
실외 주행 로봇의 위치 추정을 위한 3 차원 물체 인식

3D Object Recognition for Localization of Outdoor Robotic Vehicles

백승민, Seung-Min Baek*, 김재웅 Jaewoong Kim**, 이장원 Jangwon Lee**, Lu Zhaojin**,
이석한 Sukhan Lee***

요약 본 논문에서는 지능형 무인 운송 시스템의 위치 인식 문제를 풀기 위한 방법으로, 실외 환경에서 보이는 건물 혹은 건물 입구의 상대적 위치와 자세 추정이 가능한 파티클 필터 기반 3 차원 물체 인식 방법을 제안한다. 제안하는 파티클 필터에 기반한 인식 시스템은 다양한 인식 증거들을 연속 영상에서 융합 및 모델 매칭을 함으로써 강인한 3 차원 물체 인식 및 자세 추정이 가능하다. 제안하는 방법은, 적합한 인식 증거들을 수집/선택하고, 다양한 인식 증거들로부터 나타나는 인식 대상의 자세를 3 차원 공간 상의 확률적인 파티클로 표현하며, 파티클 필터링을 통하여 연속 영상 상의 다양한 인식 증거들을 융합하는 것을 특징으로 한다. 스테레오 카메라를 이용한 실험을 통하여, 제안하는 방법이 실외 건물의 기하학적 특징을 인식 증거로 활용한 효율적인 3 차원 인식 및 자세 추정을 수행하는 것은 보여준다.

Abstract In this paper, to solve localization problem for out-door navigation of robotic vehicles, a particle filter based 3D object recognition framework that can estimate the pose of a building or its entrance is presented. A particle filter framework of multiple evidence fusion and model matching in a sequence of images is presented for robust recognition and pose estimation of 3D objects. The proposed approach features 1) the automatic selection and collection of an optimal set of evidences 2) the derivation of multiple interpretations, as particles representing possible object poses in 3D space, and the assignment of their probabilities based on matching the object model with evidences, and 3) the particle filtering of interpretations in time with the additional evidences obtained from a sequence of images. The proposed approach has been validated by the stereo-camera based experimentation of 3D object recognition and pose estimation, where a combination of photometric and geometric features are used for evidences.

핵심어: *3D object recognition, pose estimation, multiple evidences, Out-door navigation*

This work is supported by the Intelligent Robotics Program, one of the 21st Century Frontier Program funded by the Ministry of Commerce, Industry and Energy of Korea. This work is also supported by Science and Technology Program of Gyeonggi Province as well as in part by the Sungkyunkwan University.

*주저자 : 성균관대학교 지능시스템연구센터 연구교수 e-mail: smbaek@ece.skku.ac.kr

**공동저자 : 성균관대학교 전기전자컴퓨터공학과 대학원생 e-mail: create@skku.edu, jwlee31@ece.skku.ac.kr, Luzhaojin@gmail.com

***교신저자 : 성균관대학교 전기전자컴퓨터공학과 교수 e-mail: Lsh@ece.skku.ac.kr

1. 서론

최근 미국, 유럽 등 선진국에서는 지능형 무인 운송 시스템에 대한 연구를 많이 수행하고 있으며, 그 중 일부의 결과는 실버 타운, 대학교 캠퍼스, 관광지 등 실제 사람들이 공존하는 환경에서 테스트 되고 있다[5].

실외 주행 시스템의 위치 추정을 위해서는 GPS 가 많이 적용되고 있다. GPS 는 누적위치 오차는 없지만 위치 측정의 분산이 상대적으로 크기 때문에, 특정 건물이나 건물 입구 앞에 멈추는 수 미터 이내의 정확성을 요구하는 위치 인식에는 적합하지 못하다. 이러한 환경 상의 상대적인 위치 인식을 위해 카메라 정보의 영상 처리를 이용한 방법들이 다수 제안되고 있다[1].

본 논문에서는 지능형 무인 운송 시스템의 위치 인식 문제를 풀기 위한 방법으로, 스테레오 카메라의 2 차원과 3 차원 영상을 활용하여 카메라와 인식 대상 물체 사이의 상대 위치 및 자세를 구하고, 이를 자기 위치 추정에 적용하는 방법을 제안한다. 제안하는 파티클 필터에 기반한 인식 시스템은 다양한 인식 증거들을 연속 영상에서 융합 및 모델 매칭을 함으로써 강인한 3 차원 물체 인식 및 자세 추정이 가능하다. 제안하는 방법은, 적합한 인식 증거들을 수집/선택하고, 다양한 인식 증거들로부터 나타나는 인식 대상의 자세를 3 차원 공간 상의 확률적인 파티클로 표현하며, 파티클 필터링을 통하여 연속 영상 상의 다양한 인식 증거들을 융합하는 것을 특징으로 한다. 스테레오 카메라를 이용한 실험을 통하여, 제안하는 방법이 실외 건물의 기하학적 특징을 인식 증거로 활용한 효율적인 3 차원 인식 및 자세 추정을 수행하는 것은 보여준다.

2. 3 차원 물체 인식 프레임워크

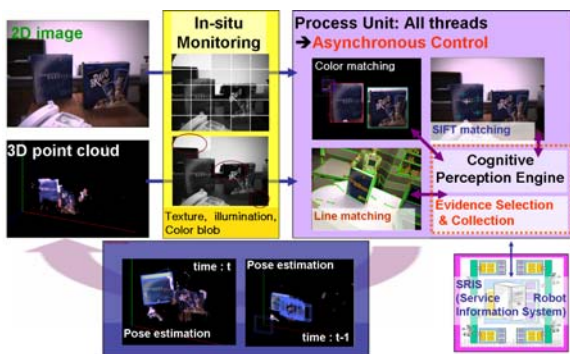


그림 1. 3 차원 물체 인식 프레임워크

본 논문에서는 [2]에서 제안한 물체 인식 방법에서 인식 증거로서 선, SIFT 를 적용하여 실외 건물인 식에 적합하도록 제안하였다. 연속 영상을 이용한 확률적인 융합 및 필터링은 [3]에서 소개한 방법을 따라 구현하였다.

그림 1 에 전체적인 3 차원 물체 인식 방법을 블록도로 소개 하였다. 2 차원 영상과 3 차원 거리 영상이 스테레오

카메라 시스템으로부터 주어지면, In-situ monitoring 부분에서 현 상황에 맞는 인식 증거의 조합을 선택한다. 예를 들어 환경이 기준보다 어둡다면 기하학적인 형상을 나타내는 선을 인식 증거로 선택하고, 텍스처가 풍부하게 검출되는 거리에서는 SIFT 를 인식 증거로 선택한다. 연산 자원이 충분하다면 다양한 조합을 선택할 수도 있다. 현 시점에서 활성화된 영상처리 경로에 따라 특징 추출과 모델 매칭이 비동기로 처리 되며, 각 인식 증거로부터 나타나는 모델 매칭 결과는 확률을 갖는 파티클로 표현된다. 최종적으로 파티클들은 필터링을 거치면서 이전 연속 영상의 정보 또는 현 영상에서 나타난 다른 인식 증거들과 융합되어 3 차원 공간 상의 물체 존재 확률 분포를 구한다.

3. SIFT 기반 모델 매칭

SIFT(Scale Invariance Feature Transform)는 2 차원 영상 기반 물체 인식에 많이 쓰이는 특징이다[6]. 그림 3 에 주어진 2 차원 영상으로부터 추출된 SIFT 특징점들을 보여주고 있다. (원의 크기는 구해진 scale, 중심점은 2 차원 위치를 나타낸다.)



그림 2. 추출된 SIFT 특징점들

본 논문에서는 [3]에서와 같이 찾고자 하는 건물의 정면 2 차원 영상과 SIFT 특징점 집합을 모델로서 저장해두고, 카메라로부터 얻어진 측정 2 차원 영상과 매칭을 하여 매칭 SIFT 특징점 집합을 찾아낸다. 이 후 이 매칭 특징점 집합을 이용하여 모델 상의 3 차원 정보와 측정된 3 차원 정보의 좌표변환 관계를 구하고, 이 변환관계가 바로 물체의 6 자유도 위치와 자세가 된다. 이렇게 매칭된 결과가 파티클로 표현되고, 각 파티클은 매칭 과정에서 계산된 오차와 특징점들의 숫자에 따라 확률 값을 갖게 되어 인식 프레임워크 상의 필터링으로 전달된다. 더 자세한 절차는 [2]와 [3]에 잘 나타나 있다.

4. 3 차원 선 기반 모델 매칭

4.1 3 차원 평행선의 추출

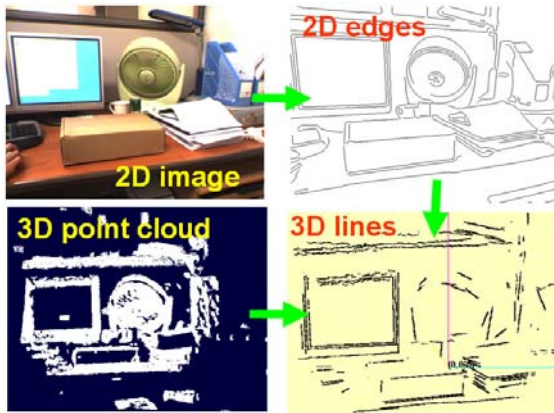


그림 3. 3차원 선 추출 과정

스테레오 카메라를 이용하여 얻어진 2 차원 영상과 3 차원 거리 영상으로부터 3 차원 선을 구하는 과정을 그림 3 에 나타내고 있다[4]. 2 차원 영상에서 Canny Edge Detector 를 이용하여 Edge Image 를 구한 후, 2 차원 선을 추출한다. 2 차원 선 주변의 화소에 해당하는 3 차원 거리 정보를 획득하여, Least Square 방식으로 3 차원 선을 구한다. 구해진 3 차원 선 간의 각도 차를 계산하여 평행선 조합들을 찾아낸다. [4]에서는 현재 환경에 맞게 튜닝된 고정적인 상수 값을 이용하여 위의 과정을 처리하였지만, 본 연구에서와 같이 물체가 카메라로부터 먼 경우에는 스테레오 카메라의 특성상 거리 정보의 불확실성이 인식 거리의 변화에 따라 크게 변한다. 따라서 불확실성을 고려한 상수 값을 찾는 방법이 매우 중요하다.

스테레오 카메라는 오른쪽과 왼쪽 카메라의 영상 평면의 Epipolar line 상의 disparity 를 이용하여 거리를 구하게 되고 disparity 변화당 측정거리의 변화(range resolution)는 다음과 같이 나타난다.

$$\frac{\Delta r}{\Delta d} = \frac{r^2}{b \cdot f} \quad (1)$$

여기서 r 은 카메라부터의 거리 이며, Δr 은 거리 해상도, Δd 는 disparity 변화, b 는 카메라 사이의 base line length, f 는 초점 거리를 의미한다.

본 논문에서는 위의 거리 해상도를 활용하여 3 차원 선을 찾을 때, 점과 선의 거리를 결정하는 반지름을 정하였다.

$$line_radius = \frac{a^2}{b \cdot f} \quad (2)$$

여기서 a 는 선을 이루는 점들의 평균 거리 값이다.

선의 평균 거리의 제곱에 비례하여 측정되는 선의 반지름이 커지게 되므로, 가까이 있는 선을 구할 때는 매우 근접한 점들만 고려하게 되고, 같은 scene 상의 멀리 있는 선을 찾을 때는 점들의 분포가 흩어져 있더라도 하나의 선을 이루는 점들로 판단한다. 즉, 측정 센서의 불확실성을 고려한 3 차원 선의 추출이 가능하며, 별도의 튜닝 과정이 생략된다.

4.2 평행선 모델 매칭

이전 단계에서 찾아낸 평행선 조합들 중에서 모델과 유사한 조합들을 찾아서 확률적 표현을 만들어 낸다. 식 (2)의 반지름 값은 여기에서도 적용되어 측정된 평행선 사이의 거리와 모델 평행선 사이의 거리가 측정된 평행선의 반지름 값의 합 보다 작을 경우 모델 매칭이 가능한 평행선 조합으로 판단한다. 측정된 선들은 일반적으로 모델의 선보다 짧게 끊어져서 나타나는 경우가 많으며, 이런 경우는 평행선의 방향으로 미끄러지면서 다수의 매칭 표현을 만들어 낸다. 물체의 자세를 연산하기 위해서는 [3]에서 SIFT 매칭 점 조합을 이용한 물체 자세 추정 방법과 같이 측정된 평행선의 4 점과 모델 평행선의 4 점을 매칭 조합으로 Quaternion 을 이용한 자세 추정을 한다. 단, 이 때 측정된 짧은 평행선은 모델 평행선의 길이로 확장하여 자세를 구한다. [2]와 [4]에서 ICP 를 적용하여 매칭 자세를 구한 방식과 달리, 제한한 모델 매칭 방식은 실시간 연산이 가능하고, 한 평행선 조합으로부터도 평행선 방향으로 다양한 표현이 나타나므로 좀더 정밀한 자세 추정이 가능한 장점이 있다.

5. 실험결과

실험을 위하여 그림 4 와 같이 base line length 는 120cm, 초점 거리 8mm 인 렌즈를 장착한 스테레오 카메라 시스템을 구성하였고, 폭 18m, 높이 12m 인 건물의 위치와 자세를 정면/오른쪽/왼쪽으로부터 약 40~50m 거리에서 scene 을 얻어서 추정해 보았다.



그림 4. Base line length 변경이 가능한 스테레오 카메라

본 논문에서 제안한 3 차원 평행선 매칭 방법을 적용한 결과 그림이 그림 5~7 에 나타나 있다. 그림에서 획득된 2 차원/3 차원 영상을 보여주고 있으며, 파란색 와이어 프레임들은 3 차원 평행선으로부터 추정된 건물의 추정 자세를 나타낸다.

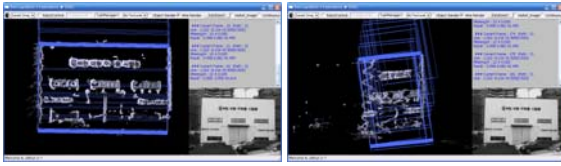


그림 5. 3차원 평행선을 이용한 자세 추정 결과
(건물 정면 scene)



그림 6. 3차원 평행선을 이용한 자세 추정 결과
(건물 우측 scene)



그림 7. 3차원 평행선을 이용한 자세 추정 결과
(건물 좌측 scene)

왼쪽의 그림들은 건물 point cloud 를 정면 방향에서 본 경우이고, 오른쪽의 그림들은 view point 를 측면으로 돌려 본 경우이다. 측면 그림을 보면 측정된 point cloud 가 카메라 방향으로 많이 흔들리는 모습이 보이는데, 이 문제 때문에 3차원 선을 거리에 적용하는 불확실성의 범위를 설정할 수 있어야 한다.

그림 5~7 의 결과를 살펴보면 건물의 추정 위치/자세가 일부 틀어진 것들도 있지만 흔들리는 point cloud 에 비해서는 추정 결과가 양호한 것을 알 수 있다.

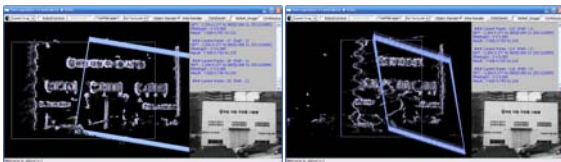


그림 8. SIFT 를 이용한 자세 추정 결과
(건물 정면 scene)

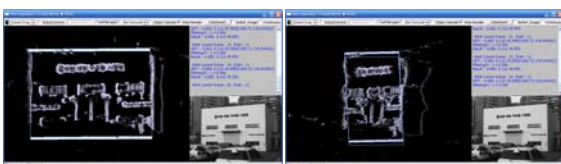


그림 9. SIFT 를 이용한 자세 추정 결과
(건물 우측 scene)

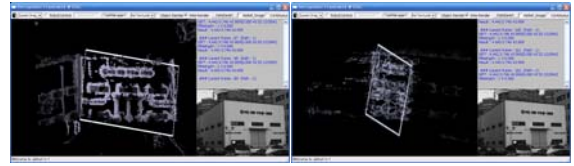


그림 10. SIFT 를 이용한 자세 추정 결과
(건물 좌측 scene)

그림 8~10 은 SIFT 매칭을 이용하여 건물의 위치와 자세를 추정한 결과이다. SIFT 는 텍스트가 잘 나타나는 영역에서 많이 매칭이 이루어지기 때문에 선을 이용한 매칭과는 다른 추정 결과를 보인다. SIFT 를 이용한 매칭에서도 측정된 point cloud 에 의존하여 3차원 위치/자세를 구하게 되므로 역시 불확실성의 영향을 받는다. 그림 9~10 에서는 추정 위치가 비교적 양호하지만, 그림 8 의 정면 영상의 경우는 잘못된 매칭이 이루어져서 추정 결과가 실제 위치와 차이가 있는 것을 볼 수 있다.



그림 11. SIFT 와 3차원 평행선을 이용한 자세 추정 결과
(건물 정면 scene)

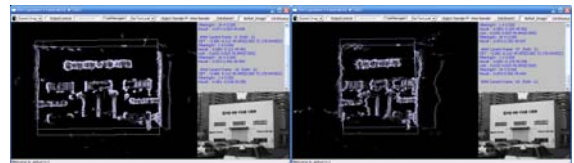


그림 12. SIFT 와 3차원 평행선을 이용한 자세 추정 결과
(건물 우측 scene)

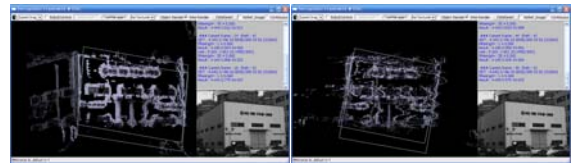


그림 13. SIFT 와 3차원 평행선을 이용한 자세 추정 결과
(건물 좌측 scene)

그림 11~13 은 SIFT 매칭을 통해 얻은 파티클들과 3차원 평행선 매칭을 통해 얻은 파티클들을 확률적으로 융합하여 가중치가 높은 파티클들을 표현한 결과이다. 그림 11 의 정면 scene 에서는 SIFT 로부터 얻어진 실제 위치에서 많이 벌어진 곳에 추정되었던 파티클은 가중치가 작아져서 사라지고 3차원 평행선 매칭결과의 파티클들이 더 큰 가중치로 남아있는 것을 알 수 있다. 그림 12 의

건물 우측 scene 에서는 SIFT, 선을 이용한 매칭 결과가 모두 안정적이라 융합 결과도 실제 위치에 가깝게 나타났다. 그림 13 의 좌측 scene 의 경우는 반대로 선을 이용한 매칭보다는 SIFT 를 이용한 매칭이 좀더 안정적이었고 융합 결과도 실제 위치에 가까운 파티클들이 가중치가 높은 것을 볼 수 있다.

6. 결론

본 논문에서는 [2]와 [3]에서 제안하였던 3 차원 물체 인식 프레임워크를 실외 주행 로봇의 3 차원 위치 추정에 적용하는 방법을 제안하였다. 제안한 방법은 실외 건물의 기하학적 크기와 texture 를 모델로 저장해 두고, 이를 이용하여 로봇의 상대 위치를 추정할 수 있기 때문에 관광지, 대학 캠퍼스, 광장 등에서 로봇이 자율 주행 시 위치 인식 문제를 해결할 수 있다.

또한 본 논문에서는 스테레오 카메라의 거리 측정 불확실성을 체계적으로 고려한 3 차원 선 추출 방법을 제안하였으며, 평행선을 이용한 매칭을 통하여 건물의 6 자유도 위치와 자세 추정이 가능함을 보였다.

SIFT 매칭을 이용한 추정 결과와 3 차원 평행선을 이용한 추정 결과가 다른 특성을 나타내는 것을 알 수 있었고, 이 두 결과를 융합하면 좀 더 안정적인 결과를 얻을 수 있었다. 다만, 각 scene 당 SIFT 매칭이 약 0.8 초, 3 차원 평행선 매칭이 약 1 초 처리 시간이 걸리므로 융합결과를 얻는 데까지는 약 1.8 초 정도가 소요된다(Centrino 2Hz 기준). 이는 성능과 처리 시간 사이의 trade-off 로 상황에 따른 적절한 판단이 필요하다.

본 위치 인식 방법은, 누적 위치 오차는 없지만 위치 인식 결과의 분산이 비교적 큰 저가의 GPS 를 통한 위치 인식 결과와 융합하였을 때, 위치 정밀도와 방향 자유도를 모두 향상시킬 수 있을 것으로 기대 된다. 그러므로 추후에는 GPS, 내부 odometry 등의 다양한 센서 정보와 융합하는 연구를 진행 할 예정이다.

참고문헌

- [1] W. Zhang and J. Kosecka, "Localization Based on Building Recognition," 2005 IEEE Computer Society Conference on Computer vision and Pattern Recognition (CVPR' 05), 2005.
- [2] Jeihun Lee, SeungMin Baek, Changyun Choi and Sukhan Lee, "A Particle Filter Based Probabilistic Fusion Framework for Simultaneous Recognition and Pose Estimation of 3D Objects in A Squence of Images," 2007 International conference on Advanced Robotics (ICAR' 07), 2007.
- [3] Sukhan Lee, Seongssoo Lee, Jeihun Lee, Dongju Moon, Eunyoung Kim, and Jeonghyun Seo, "Robust Recognition and Pose Estimation of 3D Objects Based on Evidence Fusion In a Sequence of Images," 2007 International Conference on Robotics and Automation(ICRA' 07), 2007.
- [4] Samuel H. Chang, Sukhan Lee, Dongju Moon, Woongmyung Kim and Yeunghak Lee, " Model Based 3D Object Recognition Using Line Features," 2007 International conference on Advanced Robotics (ICAR' 07), 2007.
- [5] P. Martinet, B. Thuilot, J. Bom, "Autonomous Navigation and Platooning using a Sensory Memory", invited session in Workshop "Safe Navigation in Open and Dynamic Environments: Autonomous Systems versus Driving Assistance Systems" at International IEEE Conference on Intelligent Robots and Systems, IROS'06, Beijing, China, October 2006.
- [6] D. Lowe. "Object recognition from local scale invariant features," In Proc. 7th International Conf. Computer Vision (ICCV' 99), pp. 1150 1157, Kerkyra, Greece, September 1999.