Journal of the Korea Academia-Industrial cooperation Society
Vol. 21, No. 7 pp. 285-291, 2020

https://doi.org/10.5762/KAIS.2020.21.7.285
ISSN 1975-4701 / eISSN 2288-4688

모션 인식을 위한 2D 자세 추정 알고리듬의 이미지 전처리 및 얼굴 가림에 대한 영향도 분석

노은솔[1], 이사랑[2], 홍석무[3*]
[1]공주대학교 융합기계공학과, [2]공주대학교 기계공학과, [3]공주대학교 기계자동차공학부

Investigation of image preprocessing and face covering influences on motion recognition by a 2D human pose estimation algorithm

Eunsol Noh[1], Sarang Yi[2], Seokmoo Hong[3*]
[1]Department of Mechanical Convergence Engineering, Kongju National University
[2]Department of Mechanical Engineering, Kongju National University
[3]Department of Mechanical & Automotive Engineering, Kongju National University

요 약 제조 산업에서 인력은 로봇으로 대체되지만 전문 기술은 데이터 변환이 어려워 산업용 로봇에 적용이 불가능하다. 이는 비전 기반의 모션 인식 방법으로 데이터 확보가 가능하나 이미지 데이터에 따라 판단 값이 달라질 수 있다. 따라서 본 연구는 비전 방법을 사용해 사람의 자세를 추정 시 영향을 미치는 인자를 고려해 정확성 향상 방법을 찾고자 한다. 비전 방법 중 OpenPose의 3가지 모델 MPII, COCO 및 COCO + foot을 사용했으며, CNN(Convolutional Neural Networks)을 사용한 OpenPose 구조에서 얼굴 가림 및 이미지 전처리에 미치는 영향을 확인하고자 액세서리의 유무, 이미지 크기 및 필터링을 매개 변수로 설정했다. 각 매개 변수 별 이미지 데이터를 3 가지 모델에 적용해 실제 값과 예측 값 사이 거리 오차와 PCK (Percentage of correct Keypoint)로 영향도를 판단했다. 그 결과 COCO + foot 모델은 3 가지 매개 변수에 대한 민감도가 가장 낮았다. 또한 이미지 크기는 50% (원본 3024 × 4032에서 1512 × 2016로 축소) 이상 비율이 가장 적절하며, MPII 모델만 emboss 필터링을 적용할 때 거리 오차 평균이 최대 60pixel 감소되어 향상된 결과를 얻었다.

Abstract In manufacturing, humans are being replaced with robots, but expert skills remain difficult to convert to data, making them difficult to apply to industrial robots. One method is by visual motion recognition, but physical features may be judged differently depending on the image data. This study aimed to improve the accuracy of vision methods for estimating the posture of humans. Three OpenPose vision models were applied: MPII, COCO, and COCO+foot. To identify the effects of face-covering accessories and image preprocessing on the Convolutional Neural Network (CNN) structure, the presence/non-presence of accessories, image size, and filtering were set as the parameters affecting the identification of a human's posture. For each parameter, image data were applied to the three models, and the errors between the actual and predicted values, as well as the percentage correct keypoints (PCK), were calculated. The COCO+foot model showed the lowest sensitivity to all three parameters. A 〈50% (from 3024×4032 to 1512×2016 pixels) reduction in image size was considered acceptable. Emboss filtering, in combination with MPII, provided the best results (reduced error of 〈60 pixels).

Keywords : Pose Estimation, OpenPose, Image Preprocessing, Image Filtering, Face Covering

*Corresponding Author : Seokmoo Hong(Kongju National Univ.)
email: smhong@kongju.ac.kr
Received April 8, 2020 Revised May 8, 2020
Accepted July 3, 2020 Published July 31, 2020

1. 서론

현재 제조 산업은 주 52시간 상한제 도입, 최저 임금 상승 및 산업 현장 위험성 등의 이유로 인력이 점차 로봇으로 대체되고 있다. 산업통상자원부에 따르면 국내에서 사용되는 로봇 중 약 80%가 자동차, 전기 전자 분야와 같은 제조 산업에서 사용된다. 제조 산업에서 제품 생산은 소비자 니즈 변화로 형상이 다양해져 소품종 대량 생산에서 다품종 소량 생산으로 변화되고 있다. 이에 로봇을 사용하여 제품을 생산할 때 형상마다 적절한 교시(teaching) 과정이 필요하다. 그러나 전문가 지식 기반 산업은 노하우를 수치적인 데이터로 변환 하지 못해 로봇 교시가 어렵다. 이에 Kim[1]등은 직접 교시 방법으로 작업자가 직접 로봇을 움직여 직관적으로 동작을 생성하고 복잡한 로봇 구동을 가능케 했다.

사람이 물리적인 힘을 가해 움직일 수 있는 로봇에는 적용 가능하지만 크기가 큰 자동차 및 대형 제품 제조 산업 로봇의 경우 직접 구동에 한계가 있다. 이 때 모션 인식 방법은 카메라와 센서만으로 사람 관절 및 자세를 추정하기 때문에 로봇 크기 제한 없이 전문가 기술을 데이터 전환할 수 있다[2]. 모션 인식 방법은 센서 활용 방법과 비전 방법이 있다. 센서 활용 방법[3]은 센서 및 장치를 부착하는 자기식/기계식과 키넥트(kinect) 카메라를 사용한 깊이 측정 방법 등이다(Fig. 1). 센서 활용 모션 인식은 3차원 공간 좌표로 추출이 가능하고 정확도가 높지만 센서 및 특수 장비 사용이 어렵고 공간적 제약이 있다. 반면 비전 방법은 별도의 장비 없이 자세 추정이 가능하여 사용하기 쉽다.

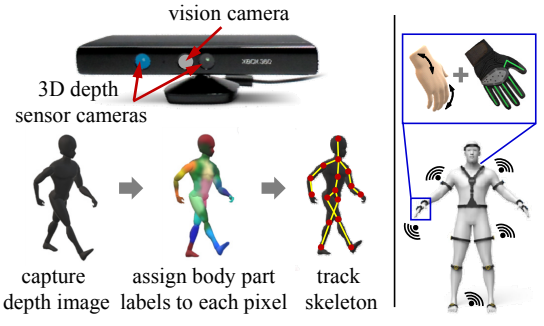

Fig. 1. Motion capture based on depth sensors

많은 분야에서 활용되는 합성곱 신경망(CNN: Convolutional Neural Networks)[4, 5]으로 빅데이터를 학습해 자세 추정을 가능케 하며, 이미지를 입력데이터로 사용해 반복적인 학습으로 오차를 줄이며 최적의 값을 도출한다. CNN은 구조상 이미지 데이터의 크기와 해상도에 따라 출력값 및 정확도가 달라질 수 있으며 이미지 필터링이나 다른 요인에 의해 차이가 발생할 수 있다.

따라서 본 연구는 기존 모션 인식 알고리즘 사용 시 관절 추출에 영향을 미치는 인자에 대하여 영향도만을 확인하고자 한다. 모션 인식 알고리즘은 오픈소스가 제공되어 쉽게 사용 가능할 수 있는 OpenPose를 사용하며 영향도만을 확인하기 위해 MPII, COCO, COCO+foot 데이터 세트로 미리 학습된 모델을 사용한다. CNN기반 OpenPose 구조를 고려하여 매개변수는 액세서리 유무, 입력 이미지 크기 및 이미지 필터링으로 정한다. 액세서리는 얼굴 특징을 감추기 위해 모자, 마스크, 안경을 사용한다. 이미지 크기는 원본 대비 70, 50, 30 및 10% 축소하고 emboss, sharpen 및 blur 필터링을 사용한다. 신체 조건이 다른 4명의 다양한 포즈를 스마트 폰으로 촬영해 이미지 데이터로 사용한다. 3가지 모델에 적용한 후 영향도 분석은 keypoint 성능 지표로 많이 사용되는 PCK(Percentage of Correct Key-point)와 기준 값, 예측 값 사이의 거리 오차 평균으로 나타낸다. 이를 통해 결과를 분석하여 각 모델에 대한 영향도를 확인하고자 한다.

2. 본론

2.1 학습 모델 설정

2.1.1 OpenPose를 활용한 포즈 추정

특수 카메라 및 장비가 필요 없는 비전 기반 모션 인식 OpenPose[6]의 구조는 Fig. 2와 같다. 입력 이미지는 VGG-19 네트워크 10개의 layer를 거쳐 하나의 특징(F)으로 출력된다. 출력된 값(F)은 다시 stage로 입력되어 branch1의 경우 confidence map의 S^t를 통해 사람 관절 위치를 예측하고, branch2는 affinity field의 L^t를 통해 추출된 관절에 해당하는 사람을 예측한다. 이러한 two-branch 구조는 반복적으로 이루어지며 confidence map과 affinity field의 결과 값들을 Eq. (1), (2)와 같은 손실함수를 적용해 다음 stage로 입력된다. 이 때 S_j^*는 confidence map의 ground truth이며 L_c^*는 affinity field의 ground truth, W는 이미지 상의 p위치에서 값이 누락된 경우 $W(p)=0$인 바이너리 마스크다. 이러한 과정을 통해 추출된 confidence map과 affinity field로 각 관절 part를 조합하여 포인트를

추출한다.

OpenPose는 CNN으로 MPII, COCO, COCO+foot 데이터 세트를 사용해 학습 및 검증을 통해 관절 keypoint 추출이 가능한 모델을 개발했다. 이 때 MPII 데이터 세트는 머리, 목, 양쪽 어깨, 팔꿈치, 손목, 양쪽 엉덩이, 무릎, 발목 및 몸통으로 0~14까지 총 15개의 keypoint 추출이 가능하다[Fig. 3(a)]. COCO는 MPII 추출 위치와 얼굴에서 양쪽 눈과 귀, 코의 0~17까지 총 18개의 keypoint 추출이 가능하다[Fig. 3(b)]. Fig. 3(c)에 나타낸 COCO+foot은 발가락, 발뒤꿈치 및 엉덩이 사이의 보간점도 확인 가능해 총 25개 keypoint가 추출된다.

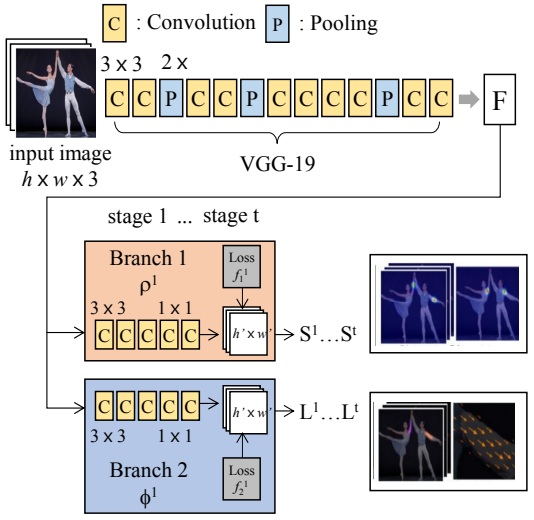

Fig. 2. Network structure of OpenPose

$$f_S^t = \sum_{j=1}^{J} \sum_{p} W(p) \cdot \| S_j^t(p) - S_j^*(p) \|_2^2 \quad (1)$$

$$f_L^t = \sum_{c=1}^{C} \sum_{p} W(p) \cdot \| L_c^t(p) - L_c^*(p) \|_2^2 \quad (2)$$

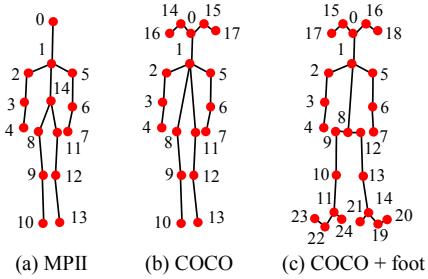

Fig. 3. "Keypoints" for MPII, COCO and COCO + foot

2.2 매개변수 설정 및 이미지 데이터 확보

OpenPose 3 모델(MPII, COCO, COCO+foot)를 사용해 자세 추정 시 영향을 미칠 수 있는 요인을 매개변수로 설정했다. 먼저 얼굴 눈, 코, 입을 가렸을 때 다른 관절 추출에 영향을 미치는지 확인하고자 액세서리 마스크, 모자, 안경을 사용했다[Fig. 4(a)]. COCO 및 COCO+foot 데이터 세트는 얼굴 keypoint 추출이 가능하지만 MPII의 경우 확인 불가능하기 때문에 이에 대한 차이를 확인하고자 한다. 두 번째로 이미지 크기를 매개 변수로 설정했다[Fig. 4(b)].

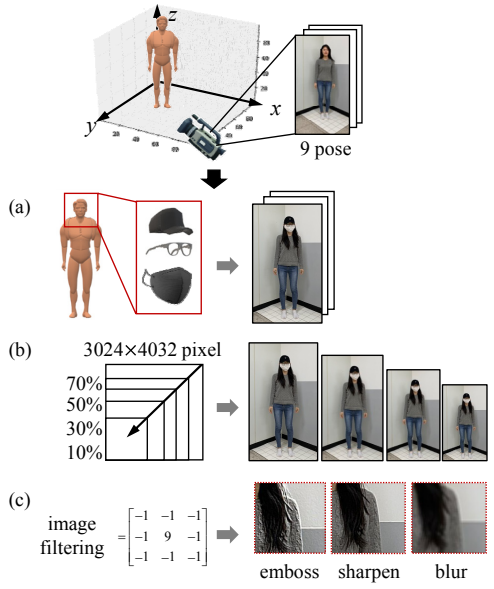

Fig. 4. Determination of parameters (accessories, image size and filtering) from captured images

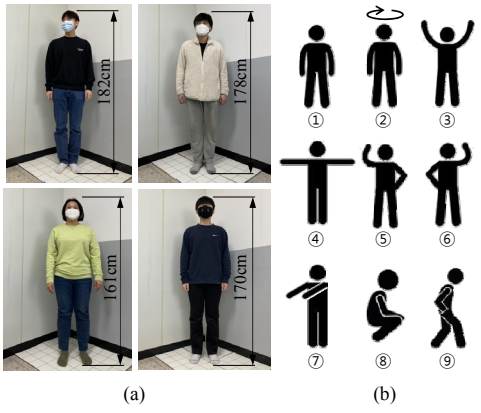

Fig. 5. The 4 people and 9 postures used for image data analysis

Table 1. Camera specification

Resolution	12 MP f/2.2 primary camera
Image Resolution	4000×3000 pixels
Camera Feature	10×digital zoom, 2×optical zoom
Physical Aperture	F2.2

CNN은 이미지를 입력 데이터로 사용할 때 층마다 커널(kernel)로 크기를 줄여가며 하나의 특징 값으로 나타내므로 입력 데이터의 크기에 따라 달라지는 경향을 확인한다. 마지막으로 배경과 인물간의 차이를 명확히 하고자 sharpen와 emboss 필터링을 사용했으며, 필터링 기법 중 자주 사용되는 blur를 사용했다[Fig. 4(c)]. 이미지 촬영은 스마트 폰 iPhone 11 Pro Max 카메라를 사용했으며, 카메라 사양을 Table 1에 나타냈다. 각 신체 조건이 다른 4명의 인물에 대해 차렷, 뒷모습, 양팔 들기, 양팔 벌리기, 한 팔 들기(왼팔, 오른팔), 앞으로 나란히, 쪼그려 앉기 및 달리는 자세인 총 9가지 포즈로 36장의 이미지를 촬영했다(Fig. 5). 촬영된 이미지 원본 크기는 3024×4032pixel 이고, 이미지 크기 변환은 원본 대비 70, 50, 30 및 10%로 축소해 사용했다. 또한 sharpen, emboss 및 blur 이미지 필터링을 적용하여 각 데이터를 확보했다.

2.3 영향도 확인 방법

각 모델(MPII, COCO, COCO+foot)은 관절 번호와 추출 가능 영역이 다르다. 따라서 공통으로 추출 가능한 영역인 목, 오른쪽 및 왼쪽 어깨, 팔꿈치, 손목, 오른쪽 및 왼쪽 엉덩이, 무릎, 발목으로 총 13개 keypoint만 확인했다[Fig. 6 (a)]. 이 때 모델로 추출된 점의 정확도 판단을 위해 이미지마다 임의로 기준점을 생성하고 모델로 예측한 값을 y_i, 기준 값을 y이라 한다. 이에 keypoint 성능 지표로 흔히 사용되는 PCK로 True와 False를 판단했다. PCK는 전체 이미지 데이터에서 추출된 모든 keypoint 중 정확히 예측된 keypoint 개수를 백분율로 나타내 평가하는 방법이다[7, 8]. 임계값 α에 따라 True와 False를 판단하며, 본 연구에서는 몸통 직경 b의 0.2배를 α로 정했다. 기준 값과 예측 값 사이 거리 d가 α이하일 때 True, 초과일 때 False로 나타낸다. 또한 True/False 이외에도 거리 오차 값을 확인하고자 d를 모든 이미지의 거리 오차를 구해 각 keypoint마다 평균을 구해 확인했다(Eq. 3).

$$error = \frac{1}{n}\sum_{i=1}^{n}|y - y_i| \tag{3}$$

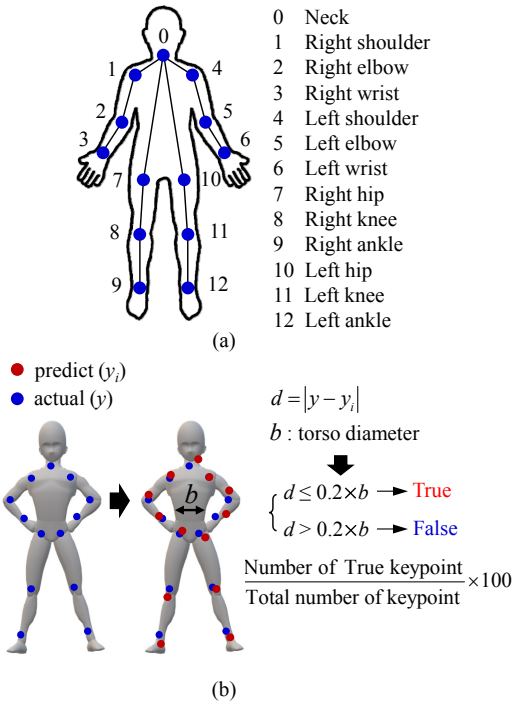

Fig. 6. (a) 13 keypoints used for comparing MPII, COCO and COCO+foot; (b) PCK based on predicted and real coordinates

2.4 매개변수 별 영향도 분석

3가지 매개변수에 따른 영향도 분석을 Fig. 6에서 제시한 keypoint에 대해 그래프로 나타냈으며 앞서 제시한 PCK와 거리 오차 평균 방법으로 영향도를 분석했다. 이 때 이미지에서 관절 위치를 감지하지 못했을 경우 해당 keypoint는 제외했다.

Fig. 7은 액세서리 유무에 따라 거리 오차 평균으로 나타낸 그래프이다. 모자, 마스크, 안경을 착용했을 때와 마스크만 착용 했을 경우, 아무것도 착용하지 않았을 경우 3가지를 그래프로 나타냈다. 이를 통해 MPII가 가장 큰 차이를 보이며, 모두 착용하지 않았을 경우와 모든 악세서리를 착용한 데이터 사이 최대 290pixel 차이를 보인다. 반면 COCO+foot의 경우 모든 keypoint가 50pixel 이하이며, 최대 12pixel로 상대적으로 차이가 작다.

이미지 크기 변환 시 마스크만 착용한 이미지를 원본 크기 대비 다양한 비율로 축소해 사용했다. 이를 각 모델로 예측하여 PCK와 거리 오차 평균을 Fig. 8에 나타냈

다. 이미지 크기 변환은 MPII가 가장 큰 오차를 보이며, COCO, COCO+foot 순으로 오차가 작아진다. 축소된 이미지 간 최대 오차는 MPII, COCO, COCO+foot 순으로 각 80, 40, 9 pixel로 확인된다. PCK도 마찬가지로 COCO+foot는 가장 근소한 차이를 보이며 상대적으로 MPII가 가장 큰 차이를 보인다. 모든 모델에서 원본 대비 70% 축소된 이미지가 다른 비율에 비해 거리 오차 평균이 대부분 가장 작으며 PCK 또한 좋다.

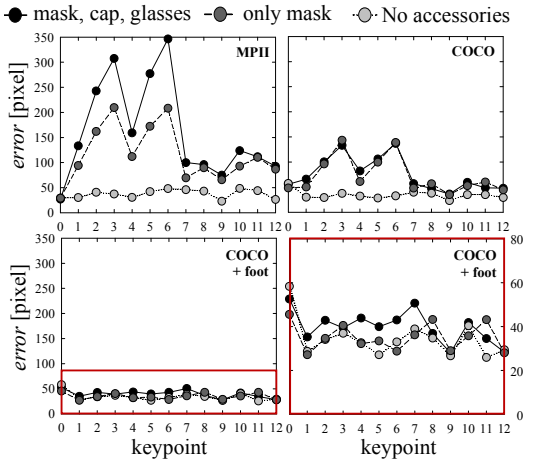

Fig. 7. Comparison of distance error for accessories

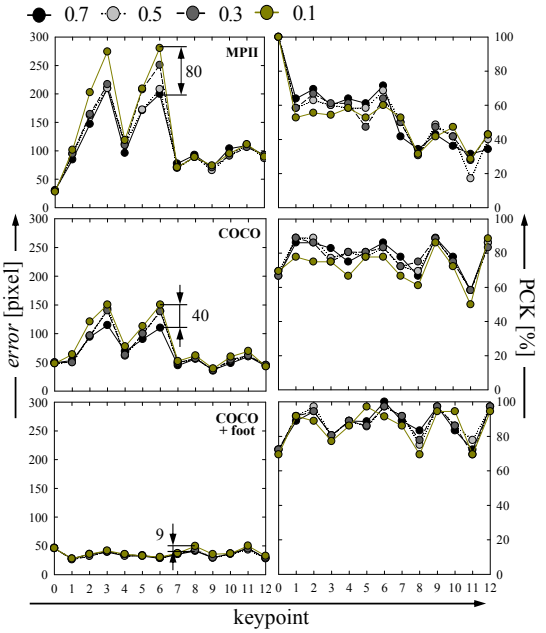

Fig. 8. Comparison of distance error and PCK for 4 different image sizes

그러나 MPII의 경우 70% (2117×2822) 비율과 거리 오차 평균의 차이를 비교할 때 50% (1512×2016)와 약 8pixel이나, 30% (900 ×1200) 및 10% (302×403)이상 축소 시 최대 80pixel로 약 10배 크게 나타난다. 따라서 MPII나 COCO의 경우 입력 데이터의 적정 크기는 약 1500×2000이상이고, 이미지 크기가 작아질수록 자세 추정 알고리즘의 정확도가 떨어질 수 있다.

마지막으로 emboss, sharpen 및 blur 필터링하여 적용 전 이미지와 비교하여 그래프로 나타냈다(Fig. 9). 앞서 두 가지 매개변수와 같은 경향을 보인다. sharpen의 경우 모든 모델에 대하여 필터링 적용 전 이미지 보다 정확도가 감소했으며 blur 필터링도 이와 유사하다. emboss의 경우 MPII에 대해서만 거리 오차 평균이 최대 60pixel 감소했다. 이를 통해 COCO, COCO+foot의 경우 이미지 필터링 적용은 좋은 효과를 얻지 못하나 MPII의 경우 emboss 필터링으로 정확도가 향상될 수 있다. 각 3가지 모델에 대한 분석 결과 추출 가능한 관절이 많을수록 매개변수에 대한 영향을 거의 받지 않았다. 그러나 정확도는 이미지 전처리나 크기 변환에 따라 향상될 수 있음을 보인다.

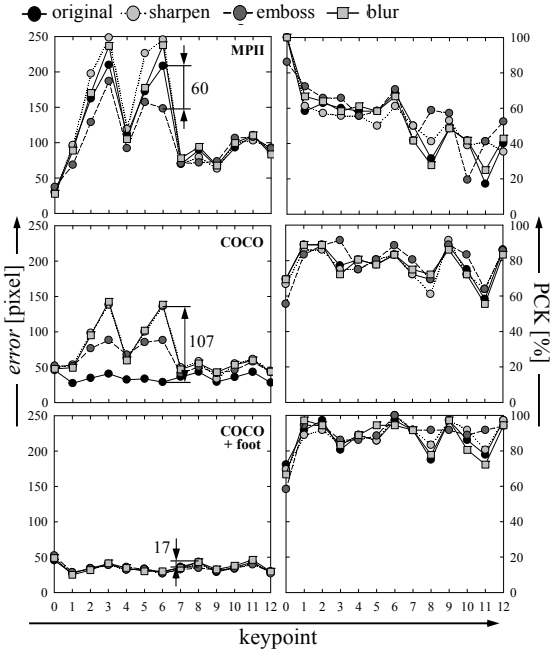

Fig. 9. Comparison of distance error and PCK according to different preprocessing of image

3. 결론

본 연구에서는 모션 인식을 위한 2D 자세 추정 알고리듬의 이미지 전처리 및 얼굴 가림에 대한 영향도를 분석했으며, 다음과 같은 결론을 얻었다.

(1) 제조 로봇에 전문가의 노하우를 적용할 때 모션 인식으로 데이터 전환이 가능하다. 이 때 비전 모션 인식은 공간상의 제약이 없으나 입력 데이터에 따라 결과 값이 달라지므로 여러 요인에 대한 영향을 확인했다.

(2) 매개변수는 액세서리(모자, 마스크, 안경) 착용 유무, 이미지 크기 및 필터링(emboss, sharpen, blur)로 정하고 OpenPose의 모델(COCO, MPII, COCO+foot 데이터로 학습)을 사용했다.

(3) 영향도 판단은 PCK와 기준 및 예측 값 사이 거리 오차 평균으로 확인했다. 모든 매개변수에 대하여 MPII가 가장 큰 차이를 보이고, COCO+foot의 경우 차이가 가장 작다. 이미지 크기 변환 시 MPII는 원본 대비 50% 축소할 때 오차 거리의 최대 차가 8pixel이나 30, 10% 축소 시 최대 80pixel로 약 10배 크게 나타난다.

(4) 이미지 필터링은 세 모델 모두 PCK가 각 포인트마다 다르다. sharpen 및 blur 필터링은 모든 모델에서 정확도가 낮아지나 emboss의 경우 MPII에서만 거리 오차가 필터링 전보다 작다.

이와 같은 연구를 확장시켜 kinect 및 다른 센서와 함께 적용할 때 더 정확한 데이터 확보가 가능할 것으로 기대된다.

References

[1] P. K. Kim, H. Park, J. H. Bae, J. H. Park, D. H. Lee, "Intuitive Programming of Dual-Arm Robot Tasks using Kinesthetic Teaching Method", *The Journal of Institute of Control, Robotics and Systems*, Vol.22, No.8 pp.656-664, 2016.
DOI: https://dx.doi.org/10.5302/J.ICROS.2016.16.0102

[2] H. H. Jung, M. K. Kim, J. Lyou, "Implementation of Hybrid Motion Capture System for Behaviour Pattern Analysis of Disaster Recovery Workers", *The Journal of Institute of Control, Robotics and Systems*, Vol.23, No.5 pp.323-331, 2017.
DOI: http://dx.doi.org/10.5302/J.ICROS.2017.17.0053

[3] J. S. Kim, H. Park, "Working Posture Analysis for Preventing Musculoskeletal Disorders using Kinect and AR Markers", *Korean Journal of Computational Design and Engineering*, Vol.23, No.1, pp.19-28, 2018.
DOI: http://dx.doi.org/10.7315/CDE.2018.019

[4] J. J. Park, C. K. Kwon, "Study on Forearm Muscles and Electrode Placements for CNN based Korean Finger Number Gesture Recognition using sEMG Signals", *Journal of the Korea Academia-Industrial cooperation Society*, Vol.19, No.8, pp.260-267, 2018.
DOI: http://dx.doi.org/10.5762/KAIS.2018.19.8.260

[5] M. J. Kang, "Comparison of Gradient Descent for Deep Learning", *Journal of the Korea Academia-Industrial cooperation Society*, Vol.21, No.2, pp.189~194, 2020.
DOI: http://dx.doi.org/10.5762/KAIS.2020.21.2.189

[6] Z. Cao, T. Simon, S. E. Wei, Y. Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields", *Proceeding of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, HI, USA, pp.7291-7299, July 2017.
DOI: http://dx.doi.org/10.1109/CVPR.2017.143

[7] M. Andriluka, L. Pishchulin, P. Gehler, B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis", *Proceeding of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, OH, USA, pp.3686-3693, June 2014
DOI: http://dx.doi.org/10.1109/CVPR.2014.471

[8] Y. Yang, D. Ramanan, "Articulated human detection with flexible mixtures of parts", *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.35, No.12, pp.2878-2890, Dec. 2013.
DOI: http://dx.doi.org/10.1109/TPAMI.2012.261

노 은 솔(Eunsol Noh) [준회원]

- 2020년 2월 : 국립공주대학교 금형설계공학과
- 2020년 3월 ~ 현재 : 국립공주대학원 융합기계공학과 석사과정

〈관심분야〉
인공지능, 머신러닝

이 사 랑(Sarang Yi) [준회원]

- 2019년 2월 : 국립공주대학교 금형설계공학과
- 2019년 3월 ~ 현재 : 국립공주대학교 기계공학과 석사과정

〈관심분야〉
인공지능, 머신러닝

홍 석 무(Seokmoo Hong) [종신회원]

- 1999년 2월 : 서강대학교 기계공학과 (기계공학 학사)
- 2001년 2월 : 서강대학교 기계공학과 (기계공학 석사)
- 2007년 3월 : Technical University of Munich, Germany, Department of Mechanical Engineering (기계공학박사)
- 2007년 4월 ~ 2015년 2월 : 삼성전자 GTC, 수석연구원
- 2015년 3월 ~ 현재 : 국립공주대학교 기계자동차공학부 교수

〈관심분야〉
금속 판재성형 및 단조, 유한요소해석, 최적 설계