE SİMÜLASYON

Gerçek hayata benzetilmeye çalışılarak simüle edilen R-ödül/ceza matrisleri aşağıdaki şekillerde haritalandırılmışlardır. Deney esnasında ceza değeri alan siyah renkli bölgelerin iterasyona dahil edilmemesi sağlanmıştır.



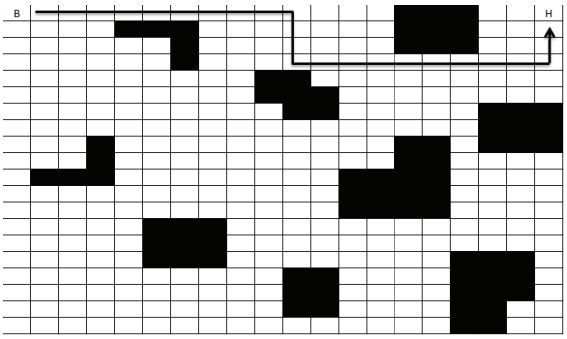
Şekil 1. 1 nolu harita (R-ödül/ceza matrisi)

Şekil 1'de görülen 1 nolu *R* matrisi üzerinde *H* hedef noktayı temsil etmektedir. 0,2 ve 0,5 olan öğrenme katsayılarının süreyi çok fazla etkilemedikleri gözlenmiştir. Buna rağmen iterasyon miktarlarının süreyi oldukça etkiledikleri görülmüştür. 0,2 öğrenme katsayısına sahip sistem 25 iterasyonda, 0,5 öğrenme katsayısına sahip sistemin ise 22 iterasyonda kararlı bir Q matrisine sahip olmuşlardır. Hedefe 3 yoldan ulaşılabilecek olan bu labirent üzerinde 2 karar noktası meydana gelmiştir. *K1* ve *K2* karar noktalarının sağında veya solunda bir başlangıç noktası seçilirse, oklarla gösterilen yolları en kısa yol olarak kabul edecektir.

Tablo 1. 1 nolu harita için sonuçlar

Deney No	Harita (R-matrisi)	Matris Boyutu	Öğrenme Katsayısı	İterasyon Miktarı	İşlem süresi(sn)
1	1	10x10	0,2	10	0,275904
2	1	10x10	0,2	20	0,588885
3	1	10x10	0,2	25	0,928174
4	1	10x10	0,5	10	0,273844
5	1	10x10	0,5	20	0,584969
6	1	10x10	0,5	25	0,934409

Şekil 2'deki 2 nolu *R* matrisi üzerinde *H* hedef noktayı, *B* başlangıç noktasını temsil etmektedir. Tüm yapılan çalışmalar en kısa yolun okla belirtildiği şekilde olduğunu göstermiştir. Boyutların daha büyük olması işlem süresini oldukça arttırmaktadır. Böylece işlem süresinin iterasyon miktarı ve boyutlara bağlı olduğu kanısına varılmıştır. Yapılan çalışmaların sonuçları Tablo 2'de görülmektedir.



Şekil 2. 2 nolu harita (R-ödül/ceza matrisi)

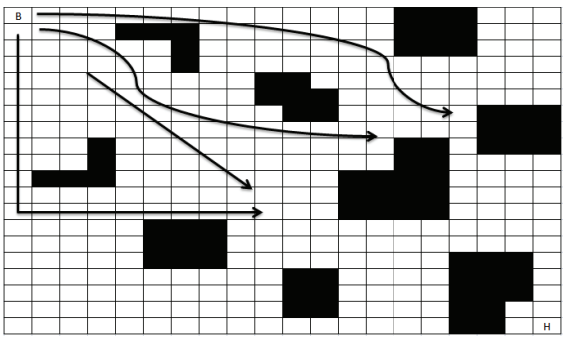
Tablo 2. 2 nolu harita için sonuçlar

Deney No	Harita (R-matrisi)	Matris Boyutu	Öğrenme Katsayısı	İterasyon Miktarı	İşlem süresi(sn)
7	2	20x20	0,2	10	2,792906
8	2	20x20	0,2	20	5,191292
9	2	20x20	0,2	25	6,485019
10	2	20x20	0,5	10	2,836962
11	2	20x20	0,5	20	5,216200
12	2	20x20	0,5	25	6,537672

Başlangıç noktası olarak B noktası seçildiği taktirde yalnızca tek bir yörünge üzerinde en kısa yol bulunmaktadır. 0,2 öğrenme katsayısına sahip sistem 16 iterasyonda, 0,5 öğrenme katsayısına sahip sistem 25 iterasyonda kararlı bir Q matrisine sahip olmuşlardır.

Şekil 3'deki R matrisinde farklı olarak çok olasılıklı durumların meydana getirdiği karmaşık durum gösterilmiştir. Bu haritaya ait Q-durum matrisinde çok sayıda kısa yol olasılığı, yani çok sayıda yörünge bulunmaktadır. Bu karmaşık durum, sadece hedef noktanın yer değiştirmesiyle elde edilmiştir. Buna bağlı olarak iterasyon süreleri artmıştır. Bu değerler Tablo 3'de gösterilmiştir.

Karmaşık bir haritanın işlem sürelerini arttırması yanında, öğrenme katsayısının Q-durum matrisi üzerindeki etkisini de görülmektedir. 1 ve 2 nolu çalışmalarda öğrenme katsayısından çok etkilenmeyen Q matrisi, 3 nolu çalışmada 0,7 öğrenme katsayısı ile 43 iterasyonda kararlı hale gelebilmiştir.



Şekil 3. 3 nolu harita (R-ödül/ceza matrisi)

Tablo 3. 3 nolu harita için sonuçlar

Deney No	Harita (R-matrisi)	Matris Boyutu	Öğrenme Katsayısı	İterasyon Miktarı	İşlem süresi(sn)
13	3	20x20	0,4	10	6,085088
14	3	20x20	0,4	20	10,921597
15	3	20x20	0,4	25	14,391622
16	3	20x20	0,7	10	5,874074
17	3	20x20	0,7	20	10,921419
18	3	20x20	0,7	25	13,622859
19	3	20x20	0,7	43	20,435169

V. SONUÇ

Yol planlama için Q-öğrenme algoritmasının uygunluğu incelenmiş, daha sonraki eş zamanlı konumlama ve haritalama çalışmaları için bir temel oluşturması amaçlanmıştır. Takviyeli öğrenme algoritmalarından biri olan Q-öğrenme ile 3 çalışma yapılmış, olası gerçek durumlara uygun ortamlar için sonuçlar elde edilmiştir. Tüm sonuçlar bilgisayar ortamında simüle edilmiştir.

Bu algoritmada sonuç elde etme süresinin, ortamın ödül/ceza matrisine uyarlanmış bir haritasının boyutlarına ve öğrenme katsayısına bağlı olduğu gözlenmiştir. Ayrıca iterasyon miktarlarının süreyi ve sonuçları nasıl etkilediği incelenmiştir. Genelde Q matrisinin kararlı bir yapıya sahip olması beklenmeden, iterasyon miktarı küçük tutularak, doğru sonuç elde edilebilmiştir. Ancak çok olasılıklı ortam haritalarında kararlı bir durum matrisi elde etmek, doğru sonuç elde etmek için önemlidir. Her bir iterasyon, hedefin bulunması için yapılmış tüm işlemleri içermektedir.

Yol planlamada takip edilecek yörüngeye karar vermek önemlidir. 3 nolu haritada görülebileceği gibi, bir robotun hangi yörüngeyi tercih etmesi gerektiğine karar vermek için, robotun kabiliyetlerini belirlemek ve bu kabiliyetlere uygun seçim yapmak gerekmektedir. Örneğin manevra kabiliyeti düşük olan bir robot için en az dönüşe sahip bir yörünge belirlemek esastır. Bu amaçla yapılacak bir inceleme sonraki çalışmalara bırakılmıştır.

KAYNAKÇA

[1] Valencia, R., Andrade-Cetto, J., and Josep M. PortaPath, J. M., "Planning in Belief Space with Pose SLAM", IEEE International Conference on Robotics and Automation (ICRA), 78-83, 2011.

[2] Guevara-Reyes, E., Alanis, A. Y., Arana-Daniel, N., Lopez-Franco, C., "Integration of an Inverse Optimal Control System with Reinforced-SLAM for path planing and mapping in dynamic environments", IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), 2013.

[3] Li, C., Zhang, J., Li, Y., "Application of Artificial Neural Network Based on Q-learning for Mobile Robot Path Planing", IEEE International Conference on Information Acquisition, 978-982, 2006

[4] Aydogdu, M. F., "Fpga-Tabanlı, Steryo Görmeye Sahip Bir Robotta Üç Boyutta Eş Zamanlı Konumlama Ve Haritalama", M.S. thesis, Institute of Natural and Applied Sciences, TOBB Economics and Technology University, Ankara, 2010.

[5] Indrani Goswami, I., Das, P. K., Konar, A., and Janarthanan, R., "Conditional Q-learning Algorithm for Path-Planing of a Mobile Robot", International Conference on Industrial Electronics, Control and Robotics (IECR), 23-27, 2010.