딥 러닝 기반의 신체방향 감지를 통한 최적의 로봇 위치 선정

최진영[0] 이범진 장병탁
서울대학교 바이오지능 연구실
(jychoi, bjlee)@bi.snu.ac.kr, btzhang@snu.ac.kr

Best Robot Positioning based on Body Orientation Estimation using Deep Learning Method

Jin-young Choi[1○], Beom-Jin Lee[2], Byoung-Tak Zhang[12]
[1] Interdisciplinary Program in Cognitive Science, [2] School of Computer Science and Engineering,
Seoul National University

요 약

최근 이미지 인식 분야에서 뛰어난 성능을 내고 있는 CNN(Convolutional Neural Network)을 이용하여 로봇에 장착된 카메라의 RGB 이미지에서 사람의 위치와 향하고 있는 방향을 인식할 수 있는 시스템과 이를 기반으로 한 로봇의 최적 위치 선정 시스템을 제안하다. 벤치마킹용 데이터와 직접 실제 환경에서 수집한 데이터를 사용하여 CNN을 학습하였다. 제안하는 방향 인식 시스템은 기존의 방법(SVM)에 비해 큰 성능 향상과 일반화 능력을 보였고, 위치 재설정 시스템은 사용자를 마주보도록 로봇을 움직여 로봇의 얼굴인식의 성공률을 크게 향상시켰다. 이러한 결과들을 통해 서비스 로봇의 사용자 중심 서비스 수행 능력 향상에 기여할 수 있을 것으로 기대된다.

1. 서 론

사람의 방향을 인지하는 능력은 최근 부각되기 시작한 서비스 로봇이 사용자를 마주보고 서비스를 제공하거나 사용자의 감정, 행동 등을 인식하는데 있어 필수적이다. 그러나 로봇에 장착된 카메라의 RGB 이미지에서 사람의 위치를 찾고 추적하는 등의 시스템이 상대적으로 많이 연구된 반면 사람의 방향까지 고려하여 로봇을 이동시키는 연구는 상대적으로 연구가 많이 이루어지지 않았다. 이에 우리는 최근 이미지 인식 분야에서 뛰어난 성능을 내고 있는 CNN(Convolutional Neural Network)을 이용하여 카메라 이미지에서 사람의 위치와 향하고 있는 방향까지 인식할 수 있는 시스템과 이를 바탕으로 사용자를 마주보도록 최적의 로봇 위치를 선정할 수 있는 알고리즘을 제안하다. 이 논문의 구성은 첫째, 사람의 방향을 인식하는 기법들과 로봇의 최적 위치 선정에 대한 기존 연구를 간략히 소개하고, 둘째, 본 시스템의 구성과 구현 방법을 설명하고, 셋째, 실험과 그 결과를 설명한 뒤 마지막으로 논의와 앞으로의 연구 계획에 대해 논하는 것으로 이루어져 있다.

2. 기 존 연 구

사람의 방향을 인식하는 시스템에 대한 연구들 중 많은 수는 신체에 부착하는 센서[1], 깊이 영상(Depth image)을 통해 얻은 포인트 클라우드[2] 등을 이용했다. 이러한 정보들을 사용하기 위해서는 잘 짜여진 실험실 환경이 필요했다. 또한, 기존 연구들은 Random Forest나 Support Vector Machine 등의 기계학습 기법을 사용하여 분류기의 성능에도 한계가 있었다.

우리는 이런 점들을 극복하기 위해 딥 러닝 방식의 CNN을 사용하여 특이점 추출과 고성능의 분류기를 이미지와 라벨만을 이용하여 한번에 학습하는 시스템을 제안하다.

한편, 로봇의 위치를 재조정하는 시스템에 대한 연구들은 주로 주변이나 장애물을 인식하고 여러 대의 카메라 중 적절한 카메라를 선택하는 연구[3], 사용자의 위치 분포를 미리 학습한 후 적절한 위치를 선택하는 연구[4] 등이 이루어졌다. 우리는 사용자를 기준으로 미리 정한 규칙에 따라 후보 위치를 선정하고 사용자의 방향과 장애물의 유무, 이동거리 등을 반영한 활성값(Utility Value)를 계산한 뒤 가장 점수가 높은 곳으로 로봇을 이동시키는 방법을 사용했다. 우리의 방법과 비슷한 연구로는 [5]가 있다. [5]는 깊이 영상과 이를 통해 얻은 관절 위치 정보를 사용해 여러 위치의 활성값 (utility value)을 계산하고 가장 점수가 높은 위치로 로봇을 이동시키는 방법을 사용하였다. 우리의 연구는 [5]와 달리 방향 인식에 깊이 센서를 필요로 하지 않고, 주변 환경에 더 강건하며 주변의 장애물까지 고려하여 시스템의 활용성을 더욱 향상시켰다.

3. 방 법 론

이 장에서는 우리가 제안하는 시스템의 구조와 구현 방법에 대해 논한다. 시스템은 세 개의 모듈로 구성되어 있다. 첫 번째 모듈은 사용자를 인식하는 모듈이다. 두 번째 모듈은 사용자의 신체 방향을 감지하는 모듈이고, 세 번째 모듈은 사용자의 방향을 바탕으로 최적의 로봇 위치를 계산하는 모듈이다. 아래에서 세 모듈의 구조와 구현 방법에 대해 설명하도록 한다. 모든 모듈은 python으로 프로그래밍 하였고, 신경망의 구현에는 tensorflow를 썼으며, 이미지 처리는 OpenCV API를 이용했다.

3.1 사용자 인식

RGB 이미지에서 사람을 인식하기 위해, 우리는 YOLO [6] 알고리즘을 사용했다. 이 알고리즘은 한 개의 CNN을 이용해 물체의 위치와 종류를 동시에 예측하여 처리 속도가 매우 빠르다는 장점이 있다. 원래의 알고리즘은 사람뿐 아니라 자동차, 항공기 등 20가지

물체를 인식할 수 있으나 우리는 이 중 사람만을 추출하여 사용했다.

3.2 방향 인식

방향 인식 모듈은 사용자 인식 모듈에 의해 선택된 박스를 CNN을 이용해 그림 1과 같은 8개의 방향 중 하나로 분류한다. 그림 2는 방향 인식 모듈에 사용된 신경망의 구조이다. 우리는 CNN에서 많이 사용하는 max-pooling을 사용하지 않았는데 이는 방향 인식에는 각각의 부분이 어디에 있는지(손의 위치, 얼굴의 위치 등)가 중요한데 max-pooling을 사용하면 이 위치 정보가 사라질 것이라고 판단했기 때문이다. 입력 이미지는 32*32픽셀로 크기를 조정하여 사용했는데 크기를 줄여도 방향을 인식하는데 필요한 특성들은 여전히 구분이 가능하고 작은 신경망을 적용하여 처리 속도를 빠르게 할 수 있기 때문이다. 또한 큰 신경망을 사용할 때의 문제인 과적합 문제를 피하기 위한 목적도 있다.

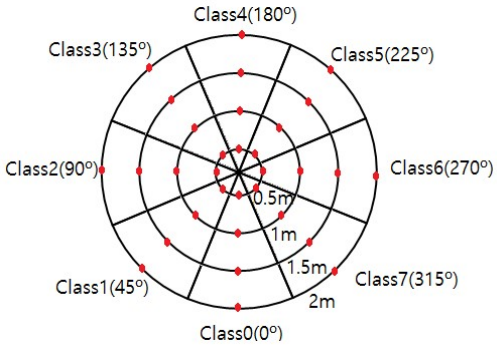

그림 1. 방향의 분류와 재배치 후보

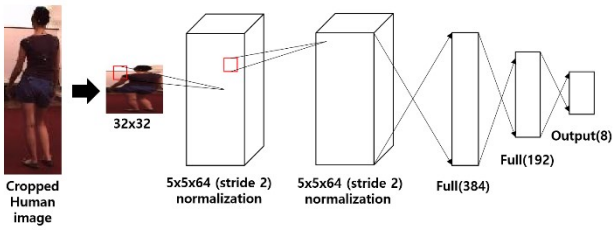

그림 2. 방향인식 CNN의 구조

3.3 최적 위치 선정 방법

최적 위치 선정 모듈은 사람의 위치와 방향을 이용하여 되도록 사용자의 정면을 관찰할 수 있는 최적의 위치를 계산한다. 우리는 그림 1의 점들과 같이 사용자를 기준으로 32곳의 후보 위치를 정한 뒤 각 위치에 식 (1)과 같은 활성화 함수 (utility function) 을 사용하여 각 위치의 점수를 계산했다.

$$U(p) = Orientation(t) \cdot Distance(p,r) \cdot Radius(p) \cdot Occupancy(p) \cdot Obstacle(p,t) \quad (1)$$

식(1)에서 p는 후보위치, t는 사용자, r은 로봇을 뜻하며, *Distance(p,r), Occupancy(p), Obstacle(p,t)*는 아래와 같이 정의된다.

$$Distance(p,r) = \max_{r'} |p - r'| - |p - r| \quad (2)$$

$$Occupancy(p) = \begin{cases} 1 & if\ p\ is\ accessible\ from\ r \\ 0 & otherwise \end{cases} \quad (3)$$

$$Obstacle(p,t) = \begin{cases} 1 & if\ obstacle\ between\ p\ and\ t \\ 0 & otherwise \end{cases} \quad (4)$$

그리고 *Orientation(t)*와 *Radius(p)*는 각각 표1, 표2와 같은 가중치를 사용하였다.

표 1. 사용자 방향에 따른 가중치

class	가중치
0	10.0
1,7	1.0
2,6	0.1
3,5	0.01
4	0.001

표 2. 후보 위치의 거리에 따른 가중치

거리	가중치
0.5m	0.5
1m	0.8
1.5m	0.8
2m	1.0

각 위치의 점수가 계산되면 점수가 가장 높은 위치를 선택하고 해당 위치로 로봇을 이동하도록 했다.

4. 실험 및 결과

이 장에서는 방향인식 모듈을 벤치마킹용 데이터셋에 대해 테스트 하고 기존 기계학습 기법과 성능을 비교한다. 또한 최적 위치 선정 모듈을 포함한 전체 시스템을 일반적인 가정 환경과 유사한 환경에서 실험한다.

4.1 방향인식 모듈의 성능

벤치마킹에는 Human 3.6M 데이터셋[7]을 사용했다. 데이터셋은 사람이 다양한 활동을 하는 상황을 촬영한 영상으로 이루어져 있다. 각 관절의 위치가 같이 제공되며, 이를 통해 사람의 방향을 계산할 수 있었다. 벤치마킹 데이터를 이용해 방향 인식 모듈을 학습한 경과는 그림 3과 같다.

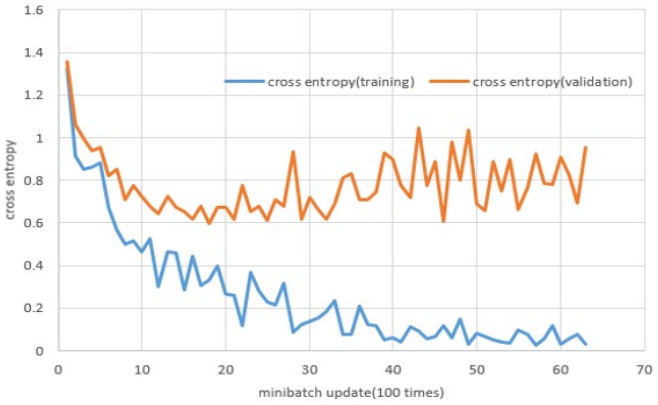

그림 3. CNN의 학습결과

학습된 모듈의 성능 평가를 위해 같은 데이터를 기존

방식인 HOG+SVM을 같은 데이터로 학습시켰다. 학습된 결과를 우리의 모델과 비교한 결과는 표3과 같다.

표 3. 벤치마킹 데이터에 대한 성능 비교

모델	데이터 셋	정확도
Convolutional neural network	Training	100.0%
Convolutional neural network	**Test**	**81.58%**
HOG+SVM (linear kernel)	Training	64.43%
HOG+SVM (linear kernel)	Test	57.26%
HOG+SVM (gaussian kernel)	Training	60.49%
HOG+SVM (gaussian kernel)	Test	77.12%

실제 환경에서의 수행능력 시험을 위해 실제 가정집과 유사한 실험실 환경에서 사람의 방향 데이터를 수집했다. 수집된 데이터에 대한 성능은 표 4와 같다.

표 4. 실제 환경에서 수집한 데이터에 대한 성능

데이터 셋	정확도
Training	97.0%
Test	**94.0%**

4.2 최적 로봇 위치선정 모듈의 성능

실제 환경 데이터를 이용해 학습한 방향인식 모듈과 최적 로봇 위치 선정 모듈을 결합하여 전체 시스템의 성능을 실험했다. 성능 평가를 위해 Microsoft Oxford API를 이용한 얼굴인식 모듈을 이용해 얼굴인식의 성공률이 얼마나 향상되는지 관찰했다. 모듈 적용 전후의 로봇 위치와 그에 따른 얼굴인식의 성공률은 각각 그림 4, 그림5과 같다.

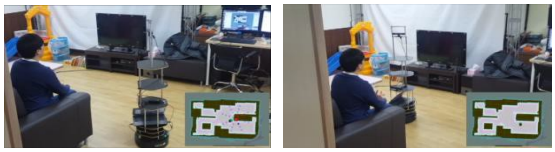

그림 4. 재배치 전후의 로봇 위치

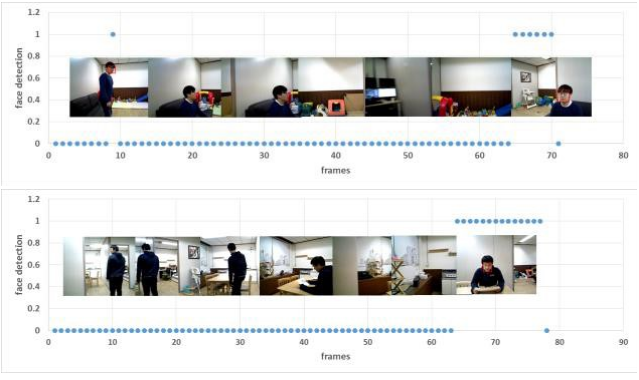

그림 5. 얼굴인식 성공률의 변화

5. 결 론

우리가 제안하는 시스템은 로봇에 장착된 카메라의 RGB 이미지에서 사용자를 인식하고 방향을 추론했으며 기존 방식에 비해 높은 정확도를 기록했다. 또한 인식한 방향 정보를 이용하여 얼굴인식 등의 서비스를 향상시키기 위해 로봇을 효율적으로 재배치하는데 성공했다. 앞으로의 연구는 강화학습을 사용해 로봇 위치선정 모듈을 더욱 향상시키는 것을 계획하고 있다.

Acknowledgement

이 논문은 2016년도 정부(미래창조과학부 국방부)의 재원으로 정보통신기술진흥센터 (R0126-16-1072-SW 스타랩), 한국산업기술평가관리원 (10044009-HRI.MFSSI 10060086-RISF), 국방과학연구소(UD130070ID-BMRR)의 지원을 받았음.

참고문헌

[1] B. Peng and G. Qian, "Binocular dance pose recognition and body orientation estimation via multilinear analysis," Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on. IEEE, 2008.

[2] C. Cheng, A. Heili, and J.-M. Odobez, "A joint estimation of head and body orientation cues in surveillance video," Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011.

[3] C. Schroeter, M. Hoechemer, et al. "Autonomous robot cameraman-observation pose optimization for a mobile service robot in indoor living space," Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009.

[4] Kessler, John, et al. "I'm still watching you: Update on observing a person in a home environment," Mobile Robots (ECMR), 2013 European Conference on. IEEE, 2013.

[5] I. Kalana, I. Lee, et al. "Mobile robotic active view planning for physiotherapy and physical exercise guidance," Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), 2015 IEEE 7th International Conference on. IEEE, 2015.

[6] R. Joseph, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," arXiv preprint arXiv:1506.02640, 2015.

[7] I. Clara, D. Papava et al. "Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," Pattern Analysis and Machine Intelligence, IEEE Transactions on 36.7 (2014): 1325-1339, 2014.