KIPS Trans. Softw. and Data Eng.
Vol.10, No.7 pp.279~286
ISSN: 2287-5905 (Print), ISSN: 2734-0503 (Online)

https://doi.org/10.3745/KTSDE.2021.10.7.279

A Study on the Estimation of Multi-Object Social Distancing Using Stereo Vision and AlphaPose

Ju-Min Lee† · Hyeon-Jae Bae†† · Gyu-Jin Jang††† · Jin-Pyeong Kim††††

ABSTRACT

Recently, We are carrying out a policy of physical distancing of at least 1m from each other to prevent the spreading of COVID-19 disease in public places. In this paper, we propose a method for measuring distances between people in real time and an automation system that recognizes objects that are within 1 meter of each other from stereo images acquired by drones or CCTVs according to the estimated distance. A problem with existing methods used to estimate distances between multiple objects is that they do not obtain three-dimensional information of objects using only one CCTV. his is because three-dimensional information is necessary to measure distances between people when they are right next to each other or overlap in two dimensional image. Furthermore, they use only the Bounding Box information to obtain the exact coordinates of human existence. Therefore, in this paper, to obtain the exact two-dimensional coordinate value in which a person exists, we extract a person's key point to detect the location, convert it to a three-dimensional coordinate value using Stereo Vision and Camera Calibration, and estimate the Euclidean distance between people. As a result of performing an experiment for estimating the accuracy of 3D coordinates and the distance between objects (persons), the average error within 0.098m was shown in the estimation of the distance between multiple people within 1m.

Keywords : Stereo Vision, Distacne Estimation, Object Detection, Skeleton Extraction, Multi-Object

Stereo Vision과 AlphaPose를 이용한 다중 객체 거리 추정 방법에 관한 연구

이 주 민† · 배 현 재†† · 장 규 진††† · 김 진 평††††

요 약

최근 COVID-19 확산 방지를 위한 공공장소에서는 최소 1m 이상을 유지하는 물리적 거리두기 정책을 실행하고 있다. 본 논문에서는 드론과 CCTV가 취득한 스테레오 영상에서 실시간으로 사람들 간의 거리를 추정하는 방법과 추정된 거리에서 1m 이내의 객체를 인식하는 자동화 시스템을 제안한다. 기존의 CCTV를 이용하여 다중 객체 간의 거리 추정에 사용되었던 방법의 문제점으로는 한 대의 CCTV만을 이용하여 객체의 3차원 정보를 얻지 못한다는 것이다. 선, 후행하거나 겹쳐진 사람 간의 거리를 구하기 위해서는 3차원 정보가 필요하기 때문이다. 또한, 일반적인 Detected Bounding Box를 사용하여 영역 안에서 사람이 존재하는 정확한 좌표를 얻지 못한다. 따라서 사람이 존재하는 정확한 위치 정보를 얻기 위해 스켈레톤 추출하여 관절 키포인트의 2차원 좌표를 획득한 후, Stereo Vision을 이용한 카메라 캘리브레이션을 적용하여 3차원 좌표로 변환한다. 3차원으로 변환된 관절 키포인트의 중심좌표를 계산하고 객체 간 사이의 거리를 추정한다. 3차원 좌표의 정확성과 객체(사람) 간의 거리 추정 실험을 수행한 결과, 1m 이내에 존재하는 다수의 사람 간의 거리 추정에서 0.098m 이내 평균오차를 보였다.

키워드 : 스테레오 비전, 거리 추정, 객체 탐지, 스켈레톤 추출, 다중객체

1. 서 론

2019년 12월에 처음 확인된 COVID-19가 1년이 지난 지금까지 전 세계적으로 범유행을 하고 있다. 세계보건기구는 사람 사이의 접촉 가능성을 감소시켜 질병의 전파를 감소시키는 물리적 거리두기를 권고한다[1]. 특히 우리나라에서는 거리두기 1단계부터 실내/외에서 최소 1m 거리를 유지해야 한다. 공공장소와 같은 밀집 지역에서 물리적 거리두기를 감시하기 위한 인력을 파견하거나 CCTV, 드론 등을 활용하고 있으나 이를 파악하기 위한 인력이 부족한 상황이다. 따라서, CCTV, 드론 등을 통해 취득된 영상에서 Computer Vision을 이용하여 실시간으로 사람 간의 거리를 인식하여 사람들 간의 거리가 1m 이내인 경우를 인식하는 방법이 필요하다[2].

※ 이 논문은 행정안전부 극한재난대응기반기술개발사업의 지원을 받아 수행된 연구임(2020-MOIS31-014).
† 비 회 원 : 성균관대학교 인공지능융합학부 학사과정
†† 준 회 원 : 차세대융합기술연구원 컴퓨터비전 및 인공지능 연구실 연구원
††† 정 회 원 : 차세대융합기술연구원 컴퓨터비전 및 인공지능 연구실 연구원
†††† 정 회 원 : 차세대융합기술연구원 컴퓨터비전 및 인공지능 연구실 선임연구원
Manuscript Received : January 19, 2021
First Revision : April 26, 2021
Accepted : April 27, 2021
* Corresponding Author : Jin-Pyeong Kim(jpkim@snu.ac.kr)

우선 객체 탐지를 통해 사람이 존재하는 정확한 2차원 좌표를 찾아야 한다. 객체를 탐지하여 Bounding Box를 생성하는 YOLO[3]와 같은 방법으로는 사람이 존재하는 정확한 좌표를 알 수 없다. 이때, Bounding Box란 이미지 내의 객체를 탐지하여 객체 주위에 표시한 직사각형이다. Detected Bounding Box는 YOLO와 같은 모델의 결과값으로 객체 탐지 결과를 보여주며, Ground Truth는 탐지된 결과와 비교하기 위한 실제 위치 정보를 의미한다. Mask R-CNN[4]과 같이 Pixel 단위로 객체 탐지를 하는 Segmentation은 연산량이 많아 실시간성에 문제가 있다. 따라서 실시간성을 확보하면서도 사람의 관절을 키포인트로 추출하기 때문에 사람의 정확한 2차원 좌표를 얻을 수 있는 AlphaPose[5]를 이용했다.

AlphaPose는 YOLO v3[6]를 통하여 객체를 탐지하고, Detected Bounding Box 안에서 사람의 관절에 대해 키포인트를 추출한다. 또한, AlphaPose는 다중 객체의 키포인트 추출에서도 우수한 성능을 보이므로 사람들이 밀집한 환경에서 사람들 간의 거리를 측정하기에 적합하다. 실제 환경에서 2차원적인 정보만으로는 선, 후행하거나 겹쳐진 사람들 간의 거리를 추정하기 어려우므로 사람이 존재하는 위치의 3차원 정보가 요구된다.

3차원 물체가 2차원 이미지에 투사되면 깊이 정보가 손실되면서 2차원 정보가 된다. 깊이 정보 복원 방법으로는 Mono Vision[7], Stereo Vision[8]이 있다. 또한, 처음부터 3차원 정보를 유지한 Point Cloud를 생성하는 Lidar[9]가 있다. 정밀한 측정이 가능한 Lidar를 사용하면 정확도는 높아지지만, 고가이기 때문에 상용화하기 힘들다는 단점이 있다. Mono Vision의 경우 카메라 한 대밖에 필요하지 않지만, Stereo Vision에 비해 정보가 부족하므로 정확도가 낮다는 단점이 있다. 따라서 실생활에서 쉽게 구할 수 있는 두 대의 카메라를 이용하여 3차원 정보를 복원하는 Stereo Vision을 사용한다.

Imran A.(2020)에서는 Brid's eye view로 얻어진 이미지상에서 YOLO를 통해 구한 Detected Bounding Box의 중심점 간의 유클리디안 거리를 이용하여 다중 객체 거리 추정을 수행하였다[10]. Mahdi R.(2020)에서는 하나의 CCTV에서 얻은 이미지를 Brid's eye view 변환시킨 후, Imran A.(2020)와 같이 다중 객체 거리 추정을 수행하였다[11]. 하지만, 위와 같은 방법에서는 한 대의 카메라를 사용하므로 객체의 3차원 좌표를 알 수 없어, 정확한 거리 추정이 불가능하다. Nikolay N.(2020)에서는 이를 추정거리 오차감소를 위해서 2대의 CCTV를 사용하여 Detected Bounding Box의 3차원 중심점 좌표를 얻었다[12]. 하지만 객체의 정확한 위치 정보를 알기 위해선 Bounding Box만을 이용하는 것보다 Skeleton의 정보를 활용하는 위치검출이 정확하다[13]. 따라서 본 논문에서는 다중 객체 거리 추정을 위해서 Skeleton을 추출하고 이를 활용한 객체별 중심좌표 계산 후, 거리 추정하는 방식을 제안한다. 수행하는 방법으로는 AlphaPose를 활용하여 좌측 프레임에서 관절 키포인트를 추출하고, 이미지 내에서 키포인트를 활용하여, 사람이 존재하는 2차원 좌표를 파악한다. 2차원 좌표의 손실된 깊이 정보를 복원하기 위해서 Stereo Vision을 이용한다. 동일한 한 점에 대해 좌/우 프레임에서 생기는 시차를 이용하여 카메라로부터 사람까지의 깊이인 z값을 얻는다. 카메라 캘리브레이션을 통하여 앞서 구한 2차원 x, y 좌표를 3차원 x, y 좌표로 변환하여 3차원 좌표를 결정한다. 이를 이용하여 사람 간의 L2 distance를 구하고, 1m 이내의 객체를 인식해서 알려준다.

2장에서는 키포인트를 추출하는 방법과 관련된 연구를 살펴본다. 키포인트는 사람의 위치를 파악할 때 활용된다. 3장에서는 추출한 2차원 키포인트를 Stereo Vision과 카메라 캘리브레이션을 통해 3차원 키포인트로 변환하는 알고리즘에 대해 살펴본다. 4장에서는 3차원 좌표의 정확성과 사람 간의 거리 측정 실험을 분석/평가한다. 5장에서는 결론 및 향후 방향을 제시한다.

2. 관련 연구

2.1 관절 키포인트 추출 모델 비교

이미지 내에서 객체의 자세 추정이 가능한 Mask R-CNN, OpenPose와 AlphaPose의 이미지의 내의 사람 수에 따른 수행시간인 Inference Time을 비교한다. Fig. 1을 보면, Inference time이 상향식의 OpenPose는 이미지 내의 사람 수와 무관하지만, 하향식인 Mask R-CNN과 AlphaPose는 이미지 내의 사람 수에 따라 선형적으로 늘어나는 것을 알 수 있다. 이때, 상향식이란 이미지에 포함된 사람의 관절 키포인트를 모두 추출하고, 관절의 상관관계를 분석하여 각 사람들의 관절 키포인트를 알아내는 방법이다. 반면에 하향식이란 이미지에서 사람을 먼저 찾고, Bounding Box 내부에서 사람의 관절 키포인트를 찾는 방법이다.

Image내에서 객체의 수가 20명 이하일 때, max accuracy를 측정하는 OpenPose보다 AlphaPose의 inference time이 짧음을 확인할 수 있다. 또한, 그 이상의 객체가 image 내에 있어도 AlphaPose가 OpenPose보다 inference time이 우수한 것을 보인다.

다음으로 Mask R-CNN, OpenPose와 AlphaPose의 정확도 비교를 위해 COCO 데이터 세트[15]에서의 정밀도(AP)를 비교한다. Table 1의 AP를 보면, 관절 키포인트 추출 성능이 AlphaPose가 71%로 다른 방법들에 비해 정확하며, Mask R-CNN이 69.2%, OpenPose가 64.2%의 정확도를 보여준다.

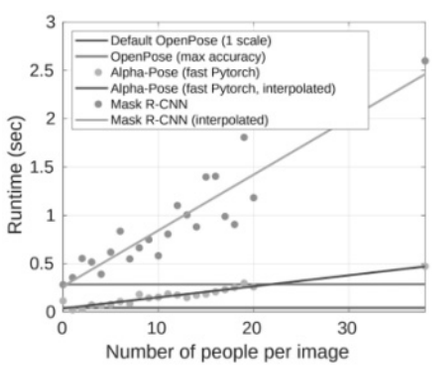

Fig. 1. Inference Time Comparison between OpenPose, Mask R-CNN, and AlphaPose[14]

Table 1. Results on COCO Test-dev Leaderboard

Method	AP	AP@ 0.5	AP@ 0.75	AP medium	AP large
Mask R-CNN	69.2	90.4	76.4	64.9	76.3
OpenPose	64.2	86.2	70.1	61.0	68.8
AlphaPose	71.0	87.9	77.7	69.0	75.2

Table 2. Results on the whole testing set of MPII Dataset

Method	Hea	Sho	Elb	Wri	Hip	Kne	Ank	mAP
OpenPose	91.2	87.6	77.7	66.8	75.4	68.9	61.7	75.6
AlphaPose	88.4	86.5	78.6	70.4	74.4	73.0	65.8	76.7

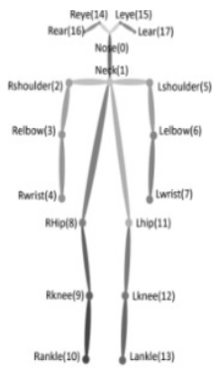

Fig. 2. The Human Skeleton Model with 18 Joints[17]

AlphaPose와 Mask R-CNN 모두 성능이 우수하지만 Mask R-CNN은 하향식 방법으로 관절 키포인트를 추출하므로 실시간으로 사용하기에는 속도가 느리다. 따라서, 추가적으로 실시간으로 사용할 수 있는 OpenPose와 AlphaPose 정확도 성능을 MPII 데이터셋[16]을 이용하여 비교한다.

Table 2의 머리, 어깨, 팔꿈치, 손목, 엉덩이, 무릎, 발목의 AP 평균인 mAP를 보면, AlphaPose의 성능이 76.7%로 OpenPose의 성능보다 우수하다.

따라서, 실내/외에서 실시간으로 사람들 간의 거리를 추정하는 방법 정확도가 가장 높으며 사람 수에 따라 선형적으로 시간이 변하는 AlphaPose를 관절 키포인트 추출 모델로 선정하였다.

2.2 AlphaPose를 활용한 관절 키포인트 추출

COCO Dataset으로 학습한 AlphaPose는 다수의 관절 키포인트 추출이 가능하며 Fig. 2와 같이 각 객체에 대해 최대 18개의 관절 키포인트를 추출한다. 하향식을 사용하므로 사람 탐지에서 발생하는 오차 즉, 부정확한 Bounding Box에 의해 발생하는 오차가 존재한다. 따라서 부정확한 Bounding Box에서도 자세를 정확하게 추정할 수 있도록 Ground Truth Bounding Box와 Detected Bounding Box 사이의 분포를 모델링하여 성능을 높인다.

3. Stereo Vision System

3.1 카메라 캘리브레이션

카메라 이미지는 3차원 공간상의 점들을 2차원 이미지 평면에 투사함으로써 얻어진다. 이때 카메라 캘리브레이션은 3차원 공간 좌표와 2차원 이미지 좌표 사이의 변환 관계를 설명하는 파라미터를 찾는 과정이다. 내부 파라미터는 우리가 동일한 장면을 같은 위치/각도에서 찍더라도 사용한 카메라에 따라서 서로 다른 영상을 얻게 되는 기구적인 요인을 뜻한다. 파라미터엔 초점거리(f_x, f_y)와 주점(c_x, c_y)이 있다. 초점거리란 렌즈 중심과 이미지 센서 사이의 거리이다. 주점은 카메라 렌즈의 중심 즉, 핀홀(Pin-Hole)에서 이미지 센서에 내린 수선의 발의 영상 좌표이다. 따라서 이 내부 파라미터에 의한 영향을 없애주는 것이 카메라 캘리브레이션의 목적이다. Equation (1)으로 이를 수행한다.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} f_x X + c_x Z \\ f_y Y + c_y Z \\ Z \end{bmatrix} = \begin{bmatrix} \frac{f_x X + c_x Z}{Z} \\ \frac{f_y Y + c_y Z}{Z} \\ 1 \end{bmatrix} \quad (1)$$

3.2 Stereo Vision System

Stereo vision은 사람의 눈과 비슷하게 좌/우 이미지 차이를 이용하여 손실된 깊이 정보인 카메라와 객체 간의 거리 정보를 복원한다. 이는 Fig. 3과 같은 방식으로 계산된다. 카메라 좌표계에 대한 물체의 좌표 ${}^{c}p = ({}^{c}x, {}^{c}y, {}^{c}z)$이고, 좌/우 이미지 평면에 투사된 좌표가 ${}^{l}p = ({}^{l}x, {}^{l}y)$, ${}^{r}p = ({}^{r}x, {}^{r}y)$이다. 시차(Disparity)란 왼쪽 이미지의 한 점의 위치와 오른쪽 이미지의 대응점 위치의 차이이다. 이때 대응점은 x축과 평행한 곳에만 위치하게 되는데 이는 좌/우 이미지의 에피폴라 라인을 평행하도록 이미지 평면을 회전/이동/스케일링하는 Rectification을 수행했기 때문이다. 시차와 깊이와의 관계는 Equation (2)과 같다. 이는 삼각측량법(Triangulation)에 의해 유도된다.

Equation (2)에서 볼 수 있듯이 시차와 깊이는 반비례 관계이다. 따라서 깊이가 깊을수록 시차가 작아지므로 좌/우 카메라의 거리인 Baseline이 멀수록 시차 편차가 크게 나타나 결론적으로 깊이 정보를 더욱 정밀하게 파악할 수 있다. 또, 시차는 깊이가 깊을수록 양자화 오차에 의해 깊이 오차가 커진다. 본 논문에서는 Baseline이 고정된 ZED2 Camera를

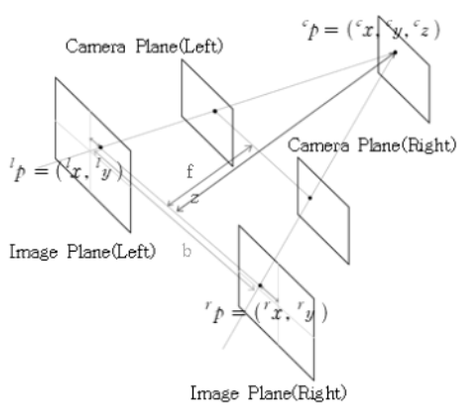

Fig. 3. Stereo Vision System

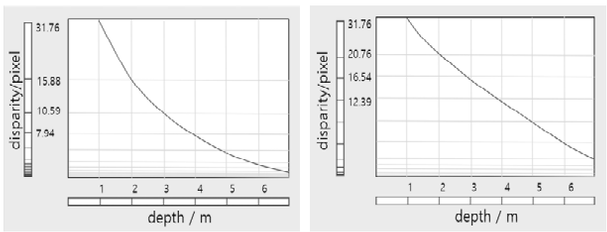

Fig. 4. Nonuniform Quantization

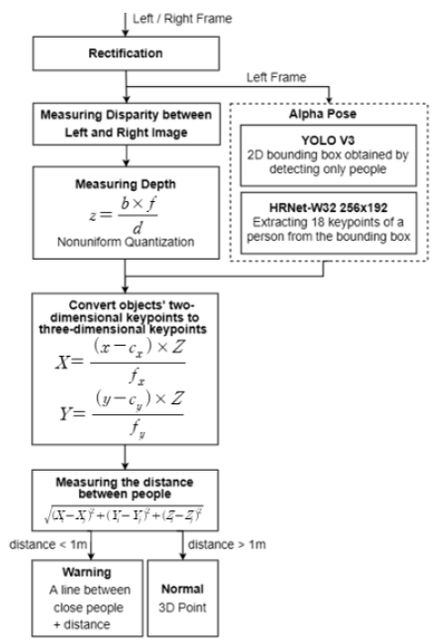

Fig. 5. Flow Chart of Algorithm

사용하므로 시차 편차를 보정하기 위해서 Fig. 4와 같이 불균일 양자화를 수행하여 깊이 오차를 보정한다.

$$depth\,(z) = \frac{focal \ \leq ngth(f) \times \square \mid (b)}{disparity} \tag{2}$$

3.3 거리 측정

본 논문의 흐름은 Fig. 5와 같이 구성된다. 좌/우 프레임에서 받은 이미지를 이용하여 시차를 계산하고 좌측 프레임에 AlphaPose[4]를 적용하여 각 객체에 대해 최대 18개의 2차원 관절 키포인트를 추출했다. 다음으로 Equation (2)에 추출된 키포인트에 해당하는 좌표의 시차를 입력값으로 대입하여 깊이를 구했다. 이후, 카메라 캘리브레이션을 수행하는 Equation (1)을 통하여 추출된 키포인트를 3차원 좌표로 변환했다. 객체를 대표하는 좌표를 구하기 위해 Equation (4)을 이용하여 중심점을 구한 후, 중심점들 간의 유클리디안 거리를 Equation (5)로 구한다.

$$N = Nmber\ of\ keypoints\ extracted \leq 18 \tag{3}$$

$$X = \frac{1}{N}\sum_{n=1}^{N} X_n \ , \ Y = \frac{1}{N}\sum_{n=1}^{N} Y_n \ , \ Z = \frac{1}{N}\sum_{n=1}^{N} z_n \tag{4}$$

$$distance_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2} \tag{5}$$

4. 실험 결과 및 평가

4.1 실험 환경

스테레오 영상 취득하기 위한 카메라는 ZED 2를 사용하였다. 카메라의 특성은 Table 3과 같다. 하드웨어 실험 환경은 3.6GHz의 Intel(R) core I7-9700K CPU와 NVIDIA Geforce GTX 1660 Ti Graphic Card, 32GB RAM이 설치된 데스크톱 PC에서 실험했다. 딥러닝 프레임워크는 PyTorch 1.1.0 버전을 사용하여 실험 시뮬레이션을 수행하였으며 Python을 프로그래밍 언어로 사용하였다.

4.2 모델의 깊이 측정 실험 결과 및 평가

시차를 측정하기에 앞서 좌/우 이미지에 대해 Rectification을 수행한다. 그 결과, 두 이미지의 대응점이 Fig. 6과 같이 x축과 평행하게 위치한다.

거리 측정의 성능을 높이기 위해서는 깊이 값 측정 성능을 먼저 검증했다. 1m 간격으로 4m까지 사람이 서 있는 상황을 10번 반복하여 실험을 진행했다. 실험 결과는 Fig. 9와 같다. 실험 환경의 예시 화면인 Fig. 8의 시각화된 좌표 (x, y, z)는 Equation (4)에 의해 구하며 z값이 깊이이다. Fig. 7은 깊이

Table 3. Feature of ZED 2

Resolution	focal length	baseline
HD2K	1000	12cm
HD1080	1000	
HD720	500	
WVGA	250	

Fig. 6. Rectified Left/right Image

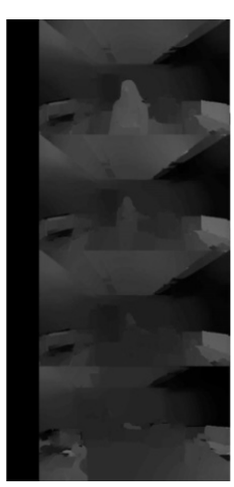

Fig. 7. Disparity Map

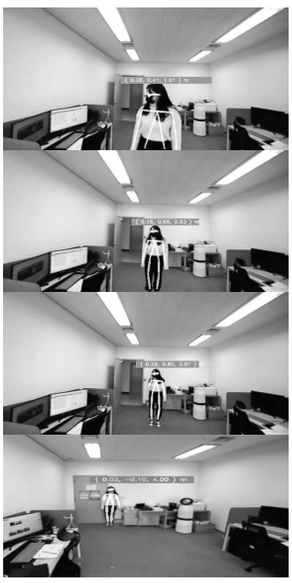

Fig. 8. Example Screens

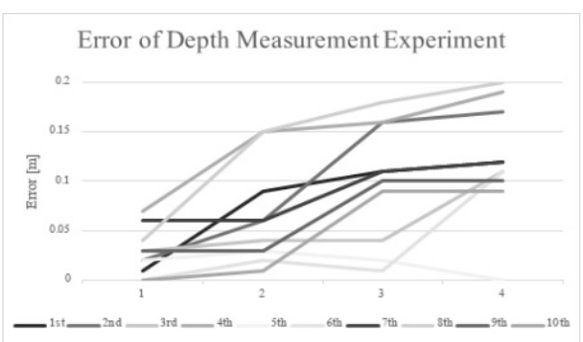

Fig. 9. Error of Depth Measurement Experiment

를 구하기 위해 먼저 구했던 시차를 0-255 사이로 정규화하여 시각화한 결과이다.

1m 이내에서는 거리추정 오차가 3cm 이하로, 높은 신뢰도를 보였다. 그러나 거리가 멀어질수록 오차가 증가하는 것을 볼 수 있다. 이는 거리가 증가함에 따라 깊이 오차가 함께 커지기 때문이다.

4.3 모델의 두 사람 간 거리 추정 실험 결과 및 평가

두 사람 간의 거리 추정 성능을 확인하기 위해서 두 가지 실험을 진행했다. 실험 환경은 Fig. 10과 같이, 사람이 45도씩 움직인 곳에 1번부터 8번까지 순번을 매겼다.

첫 번째로, 원의 중심에 있는 사람을 기준으로 여덟 방향에 다른 사람이 서 있는 상황을 10번 반복하여 1m를 측정하는 실험을 진행했다. 실험 결과는 Fig. 11과 같다. 8지점의 평균 오차는 0.098m로 우수한 성능을 보였다. ③과 ⑦ 지점에서는 두 사람이 겹쳐져서 생기는 오차가 발생한다. 두 사람이 완전히 겹쳐서 거리를 측정할 수 없는 경우는 Nan으로 표시했다. 또, 객체가 겹쳐있는 상황에서는 중심점에 있는 객체의 키포인트를 다른 객체의 키포인트로 인식하여 평균 20cm 정도의 오차가 발생한다. ⑥, ⑧ 지점에서는 3m 이상의 깊이

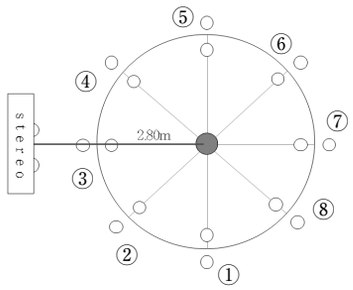

Fig. 10. Test Environment

Fig. 11. Error of 1m Measurement Experiment

측정 오차가 포함되어 있으므로 10cm 정도의 오차가 발생했다.

두 번째로 1분 동안 8지점에서 원 안과 밖을 이동하며 1m 이하/1m 초과의 두 가지 경우에 대해 분류하는 인식률 측정 실험을 진행하였다. 총 1449개의 프레임(Frame) 중 5의 배수 프레임에 대해서만 인식률을 계산했다. 8개의 지점에서 원 안/밖으로 이동하며 16개의 지점에 대해서 18프레임에 대한 측정된 결과는 Table 4와 같다. Fig. 12의 좌측 이미지는 1번에서 8번 지점에서 객체 간 1m 이상 거리가 유지된 경우이고, 우측 이미지는 1m 이하의 간격에 있는 경우로 빨간 선분과 추정된 거리로 경고 표시를 한다. 이때 선분 시작점과 끝점은 Equation (4)를 통해 구한 객체별 관절 키포인트의 중심점이다.

Table 4를 보면, 각 지점에서 18개의 프레임 중 거리가 잘못 측정된 프레임 수를 나타내었다. 2종 오류의 발생률은 0.1125, 1종 오류의 발생률은 0.06375로, 전반적으로 낮은 오차율을 보였다. 이때 1종 오류란, 실제 객체 간 거리가 1m 이하인데 추정 거리 오차로 경고가 발생하지 않은 경우이며,

Table 4. Experimental Results of 1m Recognition Rate

	①	②	③	④	⑤	⑥	⑦	⑧
Below 1m								
Misclassified Frame	0	0	5	1	1	3	2	4
Error Rate	0	0	0.28	0.06	0.06	0.17	0.11	0.22
Over 1m								
Misclassified Frame	0	0	0	0	0	3	5	1
Error Rate	0	0	0	0	0	0.17	0.28	0.06

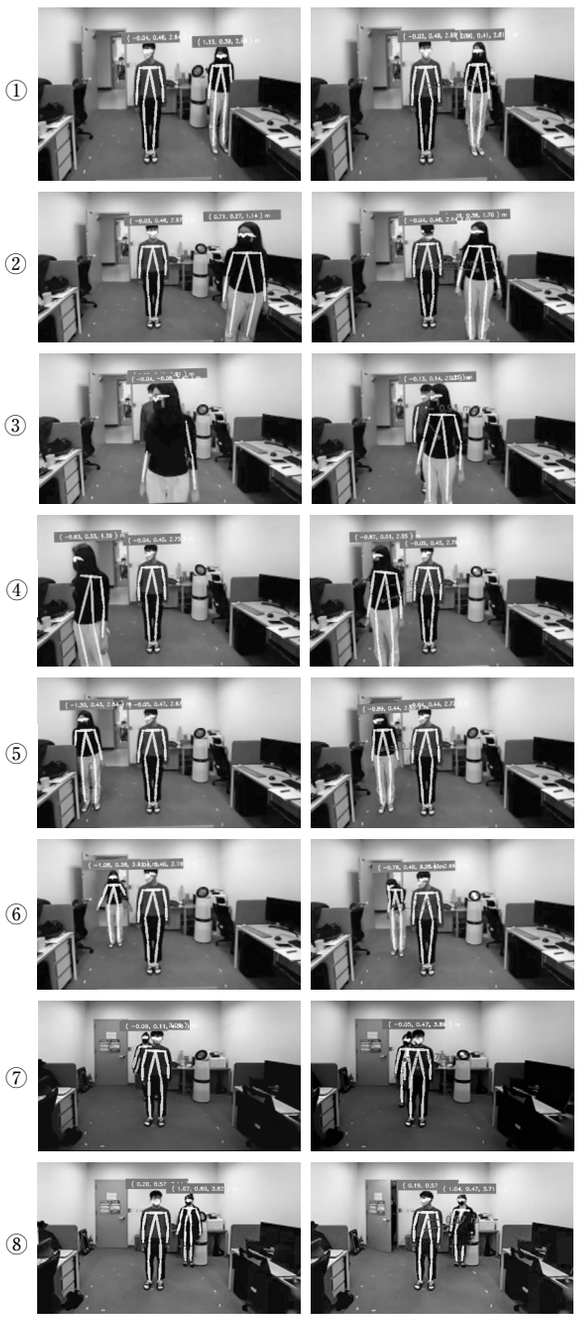

Fig. 12. Example Screens that Categorize Two Cases

2종 오류란, 실제 객체 간 거리가 1m 이상인데 추정 거리 오차로 경고가 발생한 경우이다. 첫 번째 실험과 마찬가지로 두 사람이 겹쳐지는 ③, ⑦ 지점에서의 오차율이 높았으며, 거리가 먼 ⑥, ⑦, ⑧ 지점에서 다른 지점에서보다 오차율이 높은 것을 볼 수 있다.

4.4 모델의 다중 객체 간 거리 추정 실험 결과 및 평가

마지막으로 다중 객체 간의 1m 이하/1m 초과의 두 가지 경우를 분류하는 인식률 측정 실험을 1분간 진행했다. 중심에 서 있는 사람과 60도를 이루는 곳을 찾고, Fig. 13과 같이

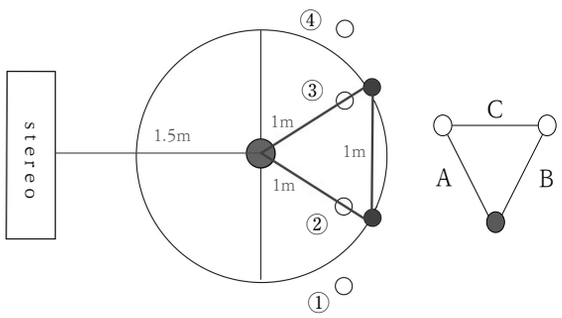

Fig. 13. Test Environment & Distance Naming

Table 5. Experimental Results of 1m Recognition Rate

		①, ④	①, ③	②, ④	②, ③	Error Rate
Mis-classified Frame	A	0	5	0	4	0.090
	B	0	0	2	1	0.021
	C	0	0	0	1	0.007
Error Rate		0	0.139	0.055	0.167	

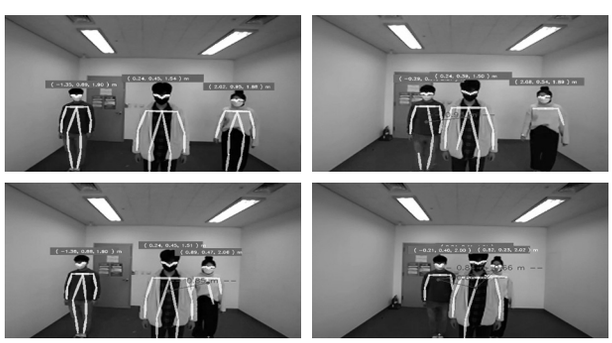

Fig. 14. Example Screens where Two People Located at (①, ④), (①, ③), (②, ④), (②, ③)

두 사람이 서 있을 곳을 1번부터 4번까지 순번을 매겼다. 원의 중점에 위치한 사람은 고정되어 있고, 두 사람의 위치를 (①, ④), (①, ③), (②, ④), (②, ③)으로 옮겨가며 세 사람 간의 거리를 측정했다. (①, ④)의 경우 세 사람 간의 거리가 모두 1m 이상이므로 경고 표시가 나타나지 않고, (①, ③)의 경우 A에서만, (②, ④)의 경우 B에서만, (②, ③)의 경우 세 사람 간의 거리가 모두 1m 이하이므로 A, B, C 모두 경고 표시가 나타나야 한다. 1449프레임 중 5의 배수 프레임으로만 인식률을 계산했다. Table 5는 각 지점의 36프레임 동안의 1m 이내 탐지 결과이다. 진행된 실험 예시 화면은 Fig. 14와 같다.

Table 5을 보면 각 지점에서 A, B, C 거리에 대한 1종 오류와 2종 오류를 나타냈다. 1m 이상인 거리를 회색 칸으로 표시했고, 그 중 2개의 프레임에서만 2종 오류가 발생했다. 이는 1m 이상의 거리 추정에서 우수한 성능을 나타낸다고 할 수 있다. (①, ④), (①, ③), (②, ④), (②, ③) 중에서 세 사람의 거리가 모두 1m보다 작은 (②, ③) 지점에서 오차가 0.167로 가장 크다. 이는 세 사람의 3차원 좌표에서 약간이

Fig. 15. Example Screen in Real Environment

라도 오차가 발생하면 거리가 1m보다 크다고 추정되기 때문이다. 또한, A, B, C 거리 중에서 A 거리의 오차가 0.090로 가장 큰 것을 볼 수 있는데 이는 어두운 옷을 입은 사람의 관절 키포인트가 제대로 추출되지 않았기 때문이다.

Fig. 15는 실제 환경에서 다중 객체의 거리 탐지 알고리즘을 적용한 결과 거리 추정이 가능함을 보여준다.

5. 결론 및 향후 방향

본 논문에서는 실시간으로 사람들 간의 거리를 추정하는 방법과 추정된 거리를 이용하여 1m 이내에 있는 객체를 인식하는 자동화 시스템을 설계하였으며 그 성능을 실험으로 입증하였다. 다중 객체의 키포인트 추출 성능이 우수한 AlphaPose를 활용하여 사람이 존재하는 정확한 2차원 좌표를 얻고, 손실된 깊이 정보를 복원하기 위해서 좌/우 프레임의 시차를 이용했다. 카메라 캘리브레이션을 통해 3차원 좌표로 변환하여 사람 간의 거리를 유클리디안 거리를 통하여 얻었다.

성능을 평가하기 위해서 깊이 추정 실험, 두 사람 간 거리 추정 실험, 다중 객체 간 거리 추정 실험을 진행했다. 첫 번째 실험에서, 1m 간격으로 4m까지 깊이를 추정한 결과, 1m까지의 오차가 3cm 이내로 가장 작았으며, 깊이가 1m씩 깊어질수록 오차가 7cm 이내, 10cm 이내, 12cm 이내로 측정되었다. 두 번째 실험에서, 두 사람 간의 1m 거리 추정이 올바른지 확인하기 위해 정확성 실험과 인식률 실험을 진행했다. 정확성 실험 결과, 8방향에 대해서 물리적 거리두기의 기준인 1m 측정 평균 오차가 0.09375m로 측정되었다. 인식률 실험 결과 2종 오류의 발생률은 0.1125이고 1종 오류의 발생률은 0.06375이다. 마지막으로 실시간으로 움직이는 다중 객체 간의 거리 인식률 실험에서 1m 이내에 있는 사람의 수가 세 명일 경우 오차율이 0.167로 실용가능할 정도의 높은 인식률을 보였다. 따라서, 세 가지 실험을 통하여 본 연구의 목적인 사람들 간의 거리를 실시간으로 추정하는 방법과 추정된 거리를 이용하여 1m 이내의 객체를 인식하는 자동화 시스템의 성능을 입증하였다.

차후 연구 계획은 베이스 라인을 넓히기 위해 두 대의 CCTV를 사용하여 4m 이상의 거리에 위치하는 객체에 대해서도 낮은 거리 추정 오차를 내는 것이다. 또한, 사람이 존재하는 정확한 좌표를 얻기 위해 키포인트를 추출하는 AlphaPose 대신, 실시간으로 Instance Segmentation을 수행하는 YOLACT[18]을 활용하여 사람 간 겹치는 영역이 커졌을 때 생기는 오차를 줄여 성능을 높일 것이다.

References

[1] T. P. B. Thu, T. P. N. H. Ngoca, N. M. Hai, and L. A.Tuanc, "Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries," *Proceeding of the Science of The Total Environment*, 2020.

[2] L. S. Liebst, W. Bernasco, M. R. Lindegaard, and E. Hoeben, "Social distancing compliance: A video observational analysis," *Proceeding of the PLoS ONE*. 2021.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE International Conference on Computer Vision*, 2016.

[4] H. Kaiming, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[5] H. S. Fang, S. Xie, Y. W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[6] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767. 2018.

[7] H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, "Deep ordinal regression network for monocular depth estimation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[8] J. R. Chang and Y. S. Chen, "Pyramid stereo matching network," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[9] Y. Wang, W. L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-LiDAR from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[10] I. Ahmed, M. Ahmad, J. P. C. Rodrigues, G. Jeon, and S. Din, "A deep learning-based social distance monitoring framwork for COVID-19," *Sustainable Cities and Society*, Vol.65, Article 10257r, 2020.

[11] M. Rezaei and M. Azarmi, "DeepSOCIAL: Social distancing monitoring and infection risk assessment in COVID-19 pandemic," *Proceedings for the Applied Sciences*, Vol.10, Article 7514, 2020.

[12] N. Neshov, A. Manolova, K. Tonchev, and V. Poulkov, "Real-time estimation of distance between people and/or objects in video surveillance," *Proceedings of the International Symposium on Wireless Personal Multimedia Communications*, 2020.
[13] L. Narupiyakul and N. Srisrisawang, "A comparison between skeleton and bounding box models for falling direction recognition," *Proceedings of the International Conference on Robotics and Machine Vision*, 2017.
[14] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *Proceedings of the IEEE International Conference on Computer Vision*, 2019.
[15] T. Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, et al. "Microsoft COCO: Common objects in context," *Proceedings of the European Conference on Computer Vision*, 2014.
[16] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: New benchmark and state of the art analysis," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2014.
[17] Human Pose Estimation Image AI Data [Internet], https://aihub.or.kr/aidata/138
[18] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++: Better real-time instance segmentation," *Proceedings of the ICCV Conference on Computer Vision,* 2019.

이 주 민
https://orcid.org/0000-0002-0565-2679
e-mail : jumin47@skku.edu
2017년 ~ 현 재 성균관대학교 인공지능융합학부 학사과정
관심분야 : 머신러닝, 딥러닝, 컴퓨터비전

배 현 재
https://orcid.org/0000-0002-2164-0125
e-mail : jason0425@snu.ac.kr
2021년 ~ 현 재 성균관대학교 소프트웨어학과(석사)
2019년 ~ 2020년 차세대융합기술연구원 인턴연구원
2020년 ~ 현 재 차세대융합기술연구원 컴퓨터비전 및 인공지능연구실 연구원
관심분야 : Pose Estimation, Object Detection, Action Recognition

장 규 진
https://orcid.org/0000-0002-1575-2796
e-mail : gjjang@snu.ac.kr
2011년 성균관대학교 전자전기컴퓨터공학과(석사)
2013년 성균관대학교 전자전기컴퓨터공학과(박사수료)
2018년 한국철도기술연구원 스마트모빌리티연구팀 주임연구원
2020년 ~ 현 재 차세대융합기술연구원 컴퓨터비전 및 인공지능연구실 연구원
관심분야 : Computer Vision & Artificial Intelligence

김 진 평
https://orcid.org/0000-0003-4840-7216
e-mail : jpkim@snu.ac.kr
2006년 성균관대학교 전자전기컴퓨터공학과(석사)
2014년 성균관대학교 전자전기컴퓨터공학과(박사)
2016년 ~ 2018년 한국철도기술연구원 선임연구원
2018년 ~ 2019년 한국도로공사 도로교통연구원 책임연구원
2019년 ~ 현 재 차세대융합기술연구원 컴퓨터비전 및 인공지능연구실 선임연구원
관심분야 : Artificial Intelligence & Computer Vision