DeepLabV3+와 단안 깊이 추정(Monocular Depth Estimation)을 결합한 위장 객체 탐지

조규성, 권현
육군사관학교 AI·데이터과학과
c21108@kma.ac.kr, hkwon.cs@gmail.com

Camouflaged Object Detection via Joint DeepLabV3+ and Monocular Depth Estimation

Jo GyuSeong, Hyun Kwon
Dept. of Artificial Intelligence and Data Science, Korea Military Academy

요 약

전장 환경에서 적군의 위장(Camouflage)은 시각 기반 감시 및 탐지를 어렵게 만들어 아군의 생존성을 위협하는 주요 요인이다. 기존의 RGB 영상 기반 객체 분할 및 탐지 기법은 색상과 질감 정보에 크게 의존하기 때문에, 배경과 시각적으로 유사한 위장복을 착용한 인원에 대해 탐지 누락이나 경계 오류가 빈번히 발생한다. 본 연구에서는 이러한 한계를 극복하기 위하여 DeepLabV3+ 기반 의미론적 분할 모델과 단안 깊이 추정(Monocular Depth Estimation)을 결합한 위장 객체 탐지 기법을 제안한다. 단안 RGB 영상으로부터 추정한 깊이 정보를 활용함으로써 색상 정보만으로는 구분이 어려운 객체와 배경을 거리 차이를 통해 분리하고, 이를 분할 결과 보정에 활용한다. 실험 결과, 제안 기법은 위장 환경에서도 안정적인 객체 경계 복원을 가능하게 하며, 시각적 분석을 통해 기존 방법 대비 탐지 성능 향상을 확인하였다.

Ⅰ. 서론

현대 전장은 복잡한 지형과 다양한 자연 환경을 포함하며, 특히 산악이나 수풀 지역에서는 전투복 및 위장을 착용한 적의 시각적 탐지는 매우 어렵다. 이러한 위장 전술은 단순한 색상 은폐를 넘어 주변 환경의 질감과 패턴을 모방함으로써 기존 영상 기반 탐지 체계의 성능을 저하시킨다. 최근 딥러닝 기반 객체 탐지 및 의미론적 분할 기술이 발전하였음에도 불구하고, 대부분의 모델은 RGB 영상의 색상 및 텍스처 정보에 의존하기 때문에 위장 환경에서 근본적인 한계를 가진다.

한편, 깊이 정보는 객체와 배경 간의 물리적 거리 차이를 제공하는 중요한 단서로서, 색상 정보와 독립적인 특성을 가진다. 실제 전장 환경에서 라이다나 스테레오 카메라와 같은 능동 센서를 활용하기에는 비용 및 운용상의 제약이 존재하므로, 단일 RGB 영상으로부터 깊이를 추정하는 단안 깊이 추정 기법은 실용적인 대안이 될 수 있다. 이에 본 연구는 DeepLabV3+ 기반 분할 결과에 단안 깊이 정보를 결합하여 위장 객체 탐지 성능을 향상시키는 것을 목표로 한다.

Ⅱ. 관련 연구 (Related Work)

의미론적 분할 분야에서 DeepLabV3+는 Atrous Convolution과 ASPP 구조를 활용하여 다중 스케일 문맥 정보를 효과적으로 반영하는 대표적인 모델이다. Encoder-Decoder 구조를 통해 세밀한 경계 복원이 가능하다는 장점이 있으나, 입력 이미지의 색상 및 질감 정보에 대한 의존도가 높아 위장 환경에서는 성능 저하가 발생할 수 있다.

단안 깊이 추정은 단일 RGB 영상으로부터 장면의 상대적 깊이를 예측하는 기술로, 최근에는 대규모 데이터셋과 딥러닝 모델을 기반으로 상당히 안정적인 성능을 보이고 있다. 특히 MiDaS와 같은 사전학습 모델은 다양한 환경에 대해 일반화된 깊이 추정을 제공할 수 있어, 추가 센서 없이도 3차원적 단서를 확보할 수 있다. 본 연구에서는 이러한 단안 깊이 추정 기법을 위장 객체 탐지 문제에 적용하였다.

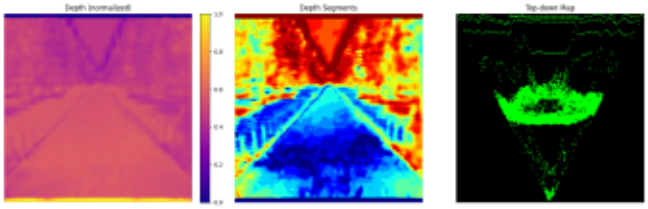

그림1) Monocular Depth Estimation 결과

Ⅲ. 연구 방법

본 프로젝트는 의미론적 분할 단계, 깊이 정보 생성 단계, 깊이 추정 모델 학습 단계, 그리고 결과 융합 및 시각화 단계로 구성된다.

먼저 입력 RGB 영상에 대해 DeepLabV3+ 모델을 적용하여 인원 및 배경에 대한 픽셀 단위 분할을 수행한다. 해당 모델은 학습 데이터셋을 기반으로 직접 학습되었으며, 출력은 클래스 인덱스 형태의 분할 마스크로 생성된다.

다음으로 단안 깊이 정보 생성을 위해 사전학습된 MiDaS 모델을 Teacher로 사용하였다. 입력 영상에 대해 MiDaS를 적용하여 연속적인 깊이 맵을 생성하고, 이를 정

규화하여 학습 데이터로 저장한다. 이 깊이 맵은 이후 DepthNet 학습의 기준 값으로 활용된다.

DepthNet은 Teacher 모델이 생성한 깊이 정보를 모사하도록 학습되는 Student 모델로, 연속적인 깊이 값을 여러 구간으로 양자화한 뒤 분류 및 회귀를 동시에 수행하는 구조를 사용하였다. 이를 통해 단안 깊이 추정의 안정성을 확보하고, 경계 영역에서의 깊이 변화 정보를 효과적으로 학습하도록 설계하였다.

마지막으로 분할 결과와 깊이 추정 결과를 결합하여 시각화하였다. 깊이 정보는 탑다운(top-down) 형태의 거리 맵으로 변환되며, 이를 통해 인원과 배경 간의 공간적 분리가 직관적으로 확인된다. 또한 분할 결과 위에 깊이 기반 보조 정보를 중첩하여 위장 객체의 경계 복원 효과를 분석하였다.

Ⅳ. 실험 및 결과 (Experiments & Results)

실험은 수풀 및 자연 환경을 포함한 데이터셋을 기반으로 수행되었다. 의미론적 분할 성능 평가는 Pixel Accuracy와 mean Intersection over Union(mIoU)을 사용하였다. 실험 결과, 분할 모델은 Pixel Accuracy 0.9176, mIoU 0.7980의 성능을 보였다. 이는 기본적인 인원 분할 성능이 안정적으로 확보되었음을 의미한다.

지난 연구 결과, 위장객체를 학습하지 않은 DeepLabV3+ 모델 단독 적용 시 위장복과 배경이 유사한 영역에서 인원 신체 일부가 배경으로 오분류되는 현상이 관찰되었다. 반면, 깊이 정보를 함께 고려한 경우 색상 차이가 미미하더라도 거리 차이에 의해 인원 영역이 상대적으로 명확히 구분되는 것을 확인할 수 있었다. 특히 수풀과 인접한 인원 경계 부근에서 깊이 기반 단서가 분할 결과 보정에 효과적으로 작용하였다.

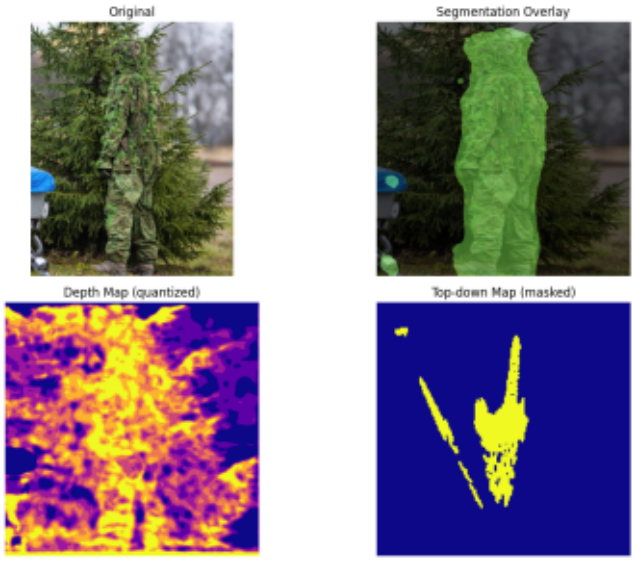

그림2) 위장 객체에 대한 DeepLabV3+모델과 Monocular Depth Estimation 적용 결과

Ⅴ. 결론 및 한계

본 연구에서는 DeepLabV3+ 기반 의미론적 분할과 단안 깊이 추정을 결합한 위장 객체 탐지 기법을 제안하였다. 색상 및 질감 정보에 의존하는 기존 분할 모델의 한계를 깊이 정보라는 보조 단서를 통해 보완함으로써, 위장 환경에서도 객체 경계 복원이 가능함을 확인하였다. 실험 결과와 시각적 분석을 통해 제안 방법의 유효성을 검증하였다. 향후 연구에서는 실시간 처리 성능 향상을 위한 모델 경량화와 함께, 야간 환경이나 악천후 조건을 고려한 열화상 영상 등 멀티모달 정보와의 융합을 통해 탐지 성능을 추가적으로 개선할 수 있을 것으로 기대된다.

참고문헌

[1]Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 801-818).

[2]Fan, D.-P., Ji, G.-P., Sun, G., Cheng, M.-M., Shen, J., & Shao, L. (2020). Camouflaged object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2777-2787).

[3]Godard, C., Mac Aodha, O., & Brostow, G. J. (2019). Digging into self-supervised monocular depth estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 3828-3838).

[4]Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2022). Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(3), 1623-1637. https://doi.org/10.1109/TPAMI.2021.3050991