

360° 스트리밍 영상에서의 객체 인식 연구

윤정록[○], 전성국^{*}, 김희민^{*}, 김운용^{*}

한국광기술원 공간광정보연구센터^{*}

전남대학교 전자컴퓨터공학과[○]

e-mail: {justin182, k612051, hmkim, kuy7023}@kopti.re.kr^{*}, 197212@jnu.ac.kr[○]

Object Recognition in 360° Streaming Video

Jeongrok Yun[○], Sungkuk Chun^{*}, Hoemin Kim^{*}, Un Yong Kim^{*}

Spatial Optical Information Research Center, Korea Photonics Technology Institute^{*}

Dept. of Electronics and Computer Engineering, Chonnam National University[○]

● 요약 ●

가상/증강현실로 대표되는 공간정보 기반 실감형 콘텐츠에 대한 관심이 증대되면서 객체인식 등의 지능형 공간인지 기술에 대한 연구가 활발히 진행되고 있다. 특히 HMD 등의 영상 시각화 장치의 발달 및 5G 통신 기술의 출현으로 인해 실시간 대용량 영상정보의 송, 수신 및 가시화 처리 기술의 기반이 구축됨에 따라, 360° 스트리밍 영상정보 처리와 같은 고자유도 콘텐츠를 위한 관련 연구의 필요성이 증대되고 있다. 하지만 지능형 영상정보 처리의 대표적 연구인 딥 러닝(Deep Learning) 기반 객체 인식 기술의 경우 대부분 일반적인 평면 영상(Planar Image)에 대한 처리를 다루고 있고, 파노라마 영상(Panorama Image) 특히, 360° 스트리밍 영상 처리를 위한 연구는 미비한 상황이다. 본 논문에서는 딥 러닝을 이용하여 360° 스트리밍 영상에서의 객체인식 연구 방법에 대해 서술한다. 이를 위해 360° 카메라 영상에서 딥 러닝을 위한 학습 데이터를 획득하고, 실시간 객체 인식이 가능한 YOLO(You Only Look Once) 기법을 이용하여 학습을 한다. 실험 결과에서는 학습 데이터를 이용하여 360° 영상에서 객체 인식 결과와, 학습 횟수에 따른 객체 인식에 대한 결과를 보여준다.

키워드: 360° 스트리밍 영상(360° Streaming Video), YOLO, 객체인식(Object Recognition)

I. Introduction

최근 HMD 등의 영상 시각화 장치의 발달 및 5G 통신 기술의 출현으로 인해 실시간 대용량 영상정보의 송, 수신 및 가시화 처리 기술의 기반이 구축됨에 따라, 360° 스트리밍 영상정보 처리와 같은 고자유도 콘텐츠를 위한 관련 연구의 필요성이 증대되고 있다. 하지만 지능형 영상정보 처리의 대표적 연구인 딥 러닝(Deep Learning) 기반 객체 인식 기술의 경우 대부분 일반적인 평면 영상(Planar Image)에 대한 처리를 다루고 있고, 파노라마 영상(Panorama Image) 특히, 360° 스트리밍 영상 처리를 위한 연구는 미비한 상황이다. 본 논문에서는 딥 러닝을 이용하여 360° 스트리밍 영상에서의 객체인식 연구 방법에 대해 서술한다. 이를 위해 360° 카메라 영상에서 딥 러닝을 위한 학습 데이터를 획득하고, 실시간 객체 인식이 가능한 YOLO(You Only Look Once) 기법을 이용하여 학습을 한다. 실험 결과에서는 학습 데이터를 이용하여 360° 영상에서 객체 인식 결과와, 학습 횟수에 따른 객체 인식에 대한 결과를 보여준다.

II. Proposed System

360° 영상에서의 객체 인식을 위해 객체에 대한 Dataset을 제작한다. 스튜디오 내부에서 3개의 객체에 대한 영상을 촬영 하였으며, 60° 화각의 6대의 카메라가 연결되어 있는 4k 360° 카메라를 이용하여 촬영 하였다. 3개의 객체는 불규칙한 거리, 각도, 위치, 가려짐 정도를 가지고 있으며, 각 객체 당 200개의 Dataset을 획득하였다. Fig. 1은 한 영상에서 3개의 객체들에 대한 Dataset의 예시이다.



Fig. 1. Examples of 360 Streaming Image Dataset

제작된 Dataset은 360° 스트리밍 영상에서의 객체 인식을 위하여 YOLO의 딥러닝 process를 이용하여 학습하였다. YOLO는 기존의 객체 인식 방법인 RCNN(Regions with CNN)[1]에 비하여 1,000배, Fast R-CNN[2]에 비하여 100배 이상 빠른 속도로 객체 인식이 가능하며 인식을 또한 우수하다. 학습 환경은 Intel Core i5-8500 4.10GHz, NVIDIA Geforce GTX 1050ti 그래픽카드, OPENCV 3.0의 환경에서 수행하였으며, 2000번과 4000번 학습하였을 경우 두 가지 경우의 가중치를 획득하였다.

III. Experiment result

실험은 학습에 사용된 영상과 학습에 사용되지 않은 영상에 대하여 진행 되었으며, 학습에 사용되지 않은 영상의 경우 2000번 학습 했을 경우와 4000번 학습 했을 경우 두 가지의 가중치에 대하여 실험하였다. 다음 Fig 2.은 학습에 사용된 영상에 대한 실험 영상이다.

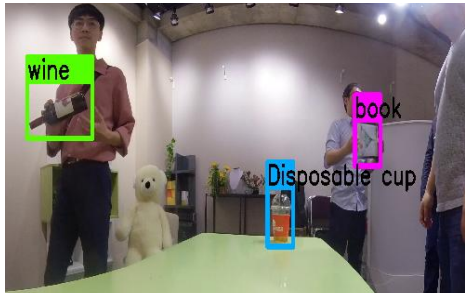


Fig. 2. Example Result of Proposed System

인식률은 평균 86%의 높은 정확도를 갖는다.

다음 Fig 3.은 2000번 학습 했을 경우와 4000번 학습한 가중치를 검증 영상에 대한 객체 인식 결과이다. 2000번 학습 했을 경우는 평균 손실값이 최소인 시점에서 중요한 가중치이며, 4000번 학습 했을 경우의 가중치는 평균 손실값 이후 2000번의 추가 학습으로 과적합(Overfitting)된 가중치 데이터이다.



Fig. 3. Object Recognition Results

와인 병의 경우 양쪽 전부 인식되지 않는다. YOLO의 특성상 세로로 길며 작은 물체는 인식에 어려움이 있으며 학습데이터의 수 또한 많지 않기 때문이다. 2000번 학습 했을 경우의 실험 결과에서는 책에 대한 인식이 정상적으로 작동한다. 그러나 4000번 학습 했을 경우의 실험 결과에선 과적합으로 책에 대한 false-positive 에러가 발생하였다.

IV. Conclusions

본 논문에서는 360° 스트리밍 영상에서의 객체 인식을 위한 딥러닝 학습 방법을 제안하였다. 향후, 제안된 데이터를 이용하여 스튜디오 내부의 빛의 강도에 대한 조절과 ROI지정 등으로 학습 데이터의 신뢰도와 객체 인식률의 향상을 기대하고 있으며, 360° VR스트리밍을 이용한 소셜 커머스 형태의 콘텐츠 연구를 진행할 계획이다.

ACKNOWLEDGEMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2018년도 문화 기술연구개발 지원사업으로 수행되었음.

REFERENCES

- [1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," P roceeding of the IEEE Conference on Computer Vision and P attern Recognition, pp. 580-587, 2014.
- [2] R. Girshick, "Fast R-CNN," P roceeding of the IEEE International Conference on Computer Vision, pp. 1440-1448, 2015.