한국정보전자통신기술학회논문지(jkiiect)'23-6, Vol.16 No.3
논문 23-16-03-117 http://dx.doi.org/10.17661/jkiiect.2023.16.3.117

응시점 추정 기반 관심 영역 내 객체 탐지

한석호*, 장훈석**

Object detection within the region of interest based on gaze estimation

Seok-Ho Han*, Hoon-Seok Jang**

요 약 사용자가 현재 응시하고 있는 위치를 자동으로 인식하는 응시점 추정과 추정된 응시점을 기반으로 객체를 탐지하는 기술을 활용한다면 사람의 시각적 행동을 파악하는데 더 정확하고 효율적인 방안이 될 수 있을 것이다. 본 논문에서는 응시점을 중심으로 관심 영역을 생성하고 해당 영역 내에서 객체를 탐지하는 방안을 제시한다. 자세하게는, 삼차원 응시점을 추정한 후에 추정된 응시점을 기반으로 관심 영역을 생성하여 관심 영역 내에서만 객체 탐지가 이루어지도록 설계한다. 실험을 통해 일반적인 객체 탐지와 제안된 관심 영역 내 객체 탐지 성능을 비교한 결과, 프레임당 처리 시간은 각각 1.4ms, 1.1ms로 관심 영역 내 객체 탐지가 처리 속도 면에서 더 우수한 것을 확인할 수 있었다.

Abstract Gaze estimation, which automatically recognizes where a user is currently staring, and object detection based on estimated gaze point, can be a more accurate and efficient way to understand human visual behavior. in this paper, we propose a method to detect the objects within the region of interest around the gaze point. Specifically, after estimating the 3D gaze point, a region of interest based on the estimated gaze point is created to ensure that object detection occurs only within the region of interest. In our experiments, we compared the performance of general object detection, and the proposed object detection based on region of interest, and found that the processing time per frame was 1.4ms and 1.1ms, respectively, indicating that the proposed method was faster in terms of processing speed.

Key Words : Computer vision, Eye tracking, Region of interest, Object detection, YOLOv5

1. 서론

응시 추적(Eye Tracking)은 현재 사용자가 응시하고 있는 위치를 자동으로 인식하는 것을 의미한다 [1]. 응시 추적은 사람이 어디를 보고 있는지, 어디에 주의를 기울이고 있는지, 무엇에 관심이 있는지 등 시각적 행동을 파악하는 데 유용하며 [2] 응시 추적은 심리학, 마케팅, 게임 등 다양한 분야에서 활용되고 있다. 한편, 이미지를 해석하기 위해서는 서로 다른 이미지를 분류하는 것에 집중하는 것뿐만 아니라 각 이미지에 객체의 개념과 위치를 정확하게 추정하려고 노력해야 한다 [3]. 컴퓨터에서 이미지를 해석하기 위해서는 객체 탐지(Object Detection) 기술이 필요하다. 객체 탐지는 이미지 혹은 동영상에서 인간, 동물, 자동차 등 시각적인 객체를 컴퓨터가 탐지하는 컴퓨터 비전(Computer Vision)의 핵심기술이다 [4]. 이러한 객체 탐지는 사람의 맨눈으로는 쉽게 수행할 수 있지만, 컴퓨터는 특정 알고리즘을 통해서만 명확하게 구분할 수 있으며, 객체 탐지 알고리즘은 주요 객체를 탐지하고 해당 객체를 중심으로 경계(Bounding Box)를 표시하

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1A1052728).
*IT Application Research Center, Korea Electronics Technology Institute (KETI)
**IT Application Research Center, Korea Electronics Technology Institute (KETI)
Received April 26, 2023 Revised May 24, 2023 Accepted June 12, 2023

여 구분한다 [5]. 앞서 설명한 응시 추적과 컴퓨터 비전의 핵심기술인 객체 탐지를 활용하면 사람의 시각적 행동을 파악하는데 더 정확하고 효율적인 방안이 될 수 있을 것이다.

따라서 본 논문에서는 응시 추적과 객체 탐지 두 가지 기술을 사용하여 사용자가 보고 있는 응시점을 기준으로 관심 영역 ROI(Region of Interest)를 생성한 뒤 관심 영역 안에서만 객체 탐지가 되도록 하는 연구를 진행하고자 한다.

2. 관련 연구

2.1 YOLO (You Only Look Once)

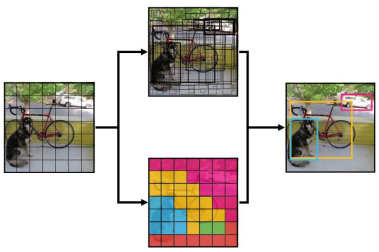

그림 1. YOLO를 사용한 객체 검출
Fig. 1. Object detection using YOLO

YOLO(You Only Look Once)는 2016년에 소개된 객체 탐지 알고리즘이다. YOLO는 들어온 이미지를 S × S 그리드로 나누고 객체의 중심이 그리드 셀에 들어가면, 해당 그리드 셀은 객체 탐지를 진행한다 [6]. YOLO는 region proposal과 classification이 동시에 이루어지는 1- Stage 방식이기 때문에 실시간 응용 프로그램에서 사용하기 적합하다.

2.2 YOLOv5 네트워크 구조

그림 2는 본 논문에서 사용할 YOLOv5 네트워크 구조이다. YOLOv5 네트워크 구조는 Backbone, Neck, Head 세 부분으로 구성된다. Backbone에서는 들어온 이미지의 특징 추출을 통해 다양한 크기의 feature map을 생성한다. CBS layer는 Con2D와 BatchNormal, SiLU로 구성된 Convolution layer이다. CSP(Cross Stage Partial Network)에서 사용되는 Bottleneck은 두 개의 CBS로 구성된 Bottleneck 1과 두 개의 CBS와 나머지 연결이 합산되는 Bottleneck 2로 구성되어 있다. CSP1_X와 CSP2_X는 CSP를 기반으로 하는 layer로 X는 사용된 Bottleneck의 수를 의미한다. 해당 구조는 CNN 학습 능력을 향상하며 계산의 복잡성을 줄인다. CSP1_X는 Backbone에서 feature 추출, CSP2_X는 Neck에서 feature 융합에 사

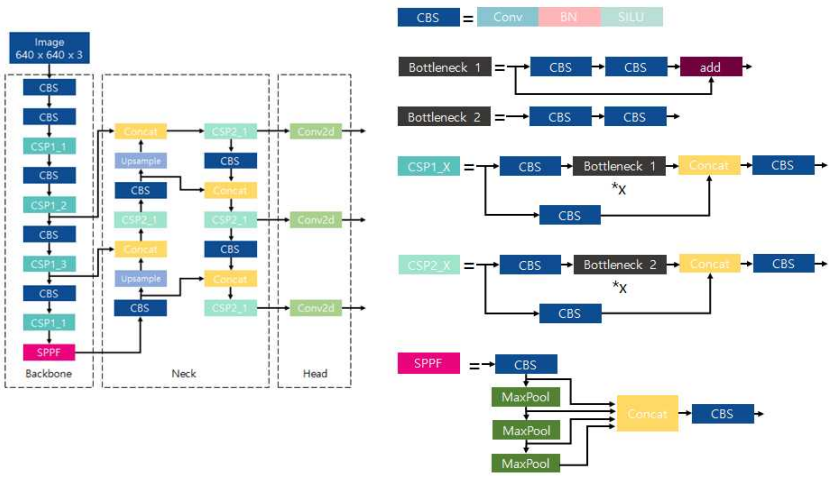

그림 2. YOLOv5 네트워크 구조
Fig. 2. YOLOv5 network structure

용된다. SPPF(Spatial Pyramid Pooling-Fast)는 3개의 MaxPool layer를 연속적으로 사용하고 그 결과 Concat 연산으로 결합하는 방식으로 다양한 스케일의 feature map을 생성하는 역할을 하며, 기존에 사용되던 SPP보다 속도가 빠르다.

Neck은 Backbone에서 생성된 feature를 융합하여 Head에 전달하는 역할을 하며, FPN(Feature Pyramid Network)과 PAN(Path Aggregation Network) 피라미드 구조가 사용된다. FPN 구조는 상위 feature map에서 하위 feature map으로 feature를 전달하고 동시에 PAN 구조는 낮은 feature map에서 높은 feature map으로 localization feature를 전달한다. 이를 통해 다른 네트워크에서 추출된 feature를 강화하여 탐지 기능을 향상시킨다.

최종적으로 Head는 탐지 단계로서, feature map에서 크기가 다른 객체를 예측하는 데 사용된다 [7,8,9].

2.3 응시 추적과 관련된 연구

응시 추적 기술의 발달로 안경형(Glass), 머리 부착형(Head mounted), 모니터 부착형(Screen based) 등 다양한 응시 추적 기기(Eye Tracker)가 생겨났으며, 응시 추적 기기가 제공하는 소프트웨어를 통해 응시점, 동공 위치, 응시 깊이 등 다양한 데이터를 획득하는 것이 가능해졌다. 이에 따라 응시 추적 기술을 활용한 관련 연구들이 진행되고 있는데, 몇 가지 연구를 소개하면 다음과 같다.

Souchet, Alexis D (2022)은 Head Mounted Displays(HMDs)를 사용하는 동안 발생하는 시각적 피로 및 높은 인지 부하를 Eye Tracking 기술을 통해 측정하는 방법에 관해 연구하였다. [10]

Lim Jia Zheng (2022)은 감정 인식 분야에서 감정 모델링, 감정 자극 방법, 응시 추적 데이터를 통해 추출할 수 있는 감정 관련 특징 등 응시 추적 기술을 이용하여 감정 인식에 관한 연구를 진행하였다. [11]

Boerman Sophie C (2023)은 마케팅 분야에서 응시 추적과 온라인 실험을 통해 인스타그램에서 인플루언서 마케팅에 대한 설득 지식수준과 인스타그램 사용자가 인플루언서 마케팅을 식별하기 위해 사용하는 단서에 대한 정보를 얻는 연구를 진행하였다. [12]

2.4 응시 추적을 활용한 객체 탐지 연구

응시 추적 기술을 활용하여 객체 탐지를 진행하는 연구들도 최근에 진행되고 있다. 다음 예시를 소개하면 다음과 같다.

Kumari Niharika (2021)는 실제 학생들의 실험 과정에서 모바일 응시 추적 데이터를 사용하여 실제 객체에 대한 객체 인식 모델을 적용하는 연구를 진행하였다. [13]

Qin Long (2022)는 응시 추적 기술을 이용하여 운전자의 시선이 집중된 응시 영역을 기반으로 눈에 띄는 물체와 중요한 객체를 탐지하는 연구를 진행하였다. [14]

Vishwakarma Sandhya(2018)는 객체 탐지 부분에서 응시 추적 데이터를 통해 사용자의 응시점을 기반으로 출입 금지 표지판을 학습하는 방법에 대해 제안했다. [15]

3. 제안된 방법

3장에서는 본 논문에서 제안된 방법에 관해 설명한다. 먼저 삼차원 응시 데이터를 획득한 후에 관심 영역을 생성하여 관심 영역 안에서만 객체가 탐지되게 설정하였다.

3.1 삼차원 응시점 추정

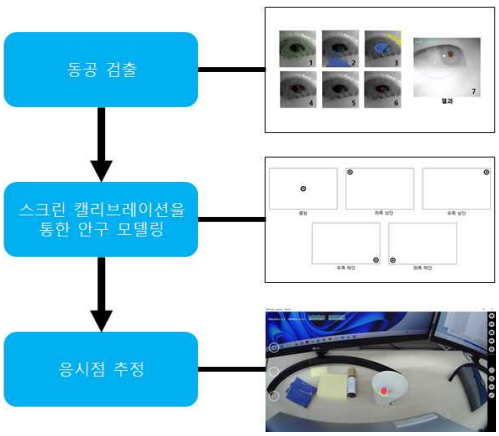

그림 3. 삼차원 응시점 추정 흐름도
Fig. 3. Flowchart for 3D gaze estimation

응시 추적을 위해서는 먼저 응시점을 추정해야 한다. 삼차원 응시점 추정은 그림 3의 과정으로 이루어진다.

3.1.1 동공 검출

동공 검출은 동공 검출 알고리즘을 통해 동공을 검출한다. 알고리즘은 1. Canny Edge Detection, 2. Histogram을 통한 동공 부위 검출, 3. 검출된 동공(파란색)을 제외한 가장자리 Filtering, 4. 곡선 연결선 기준을 통한 초기 동공 경계선 검출, 5. Ellipse Fitting을 이용한 동공 경계 후보 형성, 6. 신뢰성 판단을 통한 최종 동공 후보 결정을 끝으로 총 6단계로 진행된다 [16].

3.1.2 안구 모델링

안구 모델링은 스크린 캘리브레이션을 통해 진행된다. 스크린 캘리브레이션은 모니터에 서로 다른 위치에 5개의 마커가 나타나는데 마커를 응시하면 양안 카메라에 촬영된 영상에서 동공의 위치를 파악하여 캘리브레이션을 진행한다 [17].

3.1.3 응시점 추정

안구 모델링 후 양안 주시선 확장 및 교차점 추정 작업을 통해 실시간 교차점을 추적한다. 이를 통해 안구 운동 잡음, 유의하지 않은 응시점 형성, 눈의 깜빡임 등을 제거를 통한 최종 응시/응시 깊이 값을 추정한다.

3.2 응시 데이터 획득

삼차원 응시점 추정 후 관심 영역을 생성하는 데 필요한 응시 데이터들을 CSV 파일로 획득하였다. 획득한 데이터는 표 1과 같다.

표 1. 응시 데이터
Table 1. Gaze data

world_index	비디오 프레임
confidence	동공 감지 평가
gaze_point_3d_x	응시점 x 좌표
gaze_point_3d_y	응시점 y 좌표

3.3 응시점 기반 관심 영역 생성

사용자의 응시점을 기준으로 비디오 프레임마다 관심 영역을 생성하기 위해 응시점의 x좌표와 y좌표를 확인하여 빨간색 원으로 표시하고, 응시점을 중심으로 960x540 크기의 노란색 직사각형 관심 영역을 생성하였다. 생성된 관심 영역에서만 객체 탐지가 이루어지게 코드를 수정한 후 실험을 진행하였다.

4. 실험 결과

4장에서는 관심 영역 내에서 객체 탐지 결과와 일반적인 객체 탐지 결과를 비교하였다. 환경은 windows 10 Education 64-bit, Intel Core i5-12400F, 16GB 메모리, GeForce RTX 3070이 장착된 PC에서 진행하였다.

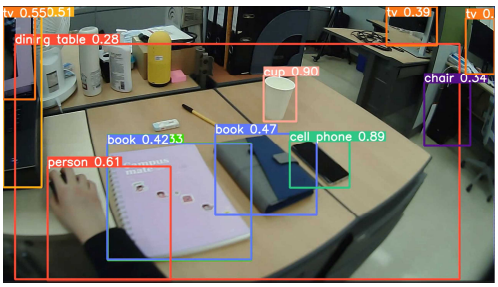

그림 4. 객체 탐지 결과
Fig. 4. Object Detection Results

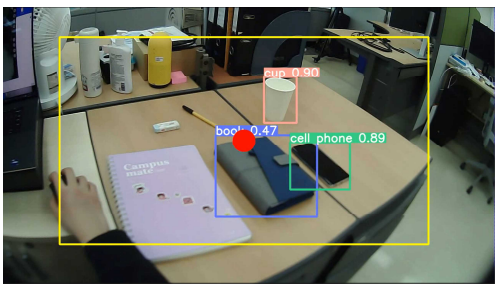

그림 5. 관심 영역 내 객체 탐지 결과
Fig. 5. Object detection results within region of interest

그림 4는 일반적인 객체 탐지했을 때 결과, 그림 5

는 응시점 기반 관심 영역 내 객체 탐지했을 시 결과이다. 그림에서 확인할 수 있듯이, 관심 영역 내에서 객체 탐지를 수행하였을 때 3개의 객체만 탐지되지만, 일반적인 탐지를 수행하면 약 10개의 객체가 탐지되는 것을 확인할 수 있었다. 또한 프레임 당 소요되는 처리 시간의 비교 결과를 살펴보면 일반적인 객체 탐지를 했을 때는 1.4ms, 관심 영역 내 객체 탐지를 진행했을 시 1.1ms인 것을 확인할 수 있었다.

5. 결론

본 논문에서는 응시점 추적과 객체 탐지, 두 가지 기술을 사용하여 사용자의 응시점을 기준으로 관심 영역을 생성하고 해당 관심 영역 안에 있는 객체들을 탐지하는 기술을 제안했다. 응시 추적 데이터는 응시 추적 기기를 사용하여 손쉽게 응시 데이터를 획득할 수 있었다. 획득된 응시점을 중심으로 가로 960, 세로 540 크기의 관심 영역을 생성하고 객체 탐지에 적용해 일반적인 객체 탐지와 관심 영역 내 객체 탐지 결과를 비교하였다. 일반적인 객체 탐지는 1.4ms의 프레임 당 처리 시간이 소요되지만 관심 영역 내 객체 탐지는 1.1ms의 프레임 당 처리 시간이 소요되므로 일반적인 탐지 결과에 비해 관심 영역 내 객체 탐지를 했을 시 탐지되는 객체의 수가 적게 나옴으로써 처리 속도가 더 빠른 것을 확인할 수 있었다.

본 연구 결과는 국방 및 엔터테인먼트 등의 분야에서의 응용 가능성을 제시하며, 적용한다면 불필요한 연산을 줄이고, 관심 있는 객체에 대한 탐지 정확도 향상을 통해 다양한 분야에서의 신뢰성 있는 결과를 제공할 수 있으리라 생각된다. 향후 추가적인 연구를 통해 응시 추적과 객체 탐지 기술의 응용 방안을 탐구할 예정이다.

REFERENCES

[1] S. K. Hwang, M. J. Moon, S. Cha, E. S. Cho, C. S. Bae, “Real Time Eye and Gaze Tracking”, Journal of Korea Institute of Information, Electronics, and communication Technology, vol. 2, no. 3, pp. 61-69, 2009.

[2] Panetta, Karen, et al., “Iseecolor: method for advanced visual analytics of eye tracking data”, IEEE Access, vol. 8, pp. 52278-52287, 2020.

[3] Zhao, Zhong-Qiu, et al., “Object detection with deep learning: A review.”, IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212-3232, 2019.

[4] Zou, Zhengxia, et al., “Object detection in 20 years: A survey.”, Proceedings of the IEEE, 2023.

[5] Y. H. LEE and Y. S. Kim, “Comparison of CNN and YOLO for Object Detection.”, Journal of the semiconductor & display technology, vol. 19, no. 1, pp. 85-92, 2020.

[6] 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.

[7] Kangshun Li, Jiancong Wang, Hassan Jalil, Hui Wang, “A fast and lightweight detection algorithm for passion fruit pests based on improved YOLOv5”, Computers and Electronics in Agriculture, vol. 204, 2023.

[8] Horvat, Marko, and Gordan Gledec, “A comparative study of YOLOv5 models performance for image localization and classification.”, Central European Conference on Information and Intelligent Systems, Faculty of Organization and Informatics Varazdin, pp. 349-356, 2022.

[9] Xue, Zhenyang, Haifeng Lin, and Fang Wang., “A small target forest fire detection model based on YOLOv5 improvement.”, Forests, vol. 13, no. 8, 2022.

[10] Souchet, Alexis D., et al., “Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review.”, International Journal of Human-Computer Interaction, vol. 38, no. 9, pp. 801-824, 2022.

[11] Lim, Jia Zheng, James Mountstephens, and Jason Teo., “Emotion recognition using eye-tracking: taxonomy, review and current challenges.”, Sensors, vol. 20, no. 8, 2020.

[12] Boerman, Sophie C., and Céline M. Müller, “Understanding which cues people use to identify influencer marketing on Instagram: a

n eye tracking study and experiment.", International Journal of Advertising, vol. 41, no. 1, pp. 6-29, 2022.
[13] Kumari, Niharika, et al., "Mobile eye-tracking data analysis using object detection via YOLO v4.", Sensors, vol. 21, no. 22, 2021.
[14] Qin, Long, et al., "ID-YOLO: Real-time salient object detection based on the driver's fixation region.", IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 15898-15908, 2022.
[15] Vishwakarma, Sandhya, D. Radha, and J. Amudha., "Effectual training for object detection using eye tracking data set.", 2018 international conference on inventive research in computing applications (ICIRCA). IEEE, 2018.
[16] Kassner, Moritz, William Patera, and Andreas Bulling. "Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction." Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication. 2014.
[17] J. H. Mun, D. W. Shin, and Y. S. Ho, "3-Dimensional Calibration and Performance Evaluation Method for Pupil-labs Mobile Pupil Tracking Device.", Smart Media Journal, vol. 7, no. 2, pp. 15-22, 2018.

저자약력

한 석 호(Seok-Ho Han) **[정회원]**

- 2023년 2월 : 원광대학교 디지털콘텐츠공학과(학사)
- 2022년 9월~현재 : 한국전자기술연구원 연구원

〈관심분야〉 신호 및 영상처리, 디지털트윈, 기계 및 심층 학습

장 훈 석(Hoon-Seok Jang) **[정회원]**

- 2014년 8월 : 광주과학기술원 기전공학과(공학석사)
- 2019년 2월 : 광주과학기술원 기전공학과(공학박사)
- 2020년 2월~현재 : 한국전자기술연구원 선임연구원

〈관심분야〉 신호 및 영상처리, 증강 및 혼합 현실, 기계 및 심층 학습