SIFT와 베이지안 네트워크를 이용한 불확실한 실내 환경에서의 위치 및 물체 인식

임승빈[0], 조성배
연세대학교 컴퓨터과학과
envymask@sclab.yonsei.ac.kr , sbcho@cs.yonsei.ac.kr

Place and Object Recognition In Uncertain Indoor Environments Using SIFT and Bayesian Network

Seung-Bin Im[0] and Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요 약

영상 정보를 통한 실내 환경의 인식은 지능형 로봇에서 매우 중요한 문제이다. 영상을 통한 실내 환경 정보는 로봇의 각도나 위치의 영향으로 불확실해질 수 있으므로 영상 인식 기법은 이러한 불확실함에 강인함을 갖고 있어야 한다. 본 논문에서는 불확실하게 들어오는 실내 환경 정보에서 PCA를 통한 위치 정보와 SIFT를 통한 물체 존재 정보를 추출하고 이를 베이지안 네트워크에 적용하여 장소 및 물체를 인식하는 방법을 제안한다. 실제 실내 환경에서의 실험을 통하여 8곳의 위치 및 20개의 오브젝트를 효과적으로 인식하는 것을 확인할 수 있었으며 위치에 따른 물체의 존재 확률 추론 및 존재 물체에 의한 위치 확률의 수정 등 다양한 방향의 추론도 가능하다.

1. 서 론

영상 시스템은 시각 정보로부터 장소를 판단하고 물체를 인식해서 고수준의 컨텍스트를 모델링할 수 있는 시스템을 말한다[1]. 사용자의 요구를 이해하고 신뢰성 있는 작업을 수행해야 하는 지능형 로봇의 범주에서 시각 센서만을 사용하여 얻어진 정보만으로 위치 및 물체를 인식하는 것은 어려우면서도 매우 중요한 문제이다. 환경을 실내 사무 환경으로 한정한 영상 인식 기법의 연구는 이런 문제를 푸는 데 있어서 하나의 중요한 접근 방식이 될 수 있다.

영상 정보는 고차원적인 특성 때문에 계산해야 할 데이터의 양이 많아 정보의 처리에 어려움을 겪는다. 따라서 중요 특성 벡터 중심으로 차원을 줄여주는 주성분 분석기법(PCA)을 이용하면 영상 정보의 효율적 처리가 가능하다[2]. 또한 환경 영상 정보는 로봇의 위치나 각도에 따라 영상이 가려지거나 충분하지 못한 조명의 영향 등에 의하여 불확실해질 수 있으므로 불확실한 정보 및 에러에 강한 영상 인식 기법이 필요하다. 이런 불확실한 정보를 처리하기 위해서는 베이지안 네트워크(Bayesian Network)를 이용한 접근방법이 좋은 성능을 보이는데, 이는 베이지안 네트워크가 불확실한 정보의 반영 및 다양한 방향의 추론 등에 강인함을 갖기 때문이다[3].

크기 불변 특성 변환기법(SIFT)은 이미지를 회전, 크기변환 등에 강인함을 갖는 지역 특성 벡터들의 집합으로 변환하는 기법이다[4]. SIFT는 이미지의 크기나 회전, 각도 변화에 강인함을 가지므로 영상에서의 물체의 존재정보 추출에 효과적으로 활용할 수 있다. SIFT로 확인한 물체의 존재정보는 베이지안 네트워크의 증거값으로 적용할 수 있으며 이 정보를 통하여 환경 인식에 대한 더 정확한 베이지안 추론이 가능하다.

본 논문에서는 PCA와 SIFT, 베이지안 네트워크를 이용한 위치 및 물체 인식을 제안한다. 이 방법은 장소와 물체들 사이의 연관성을 표현해줄 수 있는 베이지안 네트워크에 PCA를 통하여 처리된 장소 확률 정보와 SIFT를 이용한 물체 존재 정보를 적용하여 위치와 물체를 인식하는 방법으로, 가려져 보이지 않거나 발견되지 않은 물체들의 존재 추론도 가능하다. 실제 대학 실내 환경에서의 실험을 통해 PCA, SIFT와 베이지안 네트워크를 이용한 환경 인식이 효과적임을 확인할 수 있었다.

2. 관련 연구

MIT 대학의 Antonio Torralba, Kevin P. Murphy 등은 PCA를 통한 전역 특징 중심의 영상 정보를 장소간의 물리적 연관관계를 표현한 전이 행렬과 확률적 모델인 HMM에 적용하여 장소를 인식하는 시스템을 제안했다[5]. 독일 Hamburg 대학의 B. Neumann 등이 수행한 고수준 장면 인식을 위한 지식 표현 방법과 시스템 프레임워크 연구에서는 장면을 인식하기 위해서 물체 중심의 확률적 모델을 생성하여 장면에 대한 해석이 가능함을 보이고 있고, 트리 형태의 확률적 인과 관계 설계 방법을 제안하고 있다[6]. 이 연구에서 확률 추론은 Bayes' Rule을 기초로 한 확률 연산 방법이 사용되었고, 저수준 컨텍스트를 통해 고수준 컨텍스트인 장면 인식 정보를 추론하고 있다.

주위 환경에서의 물체 인식은 C. Papageorgiou와 T. Poggio가 제안한 베이지안 방법론 등 여러 가지 방법으로 얼굴이나 차 등의 인식에 대한 연구가 진행되어 왔다[6,7,8].

3. PCA, SIFT기반 베이지안 네트워크

시스템의 입력으로 들어온 영상은 1)PCA를 이용한 전역 특성 추출, 2)SIFT를 이용한 존재 물체 인식, 3)베이지안 추론이라는 세 가지 과정을 거친다.

그림 1은 PCA, SIFT기반 베이지안 네트워크를 이용한 환경 인식의 전체적인 구조를 보여준다.

3.1 주성분 분석기법을 이용한 전역 특성 추출

지역적 특성을 이용한 영상 인식 기법은 조명이나 잡음 등의 변화에 민감하다는 단점이 존재한다. 따라서 조명, 잡음 등의 변화에 강인한 물체 인식을 위해서는 영상 이미지의 지역적 특

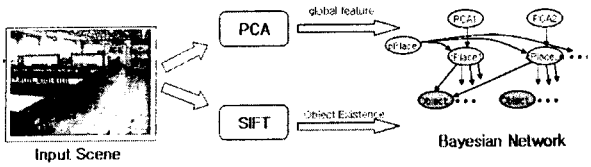

그림 1. 환경 인식의 전체적인 구조.

성과 더불어 전역적 특성을 사용하여야 한다. 이미지의 전역적 특징을 효과적으로 사용하기 위하여 Steerable Pyramid를 이용, 영상을 다양한 orientation과 scaling선분으로 분해하여 조명 등의 효과를 없애며 물체 인식에 필요한 지역적 성질을 유지하게 한다[5]. 그리고 앞서 얻어진 영상에 필터링을 수행하고 해상도를 낮추는 기법을 사용함으로서 잡음에 강인하며 전역적 특성을 유지하는 데이터를 얻을 수 있으며[5] 이러한 처리를 통하여 얻어진 데이터의 크기는 PCA를 사용하여 효과적으로 줄여줄 수 있다. 이 시스템에서는 MIT의 연구를 바탕으로 D=80 PCs를 사용하였다[5]. 이미지에서 얻어진 80개의 벡터는 미리 학습된 장소 벡터들의 가우시안 분포함수에 넣어주어 확률값을 계산해준다.

시간 t때 장소를 $Q_t \in \{1, ..., Np\}$ (Np는 장소의 수)라고 하고, 시간 t일 때의 전역 특성을 v_t^G 라고 하면, $P(Q_t|v_t^G)$는 다음과 같이 표현할 수 있다.

$$P(Q_t|v_t^G) = \exp(\frac{-1}{2\sigma_p^2}\|v_t^G - \mu_q^p\|^2)$$

본 논문에서는 기본적으로 한 장소를 네 개의 모델로 구분하고 모델 당 150장의 사진을 장소 학습 데이터로 사용하였다.

3.2 SIFT

SIFT는 물체 인식의 높은 효율성과 이미지의 크기나 회전에 대한 강인함을 얻기 위하여 이미지를 피라미드 처리하고 처리된 각 레벨에서 가우시안 함수의 차이가 최고/최저인 점을 키로 잡아 추출하는 알고리즘이다[4]. SIFT알고리즘에 사용하는 가우시안 함수는 다음과 같다.

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

SIFT가 각도나 크기에 영향을 적게 받는 기법이라고 해도 3D 오브젝트를 다양한 각도에서 모두 인식하는 것은 불가능하므로, 물체 당 각도를 변화시키면서 추출한 10개의 키를 물체의 SIFT 키 집합으로 갖는다. 입력 이미지와 키 집합을 비교하여 일치된 키가 있으면 해당 물체가 발견되었다고 판단하고 베이지안 네트워크의 증거값으로 설정해 준다.

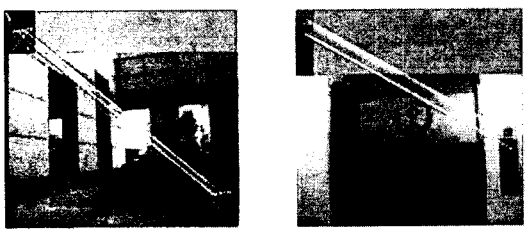

그림 2. SIFT를 사용한 물체 인식의 예. 흰색 선으로 이어진 두 점이 같은 특징점이다.

3.3 베이지안 네트워크

P를 어떤 집합 V에 있는 랜덤 변수 X들의 결합 확률 분포라고 하고, 그래프 $G=(V, E)$가 DAG(directed acyclic graph)일 때, 각각의 변수 $X \in V$에 대하여 $\{X\}$가 방향성 그래프의 모든 부모들과 조건적으로 독립이면 (G, P)를 베이지안 네트워크라고 한다[3]. 베이지안 네트워크는 에러와 다방향 추론에 강인함을 갖는 그래프 모델이며 노드 간의 아크는 확률적 인과관계를 표현한다. 본 논문에서 사용한 베이지안 네트워크는 전문가가 설계하였으며 낮은 인과 관계를 가지는 노드들은 계산의 복잡성을 줄이기 위하여 연결하지 않았다. 실험에서 실제 사용한 베이지안 네트워크 중 하나를 나타내고 있는 그림 3에서는 이전 위치는 사무실이며 이 노드는 사무실 노드와 복도 노드에 연결되어 있다. 이것은 실제 장소와의 물리적 관계를 표현한 것으로 복도를 통하지 않으면 다른 장소로 이동할 수 없음을 나타낸다. 다른 장소들의 베이지안 네트워크들도 위와 같은 룰을 기준으로 이전 장소와의 물리적 연관성을 고려하여 설계되었으며 한 장소 당 하나씩의 베이지안 네트워크가 있다.

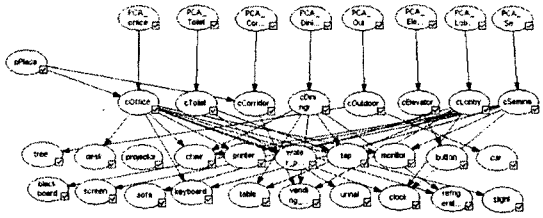

그림 3. 실제 사용한 베이지안 네트워크들 중 하나의 구조. 장소 확률값을 위한 PCA노드, 이전 상태를 위한 이전장소 노드, 현재장소 노드, 물체 노드로 구성되어 있다.

본 논문에서는 영상 정보의 불확실성을 확률적으로 처리하기 위하여 가상 증거 기술(virtual evidence technique)을 정의하여 사용하였다. 가상 증거 기술은 주어진 증거가 확률적인 특성을 가진 경우 이를 반영하기 위해 가상 노드를 자식노드로 정의하여 가변적인 확률 테이블을 사용하는 방법이다. 이 방법은 초기 확률값을 포기하고 그 대신 확률 증거를 반영하는 방법이며, 루트 노드인 경우에만 적용이 가능하다. 실제 장소 인식에서는 이곳이 어느 장소인지를 표현하는 가변적 입력값의 중요성이 초기 확률값의 중요성보다 더 크며 불확실한 증거를 베이지안 네트워크에 넣어 줄 수 있다는 장점이 있으므로 본 논문에서는 이 기술을 사용하여 장소의 확률값을 적용해 준다.

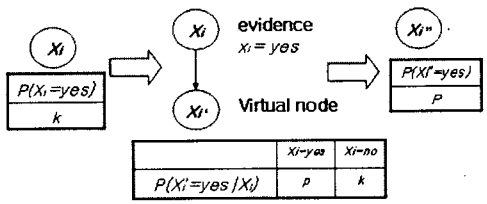

그림 4. 가상 증거 기술을 사용한 노드. 본 논문에서는 가장 오른쪽의 가상 증거 기술을 사용했다.

4. 실험 및 분석

실제 환경에서의 입력을 반영시켜 주기 위하여 실험 데이터는 대학 실내 연구 환경에서 디지털 캠코더를 이용하여 수집하였고, 실험에 사용할 장소는 8곳으로, 물체는 20가지로 한정하여 수행한다. 그러나 나무, 차 등과 같이 모양의 다양성의 문제로 SIFT키를 추출하기가 어렵거나 스크린과 같이 SIFT특징을 뽑아내기 어려운 경우 그러한 예외 물체들은 SIFT키를 통한

직접 인식은 피하고 추론에 의한 결과만으로 존재를 인식한다.

각 장소들은 장소들의 물리적 연관성을 표현해주기 위하여 이전 장소에 따라 이전 상태 노드의 연결이 다르게 구성된 베이지안 네트워크를 장소 수만큼 구성하고 이전 장소 인식의 결과에 따라 해당 베이지안 네트워크를 선택해서 사용한다.

각 장소는 기본적으로 장소 인식을 위한 4개의 학습 모델을 가지며 한 모델당 150장의 학습 데이터를 사용하였다. 그러나 식당과 같이 공간이 너무 넓은 장소의 경우에는 6개의 모델을 사용하였다. 입력 영상이 들어오면 학습 모델 모두에서 장소 인식을 시행하여 가장 높은 확률을 갖는 모델을 그 장소의 확률값으로 결정한다. 그리고 그 확률값을 가상 증거 기술을 사용하여 베이지안 네트워크에 적용시켜 주고, 이전 상태 노드와 SIFT를 통한 존재 물체 정보를 추가하여 베이지안 추론을 시행하고 현 상태 노드들 중 가장 높은 확률을 갖는 노드를 현 장소로 인식한다.

$$P(cPlace_i) = P(cPlace_i \mid PCA_i, pPlace)$$

$$P(CurrentPlace) = \max\,(P(cPlace_i)),$$
$$CurrentPlace = \arg\,\max\,(P(cPlace_i))$$
$$(i = 1, \ldots, N,\ where\ N = number\ of\ Places)$$

실험은 디지털 캠코더를 이용하여 사람의 시점높이에서 여러 장소들을 이동하며 수집한 동영상에서 초당 4장의 이미지를 추출하여 데이터로 사용한다. 동영상은 8가지 장소 중 한 장소에서 출발하여 다른 장소로 이동하는 시퀀스를 가지며 대학 연구건물의 실내 환경에서 1400초 내외의 시간으로 촬영했고 장소의 방문 순서는 랜덤하게 결정되었다.

그림 5는 실제 이동 데이터를 가지고 실험한 결과로 PCA를 통한 장소 정보를 베이지안 네트워크에 적용한 결과를 보여주고 있다. 붉은색 선이 실제 이동을 나타내며 장소 이동에 따른 인식이 잘 이루어지고 있음을 볼 수 있다. 그러나 확률을 나타내는 결과값 중 일부에서는 어느 장소인지 구분이 어렵거나 잘못 판단한 결과가 나타나기도 한다.

그림 6은 PCA를 통한 장소 정보와 SIFT를 통한 물체 존재 정보를 모두 베이지안 네트워크에 적용하여 실험한 결과를 나타내고 있다. 이 실험에서는 그림 5에서 나타났던 장소 이동 시의 불분명한 인식 결과가 많이 수정되었음을 알 수 있다. 이는 베이지안 네트워크의 증거값으로 들어간 물체 존재 정보가 물체와 관계 있는 장소들의 확률값을 올려 주었기 때문이다.

그림 7은 위 실험에서의 위치 이동에 따른 물체들의 존재 확률값의 변화를 보여주고 있다.

5. 결론

PCA와 SIFT를 통한 영상정보를 베이지안 네트워크에 이용한 환경 인식이 장소와 장소에 연관된 물체의 존재 추론에 신용할 만한 결과를 내는 것을 확인할 수 있었다. 또한, 장소를 판단하기 어려운 경우 물체 존재정보를 이용하여 역방향 추론을 하여 장소를 판단할 수 있는 것도 확인할 수 있었다. 그러나 물체의 표면에 텍스쳐 성분이 부족한 경우 SIFT 키 매칭율이 상당히 떨어지게 된다. 이를 해결하기 위해서는 온톨로지 기반으로 물체를 분해하여 각 분해된 오브젝트별로 SIFT키 매칭을 검사하거나 SIFT키 집합을 더 다양한 각도에서 추출해서 키 집합의 수를 증가하여 인식하는 방법이 있을 수 있다. 향후 연구로 이전 상태를 명확하게 표현해줄 수 있는 다이나믹 베이지안 네트워크에 영상 시스템을 적용해볼 수 있으며, 실제 로봇으로 확장하는 연구도 필요하다.

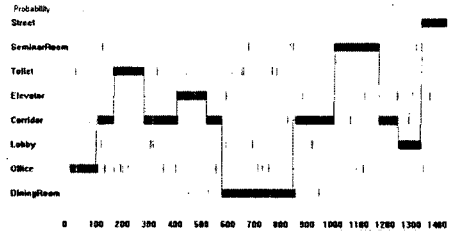

그림 5. 장소 인식 실험 결과. 색이 어두울수록 확률값이 높은 것이며 붉은색 선은 실제 장소를 나타낸다.

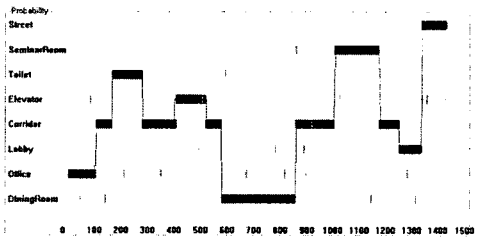

그림 6. SIFT를 적용한 장소 인식 실험 결과

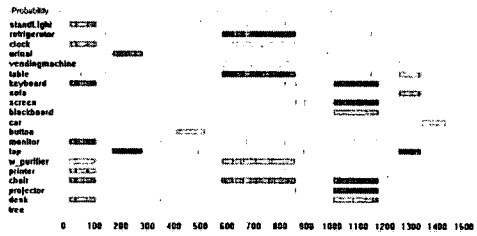

그림 7. 물체 존재 확률 그래프

참고 문헌

[1]B.Neumann, "A conceptual framework for high-level vision," *Bericht, FB Informatik*, FBI-HH-B245/02, 2002.

[2]J.Portilla, E.P.Simoncelli. "A parametric texture model based on joint statistics of complex wavelets coefficients," *Intl. J Computer Vision*, vol. 40, pp. 49-71, 2000.

[3]R.E.Neapolitan, *Learning Bayesian Network*, *Prentice hall series in Artificial Intelligence*, 2003

[4]D.G.Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Computer Vision*, vol. 60, no. 2 , pp. 91-110 , 2004.

[5]A. Torralba, K. P. Mutphy, W. T. Freeman and M. A. Rubin, "Context-based vision system for place and object recognition," *IEEE Int. Conf. on Computer Vision*, vol. 1, pp. 273, 2003.

[6]G. F.Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, pp. 309-347, 1992

[7]C. Papageorgiou and T. Poggio. "A trainable system for object detection," *Intl. J. Computer Vision*, vol. 38 no. 1, pp. 15-33, 2000.

[8]H. Schneiderman and T. Kanade. "A statistical model for 3D object detection applied to faces and cars," *In Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 746-751, 2000.

[9]P. Korpipaa, M. Koskinen, J. Peltola, S. Ma kela, T. S. nen. "Bayesian approach to sensor-based context awareness," *Personal and Ubiquitous Computing Archive*, vol. 7, pp. 113-124, July, 2003