

특집논문 (Special Paper)

방송공학회논문지 제23권 제3호, 2018년 5월 (JBE Vol. 23, No. 3, May 2018)

<https://doi.org/10.5909/JBE.2018.23.3.361>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

DNN 기반 컬러와 열 영상을 이용한 다중 스펙트럼 보행자 검출 기법

이 용 우^{a)}, 신 지 태^{a)†}

DNN Based Multi-spectrum Pedestrian Detection Method Using Color and Thermal Image

Yongwoo Lee^{a)} and Jitae Shin^{a)†}

요 약

자율주행 자동차의 연구가 빠르게 발전하는 가운데 보행자 검출에 대한 연구 또한 성공적으로 진행되고 있다. 그러나 대부분의 연구에서 사용되는 데이터셋이 컬러영상을 기반으로 있고 또한 보행자의 인식이 상대적으로 쉬운 영상이 많다. 컬러 영상의 경우 보행자가 빛에 노출되는 정도에 따라 영상에 제대로 포착이 되지 않을 수 있고 이로 인해 기존 방식들로는 이러한 보행자를 제대로 검출하지 못하는 상황이 발생한다. 따라서 본 논문에서는 DNN (deep neural network) 기반 컬러 영상과 열 영상을 이용한 다중 스펙트럼 보행자 검출 기법을 제안하고자 한다. 기존의 SSD (single shot multibox detector) 기법을 기반으로 하여 컬러 영상과 열 영상을 동시에 활용하는 퓨전 네트워크 구조를 제안한다. 실험은 KAIST의 데이터셋을 이용하여 실시하였고 제안한 기법인 SSD-H (SSD-Halfway fusion)의 방식이 KAIST 보행자 검출기준의 기준치보다 18.18% 낮은 miss rate를 획득하였고 또한 기존 halfway fusion 기법에 비해 최소 2.1% 낮은 miss rate를 획득하였다.

Abstract

As autonomous driving research is rapidly developing, pedestrian detection study is also successfully investigated. However, most of the study utilizes color image datasets and those are relatively easy to detect the pedestrian. In case of color images, the scene should be exposed by enough light in order to capture the pedestrian and it is not easy for the conventional methods to detect the pedestrian if it is the other case. Therefore, in this paper, we propose deep neural network (DNN)-based multi-spectrum pedestrian detection method using color and thermal images. Based on single-shot multibox detector (SSD), we propose fusion network structures which simultaneously employ color and thermal images. In the experiment, we used KAIST dataset. We showed that proposed SSD-H (SSD-Halfway fusion) technique shows 18.18% lower miss rate compared to the KAIST pedestrian detection baseline. In addition, the proposed method shows at least 2.1% lower miss rate compared to the conventional halfway fusion method.

Keyword : CNN, pedestrian detection, multi-spectrum, network fusion

Copyright © 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

1. 서론

보행자 검출에 대한 연구는 최근 떠오르는 자율주행 자동차 연구의 경향에 따른 다양한 연구분야 중 하나로 다양한 기업과 연구기관에서 관련 연구를 활발히 진행하고 있다. 최근 보행자 검출 기법은 DNN의 발전에 따라 그 성능이 매우 빠르게 상승하고 있지만 여전히 컴퓨터와 사람이 보행자를 인식하는 정확도의 차이가 매우 큰 상황이다^[1]. 특히 보행자가 가려지는 현상이 발생하거나 영상의 해상도가 낮고 배경이 혼란한 경우 보행자 인식률을 더욱 낮아지게 된다. 이 뿐만 아니라 현재 대부분의 보행자 검출에 대한 연구가 빛의 노출이 매우 높은 컬러 영상을 기반으로 하다 보니 실제 테스트 영상이 밤 영상이거나 보행자가 그늘로 들어가는 경우, 빛 뭉침이 심한 상황에서는 보행자 검출을 실패하는 경우가 많이 발생된다. 이러한 결점은 밤낮 구분 없이 운행되어야 하는 자율주행 자동차에 있어 매우 큰 장애물이다.

이러한 문제점은 컬러 카메라를 깊이맵 카메라나 열 영상 카메라와 같이 다양한 센서를 함께 사용하는 방식으로 해결할 수 있다. 그 중 열 영상 카메라의 경우 영상을 포착하는 과정에서 빛의 효과를 덜 받기 때문에 컬러 영상과 함께 상호보완적으로 사용될 수 있다. 그러나 대부분의 이전 연구들은 컬러 영상처럼 단일 센서로부터 획득한 데이터만을 이용하는 것에 초점을 맞추고 있다. 하지만 최근의 몇몇 연구들은 컬러 영상과 열 영상을 함께 사용해 보행자를 검출하는 방식을 소개하고 있다. [2]의 논문에서는 R-CNN (region-based convolutional neural networks)^[3]을 기반으로 한 네트워크 퓨전 구조를 소개하고 있다. 그러나

R-CNN의 경우 성능이 최근의 기법들^[4-6]에 비해 매우 떨어지는 문제점이 있으며, [2]의 퓨전 구조 또한 매우 단순한 구조를 지니는 단점이 있다. [7]의 논문에서는 R-CNN의 상향 버전인 Faster R-CNN^[5]를 기반으로 한 네트워크 퓨전 구조를 제안했다. 그러나 Faster R-CNN이 R-CNN에 비해 속도적인 측면에서 큰 개선이 있지만 정확도 측면에서는 별다른 개선이 없는 문제점이 있다. [8]의 논문에서는 SSD (Single Shot multibox Detector)를 제안하였는데 다양한 크기와 중첩비를 가지는 기본 영역 박스 (default bounding boxes)를 이용하여 기존 Faster R-CNN보다 탐지 정확도가 향상되는 결과를 보여주었다. 또한 SSD는 Faster R-CNN에서 사용하는 RPN (Region Proposal Networks)과 같이 영역 박스의 예측을 위한 샘플링 단계가 없기 때문에 매우 속도가 빠르고 자율주행과 같이 실시간으로 데이터 처리가 필요한 영역에 매우 적합한 방법이다.

따라서 본 연구에서는 [7]의 네트워크 퓨전 구조를 기반으로 한 보행자 검출 기법을 제안하고자 한다. Faster R-CNN의 단점을 보완하기 위해 본 논문에서는 SSD^[8]을 기본 DNN 구조로 삼고 [7]에서 제안한 퓨전 방식을 사용하고자 한다. 실험은 KAIST의 데이터셋^[9]을 사용하고 SSD를 각각 컬러 영상과 열 영상만으로 트레이닝한 방식과 비교하고자 한다. 또한 KAIST의 보행자 검출 기준의 기준치 방법인 ACF+T+THOG (aggregated channel features with thermal and thermal histogram of oriented gradients)^[9]와 [7]의 퓨전구조인 early fusion, halfway fusion 그리고 late fusion 방식을 적용한 SSD-E (SSD-early fusion), SSD-H (SSD-halfway fusion), SSD-L (SSD-late fusion)과 함께 비교한다. 검출 지표로는 FPPI (false positive per image) - miss rate를 사용한다. 제안 방법인 SSD-H는 ACF+T+THOG에 비해 18.18% 낮은 miss rate를 획득하였고 [7]의 halfway fusion에 비해 2.1% 낮은 miss rate를 획득하였다.

본 논문의 나머지 구조는 다음과 같다. 2장에서는 [7]의 논문에서 소개하는 네트워크 퓨전 구조와 SSD의 네트워크 구조에 대해 설명한다. 3장에서는 SSD를 컬러 영상과 열 영상을 같이 사용하는 네트워크 퓨전 구조를 설명한다. 4장에서는 제안한 방법을 비교하는 실험 방식과 실험 결과를 정리하고 마지막으로 5장에서 결론과 후속 연구에 대해 정리하고 마무리한다.

a) 성균관대학교 전자전기공학부 (School of Electronic and Electrical Engineering, Sungkyunkwan University)

‡ Corresponding Author : 신지태(Jitae Shin)

E-mail: jtshin@skku.edu

Tel: +82-31-290-7994

ORCID: <https://orcid.org/0000-0002-2599-3331>

※ 이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2017R1D1A1B03031752).

※ This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education(NRF-2017R1D1A1B03031752).

· Manuscript received March 30, 2018; Revised April 30, 2018; Accepted April 30, 2018.

II. 관련 연구

본 장에서는 [7]의 논문에서 소개하는 네트워크 퓨전 구조와 SSD의 네트워크 구조에 대해 알아본다.

1. 네트워크 퓨전 구조

그림 1에서는 [7]의 논문에서 소개한 Faster R-CNN을 기반으로 하는 3개의 네트워크 퓨전 방식을 보여준다. 각각 early fusion, halfway fusion 그리고 late fusion이다. VGG-16^[10] 구조를 기반으로 5개의 Conv. (convolutional layer)와 2개의 FC (fully connected layer)를 가지는 네트워크 구조에서 early fusion의 경우 컬러와 열 영상으로부터 획득한 특징맵 (feature map)을 첫 번째 Conv. 이후에 결합시킨다. 특징맵의 결합은 Concat. (concatenate layer)를 이용하고 그 이후 1×1 convolutional layer인 NIN (Network-in-Network)^[11]를 이용해서 새로 생성된 특징맵의 크기를 줄

반으로 줄인다. Early fusion은 첫 번째 Conv. 이후의 특징맵을 합치게 되므로 보통 낮은 레벨의 특징들이 합쳐지게 된다. Halfway fusion은 4번째 Conv.에서 특징맵을 합치게 되고 마찬가지로 NIN이 Concat. 이후에 자리한다. Halfway fusion의 경우 4번째 Conv.를 통과한 특징들을 합치게 되므로 좀 더 의미론적인 (semantic) 정보를 가지는 동시에 세밀한 디테일을 보존하는 특징이 있다. Late fusion의 경우 마지막 FC를 통과한 특징맵을 합치게 된다. 일반적으로 마지막 FC의 특징맵은 물체의 새로운 표현을 위해 사용되므로, late fusion은 높은 레벨의 특징 퓨전이라고 할 수 있다. [7]의 논문에서 추가로 제안된 score fusion의 경우 두 개의 네트워크를 직렬 연결하는 방식이기 때문에 네트워크를 퓨전 하려는 본 논문의 의도와는 맞지 않아 제외한다.

2. SSD 네트워크 구조

SSD는 [8]에서 제안된 DNN 기반 물체 검출기이다. 그림

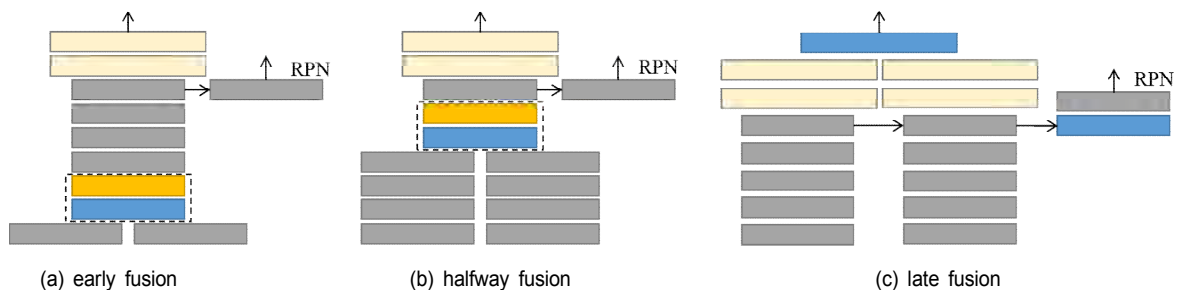


그림 1. 여러 네트워크 퓨전 구조 [7]: 회색 박스는 convolutional layer, 파란색 박스는 concatenate layer, 오렌지색 박스는 NIN layer, 밝은 노란색 박스는 fully connected layer이다.

Fig. 1. The network fusion architectures [7]: Gray and blue boxes indicate convolutional and concatenate layers, orange and bright yellow boxes represent NIN and fully connected layers.

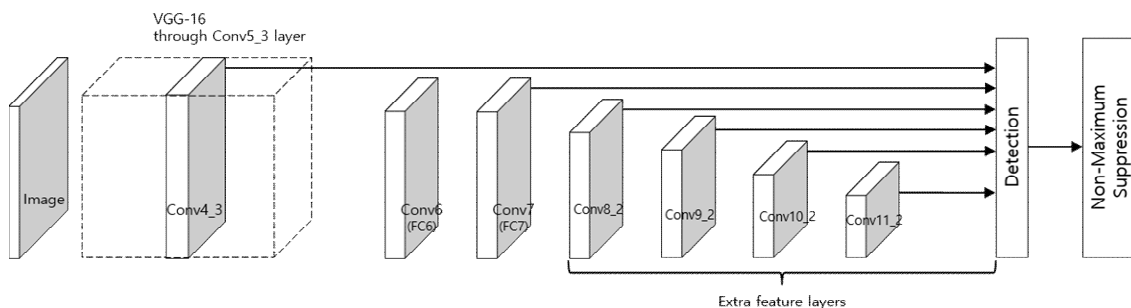


그림 2. SSD 네트워크 구조: 간단한 표현을 위해 ReLU layer, pooling layer, dropout layer 들은 그림에서 제외함

Fig. 2. SSD network structure: ReLU layers, pooling layers, dropout layers are excluded to make it simple

2에서는 [8]에서 제안된 SSD의 네트워크 구조를 보여주고 있는데 기존의 VGG-16^[10]의 FC6과 FC7를 Conv.로 교체하고 추가로 4개의 특징 layer를 연결하는 구조를 가진다. SSD는 기존 Faster R-CNN에서 사용하는 RPN과 같이 검출 영역을 제안하는 네트워크가 따로 존재하지 않기 때문에 속도 측면에서 매우 뛰어난 장점을 가진다. SSD는 물체의 영역을 예측할 때 기본 영역 박스들과 카테고리 점수를 사용한다. 이 기본 영역 박스는 Conv.를 통과하면서 생기는 다양한 스케일의 특징맵에 적용되어 검출 영역을 예측하게 된다. 또한 기본 영역 박스는 다양한 중첩비로 구성되어 있어 검출하고자 하는 물체의 영역을 좀 더 정확하게 제안할 수 있다. 정확도를 위한 손실 함수에는 지역 손실과 신뢰 손실로 나누어져 있다. 지역 손실의 경우 예측한 영역의 영역 박스와 실측 박스 (ground truth box)가 얼마나 가깝게 위치하고 있는지를 판단하는 역할을 하고 신뢰 손실의 경우 검출된 물체의 클래스가 다양한 클래스 중에서 얼마나 정확하게 자신의 클래스를 판단하는지를 측정한다.

III. 제안 방법

본 장에서는 기존 컬러 영상만을 이용한 보행자 검출의 문제점을 보완하기 위해 열 영상을 함께 사용한 보행자 검출 기법을 제안한다. 기본 네트워크 구조로는 SSD 기법을

사용하며 VGG-16을 기본 네트워크로 사용한다. 네트워크 퓨전 방식으로는 SSD-E, SSD-H, SSD-L을 적용해 어떤 구조가 최적의 구조인지를 파악한다. 그림 3에서는 각각 제안한 네트워크의 구조를 보여주고 있다. 그림 내의 ConvN_M 표현은 N번째 convolutional layer 블록의 M번째 layer를 의미한다. 기존 [7]의 논문에서는 Faster R-CNN 기반의 보행자검출 기법으로 물체가 위치한 영역을 예측하는 네트워크가 따로 존재하기 때문에 속도가 느린 단점이 존재한다. 제안하는 방식의 경우 SSD를 기반으로 하여 다양한 스케일과 중첩비를 가진 기본영역박스를 이용해 빠르고 정확하게 물체를 탐지하는 차이가 존재한다. SSD-E의 경우 먼저 컬러 영상을 입력으로 받는 네트워크와 열 영상을 입력으로 받는 네트워크를 구성한다. Conv1_2 이후 Concat.가 위치하게 되어 특징맵이 합쳐지게 되고 이후 합쳐진 특징맵은 NIN을 거치며 크기가 조절된다. 특징맵이 합쳐지게 되는 방식은 간단한 Concatenate 작업으로 예를 들어 $n \times n \times m$ 크기의 특징맵 두 개가 합쳐질 경우 $n \times n \times 2m$ 되는 형태로 특징맵을 쌓는 작업을 의미한다. 이렇게 합쳐진 특징맵의 경우 기존의 VGG-16에서 생성되는 특징맵의 크기와 달라 미리 학습된 계수를 사용할 수 없는데 NIN을 이용하면 특징맵의 크기를 다시 $n \times n \times m$ 로 줄일 수 있어 VGG-16의 미리 학습된 계수를 사용할 수 있다. 마찬가지로 SSD-H의 경우 Conv4_3 이후 특징맵이 합쳐지게 되고 SSD-L의 경우 마지막 Conv11_2 이후 특징맵이 합쳐지게

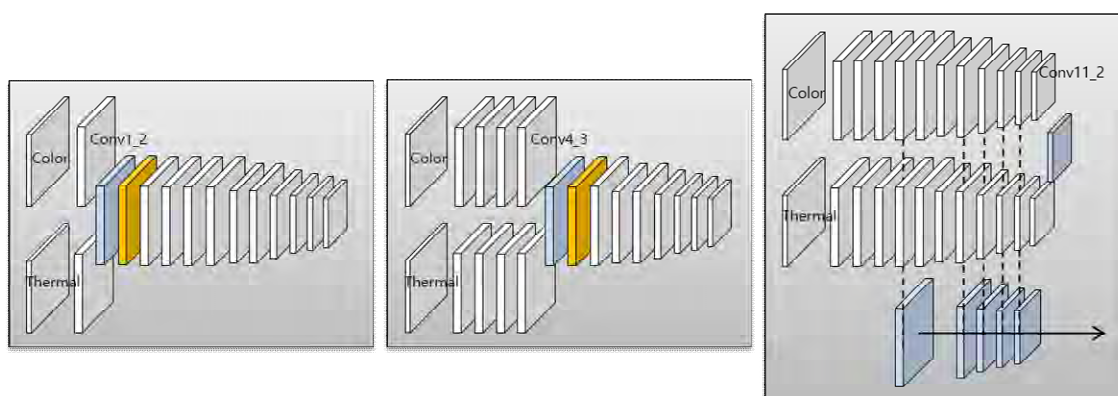


그림 3. 제안하는 SSD-E (왼쪽), SSD-H (가운데), SSD-L (오른쪽) 네트워크 구조: 입력영상을 제외한 흰색, 파란색, 노란색 박스는 각각 convolution layer, concatenate layer, NIN를 의미함

Fig. 3. Proposed SSD-E (left), SSD-H (middle), SSD-L (right) network structure: Except for the input images, white, blue, and yellow boxes indicate convolution layers, concatenate layers, and NIN respectively

된다. 기존의 SSD는 다양한 스케일의 특징맵을 사용해 물체를 탐지하는 특성 때문에 추가 특징 layer들이 존재하고 이 정보가 나중에 합쳐지게 된다. 그림 2에서 보면 총 6개의 다양한 스케일의 특징맵들이 합쳐지는 것을 확인할 수 있다. 이 때문에 SSD-L에서는 이렇게 연결되는 특징맵들을 합쳐서 마지막 layer에 전달하도록 하였다.

IV. 실험 결과

실험은 KAIST의 multispectral pedestrian dataset^[9]을 이용해 실시하였다. KAIST의 데이터셋은 총 95,328개의 컬러 영상, 열 영상 쌍이 존재하며 총 1,182명의 보행자가 103,128개의 주석으로 표기되어 있다. 트레이닝은 보행자가 가려지거나 잘려있는 경우 (occlusion), 높이가 50 픽셀 미만의 작은 보행자를 제외한 영상을 사용하였다. 또한 매 두 번째 프레임을 추출하여 총 3,357쌍의 영상을 이용해 트레이닝을 실시하였다. 테스트셋의 경우 위와 동일한 기준을 적용하고 매 세 번째 프레임을 추출하여 총 2,094쌍의 영상을 이용하였다. 트레이닝 시 사용된 파라미터의 경우

[8]에서 사용된 것과 동일하게 설정하였으며 트레이닝은 pre-training 없이 총 120,000번 반복되었다. 결과 지표로는 FPPI-miss rate를 사용하였다. 총 일곱 가지의 방법을 비교하였는데 각각 ACF+T+THOG^[9], SSD-C, SSD-T, halfway fusion^[7], SSD-E, SSD-H, SSD-L이다. ACF+T+THOG는 [9]의 데이터셋의 보행자 검출 기준의 기준치 방법으로 기본 ACF^[13]의 10개의 채널을 가진 특징에 열영상의 특징과 열영상에서 획득한 HOG (histogram of gradients)^[14] 특징을 합쳐서 사용하는 물체탐지 기법이다. SSD-C와 SSD-T는 컬러 영상과 열 영상만을 각각 이용해 트레이닝한 방식이다. SSD-E, SSD-H, SSD-L은 각각 SSD를 이용한 early fusion, halfway fusion, late fusion을 의미한다. 그림 4은 각 방법을 사용했을 때의 FPPI-miss rate 결과를 나타내고 있다. Miss rate의 범위는 [12]에서 제안된 것과 같이 $[10^{-2}, 10^0]$ 의 log 범위에서 표시하였고 기준점인 10^{-1} 에서 평균값을 표기하였다.

제안 방법인 SSD-E, SSD-H, SSD-L 모두 기존 KAIST 기준치 기법인 ACF+T+THOG에 비해 높은 결과를 획득하였다. 그 중 SSD-H가 가장 좋은 성능을 보여주었으며 18.18% 낮은 miss rate를 획득하였다. SSD-E와 SSD-L 모

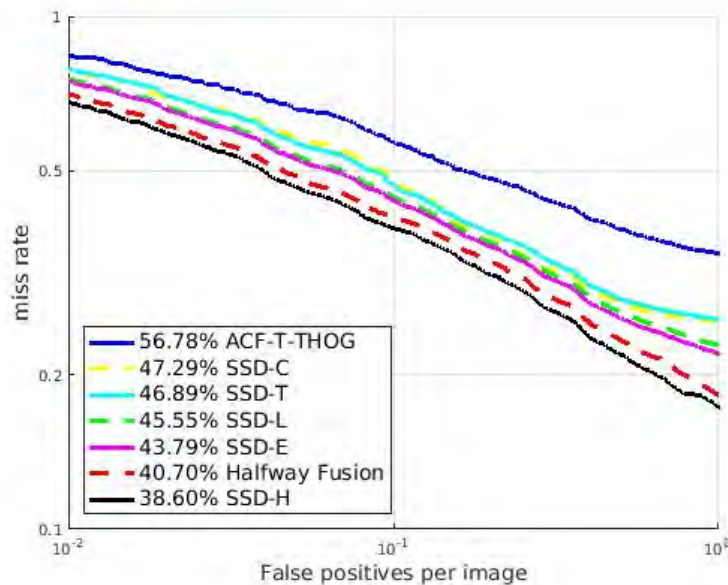


그림 4. FPPI-miss rate 결과 비교

Fig. 4. Comparison of detection results in FPPI-miss rate

두 단일 영상으로만 트레이닝했던 SSD-C와 SSD-T보다 성능이 높게 나왔지만 [7]의 *halfway fusion*보다는 좋지 않은 성능을 보여주었다. 그러나 SSD-H의 경우 2.1% 낮은 *miss rate*를 획득해 SSD가 Faster R-CNN에 비해 정확도가 높은 방법인 것을 확인할 수 있었다. SSD-C와 SSD-T의 경우 *miss rate*의 결과 차이는 미미한 수준이었다. SSD-H 방식이 가장 좋은 결과를 보이는데 대한 분석은 다음과 같다. SSD-H는 중간 단계의 *layer*에서 특징들을 퓨전했기 때문에 특징맵이 의미론적인 정보를 가지는 동시에 세밀한 디테일을 보존할 수 있다. SSD-E의 경우 픽셀 단위에 가까운 특징을 퓨전하기 때문에 네트워크 퓨전의 이점이 적을 수 밖에 없으며 SSD-L의 경우 너무 의미론적인 (*semantic*) 특징을 퓨전하기 때문에 의미론적인 특징에서 생기는 부정확한 데이터에 전체 성능이 크게 영향을 받을 수 밖에 없다. 또한 기존의 [7]의 경우 Faster R-CNN을 기반한 방식으로 실제 정확도의 경우 같은 VGG-16을 기반으로 한 SSD 보

다 낮다. SSD의 경우 Faster R-CNN 보다 속도가 빠른 것을 차치하고서라도 최종 물체를 탐지하는 판단을 하는 부분에 있어서 다양한 스케일의 특징맵을 함께 사용하기 때문에 그 정확도가 향상되고 따라서 SSD-H의 성능이 다른 비교 방식에 비해 가장 성능이 좋다.

제안한 네트워크 퓨전 구조 (SSD-H)가 컬러 영상만을 이용했을 때 (SSD-C)보다 성능이 향상되는 것을 확인하기 위해 실제 보행자 탐지 영상을 확인해 조명 환경이 좋은 낮 영상과 조명이 거의 없는 밤 영상에 테스트를 해서 결과를 비교하였다. 그림 5는 SSD-C와 SSD-H의 보행자 검출 결과를 보여준다. 그림 5의 처음 두 줄은 SSD-C의 결과이고 아래의 두 줄은 SSD-H의 결과이다. 각 결과는 컬러 영상과 열 영상을 함께 나타내었다. 다양한 영상에서 SSD-H가 SSD-C보다 정확하게 보행자를 탐지하는 것을 확인할 수 있다. 그림 5. (a)와 (c)의 낮 영상의 경우 SSD의 이점과 열 영상에서 제공되는 추가적인 정보로 인해 더 많은 보행자



그림 5. SSD-C (처음 두 줄)과 SSD-H (아래 두 줄)의 결과 비교

Fig. 5. Results from SSD-C (First two row) and SSD-H (Bottom two row)

검출이 가능하다. 또한 밤 영상의 경우 보행자가 컬러영상에서는 정확히 나타나지 않기 때문에 SSD-C의 결과에서 잘못된 보행자를 검출하거나 (그림 5. (b)와 (e)) 보행자의 위치를 검출하는 영역박스의 크기가 덜 정확한 경우 (그림 5. (d))가 발생한다. 그러나 SSD-H의 결과에서는 열 영상에서 보행자의 위치를 쉽게 판별할 수 있기 때문에 보행자의 위치를 정확히 검출하는 동시에 영역박스의 크기가 더 정확하게 예측되는 것을 확인할 수 있다.

V. 결 론

본 논문에서는 컬러 영상과 열 영상을 이용한 다중 스펙트럼 보행자 검출 기법에 대해 제안하였다. 기존의 보행자 검출 기법의 경우 컬러 영상만을 사용하기 때문에 보행자 검출의 정확도가 떨어지는 단점이 존재하지만 제안 방법에서는 컬러 영상과 열 영상의 정보를 같이 이용하기 때문에 낮과 밤 영상 모두 보행자를 정확하게 검출하는 것을 확인할 수 있었다. FPPI-miss rate 지표에서는 SSD-H가 ACF+T+THOG에 비해 18.18% 낮은 miss rate를 획득하였고 halfway fusion 대비 2.1% 낮은 miss rate를 획득해 높은 정확도를 확인하였다.

후속 연구로는 네트워크 퓨전 방법에 대해 좀 더 세밀하게 연구하고 보행자 검출 정확도를 더욱 향상하는 방향으로 진행하고자 한다. 이를 위해 기본 네트워크 구조를 새롭게 제안하고 새로운 네트워크 퓨전 방식을 제시하고자 한다.

참 고 문 헌 (References)

[1] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, "How

- far are we from solving pedestrian detection?," *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp.1259-1267, 2016.
- [2] J. Wagner, V. Fischer, M. Herman, and S. Behnke, "Multispectral pedestrian detection using deep fusion convolutional neural networks," *European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. 509-514, 2016.
- [3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, pp. 580-587, 2014.
- [4] R. Girshick, "Fast r-cnn," *arXiv preprint arXiv:1504.08083*, 2015.
- [5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Neural Information Processing Systems*, Montreal, Canada, pp. 91-99, 2015.
- [6] K. He, G. Gkioxari, P. Dollar, R. Girshick, "Mask R-CNN," *IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2980-2988, 2017.
- [7] J. Liu, S. Zhang, S. Wang, D. N. Metaxas, "Multispectral deep neural networks for pedestrian detection," *arXiv preprint arXiv:1611.02644*, 2016.
- [8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. -Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," *European Conference on Computer Vision*, Amsterdam, the Netherlands, pp. 21-37, 2016.
- [9] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, "Multispectral pedestrian detection: benchmark dataset and baseline," *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, pp. 1037-1045, 2015.
- [10] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [11] M. Lin, Q. Chen, S. Yan, "Network in Network," *arXiv preprint arXiv:1312.4400*, 2013.
- [12] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, pp. 304-311, 2009.
- [13] P. Dollar, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 8, pp. 1532-1545, Jan. 2014.
- [14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, USA, pp. 886-893, 2005.

저 자 소 개



이 용 우

- 2013년 2월 : 성균관대학교 전자전기공학부 학사
- 2013년 2월 ~ 현재 : 성균관대학교 전자전기컴퓨터공학과 석박통합과정
- ORCID : <https://orcid.org/0000-0002-2873-7122>
- 주관심분야 : 영상처리, 컴퓨터비전, 머신러닝, 딥러닝, 객체인식



신 지 태

- 1986년 : 서울대학교 전기공학과 학사
- 1988년 : 한국과학기술원(KAIST) 원자력공학과 석사
- 1988년 ~ 1991년 : 한국전력공사 전문원
- 1991년 ~ 1996년 : 한국원자력연구소 선임연구원
- 1996년 ~ 2001년 : University of Southern California 전자공학과 공학박사
- 2001년 ~ 2002년 : 경희대학교 교수
- 2002년 ~ 현재 : 성균관대학교 전자전기공학부 교수
- ORCID : <https://orcid.org/0000-0002-2599-3331>
- 주관심분야 : 영상/비디오 신호처리, 무선/이동 통신시스템상 비디오 전송, 딥러닝 응용