

# 증강현실을 위한 객체인식 기술 성능 비교

신은지 · 신광성\*

원광대학교

## Object Recognition Technology Performance Comparison for Augmented Reality

Eun-ji Shin · Kwang-seong Shin\*

Wonkwang University

E-mail : dmswl3073@naver.com / waver0920@wku.ac.kr

### 요 약

증강현실의 핵심기술은 객체인식기술이다. 최근 CNN등 다양한 인공지능 알고리즘의 개발로 인하여 영상에서 특정 객체를 효과적으로 구분하는 것이 가능해졌다. 객체를 빠르고 정확하게 인식하는 기술이 확보되어야 더욱 현실감있고 몰입감있는 증강현실 콘텐츠의 구현이 가능해진다. 본 연구에서는 SSD(single shot multibox detector)를 이용한 객체인식 모델과 YOLO를 이용한 객체 인식 모델의 비교평가를 수행하였다.

### ABSTRACT

The core technology of augmented reality is object recognition technology. Recently, due to the development of various artificial intelligence algorithms such as CNN, it has become possible to effectively distinguish specific objects from images. It is possible to realize more realistic and immersive augmented reality contents only when technology for recognizing objects quickly and accurately is secured. In this study, an object recognition model using SSD (single shot multibox detector) and an object recognition model using YOLO were compared and evaluated.

### 키워드

YOLO, SSD, AR, 객체인식

### 1. 서 론

증강현실 기술의 핵심은 증강현실 콘텐츠의 제작기술 뿐만 아니라 증강현실 콘텐츠를 증강시키기 위한 대상객체를 얼마나 정확하고 빠르게 인식하고 인식한 객체에 최대한의 sync율로 정합하는 기술이라고 할수 있다.

기존에는 객체를 인식하기 위한 기술로 카메라의 깊이값을 이용하는 방식과 텍스처 인식 기반의 영상인식 기술을 이용하였다. 최근에는 기계학습 기반의 사물인식 기술을 이용하여 대상 객체의 인식율을 현저하게 증가시켰다.

기계학습은 영상처리 분야뿐만 아니라 많은 부분에서 매우 의미 있는 역할을 할 수 있게 되었다.

지도학습 계열의 알고리즘뿐만 아니라 비지도학습 계열의 알고리즘들이 끊임없이 개발되고 연구되고 있다.

본 연구를 통해서 vuforia등에서 구현하고 있는 인공지능 알고리즘들의 성능 분석과 비교를 수행하고 최신 기계학습 알고리즘을 적용하여 객체를 정확하게 인식하고 정합도를 높이는 방법에 대해 연구한다.

또한 기계학습에 의한 객체인식 방법과 센서 기반의 객체 인식방법에 대한 기법을 개발한다.

결론적으로 본 연구에서 IoT와 증강현실 기술의 융합을 통한 효율적인 객체인식 및 정합기술을 개발하기 위해 우선적으로 SSD(single shot multibox detector)를 이용한 객체인식 모델의 비교평가를 수행하였다.

\* corresponding author

## II. 관련연구

Yolo는 2015년에 나온 논문으로 Faster R-CNN에 비하여 부러 6배 가량 빠른 속도를 보인다. 정확도는 비록 조금 낮다 하더라도 정말 비약적인 발전이라 할 수 있다. Yolo가 기존의 Object Detection과 가장 크게 구분되는 부분은 기존에 1) region proposal 2) classification 이렇게 두 단계로 나누어서 진행하던 방식에서 region proposal 단계를 제거하고 한번에 Object Detection을 수행하는 구조를 갖는다는 점이다. 이것이 어떻게 가능할까? 아래 그림1은 yolo의 1 Step 구조를 간단히 보여준다[1].

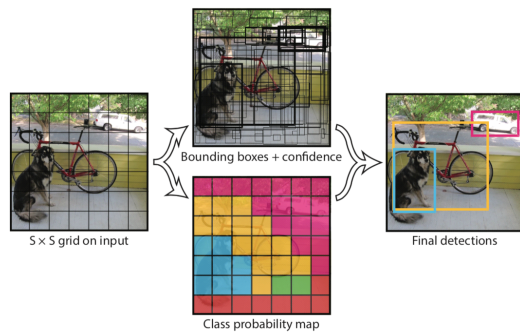


그림 1 YOLO의 1 Step 구조

YOLO는 속도 측면에서 당시 Faster R-CNN이 7FPS이었던 것을 45FPS까지 끌어올리는 비약적인 발전을 이루었다. 하지만 정확도 측면에선 다소 한계점이 존재한다. 또한 작은 물체들은 잘 잡아내지 못한다는 문제점을 가지고 있다. SSD는 바로 이러한 한계점을 극복하고자 하는 시도에서 출발하던 알고리즘이다[2].

YOLO의 문제점은 입력 이미지를 7x7 크기의 그리드로 나누고, 각 그리드 별로 Bounding Box Prediction을 진행하기 때문에 그리드 크기보다 작은 물체를 잡아내지 못하는 문제가 있었다. 그리고 신경망을 모두 통과하면서 컨볼루션과 풀링을 거쳐 coarse한 정보만 남은 마지막 단 피쳐맵만 사용하기 때문에 정확도가 하락하는 한계가 있었다.

## III. 목표시스템

표 1 다중데이터셋 별 SSD 결과

System	VOC2007 test mAP	FPS (Titan X)	Number of Boxes	Input resolution
Faster R-CNN (VGG16)	73.2	7	~6000	~1000 x 600
YOLO (customized)	63.4	45	98	448 x 448
SSD300* (VGG16)	77.2	46	8732	300 x 300
SSD512* (VGG16)	79.8	19	24564	512 x 512

영상인식을 위한 기계학습 알고리즘의 다양한 적용을 통한 성능 비교 및 평가를 수행한다. 지도 학습 계열의 알고리즘과 비지도학습 계열의 기계 학습 알고리즘들을 이용한 객체인식 성능 비교 및 평가를 수행한다. 정량적인 성능 평가를 수행한다.

SSD는 학습된 데이터를 카메라를 이용하여 검출하는 일반적인 객체인식 알고리즘이다.

이 전의 객체 검출 구조는 두 개의 구분된 단계를 가지고 있다. 물체의 위치를 제안하는 위치 제안 네트워크(region proposal network)와 제안된 위치에서 물체의 종류를 결정하는 구분자(classifier) 두 가지로 구성된다. 연산관점에서, 이런 구조는 매우 많은 연산을 필요로 하며 실제 세계에서 실시간으로 사용하기 어렵다. Single-shot models은 localization과 detection 작업을 단일 네트워크의 단일 forward 연산을 통해서 해결한다. 결과적으로 상당히 빠르게 수행할 수 있으며 연산능력이 적은 Hardware에서도 수행이 가능해진다.

## IV. 결론

기존 AR 객체 인식기법으로 마커 기반, 텍스처 기반, 예지 기반, 템플릿 기반 기법들이 사용되었으나 각각의 방법들은 고유한 한계점들을 가지고 있고 이를 해결하기 위하여 최근에는 텍스처가 부족한 객체 인식 및 추적기법과 가변형 객체추적기법과 센서를 활용하는 방법 등이 연구되고 있다. 본 연구에서는 최근 객체 인식 과정에서 연산량을

줄일 수 있는 SSD를 추가적으로 적용하여 성능을 평가하였다.

### Acknowledgement

이 논문은 한국연구재단(과학기술정보통신부)의 지원에 의함.(No. NRF- 2019R1G1A1087290)

### References

- [1] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [2] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.