장면 인식 성능 향상을 위한 베이지안 확률 및 증거의 결합

황금성°, 박한샘, 조성배

연세대학교 컴퓨터과학과
{yellowg°, sammy, sbcho}@cs.yonsei.ac.kr

Bayesian Probability and Evidence Combination For Improving Scene Recognition Performance

Keum-Sung Hwang°, Han-Saem Park, Sung-Bae Cho

Dept. of Computer Science, Yonsei University

요 약

지능형 로봇 기술이 발전하면서 영상에서 장면을 이해하는 연구가 많은 관심을 받고 있으며, 최근에는 불확실한 환경에서도 좋은 성능을 발휘할 수 있는 확률적 접근 방법이 많이 연구되고 있다. 본 논문에서는 확률적 모델링이 가능한 베이지안 네트워크(BN)를 이용해서 장면 인식 추론 모듈을 설계하고, 실제 환경에서 얻어진 증거 및 베이지안 추론 결과를 결합하여 분류 성능을 향상시키기 위한 방법을 제안한다. 영상 정보는 시간에 대해 연속성을 가지고 있기 때문에, 증거 정보와 베이지안 추론 결과들을 적절히 결합하면 더 좋은 결과를 예상할 수 있으며, 본 논문에서는 확신 요소(Certainty Factor: CF) 분석에 의한 결합 방법을 사용하였다. 성능 평가 실험을 위해서 SIFT (Scale Invariant Feature Transform) 기법을 이용하여 물체 인식 처리를 수행하고, 여기서 얻어진 데이터를 베이지안 추론의 증거로 사용하였으며, 전문가의 CF 값 정의에 의한 베이지안 네트워크 설계 방법을 이용하였다.

1. 서론

우리는 영상 정보를 분석하여 어떤 장면에서 어떤 물체가 존재하는지, 어떤 장소에 있는지, 그리고 어떤 상황인지를 알고 싶어 한다. 이렇듯 장면 시각 정보로부터 현재 장면에 대해 설명하는 컨텍스트를 추출하고 이를 해석하는 것을 장면 인식이라고 하며 아직까지 잘 풀리지 않은 문제 중에 하나이다[1]. 장면 인식을 하게 되면 사용자의 의도를 인지하거나 작업 중심의 서비스 제공, 과거의 경험 발견 및 신뢰성 있는 작업이 가능하기 때문에 중요한 연구 과제로 여겨지고 있다. 특히 실세계에서 사용자 및 환경과 상호작용이 요구되는 문제에서 시각 정보는 중요한 판단의 근거가 되기 때문에 장면 인식의 역할이 크다.

본 논문에서는 장면 인식을 위한 컨텍스트 추론 모델을 구성하는 데 있어 환경의 불확실성에 의한 성능 저하를 최소화하기 위한 방법의 하나로 사용되는 베이지안 추론 모델의 성능 향상에 초점을 둔다. 영상 정보는 시간에 대해 연속성을 가지고 있기 때문에 순차적으로 들어오는 영상 정보들은 서로 많은 연관성을 가지고 있다. 따라서 장면 인식 과정에서 얻어진 정보들을 추론해 얻은 베이지안 확률 또한 이러한 연관성을 가질 확률이 높다. 본 논문에서는 베이지안 네트워크를 통해 추론된 확률 결과들을 CF 분석에 의해 결합하여 더 좋은 성능이 나올 수 있음을 보이고자 한다.

관련 연구를 살펴보면, 최근 Torralba 등[2]에 의해 진행된 시각 센서를 이용한 장면 인식 연구에서는 장소 인식 방법으로 베이지안 규칙을 이용하였고 장소간 순서적인 연관성을 고려하기 위해 은닉 마르코프 모델(Hidden Markov Model: HMM)의 변환 행렬을 사용하고 있다.

독일 Hamburg 대학의 B. Neumann 등이 수행한 과제에서는 장면을 인식하기 위해서 개체의 존재(object)와 개체의 특성 및 개체 그룹과 클래스 정보를 계층적으로 구성하고 확률적 관계를 정의하였고[1]. 시간적인 연관성과 공간적인 연관성을 나누어서 별도의 추론 모델을 설계하였다.

2. 배경

2.1. 영상 장면 인식

장면 인식을 위한 추론 과정은 그림 1과 같다. 환경에서 수집된 센서 정보 및 기호적으로 해석된 컨텍스트 정보를 온톨로지 등의 도메인 지식과 함께 고려하여 상대적으로 수준이 높은 컨텍스트를 추론하는 과정을 거친다. 고수준 컨텍스트는 장면 인식 정보를 가리키며 본 논문에서는 장소의 종류 속성을 나타낸다.

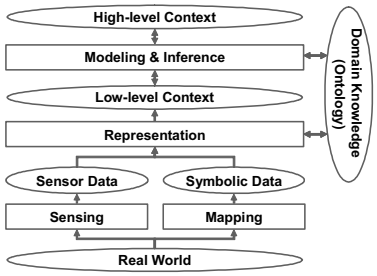

그림1. 장면 관련 고수준 컨텍스트의 인식 과정

추론 과정에는 일반적으로 확률적인 접근 방법이 많이 사용된다. 복잡한 실세계의 모든 확률 관계를 표현하거나 설계하는 것은 매우 어렵기 때문에, 확률적 인과성이 강한 정보들의 확률만 정의하는 베이지안 네트워크 모델을 이용하는 것이 일반

적이다. 베이지안 네트워크를 이용하면 분석적 추론을 이용해 결과에서 원인을 분석할 수 있을 뿐만 아니라, 구성하고 있는 노드의 종류에 상관없이 유연하게 입출력이 가능하며, 불확실성을 확률적으로 다룰 수 있는 장점이 있다.

2.2. 베이지안 네트워크

베이지안 네트워크는 변수들 간의 원인과 결과 관계를 확률적으로 모델링하기 위한 도구이다. 인과관계 네트워크를 구성한 다음 조건이나 증거가 주어진 경우의 확률, 즉 조건부 확률을 계산하여 결과 확률을 추론한다[3,4]. 그림 2는 간단한 베이지안 네트워크의 예로서 노드는 변수를, 방향성 연결선은 인과관계를 표현한다. 조건부확률 테이블(CPT: conditional probability table)은 부모에 대한 자식의 확률관계를 나타낸다.

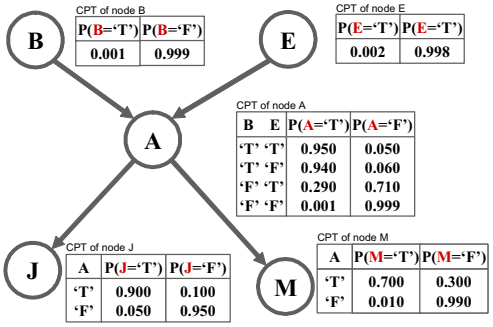

그림 2. 간단한 베이지안 네트워크 구조의 예

그림 1과 같은 확률 테이블을 구하기 위한 베이지안 확률 계산식은 수식 (1)과 같다[3].

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{\sum_{x \in \chi} P(x, y)} = \frac{P(y \mid x)P(x)}{\sum_{x \in \chi} P(y \mid x)P(x)} \quad (1)$$

3. CF를 이용한 베이지안 추론 확률 결합

3.1. 베이지안 추론 확률 결합을 이용한 장면 인식

제안하는 시스템을 도식으로 나타내면 그림 3과 같다. 먼저, 시간이 흐름에 따라 들어오는 영상 데이터를 SIFT 모듈에 입력하여 미리 학습해 놓은 개체 이미지와 특징을 비교한다. 그 결과, 얻어지는 것이 현재 영상에서 발견된 물체의 리스트이다.

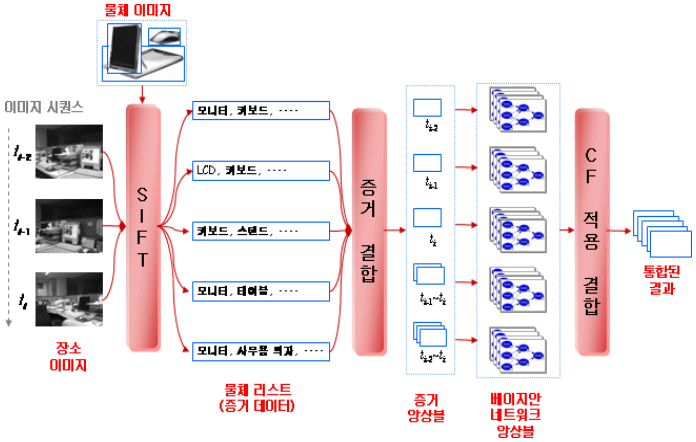

그림 3. 제안하는 장면 컨텍스트 추출 방법

두 번째 단계는 증거 결합 단계로서, 각 시간대별로 들어온 증거값(발견된 물체 정보)을 구분하여 결합하는 단계이다. 본 논문의 실험에서는 t_i, t_{i-1}, t_{i-2}, t_{i-1}~t_i, t_{i-2}~t_i가 사용되었다.

세 번째 단계는 시간대별로 결합된 증거 집합을 이용하여 베이지안 네트워크의 추론을 수행한 다음 그 결과를 확신 요소(Certainty Factor: CF) 값을 가중치로 하여 합치는 단계이다. 이렇게 통합된 결과에서 가장 높은 확률을 나타내는 장소가 현재의 장소를 나타내게 된다.

3.2. 베이지안 네트워크 설계

컨텍스트 추론을 위한 베이지안 네트워크는 다음과 같은 과정을 통해 설계되었다. 첫 번째, 인식된 각 물체의 장소에 대한 연관성을 CF 값으로 정의한다. 이때 CF는 얼마나 정확한지, 얼마나 믿을 수 있는지를 사람이 예측한 값을 나타낸다. 두 번째, 생성된 CF 값의 테이블을 이용하여 부모 조합에 의한 조건부 확률을 계산한다. 이때 CF값을 0과 1사이의 확률값으로 변환하게 된다. 마지막으로, 각 장소별로 Naive Bayesian Network를 설계한 뒤 확률 테이블을 적용한다. 이때 계산의 편의를 위해 장소 컨텍스트를 부모로, 물체 컨텍스트를 자식으로 설계하였는데, 이렇게 하면 조건부 확률 테이블(CPT)의 크기가 줄어들어 추론 연산의 속도가 빨라지고 확률 계산이 쉬워지는 장점이 있다.

그림 4는 실제 설계에 이용된 CF 테이블과 설계된 베이지안 네트워크를 보여준다. 각 물체에 대해 해당 장소에 있음 정도를 '높음/보통/낮음/매우 낮음'으로 구분하여 각각 0.9, 0.5, -0.5, -0.9의 CF값을 적용하였으며, 15개의 물체 노드를 가지는 5개의 장소 인식 베이지안 네트워크를 설계하였다.

	사무실	휴게실	화장실	엘리베이터 앞	복도
모니터	0.9	-0.9	-0.9	-0.9	-0.9
키보드	0.9	-0.9	-0.9	-0.9	-0.9
테이블	0.5	0.5	-0.9	-0.9	-0.9
의자	0.9	0.5	-0.9	-0.9	-0.9
파티션	0.5	-0.9	-0.9	-0.9	-0.9
휴지통	0.5	-0.5	-0.5	-0.5	0.5
문	0.9	0.9	0.9	-0.5	-0.5
창문	0.5	0.5	-0.5	-0.5	0.5
타일	-0.5	-0.5	0.9	-0.9	-0.9
소변기	-0.9	-0.9	0.9	-0.9	-0.9
세면대	-0.5	-0.5	0.9	-0.9	-0.9
엘리베이터 문	-0.9	-0.9	-0.9	0.9	-0.5
엘리베이터 스위치	-0.9	-0.9	-0.9	0.9	-0.5
화장실 표지판	-0.9	-0.9	0.9	-0.9	-0.5

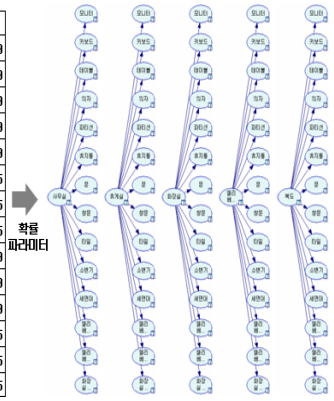

그림 4. CF 테이블(왼쪽)과 설계된 베이지안 네트워크(오른쪽)

CF 값은 사람의 판단을 효과적으로 표현하고 계산하기 위해 1980년대에 David McAllister에 의해 제안된 방법으로 -1에서 1사이의 값으로 정하며, 긍정적인 경우 0보다 큰 값을 가진다[5]. 조합된 CF 값을 계산하는 함수 f는 수식 (2)와 같다.

$$f(CF_1, CF_2) = \begin{cases} \text{when } CF_1 \geq 0 \text{ and } CF_2 \geq 0, & CF_1 + CF_2(1 - CF_1) \\ \text{when } CF_1 \leq 0 \text{ and } CF_2 \leq 0, & CF_1 + CF_2(1 + CF_1) \\ \text{when } CF_1 \times CF_2 \leq 0, & (CF_1 + CF_2)/(1 - \min(|CF_1|, |CF_2|)) \end{cases} \quad (2)$$

3.3. SIFT를 이용한 물체 인식

실세계의 이미지를 인식하려면 노이즈의 영향을 받지 않는 특징이 필요하다. SIFT (Scale Invariant Feature Transform)는 이미지를 변환, 회전, 조명 변화 등으로부터 영향을 받지 않는 스케일에 독립적인(scale invariant) 지역적인 특징 벡터들의 집합을 추출하여 이미지를 인식하는 방법이다[6,7]. 그림 5와 같이 scale-space를 생성하고, 일치하는 키포인트를 찾은 다음, 지역 최적화 키포인트를 계산하고, 각 키포인트의 방향성을 부여한 뒤, descriptor라는 3차원 벡터로 표현하는 과정을 거친다. 이렇게 얻은 SIFT 키들은 그림 6과 같이 최근접 이웃 방식(nearest-neighbour approach)을 통해 이미지로부터 대상 물체를 인식 하는 과정에 이용된다.

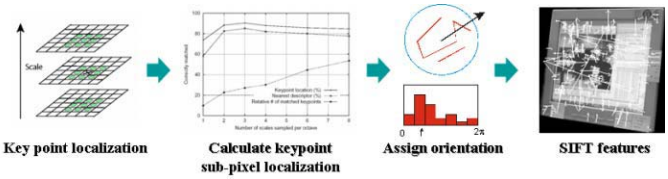

그림 5. SIFT 특징키 추출 과정

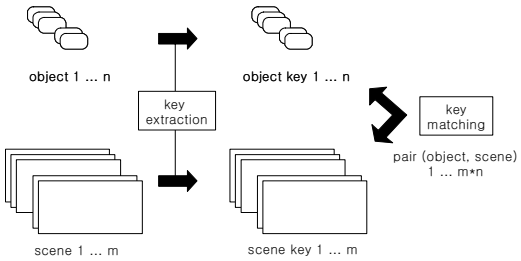

그림 6. SIFT를 이용한 키 매칭 과정

4. 실험 및 결과

연구실 주변 환경을 대상으로 5개의 환경을 선택하고, (사무실→복도→휴게실→복도→화장실→복도→엘리베이터 앞→복도→사무실)의 순서로 이동하며 총 149장의 이미지를 수집하였다. 물체의 경우 총 14가지의 물체를 가정하고, 각 물체마다 1개~7개의 이미지를 정의하여 학습하였다. 제안하는 BN 추론 확률의 결합 범위를 정하기 위해 증거 결합에 따른 성능 변화를 살펴보았다. 간단한 실험 결과 이전 이미지 포함 3개까지 사용했을 때 가장 좋은 성능을 보였기 때문에 본 논문에서는 3개의 시간대를 사용하였다. 결합 CF 값은 몇 번의 비교 실험 결과 t_i=0.9 (현재 i번째 장면의 CF값), t_{i-1}=0.75, t_{i-2}=0.6, t_{i-1}~t_i=0.8, t_{i-2}~t_i=0.7로 정해졌다. 실험 결과 비교를 위해 세 가지 방법이 사용되었다. Normal은 BN을 통한 장소 분류 결과의 결합을 수행하지 않고 현재 시간의 결과만 사용한 것을 의미하고, Voting는 다수결에 의한 결합 방법을 의미하며, CF는 CF 가중치를 이용한 결합 방법을 의미한다. 표 4는 비교 실험 결과 장소 분류 성능을 나타낸다. SIFT에 의해 발견된 물체의 수가 프레임당 평균 1개가 안 되는 수준이었기에 그다지 높은 성능은 보이지 않았지만 결합을 적용함으로 인해서 분류 성공률이 향상되었음을 알 수 있다.

표 1. BN 결합 방법별 분류 성능 비교

실험 방법	분류 성공률
Normal	0.604
Voting	0.624
CF	**0.698**

베이지안 네트워크에 의해 장소가 분류되는 과정을 좀 더 자세히 살펴보기 위해 각 프레임별 분류 결과를 그래프로 그려서 관찰해 보았다. 그림 7은 제안한 방법을 사용해서 시간대별로 장소 이동시 나타난 값의 변화를 나타낸 그래프를 Normal과 CF에 대해서만 나타낸 그림이다. Goal은 실제로 이동한 경로를 나타낸다. 그림을 보면 장소 인식이 복도로 집중되어서 많이 떨어지는 것을 알 수 있는데, SIFT에 의해 발견된 물체가 없을 경우 아무 물체도 없는 복도로 인식되기 때문이다. 그래프에서 CF에 의해 결합한 경우 좀더 안정적인 결과를 보이고 있음을 알 수 있으며, 장소가 변하는 지점에서 변화가 늦어져서 잘못된 결과를 보이는 경우가 일부분 관찰되었다. 사무실에 관련된 물체가 가장 많았기 때문에 사무실이 가장 잘 인식되었으며, 특정 물체가 거의 발견되지 않는 휴게실은 분류 성능이 떨어짐을 알 수 있다.

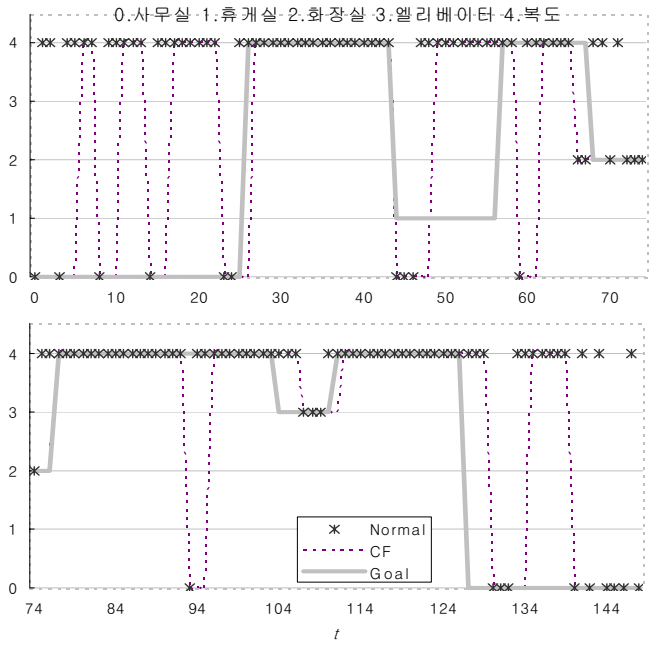

그림 7. 장소 이동에 따른 장소 컨텍스트 변화

5. 결론 및 토의

본 논문에서는 영상에서 얻어진 정보를 이용하여 장면 컨텍스트를 추론하는 시스템에서, 연관성이 있는 연속적인 증거의 축적과 각 증거별 추론 결과의 결합을 통해 성능 향상을 꾀하였다. CF 분석을 통해 결합하는 방법을 제안하였으며, 전문가의 지식을 반영한 BN 설계 기법으로 CF 테이블을 이용한 방법도 소개하였다. SIFT를 이용해 실제 장소별 증거를 수집한 실험 결과 장소 인식 성능이 상승하여 제안한 방법이 유의미함을 알 수 있었다.

실험에서 물체 인식 방법으로 사용한 SIFT의 물체 인식 성능이 생각보다 부족하여 증거의 수가 적었고, 장소 인식 성능이 많이 떨어졌다. 향후에는 물체 인식 성능을 확장하여 더 많은 증거를 확보한 실험이 필요할 것으로 예상된다.

· 참고문헌

[1] B. Neumann, *A Conceptual Framework for High-Level Vision*, Bericht, FB Informatik, FBI-HH-B245/02, 2002.

[2] A. Torralba, et al., "Context-based vision system for place and object recognition," *Int. Conf. Computer Vision*, pp. 273-280, 2003.

[3] F. Cantu, "Learning and using Bayesian networks for diagnosis and user profiling," *Technical Report CIA-RI-043, Center for Artificial Intelligence*, ITESM. Invited talk at *the Computing Int. Conference, CIC-IPN*, November 2000.

[4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA, 1988.

[5] J. C. Giarratano and G. D. Riley, *Expert Systems: Principles and Programming*, PWS-Kent Publishing Company, Boston, Mass, 1989.

[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.

[7] D. G. Lowe, "Object recognition from local scale-invariant features," *Proc. of the Int. Conference on Computer Vision (ICCV)*, pp. 1150-1157, 1999.