Proceedings of KIIS Spring Conference 2008 Vol. 18, No. 1.

SVM과 다각형 기반의 Q-learning 알고리즘을 이용한 군집로봇의 목표물 추적 알고리즘

Object tracking algorithm of Swarm Robot System for using SVM and Polygon based Q-learning

서상욱, 양현창, 심귀보
Sang-Wook Seo, Hyun-Chang Yang, and Kwee-Bo Sim

중앙대학교 전자전기공학부
(E-mail: kbsim@cau.ac.kr)

요 약

본 논문에서는 군집로봇시스템에서 목표물 추적을 위하여 SVM을 이용한 12각형 기반의 Q-learning 알고리즘을 제안한다. 제안한 알고리즘의 유효성을 보이기 위해 본 논문에서는 여러 대의 로봇과 장애물 그리고 하나의 목표물로 정하고, 각각의 로봇이 숨겨진 목표물을 찾아내는 실험을 가정하여 무작위, DBAM과 ABAM의 융합 모델, 그리고 마지막으로는 본 논문에서 제안한 SVM과 12각형 기반의 Q-learning 알고리즘을 이용하여 실험을 수행하고, 이 3가지 방법을 비교하여 본 논문의 유효성을 검증하였다.

Key Words : SVM, Dodecagon-based Q-learning, DBAM, ABAM

1. 서 론

인간 사회의 모든 환경들은 인간이 사용하기에 가장 쉽고 편리하게 느낄 수 있도록 설계되어 가고 있다. 편안한 인간의 삶을 누리기 위하여 사회의 많은 부분에서 로봇이 이용되어져 가고 그 활용 범위도 점점 넓어져 가고 있다. 근래에는 로봇을 이용한 사회 안전 분야에도 많은 연구가 진행되고 있는데, 로봇을 이용한 경비, 탐색 등은 기존의 카메라나 적외선 센서와 같은 보안 시스템과 연동을 통해 이루어지고 있다. 이런 로봇 시스템의 사회적 인프라가 구축되어 가면서 기존의 한 대의 로봇을 가지고는 많은 업무를 수행할 수 없기 때문에 군집 로봇의 필요성이 증가되어져 가고 있다. 이런 군집 로봇 시스템을 제어하기 위해서 과거에는 중앙 집중식 제어를 많이 사용하고 있는데, 중앙 집중식 제어는 중앙에서 필요한 임무에 모든 부분을 통제 할 수 있기 때문에 빠르고 정확한 제어가 가능하다는 장점을 가지고 있다. 그러나 제어해야 할 로봇 제어 시스템들이 거대화되고 복잡해짐에 따라서 로봇 제어 시스템의 유연성과 강인함이 점점 중요시 되어가고 있다. Parker는 다수 로봇의 작업 수행을 위해 heuristic 형태의 알고리즘을 제안하였다[1]. Ogasawara는 다수의 로봇을 이용해 커다란 물체를 수송하기위해 자율 분산 로봇 제어 방식을 이용하였다[2]. 본 논문에서는 다수의 로봇이 어떤 작업을 수행함에 있어 서로간의 충돌을 피하고, 자신만의 고유한 영역을 탐색하도록 하기위한 방법으로 distance-based action making and area-based action making process의 융합 모델을 제안한다.

강화 학습은 agent로 하여금 주변 환경의 탐색을 통해 능동적으로 환경에 대한 행동을 결정하도록 한다. 보상 값이 존재하는 어떤 불확실한 영역을 탐색하는 동안 agent는 연속적인 상태 공간을 따라 적절한 보상 값을 전달함으로써, 임의의 상태에 대해 어떠한 행동을 취해야 할지를 학습하게 된다[3]. 강화 학습을 구현하기위한 많은 방법 중, 본 논문에서는 SVM을 바탕으로 한 Q-learning을 이용하였다. 그 이유는 Q-learning은 불완전한 정보를 가진 Markovian 공간에서의 행동 결정에 대해, 어떤 상태와 행동으로 이루어진 Q-함수를 기본으로 하여 문제의 해결에 쉬운 방법을 제공하

감사의 글 : 본 연구는 지식경제부의 2008년도 성장동력기술개발사업인 「집단 로봇 기술을 이용한 사회안전로봇 개발(세부과제: 로봇통제 및 환경기술개발) 에 의해 수행되었습니다. 연구비 지원에 감사드립니다.

기 때문이다[4]. 본 논문에서는 distance-based action making and area-based action making process의 융합 모델을 강화하기 위해 dodecagon based Q-learning과 SVM 알고리즘을 적용한다.

본 논문의 2장에서는 DBAM과 ABAM의 융합 모델에 대해 나타내고, 3장에서는 SVM과 다각형 기반의 Q-learning 알고리즘에 대해 논한다. 4장에서는 위의 3가지 제어 알고리즘들을 적용한 목표물 탐색의 시뮬레이션 결과를 보이고 마지막으로 5장에서는 결론 및 향후 과제에 대해 논한다.

2. 로봇의 행동 결정 과정

2.1 DBAM and ABAM

Distance-based action making(DBAM) 과 Area-based action making(ABAM) 방법은 로봇이 다음 행동을 결정하는데 있어서 사용되는 방법이다. DBAM 방법은 로봇이 주위의 환경을 거리로 인식하는 방법으로 로봇과 물체 사이의 거리에 의해서 행동을 결정한다. 반면 ABAM 방법은 로봇이 자신 주변의 환경을 돌 사이의 거리가 아닌 자신 주변의 면적을 계산하여 얻어진 정보로부터 다음 행동을 결정하는 방법이다. ABAM 방법의 핵심은 로봇으로 하여금 자신 주위의 불확실성을 줄여 나가도록 한다는데 있는데, 결국 ABAM 방법은 행동 기반 방향 전환(behavior-based direction change) 방식과 많은 유사점을 가지고 있다고 할 수 있다[5][6]. 그림 1은 DBAM과 ABAM 방법이 행동을 선택하는 기준을 나타내고 그림 2는 같은 환경 아래 로봇이 존재한다고 가정하였을 때 DBAM과 ABAM 방법이 어떻게 다음 행동을 선택하는지에 대한 방법을 예를 들어 설명하고 있다. 그림 2에서 보면 같은 상황이라도 행동 결정 방식에 따라서 많은 차이점이 나타나는데, 우선 DBAM 방법의 경우에는 로봇이 취할 수 있는 행동 자체가 많지만, ABAM 방법의 경우에는 좀 더 제한적인 행동을 함으로써 좀 더 불확실한 행동을 조기에 차단할 수 있다는 장점이 있다.

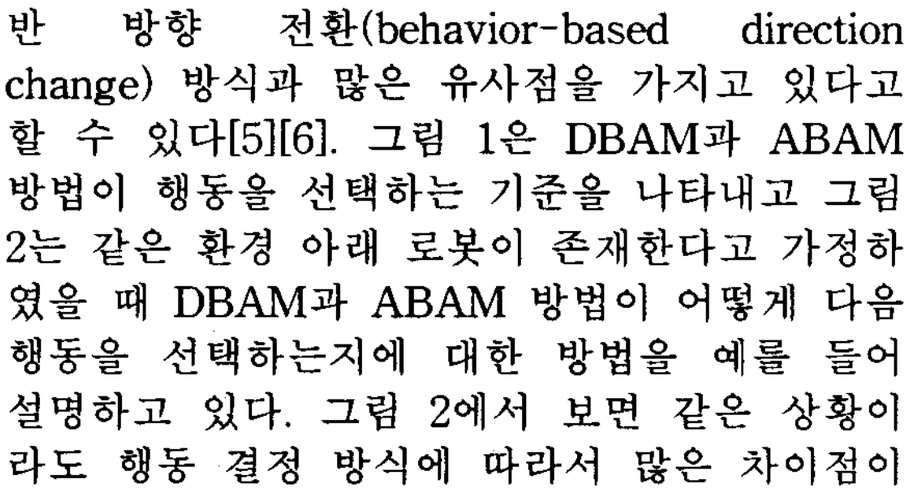

그림 1. DBAM과 ABAM의 행동 선택 기준

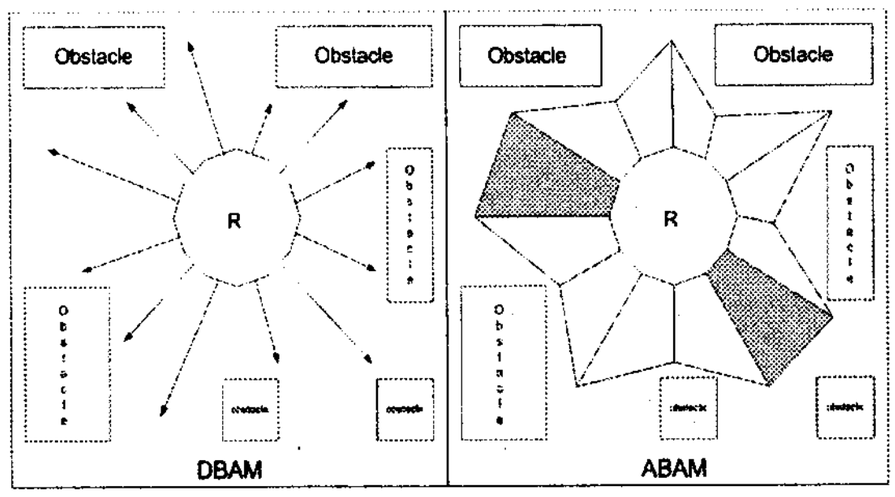

그림 2. DBAM과 ABAM의 행동 선택의 예

2.2 DBAM and ABAM의 융합모델

본 논문에서는 로봇이 다음 행동을 결정하는데 있어서 DBAM과 ABAM 방법을 융합한 모델을 사용한다. DBAM 방법을 통해서 로봇으로부터 가장 거리가 먼 방향을 선택하고, ABAM 방법으로 주위 환경에서 가장 넓이가 큰 공간을 선택하게 된다.

단순히 거리만으로 다음 행동을 결정하는 DBAM 방법은 계산 량을 줄일 수 있다는 장점이 있지만, 올바르지 못한 행동을 선택할 확률이 높은 단점이 있다. 반면 행동을 선택할 때 단순히 넓이만을 고려하는 ABAM 방법은 올바른 행동을 선택할 확률이 있다는 장점이 있으나, 계산양이 많다는 단점이 있다. 그림 3은 DBAM과 ABAM 방법의 융합 모델에 관한 그림이다.

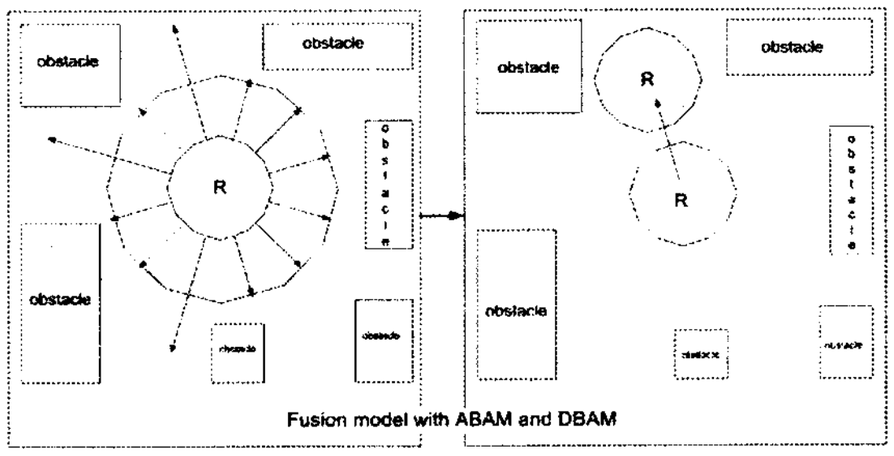

그림 3. DBAM과 ABAM의 융합 모델

3. SVM과 다각형 기반의 Q-learning 알고리즘

3.1 Q-learning

Q-learning은 강화학습으로 잘 알려진 알고리즘이다. 그리고 로봇이 효과적인 행동을 하기 위해서 보상의 개념을 이용해서 최적의 제어를 얻을 수 있다[7]. 여기서 보상은 행동 후 보상을 하게 된다. Q-learning 알고리즘은 표 1에서 설명하는 것과 같다. 여기서 s는 상태를, a는 행동을, r은 보상 값을, γ는 Q-함수 값의 조정을 위한 계수(discount factor)이다.

표 1. 모델링을 위한 초기 파라미터들

$\hat{Q}(s,a)$ 테이블의 모든 상태와 행동 s , a를 0으로 초기화 시킨다.

다음의 과정을 계속해서 반복한다.

●현재 상태 s를 인식 한다.

●현재 상태 s에 대하여 행동 a를 선택하고 행동 a를 수행한다.

●행동 a에 대해서 즉각적인 보상 값 r을 받는다.

●새로운 상태 s'을 인식한다.

●새로운 상태 s'에 대하여 행동 a'를 선택하고 행동 a'를 수행한다.

●$\hat{Q}(s,a)$ 테이블의 값들을 $s \leftarrow s'$, $a \leftarrow a'$ 로 계속해서 업데이트 시킨다.

갱신될 Q 값은 다음 식에 따라서 갱신된다.

$$\hat{Q}(s,\ a) \leftarrow r+\gamma \max_{a_2} \hat{Q}(s',\ a') \qquad (1)$$

그림 4는 Q-learning에 대한 실제적인 예를 보여준다. 각각의 정사각형은 상태를 나타낸다. R은 로봇을 나타낸다. 상태의 천이에 따른 화살표 위에 나타난 값은 그 행동을 취함에 따른 Q값을 나타낸다. 예를 들어, 초기 상태에서 오른쪽으로 상태를 천이하는데 따른 Q값은 화살표 위의 값인 $\hat{Q}(s_1, a_{right})=72$ 와 같다.

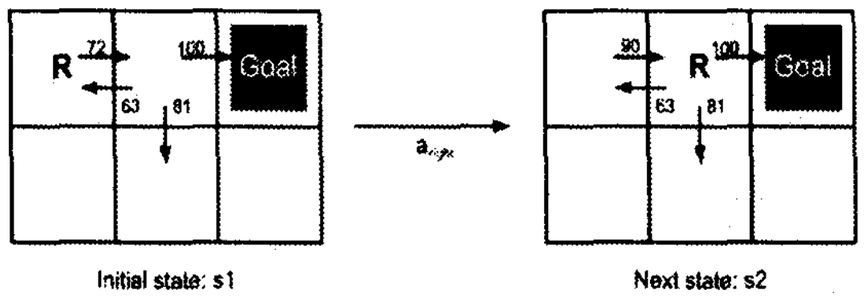

그림 4. Q-Learning의 예

초기 상태에서 만약 로봇이 오른쪽으로 행동을 취한다면, 업데이트 되는 Q값은 $r=0$, $\gamma=0$ 을 초기 값으로 할 때

$$\begin{aligned}\hat{Q}(s_1,\ a_{right}) &\leftarrow r+\gamma \max_{a_2} \hat{Q}(s_2,\ a_2)\\ &\leftarrow 0+0.9\max_{a_2}\{63,\ 81,\ 100\} \quad (2)\\ &\leftarrow 90.\end{aligned}$$

이 된다.

3.2 SVM과 12각형 기반 Q-learning 알고리즘

SVM을 이용한 Dodecagon 기반 Q-learning 알고리즘은 12개의 초음파센서를 이용하여 12방향으로 물체를 측정한다. 그리고 로봇 주위의 넓이를 12등분하여 인식한 후 각 넓이에서 물체의 넓이를 뺀 나머지 부분의 넓이와 물체와 로봇사이의 거리가 가장 긴 방향으로 행동을 취한다.

본 알고리즘은 기존의 Q-learning 알고리즘과 가장 큰 차이는 업데이트 방법이다. 기존의 Q-learning 알고리즘에서는 이전의 Q 값과 비교를 해서 최소화시킬 수 있는 방향으로 이전의 상태를 좋은 방향으로 학습해 나가는 방법이다. 하지만 본 연구에서 제안된 SVM을 이용한 SVM과 12각형 기반의 Q-learning 알고리즘은 이전의 Q 값에 의존하지 않고 그 상황에서 최적의 행동을 결정하게 된다. Q값의 갱신은 다음의 (2)식으로 이루어진다.

$$\begin{aligned}\hat{Q}(s,a) &\leftarrow r+\gamma \max_{s'} \hat{Q}(s',a)\\ &\leftarrow r+\gamma \max_{s'}\{(S_{area1}-S_{obstacle})l_{obstacle},(S_{area2}-S_{obstacle})l_{obstacle},\cdots,\\ &\qquad (S_{area12}-S_{obstacle})l_{obstacle}\}\end{aligned} \qquad (3)$$

각각의 넓이가 동일하다고 했을 경우에는 로봇과 물체사이의 거리가 가장 먼 경우를 선택하게 되고, 만약 거리가 모두 동일하다고 하였을 경우에는 로봇 주위의 동일한 넓이에서 물체의 넓이를 뺀 공간 중에서 가장 넓은 공간을 로봇은 선택하게 되어있다.

1. $(S_{areaN}-S_{obstacle})$: largeness $l_{obstacle}$: largeness
2. $(S_{areaN}-S_{obstacle})$: largeness $l_{obstacle}$: smallness
3. $(S_{areaN}-S_{obstacle})$: smallness $l_{obstacle}$: largeness
4. $(S_{areaN}-S_{obstacle})$: smallness $l_{obstacle}$: smallness

4. 시뮬레이션 및 결과

본 연구에서 제안한 알고리즘의 유효성을 보이기 위해서 본 논문에서는 3개의 알고리즘을 이용하여 모의실험을 수행하였고, 그 결과를 그림 8, 9, 10에 나타내었다. 실험 환경에는 10대의 로봇과 25개의 장애물, 그리고 하나의 목표물이 있다고 가정하였고, 탐색시간은 1회당 총60sec로 총 100번에 걸쳐서 진행하였는데 첫 번째는 무작위 탐색방법 두 번째는 DBAM과 ABAM의 융합 모델을 사용해 보았고, 마지막으로는 본 논문에서 제안한 12각형 기반의 Q-learning과 SVM을 이용한 알고리즘을 사용하여서 시뮬레이션 하였다.

총 100회의 탐색 시도에서 랜덤 탐색의 경우는 횟수가 늘어남에도 불구하고 특별한 규칙을 발견하기 힘들었다. 또한 랜덤 탐색의 특성상 통계적인 의미를 부여하기는 어려웠다. 다음으로 DBAM과 ABAM의 융합 모델인 경우, 모든 100회의 시행동안 평균적으로 4,5대 정도의 로봇이 목표물을 찾아내었다. 이것은 DBAM과 ABAM 융합모델을 통해서도 탐색의 성능이

상당히 강화 될 수 있음을 나타낸다. 또한 마지막으로 본 연구에서 제안한 12각형 기반 Q-learning과 SVM 알고리즘을 통한 탐색의 결과는 총 100회 시행에 평균적으로 7대 정도의 로봇이 목표물 탐색에 성공하였다. 이는 로봇이 자신이 처한 상황에 대해서 좀 더 정밀한 상황인식과 행동을 결정할 수 있는 방법이 될 수 있었고 기존의 2가지 방법보다는 과거의 상황에 대해서 좀 더 정확성 있게 상황인식을 하는 결과를 가지고 왔다.

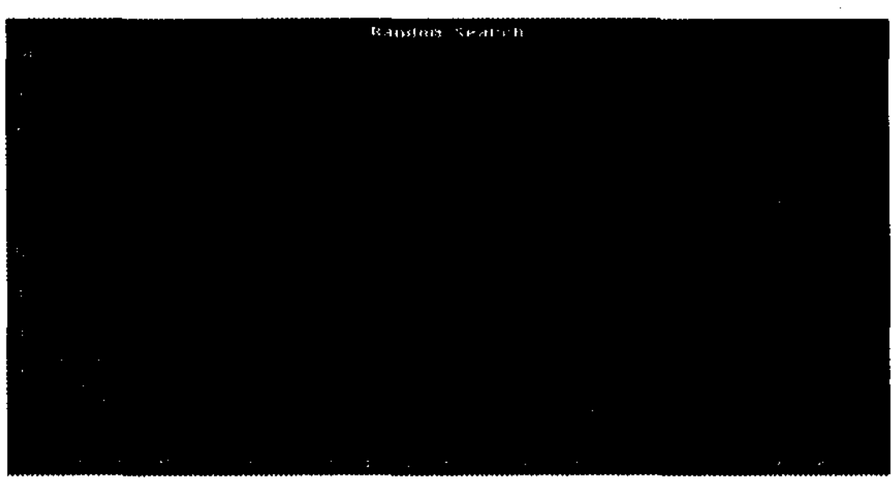

그림 6. 무작위 탐색

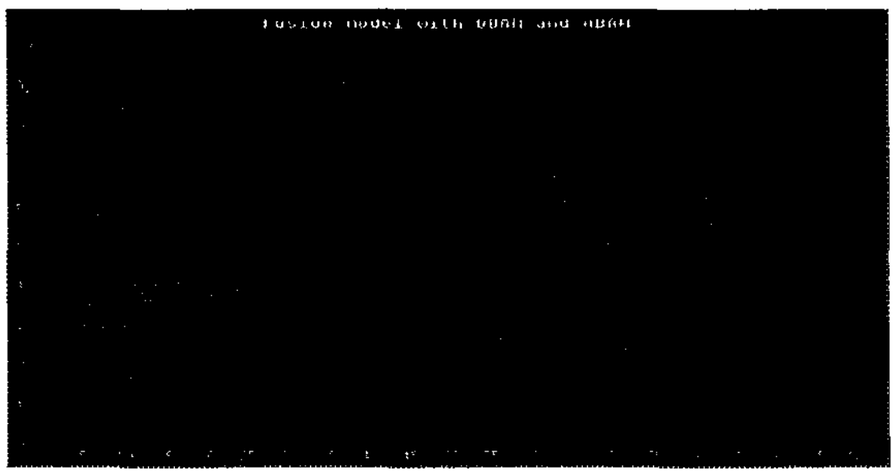

그림 7. DBAM과 ABAM의 융합모델

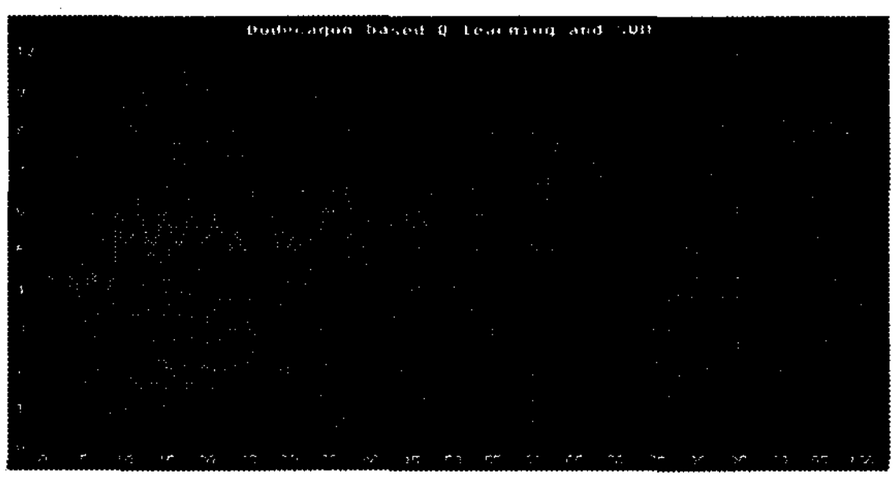

그림 8. 12각형 기반 Q-leaning과 SVM

5. 결 론

본 논문에서는 먼저 선험적 지식이 없고 장애물이 놓여있는 공간에서의 목표물 탐색 알고리즘으로 융합 모델과 SVM과 12각형 기반의 Q-learning 알고리즘을 제안하였다. 여러 대의 로봇을 통해 위에서 언급한 조건의 환경에서의 목표물 탐색을 수행하고 그 결과를 보였다. 실험의 결과를 통해 이 알고리즘이 위와 같은 환경에서의 목표물 탐색에 새로운 방법이 될 수 있음을 알 수 있다.

향후 과제로는 첫째, 로봇들의 협조 행동 학습 알고리즘의 구현을 위해 목표물 발견 후 목표물에 접근하는 알고리즘의 구현이 필요하다. 둘째, 다수 로봇에 의한 물체 수송, 대열을 갖춘 다수 로봇의 이동, object following 또는 path following 등의 로봇 기동에 관한 구현과 Fuzzy와 강화학습의 융합이나 방법의 적용과 같은 심도 있는 알고리즘의 적용에 대한 연구가 뒤따라야 할 것이다. 또한 본 논문에서 제안한 로봇의 수 자체가 군집로봇 시스템 상의 가장 기본이 되는 개수이므로, 앞으로는 수 십대 혹은 수백 대의 로봇을 기본으로 시뮬레이션 해보는 작업도 진행해야 할 것이다.

마지막으로 본 논문에서 제안한 알고리즘은 단지 로봇의 상황과 로봇의 행동에 대한 시뮬레이션을 바탕으로 모델링 한 것이므로, 향후 실제 환경에서 SVM과 12각형 기반의 Q-leaning 알고리즘을 직접 모델링 해봄으로써 앞으로 진행될 연구의 유효성을 좀 더 정밀하게 검증할 예정이다.

참 고 문 헌

[1] L. Parker, "Adaptive action selection for cooperative agent teams," *Proc. of 2nd Int. Conf. on Simulation of Adaptive Behavior*, pp. 442-450, 1992.

[2] G. Ogasawara, T. Omata, and T. Sato, "Multiple movers using distributed, decision-theoretic control," *Proc. of Japan-USA Symp. on Flexible Automation*, vol. 1, pp. 623-630, 1992.

[3] D. Ballard, *An Introduction to Natural Computation*, The MIT Press Cambtidge, 1997.

[4] J. Jang, C. Sun, and E. Mizutani, *Neuro-Fuzzy Soft Computing*, Prentice-Hall New Jersey, 1997.

[5] W. Ashley, T. Balch, "Value-based observation with robot teams (VBORT) using probabilistic techniques," *Proc. of Int. Conf. on Advanced Robotics*, 2003.

[6] W. Ashley, T. Balch, "Value-based observation with robot teams (VBORT) for dynamic targets," *Proc. of Int. Conf. on Intelligent Robots and Systems*, 2003.

[7] T. Mitchell, *Machine Learning*, McGraw-Hill, Singapore, 1997.