

다중 카메라 기반 휴머노이드 로봇 자세 추정

Humanoid robot pose estimation based on multiple cameras

신원국^{1,2} · 김준식^{1†}

Wonkook Shin^{1,2}, Jun-Sik Kim^{1†}

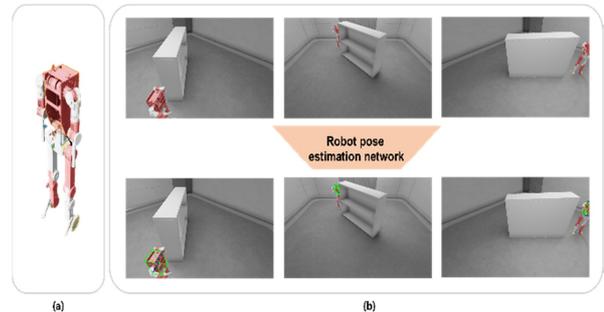
Abstract: Accurate robot localization is essential for a robot to move in a space. The location of an articulated robot can be defined by its body's pose in the world coordinate system. For articulated robots, the movement of joints and obstacles often lead to occlusion challenges. The method of estimating keypoints using a voting algorithm based on a single camera sensor can robustly estimate the pose of a partially occluded object. However single camera system has limitations in detecting the entire space due to its limited measurement range. In this paper, we propose a method to estimate the pose of a humanoid robot's body based on multiple camera systems to overcome these issues. Experimental results confirmed that the outputs of multiple camera sensors are effectively fused, and the pose can be robustly estimated even in situations where the robot is occluded by obstacles.

Keywords: Robot pose estimation, Voting based pose estimation, Multiple camera system

1. 서론

로봇이 공간에서 자유롭게 이동하기 위해서는 로봇의 위치 추정이 필수적이다. 특히, 다관절 로봇의 위치는 공간 좌표계를 기준으로 로봇 몸체의 자세로 정의된다 [1]. 휴머노이드 로봇은 다관절 운동과 주변 장애물로 인해 몸체가 가려지는 문제가 발생할 수 있다. 최근에는 카메라 센서를 활용하여 부분적으로 가려진 물체에 대한 자세 추정 방법에 관한 연구가 진행되고 있다. Voting 알고리즘을 활용한 키포인트 추정 방법은 물체의 일부분이 가려진 상황에서도 강건하게 자세를 추정할 수 있음을 보여준다 [2].

단일 카메라 센서는 제한된 측정 범위로 인해 감지 가능한 영역이 한정적이다. 휴머노이드 로봇의 자세를 전체 공간에서 추정하기 위해서는 다수의 카메라가 필요하다. 이전에 연구된 voting 알고리즘을 기반으로 한 물체 자체 추정 방법은 주로 단일 카메라 시스템에 의존하고 있어, 다수의 카메라 시스템에 대한 적용을 위해서는 별도의 접근 방식이 필요하다.



[Fig. 1] Humanoid robot pose estimation using three cameras.
(a) humanoid robot model. (b) Pose estimation process.

본 연구에서는 휴머노이드 로봇의 자세를 추정하기 위해 다수의 카메라 시스템을 활용한 새로운 방법을 제안한다. 다수의 카메라 센서 출력을 효과적으로 퓨전하여 로봇이 장애물에 가려진 상황에 대해서도 강건하게 자세를 추정할 수 있음을 실험을 통해 확인하였다. [Fig. 1] 은 휴머노이드 로봇 및 Isaac SIM 시뮬레이터 [3]로 구축된 가상 환경에서 3개의 카메라 센서를 활용하여 휴머노이드 로봇의 자세를 추정한 결과를 보여준다.

2. 본론

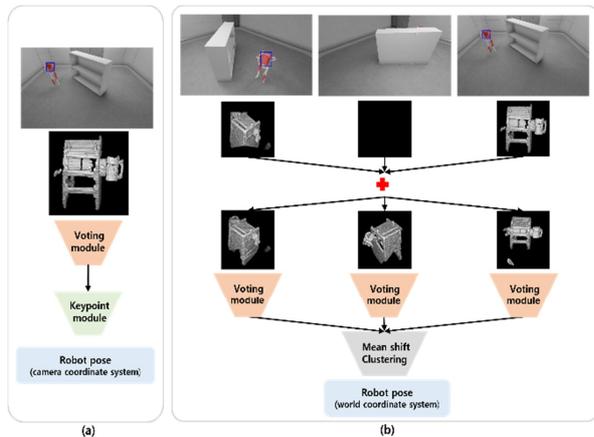
2.1 Voting 알고리즘 기반 자세 추정 네트워크 구조

G2L-Net [2] 은 단일 카메라 기반의 voting 알고리즘을 활용하여 물체의 자세를 추정한다. [Fig. 2(a)] 는 voting module 과 keypoint module 로 구성된 G2L-Net 구조를 보

※ This work was supported by the Korea Institute of Science and Technology (KIST) Institutional Program (Project No. 2E32302).

1. Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology (KIST), Seoul, Korea
2. Department of Electrical Engineering, Korea University, Seoul, Korea

† Principal Researcher, Corresponding author: Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology (KIST), Seoul, Korea (junsik.kim@kist.re.kr)



[Fig. 2] Robot pose estimation network architecture. (a) G2L-Net model architecture. (b) Multiple camera-based robot pose estimation network architecture.

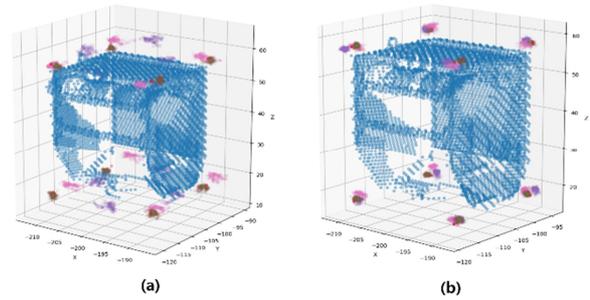
여준다. Voting module 은 입력 포인트 클라우드에 대해 키포인트를 가리키는 방향과 크기를 갖는 벡터를 추정한 후, 키포인트로 추정할 수 있는 지점들을 voting 한다. Keypoint module 은 voting 된 영역을 기반으로 키포인트를 추정 후, 최적의 물체의 자세를 연산한다.

2.2 포인트 클라우드 퓨전

[Fig. 2(b)] 는 다수의 카메라 시스템을 활용한 voting module 의 입력 포인트 클라우드 퓨전 과정을 보여준다. 단일 카메라의 경우 제한된 시야에서만 포인트 클라우드를 추출하지만, 다수의 카메라를 활용함으로써 다양한 시야에서 포인트 클라우드를 추출할 수 있는 장점을 갖는다. 실험을 통해 voting module의 경우 퓨전된 포인트 클라우드를 입력으로 사용할 때, voting 영역이 집중되는 경향을 보임을 확인하였다. [Fig. 3] 은 3개의 카메라 센서를 활용하여 입력 포인트 클라우드를 퓨전하지 않은 경우와 퓨전한 경우에 대한 voting module 의 출력을 비교한 결과를 보여준다.

2.3 비지도 학습 방식의 클러스터링

[Fig. 2(b)] 는 keypoint module을 비지도 학습 방식의 클러스터링 알고리즘으로 대체한 구조를 보여준다. Keypoint module은 학습 시 설정된 좌표계를 기준으로 물체의 자세를 추정한다. 공간 좌표계는 환경에 따라 다르게 설정되며, 특정 공간 좌표계를 기준으로 keypoint module을 학습시킬 경우, 해당 환경에서만 자세를 추정할 수 있는 문제가 발생한다. 이를 해결하기 위해, 비지도 학습 방식의 클러스터링 알고리즘을 활용하여 임의의 공간 좌표계를 기준으로 키포인트를 추정하고, 로봇의 자세를 추정할 수 있도록 한다.



[Fig. 3] Voting module results. (a) Voting result for individual point cloud without fusion. (b) Voting result with point cloud fusion.

[Table 1] Humanoid robot pose estimation results.

Model types	Rotation error	Translation error
Non fusion model	13.028 °	2.411 cm
Fusion model	5.079 °	1.838 cm

2.4 시뮬레이터를 활용한 실험

3개의 카메라 센서와 휴머노이드 로봇을 가릴 수 있는 장애물이 포함된 가상 환경을 설정하였다. 평가를 위해 로봇이 장애물 주위를 이동하며 가려지는 상황이 포함된 180장의 이미지 시퀀스 데이터를 추출하였다. 다수의 카메라에서 추출된 포인트 클라우드를 퓨전하여 로봇의 자세를 추정하는 모델과 퓨전하지 않고 자세를 추정하는 모델을 선정하고 추정 결과를 평가하였다. [Table 1]은 두 모델에 대한 휴머노이드 로봇 자세 추정 결과를 보여준다.

3. 결론

본 연구에서는 다수의 카메라에서 관측된 정보를 토대로 로봇의 자세를 추정하는데 효율적인 퓨전 방식을 제안하고, 효과를 입증하였다. 향후 제안한 방식을 실제 환경에 적용하여 평가를 수행할 예정이다.

References

- [1] Bultmann, Simon and Memmesheimer, Raphael and Behnke, Sven. "External Camera-based Mobile Robot Pose Estimation for Collaborative Perception with Smart Edge Sensors." *arXiv preprint arXiv:2303.03797*, 2023.
- [2] Chen, Wei and Jia, Xi and Chang, Hyung Jin and Duan, Jiming and Leonardis, Ales. "G2l-net: Global to local network for real-time 6d pose estimation with embedding vector features." *IEEE/CVF conference on computer vision and pattern recognition*, pages 4233-4242, 2020.
- [3] NVIDIA nvidia issac sim, [Online]. <https://developer.nvidia.com/isaac-sim>. Accessed: 2023-11-14.