

논문 2022-17-24

비디오 영상에서 2차원 자세 추정과 LSTM 기반의 행동 패턴 예측 알고리즘 (Behavior Pattern Prediction Algorithm Based on 2D Pose Estimation and LSTM from Videos)

최지호, 황규태, 이상준*
(Jiho Choi, Gyutae Hwang, Sang Jun Lee)

Abstract : This study proposes an image-based Pose Intention Network (PIN) algorithm for rehabilitation via patients' intentions. The purpose of the PIN algorithm is for enabling an active rehabilitation exercise, which is implemented by estimating the patient's motion and classifying the intention. Existing rehabilitation involves the inconvenience of attaching a sensor directly to the patient's skin. In addition, the rehabilitation device moves the patient, which is a passive rehabilitation method. Our algorithm consists of two steps. First, we estimate the user's joint position through the OpenPose algorithm, which is efficient in estimating 2D human pose in an image. Second, an intention classifier is constructed for classifying the motions into three categories, and a sequence of images including joint information is used as input. The intention network also learns correlations between joints and changes in joints over a short period of time, which can be easily used to determine the intention of the motion. To implement the proposed algorithm and conduct real-world experiments, we collected our own dataset, which is composed of videos of three classes. The network is trained using short segment clips of the video. Experimental results demonstrate that the proposed algorithm is effective for classifying intentions based on a short video clip.

Keywords : Rehabilitation exercise, Behavior pattern, Human pose estimation, Image sequence, Deep learning

1. 서론

최근 딥 러닝을 기반으로 한 컴퓨터 비전의 발전은 사람들의 생활을 편리하게 변화시키는 주요한 역할을 하고 있다. 대부분의 모바일 기기에는 고사양의 카메라가 설치되어 있고, 고화질의 영상 데이터를 쉽게 얻을 수 있다. 획득한 영상 데이터를 학습된 딥 러닝 알고리즘에 입력하여 목적에 따라 정보를 제공하는 시스템은 실생활에서 쉽게 찾아볼 수 있다. 홈 트레이닝 및 요가 동작 코칭과 같이 카메라를 활용하여 일상생활에서 사용되는 딥 러닝 알고리즘을 찾아볼 수 있다. 이는 실시간으로 사람의 동작을 학습된 알고리즘이 전문가의 수준으로 분석하여 소리나 시각적인 피드백을 통해 사용자가 올바른 자세를 취할 수 있도록 보조한다. 또한, 재활 운동이 필요한 부상자나 노인의 경우에 재활 자세나 움직임을 빠르게 하도록 허리나 팔, 다리의 각도에 피드백을 주어 치료에 효과적으로 적용할 수 있다. 이러한 비대면 헬스케어는 딥 러닝 알고리즘을 적용하여 보급에 용이하

고, 비용이 적게 들며, 공간의 제약이 줄어들어 이점이 있기 때문에 많은 사람에게 수요가 있다.

기존의 재활 치료는 접촉식 센서를 환자의 몸에 직접 부착하여 환자의 무게 중심, 기울기, 근전도 신호를 측정하여 데이터 분석 이후에 재활 치료사의 피드백을 받거나 센서 정보 기반으로 재활 보조 장치가 동작하는 방식이다 [1]. 접촉식 센서 기반의 재활 접근 방식은 환자의 움직임을 정교하게 파악할 수 있다는 장점이 있지만 재활 중인 환자의 신체에 직접 센서를 부착해야 하는 번거로움이 있다. 이러한 이유로 재활 공간의 제약이 발생할 수 있고 원시 센서 데이터는 추가적인 데이터 분석이 필요하다. 또한 전통적인 재활치료는 재활 기기가 동작한 후 환자가 기기에 따라서 움직이기 때문에 환자는 수동적인 재활치료를 받는다. 이러한 접근 방식은 환자의 효율적인 재활 치료가 어렵다는 한계점이 있어 능동적인 재활이 가능한 방법을 모색했다. 즉, 재활 환자의 동작 의도가 반영되는 치료를 위해 환자 중심의 재활 치료가 필요하고, 이는 환자의 의도를 파악한 후 기기가 보조하는 방식으로 구현할 수 있다.

접촉식 센서를 환자의 몸에 부착하고 획득한 센서 데이터를 입력으로 하는 딥 뉴럴 네트워크를 학습시켜 사용자의 재활에 도움을 주는 딥 러닝 기반의 재활 치료 방식이 연구되고 있다 [1-6]. Zhu et al. [1]은 센서 데이터를 입력으로 사용하는 MP-CNN 모델을 이용해 환자의 재활 운동 동작

*Corresponding Author (sj.lee@jbnu.ac.kr)

Received: Jul. 10, 2022, Revised: Aug. 2, 2022, Accepted: Aug. 9, 2022.

J. Choi: Jeonbuk National University. (M.S.)

G. Hwang: Jeonbuk National University. (M.S.)

S.J. Lee: Jeonbuk National University (Assist. Prof.)

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2022년도 문화기술 연구개발 사업으로 수행되었음 (과제명 : 심신안정·스트레스 완화를 위한 기능성 콘텐츠 플랫폼 개발, 과제번호 : R2020060003, 기여율: 100%).

을 4가지 class로 분류하는 재활 운동 인식 알고리즘을 제안하고 있다. 또한 사람 팔의 신진, 굴곡, 회전을 포함하는 팔의 3가지 움직임을 분류하는 네트워크인 Rehab-Net을 제안한 Panwar et al. [3]은 3 축 가속도 센서값을 이용한 딥 러닝 알고리즘을 개발했다. 위와 같은 연구들은 사람이 직접 환자의 센서값을 분석하지 않고 재활 환자의 움직임 의도를 파악하거나 분류할 수 있어 재활을 효과적으로 진행할 수 있는 장점을 보여준다. 하지만 여전히 환자의 몸에 센서를 부착하고 재활 치료를 진행해야 하는 번거로움과 치료 공간의 제약 등의 단점이 존재한다. Ajay et al. [7]에서는 비디오 영상에서 파킨슨 병 보행과 정상 보행을 분류하는 딥 러닝 기반의 알고리즘을 연구했다. 본 논문은 비디오 영상 기반의 비접촉식 재활 환자의 보행 의도 판단 딥 러닝 알고리즘 Pose Intention Network (PIN)를 소개한다.

Human pose estimation (HPE)은 영상에서 사람의 관절 좌표값을 추정하고 동작을 예측하는 딥 러닝 기반의 알고리즘이다. HPE는 2D [8, 9]와 3D 영역 [10-12]으로 나누어지며, 2D 영역에서는 이미지 내에 다수의 사람이 존재하는 경우에 2가지 접근 방식이 제안되었다. 여러 사람의 동작을 추정할 때, top-down 방식은 영상에서 먼저 사람을 감지하고 bounding box 내부에서 동작을 추정한다. 이와 다르게, bottom-up 방식은 영상 속 관절 정보를 모두 추정한 후 관절의 상관관계를 통해 개개인의 동작을 추정한다. Cao et al. [8]은 convolutional neural network (CNN)를 통해 feature map을 얻은 후 part affinity fields와 part confidence map을 얻기 위한 multi-stage CNN을 제안했다. OpenPose 알고리즘은 영상 내에 다수의 사람이 있는 경우에도 높은 성능으로 동작하며, 실시간 처리에 강한 알고리즘이다. 이 연구는 bottom-up 접근 방식인 OpenPose를 활용하여 관절의 추정값을 얻은 후, 관절 좌표의 시간에 따른 변화를 분석하는 딥 러닝 기반의 알고리즘을 통해 3가지의 동작을 분류하는 방법을 제안한다.

Recurrent Neural Network (RNN)는 시계열 데이터에서 이전 데이터와 현재 데이터의 상관관계를 분석하여 다음 데이터를 예측하고자 설계된 뉴럴 네트워크이다. 하지만 RNN은 장기 의존성 문제의 한계점이 있기 때문에 이를 해결할 수 있는 long short term memory (LSTM) [13]의 연구가 진행되었다. LSTM은 forget, input, update 그리고 output gate로 구성되어 있으며, 이러한 구조는 직전 데이터뿐만 아니라 더 오래된 데이터를 함께 고려하여 다음 데이터를 예측하는데 용이하다. 우리는 LSTM 연구에서 영감을 받아 재활 환자의 이전 동작 데이터를 통해 다음 동작을 예측할 수 있는 네트워크를 구성한다. 즉, 짧은 시간 동안의 OpenPose 출력을 순차적으로 이어 영상 데이터의 시계열 특성을 분석하기 위한 도구로 LSTM을 사용한다.

본 논문의 핵심은 사용자의 의도 분석이라는 실제 문제를 풀기 위해서 적은 양의 데이터로 의도 분석을 위한 딥 러닝 모델을 학습하는 방법에 있다. 실제 문제를 푸는 과정에서 데이터 획득과정에 어려움이 있기 때문에, 의도분석을 위한 딥 러닝 모델을 HPE와 intention classification이라는 두 단

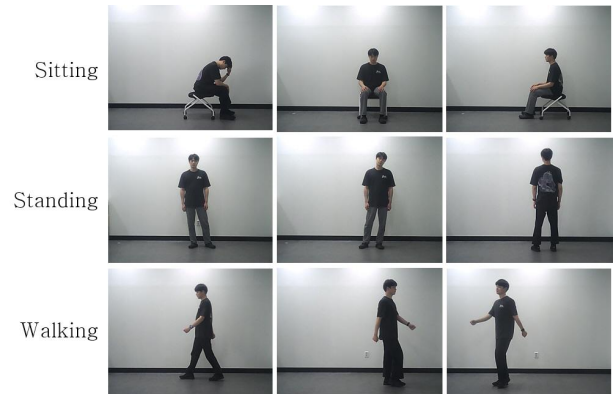


그림 1. 자체 제작한 데이터셋의 샘플 영상
Fig. 1. Example images of the collected image data

계로 나누었다. HPE 단계에서는 빅데이터에서 학습된 OpenPose의 pre-trained parameter를 사용하였으며, 개별 프레임 영상들로부터 획득된 joint 특징맵의 시계열 데이터를 효과적으로 활용하기 위하여, LSTM 기반의 다운스트림 네트워크를 구성하였다. 자체적으로 구성된 데이터를 활용하여 실험한 결과, 제안하는 딥 러닝 모델을 기반으로 한정적인 양의 실제 데이터만을 활용하여 사용자 의도분석이 가능함을 확인할 수 있었다. 일반적으로 HPE 과정에서 joint 좌표를 얻기 위해서는 argmax와 같은 미분 불가능한 연산을 필요로 한다. 하지만, 본 논문에서 제안하는 방법은 joint 좌표를 직접적으로 계산하지 않기 때문에 전체 과정이 미분 가능한 연산으로만 이루어져 있으며, 따라서 주어진 실제 데이터에 대하여 end-to-end로 fine-tuning이 가능한 장점이 있다.

II. 본 론

1. 데이터 세트 구성

본 논문은 동작 의도 파악 알고리즘을 실험하기 위해 데이터 세트를 제작하였다. 데이터 distribution이 균일하도록 class 별 영상을 같은 시간 동안 640x480 크기와 19 FPS로 촬영하였다. 촬영 장비는 Webcam Bos-c100이 사용되었으며 하얀색 배경의 실내에서 데이터를 획득했으며, 그림 1에 데이터 예시를 볼 수 있다. Class 0은 앉는 동작으로 3분간 등받이가 없는 의자에 앉아 카메라를 기준으로 1분씩 정면, 왼쪽과 오른쪽을 보고 여러 가지 상체의 각도와 다리를 다양한 자세로 취하며 데이터를 수집하였다. 서 있는 동작은 Class 1로 정의하고 3분간 카메라를 기준으로 좌, 우, 뒤와 정면 방향으로 차례 자세, 짝다리를 짚는 자세 등으로 데이터를 수집하였다. Class 2는 걷기 데이터로 카메라의 화각 끝과 끝까지 걷기, 제자리 걷기, 뒤돌기, 사선으로 걷는 방법으로 동일하게 3분 동안 수집하였다.

2. Pose Intention Network 구조

2.1. 전처리 과정

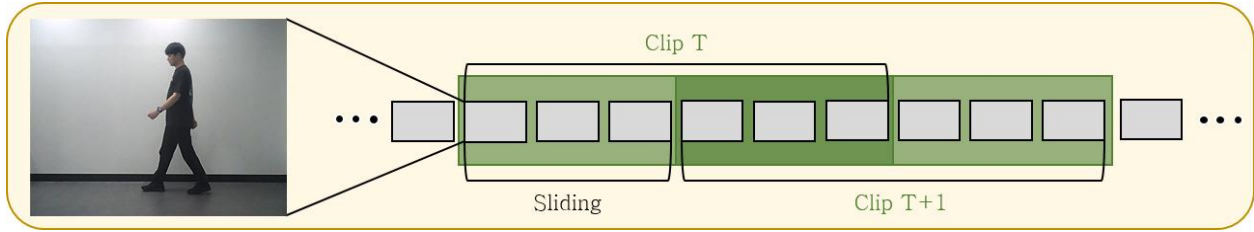


그림 2. 데이터 전처리 과정
Fig. 2. Data pre-processing

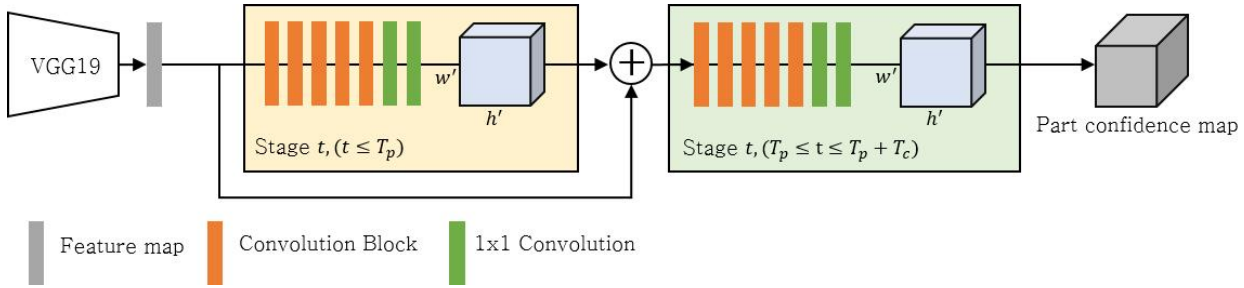


그림 3. OpenPose 구조
Fig. 3. Architecture of OpenPose

각 class의 영상은 500 ms 단위의 작은 영상 segment (Clip)로 분할되며 19 FPS 기준 9 장의 연속된 이미지이고, 순간적인 움직임을 학습하기 위한 입력 영상으로 사용하였다. 학습 시 데이터양의 증가와 데모 영상에 적용할 시 분류 결과를 빠르게 최신화하기 위해 clip에 약 200 ms 시간 차의 sliding time을 적용하였다. T 번째와 T+1 번째 clip이 clip - sliding 시간의 frame image 들을 공유하며 모델이 데이터의 시간적 특성을 학습하도록 하였다. 그림 2에서 이미지 sequence를 전처리하는 과정을 확인할 수 있다. Clip을 구성하는 각 frame image는 pre-trained OpenPose 모델에 입력되고 part confidence map으로 출력된다. Part confidence map은 60x80의 크기로 38의 channel을 갖는데, 19개의 관절 좌표의 confidence를 각 channel에담고 있다. OpenPose 네트워크로 관절 좌표를 얻으면서 촬영 배경의 변화나 사람의 원근감 등 여러 환경변수를 배제하여 Pose Intention Network가 환경변수에 강인하도록 데이터를 전처리한다.

2.2. OpenPose

OpenPose는 backbone network로 VGG19를 채택하고 있으며, 이는 다른 CNN 모델 구조로 변경이 가능하다. 영상은 각 프레임별로 VGG19 네트워크의 입력으로 들어가게 되고 특징 맵 (feature map)을 생성한다. 그림 3은 OpenPose의 구조를 나타낸다. 특징 맵 F는 관절 간의 관련성을 인코딩하는 단계를 거쳐서 part affinity fields를 반복적으로 예측하고 최종 PAF와 특징 맵 F가 더해져 confidence map을 추론하는 단계의 입력으로 사용된다. Part confidence map은 관절의 위치를 감지하고 픽셀마다 관절의 confidence를 인코딩하여 관절이 위치할 수 있는 가능성에 대한 정보를

담고 있다. Part affinity fields는 감지된 특정 관절에서 다른 관절로 향하는 방향성과 위치에 대한 관계를 추정하여 서로 연결되는 두 관절을 예측한다. 본 논문에서는 짧은 시간 동안 각 관절 위치 정보의 이동을 분석하여 동작을 예측하고자 했기 때문에 OpenPose network의 part confidence map만을 이용하여 intention network의 입력 데이터로 사용했다.

2.3. Intention Network

순간적인 몸의 움직임을 영상 데이터로 분석해 동작의 의도를 분류하고자 FPS/2 길이의 영상 Clip을 학습하고 동작을 분류하는 딥 러닝 네트워크 intention network를 설계했으며, 그림 4에서 확인할 수 있다. 각 part confidence map의 local feature를 찾기 위해 4개의 층으로 구성된 2D CNN에 입력한다. 각 층은 batch normalization, ReLU activation function과 maxpooling을 포함하고 있으며, Global average pooling을 통해 128 channel의 local feature vector를 얻는다. Local feature를 정확하게 찾는 과정은 시간적인 관절 위치의 변화와 관계가 없으므로 각 frame이 입력되는 CNN이 같은 가중치를 갖도록 parameter sharing 하였다. 이는 모든 part confidence map에 CNN을 적용하는 것에 비하여 딥 뉴럴 네트워크의 연산량을 줄이고 실시간 영상에 적용할 경우 지연 시간을 짧게 하는 이점이 있다. 이후에 Clip의 part confidence map 들의 local feature를 순차적으로 결합하여 time sequence vector를 생성한다. Clip의 시간적 특성을 통한 동작의 분류가 목표이므로, 관절의 변화량을 분석하기 위해 time sequence vector를 순환신경망 LSTM에 입력하고 마지막 hidden layer의 출력을 fully-connected layer와 softmax function을 이용해 분류할 class 별 확률을 얻는다.

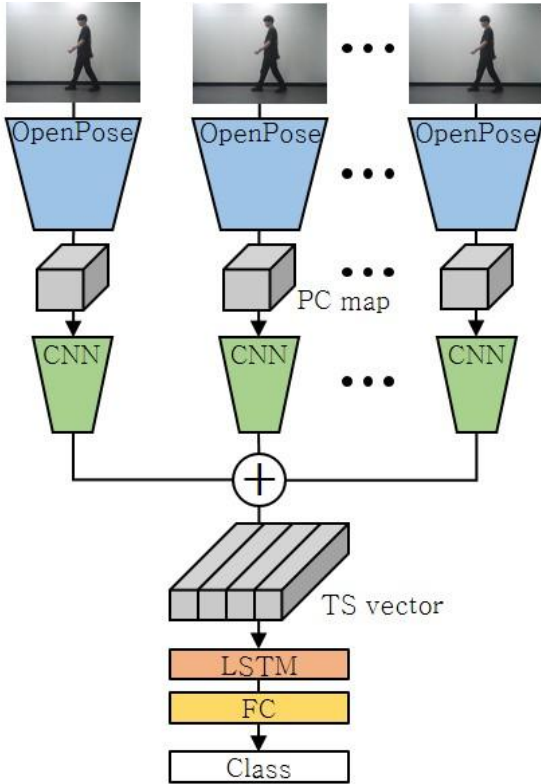


그림 4. PIN 구조, PC map과 TS vector는 각각 part confidence map과 time sequence vector를 의미

Fig. 4. Architecture of PIN, PC map and TS vector represent part confidence map and time sequence vector, respectively

즉, confidence는 softmax output으로 얻을 수 있으며 아래 수식으로 표현할 수 있다.

$$Confidence = Softmax(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)} \quad (1)$$

PIN은 OpenPose와 intention network를 end-to-end로 연결한 네트워크 구조이다. PIN은 cross entropy loss를 사용하며 다음과 같이 표현된다.

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \quad (2)$$

위 수식에서 N은 class의 개수를 의미하고 본 논문에서는 3으로 설정한다. 이에 따라 intention network는 비디오 영상에서 보행자의 행동 패턴을 앉기, 서기, 걷기의 클래스 별 확률을 추론한다.

III. 실험

1. 실험 환경

실험은 Intel Core i9-10900K CPU, 64GB DDR4 RAM,

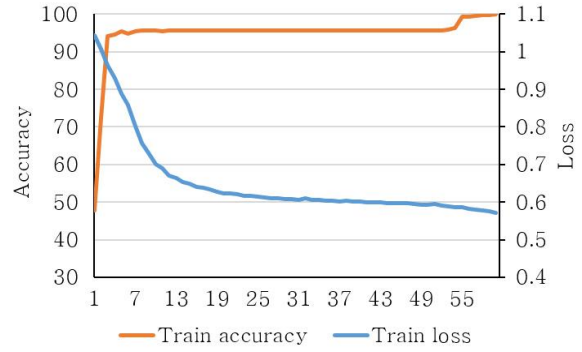


그림 5. 학습 정확도 및 손실
Fig. 5. Accuracy and loss of training

NVIDIA RTX 3090의 하드웨어 스펙에서 모델 훈련과 평가를 했으며, 운영 체제는 Ubuntu 18.04, 프로그래밍 언어는 Python 3.8, 프레임워크는 Pytorch를 사용하였다. Adam optimizer는 0.0001의 learning rate로 가중치를 업데이트하였다. 3가지 class 분류를 위해 각 class의 확률에 대하여 cross entropy loss를 계산하고 batch 크기는 3으로 설정하여 학습을 진행하였다.

2. 네트워크 정확도 및 어려움

19 FPS로 class 당 3 분씩 녹화된 영상은 500 ms의 clip과 약 200 ms의 sliding time으로 총 2,587 개의 clip을 생성하였다. Training과 test 데이터의 비율은 7:3이고, 각각 train clip은 1,810개, test clip은 777개로 구성된다. 걷기, 앉기 및 서기의 3 가지 출력 class에 대하여 그림 5에서 보이는 것과 같이 training 정확도는 100%를 달성하였으며, 학습이 진행됨에 따라 안정적으로 손실 값이 감소하였다. 표 1에서 test 오차행렬을 확인할 수 있다. Test 정확도는 단일 동작으로 구성된 데이터에 대해서는 100%의 정확도를 보였지만, 이후에 논의할 데모 영상 및 연속 동작에 대해서는 상대적으로 낮은 confidence를 보인다.

3. Ablation study

OpenPose의 관절 좌표 정보가 PIN 성능에 효과적임을 증명하기 위해 ablation study를 진행하였다. 모든 실험환경은 동일하게 설정하였고 intention network의 입력은 동영상

표 1. 오차행렬
Table 1. Confusion matrix

		Prediction class		
		Sitting	Standing	Walking
Actual class	Sitting	270	0	0
	Standing	0	250	0
	Walking	0	0	257

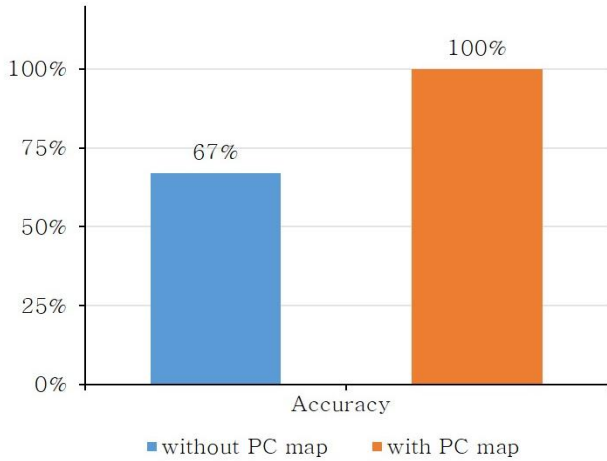


그림 6. PC map의 유무에 따른 실험결과
Fig. 6. Ablation study of with and without PC map

clip으로 하였다. 그림 6에서 볼 수 있듯이 part confidence map의 유무에 따라 test data에 대하여 33%의 성능 차이가 있다. Part confidence map이 원 영상보다 함축적인 관절 정보를 갖고 있어 PIN의 동작 의도 파악에 효과적임을 알 수 있다.

4. 데모 영상

데모 영상은 피 실험자가 앉기, 서기, 걷기 동작을 연속적으로 수행하며 1 분 동안 촬영되었다. PIN은 데모 영상에서 동작하는 사람의 행동을 즉시 판단해야 하므로 clip이 학습된 네트워크의 예측값이 출력되는 지연시간은 짧아야 한다.

이는 model complexity (복잡도)가 얼마나 크지에 따라 결정되고 더 복잡할수록 computing 장비에서 필요한 연산량 또한 높아진다. 짧은 지연시간과 높은 분류 정확도는 trade-off 관계에 있으므로 두 변수를 적절히 조절하는 것이 중요하며, 제안하는 모델은 의도에 따라 효율이 좋은 딥 뉴럴 네트워크를 설계하는 것이 중요하다. PIN으로 구성된 end-to-end 네트워크는 clip 영상의 길이인 500 ms가 짧은 시간임에도 불구하고 19 FPS의 데모 영상에서 sliding time에 따라 즉시 confidence map을 출력하였다. 이는 카메라의 영상에 따라 보행 재활을 위해 작동하는 기계에 confidence를 빠르게 제공하고 재활하는 사람의 움직임에 따라 즉각적인 보조를 할 수 있다는 점을 의미한다. 그림 7에서 PIN을 데모 영상에서 추론했을 때 confidence를 그래프 형식으로 영상과 같이 표시하였다.

5. 고찰

PIN을 학습시키기 위해 직접 제작한 데이터는 실험실 환경에서 제작되어 배경의 다양성이 부족하고 피실험자가 한정적이다. 또한 딥 뉴럴 네트워크를 일반화하기에는 데이터의 개수가 부족했다는 한계점이 존재한다. 네트워크의 정확도가 빠르게 수렴하는 현상은 위와 같은 원인으로부터 기인한다. Test 결과는 높게 측정되었지만, 여러 가지 class가 연속적으로 변화하는 데모 영상에서는 상대적으로 아쉬운 성능을 보였다. 특히, 서 있는 동작은 다른 동작에 비해 낮은 confidence 값을 가졌는데, 해당 동작의 데이터 다양성이 부족하여 야기된 문제로 보인다. 이러한 문제들은 다양한 피실험자와 촬영 환경으로 구성된 규모가 큰 데이터 세트로 학습하여 해결할 수 있을 것이다.

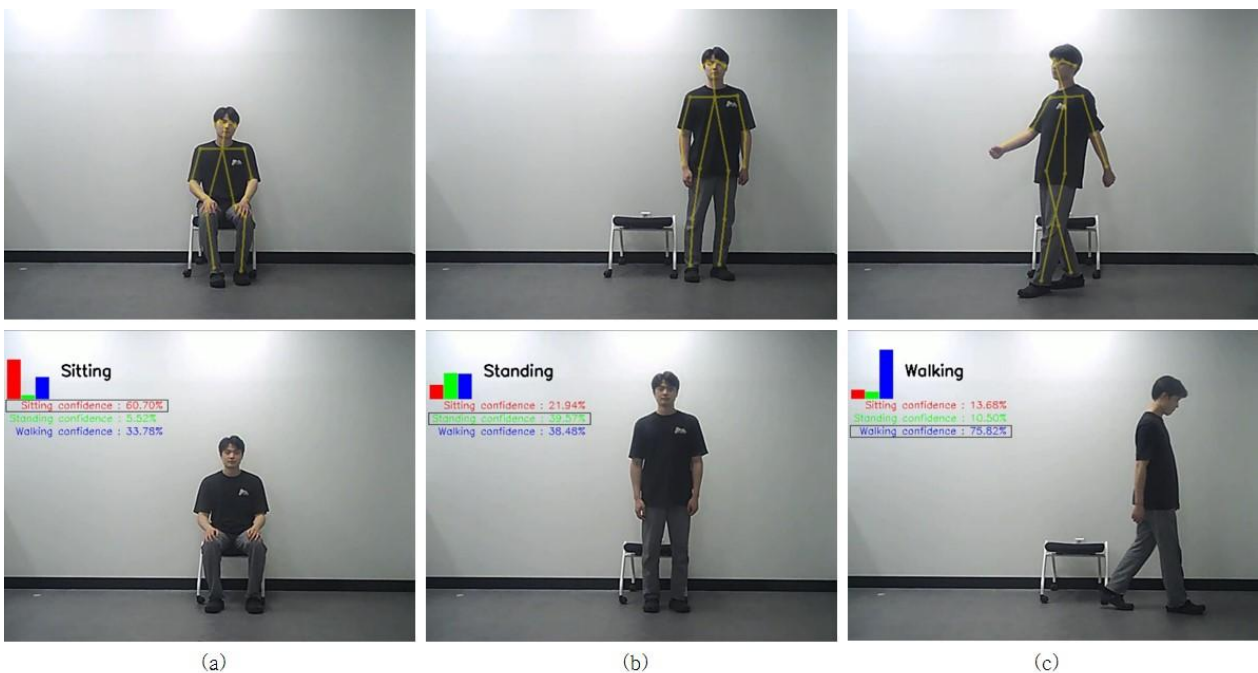


그림 7. 데모 OpenPose 및 PIN 결과 영상, 각각 (a) 앉기, (b) 서기 그리고 (c) 걷기 동작을 의미

Fig. 7. Real-time result of OpenPose and PIN

IV. 결론

이 연구는 영상 기반의 재활 환자의 의도 파악을 위한 PIN 알고리즘을 제안한다. OpenPose를 사용하여 환자의 관절을 추정하고 관절 정보 sequence를 intention network의 입력으로 사용한다. Intention network는 환자의 동작 의도를 추정하는 분류기로 짧은 시간 동안의 데이터를 이용하여 의도에 따른 관절의 변화를 학습한다. 실험을 통해 제안하는 알고리즘이 사용자의 동작을 높은 성능과 함께 분류하는 것을 확인했으며, 데모 영상에서도 알고리즘이 동작하는 것을 보였다. 본 논문에서 제안하는 PIN은 영상을 기반으로 재활 환자의 능동적인 재활의 가능성을 보이는 연구이다. 향후 연구에서는 큰 규모로 구성된 데이터 세트로 네트워크를 학습시켜서 다양한 환경에서도 동작할 수 있는 알고리즘 개발을 할 예정이다.

References

- [1] Z. A. Zhu, Y. C. Lu, C. H. You, C. K. Chiang, "Deep Learning for Sensor-based Rehabilitation Exercise Recognition and Evaluation," *Sensors*, Vol. 19, No. 4, pp. 887, 2019.
- [2] T. Zhong, D. Li, J. Wang, J. Xu, Z. An, Y. Zhu, "Fusion Learning for sEMG Recognition of Multiple Upper-limb Rehabilitation Movements," *Sensors*, Vol. 21, No. 16, pp. 5385, 2021.
- [3] M. Panwar, D. Biswas, H. Bajaj, M. Jöbges, R. Turk, K. Maharatna, A. Acharyya, "Rehab-net: Deep Learning Framework for Arm Movement Classification Using Wearable Sensors for Stroke Rehabilitation," *IEEE Transactions on Biomedical Engineering*, Vol. 66, No. 11, pp. 3026-3037, 2019.
- [4] L. M. S. D. Nascimento, L. V. Bonfati, M. L. B. Freitas, J. J. A. Mendes Junior, H. V. Siqueira, S. L. Stevan Jr, "Sensors and Systems for Physical Rehabilitation and Health Monitoring – A Review," *Sensors*, Vol. 20, No. 15, pp. 4063, 2020.
- [5] A. Kos, A. Umek, "Wearable Sensor Devices for Prevention and Rehabilitation in Healthcare: Swimming Exercise with Real-time Therapist Feedback," *IEEE Internet of Things Journal*, Vol. 6, No. 2, pp. 1331-1341, 2018.
- [6] S. Qiu, L. Liu, H. Zhao, Z. Wang, Y. Jiang, "MEMS Inertial Sensors Based Gait Analysis for Rehabilitation Assessment via Multi-sensor Fusion," *Micromachines*, Vol. 9, No. 9, pp. 442, 2018.
- [7] J. Ajay, C. Song, A. Wang, J. Langan, Z. Li, W. Xu, "A Pervasive and Sensor-free Deep Learning System for Parkinsonian Gait Analysis," In *Proceedings of 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 108-111, 2018.
- [8] Z. Cao, T. Simon, S. E. Wei, Y. Sheikh, "Realtime Multi-person 2d Pose Estimation Using Part Affinity Fields," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291-7299, 2017.
- [9] A. Toshev, C. Szegedy, "DeepPose: Human Pose Estimation

Via Deep Neural Networks," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1653-1660, 2014.

- [10] J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, L. Shao, "Deep 3D Human Pose Estimation: A Review," *Computer Vision and Image Understanding*, Vol. 210, pp. 103225, 2021.
- [11] Y. He, R. Yan, K. Fragkiadaki, S. I. Yu, "Epipolar Transformers," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7779-7788, 2020.
- [12] K. Isakov, E. Isakov, V. Lempitsky, Y. Malkov, "Learnable Triangulation of Human Pose," In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7718-7727, 2019.
- [13] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," *Neural computation*, Vol. 9, No. 8, pp. 1735-1780, 1997.

Jiho Choi (최지호)



2017 Electronic Engineering from Jeonbuk National University (B.S.)

2021 Electronic Engineering from Jeonbuk National University (M.S.)

Field of Interests: Generative learning, Image-to-image translation, Weakly supervised learning, Time-series signal analysis

Email: jihochoi@jbnu.ac.kr

Gyutae Hwang (황규태)



2015 Electronic Engineering from Jeonbuk National University (B.S.)

2022 Electronic Engineering from Jeonbuk National University (M.S.)

Field of Interests: Biosignal analysis, Biometric estimation, Affective computing

Email: gyutae741@gmail.com

Sang Jun Lee (이 상 준)



2007 Electrical Engineering from POSTECH
(B.S.)

2011 Electrical Engineering from POSTECH
(Ph.D. & M.S.)

Career:

2018~2020 Samsung Advanced Institute of Technology
(Senior Researcher)

2020~ Jeonbuk National University (Assist. Prof.)

Field of Interests: Deep learning, Computer vision, Image
processing, Robotics

Email: sj.lee@jbnu.ac.kr