KIPS Trans. Softw. and Data Eng.
Vol.10, No.5 pp.187~194
ISSN: 2287-5905 (Print), ISSN: 2734-0503 (Online)

https://doi.org/10.3745/KTSDE.2021.10.5.187

LSTM(Long Short-Term Memory)-Based Abnormal Behavior Recognition Using AlphaPose

Hyun-Jae Bae† · Gyu-Jin Jang†† · Young-Hun Kim††† · Jin-Pyung Kim††††

ABSTRACT

A person's behavioral recognition is the recognition of what a person does according to joint movements. To this end, we utilize computer vision tasks that are utilized in image processing. Human behavior recognition is a safety accident response service that combines deep learning and CCTV, and can be applied within the safety management site. Existing studies are relatively lacking in behavioral recognition studies through human joint keypoint extraction by utilizing deep learning. There were also problems that were difficult to manage workers continuously and systematically at safety management sites. In this paper, to address these problems, we propose a method to recognize risk behavior using only joint keypoints and joint motion information. AlphaPose, one of the pose estimation methods, was used to extract joint keypoints in the body part. The extracted joint keypoints were sequentially entered into the Long Short-Term Memory (LSTM) model to be learned with continuous data. After checking the behavioral recognition accuracy, it was confirmed that the accuracy of the "Lying Down" behavioral recognition results was high.

Keywords : Safety Management, Action Recognition, Pose Estimation, LSTM, Deep Learning

AlphaPose를 활용한 LSTM(Long Short-Term Memory) 기반 이상행동인식

배 현 재† · 장 규 진†† · 김 영 훈††† · 김 진 평††††

요 약

사람의 행동인식(Action Recognition)은 사람의 관절 움직임에 따라 어떤 행동을 하는지 인식하는 것이다. 이를 위해서 영상처리에 활용되는 컴퓨터 비전 태스크를 활용하였다. 사람의 행동인식은 딥러닝과 CCTV를 결합한 안전사고 대응서비스로서 안전관리 현장 내에서도 적용될 수 있다. 기존연구는 딥러닝을 활용하여 사람의 관절 키포인트 추출을 통한 행동인식 연구가 상대적으로 부족한 상태이다. 또한 안전관리 현장에서 작업자를 지속적이고 체계적으로 관리하기 어려운 문제점도 있었다. 본 논문에서는 이러한 문제점들을 해결하기 위해 관절 키포인트와 관절 움직임 정보만을 이용하여 위험 행동을 인식하는 방법을 제안하고자 한다. 자세추정방법(Pose Estimation)의 하나인 AlphaPose를 활용하여 신체 부위의 관절 키포인트를 추출하였다. 추출된 관절 키포인트를 LSTM(Long Short-Term Memory) 모델에 순차적으로 입력하여 연속적인 데이터로 학습을 하였다. 행동인식 정확률을 확인한 결과 “누워있기(Lying Down)” 행동인식 결과의 정확도가 높음을 확인할 수 있었다.

키워드 : 안전관리, 행동인식, Pose Estimation, LSTM, 딥러닝

1. 서 론

사람의 행동인식기술은 다양한 센서를 활용하여 사람 동작 데이터를 수집하고 해석하여 행동을 인식하는 기술이다. 사람의 행동을 인식하는 기술 중에서 가장 핵심적인 부분은 입력 신호에서 분석해야 할 행동들의 특징(Feature) 정보를 추출하는 것이다. 행동들의 특징 정보는 사람의 관절에 해당하는 키포인트를 추출하고 이를 통해 자세추정(Pose Estimation) 수행한다[1].

자세추정방법은 일반적으로 사람 관절에 해당하는 키포인트와 키포인트를 연결한 집합(스켈레톤)으로 자세를 추정한다. 자세추정방법은 크게 상향식 방법과 하향식 방법으로 구분된다. 상향식 방법은 CCTV 영상에 포함된 사람의 관절을 모두 추출하고, 관절의 상관관계를 분석하여 자세를 추정하는 방법이며 대표적으로 OpenPose가 있다. 하향식 방법은 영상에서

※ 이 연구는 차세대융합기술연구원의 지원으로 수행되었음(AICT-2020-0001).
※ 이 논문은 행정안전부 극한재난대응기반기술개발사업의 지원을 받아 수행된 연구임(2020-MOIS31-014).
† 준 회 원 : 차세대융합기술연구원 연구원
†† 정 회 원 : 차세대융합기술연구원 연구원
††† 비 회 원 : 차세대융합기술연구원 선임연구원
†††† 정 회 원 : 차세대융합기술연구원 선임연구원
Manuscript Received : December 18, 2020
First Revision : February 26, 2021
Accepted : March 6, 2021
* Corresponding Author : Jin-Pyung Kim(jpkim@snu.ac.kr)

사람을 먼저 찾고, 바운딩박스(BoundingBox) 내부에서 자세를 추정하는 방법으로서 AlphaPose[1]와 Mask-RCNN이 주로 사용된다. 여러 방법 중 관절 추출정확도는 AlphaPose가 가장 높으며[1], 다수의 사람을 검출할 수 있고, CCTV 높이에서 사람의 이상행동인식이 가능하다. 그러나 AlphaPose는 작업현장에 적용된 사례는 기존 연구에서는 발견할 수 없었다. 따라서 본 연구에서는 AlphaPose를 활용하여 작업현장 내 안전사고를 예방하려는 연구를 시도하였다. AlphaPose는 현장 작업자들의 행동인식 시스템으로 활용 가능하며 이는 현장 작업 중 이상 상태를 감시하여 안전사고를 예방하기 위한 목적이 있다.

CCTV를 활용하여 작업현장 내 사람의 관절 움직임으로 자세를 추정하기 위한 연구는 부족하였으며, 다른 방법은 고가의 장비 또는 복잡한 장비 구성으로 인해 현장 적용에 어려움이 있다. 따라서 영상의 프레임만을 이용하여 사람의 자세를 추정하는 방법은 작업현장 내 이상행동을 인식하여 안전사고를 예방하기 위한 목적으로 필요하다.

본 논문에서는 HRNet[2] 기반의 관절 키포인트 추출을 통하여 AlphaPose를 활용하고, 획득한 관절 키포인트를 관절의 움직임을 통하여 행동을 인식하는 방법을 제안한다. 연속된 데이터를 활용하여 영상기반 시계열 예측으로 행동인식을 나타낸다. 이러한 시계열 예측을 위해 LSTM(Long Short-Term Memory)[3]을 사용하였다. 3가지 자세추정방법(Tf-Pose-Estimation, OpenPose, AlphaPose)을 사용하여 관절 키포인트 추출에 대한 성능을 비교하였다.

2장에서는 자세추정과 행동인식 관련 연구를 살펴보고, 3장에서는 행동인식 모델을, 4장에서는 작업현장에서 활용될 데이터셋과 위험 행동인식 실험을 분석하고 평가하였으며, 5장에서는 결론 및 향후 방향을 제시하였다.

2. 관련 연구

2.1 AlphaPose를 활용한 관절 키포인트 추출

상하이 대학교의 Fang 팀은 부적절한 바운딩박스가 있는 상황에서 지역 다중 자세추정방법을 가능하게 하는 AlphaPose 모델을 제안하였다[1]. AlphaPose는 다중 사람에 대한 자세추정방법을 실시간으로 할 수 있으며[4], Fig. 1과 같이 총 18개의 관절 키포인트를 추출하였다. AlphaPose는 하향식 방법으로 상향식 방법 대비 정확도와 효율성 측면에서 State-of-The-Art(SOTA) 결과에 준하는 성능을 보였다.

Equation (1)은 AlphaPose의 관절 추출에 대한 식이다. 객체의 모든 관절 키포인트에 대한 분포가 낮을수록 각각의 컨피던스(Confidence) 점수가 높을수록 좋다. tanh는 신뢰도가 낮은 값을 필터링하며 Pi와 Pj의 두 좌표가 일치할수록 tanh의 미분된 값은 1에 근접해진다.

Fang 팀은 Table 1에서 AlphaPose와 다른 방법들의

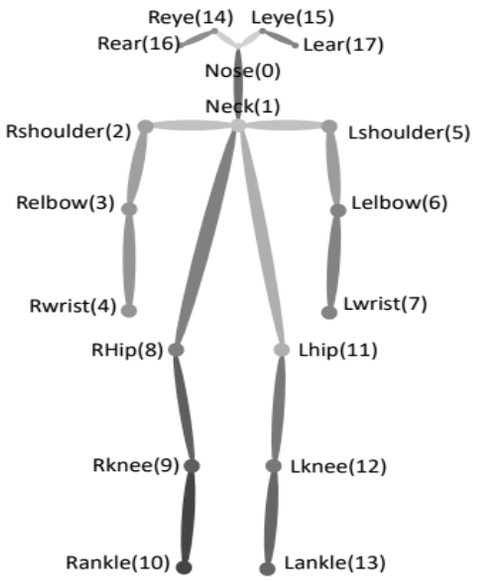

Fig. 1. The Human Skeleton Model with 18 Joints

$$K_{Sim}(P_i,P_j \mid \sigma_1) = \begin{cases} \sum_n \tanh\frac{c_{i^n}}{\sigma 1} \cdot \tanh\frac{c_{i^n}}{\sigma 1}, & \text{if } k_j^n \text{ is within } \beta(k_j^n) \\ 0 & otherwise \end{cases} \quad (1)$$

K_j : $Coordinate\ of\ Key\ Point$
P_i : $Coordinate\ of\ BoundingBox$
P_j : $Coordinate\ of\ BoundingBox$

정확률을 비교하기 위해 COCO 데이터셋[4]에서의 정밀도(AP)를 측정하였다. AP@0.5:0.95는 IoU(Intersection over Union: 교집합 영역 넓이 / 합집합 영역 넓이)의 이상치(Threshold)를 0.5부터 0.95까지 0.05의 간격으로 달리 줬을 때의 AP 평균을 의미한다. AlphaPose는 AP@0.5:0.95 측정 결과 73.3%로서 다른 방법들보다 정확한 것으로 판명되었다.

Table 1의 관절 키포인트 추출 성능을 비교해 본 결과, Fig. 2와 같이 OpenPose는 상향식방법으로 속도는 빠르지만 정확성은 떨어지고, Mask-RCNN은 속도는 느리나 정확성은 높다. 본 논문에서는 기존 방법인 OpenPose[5]와 Detectron[6]보다 정밀하게 관절 키포인트를 추출하는 AlphaPose를 채택하였다.

Table 1. Results on COCO Test-dev 2015[1]

Method	AP@ 0.5:0.95	AP@ 0.5	AP@ 0.75	AP medium	AP large
OpenPose (CMU-Pose)	61.8	84.9	67.5	57.1	68.2
Detectron (Mask R-CNN)	67.0	88.0	73.1	62.2	75.6
AlphaPose	73.3	89.2	79.1	69.0	78.6

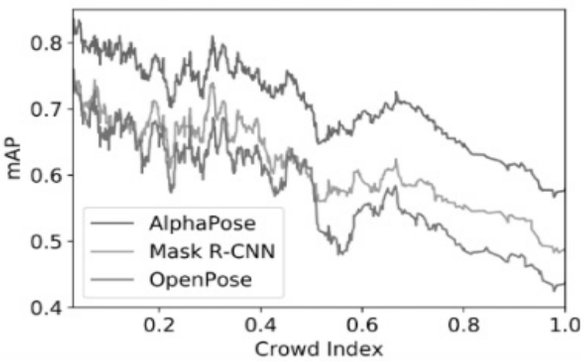

Fig. 2. Relationship between Crowd Index and Landmark Average Precision on COCO Dataset[7]

2.2 LSTM을 활용한 행동인식 구조

프레임마다 사람의 관절에서 34(17*2)개의 키포인트 값을 추출하여 벡터로 변환하고 연속된 프레임을 모아 데이터를 만들었다. 형성된 관절 키포인트 데이터는 학습 데이터와 테스트 데이터를 8:2의 비중으로 나누었다.

추출된 관절의 키포인트들은 영상의 프레임에 대해 시계열 예측이 가능한 RNN(순환신경망:Recurrent Neural Network)의 은닉층(Hidden State)에 기억층(Cell State)을 추가한 구조인 LSTM 모델에 입력 데이터로 사용하였다[8]. RNN의 기울기 소실 문제를 극복하기 위해 고안된 LSTM을 프레임마다 관절의 움직임을 기억하도록 내부 파라미터를 수정하였다. 입력 데이터를 평균이 0, 분산이 1인 데이터 분포로 바꾸고 균일한 분포의 데이터로 전환하여 손실함수에서 학습이 잘 진행하도록 하였다. 여러 클래스(Class)를 분류하기 위해 크로스 엔트로피(Cross Entropy) 함수를 사용하였다. Equation은 크로스 엔트로피 식이다(2). 실제 환경의 값은 q이고 모델의 예측 값은 p이다. 실제 분포 q에 대해 알지 못한 상태 중 p를 통하여 q를 예측하는 것이다. 주목적은 실제 값과 예측값의 차이를 줄이기 위함이다.

$$H_p(q) = -\sum_{i=1}^{n} q(x_i)\log p(x_i) \quad (2)$$

본 논문에서는 시계열 데이터를 다룰 수 있는 순환 신경망 모델 중 LSTM 모델을 활용하였으며 9프레임에 따라 사람 관절의 움직임을 활용해 행동인식 연구를 진행하였다. 미국직업안정 및 건강관리청(OSHA: Occupational Safety and Health Administration)의 안전보건사고통계목록 기반[9]으로 "쓰러지기(Falling Down)" 행동을 일련의 불안전한 행동으로 정의하였다. 영상인식 기반으로 작업자들의 안전을 관리할 수 있는 적정 기술을 검토하였다. 영상의 일정 프레임을 확인하고 관절의 움직임을 관찰하여 현장 내 이상 행동인식을 판단하기 위함이다[9].

3. 현장 안전관리를 위한 행동인식 모델

행동인식 모델을 활용하여 현장 안전을 체계적으로 관리하기 위한 목적으로 연구를 수행하였으며, 현장 안전관리를 위한 행동인식 모델의 연구 진행순서는 Fig. 3와 같다. 1단계 자세추정방법은 영상에서 HRNet을 활용하여 사람의 관절 키포인트를 추출한다. 2단계 LSTM 방법은 추출된 키포인트가 LSTM의 데이터로 입력되고 9프레임 단위로 행동을 인식한다. 이로써, 행동인식 모델을 설정하였다.

본 논문에서는 HRNet을 활용하여 20만개 이상의 관절 키포인트를 추출하고 사람의 행동인식 목적으로 구축한 LSTM 기반의 모델을 제안한다. 전통적인 머신러닝 방법인 랜덤포레스트(Random Forest)와 SVM(Support Vector Machine) 방법도 클래스에대해 다중분류가 가능하지만 데이터의 수가 많아질수록 랜덤 포레스트와 SVM 방법은 속도가 크게 떨어지는 문제점이 있다. 그러므로 속도가 더 빠르고 시계열 예측 성능이 좋은 LSTM으로 연구를 수행하였다.

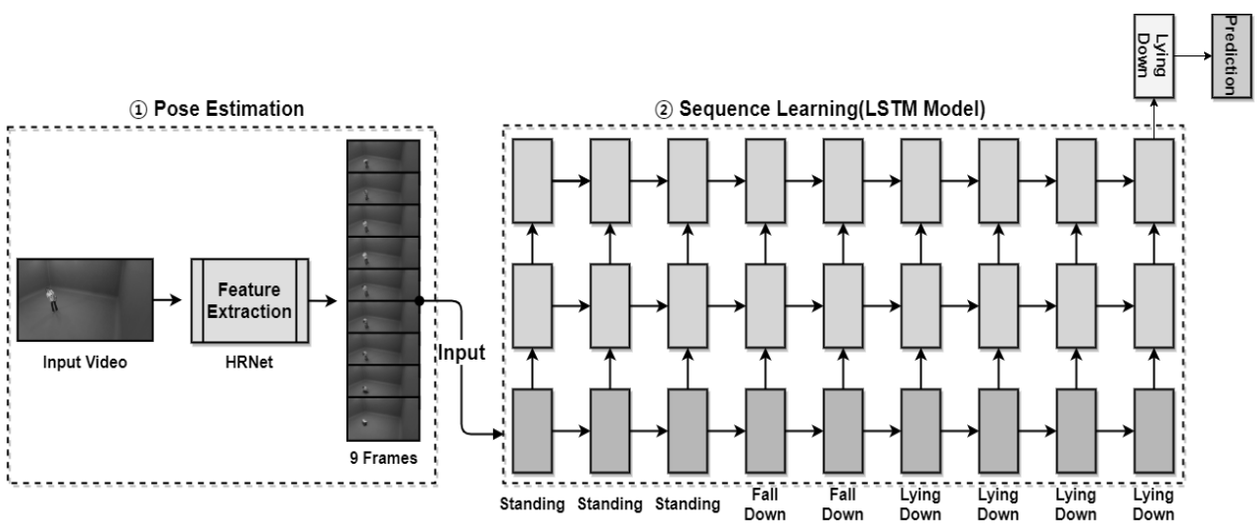

Fig. 3. The Proposed HRNet-LSTM Model

Table 2. Parameter Values of the Proposed LSTM Model

Parameters	Values(functions)
Keypoints(X,Y)	17 * 2
Num Layers	3
Hidden Size	128
Loss Function	Cross Entropy
Optimizer	Adam
Sequence Length	9
Epochs	200
Shuffle	True
Pin Memory	True

기존 방식은 입력 이미지를 저해상도로 인코딩한 상태에서 다시 고해상도로 복구하는 경우였다. 본 연구에서 활용한 HRNet은 피처맵(Feature Map)의 크기가 작아지고 깊이가 깊어질수록 고해상도 표현이 좋아진다는 장점이 있다[2]. HRNet은 고해상도 표현을 향상하기 위해 멀티 스케일 퓨전(Multi Scale Fusions)을 수행하고, 동일한 깊이와 유사한 수준의 저해상도 표현도 사용하였다. 그 결과 고해상도 표현은 자세추정방법에 좋은 성능을 보였다.

한국정보화진흥원(NIA) 주관으로 구축한 사람의 행동 6가지 영상데이터들로부터 관절의 키포인트 값을 추출 및 저장한다. 관절의 좌표 X, Y를 각각 17개씩 추출하여 관절 좌표의 변화에 따라 물건 옮기기, 쓰러지기, 걷기, 물건 들기, 서있기, 누워있기의 총 6가지 행동을 인식하는 모델이다. 현장 안전관리를 위한 모델은 현장 내 이상행동 4가지(물건 옮기기, 걷기, 물건 들기, 서있기)와 일반적인 행동 2가지(쓰러지기, 누워있기)로 분류되어 인식한다. 모델의 적합한 계산을 위해 아래 Table 2와 같이 세부 파라미터를 수정하고 최적화한다. 모델이 전달받는 시간 정보의 범위를 담당하는 프레임 수를 최적의 파라미터 값으로 조정한다.

LSTM 파라미터 중 손실함수를 Cross Entropy로 설정한 경우는 6가지 행동 클래스를 다중 분류하여 인식하기 위함이다. Optimizer를 Adam[11]으로 설정한 경우는 기존에 사용되었던 RMSProp보다 세밀하게 학습되며, 보다 학습속도를 빠르게 한다. 학습률은 RMSprop과 마찬가지로 Gradient 제곱의 이동평균에 반비례하도록 설정된다. Adam은 RMSprop과 Momentum의 장점을 모두 지니고 있어 학습 시 알고리즘에 주로 사용되고 있다. 데이터 구성 시 초당 30프레임이고, 9개의 프레임을 추출하는 방법은 장면마다 사람의 관절 키포인트를 추출하여 순차적으로 키포인트를 데이터 로더(DataLoader)로 불러오는 방법을 활용하였다. 장면의 연속되는 데이터 길이가 9일 때 한가지 행동으로 인식하는 것이다. 사람의 관절 키포인트 값들을 과적합 시키지 않기 위해 셔플(Shuffle) 파라미터를 True로 지정하였다. 메모리 가속화 목적으로 텐서(Tensor)를 CUDA 고정 메모리에 할당시키는 파라미터 설정으로 핀메모리(Pin Memory)를 True로 설정하였다. 추출된 데이터는 LSTM의 입력 데이터 길이와 같게 입력된다. 관절의 키포인트는 17개씩 2개로 구성되어 34개씩 순차적으로 입력된다. 34개의 관절 움직임을 9프레임 동안 확인하고 행동을 인식하는 모델로 Fig. 5와 같은 방식이다. 영상 프레임에 대해서 하나의 출력값인 행동을 인식하기 위해 Many-To-One 방식인 모델로 구축하였다.

Table 3. Number of Keypoints for 6 Action Classes

Actions	Train Set	Test Set
Carrying[물건 옮기기]	35,898	8,865
Falling Down[쓰러지기]	33,334	8,357
Walking[걷기]	33,303	8,151
Lifting[물건 들기]	27,086	6,772
Standing[서있기]	19,520	4,880
Lying Down[누워있기]	18,745	4,687
계	167,886	41,712

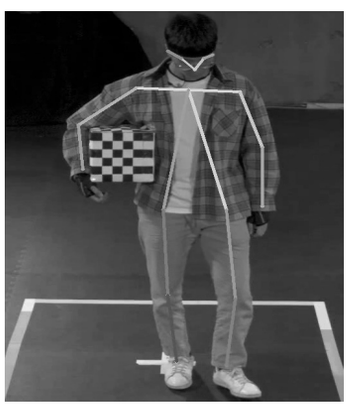

Fig. 4. Example of Joint Keypoint Extraction

4 실험 결과 및 평가

4.1 AI-Hub 사람 동작 데이터셋

한국정보화진흥원(NIA) 주관으로 구축한 AI 기술 및 제품 서비스 인프라인 AI-Hub의 데이터를 활용하였다[12]. 영상 데이터는 작업현장에서 활용되고 있는 물건 옮기기, 쓰러지기, 걷기, 물건 들기, 서있기, 누워있기의 총 6가지 사람 동작 영상을 수집하였다. 수집된 영상의 데이터로부터 사람의 관절을 추출하여 관절 키포인트 값을 저장하였다.

위 Table 3과 같이 영상으로부터 사람의 관절을 추출한 데이터셋 비율은 학습 데이터를 8, 평가 데이터를 2의 비중도로 나누었다. 학습 데이터의 총 데이터 수는 167,886개이고, 평가 데이터의 총 데이터 수는 41,712개이다. 데이터값은 카테고리 수가 아닌, 객관적 특징점의 좌표를 의미하는 것이다. 위 Table 3의 데이터 수는 단순히 영상으로부터 관절 키포인트를 추출한 값이다.

위 Fig. 4와 같이 행동의 영상은 각각 5~10초 길이이며,

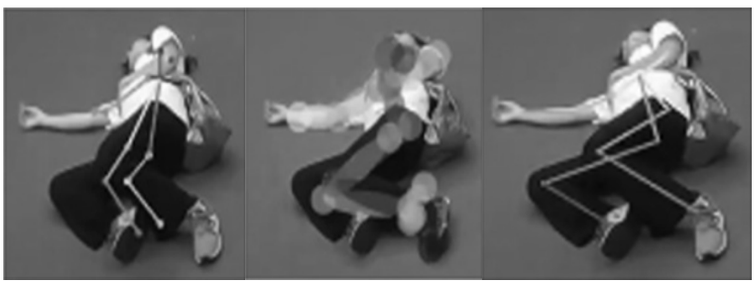

Fig. 5. (a) Tf-Pose-Estimation, (b) OpenPose, (c) AlphaPose. Comparison of Pose Estimation Performance

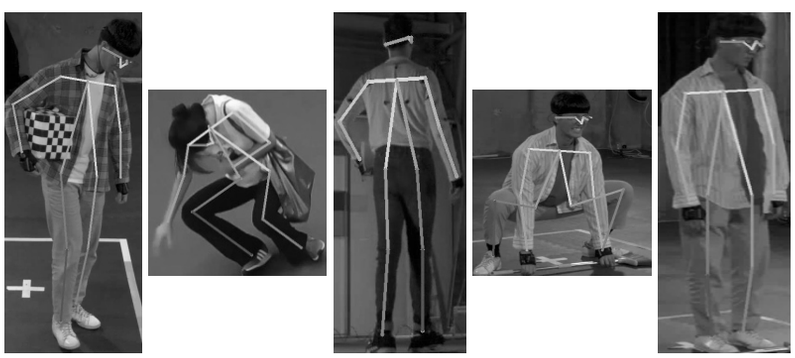

Fig. 6. Pose Estimation Experiments on Various Postures (a) Carrying, (b) Falling Down, (c) Walking, (d)Lifting, (e) Standing

영상의 크기는 1,920 x 1,080픽셀과 프레임 속도는 30Fps (Frame Per Second)이다.

한국정보화진흥원에서 운영하는 영상 데이터 촬영 방법은 동작 캡처 촬영을 위한 대략 8m x 8m 공간을 12대의 고속 카메라를 동기화하여 촬영을 진행하였다고 명시되어 있다[12]. 20대 배우들이 자이로 슈트를 착용하고 그 슈트 위에는 평상복을 입어 여러 명의 연기자의 동작을 수집하였다. 고속 카메라 12대는 서로 간에 동기화와 트리거를 연결하고 동시에 촬영하여 영상 클립의 싱크를 맞추는 방식으로 타임 동기화를 진행하였다.

4.2 실험 환경

하드웨어 실험 환경은 3.60GHz의 Intel(R) core I7-9700K CPU와 NVIDIA RTX 2070 SUPER Founders Edition D6 8GB GPU, 32GB RAM, Linux Ubuntu 18.04 운영체제가 설치된 데스크톱 PC에서 실험하였다. 딥러닝 프레임워크 중 최근 많이 사용되는 PyTorch 1.1.0 버전을 사용하여 실험 시뮬레이션을 수행하였으며 Python을 프로그래밍 언어로 사용하였다.

4.3 모델의 실험 결과 및 평가

행동인식의 성능을 높이기 위해서는 자세추정방법에서 관절의 키포인트 추출에 대한 성능을 먼저 확인해야 한다[13,14]. 재난안전분야에서 사람이 특정 자세를 취하였을 때 관절의 시각화(Visualization)가 정상적으로 되는지와 관절의 관계성을 비교할 목적으로 3가지 자세추정방법을 사용하였다. 테스트 영상은 비정상적인 행동인 '누워있기(Lying Down)'로 진행하였다[15]. 첫 번째로 Tf-Pose-Estimation, 두 번째로 OpenPose 마지막으로 AlphaPose 순으로 진행하였다. 각각 추출된 관절의 개수는 18개, 25개, 18개이다. 실험 결과는 Fig. 5와 같다.

실험의 첫 번째 Tf-Pose-Estimation의 '누워있기' 인식 결과로는 눈, 코, 귀, 골반, 팔과 다리의 관절 위치와 관절과 관절 간 길이의 대응 여부를 확인하였다. 두 번째 OpenPose의 '누워있기' 인식 결과로는 눈, 코, 귀, 팔, 골반의 관절 위치와 관절과 관절 간 길이는 정상적으로 대응되나 오른쪽 다리 관절의 위치가 왼쪽 다리가 비정상적으로 겹치는 결과가 나왔다. 마지막 AlphaPose의 '누워있기' 결과로는 모든 관절의 위치가 다른 방법들보다 정상적으로 위치하며 관절과 관절 간 길이가 정확히 대응됨을 확인하였다.

'누워있기' 행동에 대한 AlphaPose 성능을 확인한 후, Fig. 6과 같이 나머지 행동 5가지(물건 옮기기, 쓰러지기, 걷기, 물건 들기, 서있기)에 대해서도 관절의 길이가 정확히 대응되는지 실험을 순차적으로 진행하였다.

총 6가지의 행동 데이터를 학습 데이터 8, 평가 데이터 2로 나누어 활용하였다. 모델 학습 시 옵티마이저(Optimizer)를 RMSProp으로 설정한 경우 Fig. 7A와 같은 실험 결과를 보였다. 학습결과의 정확도는 96.69%이고, 평가결과의 정확도는 92.48%이다. Optimizer를 Adam으로 설정한 경우 Fig. 7B와 같은 결과를 보였다. 학습결과의 정확도는 96.75%이고, 평가결과의 정확도는 95.63%이다. 텐서보드(Tensorboard)를 활용하여 실험 결과를 그래프로 그려 시각화하였다. 학습결과의 정확도 부분에서는 차이가 작았으나, 평가결과의 정확도에서는 Adam을 사용했을 때, 학습방향과 스텝 크기를 최적화

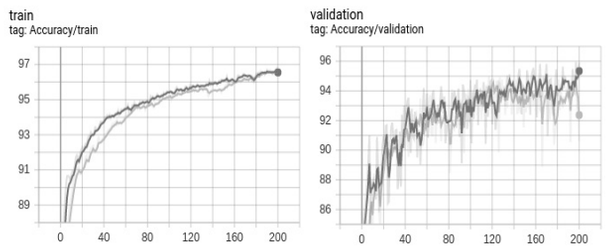

Fig. 7A. Accuracy of Training and Testing using RMSProp

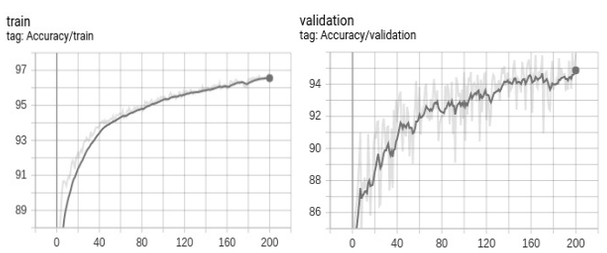

Fig. 7B. Accuracy of Training and Testing using Adam

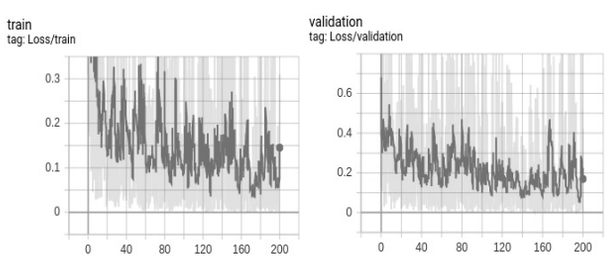

Fig. 8. Loss of Training and Testing using Adam

해주기 때문에 4.27% 정도 성능이 더 좋은 결과가 나왔다.

모델 학습 시 교차 엔트로피(Cross Entropy)는 두 확률 분포의 차이를 구하기 위해서 사용하였다. Cross Entropy는 실제 데이터의 확률 분포와 계산한 확률 분포의 차이를 구하는데 사용된다. Cross Entropy 손실함수를 사용했을 때 학습 결과 손실은 0.1532이고, 평가결과 손실은 0.1845이다. 모델 학습 시 Adam의 손실 결과는 Fig. 8과 같다. 손실 값 0.3과 정

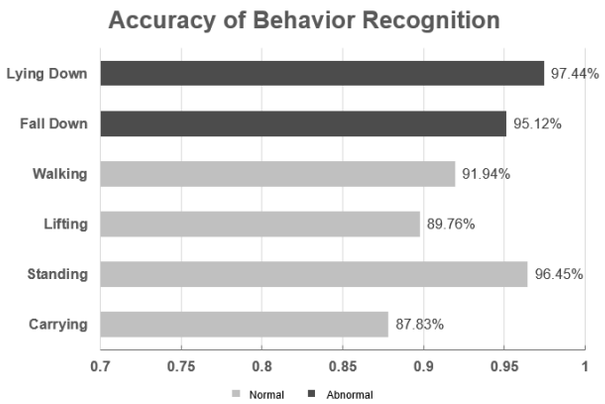

Fig. 10. Result of Fall Down Behavior

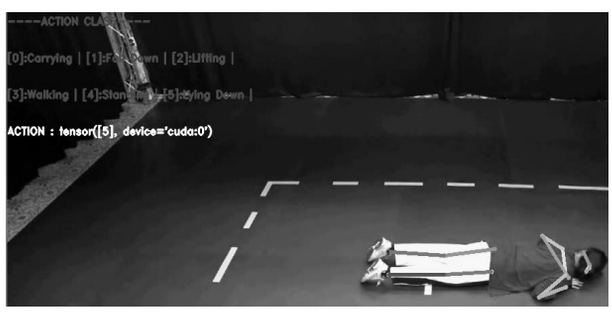

Fig. 11. Result of Fall Down Behavior

확도 95%를 달성하는 Epoch 80일 때, 성능이 좋아졌다.

행동인식은 Fig. 9와 같이 표현된다. 실험분석 결과로 재난안전관리 분야에서의 행동 중 누워있기와 쓰러지기와 같은 이상행동과 물건 옮기기, 걷기, 물건 들기, 서있기와 같은 일반적인 행동을 비교해 보았다. 행동의 인식률 결과는 Fig. 10과 같이 확인됐다.

그 결과, Fig. 11과 같이 누워있기 행동의 관절 변화량과 관절이 움직이는 각도가 다른 일반적인 행동들보다 크기 때문에 쓰러지는 행동의 인식률이 일반적인 행동의 인식률보다 10~12% 더 높게 측정되었다.

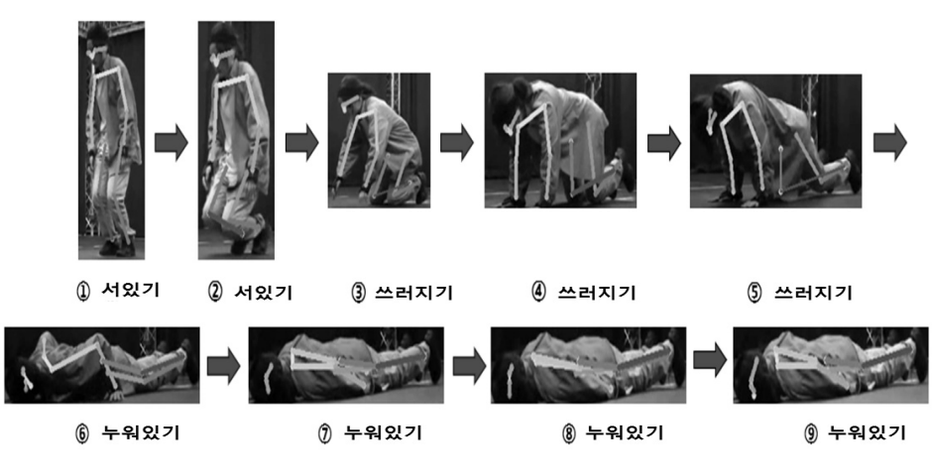

Fig. 9. Behavior Recognition over 9 Frames

5. 결론 및 향후 방향

본 논문에서는 행동인식 연구를 목적으로 자세추정 모델들의 관절 키포인트 추출에 대한 성능 평가 후, 성능이 좋은 AlphaPose와 LSTM을 활용하여 이상행동인식 모델을 구축하였다.

현장안전 분야에서의 행동인식 데이터셋 구축방법을 소개하였고, 다른 여러 산업 분야에서 CCTV를 활용해 행동인식 태스크를 진행할 때 도움을 주고자 한다. 영상 데이터로부터 프레임마다 관절의 키포인트를 추출하고 추출된 값을 적용시켰다. LSTM 내부 파라미터들에 대해 fine-tuning을 함으로써 대체로 성능 향상을 이루었다. 관절 키포인트를 추출하는 부분에서는 고해상도 유지 목적으로 HRNet을 사용하였고, 추출된 관절 데이터를 LSTM에 인풋 데이터로 적용한 결과 정확도는 95.63%, 손실은 0.1845로 성능이 가장 높게 나왔다.

향후 연구계획으로는 사람의 관절 값을 3차원으로 추출하여 CCTV를 통해 실시간으로 사람의 행동을 인식하는 연구를 지속하고자 한다. 다른 각도에서는 보이지 않는 관절을 추출할 수 있으며, 복잡한 관절에 대한 인식률 성능 향상을 기대할 수 있다.

References

[1] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu, "Rmpe: Regional multi-person pose estimation," *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[2] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang, "Deep high-resolution representation learning for human pose estimation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[3] Felix A. Gers, Jürgen Schmidhuber, and Fred Cummins, "Learning to forget: Continual prediction with LSTM," (1999): 850-855.

[4] Tsung-Yi Lin, et al., "Microsoft coco: Common objects in context," *European Conference on Computer Vision*, Springer, Cham, 2014.

[5] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," arXiv preprint arXiv:1812.08008, 2018.

[6] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick, "Mask r-cnn," *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[7] Jiefeng Li, Can Wang, Hao Zhu, Yihuan Mao, Hao-Shu Fang, and Cewu Lu, "Crowdpose: Efficient crowded scenes pose estimation and a new benchmark," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[8] Zaremba, Wojciech, Ilya Sutskever, and Oriol Vinyals, "Recurrent neural network regularization," arXiv preprint arXiv:1409.2329 (2014).

[9] Statistic, US Bureau of Labor, "Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work, 2011," UDo Labor, Editor (2012).

[10] Lieyun Ding, Weili Fang, Hanbin Luo, Peter E. D. Love, Botao Zhong, and Xi Ouyang, "A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory," *Automation in Construction*, Vol.86, pp.118-124, 2018.

[11] D. P. Kingma, and B. Jimmy, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[12] Human Pose Estimation Image AI Data [Internet], https://aihub.or.kr/aidata/138

[13] Toshev, Alexander, and Christian Szegedy, "Deeppose: Human pose estimation via deep neural networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[14] Xiao, Bin, Haiping Wu, and Yichen Wei, "Simple baselines for human pose estimation and tracking," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[15] Yan, Sijie, Yuanjun Xiong, and Dahua Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol.32. No.1. 2018.

배 현 재

https://orcid.org/0000-0002-2164-0125
e-mail : jason0425@snu.ac.kr
2019년 ~ 2020년 차세대융합기술연구원 인턴연구원
2020년 ~ 현 재 차세대융합기술연구원 컴퓨터비전 및 인공지능연구실 연구원
2021년 ~ 현 재 성균관대학교 소프트웨어학과 석사과정
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 인공지능 연구실 연구원
관심분야 : Computer Vision & Artificial Intelligence

김 영 훈

https://orcid.org/0000-0002-4151-7933
e-mail : toya84@snu.ac.kr
2008년 서울대학교 기계항공공학부(학사)
2014년 서울대학교 기계항공공학부
(석·박사통합과정)
2014년 ~ 2016년 한국과학기술연구원
바이오닉스연구단 박사후연구원
2016년 ~ 2018년 엔씨소프트 소프트웨어 엔지니어
2018년 ~ 2018년 삼성전자 글로벌기술센터 책임엔지니어
2018년 ~ 현 재 차세대융합기술연구원 선임연구원
관심분야 : Signal Processing & Machine Learning

김 진 평

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