특집논문 (Special Paper)
방송공학회논문지 제23권 제6호, 2018년 11월 (JBE Vol. 23, No. 6, November 2018)
https://doi.org/10.5909/JBE.2018.23.6.780
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

딥러닝 기반의 무기 소지자 탐지

김 건 욱[a)], 이 민 훈[a)], 허 유 진[a)], 황 기 수[a)], 오 승 준[a)‡]

Armed person detection using Deep Learning

Geonuk Kim[a)], Minhun Lee[a)], Yoojin Huh[a)], Gisu Hwang[a)], and Seoung-Jun Oh[a)‡]

요 약

전 세계적으로 총기 사고는 인적이 드문 장소뿐만 아니라 사람들이 많이 모여 있는 공공장소에서도 빈번하게 일어난다. 특히, 권총과 같은 소형 총기 사고의 빈도수가 매우 높다. 그러므로 사람에 비해 상대적으로 매우 작은 크기의 객체인 권총을 가진 사람을 탐지하는 것은 사고의 피해를 최소화하는데 핵심적이다. '권총 든 사람'을 탐지하는 연구가 수행되고 있지만, 사람보다 권총은 상대적으로 크기가 작기 때문에 단일 객체만을 탐지하는 기존 객체 탐지 방법으로 '권총 든 사람'을 탐지하면 오류 발생 빈도수가 매우 높다. 이러한 문제점을 해결하기 위하여 권총으로 무장한 사람을 탐지하는 방법으로 APDA(Armed Person Detection Algorithm)를 제안한다. APDA는 입력 영상에서 합성곱신경망(Convolutional Neural Network, CNN) 기반의 인체 특징점 탐지 모델과 객체 탐지 모델을 병행하여 획득한 양 손목과 권총의 위치를 후처리 작업에서 이용하여 '권총 든 사람'을 탐지한다. APDA는 기존 방식보다 객관적 평가에서 재현율이 46.3% 향상되었고, 정밀도는 14.04% 향상되었다.

Abstract

Nowadays, gun crimes occur very frequently not only in public places but in alleyways around the world. In particular, it is essential to detect a person armed by a pistol to prevent those crimes since small guns, such as pistols, are often used for those crimes. Because conventional works for armed person detection have treated an armed person as a single object in an input image, their accuracy is very low. The reason for the low accuracy comes from the fact that the gunman is treated as a single object although the pistol is a relatively much smaller object than the person. To solve this problem, we propose a novel algorithm called APDA(Armed Person Detection Algorithm). APDA detects the armed person using in a post-processing the positions of both wrists and the pistol achieved by the CNN-based human body feature detection model and the pistol detection model, respectively. We show that APDA can provide both 46.3% better recall and 14.04% better precision than SSD-MobileNet.

Keyword : Object-related human detection, Pose estimation, Object detection, CNN, Deep learning

a) 광운대학교 전자공학과(Department of Electronic Engineering, Kwangwoon University)
‡ Corresponding Author : 오승준(Seoung-jun Oh)
E-mail: sjoh@kw.ac.kr
Tel: +82-2-940-5102
ORCID: https://orcid.org/0000-0002-5036-3761

※이 논문의 연구결과 중 일부는 "한국방송·미디어공학회 2018년 하계학술대회"에서 발표한 바 있음.
※본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과 (IITP-2018-2016-0-00288)와 2017년도 광운대학교 교내 학술연구비 지원에 의해 연구되었음.
· Manuscript received September 7, 2018; Revised November 1, 2018; Accepted November 1, 2018.

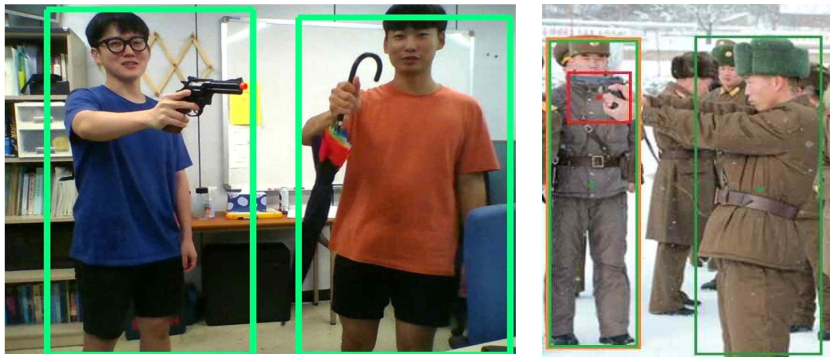
Ⅰ. 서 론

미국이나 유럽 몇몇 나라들과 같이 총기 소지가 가능한 나라를 포함하여 전 세계적으로 총기 관련 사고가 빈번하게 발생함을 뉴스를 통해서 쉽게 접할 수 있다. 총기 사고들의 특징들을 살펴보면 범죄자들이 은행이나 학교, 백화점과 같은 공공장소에 총기를 들키지 않고 반입하기 위하여 외투나 가방에 숨길 수 있을 정도로 작은 크기의 권총을 소지하는 경우뿐만 아니라 골목길같이 인적이 드문 곳에서 금품을 갈취하기 위해 권총이나 칼 등을 사용하여 사람들을 위협하는 경우도 빈번하다. 이에 대응하기 위해서 작은 크기의 흉기를 든 사람을 탐지하는 것은 인명 피해를 줄이는데 핵심적인 역할을 할 수 있다.

컴퓨터 비전 분야에서 관심 영역 및 객체를 탐지하는 연구는 오래전부터 지속적으로 수행되어왔다. 객체 탐지를 수행하는데 HOG(Histogram of Oriented Gradients), Harr-like feature 알고리즘, 딥 러닝(Deep learning) 등 다양한 방법들이 존재한다[1]-[3]. 이들 중에서 CNN 기반의 딥 러닝 방식은 특히 객체 탐지 분야에서 성능을 크게 향상시켰다. ILSVRC (ImageNet Large Scale Visual Recognition) 2015에서 CNN을 이용한 ResNet(Residual Network)[4] 모델은 3.6% 오차율을 기록하여 처음으로 훈련을 받은 사람이 영상을 분류할 때 발생한 오차율인 5%를 앞섰고, 최근 ILSVRC 2017에서는 CNN을 이용한 SENet(Squeeze-and-Excitation Networks)[5] 모델이 2.3%의 오차율을 기록하였다.

이와 같이 객체 탐지 분야에서 성능 향상이 크게 이루어졌지만, 여전히 한계점이 존재한다. 기존의 객체 탐지는 '사람', '총', '칼', '총을 든 사람', '칼을 든 사람' 등 탐지하려는 객체를 단일 객체로 탐지하기 때문에 '사람', '총', '칼' 등과 같은 각각의 객체에 대해서는 좋은 성능을 보이지만, '총을 든 사람', '칼을 든 사람' 등과 같이한 객체와 상대적으로 매우 작은 객체로 결합한 복합적인 요소의 객체를 탐지하는 경우에는 성능이 눈에 띄게 저하된다. 따라서 복합적인 요소의 객체를 정확하게 탐지하는 방법을 연구하는 것은 매우 중요하다.

그림 1(a)는 '권총 든 사람'이라는 복합적인 요소의 객체를 SSD-MobileNet[6][7] 방법으로 탐지하였을 때 발생한 오류를 보여준다. '권총'과 같은 작은 형태를 가진 물체는 전체 입력 특징 지도(Feature map)에서 차지하는 비율이 낮으므로 학습 시 크게 반영되지 않아 모든 형태의 사람을 '권총 든 사람'으로 탐지한다. 따라서 이 문제를 해결하기 위하여 YOLO(You Only Look Once)[8] 방법으로 '사람'과 '권총'을 따로 학습시켜 각각을 경계 상자(Bounding Box)로 표현하여 상자 간의 거리로 '권총 든 사람'을 탐지하는 방법을 적용하였다. 그러나 사람에 대한 경계 상자의 중심과 권총에 대한 경계 상자의 중심 간의 거리로 관계를 정립하면 그림 1(b)와 같은 영상에서 오 탐지가 발생한다. 그러므로 '권총 든 사람'을 하나의 객체로 학습시키는 방법이나 '사람'과 '권총'을 경계 상자로 탐지하는 방법으로는 원하는 결과를 얻을 수 없다.



(a) (b)

그림 1. '권총 든 사람' 탐지를 위한 기존 객체 탐지 방법의 오 탐지 결과 예
Fig. 1. Examples of false detection in conventional object detection methods for detection of the armed person

상기한 문제를 해결하기 위하여 본 논문에서는 CNN을 기반으로 인체의 특징점과 그 주변의 권총을 각각 탐지하고, 탐지된 인체 양 손목과 권총 사이의 거리 정보를 활용하여 '권총 든 사람'을 탐색하는 알고리즘으로 APDA(Armed Person Detection Algorithm)를 제안한다. APDA는 CMU-Pose[9]와 SSD-MobileNet을 이용하여 인체 특징점과 권총을 각각 탐지한 후, 후처리 작업(Post processing)을 통해 손목과 권총 사이의 거리 정보를 활용하여 복합적 객체 즉 '권총 든 사람'을 탐지한다.

본 논문은 다음과 같이 구성된다. 2장에서는 관련 이론에 대하여 그리고 3장에서는 제안하는 알고리즘에 대하여 자세히 설명한다. 4장에서는 제안한 알고리즘의 실험 결과와 이전 연구들을 비교하여 성능을 평가하고 5장에서 결론을 맺는다.

Ⅱ. 관련 이론

1. 객체 탐지(Object Detection)

영상 내 객체 탐지는 컴퓨터 비전 분야의 큰 연구 분야이고, 많은 연구가 이루어져 왔다. [10]은 최초로 합성곱신경망을 이용하여 합성 곱 연산이 이루어진 지역적 정보들을 종합하여 영상을 분석하는 알고리즘을 제안하였다. 이후 합성곱신경망은 영상 분석에 높은 성능을 보이며 이를 이용한 객체 탐지에 대한 연구들이 뒤따랐다. YOLO는 단일 CNN(Single CNN)을 통해 다중 경계 상자에 대한 클래스 확률을 계산하는 방식이다. 네트워크는 입력 영상을 448 x 448의 크기로 변경한 후, 단일 CNN을 진행하고, 모델의 신뢰에 의한 결과 탐지를 임계값으로 설정하여 객체의 위치와 종류를 알아낸다. YOLO는 초당 45프레임으로 복잡한 처리 과정을 가진 R-CNN과 같은 탐지 시스템들에 비해 빠른 처리 속도를 가진다. 하지만 작은 객체에 대해 상대적으로 낮은 정확도를 보인다. SSD는 순방향 신경망(Feed-Forward Neural Network, FFNet) 기반의 단일 심층 신경망을 사용하며, 후보 상자를 탐지하기 위해서 픽셀이나 특징을 다시 추출하지 않는 최초의 깊은 신경망 기반의 객체 탐지 네트워크다. 이전의 작업에서 수행되던 영역 후보 생성과 클래스 분류를 하는 2단계 작업을 한 번에 처리하였고, 깊이가 다른 복수의 특징 지도를 사용하여 얕은 쪽은 작은 객체를 탐지하고, 깊은 쪽은 큰 객체를 탐지할 수 있도록 깊이에 따라 상자의 크기가 바뀐 후 탐지한다. PASCAL VOC2007(PASCAL Visual Object Classes Challenge 2007)에서 73.2% mAP와 초당 7프레임의 성능을 가진 Faster-RCNN[11]과 63.4% mAP와 초당 45프레임의 성능을 가진 YOLO에 비해 SSD는 300 x 300 입력의 경우 초당 59프레임과 74.3% mAP를 얻었고, 512 x 512 입력의 76.9% mAP를 달성하여 탐지 정확도 및 속도가 향상됐다. ResNet은 기존 CNN의 합성곱 층이 깊어질수록 성능이 저하되는 문제를 해결하기 위한 네트워크다. ResNet 네트워크는 기존의 합성곱 층에 층의 입력과 층의 출력을 바로 연결하는 단축 연결(Shortcut connection)을 추가하였다. 즉, 참조되지 않은 함수를 학습하는 대신 층의 입력을 참조하여 잔차(Residual) 함수를 학습하는 것으로 층을 명시적으로 재구성한다. 그 결과 잔차 네트워크가 최적화하기 쉽고 상당히 깊어진 층에서 정확성을 얻을 수 있음을 보였다.

2. 무기 소지 탐지(Weapon detection)

객체 탐지의 응용 분야 중에서 보안과 관련된 분야에서는 권총이나 칼과 같이 무기를 탐지하는 것이 활발히 연구되고 있다. [12]는 Active Appearance Models(AMMs)을 사용하여 칼을 탐지하는데, 이는 Principle Component Analysis(PCA)[13]를 사용하여 칼의 형태를 추정하고, Harris corner-algorithm[14]을 사용하여 가장자리 특징점을 얻어내는 과정으로 구성되어있다. [15]는 CCTV 영상을 활용하여, 칼과 총 각각의 특성에 맞는 탐지 방법을 제안하였다. 우선 칼에 대한 탐지는 MPEC-7 특징 추출 방법을 사용하여 특징 지도를 구성하였고, 이를 사용하여 머신러닝의 한 종류인 SVM을 통해 칼에 대한 탐지를 진행하였다. 다음으로 총에 대한 탐지는 모폴로지(Mophology)를 이용한 배경 탐색과 캐니 에지(Canny Edge) 탐색[16], PCA, 3개의 층으로 구성된 Neural Network, MPEC-7 Visual Descriptor[17] 등을 이용한 칼의 형태 탐색을 통해 진행되었다. [18]은 Faster R-CNN의 가장 최신 모델인 VGG-16을 사용하여 손에 쥔 총에 대한 탐지를 진행하였다. 학습 및 평가는

IMFDB[19]를 사용하여 이루어졌으며 탐지의 실험 결과는 SVM, K-Nearest Neighbor(KNN)[20] 모델과의 비교를 통해 성능을 입증하였다.

3. 자세 추정(Pose Estimation)

많은 컴퓨터 비전 분야와 마찬가지로 사람의 각 신체 부위를 의미 있는 특징으로 보는 자세 추정은 중요한 가치를 지니고 있다. [21]은 두 단계로 구성된 하향식(Top-Down) 접근 방식을 사용한다. 첫 번째 단계에서는 Faster R-CNN을 이용하여 사람들을 포함할 가능성이 있는 경계 상자의 크기와 위치를 예측한다. 두 번째 단계에서는 제안된 각 경계 상자에 위치하는 사람의 키포인트를 추정한다. 각 인체 특징점에 대해 FCN인 ResNet을 사용하여 밀도가 높은 히트 맵(Heat map) 오프셋(Offset)을 예측한다. 이후 두 가지 출력을 결합하여 높은 키포인트 예측을 얻는다. 이 알고리즘의 정확도는 COCO(Common Objects in Context) 테스트 셋에서 59.8% mAP를 달성하였다. 한편, CMU-Pose는 상향식(Bottom-Up) 접근 방식으로 2-branch로 구성된 다중 단계 CNN을 사용한다. 입력 영상에 대해 하나의 CNN에서는 각 관절의 위치를 히트 맵 형태로 예측하여 그림 2의 번호로 나타내고, 다른 하나의 CNN에서는 각 신체 부위의 연관성을 표현하는 벡터 맵인 부위 선호도 필드(Part Affinity Fields, PAFs)를 예측한다. 그리고 각각 예측된 정보를 결합하여 각 관절을 연관 지어 그림 2와 같이 최종 출력을 뽑아낸다. 이 알고리즘의 정확도는 COCO 테스트 셋에서 60.5% mAP를 달성했으며, 이전의 작업보다 향상된 정확성을 보인다. 그리고 초당 200프레임의 성능을 가져 실시간 자세 추정에 적합하다.

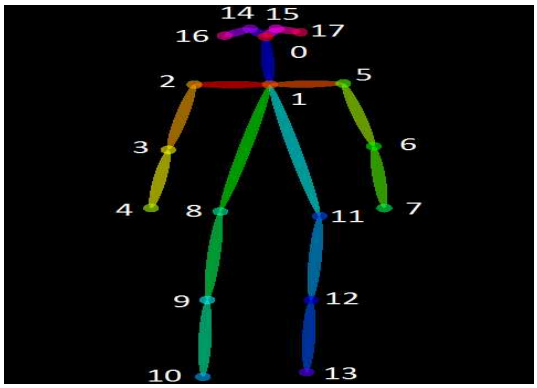


그림 2. CMU-pose의 인체 특징점 지도
Fig. 2. A Keypoint map of CMU-pose

Ⅲ. APDA(Armed Person Detection Algorithm)

인체 특징점을 기반으로, '권총 든 사람'을 찾기 위한

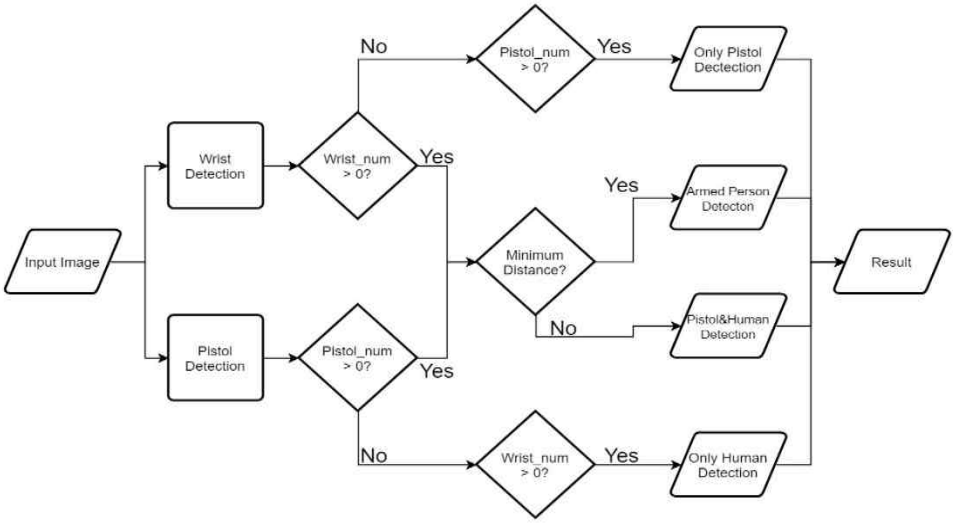


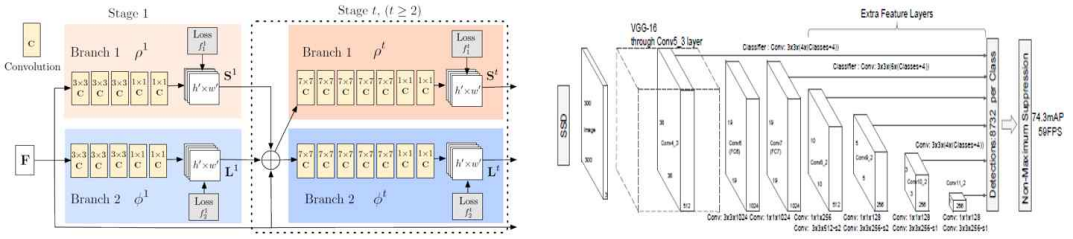
그림 3. APDA의 순서도
Fig. 3. A Flowchart of APDA

APDA는 그림 3과 같다. '권총 든 사람'을 탐지하는 것이므로 양 손목의 정보만 사용한다. 영상이 입력되면 CMU-pose를 이용하여 그림 4(a)와 같이 입력 영상에 대한 인체 열 지도(Heat map)와 방향 지도(Vector map)를 추정하여 스켈레톤 형태의 사람을 탐지한다. 이는 Bottom-up 방식을 사용하기 때문에 서로 다른 사람의 특정한 인체 특징점이 근접할 때 발생하는 오류로부터 복원되는 성능이 뛰어나다. 그리고 동일 입력 영상에 대해 SSD-MobileNet을 이용하여 권총을 탐지한다. SSD-MobileNet은 그림 4(b)와 같이 깊이가 다른 복수의 특징 지도를 사용하기 때문에 작은 인체 주변의 객체 탐지에 적합하고, 분리 가능한 합성곱신경망을 사용하기에 적은 매개변수를 효율적으로 학습한다.

권총과 사람의 손목 사이의 최소 거리는 식 (1)을 통해 구할 수 있다.

$$argmin(d(m,k)) = \|w_m - p_k\|_2, \quad 1 \le m \le M, \quad 1 \le k \le K \tag{1}$$

여기서 w_m은 탐지된 사람 M명 중 m번째 사람의 오른쪽 손목 b_0와 왼쪽 손목 b_1으로 이루어져 있으며 각각의 손목은 2차원 좌표를 가지고 있다. 따라서 $w_m = (b_0, b_1)$로 표현한다. p_k는 탐지된 K개의 총 중 k번째 총의 경계 상자 중심점의 2차원 좌표이다. $\|\cdot\|_2$ 은 두 점 사이의 거리를 구하는 유클리디안(Euclidean) 거리 표현 방식이다. k번째 총에 대해 탐지된 모든 사람의 양 손목과의 거리를 각각 구한 후 거리가 최소가 될 때의 변수 m을 찾아, k번째 총을 든 사람을 탐지한다. 예를 들어, 그림 5와 같이 입력 영상에 대해 인체 추정과 권총 탐지를 진행한다. 이후 탐지된 권총의 경계 상자 중심점으로부터 탐지된 모든 손들 사이의 거리를 각각 계산하여 영상 내 권총과 가장 가까운 사람의



(a) (b)

그림 4. APDA의 두 가지 탐지 모듈 구조 (a) CMU pose, (b) SSD
Fig. 4. Two major blocks for object detection in APDA: (a) CMU pose, and (b) SSD

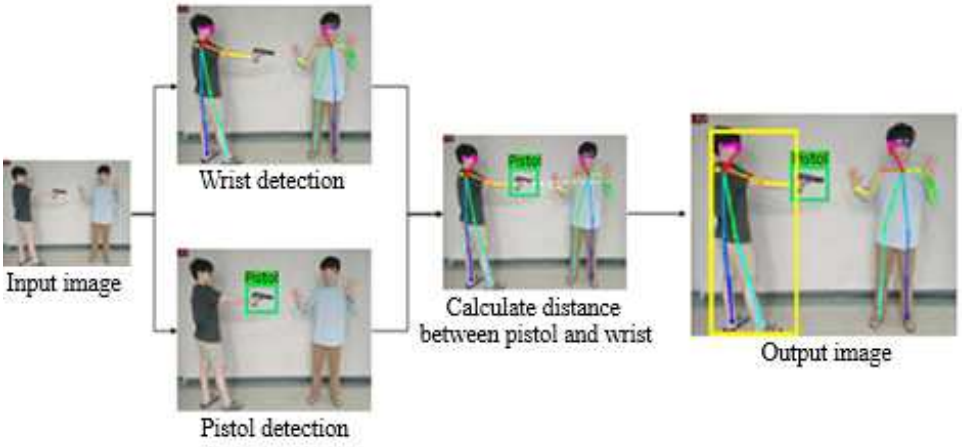


그림 5. APDA 처리 예
Fig. 5. An APDA processing example

손을 찾는다. 그 결과 왼쪽 사람의 손이 오른쪽 사람의 손보다 권총과의 거리가 짧으므로, 왼쪽 인물이 '권총 든 사람'으로 탐지되었다.

Ⅳ. 실험결과

본 논문에서 사용하는 학습 데이터셋은 ImageNet, IMFDB 및 직접 촬영한 영상을 사용하였으며 각각의 데이터셋들을 'TFrecord' 형식으로 변환하여 학습에 사용하였다. 실험은 두 가지로 진행되는데, 이는 '권총' 데이터셋 형태를 결정하기 위한 첫 번째 실험과 APDA의 성능 평가를 위한 두 번째 실험으로 구성된다.

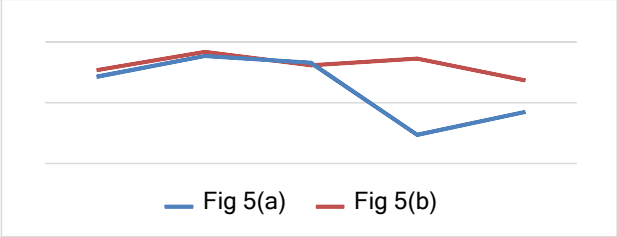
1. 권총 데이터셋 형태 결정 과정

'권총 든 사람'을 탐지할 때 일반적으로 권총의 손잡이가 사람의 손에 가려져 있으므로, '권총' 데이터셋을 그림 6과 같이 '권총'과 '손에 쥔 권총' 두 형태로 나누어 각각 SSD-MobileNet을 통해 학습시키고 그 성능을 비교하였다. 이때 경계 상자는 그림 6과 같이 각 데이터셋의 형태에 맞추어 지정하였다.

표 1은 '권총' 데이터 500개와 '손에 쥔 권총' 데이터 500개를 가지고 단계별로 학습한 후 각 데이터셋 형태에 따라 단계별로 정밀도를 평가한 것을 나타낸다. 이때, 직접 구축한 455개의 입력 영상을 통해 이에 대한 탐지 정밀도를 평가하였다. 실험 결과 '4500'단계 이후로 오버피팅(Overfitting)이 발생하여 탐지 정밀도가 감소하였다. 한편, '손에 쥔 권총' 데이터셋으로 학습시킬 경우 '권총' 데이터셋으로 학습했을 경우에 비해 정밀도가 평균적으로 5.73% 높았다.

표 1. '권총'과 '손에 쥔 권총'의 평균 정밀도
Table 1. Average Precision for both 'pistol' and 'pistol held by hands' data

Epoch	4000	4500	5000	5500	6000	Average
Fig 5(a)	91.49%	96.52%	94.83%	77.10%	82.67%	88.52%
Fig 5(b)	93.07%	97.51%	94.21%	95.88%	90.59%	94.25%



2. APDA의 성능평가

위 실험 결과를 바탕으로 '손에 쥔 권총' 형태의 데이터 1000개를 이용하여 APDA의 SSD-MobileNet 모듈에 학습을 진행하였다. 그리고 표 2와 같이 동일한 450개의 실험 영상에 대해 각 단계 학습 결과의 평균 정밀도를 평가하였다. 이때 '6000'단계일 때의 평균 정밀도가 93.10%로 가장 높았다. 이후 학습 단계에서 정밀도가 떨어지는 오버피팅이 발생하였기 때문에, '6000' 단계의 학습결과를 사용하여 APDA의 SSD-MobileNet 모듈을 확정하였다. 한편, APDA

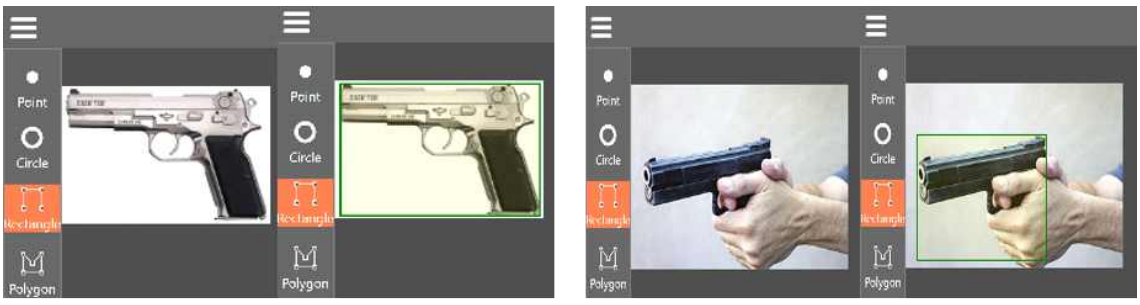


그림 6. 훈련에 사용한 두 종류 데이터 예 (a) '권총' 데이터 (b) '손에 쥔 권총' 데이터
Fig. 6. Examples of two kinds of pistol training data in APDA (a) 'pistol', and (b) 'pistol held by hands'

의 CMU-pose 모듈의 매개변수는 CMU-pose에서 제공한 값을 사용하였다.

표 2. '손에 쥔 권총' 학습 시 단계 별 평균 정밀도
Table 2. Average Precision per step in learning for 'pistol held by hands'

Epoch	4500	5000	5500	6000	6500
Average Precision	87.95%	89.94%	90.27%	93.10%	91.29%

이후 APDA의 성능 평가를 위하여 '권총 든 사람' 데이터 1000개를 학습한 SSD-MobileNet에 대한 실험과 APDA에 대한 실험을 진행하여 성능을 비교하였다. 이때 객관적인 성능 평가를 위해 두 실험 모두 직접 구축한 500개의 입력 영상에 대해 동일한 실험을 진행하였다. 그림 7은 APDA에 대한 결과의 한 예이다. 먼저 왼쪽의 권총이 탐지되고 영상 내 두 인물에 대한 각각의 인체 특징점이 탐지된다. 이후 식 (1)을 사용하여, 왼쪽 사람이 '권총 든 사람'으로 탐지되었다. 이는 기존 방식으로 탐지한 결과인 그림 1과 달리 작은 객체인 권총을 정확하게 탐지한 결과이다. 한편 식 (2)와 같이 정의된 정밀도(*precision*)와, 재현율(*recall*)을 측정한 결과를 표 3에 정리하였다.

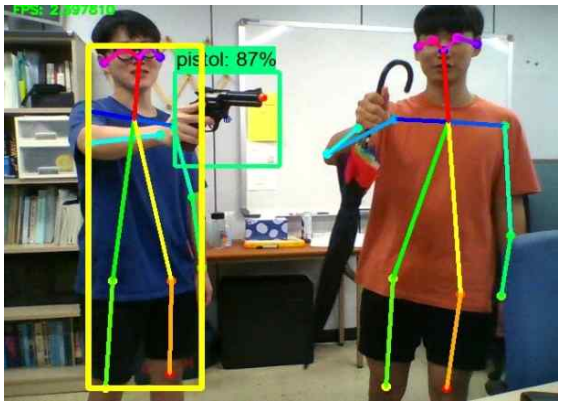


그림 7. APDA를 이용한 실험결과의 예
Fig. 7. An example of experimental result with APDA

$$precision = \frac{TP}{(TP+FP)}\ ,\ recall = \frac{TP}{(TP+FN)} \quad (2)$$

여기서, TP, FN, FP, TN은 각각 '권총 든 사람'을 탐지한 경우, '권총 든 사람'을 탐지하지 못한 경우, '권총'을 들고 있지 않은 사람을 탐지한 경우, '권총'을 들고 있지 않은 사람을 탐지하지 않은 경우의 수이다.

표 3. SSD-MobileNet과 APDA에 대한 정밀도와 재현율
Table 3. Precision and recall values for SSD-MobileNet and APDA

Methods	Precision	Recall
SSD-MobileNet	18.24 %	43.02 %
APDA	32.28 %	89.30 %

SSD-MobileNet에 대한 정밀도와 재현율은 각각 18.24 %, 43.02% 이고, APDA에 대한 정밀도와 재현율은 각각 32.28%, 89.3% 이다. APDA는 SSD-MobileNet보다 46.3% 향상된 재현율을 보이고, 14.04% 향상된 정밀도를 보였다. 하지만 낮은 정밀도를 보였는데, 그 이유는 APDA의 경우 작은 객체인 '권총'의 특징 지도가 아주 작기 때문에 탐지가 정확하게 이루어지지 않았기 때문이다. 그리고 APDA에도 오 탐지가 발생했다. 그 이유는 APDA의 후처리 과정에서 깊이 정보를 배제한 이차원 거리만을 사용하기 때문이다. 그림 8은 이와 같은 원인으로 오 탐지가 발생한 한 예이다. 그림 8에서 '권총'과 인체 특징점은 정확하게 탐지되었지만 '권총'과 '사람'간의 깊이 정보를 반영하지 못하여 오 탐지하였다. 향후 이차원 영상에서 깊이 정보를 추출할 수 있는 방법[22]을 이용하여 깊이 정보를 고려하면 오 탐지율을 줄일 수 있을 것이다.

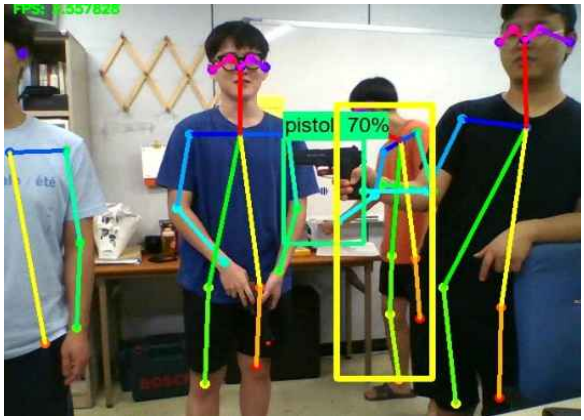


그림 8. APDA로 오 탐지된 예
Fig. 8. An example of false detection in APDA

V. 결 론

전 세계적으로 총기사고는 인적이 드문 장소뿐만 아니라 사람들이 많이 모여 있는 공공장소에서도 빈번하게 일어난다. 특히, 권총과 같은 소형 총기 사고의 빈도수가 매우 높다. 그러므로 사람에 비해 상대적으로 매우 작은 크기의 객체인 권총을 가진 사람을 탐지하는 것은 사고의 피해를 최소화하는데 핵심적이다. 기존 객체 탐지방법은 탐지하려는 객체를 단일 객체로 탐지하기 때문에 각각의 객체에 대해서는 좋은 성능을 보인다. 하지만 한 객체와 상대적으로 매우 작은 객체로 결합한 복합적인 요소의 객체를 탐지하게 되는 경우에는 객체 탐지 성능이 눈에 띄게 저하된다. 본 논문에서는 이러한 문제점을 해결하기 위하여 권총으로 무장한 사람을 탐지하는 방법으로 APDA를 제안하고 실험을 통하여 그 성능을 검증하였다. APDA는 CMU-pose 방법으로 양 손목의 위치를 찾는 모듈과 SSD-MobileNet으로 권총의 경계 상자의 위치를 찾아내는 모듈이 병렬적으로 배치되고, 이 두 모듈에서 제공되는 정보를 이용하는 후처리 모듈로 구성되었다. SSD-MobileNet 모듈에서는 '권총 든 사람'의 '권총'은 사람의 손에 가려진다는 점을 고려하여 '손에 쥔 권총' 데이터셋으로 학습시켜 권총의 탐지 정확도를 높였다. 권총과 양 손목 사이의 거리가 최소가 되는 사람을 '권총 든 사람'으로 탐지하였다. 객관적 성능을 평가하기 위한 지표로 정밀도, 재현율을 사용하였다. APDA의 재현율은 기존 방법 대비 46.3% 향상된 89.3%를 보였고, 정밀도는 기존 방법 대비 14.04% 향상된 32.28%를 보이며 기존 방법보다 강인하게 객체를 탐지하였다.

참 고 문 헌 (References)

[1] Kwangsoo Kim, Ungtae Kim and Sooyeong Kwak, "Real-time Violence Video Detection based on Movement Change Characteristics" JBE, Vol.22, No. 2, pp. 234-239, March 2017, http://dx.doi.org/10.5909/JBE.2017.22.2.234 (accessed Aug. 1, 2018).

[2] Sanggi Kim and Dongseog Han, "Real Time Traffic Light Detection Algorithm Based on Color Map and Multilayer HOG-SVM" JBE, Vol. 22, No. 1, pp. 62-69, Jenuary 2017, http://dx.doi.org/10.5909/JBE.2017.22.1.62 (accessed Aug. 3, 2018).

[3] Seulbeen Kim and Wonjun Kim, "User Identification Method using Palm Creases and Veins based on Deep Learning" JBE, Vol. 23, No. 3, pp. 395-402, May 2018, http://dx.doi.org/10.5909/JBE.2018.23.3.395 (accessed Aug. 3, 2018).

[4] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition" *In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 770-778, 2016, https://doi.org/10.1109/cvpr.2016.90 (accessed Aug. 10, 2018).

[5] J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Network" arXiv: 1709.01507, 2017, https://arxiv.org/pdf/1709.01507 (accessed Aug. 10, 2018).

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, Cheng-Yang Fu and Alexander C. Berg, "SSD: Single Shot MultiBox Detector" In Proceeding of the European Conference on Computer Vision(ECCV), pp.21-37, 2016, https://doi.org/10.1007/978-3-319-46448-0_2 (accessed Sep 8, 2018).

[7] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, "MobileNets: Efficient Convolutional Neural Network for Mobile Vision Applications" arXiv: 1704.04861, 2017, https://arxiv.org/abs/1704.04861 (accessed Sep 20, 2018).

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection" In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016, https://doi.org/10.1109/cvpr.2016.91(accessed Sep 8, 2018).

[9] Z. Cao, T. Simon, Shih-E. Wei and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields" In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1302-1310, 2017, https://doi.org/10.1109/cvpr.2017.143 (accessed Sep 8, 2018).

[10] Y. Lecun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition" Neural Computation, vol. 1, no. 4, pp 541-551, Winter 1989, 10.1162/neco.1989.1.4.541 (accessed Aug. 5, 2018).

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, June 2017, https://doi.org/10.1109/tpami.2016.2577031 (accessed Aug. 10, 2018).

[12] A. Glowacz, M. Kmiec and A. Dziech, "Towards Robust Visual Knife Detection in Images: Active Appearance Models Initialised with Shape-specific Interest Points" In Multimedia Communications, Services and Security : 5th International Conference. vol. 287, pp. 148-158, 2012, https://doi.org/10.1007/978-3-642-30721-8_15 (accessed Aug 9, 2018).

[13] L. Malagón-Borja, and O. Fuentes, "Object detection using image reconstruction with PCA" *Image and Vision Computing*, vol. 27, no. 1-2, pp. 2–9, 2009, https://doi.org/10.1016/j.imavis.2007.03.004 (accessed Aug 11, 2018).

[14] Derpanis KG, "The Harris corner detector" http://www.cse.yorku.ca/~kosta/CompVis_Notes/harris_detector.pdf(accessed Aug. 10, 2018)

[15] M. Grega, A. Matiolanski, P. Guzik and M. Leszczuk, "Automated Detection of Firearms and Knives in a CCTV Image" *Sensors*, vol. 16,

no. 1. Jan 2016, https://doi.org/10.3390/s16010047 (accessed Aug 2, 2018).

[16] J. Canny, “A Computational Approach to Edge Detection” *In IEEE Trans. Pattern Anal, Machine Intell.*, vol. PAMI-8, issue 6, pp. 679-698, Nov 1986, https://doi.org/10.1016/b978-0-08-051581-6.50024-6 (accessed Aug 1, 2018).

[17] B.S. Manjunath, Philippe Salembier and Thomas Sikora, *Introduction to MPEG-7, Multimedia Content Description Interface.* Wiley, USA, 2002, https://doi.org/10.1007/springerreference_72884 (accessed Aug 15, 2018).

[18] Gyanendra K. Verma and Anamika Dhillon, “A HandHeld Gun Detection using Faster R-CNN Deep Learning” *In Proceeding of the 7th International Conference on Computer and Communication Technology*, pp. 84-88, November 2017, https://doi.org/10.1145/3154979.3154988 (accessed Aug 8, 2018).

[19] IMFDB: Internet Movie Firearms Database, http://www.imfdb.org/wiki/Main_Page (accessed Aug.15, 2018).

[20] J.M. Keller, M.R. Gray and J.A. junior, “A Fuzzy K-Nearest Neighbor Algorithm” *In IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-15, issue 4, pp. 580-585, 1985, https://doi.org/10.1109/tsmc.1985.6313426 (accessed Aug 9, 2018).

[21] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler and K. Murphy, “Towards Accurate Multi-person Pose Estimation in the Wild” *In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2017, https://doi.org/10.1109/cvpr.2017.395 (accessed Sep 5, 2018).

[22] A. Saxena, S. H. Chung and A. Y. Ng, “3-D Depth Reconstruction from a Single Still Image” *International Journal of Computer Vision*, Vol. 76 Issue 1, pp. 53-69, January 2008, https://doi.org/10.1007/s11263-007-0071-y (accessed Sep 1, 2018).

저 자 소 개

김 건 욱

- 2015년 3월 ~ 현재 : 광운대학교 전자공학과 학사과정
- ORCID : https://orcid.org/0000-0002-5055-3338
- 주관심분야 : 신경망

이 민 훈

- 2013년 3월 ~ 현재 : 광운대학교 수학과 학사과정
- ORCID : https://orcid.org/0000-0001-8165-5380
- 주관심분야 : 컴퓨터비전, 영상압축, 영상처리

허 유 진

- 2015년 3월 ~ 현재 : 광운대학교 전자공학과 학사과정
- ORCID : https://orcid.org/0000-0003-0250-8986
- 주관심분야 : 컴퓨터비전, 영상압축, 딥러닝

저 자 소 개

황 기 수

- 2012년 3월 ~ 현재 : 광운대학교 전자공학과 학사과정
- ORCID : https://orcid.org/0000-0003-1046-9286
- 주관심분야 : 컴퓨터비전, 영상압축, 영상처리

오 승 준

- 1980년 2월 : 서울대학교 전자공학과 학사
- 1982년 2월 : 서울대학교 전자공학과 석사
- 1988년 5월 : 미국 Syracuse University 전기/컴퓨터공학과 박사
- 1982년 3월 ~ 1992년 8월 : 한국전자통신연구원 근무
- 1986년 7월 ~ 1986년 8월 : NSF Supercomputer Center 초청 학생연구원
- 1987년 5월 ~ 1988년 5월 : Northeast Parallel Architecture Center 학생연구원
- 1992년 3월 ~ 1992년 8월 : 충남대학교 컴퓨터공학부 겸임교수
- 2002년 3월 ~ 2017년 12월 : SC29-Korea 전문위원회 위원장
- 1992년 9월 ~ 현재 : 광운대학교 전자공학과 교수
- 2002년 3월 ~ 현재 : MPEG 뉴미디어 포럼 부의장
- ORCID : https://orcid.org/0000-0002-5036-3761
- 주관심분야 : 비디오데이터처리, 컴퓨터비전, 머신러닝, 딥러닝