로봇 항법을 위한 영상 기반 사람 충돌 방지용 심층 학습 망 개발

강 희근[1], 정 봉혁[1], 김 태진[1], 김 재열[1], 이 기범[1], 김 행근[1,2,*], 손 영돈[1,2,*]
[1]가천대학교 가천융합의과학원 융합의과학과 의용생체공학전공
[2]가천대학교 의용생체공학과

Development of Image-Based Human Anti-Collision Network for Robot Navigation using Deep learning

H. G. Kang[1], B. H. Jeong[1], T. J. Kim[1], J. Y. Kim[1], K. B. Lee[1], H. K. Kim[1,2,*] and Y. D. Son[1,2,*]
[1]Department of Health Sciences and Technology, GAIHST, Gachon Univ., Incheon, Korea
[2]Department of Bio-medical Engineering, Gachon University, Incheon, Korea
* ydson@gachon.ac.kr

Abstract

This study proposes a method for predicting the robot actions, whether robot to stop or go, with human position using deep learning. We used VGG16 network which was pre-trained by Imagenet dataset. Labels which were pre-processed with human pose data extracted from 'OpenPose' are used for input data with raw images. For predicting robot's actions, labels were defined based on the viewing angle of robot. This trained network predicted the robot's actions with test accuracy up to 86.21%. In addition, this method can be applied to detect a human position to derive various robot actions.

1. 연구 배경

최근 병원에서는 의료 외적인 서비스를 수행하는데 있어 부족한 인력 문제를 해결하기 위해, 로봇 도입을 시도하고 있다. 특히 규모가 큰 병원은 내원 환자들의 많은 이동을 필요로 하기 때문에 길 안내 로봇에 대한 필요성을 제기하고 있다.

로봇 분야에서는 ROS (Robot Operating System)[1]를 이용한 길 안내 시스템이 꾸준히 발전되어 왔다. 길 안내 시스템은 기본적으로 목적지까지 경로를 설정하는 알고리즘(Global path planner)과 로봇의 거리 센서와 깊이 센서에서 출력되는 데이터를 통해 동적 장애물을 실시간으로 인식하고 대처하는 알고리즘(Local path planner)의 조합으로 이루어져 있다. 본 연구는 거리 센서와 깊이 센서를 사용하지 않고, 딥 러닝을 통해 단안 카메라 영상으로 경로상의 동적 장애물 특히 사람에 대해 회피할 수 있는 네트워크를 개발하는 것에 초점을 두었다.

2. 연구 방법

입력 데이터 전 처리

카메라의 시야 내에 존재하는 사람들의 위치 정보를 토대로 로봇의 움직임 예측하는 네트워크를 개발하기 위해, 1 인칭 Youtube 동영상[2-7]들을 사용하여 데이터셋을 구축하였다. 위 6 개의 영상은 초당 30 프레임을 가지는 영상이지만, 데이터 간의 유사도를 낮추기 위해, 초당 15 프레임으로 변경해 사용하였다. 또한 이 입력 영상들의 해상도를 VGG16[8] 네트워크의 입력 크기에 맞도록 224 x 224 로 축소하였다.

레이블을 생성시 OpenPose[9] 네트워크를 사용하였는데, 이 네트워크는 동영상 또는 영상에서 실시간으로 여러 사람들의 머리, 손, 발 등 사람들의 자세를 좌표 형태로 추출할 수 있는 Open source 패키지이다. 이 패키지를 통해 앞서 처리한 영상들을 입력으로 넣고, 출력으로 영상 속에 나타나는 모든 사람들의 신체 관절 좌표 데이터를 받는다. 해당 데이터를 빈 이미지에 뿌려 사람의 위치정보를 갖는 이진 영상을 만든다. 그리고 로봇 카메라의 시야각을 통해 실제 로봇이 특정 범위 내에 장애물이 감지되면 멈춰야 하는 Safety zone과 유사한 관심 영역을 도출하여, 해당 영역을 마스킹한 멈춤 영역 마스킹 영상을 만든다. 생성한 두 영상을 비교하여, 해당 프레임의 영상에서 사람의 위치가 멈춤 영역 안에 있을 경우 'STOP', 밖에 있을 경우, 'GO' 레이블을 생성하였다. (그림 1).

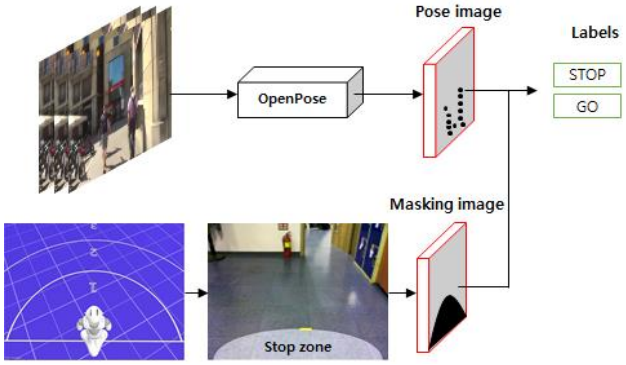

그림1 두 이진 영상들의 비교를 통한 레이블 생성 과정.

전 처리를 통해 생성한 영상과 레이블을 쌍으로 추출하여, 총 140,500개의 데이터를 생성하였다. 해당 입력 데이터 쌍을 임의로 선택해 가지고 오도록 random- shuffle generator를 사용하였다. 그리고 네트워크의 훈련 과정 중 과대적합을 방지하기 위해, 임의로 영상들의 대조도와 밝기를 조절하고, 좌우반전 영상을 만들어 내도록 하였다.

로봇 동작 예측

로봇의 동작을 예측하기 위한 네트워크로는 ImageNet dataset으로 사전 학습된 VGG16의 feature extractor로 사용하였고, Global Average Pooling을 사용해 1D Feature로 변환하여 Fully connected layer에 입력으로 사용하였다. (그림 2) 또한 과대적합을 방지하기 위해 Dropout을 추가하였다. 생성한 입력 데이터 총 140,500장 중 학습에 107,500장, validation으로 18,000장, test로 15,000장을 사용하였다.

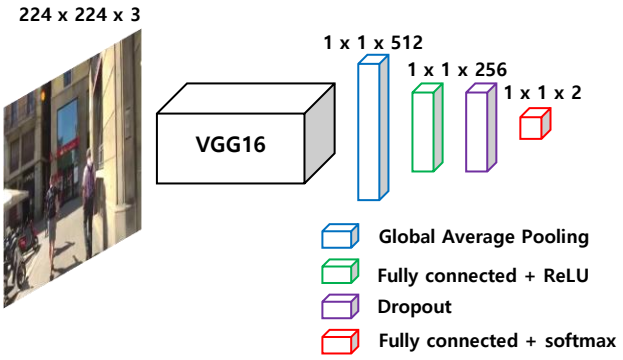

그림2 로봇 동작 예측을 위한 네트워크.

3. 연구 결과

네트워크의 훈련 결과 최대 train accuracy(top 1)는 92.41%, validation accuracy(top 1)는 87.75%를 가지고, 이를 통해 test 영상들을 평가한 결과 test accuracy(top 1)는 86.21%를 가진다. 이 훈련된 네트워크가 이미지의 어떤 부분을 보고 특정 label 로 판단을 내렸는지 Grad-CAM[10]을 통해 확인하였다.

그 결과 사람의 위치 정보가 로봇의 시야에 따라 정의된 멈춤 영역에 속하는 영상이 입력으로 들어온 경우 (그림 3-a, 3-b), 훈련된 네트워크는 멈춤 영역에 들어온 사람의 위치를 보고 'STOP' 동작을 해야 한다는 판단을 내린다. (그림 3-c) 이에 반해, 사람의 위치 정보가 멈춤 영역에 속하지 않는 영상이 입력으로 들어온 경우, 'GO' 동작을 해야 한다는 판단을 내린다.

이와 같은 방법을 로봇에 적용한다면, 실시간으로 영상 정보만을 가지고 동적 장애물인 사람에 대해 'STOP' 혹은 'GO' 등의 대처하는 판단을 내릴 수 있을 것으로 사료된다.

그림3 (a)는 네트워크 입력으로 들어가는 영상. (b)는 OpenPose 결과 영상과 멈춤 영역 영상의 합성 영상. (c) 훈련된 네트워크의 Grad-CAM test 영상.

향후에 영상 정보만을 가지고 멈추고 가는 두가지 동작 뿐만 아니라, 실시간으로 여러 동작들을 수행할 수 있는 네트워크를 개발을 할 예정이다. 또한 RNN(recurrent neural network)으로 확장해 동적 장애물들이 이동하는 경로를 함께 tracking 하여, 성능을 향상시키는 연구를 진행할 예정이다.

4. Acknowledgements

이 논문은 2019년도 정부의 재원으로 한국보건산업 진흥원의 지원을 받아 수행된 기초연구 사업임(캡슐내시경 위치 인식 및 실시간 병변 인식 기술 개발, No. HI19C0 656030019).

참고 문헌

[1] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler A., and Ng. "ROS: an open-source robot operating system." In ICRA Workshop on OSS, 2009.

[2] VisualJ Walks. China. (2018, November 18). A walk in Shichahai, Beijing, Hutong tour | What's it like in China? 《 4K 》 [Video file]. Retrieved from https://www.youtube.com/watch?v=qztRJrWt-Es&t=976s

[3] Wanna Walk. (2019, March 21). Buenos Aires, Argentina — City Walking Tour 【4K】ᴀʀ [Video file]. Retrieved from https://www.youtube.com/watch?v=SaDjNMF7hzo&t=505s

[4] Wanna Walk. (2018, October 24). MEXICO CITY — Zona Rosa, Video Walk 【4K】ᴍx [Video file]. Retrieved

from https://www.youtube.com/watch?v=2vMOyxJhpS8&t=480s

[5] Watched Walker. (2019, August 21). Regent Street Summer Streets 2019 (Narrated) London Walk Tour [Video file]. Retrieved from https://www.youtube.com/watch?v=8nfd-prM51g&t=2636s

[6] Wind Walk Travel Videos. (2017, May 19). Walking tour of Times Square in Midtown Manhattan, New York City Travel Guide 【4K】 [Video file]. Retrieved from https://www.youtube.com/watch?v=u68EWmtKZw0&t=2s

[7] Watched Walker. (2019, April 12). Gothic Quarter of Barcelona - Summer Walk Tour [Video file]. Retrieved from https://www.youtube.com/watch?v=RgJXH9KgL2E&t=1497s

[8] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[9] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 3D pose estimation using Part Affinity Fields. In arXiv preprint arXiv:1812.08008, 2018.

[10] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. CoRR arXiv:1610.02391 (2016)