

# RGB-D 정보를 이용한 객체 탐지 기반의 신체 키포인트 검출 방법☆

## A Method for Body Keypoint Localization based on Object Detection using the RGB-D information

박 서 희<sup>1</sup>                      진 준 철<sup>1\*</sup>  
Seohee Park                      Junchul Chun

### 요 약

최근 영상감시 분야에서는 영상에서 움직이는 사람을 탐지하고, 탐지된 사람의 행위를 분석하는 방식에 딥러닝 기반 학습방법이 적용되기 시작했다. 이러한 지능형 영상분석 기술을 적용할 수 있는 분야 중 하나인 인간 행위 인식은 객체를 탐지하고 탐지된 객체의 행위를 인식하기 위해 신체 키포인트를 검출 하는 과정을 거치게 된다. 본 논문에서는 RGB-D 정보를 이용한 객체 탐지 기반의 신체 키포인트 검출 방법을 제시한다. 먼저, 두 대의 카메라로 생성된 색상정보와 깊이정보를 이용하여 이동하는 객체를 배경으로부터 분할하여 탐지한다. RGB-D 정보를 이용하여 탐지된 객체의 영역을 재조정하여 생성된 입력 데이터를 한 사람의 자세 추정을 위한 Convolutional Pose Machines(CPM)에 적용한다. CPM을 이용하여 한 사람당 14개의 신체부위에 대한 신념 지도(Belief Map)를 생성하고, 신념 지도를 기반으로 신체 키포인트를 검출한다. 이와 같은 방법은 키포인트를 검출할 객체에 대한 정확한 영역을 제공하게 되며, 개별적인 신체 키포인트의 검출을 통하여 단일 신체 키포인트 검출에서 다중 신체 키포인트 검출로 확장 할 수 있다. 향후, 검출된 키포인트를 이용하여 인간 자세 추정을 위한 모델을 생성할 수 있으며 인간 행위 인식 분야에 기여 할 수 있다.

☞ 주제어 : 영상감시, 객체 탐지, 신체 키포인트 검출, 컨볼루션서널 포즈 머신, 신념 지도, 인간 행위 인식

### ABSTRACT

Recently, in the field of video surveillance, a Deep Learning based learning method has been applied to a method of detecting a moving person in a video and analyzing the behavior of a detected person. The human activity recognition, which is one of the fields this intelligent image analysis technology, detects the object and goes through the process of detecting the body keypoint to recognize the behavior of the detected object. In this paper, we propose a method for Body Keypoint Localization based on Object Detection using RGB-D information. First, the moving object is segmented and detected from the background using color information and depth information generated by the two cameras. The input image generated by rescaling the detected object region using RGB-D information is applied to Convolutional Pose Machines for one person's pose estimation. CPM are used to generate Belief Maps for 14 body parts per person and to detect body keypoints based on Belief Maps. This method provides an accurate region for objects to detect keypoints and can be extended from single Body Keypoint Localization to multiple Body Keypoint Localization through the integration of individual Body Keypoint Localization. In the future, it is possible to generate a model for human pose estimation using the detected keypoints and contribute to the field of human activity recognition.

☞ keyword : Video Surveillance, Object Detection, Body Keypoint Localization, Convolutional Pose Machines, Belief Map, Human Activity Recognition

## 1. 서 론

1 Department of Computer Science, Kyonggi University, Gyeonggi-do, 443-760, Korea.

\* Corresponding author (jchun@kyonggi.ac.kr)

[Received 25 September 2017, Reviewed 10 October 2017, Accepted 24 October 2017]

☆ 본 연구는 2017학년도 경기대학교 대학원 연구원장학생 장학금 지원에 의하여 수행되었음.

☆ 본 논문은 2017년도 한국인터넷정보학회 춘계학술발표대회 우수 논문 추천에 따라 확장 및 수정된 논문임

최근 영상감시 분야에서는 인공지능 기술과 컴퓨팅 능력 향상으로 인하여 다양한 지능형 영상분석이 가능하게 되었고, 이러한 지능형 영상감시 시스템에 딥러닝(Deep Learning) 기반 학습 방법이 적용되어 범죄, 화재, 이상 현상과 같은 다양한 이상행위들을 강건하게 탐지할 수 있게 되었다[1]. 또한 사람과 같은 객체의 전반적인 행위를 인식하기 위해서는 영상에서 객체의 움직임을 탐지 하는 과정과 탐지된 객체의 각 신체부위를 검출하는 과



(그림 1) RGB-D 정보를 이용한 객체 탐지 기반의 신체 키포인트 검출 방법  
 (Figure 1) A Method for Body Keypoint Localization using the RGB-D information

정이 선행됨으로써 인간의 자세를 추정하고, 행위를 인식 할 수 있다[2]. 따라서 본 논문에서는 기존 RGB 정보에 객체의 위상학적 관계를 고려하기 위한 깊이(Depth) 정보를 추가하여 객체를 탐지하고, 인간의 신체부위를 나타내는 키포인트(Keypoint)를 검출하기 위한 방법을 제시한다.

본 논문은 다음과 같이 구성된다. 먼저 2장에서는 본 논문에서 제시한 방법과 관련된 딥러닝 기반의 인간의 자세를 추정하는 다양한 연구를 소개하고, 제안된 접근법을 통하여 해결하고자 하는 문제를 설명한다. 3장에서는 본 논문에서 제안하는 RGB-D 정보를 이용한 객체 탐지 기반의 키포인트 검출 방법에 대한 알고리즘을 제시한다. 4장에서는 RGB-D 정보를 이용한 객체 탐지 방법을 소개하고, 5장에서 객체 탐지 기반의 신체 키포인트 검출(Body Keypoint Localization) 방법을 소개한다. 6장 실험결과에서는 객체 탐지 기반의 신체 키포인트 검출 결과를 소개한다. 마지막으로 7장에서는 결론을 맺는다.

## 2. 관련 연구

최근 영상감시 분야에서는 인공지능 기술의 발달로 인하여 딥러닝 기반 학습 방법을 적용하여 다양한 지능형 영상 분석이 가능하게 되었다. 영상감시 분야에서 사람과 같은 객체의 움직임을 탐지하고, 행동을 인식하는 인간 행위 인식(Human Activity Recognition) 분야는 딥러닝을 이용한 지능형 영상분석 기술을 적용하여 다양한 행위를 인식 할 수 있게 되었다[3]. 이러한 인간의 행위를 인식하기 위하여 선행되어야 하는 객체 탐지(Object Detection), 신체부위 검출(Body Parts Detection), 인간 자세 추정(Human Pose Estimation) 등의 연구가 활발히 이루어지고 있다[4-6]. 최근 Carnegie Mellon University의 The Robotics Institute에서 단일 이미지에서 여러 사람의 Body, Hand, Face를 나타내는 총 130개의 키포인트를 검출하는 최초의 실시간 시스템인 OpenPose 라이브러리가 발표되었다 [7]. 이러한 시스템은 Convolutional Pose Machine(CPM) 기

반으로 모델을 학습하여 키포인트를 검출하고 있으며, 기존의 탐지된 객체의 영역 안에서 키포인트를 검출하는 하향식(Top-down) 방식에서 키포인트를 먼저 검출하고 검출된 키포인트들 간의 연관성을 학습하여 인간의 자세를 추정하는 방식인 상향식(Bottom-up) 방식을 사용하고 있다[8-10].

키포인트를 검출하기 위해서 사용되는 CPM을 여러 사람이 있는 이미지에 적용 했을 때 한 사람에 대한 키포인트만을 검출하고, 나머지 사람들에 대해서는 검출을 수행하지 않는다. 이러한 단일 신체 키포인트만을 검출하는 CPM을 사용하여 다중 신체 키포인트 검출을 수행하려면 기존의 하향식 방식을 채택하여 목표 객체를 먼저 탐지하는 과정을 거치고, 탐지된 영역 내부에서 각 신체부위를 나타내는 키포인트의 위치를 검출 한 뒤, 개별적으로 검출된 키포인트를 위치들을 병합함으로써 다중 신체 키포인트 검출을 시도해 볼 수 있다. 또한, CPM의 키포인트를 검출하는 방식은 높은 정확도를 달성하지만, 주로 여러 사람이 가까이 있을 때 실패 사례를 관찰 할 수 있다. 그러므로 여러 사람이 가까이 있을 경우와 목표 객체에 대한 가려짐(Occlusion)이 발생하는 문제를 해결하기 위해서 본 논문에서는 깊이 정보를 추가하여 RGB-D 정보 기반의 객체 탐지를 수행하고, 탐지된 객체의 영역 내부에서 신체 키포인트를 검출함으로써 다중 신체 키포인트 검출을 위한 방법을 제안 할 수 있다.

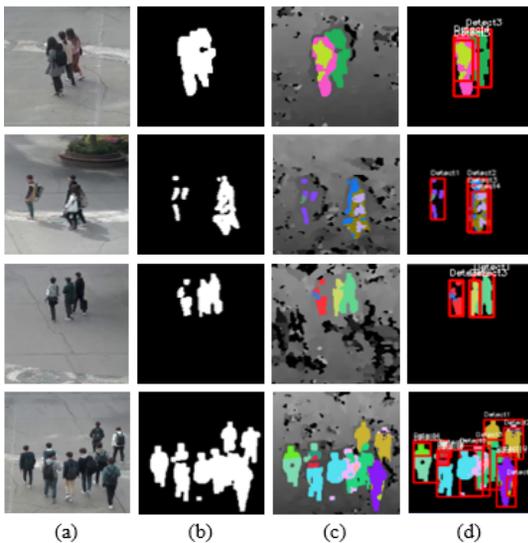
## 3. RGB-D 정보를 이용한 객체 탐지 기반의 신체 키포인트 탐지 방법

인간의 행위를 인식하는 분야에서 객체를 탐지하는 과정이 선행되면, 이를 기반으로 신체부위인 키포인트를 검출하는 과정을 거칠 수 있다. 본 논문에서는 RGB-D 정보를 이용한 객체 탐지 기반의 신체 키포인트 검출 방법을 제시한다. 본 논문에서는 깊이 정보를 얻기 위해 두 대의 카메라를 이용하였고, 두 영상 간의 차이를 계산하여 깊이 영상을 생성하였다. 원 영상에 혼합 가우시안(Mixture

of Gaussians) 기법을 사용하여 배경 모델링을 수행하고, 배경에서 움직이는 객체를 분할하게 된다. RGB 정보를 이용하여 분할된 객체의 영역은 깊이 정보 기반의 분할을 수행하기 위한 도메인으로 설정되고, 설정된 영역 내부에서 깊이 값을 이용한 분할을 수행하여 객체를 탐지한다. RGB-D 정보를 이용하여 객체 탐지된 결과를 기반으로 라벨링 된 영역을 재조정하고, CPM에 입력되기 위한 입력 데이터를 생성한다. 이를 CPM에 적용하여 한 사람의 각 신체부위들을 검출하기 위한 14개의 신념 지도를 생성하여 이를 기반으로 신체 키포인트 검출과정을 거치게 된다. 키포인트 탐색 절차는 그림1과 같다.

#### 4. RGB-D 정보를 이용한 객체 탐지

##### 4.1 객체 탐지



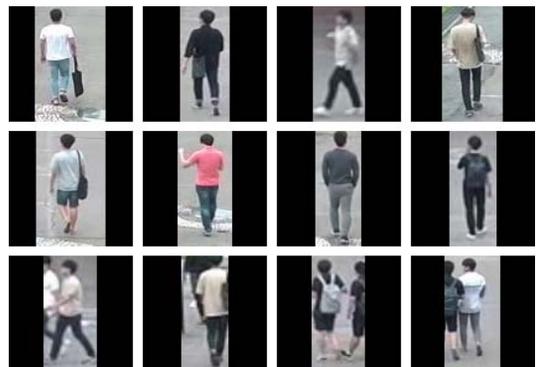
(그림 2) RGB-D 정보를 이용한 객체 탐지 결과  
(Figure 2) Object Detection Result using RGB-D information

사람의 행위를 인식하기 위해서는 영상에서 움직이는 객체를 배경영상으로부터 변화하는 영역을 추출함으로써 탐지 할 수 있다. 또한 3차원 실세계가 카메라를 통해 2차원 영상으로 변환되면서 생기는 위상학적 정보의 손실 한계를 개선하고, 목표 객체가 가려지는 문제를 해결하기 위하여 본 논문에서는 깊이 정보를 이용하여 객체 탐지를 수행하였다. 먼저, 원 영상에 혼합 가우시안 기법을 이용하여 배경모델을 계산하고, 지속적으로 배경모델

을 업데이트하여 배경 모델링을 수행한 후, 움직이는 객체에 차 연산을 수행하여 객체를 분할한다.

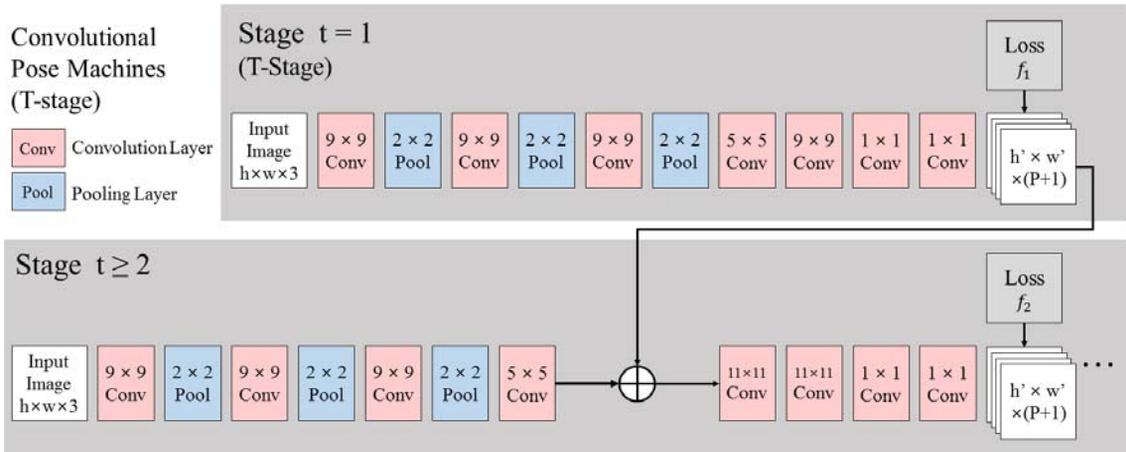
그 다음 RGB 기반으로 분할된 결과를 이용하여 깊이 정보를 이용한 객체 탐지를 수행하기 위해 깊이지도 (Depth Map)을 생성하여 깊이 값을 추출해야 한다. 좌, 우 카메라에서 입력되는 두 영상 간의 차이를 계산하기 위한 유사도를 측정하여 블록 단위로 매칭 한 뒤 깊이지도를 생성한다. 깊이 정보를 이용하여 분할을 수행하기 위해 RGB 기반으로 분할된 영역을 깊이 값을 추출할 도메인으로 설정하고, 도메인 내부에서 일정한 깊이 값 범위에 따라 객체 분할을 수행한다. 깊이 값은 0에서 255 사이의 명암 값으로 나타난다. RGB-D 정보를 이용하여 객체를 탐지한 결과는 그림 2에 나타나있으며, 그림 4a는 원 영상, 그림 4b는 RGB 정보를 이용한 객체 분할, 그림 4c는 RGB-D 정보를 이용한 객체 분할, 그림 4-d는 RGB-D 정보를 이용한 객체 탐지 라벨링 결과이다[11].

##### 4.2 탐지된 영역 재조정



(그림 3) 탐지된 영역 재조정 및 자르기  
(Figure 3) Rescale and crop detected regions

RGB-D 정보를 이용하여 탐지된 객체의 영역을 CPM에 입력 데이터로 제공하기 위해 각기 다른 크기로 탐지된 객체의 영역을 재조정 하는 과정을 거친다[12,13]. RGB-D 기반으로 탐지된 객체의 중심을 기준으로 생성된 라벨링 된 영역을 이용하여 영상을 자른다. 그 다음 영상의 비율을 모두 동일하게 1:1의 비율로 재조정하고, 영상의 여백 부분인 패딩을 추가하여 100x100 픽셀의 해상도로 입력 데이터를 생성한다. 이러한 입력 데이터는 단일 입력 프레임에 대한 포즈 추정기인 CPM에 적용되며, 탐지된 영역을 재조정 한 결과는 그림 3과 같다.



(그림 4) Convolutional Pose Machines의 아키텍처  
(Figure 4) Architecture of Convolutional Pose Machines

## 5. 신체 키포인트 탐지

### 5.1 Convolutional Pose Machines

CPM은 합성곱 신경망 (Convolutional Neural Network) 기반의 순차 예측 프레임워크(Sequential Prediction Framework)이며, 기존 포즈 머신(Pose Machines)의 포즈 추정 과정을 합성곱 신경망 구조를 이용하여 구현한 것이다[14]. CPM은 신체부위들에 대한 주석이 달린 25,000 개의 이미지로 이루어져있고, 400개가 넘는 인간의 활동들이 담겨져 있는 MPII Human Pose Dataset 등을 이용하여 사전 훈련(Pre-training)되었다[15].

CPM의 네트워크 구조는 입력 영상에서 의미 있는 특징(Feature)을 추출하는 합성곱 층(Convolution Layer)과 특징을 줄이는 서브 샘플링(Subsampling)을 하는 풀링 층(Pooling Layer)으로 구성되어 있다. 출력 값을 위한 활성화 함수(Activation Function)는 입력 값이 0보다 작으면 0, 0보다 크면 입력 값 그대로를 출력하는 ReLU(Rectified Linear Unit) 함수를 사용한다. 그림 4의 CPM 아키텍처에서 영상은 368x368 픽셀의 해상도로 정규화 되어 입력되고, 필터크기와 스트라이드(Stride)가 2x2로 3개의 맥스 풀링 층(Max-Pooling)을 거치게 되면 368x368 픽셀의 해상도에서 184x184, 92x92, 46x46의 해상도로 계층을 거치면서 x8 만큼 다운 샘플링 되며, 신체부위의 특징을 추출하게 된다.

이러한 CPM의 아키텍처는 신념 지도를 생성하는 여러 단계들  $t \in \{1, \dots, T\}$ 로 구성되어있으며, 이전 단계의

예측 결과를 다음 단계에 사용하여 순차적으로 신체부위  $p$ 를 예측하게 된다. 이처럼 단계  $t$ 에서는 모든 신체부위  $P$ 의 이미지상의 위치  $Y = (Y_1, \dots, Y_P)$ 를 예측하기 위해서 신체부위  $p$ 에 대한 신념 지도를 반복적으로 생성하도록 훈련된다. 한 이미지에 있어서의 모든 위치  $z = (u, v)^T$  ( $Z$ 는 이미지의 모든 위치  $(u, v)$ 의 집합,  $\forall z \in Z$ )에서 신체부위  $p$ 에 대한 이상적인 신념 지도  $b_p^*(Y_p = z)$ 는 각 신체부위  $p$ 의 이상적인(Ground Truth) 위치에 가우시안 피크(Gaussian peak)를 놓음으로써 생성된다.

또한 각 신체부위에 대한 예측 결과와 이상적인 신념 지도 사이의 거리를 최소화 하는 각 단계  $t$ 의 출력에서 손실 함수(Loss Function)를 다음과 같이 정의함으로써 네트워크가 반복적으로 정확한 신체부위 위치에 접근하도록 한다. 그러므로 각 단계의 결과에서 최소화 하고자 하는 손실 함수는 다음 식1과 같으며, 식에서  $P+1$ 은 신체부위  $p$ 가 포함된 배경을 위한 것이다.

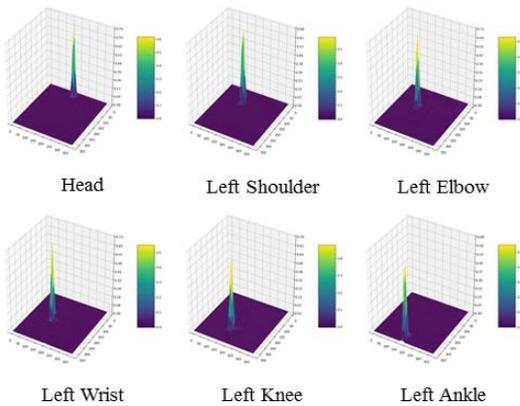
$$f_t = \sum_{p=1}^{P+1} \sum_{z \in Z} \|b_p^t(z) - b_p^*(z)\|_2^2 \quad (1)$$

이러한 손실 함수를 이용하여 각 단계에서 손실을 더함으로써 정확한 신체부위의 위치가 얻어지며 다음 식2와 같다.

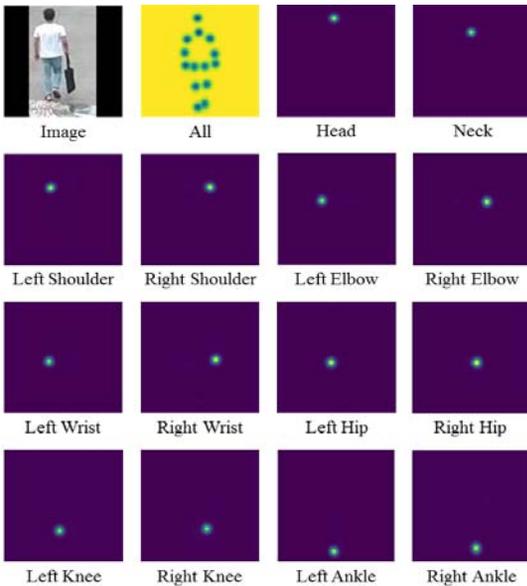
$$F = \sum_{t=1}^T f_t \quad (2)$$

### 5.2 신념 지도 생성

각 신체부위  $p$ 에 대한 신념 지도는 각 단계  $t$ 의 출력에서 손실 함수를 통하여 순차적으로 신체부위  $p$ 를 예측하는 과정을 통해 생성된다. 이 때 각 신념 지도에 표시된 지점은 CPM에 의해 생성된 신뢰도(Confidence)의 최대 높이 값을 뜻하며, 각 신뢰도의 최대값을 나타내는 신념 지도는 각 신체부위에 대한 예측 결과를 나타낸다[16].



(그림 5) 신뢰도 값의 분포도  
(Figure 5) Distribution of confidence values



(그림 6) 신념 지도  
(Figure 6) Belief Map

신념 지도를 나타내기 위한 신뢰도 값의 분포도에서 x축과 y축은 CPM의 입력 영상 크기인 368x368의 해상도를 뜻하고, z축은 신뢰도 값을 뜻한다. 이러한 신념 지도를 나타내는 신뢰도 값의 분포도는 그림 5에 나타나 있으며 신뢰도가 낮을수록 어두운 보라색으로 나타내고, 신뢰도가 높을수록 밝은 노란색으로 나타난다. 본 논문에서는 RGB-D 정보를 이용한 객체 탐지 된 영역을 재조정 한 이미지에 CPM에 의해 사전 훈련된 모델을 사용하여 신뢰도 값을 예측하였고, 그림 5에서와 같이 각 신체의 관절 부위에 대해 예측된 신뢰도의 최대값을 이용하여 한 사람당 14개의 관절을 나타내는 키포인트에 대한 신념 지도를 생성하였다. 신념 지도를 생성함으로써 신체 부위에 대한 키포인트의 위치를 예측한 결과는 그림 6과 같다.

### 5.3 키포인트 탐지

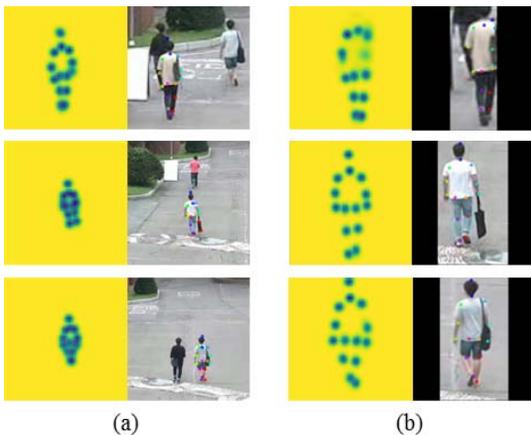
생성된 신념 지도는 인간의 각 신체부위를 나타내며, 이를 기반으로 머리(Head), 목(Neck), 어깨(Shoulder), 팔꿈치(Elbow), 손목(Wrist), 엉덩이(Hip), 무릎(Knee), 발목(Ankle)의 위치를 나타 낼 수 있다. 신념 지도를 이용하여 각 14개의 키포인트의 위치를 검출하기 위해서 신념 지도에 나타난 포인트의 중심점을 계산하여 신체 키포인트의 위치를 출력 할 수 있다[17]. 한 사람당 신체의 키포인트는 14개가 검출되며, 키포인트는 각 Head, Neck, Left Shoulder, Right Shoulder, Left Elbow, Right Elbow, Left Wrist, Right Wrist, Left Hip, Right Hip, Left Knee, Right Knee, Left Ankle, Right Ankle으로 구분된다. 신체 키포인트 검출 결과는 다음 그림 7과 같다.



(그림 7) 신체 키포인트 검출 결과  
(Figure 7) Body Keypoint Localization

## 6. 실험결과

본 논문에서는 실제 CCTV 환경에서 해상도 800x450 픽셀 크기의 영상을 자체적으로 획득하여 실험하였다. 실험 환경은 AMD Ryzen 7 1700 3GHz CPU, 16GB RAM, NVIDIA GeForce GTX 1080 Ti GPU, Windows 10 에서 딥러닝 오픈소스 라이브러리인 텐서플로우(Tensorflow)를 이용하여 수행하였다[18].

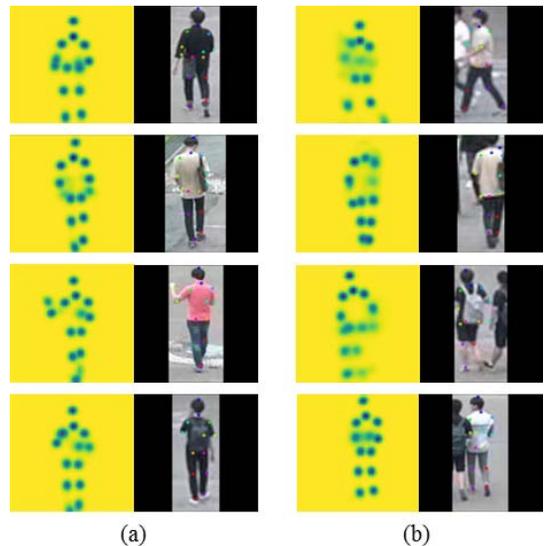


(그림 8) 객체 탐지과정 유무에 따른 신체 키포인트 검출 결과 (Figure 8) Body Keypoint Localization Result with and without Object Detection Process

그림 8에서는 객체 탐지과정 유무에 따른 신체 키포인트 검출 결과를 나타낸다. 그림 8-a에서 객체 탐지 과정 없이 키포인트 검출을 수행한 결과, 한 사람에 대해서만 키포인트 검출을 수행하며, 검출하고자 하는 목표 객체 (Target Object)에 대해서 검출을 수행하지 않을 수 있다. 본 연구에서는 특정한 한 사람, 즉 목표 객체에 대하여 키포인트 검출을 수행하기 위하여 CPM에 객체 탐지 과정을 추가하였다. 그 결과, 목표 객체에 대한 정확한 영역을 제공함으로써 한 사람에 대한 키포인트 검출을 수행하게 되었고, 키포인트 검출 결과는 그림 8-b에 나타나있다. 이는 목표 객체에 대한 개별적인 키포인트 검출이 가능함에 따라 영상 내 존재하는 모든 움직이는 사람에 대하여 개별적인 키포인트 검출을 각각 수행하고, 개별적으로 검출된 키포인트의 위치를 전체 영상에 매핑 함으로써 여러 사람에 대한 키포인트 검출을 시도 할 수 있다.

그림 9에서는 가려짐 유무에 따른 신체 키포인트 검출 결과를 나타낸다. 그림 9-a에서는 다른 사람에 의해 목표

객체가 가려지지 않았으므로 정상적으로 키포인트 검출이 수행되었다. 그러나 목표 객체와 다른 객체가 가까이 있거나, 다른 객체에 의해서 목표 객체가 가려진 경우는 객체가 하나로 탐지되는 문제가 발생 할 수 있다. 이러한 경우는 부정확한 영역에서 키포인트 검출이 수행됨으로써 키포인트 검출의 정확성을 보장 할 수 없게 된다. 그러므로 정확한 영역을 제공함으로써 제공된 영역 내부에서 키포인트 검출을 수행해야한다. 이는 본 논문에서 제안한 RGB-D 기반의 객체 탐지를 수행함으로써 붙어 있는 객체를 하나로 탐지하지 않고, 목표 객체의 정확한 영역을 탐지함으로써 CPM의 입력 데이터로 제공하였다. 그 결과 그림 9-b와 같이 개별적인 키포인트 검출을 수행할 수 있었다.



(그림 9) 가려짐 유무에 따른 신체 키포인트 검출 결과 (Figure 9) Body Keypoint Localization Result with and without Occlusion

## 7. 결 론

본 논문에서는 RGB-D 정보를 이용한 객체 탐지 기반의 신체 키포인트 검출 방법을 제시하였다. 그 결과, 영상 내에서 목표 객체가 아닌 불특정 한 사람만을 탐지하던 CPM에 객체 탐지 과정을 추가함으로써 목표하는 객체에 대한 키포인트 검출을 할 수 있었다. 또한, 기존 RGB 정보에 깊이정보를 추가하여 객체 탐지를 수행함으

로써 가려짐이 발생하는 목표 객체에 대해 명확하게 탐지하여 키포인트 검출을 위한 정확한 영역을 제공 할 수 있었다. 향후 이러한 RGB-D 기반의 객체 탐지 과정을 추가하여 개별적인 목표 객체에 대해 신체 키포인트 검출함으로써 여러 사람의 키포인트 검출이 가능해진다. 다중 신체 키포인트 검출(Multi-Keypoint Localization)은 여러 사람이 포함되어있는 CCTV 영상 내에서 움직이는 모든 사람들에 대하여 탐지를 수행한다. 각 탐지된 객체의 영역을 기반으로 이미지를 재조정 할 때 원 영상에서 탐지된 영역의 위치 정보  $Q$ 를 저장한다. 그 다음, 재조정된 영역들의 신체 키포인트를 순차적으로 모두 검출하고, 각 검출된 키포인트의 위치  $P$ 를 원 영상에 매핑 함으로써 다중 신체 키포인트 검출을 수행 할 수 있다. 또한, 신체 키포인트 탐지 결과를 이용하여 2D 기반의 스켈레톤 모델 (Skeleton Model) 또는 3D 기반의 스켈레톤 모델을 구축 함으로써 인간 행위 인식을 위한 인간 자세 추정(Human Pose Estimation)으로 연구의 범위를 확장 할 수 있다 [19,20].

## 참고문헌(Reference)

- [1] Grant, Jason M., and Patrick J. Flynn., "Crowd Scene Understanding from Video: A Survey," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Vol 13, No. 2, pp. 19, 2017. <https://doi.org/10.1145/3052930>
- [2] Zhang, Shugang, et al., "Vision-Based Human Activity Recognition: A Review," Journal of Healthcare Engineering, Vol 2017, pp. 1-31, 2017. <https://doi.org/10.1155/2017/3090343>
- [3] Vrigkas, Michalis, Christophoros Nikou, and Ioannis A. Kakadiaris, "A review of human activity recognition methods," Frontiers in Robotics and AI, Vol 2, article 28, 2015. <https://doi.org/10.3389/frobt.2015.00028>
- [4] Paul, Manoranjan, Shah ME Haque, and Subrata Chakraborty., "Human detection in surveillance videos and its applications-a review," EURASIP Journal on Advances in Signal Processing, Vol 176, No. 1, pp.1-16, 2013. <https://doi.org/10.1186/1687-6180-2013-176>
- [5] Toshev, Alexander, and Christian Szegedy., "DeepPose: Human pose estimation via deep neural networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1653-1660, 2014. <https://doi.org/10.1109/cvpr.2014.214>
- [6] Pishchulin, Leonid, et al., "Deepcut: Joint subset partition and labeling for multi person pose estimation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4929-4937. 2016. <https://doi.org/10.1109/cvpr.2016.533>
- [7] OpenPose: A Real-Time Multi-Person Keypoint Detection And Multi-Threading C++ Library, 2017. <https://github.com/CMU-Perceptual-Computing-Lab/openpose>
- [8] Cao, Zhe, et al., "Realtime multi-person 2d pose estimation using part affinity fields," arXiv preprint arXiv:1611.08050, 2016. <https://arxiv.org/abs/1611.08050>
- [9] Simon, Tomas, et al., "Hand Keypoint Detection in Single Images using Multiview Bootstrapping," arXiv preprint arXiv:1704.07809, 2017. <https://arxiv.org/abs/1704.07809>
- [10] Wei, Shih-En, et al., "Convolutional pose machines," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4724-4732, 2016. <https://doi.org/10.1109/cvpr.2016.511>
- [11] Seohee Park, Junchul Chun., "A Robust Object Detection and Tracking Method using RGB-D Model", Journal of Internet Computing and Services (JICS), Vol 18, No. 4, pp. 61-67, 2017. <http://dx.doi.org/10.7472/jksii.2017.18.4.61>
- [12] Papandreou, George, et al. "Towards Accurate Multi-person Pose Estimation in the Wild." arXiv preprint arXiv:1701.01779, 2017. <https://arxiv.org/abs/1701.01779>
- [13] Linna, Marko, Juho Kannala, and Esa Rahtu., "Real-time human pose estimation from video with convolutional neural networks," arXiv preprint arXiv:1609.07420, 2016. <https://arxiv.org/abs/1609.07420>
- [14] Ramakrishna, Varun, et al., "Pose machines: Articulated pose estimation via inference machines," European Conference on Computer Vision, pp.

- 33-47, 2014.  
[https://doi.org/10.1007/978-3-319-10605-2\\_3](https://doi.org/10.1007/978-3-319-10605-2_3)
- [15] Andriluka, Mykhaylo, et al., "2d human pose estimation: New benchmark and state of the art analysis," Proceedings of the IEEE Conference on computer Vision and Pattern Recognition, pp. 3686-3693, 2014.  
<https://doi.org/10.1109/cvpr.2014.471>
- [16] Bulat, Adrian, and Georgios Tzimiropoulos., "Human pose estimation via convolutional part heatmap regression," European Conference on Computer Vision, pp. 717-732, 2016.  
[https://doi.org/10.1007/978-3-319-46478-7\\_44](https://doi.org/10.1007/978-3-319-46478-7_44)
- [17] Belagiannis, Vasileios, and Andrew Zisserman., "Recurrent human pose estimation," Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on. IEEE, pp. 468-475, 2017. <https://doi.org/10.1109/fg.2017.64>
- [18] Google, "MNIST For ML Beginners,".  
<https://www.tensorflow.org>
- [19] Mehta, Dushyant, et al., "VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera," arXiv preprint arXiv:1705.01583, 2017.  
<https://arxiv.org/abs/1705.01583>
- [20] Ramakrishna, Varun, Takeo Kanade, and Yaser Sheikh., "Reconstructing 3d human pose from 2d image landmarks," Computer Vision - ECCV 2012, pp. 573-586, 2012.  
[https://doi.org/10.1007/978-3-642-33765-9\\_41](https://doi.org/10.1007/978-3-642-33765-9_41)

## ● 저 자 소 개 ●

### 박 서 희(Seohee Park)

2017 B.S. in Computer Science, Kyonggi University, Suwon, Korea  
 2017~Present : M.S. Student in Computer Science, Kyonggi University, Suwon, Korea  
 Research Interests : Computer Vision, Human Activity Recognition  
 E-mail : eehoeskrap@kgu.ac.kr



### 전 준 철(Junchul Chun)

1984 B.S. in Computer Science, Chung-Ang University, Seoul, Korea  
 1986 M.S. in Computer Science(Software Engineering), Chung-Ang University, Seoul, Korea  
 1992 M.S. in Computer Science and Engineering (Computer Graphics), The Univ. of Connecticut, USA  
 1995 Ph.D. in Computer Science and Engineering (Computer Graphics), The Univ. of Connecticut, USA  
 2001.02~2002.02 Visiting Scholar, Michigan State Univ. Pattern Recognition and Image Processing Lab.  
 2009.02~2010.02 Visiting Scholar, Univ. of Colorado, Wellness Innovation and Interaction Lab.  
 Research Interests : Augmented Reality, Computer Vision, Human Computer Interaction  
 E-mail : jchun@kgu.ac.kr

