

드론 영상에서 재난 상황인지를 위한 딥러닝 기반 다중 객체 추적 시스템

김찬란 송제인 이재훈

줌인터넷 부설연구소

chanranhari@zuminternet.com

Multi-object Tracking System for Disaster Context-aware using Deep Learning

Chanran Kim Jein Song Jaehoon Lee

Zuminternet Research Institute

요약

고위험의 재난 상황에서 사람이 상황을 판단하고, 요구조자를 탐색하며, 구조하는 것은 추가 피해를 발생시킬 수 있다. 따라서 재난 상황에서도 이동과 접근이 용이한 무인항공에 관한 연구와 개발이 활발히 이루어지고 있다. 재난 상황에서 신속하게 대처하기 위해서는 선제적 상황인지 기술이 필요하다. 이에 본 논문은 구조 및 대피를 위해 사람, 자동차, 자전거 등의 객체를 인식하고 중복 인식을 피하기 위해 추적하는 딥러닝 기반 다중 객체 추적 시스템을 제안한다. 2019 인공지능 R&D 그랜드 챌린지 상황인지 부문에서의 대회 결과로 실험 성능을 증명한다.

1. 서론

지진, 화재 등 대형복합재난으로 인한 피해를 최소화하기 위해서 인공지능과 로봇틱스를 이용한 기술이 연구 개발되고 있다[1, 2]. 무인항공기는 재난 상황에서 사람의 접근이 어려운 환경 속에서 요구조자를 탐지하고 추적하여 상황을 파악하는데 활용된다. 재난 상황에 더불어 2차적으로 추가 발생할 수 있는 구조 사고 피해를 감소시킴으로써 무인항공기는 효율적인 재난 구조에 도움을 줄 수 있다. 하지만 수많은 무인항공기가 촬영한 영상을 일일이 확인하며 재난 상황을 판단하는 것은 매우 어려운 일이다. 이에 본 논문은 재난 상황 인지를 위한 인공지능 시스템을 제안한다.

제안하는 시스템이 해결하는 문제를 크게 3가지로 나누어 정의할 수 있다. 첫 번째는 높은 자유도이다. 드론을 이용한 영상 촬영에서는 카메라의 이동과 회전이 동시에 가능하며, 추적하고자 하는 대상도 동시에 움직인다. 따라서 카메라와 객체 관계에서 움직임의 자유도가 상당히 높다. 두 번째는 다양한 위치와 각도에서 촬영한 영상이기 때문에 동일한 객체라고 하더라도 크기와 형태가 다양하다는 점이다. 세 번째는 다중 객체를 교차, 폐색, 손실 상황에도 강인하게 추적해야 한다는 것이다.

2. 관련 연구

딥러닝 기반의 객체 인식 기술은 크게 one-stage 구성과 two-stage 구성으로 나뉜다. Two-stage 구성은 객체가 될 수 있는 후보를 제안하는 과정과 분류하는 과정을 구분하여 시스템을 구성한 것이고, one-stage 구성은 그 두 과정을 동시에 연산하도록 구성한 것이다. 가장 대표적인 two-stage 객체 인식 기술은 R-CNN[3]으로, 빠른 연산을 위해 Fast R-CNN[4], Faster R-CNN[5] 등의 기술들이 추가적으로 제안되었다. one-stage 객체 인식 기술은 SSD[6], YOLO[7], RetinaNet[8] 등이 있으며, 상대적으로 정확도는 낮지만 연산 속도는 빠르다. 신속한 대응이 필요한 상황에는 one-stage 객체 인식 기술이 적

합하다.

드론 영상에서의 움직이는 객체의 이동 거리, 속도 측정 기술들은 객체를 추적하는데 활용될 수 있다[9]. 특징점 기반의 알고리즘을 통해 영상간의 변환 관계를 호모그래피 행렬로 계산할 수 있으며, 이를 통해 카메라의 움직임을 계산한다. 대표적인 움직임 기반 기술로는 SORT[10] 추적기가 있다. 움직임 기반 기술과 다르게 딥러닝 기반의 객체 추적 기술은 순차적 정보를 활용하기 위해서 재귀적인 구조를 활용한다. 대표적으로 RNN 계열 신경망 구조인 LSTM을 활용한 Re3[11] 추적기가 있다.

3. 제안하는 시스템

제안하는 시스템의 아키텍처는 그림 1의 구성으로 이루어진다. 크게 객체 검출과 객체 추적 모듈로 나뉜다. 드론으로 촬영한 동영상 프레임으로부터 딥러닝 기반의 객체 검출 기술로 지정된 종류의 객체들을 검출하고, 카메라의 움직임을 파악하여 위치 정보를 보완하며 객체를 추적한다. 추적하는 객체들의 프레임간 이동 속도를 계산하고, RNN 기반의 모델을 활용하여 연속된 프레임으로부터 객체의 형태 정보를 파악하여 영상 속에서 객체가 폐색되거나 교차하여 손실된 후에 재등장하였을 경우 예측 및 매칭 과정을 통해 객체를 재인식하는 기술을 활용하였다. 최종적으로 물체를 추적 및 식별하여, 요구조자의 수와 활용할 수 있는 이동수단 등의 재난 상황을 파악할 수 있다.

객체 검출은 CNN 기반의 딥러닝 모델인 RetinaNet을 활용하였다. 기존 모델에서 학습되지 않은 작은 크기의 객체에 대해서도 강인할 수 있도록 입력 영상을 분할하여 각각에 대하여 객체를 검출하였다. 각 분할 영상에서의 검출 결과를 NMS(Non-Maximum Suppression)으로 통합하여 최종 검출 결과를 도출하였다. RetinaNet은 ResNet-50+FPN을 백본(Back Bone) 모델로 사용하며, focal loss를 활용하여 one-stage 방법임에도 불구하고 높은 정확도를 보인다. RetinaNet에

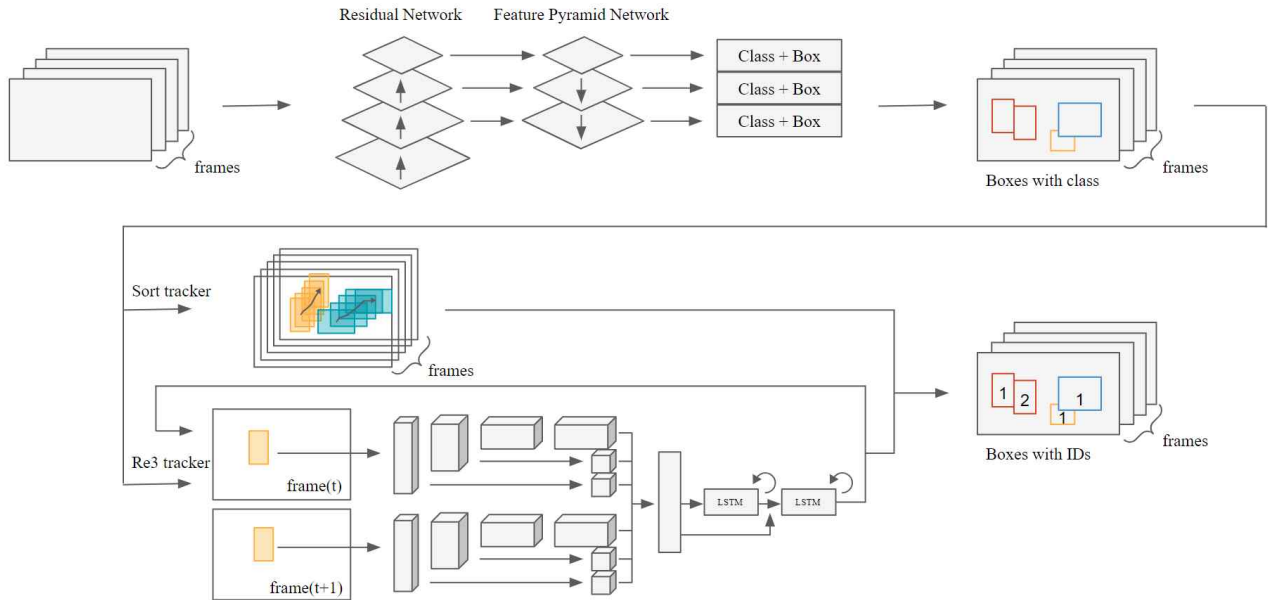


그림 1. 제안하는 다중 객체 검출 및 추적 시스템 흐름도

서는 다양한 크기와 비율의 이동 불변 앵커 박스와 multi-scale feature pyramids[12]를 사용한다.

다중 객체 추적 알고리즘은 기본적으로 이전 영상 프레임에서 검출한 객체 정보와 현재 영상 프레임에서 검출한 객체 정보의 겹치는 정도를 계산하는 IoU(Intersection of Union)를 활용한다. 먼저 카메라의 움직임을 고려한 영상 좌표계를 설정하기 위하여 배경 요소들을 이용한다. 배경 요소들에서 ORB(Oriented FAST and Rotated BRIEF)[13] 방식으로 특징을 추출 및 기술하고, FLANN(Fast Library for Approximate Nearest Neighbors)[14] 알고리즘으로 매칭하여 프레임간 변환에 대한 호모그래피 행렬을 구한다. 배경의 움직임을 영상의 움직임으로 근사하고, 이를 이용해 프레임 간의 상대적인 위치를 보정한다. 보정한 영상을 이용하여 겹침의 정도를 판단한다면 카메라의 상대적인 움직임에도 동일한 좌표계에서 이동 경로를 계산할 수 있다. 따라서 객체의 폐색, 교차에 강한 추적을 할 수 있다.

본 논문에서 제안하는 다중 객체 추적은 두 알고리즘을 기반으로 하고 있다. 첫 번째 다중 객체 추적 알고리즘은 움직임 기반의 SORT(Simple Online and Realtime Tracking) 방법이다. SORT는 기초적인 데이터 관계 및 상태 추정 기술에 기반한 시각적 다중 객체 추적 기술이다. 기존의 SORT 추적 방법에서는 Faster R-CNN으로 객체를 인식하고 추적하였지만, 본 논문에서 제안하는 시스템에서는 RetinaNet으로 대체하였다. 따라서 SORT 추적기의 칼만필터와 헝가리안 알고리즘을 활용하는 추적 부분만을 활용한다. 두 번째 다중 객체 추적 알고리즘은 딥러닝 기반의 Re3(Real-time Recurrent Regression) 방법이다. 객체의 모양, 동작 및 시간에 따른 변화에도 강한 객체 추적 기술이다. 150 fps(frame per second)으로 실시간 처리가 가능하면서 높은 정확도를 보인다. 본 논문에서 제안하는 시스템은 수정한 SORT와 Re3 방법을 앙상블하여 최종적인 결과를 도출한다.

4. 실험

최소 5m 이상의 높이에서 -30~30도 사이의 다양한 각도로 촬영한 FHD급의 30초 길이 드론 영상(10 fps) 500개에 대해서 실험을 수행하였다. 2019 인공지능 R&D 그랜드 챌린지에서 총 연산시간은 5시간으로 제한되었다. 제한 시간 내에 모든 연산을 하기 위하여 프레임 감산을 하였지만, 이는 추적의 성능을 높이기 위해서도 사용되는 방법이다. Keras와 TensorFlow 프레임워크를 사용하여 개발하였으며, 모듈 간의 대기 시간을 낭비하지 않기 위해 병렬적으로 시스템을 구성하였다. Geforce RTX 2080 Ti 11GB 2개 사용하여 GPU 연산을 수행하였다. 자유도가 높은 드론 영상으로부터 단축기준 32픽셀 이상의 관심 객체를 식별하고 계수한다. 해당하는 관심 객체는 사람, 소화기, 소화전, 차량, 자전거, 오토바이이다. 등장하는 계수 대상의 추정한 등장 횟수와 정답 사이의 가중 유클리디안 거리로 성능을 평가한다.

복잡재난상황에서는 요구조자의 수를 파악하는 것이 가장 중요하므로, 이에 대해 상대적으로 높은 가중치, 5로 계산된다. 나머지 대상에 대한 가중치는 1로 주어진다. 이러한 상대적 가중치가 적용된 유클리디안 거리는 식(1)로 나타낼 수 있다. 주어진 데이터셋에 대한 실험결과를 식(1)로 계산한 결과를 표 1에 나타내었다. 표 1은 상위 3팀, 5팀, 10팀, 20팀의 점수 평균과 제안한 방법의 점수를 나타낸다. 해당 결과는 2019 인공지능 R&D 그랜드 챌린지 상황인지 부문에서의 공식 결과로 본 시스템이 제안한 방법으로 전체 참가팀 41팀 중에 2등의 성적을 달성했다.

$$d(\vec{p}_1, \vec{q}_1) = \sqrt{w_1(p_1 - q_1)^2 + w_2(p_2 - q_2)^2 + \dots + w_n(p_n - q_n)^2} \quad (1)$$

표 1. 상위 n팀의 점수 평균과 제안한 시스템의 점수

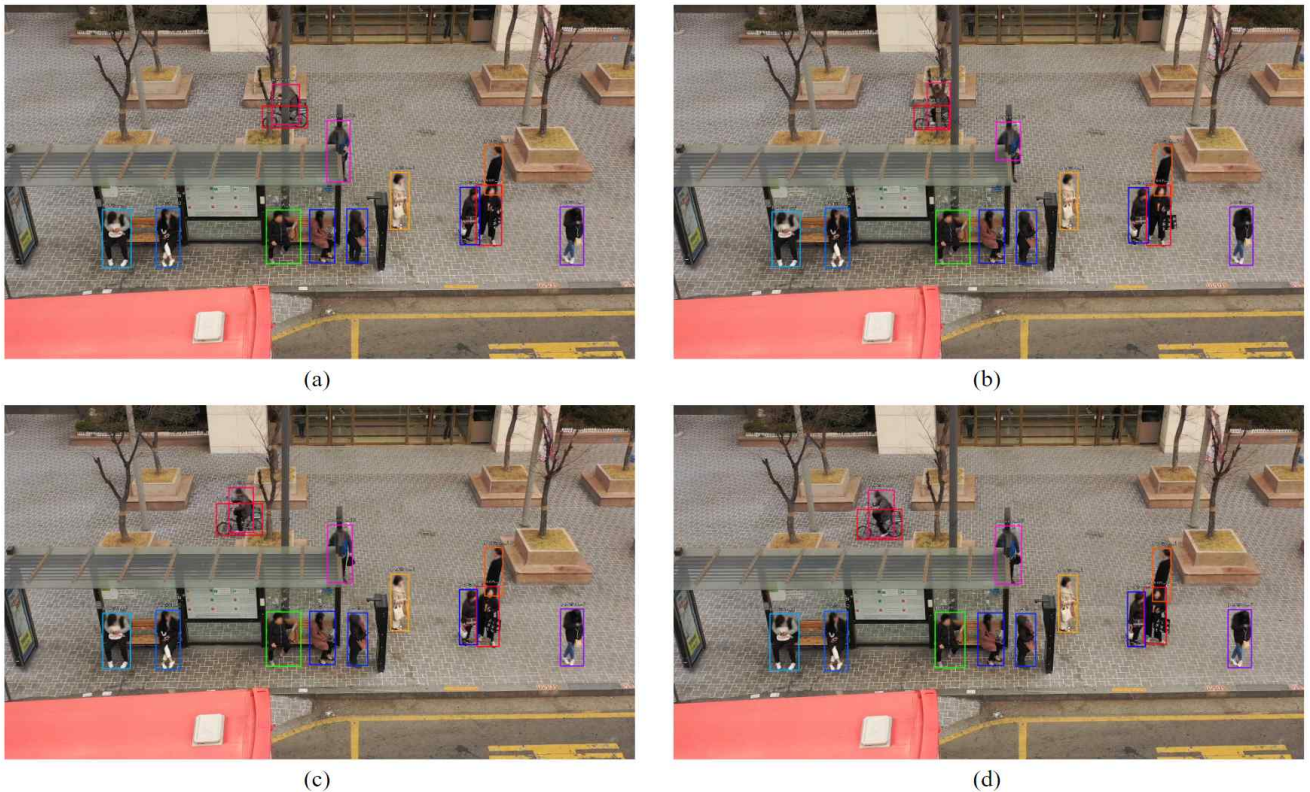


그림 2. 제안하는 다중 객체 검출 및 추적 시스템의 추적 결과

| Team | Score |
|----------------|-------------|
| Ours (top-2) | 9.906489317 |
| top-3 average | 9.862762437 |
| top-5 average | 10.4932817 |
| top-10 average | 12.39941792 |
| top-20 average | 15.94229937 |

5. 결론

본 논문은 2019 인공지능 R&D 그랜드 챌린지 상황인지 부문 참가 과정에서 구성된 드론 영상에서의 딥러닝 기반 다중 객체 추적 시스템을 제안한다. 드론 영상을 분석하기에 적합한 one-stage 구성의 객체인식 기술과 움직임 기반 객체추적 기술 및 딥러닝 기반 객체 추적 기술을 앙상블 하여 활용하였다. 학습데이터가 적은 상황이기에 데이터 증강, 준 지도학습 등의 방법을 활용하면 성능이 향상될 것으로 기대된다.

감사의 글

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.IITP20190017620021001, 드론 영상에서의 다중 객체 추적을 위한 딥러닝 기반 영상 처리 기술 개발)

참고문헌

- [1] Restas, Agoston. "Drone applications for supporting disaster management." *World Journal of Engineering and Technology* 3, no. 03 (2015): 316.
- [2] Hassanalian, Mostafa, and Abdessattar Abdelkefi. "Classifications, applications, and design challenges of drones: A review." *Progress in Aerospace Sciences* 91 (2017): 99-131.
- [3] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580-587. 2014.
- [4] Girshick, Ross. "Fast r-cnn." In *Proceedings of the IEEE international conference on computer vision*, pp. 1440-1448. 2015.
- [5] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." In *Advances in Neural Information Processing Systems(NIPS)*, pp. 91-99. 2015.
- [6] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In *European conference on computer*

vision, pp. 21-37. Springer, Cham, 2016.

[7] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788. 2016.

[8] Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal loss for dense object detection." In Proceedings of the IEEE International Conference on Computer Vision(ICCV), pp. 2980-2988. 2017.

[9] Risse, Benjamin, Michael Mangan, Luca Del Pero, and Barbara Webb. "Visual tracking of small animals in cluttered natural environments using a freely moving camera." In Proceedings of the IEEE International Conference on Computer Vision(ICCV) Workshops, pp. 2840-2849. 2017.

[10] Bewley, Alex, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. "Simple online and realtime tracking." In 2016 IEEE International Conference on Image Processing (ICIP), pp. 3464-3468. IEEE, 2016.

[11] Gordon, Daniel, Ali Farhadi, and Dieter Fox. "Re3:Real-TimeRecurrentRegressionNetworksforVisualTrackingofGenericObjects."IEEE Robotics and Automation Letters(RA-L) 3, no. 2 (2018): 788-795.

[12] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature pyramid networks for object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR), pp. 2117-2125. 2017.

[13] Rublee, Ethan, Vincent Rabaud, Kurt Konolige, and Gary Bradski. "ORB: An efficient alternative to SIFT or SURF." In 2011 International conference on computer vision, pp. 2564-2571. Ieee, 2011.

[14] M. Muja and D. G. Lowe. "FLANN - Fast Library for Approximate Nearest Neighbors." <http://people.cs.ubc.ca/~mariusm/index.php/FLANN/FLANN>.