

모바일 비전-위치 데이터 학습 기반 장소 인식*

이충연¹○, 이범진², 하정우², 장병탁^{1,2}

{cylee, bjlee, jwha, btzhang}@bi.snu.ac.kr

¹서울대학교 뇌과학협동과정, ²서울대학교 컴퓨터공학부

Place Recognition by Learning Mobile Vision-Location Data

Chung-Yeon Lee¹○, Beom-Jin Lee², Jung-Woo Ha², Byoung-Tak Zhang^{1,2}

¹Brain Science Program, Seoul Nat'l Univ., ²School of Computer Science and Engineering, Seoul Nat'l Univ.

요약

추가적인 하드웨어나 인프라의 도입 없이 개별 사용자의 건물 내 장소에 대한 추상화된 정보 획득이 가능하다면, 보다 효율적이고 다양한 위치기반서비스를 제공할 수 있다. 위치정보 데이터베이스에 기반한 기존 GPS 기반 위치 인식 시스템과 달리 비전 정보를 위치 인식에 사용할 경우 음영지역이나 데이터베이스에 명시되지 않은 위치에 대해서 추상화된 사용자의 현재 장소를 인식할 수 있다. 본 논문에서는 서포트벡터머신(SVM)을 이용하여 스마트폰에 장착된 카메라와 GPS 센서를 통해 입력받은 영상 데이터와 GPS 좌표의 통합정보로부터 장소를 인식하는 방법을 제안한다. 제안하는 방법을 평가하기 위해 네 종류의 실내외 장소에서 촬영한 약 4000개의 영상-GPS 데이터로부터 사용자의 현재 장소를 인식하였다. 실험 결과는 제안하는 방법을 통해 위치관련 정보가 없는 상황에서도 영상정보 만으로도 사용자가 현재 위치한 장소를 인식하는 것이 가능함을 보여준다. 또한 영상 데이터와 동시에 기록된 GPS 데이터를 SVM에 입력되는 벡터의 가중치로 사용함으로써 인식 정확도를 향상시킬 수 있음을 확인하였다.

1. 서론

최근 스마트폰 사용이 대중화됨에 따라 사용자의 위치에 따른 특정 정보를 활용하여 다양한 생활편의를 제공하는 위치기반서비스(Location-based Service, LBS)가 부상하고 있다. 이에 따라 LBS에서의 보다 정밀한 위치 정보 수집을 위하여 현재 선진 각국에서 위치 인식 시스템 기술에 대한 활발한 연구가 진행되고 있다.

위치 인식 시스템은 인식 가능 영역에 따라 매크로 위치 인식, 마이크로 위치 인식, 그리고 Ad-hoc 위치 인식 시스템으로 분류할 수 있다 [1].

매크로 위치 인식 시스템은 광범위의 위치 인식 영역을 제공하며, GPS 및 이동통신망 기반 위치 인식 시스템, 그리고 이동통신망과 GPS를 복합적으로 활용하는 A-GPS(Assistance-GPS)가 활용되고 있다. 하지만 실내 및 건물 밀집 지역과 같은 음영지역에서는 신호의 강도로 정확한 위치 인식이 어렵다는 단점이 있다.

마이크로 위치 인식 시스템은 매크로 위치 인식 시스템이 커버하지 못하는 지역에서 위치 인식을 제공하며, 위치 인식을 위해 사용되는 매체에 따라 적외선 기반 시스템(Active Badge), 초음파 기반 시스템(Active Bat, Cricket), RF 신호 기반 시스템(RADAR, 3D-iD) 등이 있다. 본 시스템들은 위치 인식을 위해 추가적인 하드웨어

를 필요로 하는 제약을 가진다.

Ad-hoc 위치 인식 시스템은 위치 인식을 위해 고정된 인프라에 대한 접근이 어려운 환경에서 절대적 위치 좌표를 제공하는 특정 레퍼런스 노드에 의존하지 않고, 이동 레퍼런스 노드들 간의 연결성만으로 위치를 계산한다 [2]. 하지만 이 방식은 대부분 상대적이고 논리적인 위치만이 판별되므로 지리적으로 절대적인 위치를 얻어내는 것은 어렵다.

이밖에도 지능형 모바일 로봇의 자율이동기능 구현을 위해서는 영상 정보를 분석하여 위치를 파악하는 기술인 비전(Vision) 기반 위치 인식 시스템이 활발히 연구되고 있다 [3-5]. 로봇이 사람처럼 집안의 가구 위치와 벽면 각도, 바닥이나 천정, 조명의 특징 등을 인식해서 스스로 이동경로를 설정하는 가장 고차원적인 기술로, 사람처럼 순수한 영상정보로만 위치를 인식한다는데 의의가 있다.

위치 인식 시스템을 LBS에서 활용하기 위해서는 수집된 좌표 데이터를 실제 서비스에서 이용 가능한 형태의 논리 위치 정보, 즉 장소(place) 정보로 추상화 맵핑하는 과정이 필요하다. 광범위의 위치 영역에서는 이미 지도 정보를 구축하여 활용함으로써 주소 및 지명을 제공하고 있지만, 특정 건물 내부에서 각 층(floor) 및 방(room) 단위로 구분하여 서비스를 제공하지는 않는다. 이러한 서비스를 위해 각 건물마다의 추상화된 논리 위치 정보를 맵핑하여 활용하는 것은 맵핑의 수가 너무 많고, 교체될 때마다 정보를 갱신해야 하며, 또한 개인이나 기업의 프라이버시가 침해될 수 있기 때문에 현실적으로 불가능에 가깝다.

* 이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. 2012-0005643, Videome, No. 2012-0005801, BrainNet), 교육과학기술부의 BK21-IT 프로그램에서 일부 지원되었음.

따라서 위치 인식을 위한 추가적인 하드웨어나 인프라의 도입 없이 건물 내 위치에 대한 추상화된 정보를 개별 사용자별로 얻어낼 수 있다면, 보다 효율적이고 다양한 LBS를 제공할 수 있을 것이다. 이를 위해 본 논문에서는 스마트폰에 장착된 카메라를 통해 입력된 영상 데이터를 기반으로 실내외의 장소를 인식하는 방법을 제안하고, 나아가 영상 데이터와 GPS 좌표의 통합정보를 이용하여 보다 정확히 장소를 인식하는 방법을 제안한다.

2. 연구 내용 및 방법

먼저 스마트폰에 장착된 카메라를 이용하여 장소 인식에 사용할 시험 영상들을 촬영한다. 본 연구의 목적은 자동화된 LBS에서의 장소 인식이므로, 사진 촬영이 아닌 사용자가 일반적으로 스마트폰을 사용할 때 세로 방향으로 카메라가 아래쪽 사선방향으로 파지하고 있는 경우를 상정하여, 촬영 시 해상도는 480×640 pixels(72 dpi)로 하고, 정면 방향과 아래쪽 사선 방향으로 각각 촬영한다. 조리개 개방 및 노출 수치는 카메라가 자동으로 조정하며, 플래시나 손떨림 보정, 영상 후보정과 같은 부가적인 영상처리 기법들은 사용하지 않는다.

이후 영상 데이터의 크기를 120×160 pixels로 축소하고, 그레이스케일(grayscale) 영상으로 변환한 후, SIFT(Scale Invariant Feature Transform)[6] 기술자를 사용하여 특징 벡터를 다음 수식 1과 같이 추출한다.

$$Descriptor(x,y) = [\theta(x,y), m(x,y)] \quad (1)$$

$$\theta(x,y) = \tan^{-1} \left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \right)$$

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

수식 1에서 Descriptor(x,y)로 표현되는 SIFT 기술자는 128차원의 벡터 집합으로 구해지는데, 이것은 특징점 부근의 4×4 격자 영역 내부 픽셀의 기울기 값 θ 를 8방향으로 양자화하고, 각 픽셀의 벡터 방향에 따라 크기 m 을 정렬한 히스토그램으로 나타낸 것이다.

구해진 특징 벡터를 모두 사용할 경우 영상 데이터의 수에 비례하여 학습에 필요한 비용이 급격히 늘어날 수 있기 때문에, 본 연구에서는 k -means 클러스터링을 적용하여 차원을 축소한 특징 벡터를 실험에 사용한다. k -means는 사전에 정한 클러스터 수 k 에 기초하여 전체 데이터를 상대적으로 유사한 k 개의 클러스터로 구분하는 방법이다. 즉, k 값에 따라 특징 벡터의 크기가 결정되므로, 효율적인 학습을 위해 최적의 k 값을 구한다.

다음으로 Support Vector Machine(SVM) 분류기를 이용하여 시험 영상으로부터 추출한 특징 벡터를 영상에 촬영된 장소에 대하여 분류한다. 본 논문에서는 2차 프로그래밍(Quadratic Programming Problem)의 해로 널리 사용되는 Sequential Minimal Optimization(SMO) 방식의 LIBSVM [7]을 사용한다.

SVM은 입력되는 벡터에 가중치를 적용함으로써 중요도가 높다고 생각하는 특징에 더 많은 가중치를 주어 학

습이 가능하게 할 수 있다. 따라서 장소 인식 성능을 강화하기 위해 영상 촬영과 동시에 입력 받은 GPS 데이터를 SVM 분류시 가중치로 사용한다. GPS 데이터 획득을 위해 영상 데이터의 EXIF(Exchangeable Image File format) 메타 데이터에 기록되어 있는 도분초($dd^{\circ}mm'ss$) 단위의 GPS 좌표를 추출하고, 수식 2를 이용하여 도 단위(DD.DDDD)의 데이터로 변환한다.

$$DD = dd + (mm/60) + (ss/(60)^2) \quad (2)$$

이때 이상치(outlier)로 인한 성능 저하를 막기 위해 획득한 GPS 데이터 중 실험 장소로 사용된 건물의 실제 좌표 범위를 크게 벗어나는 것은 학습에서 제외시킨다. 추출한 GPS 데이터를 가중치 벡터 W 로 사용한 SVM의 판별 함수는 수식 3과 같다.

$$f(X) = \sum_{i=1}^M W_i \alpha_i^* K(X_i^*, X) + b^* \quad (3)$$

$$W = DD_E \times DD_N$$

$$K(X_i^*, X) = \exp(-\gamma \|X_i^* - X\|^2)$$

X_i^* 는 학습에 의해 구해진 M 개의 서포트 벡터(support vector) 중 i 번째 벡터이며, 최적화 바이어스(optimization bias) a^* 와 b^* 는 학습에 의해 구해지는 2차 프로그래밍 문제의 해이다. 커널 함수로는 Radial Basis Function(RBF)을 사용한다.

3. 실험 결과

실험에 사용된 영상 데이터는 스마트폰(삼성 갤럭시 S3, 8M 화소 전면 카메라)을 이용하여 교내 강의실과 식당, 복도 그리고 실외에서 표 1과 같이 촬영하였다. 강의실은 서로 다른 층에서 2곳씩, 총 4곳, 복도는 같은 건물의 1~6층, 식당은 1곳, 실외는 건물 주위이다.

촬영된 영상들로부터 SIFT와 k -means 클러스터링을 통해 k 차원의 특징 벡터들을 추출하고, SVM 분류기를 사용하여 10-fold 교차 검증한 결과는 표 2와 같다.

표 1. 장소별 시험 영상 데이터 크기 및 예 (단위: 개)

	강의실	복도	식당	실외
정면방향	430	497	548	553
사선방향	544	494	520	518

표 2. 특징벡터 크기별 비전 기반 장소 인식 결과

k	정면방향		사선방향	
	정확도(%)	수행시간(s)	정확도(%)	수행시간(s)
50	69.34	4.896	75.64	4.096
100	72.45	5.799	78.75	4.218
200	74.61	9.740	80.03	6.964
500	80.25	29.279	80.87	23.613

사선방향에서 촬영한 영상 데이터의 인식률이 평균적으로 정면방향의 데이터보다 높은 가운데, k가 500일 때 사선방향의 데이터가 가장 높은 인식률을 나타내었다. 하지만 정면방향 데이터를 함께 사용하는 것과 전체 수행 시간을 고려한다면, 향후 LBS 시스템에서는 k가 200인 특징 벡터를 사용하는 것이 효율적일 것이다.

스마트폰에 탑재된 A-GPS 시스템을 이용하여 영상 데이터와 함께 기록된 GPS 데이터 중 가용한 것은 실험 위치의 좌표 범위를 북위(N) 37.4480 ~ 37.4490도, 동경(E) 126.9515 ~ 126.9530으로 설정하여 전처리를 수행한 결과, 강의실, 복도, 식당, 실외에서 각각 사선방향에 대한 것이 89, 494, 326, 493개, 정면방향에 대한 것이 58, 197, 287, 322개이다.

그림 1은 전처리를 거친 사선방향의 GPS 데이터 좌표에 대한 히스토그램이다. 두 세 곳의 위치가 특히 실내에서의 좌표가 많이 겹치는 것을 확인할 수 있으며, 해당 좌표들은 실제 위치와 상당수 차이를 확인하였다. 이러한 현상은 실내에서 GPS 신호가 감쇄되어 정확하지 않고, 층간 구분이 되지 않기 때문인 것으로 볼 수 있다.

전처리된 GPS 데이터만을 사용하여 결정트리 알고리즘 중 하나인 J48로 분류할 경우 80~90%의 높은 분류도를 나타내었지만, 데이터의 신뢰도가 낮고, 명시되지 않은 위치에 대한 논리적인 위치 인식이 불가능하다는 문제점이 있다. 하지만 본 GPS 데이터의 불연속적인(discrete) 특성으로 인하여, 연속적인(continuous) 영상 데이터의 특징 벡터를 분류하는데 가중치로 사용할 경우 동일한 장소에 대한 특징 벡터 간에 일관성을 부여하는 데 도움이 될 것으로 예상된다.

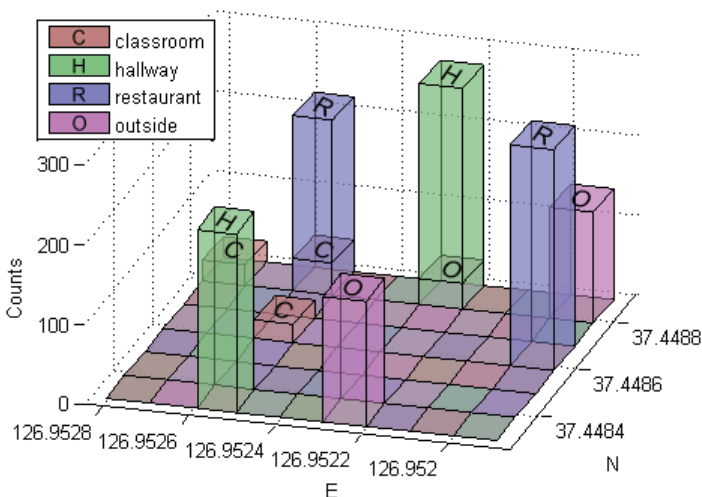


그림 1. 위치별 GPS 데이터 히스토그램

표 3. 분석 데이터별 장소 인식 결과

분석 데이터	정확도 (%)	
	정면방향	사선방향
영상	78.6296	80.0718
영상 + GPS	81.8287	80.9964

전처리를 거친 GPS 데이터를 가지는 영상 데이터의 특징 벡터를 재추출하고, 가중치를 적용하여 장소 인식을 수행한 결과는 표 3과 같다. GPS 데이터를 가중치로 사용하여 분류한 정확도가 영상 데이터만을 사용하여 분류한 것보다 높게 나타나는 것을 확인할 수 있다.

4. 결론 및 향후 연구

스마트폰에 장착된 카메라와 GPS 센서를 통해 입력받은 영상 데이터와 GPS 좌표를 통합하여 장소 인식을 시도하였다. 기존 GPS 기반 위치 인식 시스템과 달리 비전 기반 위치 인식의 경우 명시되지 않은 위치에 대해서 논리적 위치 정보를 인식해낼 수 있으며, GPS 데이터를 가중치로 사용함으로써 분류 정확도를 다소 높일 수 있음을 확인하였다.

또한 GPS 데이터를 이용할 경우 광범위한 영역에서의 위치 인식이 가능하므로, 향후 한 건물이 아닌 다수의 건물 내 장소를 인식함으로써 보다 효과적인 LBS를 제공할 수 있을 것으로 기대된다.

참고 문헌

- [1] 박종태, 이위혁, 조영훈, 나재욱, "유비쿼터스 센서 네트워크에서 위치 측정 기술," 전자공학회지 제 32권 7호, pp. 81-94, 2005.
- [2] L. Blazevic, JY. Le Boudec, and S. Giordano, "A Location-Based Routing Method for Mobile Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, Vol. 4, No. 2, pp. 97-110, 2005.
- [3] DC. Herath, S. Kodagoda and G. Dissanayake, "A Two-tier Map Representation for Compact Stereo-Vision-based SLAM," *Robotica*, Vol. 30, No. 2, pp. 245-256, 2012.
- [4] A. Stelzer, H. Hirschmüller and M. Görner, "Stereo-vision-based navigation of a six-legged walking robot in unknown rough terrain," *The International Journal of Robotics Research*, Vol. 31, No. 4, pp. 381-402, 2012.
- [5] A. Hornung, KM. Wurm, M. Bennewitz, "Humanoid robot localization in complex indoor environments," *Proc. of International Conference on Intelligent Robots and Systems*, pp. 1690-1695, 2010. 10.
- [6] DG. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.
- [7] CC. Chang and CJ. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, Article 27, pp. 1-27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>