

# DNN을 이용한 응시 깊이 보정

한석호\*, 장훈석\*\*

## Correcting the gaze depth by using DNN

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

**요약** 응시점을 통해 어떤 것을 보고 있는지 알 수 있다면 많은 정보를 얻을 수 있다. 응시 추적 기술의 발달로 응시 점에 대한 정보는 다양한 응시 추적 기기에서 제공해주는 소프트웨어를 통해 얻을 수 있다. 하지만 실제 응시 깊이와 같은 정확한 정보를 추정하기란 어렵다. 응시 추적 기기를 통해 만약 실제 응시 깊이로 보정할 수 있다면 시뮬레이션, 디지털 트윈, VR 등 다양한 분야에서 현실적이고 정확한 신뢰성 있는 결과를 도출하는 것이 가능해질 것이다. 따라서 본 논문에서는 응시 추적 기기와 소프트웨어를 통해 원시 응시 깊이를 획득하고 보정하는 실험을 진행한다. 실험은 Deep Neural Network(DNN) 모델을 설계한 후 소프트웨어에서 제공하는 응시 깊이 값을 300mm에서 10,000mm까지 지정한 거리별로 획득한다. 획득한 데이터는 설계한 DNN 모델을 통해 학습을 진행하여 실제 응시 깊이와 대응하도록 보정하였다. 보정한 모델을 통해 실험을 진행한 결과, 300mm에서 10,000mm까지 지정한 거리별 297mm, 904mm, 1,485mm, 2,005mm, 3,011mm, 4,021mm, 4,972mm, 6,027mm, 7,026mm, 8,043mm, 9,021mm, 10,076mm로 실제와 비슷한 응시 깊이 값을 획득할 수 있었다.

**Abstract** if we know what we're looking at, we can get a lot of information. Due to the development of eye tracking, Information on gaze point can be obtained through software provided by various eye tracking equipments. However, it is difficult to estimate accurate information such as the actual gaze depth. If it is possible to calibrate the eye tracker with the actual gaze depth, it will enable the derivation of realistic and accurate results with reliable validity in various fields such as simulation, digital twin, VR, and more. Therefore, in this paper, we experiment with acquiring and calibrating raw gaze depth using an eye tracker and software. The experiment involves designing a Deep Neural Network (DNN) model and then acquiring gaze depth values provided by the software for specified distances from 300mm to 10,000mm. The acquired data is trained through the designed DNN model and calibrated to correspond to the actual gaze depth. In our experiments with the calibrated model, we were able to achieve actual gaze depth values of 297mm, 904mm, 1,485mm, 2,005mm, 3,011mm, 4,021mm, 4,972mm, 6,027mm, 7,026mm, 8,043mm, 9,021mm, and 10,076mm for the specified distances from 300mm to 10,000mm.

**Key Words** : Eye tracking, Gaze depth, Gaze correction, Deep Neural Network (DNN)

### 1. 서론

응시점에는 여러 가지 정보가 담겨 있다. 응시점을 통해 무엇을 보는가를 알 수 있다면 어떤 것이 응시점

을 집중시키고 있는지, 특정 요소가 얼마나 오래 주의를 집중시키고 있는지 등을 알 수 있으며 어떤 공간 요소가 응시점을 유도하는지, 구성 요소의 시각적 상호

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 28, 2023

Accepted June 05, 2023

영향까지도 알아낼 수 있다 [1]. 이러한 응시 추적에 관한 연구는 심리학, 인지 언어학, 기계 등 여러 학계에서 활발하게 진행되고 있다. 응시 추적을 위해서는 인간의 시각적 활동 연구 및 모든 시각적 활동이 일차적으로 시각 정보의 유입이 우선시 되므로 응시점을 객관적으로 분석할 수 있어야 한다. [2,3] 이러한 응시 정보를 획득 및 분석을 위해서는 응시 추적 기술이 필요하다. 응시 추적 기술은 눈동자의 움직임을 감지하여 응시점의 위치를 추적하는 방법을 의미한다 [4]. 현재는 응시추적 기술의 발달로 안경형(Glasses) 응시 추적 기기, 머리 부착형(Head mounted) 응시 추적 기기, 모니터 부착형(Screen based) 응시 추적 기기 등 다양한 장비에서 제공하는 소프트웨어를 통해 응시점, 동공 위치, 응시 깊이 등 다양한 정보를 획득하는 것이 가능해졌다. 하지만 현실의 실제 응시 깊이와 같은 정확한 정보를 추정하기란 어렵다. 정확한 응시 깊이를 획득하기 위해서는 소프트웨어에서 추정된 응시 깊이 값을 획득한 뒤 사용자 스스로 보정해야 하는 작업이 필요하다. 따라서 본 논문에서는 응시 추적 기기 Pupil Core와 응시 추적 소프트웨어를 통해 원시 응시 깊이를 획득하기 위한 연구를 진행하였다.

## 2. 개발 환경 구축

### 2.1 하드웨어 구성

그림 1은 응시 추적 기기 Pupil Core이다. 양쪽 눈을 촬영하는 2개의 Eye 카메라와 사용자의 시야가 될 World 카메라로 구성되어 있다. 성공적인 응시 추적 및 보정을 위해서는 동공 감지와 World camera의 FOV가 제대로 조정되어야 하는데 이를 위해 Eye camera와 연결된 Eye camera arm을 안쪽으로 밀거나 연결된 Ball joint를 중심으로 회전시키면서 카메라 조정할 수 있다. World camera의 경우 위아래로 회전하여 FOV를 조정할 수 있다. Pupil Core는 USB를 통해 노트북, 컴퓨터, 모바일 컴퓨터 등 다양한 플랫폼에 연결하여 제공하는 응시 추적 소프트웨어 Pupil Capture와 함께 사용할 수 있다.



그림 1. 응시 추적 기기

Fig. 1. Eye tracker

### 2.2 소프트웨어 구성

응시 추적 소프트웨어 Pupil Capture는 Eye camera와 World camera에서 들어오는 비디오 스트림을 통한 동공 감지, 응시 추적, 비디오 및 이벤트 기록 등 데이터를 실시간으로 전송한다. 그림 2는 Pupil Core 연결 후 Pupil Capture를 실행한 화면이다. World video 창과 2개의 Eye video 창이 나오는 것을 볼 수 있다.

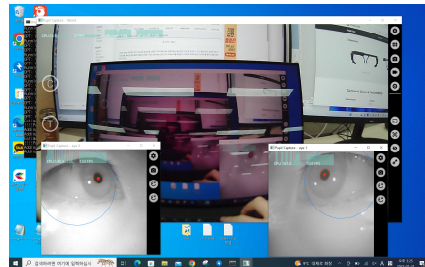


그림 2. 응시 추적 소프트웨어

Fig. 2. Software for eye tracking

World video 상단에는 CPU, FPS, 동공 알고리즘 신뢰도가 표시되며 좌측 단축키에는 동공 보정, 녹화 등 클릭할 수 있는 플러그인이 있으며 좌측 사이드바를 통해 설정과 플러그인을 추가할 수 있다. Eye video 상단 역시 CPU와 FPS를 확인할 수 있으며, 우측에 있는 사이드바를 통해 설정을 변경할 수 있다.

### 3. 삼차원 응시점 추정

#### 3.1 요약

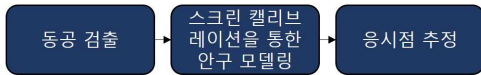


그림 3. 삼차원 응시점 추정 흐름도  
Fig. 3. Flowchart for 3D gaze estimation

그림 3은 삼차원 응시점 추정 흐름도이다. 삼차원 응시점 추정 흐름도는 동공 검출, 스크린 마커의 위치 촬영을 통한 안구 모델링, 응시점 추정 과정으로 진행된다.

#### 3.2 동공 검출

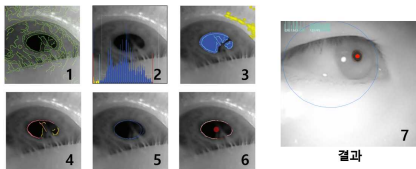


그림 4. 동공 검출 알고리즘의 단계별 결과  
Fig. 4. Step by step results of the pupil detection algorithm

동공 검출 알고리즘의 단계별 결과는 그림 4를 통해 확인할 수 있다. 1. Canny Edge Detection 진행 후 2. Histogram을 통한 동공 부위 검출, 3. 동공(파란 부분)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 결정 과정을 통해 동공 검출이 이루어진다 [5].

#### 3.3 안구 모델링

안구 모델링은 Pupil Capture의 World video에

서 좌측 단축키에 C라고 표시된 Calibration 플러그인 클릭을 통해 안구 모델링을 진행할 수 있다.

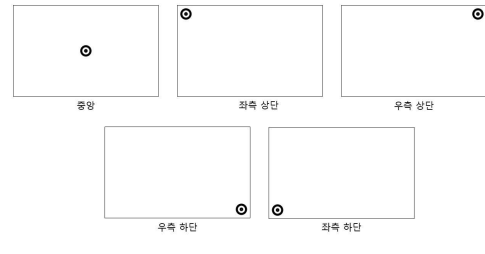


그림 5. 스크린 캘리브레이션  
Fig. 5. Screen calibration

안구 모델링은 Screen calibration, Single calibration, Natural Features 3가지 방식이 있는데 본 연구에서는 기본으로 설정된 Screen calibration 방식을 사용하여 응시점을 월드 카메라 좌표 축으로 매핑한다. 그림 5는 Screen calibration을 이용한 응시점 매핑 방식으로 중앙, 왼쪽 위, 오른쪽 위, 오른쪽 아래, 왼쪽 아래 순서대로 총 5단계로 마커가 나타난다. 마커를 응시하면 양쪽 눈을 촬영하는 2개의 Eye 카메라에서 촬영된 동공의 위치를 파악하여 캘리브레이션을 수행한다. 각 단계에서 중앙이 빨간색인 마커가 나타나는데 마커를 응시하면 빨간색에서 초록색으로 변하고 일정 시간이 지난 후에는 사라진다. 이러한 과정을 5번 반복하면 안구 모델링이 완료된다 [6].

#### 3.4 응시점 추정

안구 모델링 후에 양안 주시선 확장 및 교차점 추정 작업이 들어가는데 이를 통해 실시간 교차점을 추적하며 안구 운동 잡음, 유의하지 않은 응시점 형성, 눈의 깜빡임 등을 제거를 통한 최종 응시/응시 깊이 값을 추정하게 된다.

성공적으로 안구 모델링 및 삼차원 응시점 추정이 끝나고 World video를 확인해보면 그림 6과 같이 노란색 빈 원안에 동공의 현재 응시점을 보여주는 빨간색 원을 확인할 수 있으며 빨간색 원 위에는 파란색으로 보정되지 않은 소프트웨어에서 제공하는 응시 깊이

값이 숫자로 표기된다.



그림 6. 삼차원 응시점 추정 결과  
Fig. 6. 3D gaze estimation results

## 4. 제안된 방법

### 4.1 Deep Neural Network

Artificial Neural Network(ANN)는 인간의 중추 신경계 구조에 착안하여 물체를 인식하는데 데이터를 입력받고 조정할 수 있는 가중치의 집합인 학습 알고리즘에 의해 학습한다 [7]. 그 중 Deep Neural Network(DNN)는 입력층(Input layer)과 출력층(Output layer) 사이에 많은 수의 은닉층(Hidden layer)들로 이루어진 Artificial Neural Network(ANN)의 일종으로 비선형문제를 해결하는 데 사용되는 기계학습 방법이다 [8]. 각 노드는 각자 다른 특징을 추출하며 설계되어 여러 가지 복잡한 특징을 추출할 수 있으며, 층마다 다른 층위의 특징을 학습할 수 있다 [9]. 이러한 DNN에 관한 연구는 기계, 의학, 농업 등 여러 학계에서 활발하게 진행되고 있다. [10,11,12]

### 4.2 훈련 데이터 수집

DNN 모델에 사용할 응시 깊이에 관한 훈련 데이터 수집을 위해 거리는 단거리부터 증장거리까지 mm 단위로 300mm, 900mm, 1,500mm, 2,000mm, 3,000mm, 4,000mm, 5,000mm, 6,000mm, 7,000mm, 8,000mm, 9,000mm, 10,000mm 총 12개의 거리를 설정하였다. 그림 7과 같은 객체를 앞에서 설정된 거리에 위치시켜서 사용자에게 해당 객체를 응시하게 함으로써 설정된 거리마다 소프트웨어를 통해 출력되는 깊이 값들을 획득하여 훈련 데이터를 수집하였다.

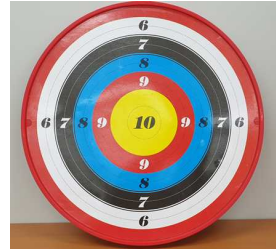


그림 7. 응시 대상 객체  
Fig. 7. Target object

### 4.3 DNN 모델 설계

그림 8과 같이 설계된 DNN 모델은 실시간으로 추정되는 하나의 응시 깊이 값이 보정될 수 있도록 Input Layer와 Output layer는 각각 1개의 노드로 구성하였다. Hidden layer는 3개의 층으로 첫 번째 층은 16개 노드, 두 번째 층은 4개 노드, 세 번째 층은 2개의 노드로 구성하였다.

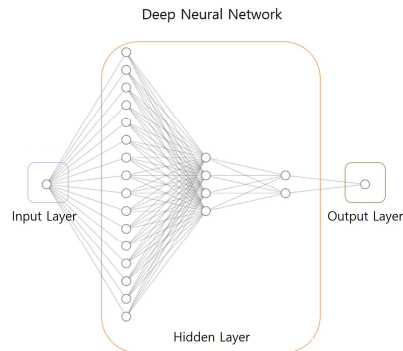


그림 8. 심층 신경망 모델  
Fig. 8. Deep Neural Network Model

Hidden layer의 첫 번째 층에서 활성화 함수(Active Function)는 입력값이 0보다 작으면 0으로 출력, 0보다 크면 입력값 그대로 출력하는 ReLU 함수를 사용하였고 최적화 기법(Optimizer)은 Momentum과 RMSProp의 장점을 결합한 Adaptive Moment Estimation(Adam)과 손실함수(Loss Function)는 Mean Squared Error(MSE)를 사용하였다.

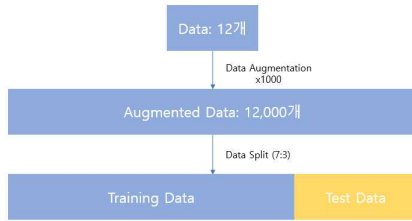


그림 9. 데이터 증강

Fig. 9. Data Augmentation

다양한 유형의 학습 데이터를 학습시키기 위해 데이터양을 그림 9와 같이 증강하였다. 획득한 훈련데이터 12개에 각각 평균 0과 표준편차가 1인 표준정규분포로부터 1,000번 샘플링된 난수들을 더하여 총 12,000개의 샘플 데이터로 증강하였으며, 증강된 데이터를 7:3 비율로 훈련 데이터(8,400개)와 검증 데이터(3,600개)로 무작위로 나눴다.

## 4. 실험 결과

### 4.1 모델 학습 및 검증

모델 학습 및 검증을 위해 하이퍼 파라미터를 Epoch는 30, Batch size는 10으로 설정하여 훈련 데이터(8,400개)를 통해 모델을 학습하고 검증 데이터(3,600개)로 모델 검증을 하였다.

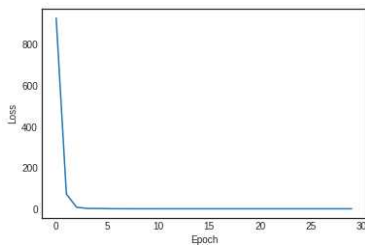


그림 10. 모델 학습 결과

Fig. 10. The result for model training

그림 10은 모델 학습 결과이다. 처음 학습했을 시 Loss 값은 800이 넘는 수치를 보였지만 7번째 학습 이후로는 Loss 값이 1 이하로 내려가고 9번째 학습 이후로는 0.3 미만으로 수렴하는 것을 확인할 수 있었다.

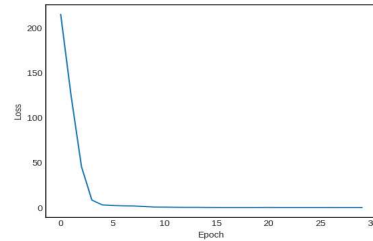


그림 11. 모델 검증 결과

Fig. 11. The result for model validation

그림 11은 검증 데이터를 통한 모델 검증 결과이다. 결과를 확인해보면 처음 학습했을 시 Loss 값은 200이 넘는 수치를 보였지만 10번째 학습 이후로는 Loss 값이 1 이하로 내려가고 15번째 학습 이후로는 그림 10과 같이 0.3 미만으로 수렴하는 것을 확인할 수 있었다.

### 4.2 성능 평가

성능 평가를 위해 훈련 데이터를 수집하였을 때와 같은 환경으로 300mm에서 10,000mm 사이의 응시 깊이 데이터 12개를 수집하였다.

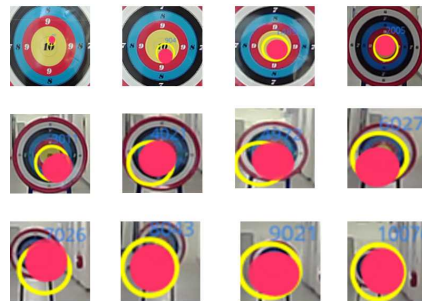


그림 12. 실제 깊이에 따른 보정된 응시 깊이 시각화

Fig. 12. Visualization of corrected gaze depth at actual depth

그림 12와 같이 실제 깊이에서의 보정된 응시 깊이 값을 확인해보면 300mm에서 10,000mm까지 거리별 297mm, 904mm, 1,485mm, 2,005mm, 3,011mm, 4,021mm, 4,972mm, 6,027mm, 7,026mm, 8,043mm, 9,021mm, 10,076mm로 실제와 비슷한 응시 깊이 값을 획득할 수 있었다.

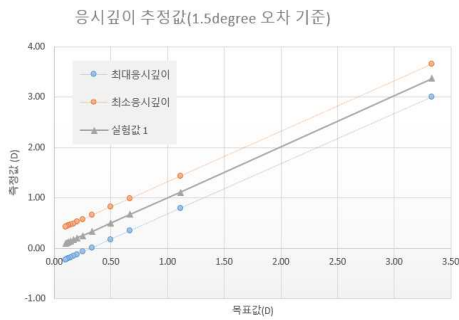


그림 13. 보정된 응시 깊이 성능 평가  
 Fig. 13. Performance evaluation for corrected gaze depth

300mm에서 10,000mm 범위에서 응시 깊이 허용 범위 (visual angle 관점에서 1.5도 이내)를 만족하는 것을 그림 13을 통해 확인할 수 있었다.

### 5. 결론

응시 추적 기기와 소프트웨어를 통해 동공 검출 및 안구 모델링, 응시점 추정을 할 수 있었다. 하지만 응시 깊이 값은 소프트웨어에서 제공하는 값만 나올 뿐 직접 보정해야 할 필요성이 있었다. 본 논문에서는 소프트웨어에서 제공한 응시 깊이 값을 획득하여 DNN 모델을 설계한 후 학습시킨 뒤 성능 평가를 진행하였다. 성능 평가 결과 지정한 거리 300mm, 900mm, 1,500mm, 2,000mm, 3,000mm, 4,000mm, 5,000mm, 6,000mm, 7,000mm, 8,000mm, 9,000mm, 10,000mm를 각각 297mm, 904mm, 1,485mm, 2,005mm, 3,011mm, 4,021mm, 4,972mm, 6,027mm, 7,026mm, 8,043mm, 9,021mm, 10,076mm로 실제 응시 깊이와 비슷한 응시 깊이 값을 획득할 수 있었다.

향후 작업으로 본 실험에서 획득한 결과를 토대로 응시 깊이 값에 따른 ROI(관심 영역) 생성 및 YOLO 알고리즘 적용을 통한 추가적인 실험을 진행할 예정이다.

### REFERENCES

- [1] J. H. Kim, "Eye-tracking and Perception", Review of Architecture and Building Science, vol. 58, no. 9, pp 21-26, 2014.
- [2] E. S. Seo, "Mobile Eye Tracker and for Use of the Same for Revitalizing Studies on Eye Tracking", The Journal of the Korea Contents Association, vol. 16, no. 12, pp 10-18, 2016.
- [3] 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.
- [4] J. H. Mun, D. W. Shin, 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.
- [5] 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.
- [6] Pupil Captrue | Pupil Labs, Pupil Labs, last modified Jan 27, 2023, accessed Jan 31, 2023, docs.pupil-labs.com/core/software/pupil-capture/
- [7] S. H. Park, T. J. Jeon, S. H. Kim, S. Y. Lee, J. W. Kim, "Deep learning based symbol recognition for the visually impaired.", Journal of Korea Institute of Information, Electronics, and communication Technology, vol. 9, no. 3, pp. 249-256, 2016.
- [8] D. J. Park, B. W. Kim, J. Y. Jeong, W. A. Chang, "Deep Neural Network Based Prediction of Daily Spectators for Korean Baseball League : Focused on Gwangju-KIA Champions Field.", Smart Media Journal vol. 7, no. 1, pp. 16-23, 2018.
- [9] H. J. Moon, K. B. Min, "Introduction to Deep Learning in 4<sup>th</sup> Industrial Revolution", Journal of The Korean Association For and Spatial Structures vol. 17, no. 4, pp. 4-9, 2017.

- [10] THAMBA, Narendiranath Babu, et al., "Application of EMD, ANN and DNN for self-aligning bearing fault diagnosis.", Archives of acoustics, vol. 43, no. 2, pp. 163-175, 2018.
- [11] Gawehn, Erik, et al., "Advancing drug discovery via GPU-based deep learning.". Expert opinion on drug discovery, vol. 13, no. 7, p. 579-582, 2018.
- [12] Koklu, Murat, Ilkay Cinar, and Yavuz Selim Taspinar., "Classification of rice varieties with deep learning methods.", Computers and electronics in agriculture, vol. 187, 2021.

---

## 저자약력

---

### 한 석 호(Seok-Ho Han)

[정회원]



- 2023년 2월 : 원광대학교 디지털 콘텐츠공학과(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