

rPPG를 이용한 감정 인식

남유상^{1*}, 이정환¹, 손량희², 권새힘², 이현태³, 박철수¹

광운대학교 컴퓨터공학과, ㈜엠마헬스케어

Emotion Recognition with rPPG

Yusang Nam^{1*}, Jeonghwan Lee¹, Ryanghee Sohn², Sachim Kwon², Hyuntae Lee³, Cheolsoo Park¹

Computer Engineering, Kwangwoon University, Korea, Emma Healthcare Co. LTD, Korea

*usang0821@naver.com

Abstract

With the emergence of deep learning, significant advancements have been witnessed in the field of computer vision. Particularly, Convolutional Neural Networks (CNNs) have emerged as models bringing substantial performance improvements over traditional algorithms in computer vision. Moreover, CNNs have greatly influenced the accuracy enhancement of Facial Expression Recognition (FER). The utilization of CNNs in FER tasks has demonstrated remarkable improvements on benchmark datasets. However, these models have been limited to recognizing facial expressions, overlooking other physiological signals obtainable from video. Additionally, studies addressing variations in facial expression intensity across individuals or scenarios with limited facial expressions have been overlooked. This study proposes a method to enhance the accuracy of existing FER systems by leveraging remote-Photoplethysmography (rPPG), a physiological signal obtainable from video.

1. 연구 배경

딥러닝의 발전과 Convolutional Neural Networks(CNN)의 등장은 컴퓨터 비전에 지대한 영향을 끼쳤다. 특히 Facial Emotion Recognition (FER)은 CNN의 등장과 딥러닝의 발전 이전에 비해 크게 발전했다^[1]. 이러한 FER 모델들은 영상으로부터 얻을 수 있는 생체신호를 고려하지 않았고, 그로 인해 모델의 설계 단계에서부터 한 장의 이미지에서 감정을 분류하는 형태를 취한다. 이 때문에 영상 속 사람 별 감정에 따른 표정 변화에 그 성능이 좌우될 수밖에 없다는 의존성이 있다. 한편, 감정 인식 모델에 영상을 입력으로 사용할 경우 각 프레임에서 기존처럼 감정 인식을 수행할 수 있을 뿐만 아니라, 영상에서 추가적인 생체신호 remote-Photoplethysmography(rPPG)를 얻을 수 있다. 영상에서 획득한 rPPG는 표현으로 드러나지 않는 감정 상태를 유추할 수 있는 지표 중 하나가 될 수 있다.^[2] 이러한 근거를 바탕으로 본 연구에서는 딥러닝 모델을 이용하기 전에 설명 가능한 신호처리 기술에 기반한 연구를 선행함으로써 설명 가능한 End-to-End FER&rPPG 감정 분류 딥러닝 모델의 기반을 쌓고자 한다.

2. 연구 방법

연구를 위해서는 분류해야 할 감정 종류를 설명하는 모델이 전제되어야 한다. 이를 위해 이 연구에서는 총 2가지 축, Valence와 arousal으로 각 감정들을 총 4분면으로 나눈 모델을 참고하였다^[3]. 본 연구에서는 PPG가 아니라 rPPG를 사용하므로, 접촉식 센서를 통해 값을 얻는 PPG보다 노이즈가 많은 데이터를 수집할 것이라 예상된다. 따라서 본 연구의 트러블 슈팅을 위해서는 보다 간단한 모델부터 시작하는 것이 알맞다고 판단, 해당 모델에서 제안하는 8가지 감정 중 1사분면에 해당하는 행복, 2사분면에 해당하는 분노, 3사분면에 해당하는 우울, 4사분면에 해당하는 중립, 총 4가지의 감정으로 각 사분면당 하나의 대표 감정으로 클래스를 단순화시켰다.

본 연구의 목적이 CNN 기반 FER 모델의 성능을 보다 다양한 환경에서 높이는 것이므로, 최종적으로 앙상블 모델을 구현하였다. 기존 SOTA FER 모델은 DDAMFN모델^[1]

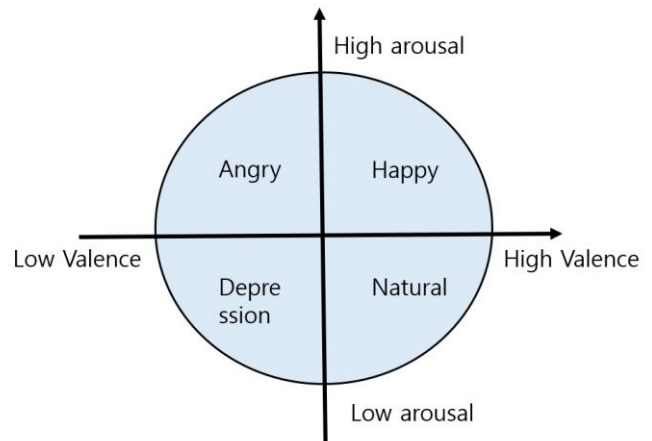


그림 1. 감정 모델

이 사용되었다.

영상에서 rPPG를 추출하는 과정은 카메라 성능, 인물의 움직임, 주변 조명 등에 큰 영향을 받는다. 해당 노이즈로 인한 rPPG 자체의 정확도 저하를 막기 위해 영상 내 노이즈 정도를 미리 판단하고 노이즈가 많은 픽셀을 미리 제외하는 전처리 과정이 OMIT 알고리즘을 제안한 논문^[8]의 파이프라인을 구현하였다. 다만 본 연구에서는 프레임을 격자 모양으로 분할하는 것이 아닌 mediapipe^[9] 라이브러리의 face mesh 기능을 사용했다. 해당 기능이 제공하는 478개의 랜드마크를 눈과 입술을 제외한 총 621개의 폴리곤으로 나누어 각 폴리곤에 속하는 픽셀의 RGB 신호를 사용하였다.

PPG 신호를 사용한 감정 분류 방식은 크게 두 가지가 있다. 첫 번째는 생체신호 값 자체를 그대로 입력으로 사용하는 방법^[3]과 생체신호에서 feature를 추출해 그 값을 사용하는 방법^{[4][5][6][7]}이다. PPG와 rPPG는 그 기반 원리가 같으므로, PPG를 이용한 감정 인식 방법이 rPPG를 이용한 감정 인식에도 적절하다는 가정 하에 본 연구에서는 feature를 추출해 그 값을 사용하는 방법^[4]을 참고해 진행하였다. feature로는 총 2가지 종류가 있으며, Time parameters와 Frequency parameters이다. Time

parameters에는 min-max 정규화된 전체 신호의 Time window 6초, Stride 1초로 계산된 평균, 분산과 전체 데이터의 1,2차 차분 유클리디안 거리, KFD가 포함되었다. Frequency domain에는 Time window 6초로 계산된 STFT, Spectral Coherence Function이 포함된다.

실험을 위한 데이터셋은 ㈜엠마헬스케어와 함께 자체 수집한 데이터이다. 해당 실험은 각 감정에 해당되는 영상을 시청하는 동안 PPG를 측정하며, 스스로 느낀 감정의 정도를 설문지에 1~5 사이의 점수로 표기하는 방식으로 진행되었다. 총 20명의 남녀를 대상으로 수집되었다.

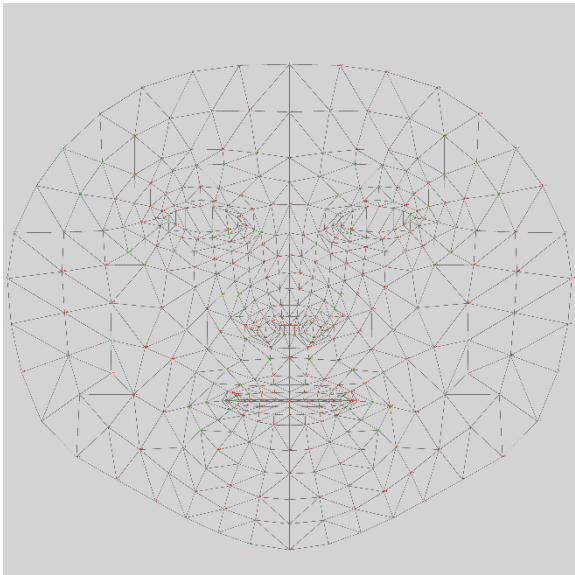


그림 2. mediapipe face mesh

FER 모델과 rPPG를 이용한 감정 분류 모델 각각의 결과는 SVM과 Random Forest 모델을 사용해 최종 결과를 출력하였다.

3. 연구 결과

이상적인 rPPG 값에 해당하는 감정 분류 결과를 우선적으로 분석했으며, 위 감정 모델의 4분면에 해당하는 각 감정의 Precision, Recall, F1-score는 아래와 같다.

	Precision	Recall	F1-score
Happy	0.99	0.92	0.95
Angry	0.95	0.99	0.97
Depression	0.37	0.33	0.35
Natural	0.53	0.56	0.54

앞서 제안한 연구 방법에 따른 FER-rPPG 양상불 모델의 결과와 FER 모델 단독의 성능 결과는 추후 연구를 통해 비교될 수 있을 것이다.

4. Acknowledgements

이 연구는 2024년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (P0017124, 2024년 산업혁신인재성장지원사업) 또한 이 연구는 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원을 받아 수행된 연구임 (No.TS-2023-00229800, 비대면 환경 사용자의 정량적 감정-감성 평가모델 개발 및 양방향 디지털 콘텐츠 적용 상용화) 또한 이 연구는 광운대학교 생명윤리심의위원회의 생명윤리심의결과 승인받았음 (7001546-202301201-HR(SB)-011-03)

5.참고 문헌

[1] Zhang, Saining, et al. "A Dual-Direction Attention Mixed Feature Network for Facial Expression Recognition." Electronics 12.17 (2023): 3595.

[2] Egger, Maria, Matthias Ley, and Sten Hanke. "Emotion recognition from physiological signal analysis: A review." Electronic Notes in Theoretical Computer Science 343 (2019): 35-55.

[3] Lee, Min Seop, et al. "Fast emotion recognition based on single pulse PPG signal with convolutional neural network." Applied Sciences 9.16 (2019): 3355.

[4] Gouizi, Khadidja, F. Bereksi Reguig, and Choubeila Maaoui. "Emotion recognition from physiological signals." Journal of medical engineering & technology 35.6-7 (2011): 300-307.

[5] Goshvarpour, Atefeh, and Ateke Goshvarpour. "Evaluation of novel entropy-based complex wavelet sub-bands measures of PPG in an emotion recognition system." Journal of medical and biological engineering 40.3 (2020): 451-461.

[6] Rakshit, Raj, V. Ramu Reddy, and Parijat Deshpande. "Emotion detection and recognition using HRV features derived from photoplethysmogram signals." Proceedings of the 2nd workshop on Emotion Representations and Modelling for Companion Systems. 2016.

[7] Rakshit, Raj, V. Ramu Reddy, and Parijat Deshpande. "Emotion detection and recognition using HRV features derived from photoplethysmogram signals." Proceedings of the 2nd workshop on Emotion Representations and Modelling for Companion Systems. 2016.

[8] Casado, Constantino Alvarez, and Miguel Bordallo López. "Face2PPG: An unsupervised pipeline for blood volume pulse extraction from faces." IEEE Journal of Biomedical and Health Informatics (2023).

[9] Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." arXiv preprint arXiv:1906.08172 (2019).