

## Deep learning based Person Re-identification with RGB-D sensors

Min-Kim\*, Dong-Hyun Park\*

\*Student, Dept. of Industrial &amp; Management Engineering, Inha University, Incheon, Korea

\*Professor, Dept. of Industrial &amp; Management Engineering, Inha University, Incheon, Korea

## [Abstract]

In this paper, we propose a deep learning-based person re-identification method using a three-dimensional RGB-Depth Xtion2 camera considering joint coordinates and dynamic features(velocity, acceleration). The main idea of the proposed identification methodology is to easily extract gait data such as joint coordinates, dynamic features with an RGB-D camera and automatically identify gait patterns through a self-designed one-dimensional convolutional neural network classifier(1D-ConvNet). The accuracy was measured based on the F1 Score, and the influence was measured by comparing the accuracy with the classifier model (JC) that did not consider dynamic characteristics. As a result, our proposed classifier model in the case of considering the dynamic characteristics(JCSpeed) showed about 8% higher F1-Score than JC.

▶ **Key words:** person re-identification, surveillance system, deep learning, human action recognition, multi class identification

## [요 약]

본 연구에서는 3차원 RGB-D Xtion2 카메라를 이용하여 보행자의 골격좌표를 추출한 결과를 바탕으로 동적인 특성(속도, 가속도)을 함께 고려하여 딥러닝 모델을 통해 사람을 인식하는 방법을 제안한다. 본 논문의 핵심목표는 RGB-D 카메라로 손쉽게 좌표를 추출하고 새롭게 생성한 동적인 특성을 기반으로 자체 고안한 1차원 합성곱 신경망 분류기 모델(1D-ConvNet)을 통해 자동으로 보행 패턴을 파악하는 것이다. 1D-ConvNet의 인식 정확도와 동적인 특성이 정확도에 미치는 영향을 알아보기 위한 실험을 수행하였다. 정확도는 F1 Score를 기준으로 측정하였고, 동적인 특성을 고려한 분류기 모델(JCSpeed)과 고려하지 않은 분류기 모델(JC)의 정확도 비교를 통해 영향력을 측정하였다. 그 결과 동적인 특성을 고려한 경우의 분류기 모델이 그렇지 않은 경우보다 F1 Score가 약 8% 높게 나타났다.

▶ **주제어:** 사람 재인식, 감시 시스템, 딥러닝, 합성곱신경망, 사람 행동 인식, 다중 분류 인식

- 
- First Author: Min-Kim, Corresponding Author: Dong-Hyun Park
  - Min-Kim (min1106.kim@gmail.com), Dept. of Industrial & Management Engineering, Inha University
  - Dong-Hyun Park (dhpark@inha.ac.kr), Dept. of Industrial & Management Engineering, Inha University
  - Received: 2021. 01. 13, Revised: 2021. 02. 18, Accepted: 2021. 02. 19.

## I. Introduction

사람 재인식(Person Re-Identification; ReID)이란 여러 대의 카메라를 통해 서로 중첩이 없는 영역을 촬영한 영상 중에서 동일 인물을 탐지하는 과정을 의미한다. 최근 사람 재인식 기술은 감시분야에 적용되어, 감시가 필요한 기관 및 CCTV(Close-Circuit Television)가 설치된 장소에 도입하려는 시도가 점차 늘고 있으며, 24시간 사람의 이상행동을 탐지해야 하는 공항이나 공공시설에서도 이러한 기술을 이용하여 노동 집약적인 업무를 대체할 수 있다[1-3].

특히, 보안을 목적으로 한 자동화 감시시스템의 경우에는 탐지하는 대상을 지속적으로 추적하거나 재인식 할 수 있어야 하며, 가장 먼저 사람을 탐지하고, 찾아낸 사람의 신원이 데이터베이스에 등록되어있는지 찾아낼 수 있어야 한다[4-6].

최근, 사람 재인식을 위한 여러가지 방법들이 제시되고 있다. 사람의 외형적 특징 중 하나인 색상정보를 학습하여 특정 사람에게 고유성을 부여하는 방법과[7], Superpixel 기반으로 색상과 텍스처 정보를 가우시안 혼합모델(Gaussian Mixture Model; GMM)로 구분하는 방법[8], 사람의 신체부위를 범주화하여 학습하는 방법을 제시하였다[9]. 이러한 논문들의 공통점은 사람을 구별할 수 있는 특징으로 색상을 이용했다는 것이다. 실제로 사람을 인식하는 대부분의 경우에도 그 사람의 외형 즉, 옷의 색상, 옷의 종류 등을 기억하게 된다. 마찬가지로 지능형 영상감시 시스템으로 영상 속 사람을 인식하는 가장 효과적인 방법 역시 외형적인 옷 색상을 활용하는 것이다. 하지만, 색상 정보를 이용하는 방법은 환경변화에 따른 여러 가지 한계점이 존재하는데 예를 들어 조명 변화, 카메라 각도의 변화, 객체 간의 중첩에 의한 가려짐 등으로 정확도가 떨어지는 문제점이 있을 수 있다. 따라서 객체 간의 중첩, 가려짐, 외형 및 환경적인 변화가 있는 상황에서도 사람을 정확하게 재인식하기 위해서는 사람의 보행(Gait) 정보를 이용하는 것이 유리하다[10].

현재 딥러닝 기술이 발전함에 따라 객체의 영상정보를 딥러닝 모델로 학습시켜 신체부위 정보를 추출하고 보행 패턴을 인식하는 데 초점을 맞춘 연구들이 소개되고 있다. 하지만 이미지를 실시간으로 처리함과 동시에 학습된 딥러닝 모델을 구동하기 위해서는 고 사양의 컴퓨팅장비가 필요하기 때문에 일반적인 환경에서는 실시간으로 구동하는 데 한계가 있다[11].

본 논문은 이러한 한계점을 해결하기 위해 RGB-Depth 센서를 탑재한 ASUS Xtion2 카메라를 이용하여 실시간

으로 보행자의 골격(Skeleton) 정보를 수집하였고, 동적인 특징을 새롭게 생성하였다. 이러한 데이터를 기반으로 자체적으로 고안한 1차원 합성곱 신경망(One-Dimensional Convolutional Neural Network; 1D-ConvNet)모델을 학습시켜 기존의 방법보다 향상된 인식 성능을 확인하고자 하였다.

본 연구에서 제안하는 사람 재인식 프로세스는 그림 1과 같으며 다음과 같은 성과가 예상된다

1. RGB-Depth 센서를 탑재한 Xtion2 카메라를 사용하여 사람을 탐지하고, 재인식을 위한 자동화 알고리즘의 고안.
2. 식별 정확도를 높이기 위한 보행자의 동적인 특성(속도, 가속도)의 추가적인 생성을 통한 딥러닝 모델의 정확도 개선.
3. 실시간으로 측정되는 골격 정보를 효율적으로 처리하는데 최적화된 지도학습 기반 1D-ConvNet 모델의 고안.

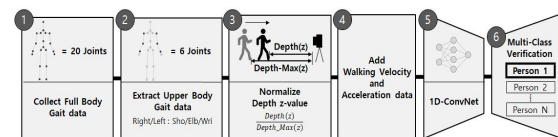


Fig. 1. An overview of person verification process.

## II. Related works

보행 패턴을 인식하는 방법은 대표적으로 모델기반(Model-based) 방법과 비-모델기반(Model-free-based) 방법으로 구분된다. 모델기반 방법은 상지 신체부위와 하지의 움직임 정도(각도)를 분석하여 특성(Feature)을 추출하여 보행패턴을 인식하는 방법이다[12-14]. 각 구성요소의 파라미터를 정확하게 추출하기 어렵다는 단점을 가지고 있지만 정확하게 추출된다면 높은 인식률을 보이는 장점이 있다. 반면, 비 모델기반 방법은 보행 영상 전체가 하나의 패턴으로 사용되는 방식으로 비교적 패턴을 인식하는 과정이 모델기반 방법보다 간단하며 데이터의 부분적 손실이 상대적으로 적다는 장점이 있다[15]. 또한, 두 방법을 결합한 형태로 보행 에너지 영상(Gait Energy Image: GEI)을 상체, 몸통, 하체 등의 3 등분으로 나누어 각 구성요소의 특성을 추정하는 방식을 통해 인식률을 높이는 시도가 있었다[16].

보행 패턴 인식은 주로 감시시스템, 범죄자 식별에도 응용되고 있다[17]. 특히 야간에는 가시광 카메라로 촬영된 영상을 토대로 보행 패턴을 인식해야 하며, 이는 조명변화

에 따라 측정값이 변동하는 문제가 있으므로 보행 패턴을 분석하는데 큰 제약 조건으로 작용한다. 이러한 이유로 최근 보행 패턴을 인식하는데 열화상 카메라를 이용하는 방식이 소개되고 있다. [18]은 열화상 카메라로 촬영된 영상을 이용한 HTI 모델을 제시하였다. 이 연구는 보행 영상 속 얼굴에서 허벅지 부분을 추출하여 인식 정확도를 유지하면서 인식 속도를 높이는 방법을 제시하였다. [19]에서는 웨이블릿 변환을 이용한 다 해상도의 특징들과 골격추출 알고리즘을 이용하여 높은 인식 성능을 보였지만 인식 과정이 복잡하고 구동하는 데 시간이 많이 소요되는 것으로 나타났다. [20]에서는 보행자의 머리, 몸통, 다리 부분을 분리한 다음 신체 부위를 좌우로 이동시키거나 회전시켜 SEI 모델을 생성하였고 GSP 방법을 통해 다양한 자세로 인한 구조적인 변화에도 정확하게 인식하는 방법을 제시하였다. 그러나 이 연구에서 사람의 정면이나 대각선 방향에서 촬영된 경우는 고려하지 않았고 오직 옆 모습 영상만을 활용했다는 점에서 한계가 있다. [13]에서는 GEI와 GENs를 이용하여 각 영상을 아주 작은 Window 형태로 분할한 후 Random Window Subspace Method(RWSM) 기법을 활용하여 옷차림의 변화에도 인식률을 유지하는 방법을 제시하였다. 하지만 영상을 각각의 작은 Window로 분할하여 연산이 진행되기 때문에 객체를 인식하는데 시간이 오래 걸리는 한계점이 존재한다.

최근, 딥러닝(Deep Learning)을 적용하여 기존 재인식 기법의 한계점을 해결하여 식별(Identification) 성능을 개선하고자 하는 연구가 수행되고 있다[21,22]. 식별 방법은 딥러닝 네트워크 훈련 시 가장 말단에 위치한 출력계층(Layer)의 이전 계층에서 출력된 특징을 활용하여 유사도(Similarity)를 비교한다. [23]는 Siamese Network 구조를 통해 두 이미지가 투입되면 동일한 가중치를 가진 합성곱 신경망(Convolutional Neural Network; CNN)으로 특징정보를 추출한 후 CIND(Cross-Input Neighborhood Differences)를 통해 특징정보를 서로 비교하며, 최종적으로 두 이미지가 동일한 사람인지를 완전연결 계층(Fully Connected Layer)에서 판단하게 된다. [24]는 이미지 속 사람의 신체부위(팔, 다리) 정보를 추출하는 단계를 추가함으로써 이미지마다 다른 사람의 자세와 이로 인해 발생하는 오차를 줄이는 방식을 제안하였다. 사람의 신체부위를 탐지하는 Pose Estimation 네트워크와 이를 통해 도출된 정보를 토대로 식별하는 네트워크로 구성되어 있다. 하지만, 딥러닝 기법을 통해 영상을 인식하고 특성 정보를 통해 사람을 식별하는 시스템의 경우 매 프레임마다 픽셀

단위의 연산이 가능한 고사양의 컴퓨팅 장비가 요구되고 딥러닝 모델의 학습을 위한 데이터셋이 마련되어야 한다.

본 논문에서는 RGB-D 카메라를 사용하여 조명변화와 옷차림, 가려짐과 같은 상황에서도 강건하게 사람을 식별하는 방법에 대해 제시하고자 하였다. 비교적 저렴한 3차원 RGB-Depth Xtion2 카메라를 사용하여 손쉽게 실시간으로 골격 정보를 추출하고 보행자의 동적인 특징인 움직임 속도와 가속도를 계산한 다음 자체적으로 고안한 1D-ConvNet 모델에 투입시켜 자동으로 사람을 식별하는 방법을 통해 기존의 한계점을 극복하고자 하였다.

### III. Preliminaries

#### 1. Data Preprocessing

본 연구에서는 3차원 RGB-Depth Xtion2 카메라를 사용하여 추출한 보행자의 인체골격 데이터(Skeletal data)를 활용한 딥러닝 모델을 구축하고자 한다. 1.1, 1.2, 1.3, 1.4에서는 딥러닝 모델의 단계별 구축 방법과 사용된 기법에 관하여 기술하고 있다.

##### 1.1 Structure of skeletal data

본 논문에서는 1대의 ASUS사의 3차원 RGB-Depth Xtion2 카메라를 사용하여 정면으로 걸어오는 사람의 보행 자세를 촬영한 뒤 Nuitrack SDK를 통해 인체골격 데이터를 추출하였다. Nuitrack SDK를 통해 추출할 수 있는 인체골격은 그림 2에 나타난 바와 같이 총 20개이다. 하지만, 본 논문에서는 딥러닝 모델 학습 파라미터 수를 줄여 경량화하기 위하여 신체 양측의 가장 대표적인 어깨(Shoulder), 팔꿈치(Elbow), 손목(Wrist)와 같은 상지(Upper-body) 정보만을 활용하여 딥러닝 모델을 학습하였다.

각각의 신체좌표는 x축, y축, z축으로 구성된 3차원 좌표로 구성되며 x축, y축 좌표의 경우 카메라 화면의 최대 너비 및 높이로 나눈 값으로 표현된다. 예를 들어 특정 신체좌표가 화면의 정 중앙에 위치한 경우 x축, y축 값은 모두 0.5의 값을 갖게 된다. 여기서 z축 값은 보행자와 카메라와의 거리를 나타낸다. 따라서 카메라의 인식 범위가 증가할수록 z축 값의 범위도 비례하여 증가한다.

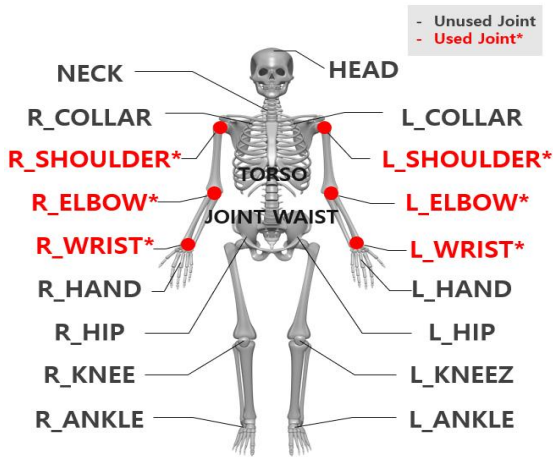


Fig. 2. Anatomical joint landmarks returned by Xtion 2 for NuiTrack SDK. Red points are used for model training (Left/Right Shoulder, Left/Right Elbow, Left/Right Wrist).

1.2 Data normalization

측정된 x축, y축 값은 화면의 최대크기로 값을 나누는 최대-최소 정규화(Min-Max Normalization)가 자동으로 적용된다. 하지만, z축 값의 경우는 해당하지 않기 때문에 이에 대한 별도의 정규화(Normalization) 작업을 수행해야 한다. z축 값의 경우 Xtion2 카메라에서 측정된 최댓값으로 나눈 방식을 통해 정규화하였다.

본 연구에서는 학습시킬 딥러닝 모델의 경량화 및 연산을 최소화하기 위하여 20개의 탐지 가능한 전체 신체좌표 중 6개의 상지 좌표만 선별하는 기능을 추가하였고 데이터를 선별하여 사용하였다.

1.3 Zero-padding metric

본 논문에서는 1.2절에 소개된 방법을 통해 최대-최소 정규화가 완료된 6개의 상지 신체좌표를 활용하여 속도 및 가속도 등의 신규 특성을 추출하여 학습에 활용하는 방식을 도입하였다. 사용한 속도 및 가속도 변환공식은 식(1), (2)와 같다.

$$V_t = \frac{D_t - D_{t-1}}{T_t - T_{t-1}} \tag{1}$$

$$A_t = \frac{V_t - V_{t-1}}{T_t - T_{t-1}} \tag{2}$$

보행자의 보행 영상을 프레임(Frame) 단위로 분리했을 때 특정 프레임(t)과 직전 프레임(t-1)을 비교하면 각 신체부위(Joint)의 움직임 속도를 구할 수 있으며 이를 통해 가속도 정보 역시 생성할 수 있다. 즉, 한 개의 신체부위에서 총 9개

의 특성(Feature)이 생성되고 총 6개의 신체부위를 통해 각 프레임 당 54개의 신규 특성 데이터를 생성할 수 있다.

다만 각각의 피실험자마다 보행자의 보행속도가 동일하지 않는다면 측정된 데이터의 수가 일치하지 않게 된다. 이러한 문제는 딥러닝 모델이 학습할 때 오류를 발생하게 되고 원활한 학습이 어렵게 된다. 따라서 본 연구에서는 제로패딩기법(Zero-padding)을 활용하여 문제를 해결하고자 하였다[27]. 제로 패딩기법이란 측정된 데이터의 잔여 부분에 0을 추가하는 방식으로, 이 방법으로 데이터의 크기를 균일하게 유지할 수 있는 장점이 있다.

본 논문에서는 전체 데이터 중 가장 길이가 큰 데이터 (T-max)를 기준으로 나머지 데이터의 길이를 제로패딩 기법(Zero-padding)으로 일치시켰으며, 이에 대한 예시를 그림 3에 정리하였다.

예를 들어 보행자 A(Subject A)는 총 4번의 보행이 있었고, 보행자 B(Subject B)는 3번 보행했다고 가정했을 때 보행자 B의 마지막 4번째 보행 데이터를 모두 0으로 추가하여 데이터의 길이를 맞추게 된다. 이러한 방식을 통해 보행자 A와 B의 걸음 수가 다르더라도 딥러닝 모델의 학습 데이터로 활용할 수 있게 된다.

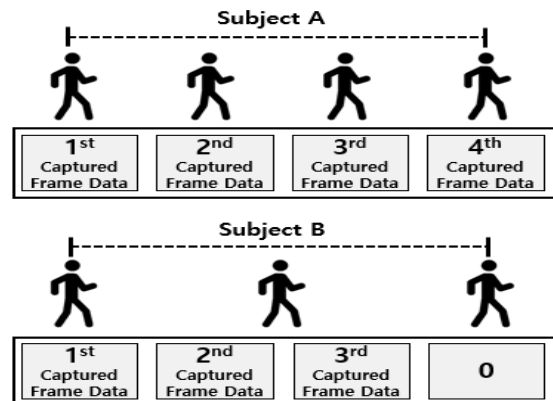


Fig. 3. Frame data normalization using zero-padding method.

1.4 Deep learning based 1D-ConvNet architecture

본 논문에서는 ASUS사의 3차원 RGB-Depth Xtion2 카메라를 통해 수집한 인체관절 데이터(Skeletal data)를 자체적으로 구현한 1차원 합성곱 신경망(One Dimensional Convolutional Neural Network: 1D-ConvNet)기반의 딥러닝 모델에 대입하여 보행자의 신원을 예측하는 모델을 설계하였다.

합성곱 신경망은 이미 많은 연구 분야에서 뛰어난 성과를 입증받았으며, 특히 컴퓨터 비전 분야에서 주로 객체탐지, 이미지 분류, 자세 분석용으로 사용되고 있다[25-26].

합성곱 신경망은 가장 기본적인 신경망 구조를 띠며, 훈련 가능한 가중치(Weight)와 편향(Bias)으로 구성되어 있다. 모든 뉴런(Neuron)은 투입된 정보(Input data)를 활용하여 벡터의 내적(Dot Product)과 같은 연산을 수행한다. 그다음 연산결과를 비선형 변환함수(ReLU)를 통해 다음 뉴런으로 전달한다. 모든 뉴런은 각각의 결과를 손실함수(Loss function)를 통해 점수화한다. 대표적인 손실함수의 예로는 소프트맥스(Softmax), 서포트벡터머신(Support Vector Machine; SVM)이 있으며, 각 클래스로 분류된 점수를 최종적으로 완전연결 계층(Fully connected layer)에서 평가하는 역할을 한다.

본 논문에서 제안하는 1차원 합성곱 신경망으로 구성된 모델은 총 2 개의 합성곱 층(Convolution Layer; Conv1D)과 추론을 담당하는 한 개의 Flatten 층, 그리고 두 개의 Dense 층으로 구성되어 있다(그림4). 마지막 층에서는 투입된 데이터가 어느 보행자의 것인지 확률로서 나타내게 된다.

비교적 간단한 1차원 합성곱 신경망 구조(1D-ConvNet)의 모델을 구성하여 추출된 보행 데이터를 학습시키고, 이를 통해 효율적이고 높은 정확도의 다중 분류기 모델(Multi-Class Classification Model)을 얻을 수 있다.

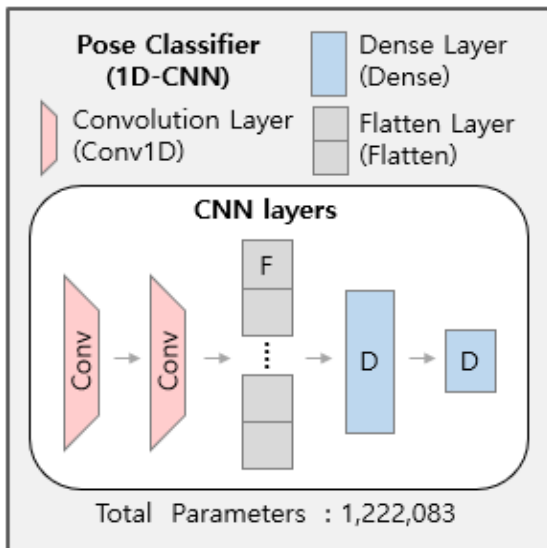


Fig. 4. An overview of 1D-ConvNet

## IV. Experiment

본 연구에서는 총 34명의 보행자(남: 23명, 여: 11명)를 대상으로 수집한 인체골격 데이터를 활용하여 모델을 학습시키고, 식별 정확도를 검증하는 실험을 수행하였다. 측

정된 인체골격 데이터를 전수 검사한 결과 오직 신체좌표 정보만으로는 특성(Feature)의 수가 부족하다고 판단하였고, 식별 성능을 개선하고자 각 관절의 속도(Velocity) 및 가속도(Acceleration) 특성을 새롭게 생성하였다. 따라서, 첫 번째 실험에서는 골격좌표(Joint Coordinate)뿐만 아니라 속도 및 가속도를 총괄적으로 학습한 모델(JCSpeed)의 결과를 살펴보고, 그 다음 골격좌표만 학습한 모델(JC)의 결과에 대하여 비교 분석하였다.

### 1. Equipment

실시간으로 인체골격 데이터를 측정하기 위해 Xtion2 카메라를 데스크탑 PC에 연결하였고, CPU는 Intel Core i5-6600 3.3GHz, RAM은 16GB, GPU는 RTX 2080ti NVIDIA, OS는 Windows 10, Programming Language는 Python, Environment는 Jupyter Notebook을 사용하였다.

### 2. Procedure

피실험자로 하여금 약 1.5m 거리에서부터 카메라를 정면으로 두고 평상시 보행속도를 유지하며 걷도록 지시하였고, 이 과정을 촬영하였다. 카메라는 지면으로부터 180cm 높이에 설치하였고, 그 외 부수적인 실험계는 선행 연구 [28]를 참고하여 구성하였다.

Xtion2 카메라는 300ms 메모리 버퍼(30FPS)로 측정이 가능한 것이 특징이다. 따라서 1초마다 측정된 30개의 프레임을 1개로 간주하여 그룹화한 다음 배치(Batch) 단위로 처리하는 과정을 통해 연산에 소요되는 시간을 획기적으로 단축시켰다. 마지막으로 배치화 된 각 데이터를 어느 보행자의 것인지 명시하는 라벨링 작업이 수행되었다.

본 연구에서는 앞서 그림 2에 나타난 바와 같이 Nitrack SDK를 통해 20개의 골격좌표를 추출하였다. 그 다음 6개의 상지 골격좌표만을 활용하여 속도, 가속도를 계산하고 딥러닝 모델에 투입하여 식별하도록 학습을 진행하였다. 이때 속도, 가속도가 식별 정확도에 미치는 영향을 알아보기 위해 인체골격 데이터만 가지고 딥러닝 모델을 학습시켜 동일한 방식으로 식별 정확도를 구한 다음 두 모델을 서로 비교해보았다.

두 방식 모두 자체적으로 고안한 1차원 합성곱 신경망 모델(1D-ConvNet)을 통해 34명의 보행자를 식별하도록 훈련하였다. 앞서 언급한 바와 같이 1차원 합성곱 신경망 모델은 데이터의 주요한 특징을 학습하는 2개의 1차원 합성곱 층(Conv1D)를 거치게 된다. 그다음, Flatten 층을 거쳐 추론(Regression)을 담당하는 2개의 Dense 층을 통해 학습이

진행된다. Dense 층은 비선형 함수인 ReLU와 Softmax를 사용하였고, 최적화 알고리즘(Optimization)은 Adam을, 손실함수(Loss function)은 Categorical-cross-entropy, 학습속도(Learning rate)는 0.000045로 설정하였다. 훈련 데이터셋(Training dataset)과 검증 데이터셋(Validation dataset)은 80:20 비율로 나누어 총 4000번의 학습을 수행하였다.

### 3. Data Analysis

학습된 모델의 성능은 전체 데이터 세트에서 학습에 사용하지 않은 20%의 테스트 데이터 세트를 통해 검증하였다. 학습된 모델에 테스트 데이터를 투입하여 예측한 결과를 범주별로 정리했고 결과는 표 1에 정리하였다. 예측한 정답 중에서 실제 정답의 비율인 정밀도(Precision)와 실제 정답 중에서 예측한 정답의 비율인 재현율(Recall)을 구한 후 조화평균을 사용한 F1 score를 구하여 모델의 성능을 평가하였다. 정밀도, 재현율, F1 Score를 구하는 식은 (3), (4), (5)와 같다.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (3)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (4)$$

$$F1score = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \quad (5)$$

## V. Results

앞에서 명시한 수식들로 인체골격 좌표, 속도, 가속도 데이터를 모두 학습한 JCSpeed 모델과 인체골격 좌표만 학습한 JC 모델의 범주별 Precision과 Recall 값을 각각 구하였고, 표 1에 정리하였다. 두 모델은 학습데이터의 특성만 다를 뿐 그 외의 모든 파라미터는 동일하게 적용하였다.

Precision과 Recall 값들의 전체적인 평균을 구했을 때 JCSpeed 모델의 훈련(검증) Precision은 0.98(0.93), Recall은 0.95(0.93)로 확인되었다. 이를 토대로 모델의 전체적인 성능을 나타내는 F1 Score를 계산했을 때 훈련(검증) F1 Score는 0.98(0.95)로 각각 나타났다. 반면, JC 모델의 경우 훈련(검증) Precision은 0.88(0.86), Recall은 0.89(0.87)의 값이 확인되었다. 두 값을 토대로 F1 Score를 계산했을 때 0.89(0.87)의 수치를 보였다.

그림 5의 (a), (b)의 학습 진행 과정을 살펴보면 JCSpeed, JC 모두 모델의 정확도가 우상향하는 그래프를 보였다. 이 중 JCSpeed의 경우 손실(Loss) 그래프(그림 5

의 c)에서 훈련 데이터셋과 검증 데이터셋의 차이가 두드러지게 나타났고, 이러한 차이는 훈련 시와 검증 시 식별 정확도에 차이가 있음을 알려준다. 반면, JC의 경우 그림 5, (d)와 같이 올바르게 수렴하는 손실(Loss) 그래프를 보이며, 검증 데이터셋에 올바르게 수렴하는 양상을 보였다.

두 모델의 정확도를 비교해보면 속도와 가속도 특성을 모두 학습한 JCSpeed이 약 0.09 높은 수치가 확인되었다. 이 결과를 통해 보행자의 두드러진 보행 패턴 특성이 사람의 신원을 식별하는 데 더욱 큰 영향을 미쳤다고 해석할 수 있다.

Table 1. Verification accuracies of classifiers based on different feature sets

Model	Precision (train/val)	Recall (train/val)	F1-Score (train/val)
JCSpeed	0.98(0.93)	0.95(0.93)	0.98(0.95)
JC	0.88(0.86)	0.89(0.87)	0.89(0.87)

## VI. Conclusion

본 연구에서는 딥러닝 기반으로 보행자의 보행 패턴을 인식하여 식별하는 방법을 제안하였다. 제안된 방법은 3차원 RGB-Depth Xtion2 카메라를 이용하여 인체골격 좌표 뿐만 아니라 속도와 가속도 특성을 함께 고려하여 자체적으로 고안한 1차원 합성곱 신경망 알고리즘(1D-ConvNet)을 학습시켜 효율적인 데이터처리 프로세스와 향상된 인식 성능을 확인하였다.

학습된 모델의 성능을 검증하기 위하여 총 34명의 피실험자를 섭외하여 30FPS 속도로 골격좌표 정보를 추출하였다. 그다음 측정된 데이터에 피실험자의 이름이 기록하고, 6개의 상지 움직임 속도와 가속도 특성정보를 포함하여 데이터셋을 구축하였다.

훈련용 데이터로 측정된 모델(JCSpeed)의 F1-score는 0.98을 보였고, 검증 데이터셋에 대한 정확도는 0.95를 보였다. 이처럼 상지 신체부위의 동적인 특성을 학습한 모델(JCSpeed)이 그렇지 않은 경우(JC)보다 약 0.08정도 높은 정확도를 보였다. 여기서 0.08의 정확도 개선효과는 총 34명의 피실험자를 대상으로 식별하는 경우에 약 2명에 대한 판단결과의 오류를 개선한 수준이라고 이해할 수 있으며, 개선 결과를 통해 주요 상체 부위의 움직임 특성이 사람을 식별하는 데 중요한 역할을 했다고 해석할 수 있다.

본 연구에서 제안하는 JCSpeed 모델을 출입 게이트(Gate)에 적용한다면 데이터베이스에 동적인 특징을 등록



한 사람은 통과 시 별도의 카드나 신분증을 표찰하지 않고도 출입이 가능할 것으로 예상된다. 또한, 감시가 필요한 지역에서 보행자의 수상한 행동을 탐지 및 분류하도록 응용한다면 소수의 감시인력으로 더욱 효율적인 보안시스템 운영이 가능할 것으로 예상된다.

향후 다양한 환경과 복잡한 환경에서 데이터베이스를 구축하여 좀 더 일반적인 실험 결과 입증에 관한 연구가 필요할 것으로 보인다. 동시에 여러 명이 움직이는 상황에서 겹쳐지는 현상이 발생할 때 인식률의 저하를 최소화하는 방안에 관한 연구가 필요하다고 판단된다. 또한, 본 연구는 실험실의 제한된 조건에서 실험을 통해 제안하고자 하는 방법의 우수성을 입증하였으나, 향후 다양한 조건의 환경과 실제 적용현장에 적용하여 실험 결과의 객관성과 일관성을 확보할 수 있는 연구를 수행할 계획이며, 실시간 신체골격정보 추출이 가능한 딥러닝 모델을 활용하여 센서 장비 없이도 정확하게 사람을 식별하는 방법에 관한 연구를 수행할 계획이다.

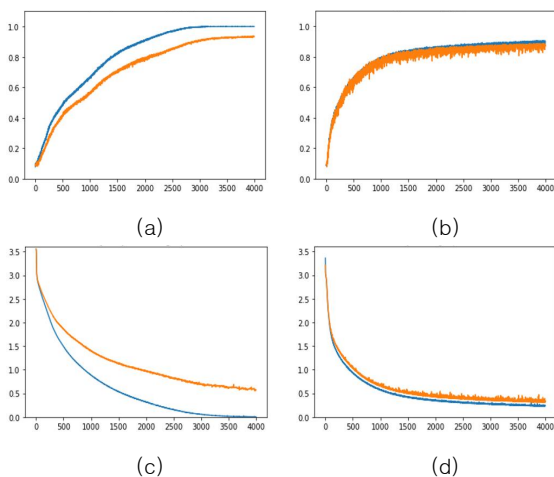


Fig. 5. Accuracy curves for JCSpeed model (a), JC model (b) and convergence curves for JCSpeed model (c), JC model (d).

## REFERENCES

- [1] R. Collins, A. Lipton, T. Kanade, "A System for Video Surveillance and Monitoring", Proc. Am. Nuclear Soc. (ANS) Eighth Int'l Topical Meeting Robotic and Remote Systems, April, 1999.
- [2] B. Huang et al., "Sparsity-based occlusion handling method for person re-identification" in Multimedia Modeling, Cham, Switzerland:Springer, 2015.
- [3] Y. Huang, S. Luo, E. Chen, "An Efficient Iris Recognition System", Proc. First Int'l Conf. Machine Learning and Cybernetics, pp. 450-454, Nov 2002.
- [4] S. Modi, S. Elliott, J. Whetsone, and H. Kim, "Impact of age groups on fingerprint recognition performance," in IEEE Workshop on Automatic Identification Advanced Technologies, pp. 19-23, 2007.
- [5] L. Deng, G. Hinton, B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview", Proc. IEEE Int. Conf. Acoust. Speech Signal Process., pp. 8599-8603, 2013.
- [6] G. Johansson, "Visual Motion Perception", Scientific Am., vol. 232, pp. 76-88, 1975.
- [7] Z. Wu, Y. Li and R. J. Radke, "Viewpoint Invariant Human Re-Identification in Camera Networks Using Pose Priors and Subject-Discriminative Features," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 5, pp. 1095-1108, 1 May 2015, doi: 10.1109/TPAMI.2014.2360373.
- [8] Shu, Guang. "Human detection, tracking and segmentation in surveillance video." (2014).
- [9] S. Zhiyuan, M. Timothy, "Transferring a Semantic Representation for Person Re-Identification and Search," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4184-4193
- [10] T. Dutta, "Evaluation of the Kinect sensor for 3-D kinematic measurement in the workplace", Appl. Ergonom., vol. 43, no. 4, pp. 645-649, Jul. 2012.
- [11] B. Bonnechère, B. Jansen, P. Salvia, H. Bouzahouene, L. Omelina, F. Moiseev, V. Sholukha, J. Cornelis, M. Rooze, S. Van Sint Jan, "Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry", Gait & Posture, vol. 39, no. 1, pp. 593-598, 2014.
- [12] C. BenAbdelkader and R. Cutler, "View invariant estimation of height and stride for gait recognition", Workshop on Biometric Authentication ECCV, pp.15-167, May 2002.
- [13] M. S. Islam, and M. R. Islam, "Window based clothing invariant gait recognition", International Conference on Advances in Electrical Engineering(ICAEE), pp. 411-414, 2013
- [14] J. J. Litle and J. E. Boyd, "Recognition people by their gait: the shape of motion", Videre, vol. 1, no. 2, 1998.
- [15] L. Le, and W. E. L. Grimson, "Gait appearance for recognition", Workshop on Biometric Authentication ECCV, pp. 143-154, 2002.
- [16] Q. -S. Li, Z. -T. Lu and D. -D. Zhang, "Integration of Gait and Side Face for Human Recognition in Video," 2009 Second International Symposium on Electronic Commerce and Security, Nanchang, 2009, pp. 65-69, doi: 10.1109/ISECS.2009.202.
- [17] A. Kale, A. K. R. Chowdhury, R. Chelapa, "Towards a view invariant gait recognition algorithm", Advanced video and signal based surveillance IEEE Conference on, pp. 143-150, July 2003.
- [18] D. Tan, K. Huang, S. Yu, and T. Tan, "Efficient night gait recognition based on template matching", The 18th International

- Conference on Pattern Recognition(ICPR), vol. 3, pp. 100-103, August 2006.
- [19] Z. Xue, D. Ming, W. Song, B Wan, and S. Jin, "Infrared gait recognition based on wavelet transform and suport vector machine", *Patern Recognition*, vol. 43, no. 8, pp. 2904-2910, August 2010.
- [20] Xiayi Huang, N. V. Boulgouris, "Gait recognition with shifted energy image and structural feature extraction", *Image Procening, IEEE Transaction on*, Vol. 21, pp. 256-268, April 2012.
- [21] S.D. Khan and H. Ullah, "A Survey of Advances in Vision-based Vehicle Re-identification," *Computer Vision and Image Understanding*, Vol. 182, pp. 50-63, 2019.
- [22] Q. Leng, M. Ye, and Q. Tian, "A Survey of Open-world Person Re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 30, No. 4, pp. 1092-1108, 2019
- [23] E. Ahmed, M. Jones, and T. K. Marks, "An Improved Deep Learning Architecture for Person Re-Identification," *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp.3908-3916, 2015.
- [24] L. Zheng, Y. Huang, H. Lu and Y. Yang, "Pose Invariant Embedding for Deep Person Re-Identification," *IEEE Transactions on Image Processing*, pp.4500-4509, 2019. DOI: 10.1109/TIP.2019.2910414
- [25] D. Wang, N. Canagarajah, D. Redmill and D. Bull, "Multiple description video coding based on zero padding," 2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512), Vancouver, BC, 2004, pp. II-205, doi: 10.1109/ISCAS.2004.1329244.
- [26] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth and H. Bischof, "Large scale metric learning from equivalence constraints," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 2288-2295, doi: 10.1109/CVPR.2012.6247939.
- [27] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan and M. Shah, "Visual Tracking: An Experimental Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442-1468, July 2014, doi: 10.1109/TPAMI.2013.230.
- [28] Manghisi, Vito Modesto, et al. "Real time RULA assessment using Kinect v2 sensor." *Applied ergonomics* 65 (2017): 481-491.

## Authors



Min-Kim received M.S degree in Electrical Engineering from the Graduate School of Engineering, Inha University, in 2018. Currently he is Ph.D candidates from the Graduate School of Industrial Engineering,

Inha University. He has been with the Department of Industrial Engineering at Inha University as a visiting assistant professor since 2021. His current research interests are human factors, 3D human pose estimation, computer vision, deep learning, action recognition, gait analysis.



Dong-Hyun Park received his Ph.D degree in Industrial and Manufacturing Engineering from Pennsylvania State University. He has been with the Department of Industrial Engineering at Inha University as a Professor

since 1994. His current research interests are industrial ergonomics, WMSDs and industrial safety.