딥러닝 기반의 눈 랜드마크 위치 검출이 통합된 시선 방향 벡터 추정 네트워크

주희영 고민수 송혁
한국전자기술연구원
{jhycj, kmsqwet, hsong}@keti.re.kr

Deep Learning-based Gaze Direction Vector Estimation Network Integrated with Eye Landmark Localization

Joo, Hee Young Ko, Min Soo Song, Hyok
Korea Electronics Technology Institute

요약

본 논문은 눈 랜드마크 위치 검출과 시선 방향 벡터 추정이 하나의 딥러닝 네트워크로 통합된 시선 추정 네트워크를 제안한다. 제안하는 네트워크는 Stacked Hourglass Network[1]를 백본(Backbone) 구조로 이용하며, 크게 랜드마크 검출기, 특징 맵 추출기, 시선 방향 추정기라는 세 개의 부분으로 구성되어 있다. 랜드마크 검출기에서는 눈 랜드마크 50개 포인트의 좌표를 추정하며, 특징 맵 추출기에서는 시선 방향 추정을 위한 눈 이미지의 특징 맵을 생성한다. 그리고 시선 방향 추정기에서는 각 출력 결과를 조합하고 이를 통해 최종 시선 방향 벡터를 추정한다. 제안하는 네트워크는 UnityEyes[2] 데이터셋을 통해 생성된 가상의 합성 눈 이미지와 랜드마크 좌표 데이터를 이용하여 학습하였으며, 성능 평가는 실제 사람의 눈 이미지로 구성된 MPIIGaze[3] 데이터 셋을 이용하였다. 실험을 통해 시선 추정 오차는 0.0396 MSE(Mean Square Error)의 성능을 보였으며, 네트워크의 추정 속도는 42 FPS(Frame Per Second)를 나타내었다.

1. 서론

최근 증강 현실 기술의 발전에 힘입어 시선 추정 기술이 활용될 수 있는 분야가 점점 확장되고 있다. 기존 전통적인 패턴 분석 기반의 시선 추정 기술은 조명, 머리의 위치, 안경 착용과 같은 변화에 강인하지 못하다는 한계가 있었다. 최근에는 딥러닝 기술을 기반으로 다양한 환경에서의 학습을 통하여 변화에 강인한 시선 추정 기술들이 소개되고 있다. 시선 추정(Gaze Estimation)이란 사용자의 시선 방향을 포함한 시선 정보를 추정하는 것을 말한다[4]. 시선 정보를 바탕으로 개인의 시각적 관심을 파악할 수 있기 때문에 시선 추정 기술은 인간-컴퓨터 상호작용, 마케팅 등 다양한 분야에서 활용되어 왔다. 최근에는 사용자의 시선 정보를 콘텐츠 산업에 활용하고자 하는 움직임이 커지고 있으며 이에 따라 그 중요성이 점점 커지고 있다[5].

본 논문에서 제안하는 네트워크는 눈 랜드마크 좌표와 시선 방향 벡터가 하나의 네트워크에서 추론되는 통합된 구조를 갖는다. 인간 포즈 추정 기법에 자주 활용되는 딥러닝 기반의 모델 Stacked Hourglass Network를 백본 네트워크로 이용한다. 눈 영역 이미지가 제안하는 네트워크를 통과하면 이미지 공간에서의 50개의 눈 랜드마크의 좌표와 3차원 구면 좌표계에서의 정규화된 시선 방향 벡터가 추정된다. 추정된 벡터는 좌표 변환된 후 화살표 선과 함께 시각화된다.

2. 관련 연구

시선 추정(Gaze Estimation) 연구는 특징-기반(Feature-based), 모델-기반(Model-based), 외형-기반(Appearance-based)이라는 세 가지의 큰 흐름으로 발전해왔다[5].

특징-기반 방법은 인간의 눈에 대해 수작업으로 얻은 특징 벡터(Handcrafted Feature Vector)를 시선 추정에 활용한다[5][6]. 모델-기반 방법은 사람의 안구를 두 개의 구가 교차하는 형태로 모델링 하는 데에서 출발한다[7]. 3차원 공간에 안구를 배치하고 이 보다 크기가 작은 가상의 구가 안구와 적정 범위 내에서 교차하며 움직이는 것으로 눈동자의 움직임을 모델링하는 것이다. 외형 기반 방법은 눈 영역에서 추출한 특징을 추출하여 동공의 위치를 검출하는 방법으로 딥러닝 기반의 시선 추정 기법들이 여기에 속한다고 볼 수 있다[5]. 특히 Stacked

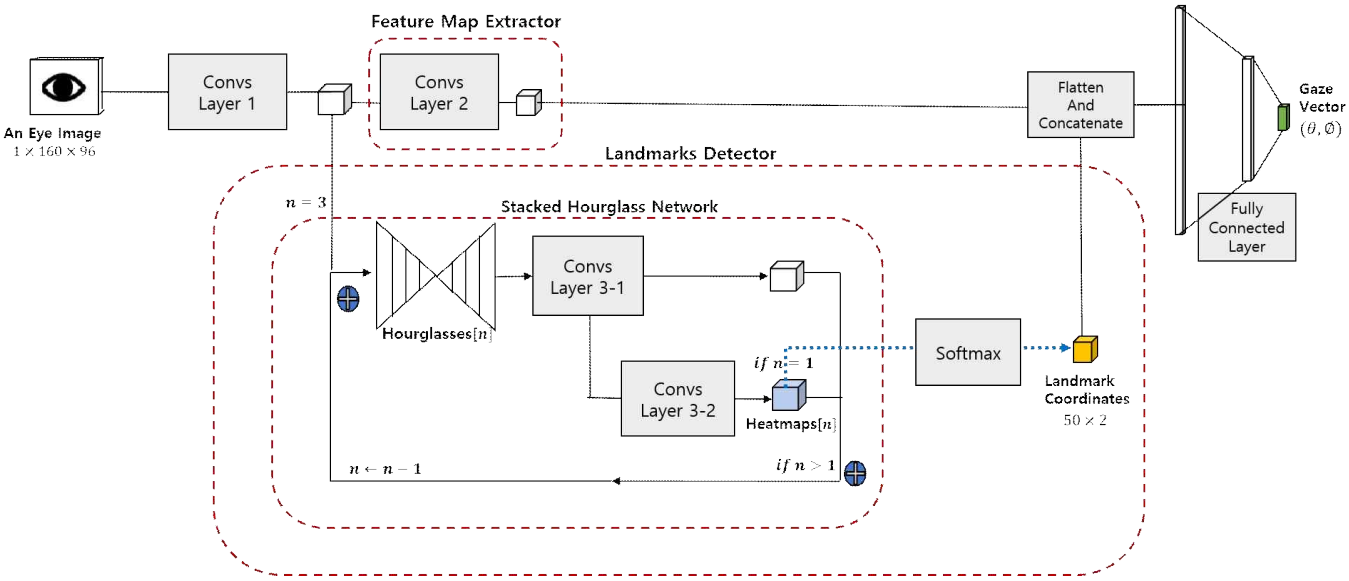

그림 1. 제안하는 시선 방향 벡터 추정 네트워크 구조(Architecture).

Hourglass Network 기반 딥러닝 기반 연구에는 S. Park et al.의 연구[8]가 있다. 이는 3개의 Hourglass Network를 쌓은 구조를 기반으로 눈 랜드마크를 검출한 후, SVR(Support Vector Regressor)[9]을 적용해서 시선 벡터를 얻는다.

본 연구는 눈 랜드마크 검출과 시선 방향 벡터가 하나의 네트워크에서 추론된다는 점에서 이전의 연구와 차별점이 있다. 제안하는 기법은 모델-기반 방법에서 사용한 안구 모델을 기반으로 모델링 된 시선 방향 벡터를 딥러닝 네트워크를 기반으로 추론한다는 점에서 모델-기반과 외형-기반의 방법론을 모두 적용하였다고 볼 수 있다.

3. 제안하는 기법

본 논문은 인간 포즈 추정에 자주 활용되는 Hourglass Network를 백본 네트워크로 하여 눈 랜드마크 위치와 시선 방향 벡터를 추정할 수 있는 모델을 제안한다. 백본 네트워크는 재귀 구조를 갖는 Hourglass Network 3개를 쌓은 Stacked Hourglass Network이다. 제안하는 모델은 눈 영역 이미지에 대해 이미지 공간에서의 랜드마크 좌표와, 구면 좌표계에서 정규화된 시선 방향 벡터를 추정한다. 추정된 결과를 직교 좌표계에서의 벡터로 변환한 후 이를 홍채 중심점을 시작점으로 하는 화살표 선으로 시각화한다.

제안하는 네트워크는 그림 1이 나타내는 바와 같이, 흐름 구조상 랜드마크 검출기(Landmark Detector)와 특징 맵 추출기(Feature Map Extractor), 시선 방향 벡터 추정기(Gaze Direction Vector Estimator)라는 세 개의 부분으로 구성되어 있다. 랜드마크 검출기는 히트맵(Heatmap)과 랜드마크 좌표를 추정하며 특징 맵 추출기는 시선 방향 벡터 추정을 위한 눈 이미지의 특징 맵을 생성한다. 시선 방향 벡터 추정기는 각 출력 결과를 픽셀-단위 덧셈(Pixel-wise Addition) 연산을 적용하여 최종적으로 정규화된 시선 방향 벡터 (θ, ϕ)를 추정한다. 이 때, 각 추정 결과인 히트맵, 랜드마크 좌표, 시선 방향 벡터에 대한 손실 값은 평균 제곱 오차(Mean Square Error)로 계산하며 최종 손실 함수는 이들의 선형 결합으로 설정하였다.

4. 실험 결과

4-1. 학습

제안하는 네트워크 학습에 사용한 데이터 셋은 UnityEyes[2]로서 실제 사람의 눈을 모방한 합성(Synthesized) 눈 이미지 데이터이다. 학습용 데이터 30791장, 검증용 데이터 3849장을 사용하였다. 옵티마이저(Optimizer)는 Adam(Adaptive Moment Estimator)[10]을 사용하였으며, 초기 학습률(Learning Rate) 4×10^{-4}에 대해 매 25 에포크(Epoch)마다 0.1배씩 감소시키며 이를 조정하였다.

4-2. 성능 평가

제안하는 네트워크의 성능 평가에는 MPIIGaze[3] 데이터 셋을 사용하였다. 이 데이터 셋은 제약조건이 없는(Unconstrained) 실제 사람의 눈 이미지를 수집한 것이다. 제안하는 네트워크는 실제 사람의 눈이 아닌 합성 이미지에 대해서만 학습이 이루어졌기 때문에 MPIIGaze 데이터 셋에 대한 성능 평가 결과는 모델의 강인성(Robustness)을 보여준다고 할 수 있다.

MPIIGaze데이터 셋 37767장에 대한 모델 성능 지표는 정답(Label) 시선 방향 벡터에 대한 MSE를 사용하였으며 0.0396을 보였다. 추론 시간은 FPS 42를 나타내었다. 이는 S. Park et al.의 연구[8]에서 제시한 오차 0.046보다 13.9% 감소된 수치이다.

5. 결론

본 논문은 눈 랜드마크 위치 검출과 시선 방향 벡터를 추정하는 딥러닝 기반 네트워크를 제안한다. 제안하는 기법은, 랜드마크 좌표를 추정하는 단계까지만 딥러닝 학습을 수행하였던 기존 연구[8]와는 달리, 랜드마크 위치 검출과 시선 방향 벡터 추정을 하나의 네트워크로 한 번에 학습이 가능하다는 점에서 기존의 연구와 차별점이 있다. 이러한 통

합된 구조를 갖는 네트워크의 손실함수는 랜드마크 좌표와 시선 방향에 대한 추정 오차를 모두 반영한다. 따라서 제안하는 기법으로 학습된 시선 방향 벡터 추정 모델은 기존의 기법보다 더 높은 성능을 보이는 것으로 판단된다.

ACKNOWLEDGMENT

본 논문은 2021년 과학기술정보통신부 비대면 비즈니스 디지털혁신 기술개발사업[과제번호 2020-0-01982]의 지원을 받아 수행한 결과입니다.

참고문헌

[1] A. Newell et al., "Stacked hourglass networks for human pose estimation," European conference on computer vision, pp.483-499, 2016.

[2] E. Wood et al., "Learning an appearance-based gaze estimator from one million synthesised images." In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, pp. 131-138. 2016.

[3] X. Zhang et al., "MPIIGaze: Real-world dataset and deep appearance-based gaze estimation," IEEE transactions on pattern analysis and machine intelligence Vol.41, pp.162-175, 2017.

[4] A. Kar et al., "A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms," IEEE Access, Vol.5, pp.16495-16519, 2017.

[5] 조성현, “시선 추적 기술의 소개“, 전자공학회지, Vol. 45, pp.23-32, 2018.

[6] L. Sesma et al., “Evaluation of pupil center-eye corner vector for gaze estimation using a web cam,“ Proceedings of the Symposium on Eye Tracking Research and Applications, pp.217-220, 2012.

[7] C. Nitschkec et al., “Display-camera calibration using eye reflections and geometry constraints,“ Computer Vision and Image Understanding, Vol.115, pp.835-853, 2011.

[8] S. Park et al., "Learning to find eye region landmarks for remote gaze estimation in unconstrained settings," Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, pp.1-10, 2018.

[9] M. Awad et al., "Support vector regression," Efficient learning machines, pp.67-80, 2015.

[10] DP. Kingma et al., "Adam: A method for stochastic optimization," arXiv preprint, 2014.