눈 특징점 탐지 기반의 실시간 시선 추적 방법에 관한 연구

이영근, 김준식, 유지상, 권순철*

광운대학교

yklee1308@gmail.com, wnstlr5602@naver.com, jsyoo@kw.ac.kr, *ksc0226@kw.ac.kr

Real-Time Gaze Estimation Based on Eye Landmarks Detection

Youngkeun Lee, Junsik Kim, Jisang Yoo, Soonchul Kwon*

Kwangwoon Univ.

요 약

본 논문은 눈 특징점을 탐지하고 탐지된 특징점들을 기반으로 하여 실시간으로 시선을 추적하는 방법을 제안한다. 본 연구는 특징점 추출에 특화된 인공 신경망과 서포트 벡터 회귀를 사용하여 정확하고 효율적으로 눈 특징점을 탐지하고 시선 벡터를 예측한다. 또한, 데이터 증강을 위한 합성 눈 이미지 데이터셋 구축 및 전처리 과정과 EAR 도입을 통한 오탐 감소를 활용하여 약 2.69°의 피치 오차, 약 2.21°의 요 오차의 성능을 보인다.

Ⅰ. 서 론

시선 추적 방법은 사람이 흥미를 보이는 시각적 요소들을 파악하는 데 필수적인 정보를 제공한다는 점에서 컴퓨터 비전의 가장 핵심적인 분야 중 하나이다. 본 논문에서는 눈 특징점 탐지를 기반으로 한 실시간 시선 추적 방법을 제안한다. 신체 특징점 추출에 우수한 성능을 보이는 HRNet[1]을 사용하여 눈 특징점들의 위치를 예측한 후, 서포트 벡터 회귀(Support Vector Regression)[2]를 사용하여 특징점들의 분포를 기반으로 시선 벡터를 예측한다. 인공 신경망 및 서포트 벡터 회귀 모델을 학습시키는 과정에서는 실제 이미지에 가까운 합성 눈 이미지 데이터셋인 UnityEyes[3]를 사용한다. 최종적으로, 눈이 닫혀 있을 시에 시선이 추적되는 에러를 줄이기 위하여 EAR(Eye Appearance Ratio)[4]을 도입한다. 측정된 EAR 값이 설정된 값보다 작을 경우, 시선 벡터와 홍채 중심 좌표를 출력하지 않음으로써 오탐을 줄여 시선 추적 정확도를 향상시킨다.

Ⅱ. 본론

본론에서는 정확하고 사용자 변화에 강인한 실시간 시선 추적 방법을 제안한다. 눈 특징점 탐지 기반의 실시간 시선 추적 방법의 전체 흐름도는 그림 1과 같다. 먼저, 카메라로부터 전달받은 눈 프레임이 인공 신경망인 HRNet에 입력된다. HRNet은 상향식 경로와 하향식 경로의 효과적인 융합을 통해 눈 특징점 예측을 위한 특징맵을 추출한다. 추출된 특징맵의 각 채널에서는 히트맵 방식의 눈 특징점들의 위치에 대한 예측이 이루어지고, 예측된 특징점의 좌표들은 서포트 벡터 회귀를 통한 시선 벡터를 예측하는 데 사용된다. 최종적으로 시선 추적 시스템은 시선 벡터와 홍채 중심 좌표를 출력하기 이전에 EAR 값을 계산하고 사전에 설정한 값과 대소를 비교함으로써 예측 결과의 출력 여부를 결정한다.

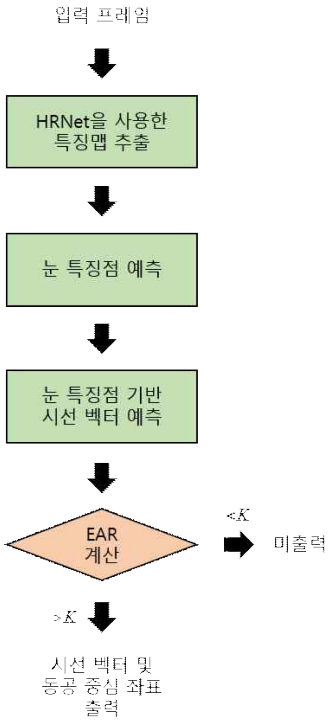

그림 1. 실시간 시선 추적 방법의 전체 흐름도

HRNet[1]은 상향식 경로와 하향식 경로 모두에 중점을 두어 특징맵들을 융합하는 방식을 사용하여 정확도를 높이는 동시에 부동 소수점 연산량(FLOPS)을 절반으로 줄이는 효과를 가져온다. HRNet은 매 단계에서 저해상도 특징맵들을 생성하는 동시에 고해상도 특징맵에 1×1 컨볼루션을 적용하여 유지함으로써 작은 영역에 대한 특징들도 강화해나간다. 그림 2는 residual block을 생략한 시선 추적 방법에 적용하는 HRNet 구조를 나타낸다. 실시간 동작을 위해 가벼운 모델인 HRNet-W32를 사용하며, 가장 고해상도 특징맵부터 저해상도 특징맵까지의 각 채널 수는 차례대로 32, 64, 128, 256으로 설정한다. 출력으로는 가장 해상도가 높은 특징맵을 출력하고, 최종 특징맵의 채널 수는 눈 특징점의 개수인 50으로 설정한다.

HRNet으로부터 추출한 출력 특징맵의 채널은 각각 눈 특징점 위치의 히트맵을 나타낸다. 총 50개의 채널 중 16개는 눈 가장자리 점들의 좌표에 해당하고, 32개는 홍채 가장자리 점들의 좌표에 해당한다. 나머지 2개의 채널은 각각 눈 중심의 좌표와 홍채 중심의 좌표를 나타낸다. 각 히트맵에서 가장 높은 값을 갖는 지점이 눈 특징점이 존재할 확률이 가장 큰 점을 의미하므로 최종적으로 해당 점의 좌표를 예측값으로 출력한다. 인공 신경망의 학습 과정에서는 각 채널에 눈 특징점을 중심으로 하고, 1의 표준편차를 갖는 2D 가우시안 분포를 ground truth로 입력한다. 출력 히트맵

과 ground truth 간의 손실 함수로는 L2 loss인 평균 제곱 오차(Mean Square Error)를 적용한다.

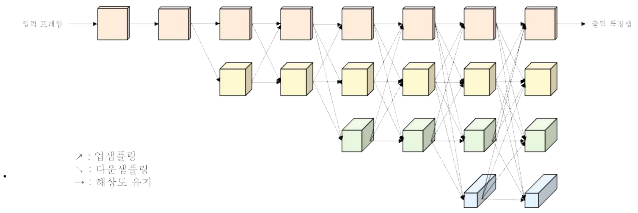

그림 2. 시선 추적 방법에 적용하는 HRNet 구조

본 연구에서는 appearance-based 방법의 인공 신경망을 사용하여 특징 맵을 추출하고, 예측된 눈 특징점 좌표들에 feature-based 방법을 사용하여 시선을 추적한다. 본 연구에서 적용하는 feature-based 방법은 다음과 같다. 먼저, 히트맵을 통해 예측된 눈 특징점 좌표들을 눈 중심 좌표를 중심으로 하고, -1부터 1의 값을 갖도록 정규화한다. 정규화된 좌표들은 서포트 벡터 회귀(Support Vector Regression)[2] 모델에 입력되고, 모델은 3차원의 시선 벡터를 예측한다. 서포트 벡터 회귀 모델의 학습 과정에서 ground truth로는 시선 벡터의 피치(pitch, θ)와 요(yaw, ϕ)를 입력한다.

시선 추적을 위한 인공 신경망을 학습시키기 위한 데이터셋으로 UnityEyes[3]로 생성한 합성 눈 이미지 데이터셋를 사용한다. 그림 3은 UnityEyes로 생성한 합성 눈 이미지의 예와 눈 특징점 좌표를 나타낸다. 눈 특징점 좌표는 총 55개로, 16개의 눈 가장자리, 7개의 caruncle, 32개의 홍채 가장자리 좌표들을 포함한다. 본 연구에서는 caruncle 좌표들을 제외한 눈 및 홍채 가장자리 좌표들만을 사용한다. 또한, 눈 중심 좌표 및 홍채 중심 좌표를 가장자리 좌표들의 평균으로 계산 및 추가하여 눈 특징점 위치 히트맵의 ground truth로 사용한다. 합성 눈 이미지는 640×480의 크기로 생성한 후, 사용하는 카메라의 해상도에 맞추어 눈 중심 좌표를 중심으로 크롭한다.

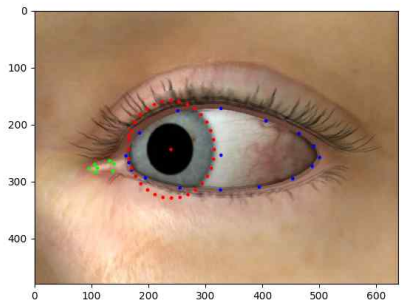

그림 3. UnityEyes로 생성한 합성 눈 이미지 및 눈 특징점 좌표

사용자의 눈이 닫혀 있을 시에 시선을 추적하는 오탐이 발생하는 문제점을 해결하기 위하여 EAR(Eye Aspect Ratio)[4]을 도입한다. EAR은 눈 가장자리의 가로 양 끝점 간의 거리와 세로 양 끝점 간의 거리의 비율로 눈의 닫힌 정도를 측정하는 데 사용된다. 눈이 완전히 열려 있을 시의 EAR은 사용자에 따라 크게 변하기 때문에 본 시선 추적 시스템에서는 프로그램 실행 시 약 2초간 사용자의 EAR 값들을 수집한 후, 중앙치(median)의 0.3배 한 값을 대소 비교 값으로 설정한다. 이후, 측정된 EAR이 설정 값보다 작을 경우, 눈이 닫혀 있다고 판단하여 예측 결과를 출력하지 않는다.

표 1은 본 시선 추적 시스템의 피치와 요의 평균 오차를 나타낸다. 본 시스템은 피치는 약 2.69°의 오차를, 요는 약 2.21°의 평균 오차를 보이며 상하 움직임이 비하여 좌우 움직임에 보다 우수한 성능을 보인다.

표 1. 시선 추적 시스템의 피치와 요의 평균 오차

	Pitch (θ)	Yaw (ϕ)
Avg. Error	2.69°	2.21°

그림 4는 시선 추적 시스템의 실제 눈 이미지에 대한 눈 특징점 탐지 및 시선 벡터 예측 결과를 나타낸다. 16개의 눈 가장자리 점들은 빨간색으로, 32개의 홍채 가장자리 점들은 파란색으로, 2개의 중심 좌표들은 흰색 및 초록색 점으로 각각 표시되어 있다. 또한, 3D 시선 벡터는 평면에 정사영된 2D 벡터로 시각화되었으며, 그림 4에 노란색으로 표시되어 있다.

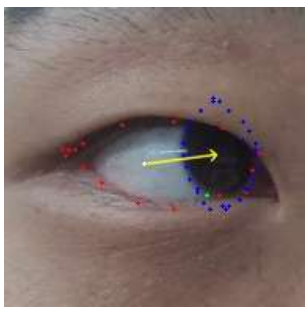

그림 4. 실제 눈 이미지에 대한 눈 특징점 탐지 및 시선 벡터 예측 결과

Ⅲ. 결론

본 논문에서는 눈 특징점 탐지를 기반으로 하여 실시간으로 시선을 추적하는 방법을 제안하였다. 본 연구의 시선 추적 시스템은 약 2.69°의 피치 평균 오차를, 약 2.21°의 요 평균 오차의 성능을 보였다. 다양한 학습 방법 및 손실 함수를 적용하여 시스템의 성능을 비교하고, 많은 스케일에 대한 합성 눈 이미지 데이터셋을 추가 구축한다면 시선 추적 정확도를 더욱 향상시킬 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2020R1F1A1069079)

참 고 문 헌

[1] Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15 - 21 June 2019; pp. 5686 - 5696.

[2] Smola, A.J.; Scholkopf, B. A tutorial on support vector regression. Stat. Comput. 2004, 14, 199 - 222.

[3] Wood, E.; Baltrušaitis, T.; Morency, L.P.; Robinson, P.; Bulling, A. Learning an appearance-based gaze estimator from one million synthesised images. In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, Charleston, SC, USA, 14-17 March 2016; pp. 131 - 138.

[4] Cech, J.; Soukupova, T. Real-time eye blink detection using facial landmarks. Cent. Mach. Perception, Dep. Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague, 2016, 1-8.