멀티모달 센서를 활용한 객체 검출 프레임워크

Object detection framework using multimodal sensors

이호준 [1] · 정인섭 [2] · 곽노준 †

Hojun Lee[1], Inseop Chung[2] Nojun Kwak †

Abstract: Object detection methods using RGB images have been studied a lot. However, there may be some problems when there are limitations in illumination, such as at night, or when there are limitations in the field of view due to smoke or fog, etc. In order to overcome these limitations, we propose a framework of using multimodal sensor images that contain a variety of information that RGB does not have. To use the multimodal sensor image, we design a Multimodal Attention Network that focuses on the regions useful for object detection in multimodal sensor images. Through qualitative and quantitative experiments, we show the effectiveness of using multimodal sensor images and the performance can be improved by using the Multimodal Attention Network.

Keywords: Deep Learning, Object detection, Multimodal sensor image

1. 서 론

Convolutional Neural Network (CNN) [1] 이 발전하며 RGB 영상[2] 기반의 객체 검출 기술[3, 4] 도 발전해왔다. RGB 영상 기반 객체 검출 알고리즘은 충분한 조도, 시야 장애가 없는 상황에서는 잘 동작할 수 있다.

그러나 빛이 없는 야간이나 안개, 화재로 인한 연기 등으로 인해 시야가 제한된 상황에서는 RGB 영상에서 객체 검출 알고리즘에 제한이 생길 수 있다. [Fig. 1] 의 실내 화재 시뮬레이션 예시에서 볼 수 있듯이, RGB 영상에서는 시야가 제한될 수 있다. 이런 경우에 다른 센서 데이터를 함께 이용한다면 RGB 영상만 이용했을 때보다 더 효과적으로 객체를 검출할 수 있다.

본 연구에서는 RGB 뿐만 아니라, Nightvision, Infrared, Thermal 등의 멀티모달 센서 데이터를 함께 이용하는 객체 검출 딥러닝 프레임워크를 제안한다. 멀티모달 센서 영상에서 객체 검출에 도움이 되는 영역을 집중하여 효과적으로 객체 검출할 수 있도록 돕는 세부 네트워크를 소개한다. 또한, 프레임워크의 정성적/정량적인 실험 결과를 제공한다. 정성적 실험의 경우 자체 수집 데이터에서, 정량적 실험은 MPD 데이터[5] 를 진행했다.

[Fig. 1] Fire situation simulation. The left column shows RGB images, and the right column shows thermal images and the object detection results using the thermal images. Unlike in RGB images, where the field of view is limited due to the smoke, thus humans are not visible, in thermal images, the shape of humans are relatively well revealed.

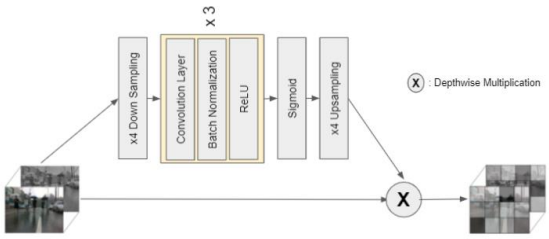
2. Framework

[Fig. 2]는 전체 구조이고, 전체 구조의 Multimodal Attention CNN에 대한 세부 구조는 [Fig. 3] 와 같다. Multimodal Attention CNN 은 멀티모달 센서 영상 데이터에서 객체 검출에 도움이 되는 영역의 값을 강화, 그렇지 않은 영역의 값을 약화한다. 이 과정으로 Attention이 된 영상이 생성되고 이 영상이 객체 검출기에 입력되면 객체 검출기가 bounding box를 추론한다.

※ This work was supported by ICT Research and Development Program of MSIP/IITP, Korean Government (2017-0-00306)

1. Ph.D. student, Seoul National University, Seoul, Korea (hojun815@snu.ac.kr)
2. Ph.D. student, Seoul National University, Seoul, Korea (jis3613@snu.ac.kr)
† Professor, Corresponding author: Seoul National University, Seoul, Korea (nojunk@snu.ac.kr)

[Fig. 2] Schematic of the multimodal object detector framework. The multimodal sensor images can be configured in various ways such as RGB, thermal image and foreground separation result. Yolov4 is used as the object detector.



[Fig. 3] Detailed schematic of the multimodal attention network. Using this structure, it is possible to focus on the regions useful for object detection in the multimodal sensor image data.

3. 실험 결과

3.1 정성적 실험 결과

[Fig. 4] 는 자체 수집된 멀티모달 영상 데이터에서의 검출 결과다. 데이터는 RGB, Depth, Infrared, Nightvision, Thermal, Lidar로 구성돼 있고, RGB와 Nightvision 두 영상에서 실험했다. 낮은 조도로 인해 검은 물체를 식별하기 어려워졌으나 Nightvision을 같이 썼을 경우 RGB에서 검출되지 않았던 객체를 검출할 수 있었다.

[Fig. 4] Qualitative results. The left column shows the detection results using RGB images, and the right column shows the results using RGB and Nightvision together. Note that using only RGB at night leads to missing objects (yellow rectangles).

3.2 정량적 실험 결과

Multimodal image 와 Multimodal Attention CNN의 유무에 따른 정량 비교 실험을 진행했다. 정량적 실험은 MPD 데이터에서 자체 평가로 진행했다. MPD 데이터는 주/야간의 RGB와 Infrared영상으로 구성돼 있고, person, people, cyclist의 class와 bounding box 로 annotation 되어 있다. [Table 1]에서 볼 수 있듯이 멀티모달 Attention CNN을 사용했을 때 성능향상을 확인할 수 있었다.

[Table 1] Quantitative results in MPD dataset. Unit : Average Precision (AP)

	Day	Night
RGB Unimodal	61	45
Multimodal Without Multimodal Attention network	63	62
Multimodal With Multimodal Attention network	65	64

4. 결론

본 연구는 멀티모달 센서 영상을 활용한 객체 검출 프레임워크를 제안한다. Multimodal Attention CNN으로 멀티모달 영상을 Attention하여 성능 향상이 가능하다는 것을 보였다. 본 연구의 프레임워크로 시야가 제한된 환경에서 객체 검출을 더 효과적으로 할 수 있다.

References

[1] Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton “ImageNet classification with deep convolutional neural networks,” NIPS12: Proceedings of the 25th International Conference on Neural Information Processing Systems, Nevada, USA, pp. 1097–1105, 2012.

[2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence Zitnick. “Microsoft coco: Common objects in context,” In European conference on computer vision, pp. 740–755, 2014.

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. “Faster R-CNN: towards real-time object detection with region proposal networks,” In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'15), Cambridge, USA, pp. 91–99, 2015.

[4] Bochkovskiy Alexey, Wang Chien-Yao, Liao Hong-yuan, “YOLOv4: Optimal Speed and Accuracy of Object Detection”, arXiv preprint arXiv:2004.10934, 2020

[5] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi and In So Kweon, “Multispectral Pedestrian Detection: Benchmark Dataset and Baseline”, CVPR 2015 : 28th IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, pp. 1037-1045, 2015.