

랜드마크 기반 앙상블 네트워크를 이용한 얼굴 표정 분류

안 영 은¹ · 이 지 민² · 김 민 구¹ · 반 성 범^{3*}

¹조선대학교 IT연구소 연구교수

²조선대학교 SW중심대학사업단 초빙객원교수

^{3*}조선대학교 전자공학과 교수

Classification of Facial Expressions Using Landmark-based Ensemble Network

Young-Eun An¹ · Ji-Min Lee² · Min-Gu Kim¹ · Sung-Bum Pan^{3*}

¹Research Professor, IT Research Institute, Chosun University, Gwangju 61452, Korea

²Invited guest professor, SW Convergence Education Institute, Chosun University, Gwangju 61452, Korea

^{3*}Professor, Department of Electronic Engineering, Chosun University, Gwangju 61452, Korea

[요 약]

얼굴 표정 분류 연구는 얼굴의 구성 요소, 특징 추출 등과 관련되어 꾸준히 연구되어 왔다. 본 연구에서도 이러한 변화를 이용하여 효과적인 얼굴 표정 분류를 하기 위해 랜드마크 기반 앙상블 네트워크를 이용하여 얼굴 표정을 분류하는 알고리즘을 제안한다. 먼저 전처리 과정으로 얼굴 전체 영상에서 얼굴 부분만 추출한 후, 얼굴 이미지만 이용하여 특징 정보를 구성한다. 그리고 얼굴 이미지에서 랜드마크 정보를 이용하여 새로운 랜드마크 기반 특징 정보를 추출한다. 얼굴 전체 이미지 특징 정보와 랜드마크 기반 특징 정보를 각각의 CNN 네트워크에 학습한 후, 앙상블 학습을 통해서 얼굴 표정을 분류한다. 실험 결과, 제안한 알고리즘은 기존 얼굴 이미지만 이용하여 표정 분류를 한 기술보다는 1.14%, 랜드마크 특징 정보만 이용하여 표정 분류를 한 기술보다는 0.57% 더 좋은 분류 성능을 나타내는 것을 확인할 수 있다.

[Abstract]

Research on facial expression classification has been steadily studied in relation to facial components and characteristics extraction. In this study, an algorithm for classifying facial expressions using a landmark-based ensemble network is also proposed to effectively classify facial expressions using these changes. First, only the face part is extracted from the entire face image through the preprocessing process, and then feature information is constructed using only the face image. And new landmark-based feature information is extracted using landmark information from the face image. After learning face-wide image feature information and landmark-based feature information on each CNN network, facial expressions are classified through ensemble learning. As a result of the experiment, it can be seen that the proposed algorithm shows 1.14% better classification performance than the technology that classified facial expressions using only existing facial images and 0.57% better classification performance than the technology that classified facial expressions using only landmark feature information.

색인어 : 얼굴 표정 분류, 앙상블 네트워크, 랜드마크, 얼굴 부분 특징, 딥러닝

Keyword : Classification of facial expressions, Ensemble network, Landmark, Face feature, Deep learning

<http://dx.doi.org/10.9728/dcs.2022.23.1.117>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 16 November 2021; **Revised** 17 December 2021

Accepted 17 December 2021

***Corresponding Author; Sung-Bum Pan**

Tel: +82-62-230-6897

E-mail: sbpan@chosun.ac.kr

I. 서론

얼굴 표정 분류는 학문적, 상업적 잠재력이 매우 크기 때문에 컴퓨터 비전 및 인공 지능 분야에서 중요한 연구 주제이다. 얼굴 표정 분류는 시각적 표현이 커뮤니케이션의 주요 정보 채널 중 하나이기 때문에 얼굴 이미지를 이용하는 연구에 중점을 둔다. 얼굴 표정은 타인의 의도를 이해하는 데 도움을 주는 인간 의사소통의 중요한 요소이다. 일반적으로, 사람들은 얼굴 표정과 목소리 톤을 사용하여 기쁨, 슬픔, 분노와 같은 감정 상태를 추론한다. 연구[1-2]에 따르면, 언어 구성요소는 인간 의사소통의 1/3을 전달하며, 비언어 구성요소는 2/3를 전달한다. 여러 비언어적 요소들 중에서, 감정적인 의미를 전달하는 얼굴 표정은 의사소통의 주요 정보 통로 중 하나이다. 따라서 얼굴 표정에 대한 연구가 지난 수십 년 동안 지각 및 인지 과학뿐만 아니라 감성 컴퓨팅과 컴퓨터 애니메이션에도 적용되면서 많은 관심을 얻고 있는 것은 당연하다[3-5].

따라서 본 논문에서는 이러한 비언어 의사소통의 주요 방법 중 하나인 얼굴 표정을 이용하여 감정 상태를 추론할 수 있도록, 랜드마크 기반 앙상블 네트워크를 이용하여 얼굴 표정을 분류한다. 먼저 얼굴 전체 이미지에서 얼굴 부분만 검출하여 특징 벡터를 구성한다. 그리고 얼굴 부분만 검출한 이미지에서 랜드마크를 추출한 후, 랜드마크 중 얼굴 표정의 특징에 따라 눈, 눈썹과 코 그리고 입에 관한 정보를 이용하여 새로운 랜드마크 특징 벡터를 추출하고, 얼굴 전체 특징벡터와 새로운 랜드마크 특징들을 각각 CNN 네트워크에 학습 시킨 후 앙상블 학습을 통해 얼굴 표정을 분류한다.

2장에서는 기존 얼굴 표정 관련 연구에 대해서 고찰하고, 3장에서는 본 논문에서 제안한 알고리즘에 대해서 설명한다. 4장에서는 2장에서 언급한 기존 연구와 본 논문에서 제안한 알고리즘의 비교 분석을 통해 얼굴 표정 분류 정확도를 검증하고, 5장에서 결론을 맺는다.

II. 기존 얼굴 표정 분류 연구

Jyoti Kumari 등은 그림 1에서와 같이 얼굴 전체 이미지를 전처리를 통해서 얼굴 부분만 추출해낸다. 그런 후 눈, 입 등의 얼굴 구성요소를 추출해서 얼굴 특징 정보를 구성하고 이를 이용해서 얼굴 표정을 분류한다[6, 7]. 이 알고리즘은 얼굴 전체 이미지를 이용한 특징으로 표정을 분류하기 때문에 조명이나 카메라 각도에 많은 영향을 받는다.

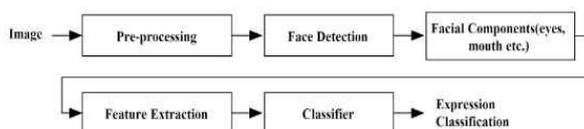


그림 1. 얼굴 표정 분류 블록도
Fig. 1. Facial expression classification block diagram

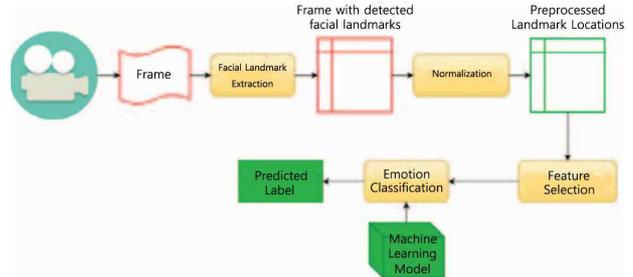


그림 2. 얼굴 표정 분류 블록도
Fig. 2. Facial expression classification block diagram

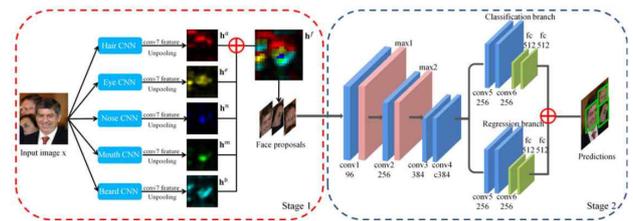


그림 3. Faceness-Net 블록도
Fig. 3. The pipeline of the baseline Faceness-Net

Binh T. Nguyen 등은 그림 2에서와 같이 실시간 영상에서 얼굴을 추출한 후 랜드마크를 추출한다. 그리고 정규화 과정을 통해서 랜드마크를 이용하여 얼굴의 특징 정보를 추출하고, 이를 이용하여 얼굴 표정을 분류한다[8, 9]. 이는 전처리를 통해 얼굴 이미지를 정규화하고, 얼굴의 특징점을 이용하여 얼굴 표정 분류를 하기 때문에, 얼굴 전체 이미지를 이용하여 검색한 것 보다는 더 좋은 검색 성능을 보이나, 슬픈 감정에 대해서는 다른 감정들에 비해 상대적으로 낮은 인식률을 보인다.

Shuo Yang 등은 그림 3에서와 같이 얼굴 전체 영상에서 얼굴 주요 특징 정보 중 헤어, 눈, 코, 입 등에 관한 특징을 따로 추출하여 딥러닝 네트워크에 학습한 후, 특징 표현이 잘 된 이미지를 이용하여 새로운 얼굴 이미지를 구성하고, 이를 다시 CNN 네트워크에 학습시켜서 얼굴을 인식한다[10]. 이 알고리즘은 얼굴이 절반 이상 가려진 경우에도 잘 검출해내나, 얼굴 부분 특징을 이용해서 새로운 얼굴을 구성하기 때문에, 얼굴 표정 분류에 대해서는 검색 성능을 증명하기 어렵다.

III. 랜드마크 기반 앙상블 네트워크를 이용한 얼굴 표정 분류

본 논문에서는 얼굴 표정 분류를 위해 랜드마크 기반 앙상블 네트워크를 이용한 얼굴 표정 분류 기법을 제안한다. 그림 4는 본 논문에서 제안한 알고리즘의 블록도이다.

먼저, 정규화 과정으로 Viola와 Jones에 의해 처음 제시된 Haar cascade 기반 형태인식 아이디어[11-13]를 이용하여 얼굴 표정 이미지에서 얼굴 영역만 검출하여 특징 벡터로 구성한다.

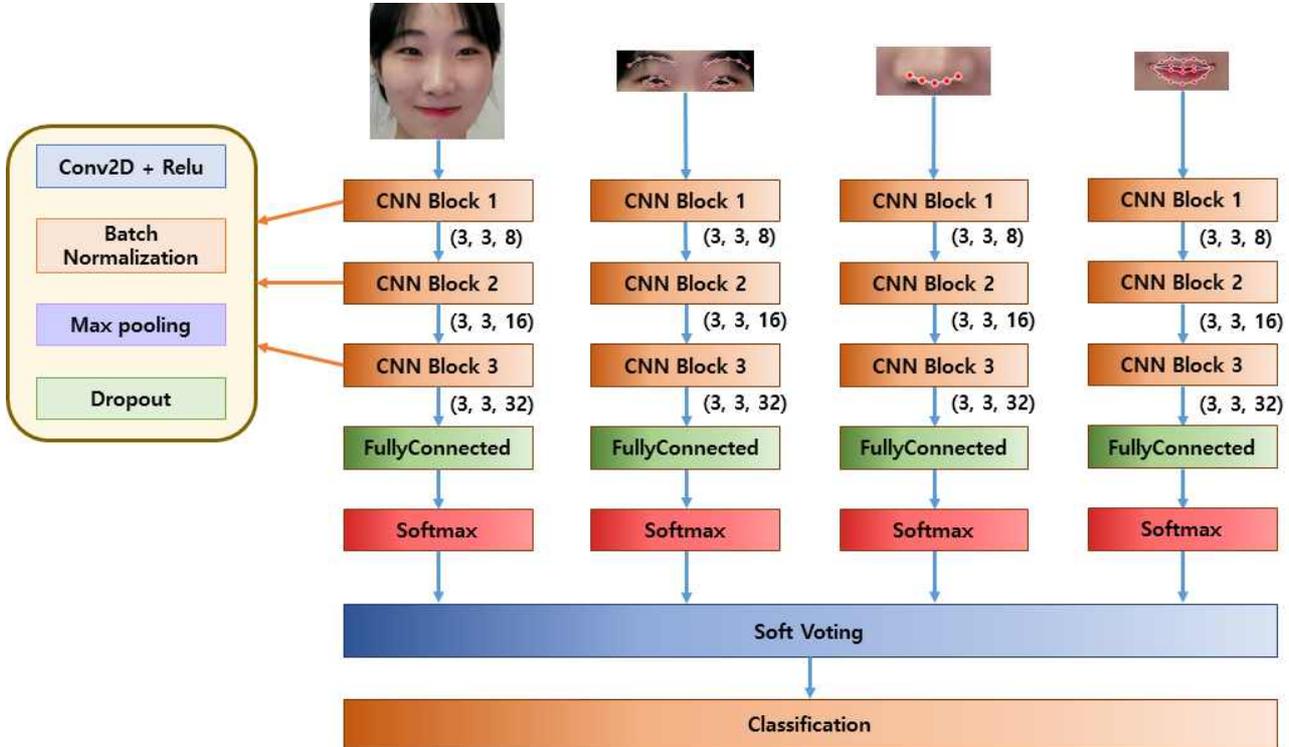


그림 4. 제안한 알고리즘 블록도
 Fig. 4. Proposed algorithm flowchart

검출된 얼굴 영역에서 얼굴 표정을 구분할 수 있도록 랜드마크를 추출하고[8], 추출된 랜드마크 중 얼굴 표정에 따른 특징에 의해서, 눈, 눈썹과 코 그리고 입에 관련된 랜드마크들의 랜드마크 간 거리 정보를 분석하여 새로운 특징 벡터를 구성한다. 얼굴 영역의 특징 벡터와 눈, 눈썹과 코 그리고 입에 관련된 랜드마크를 이용하여 구성된 새로운 특징 벡터들을 각각의 CNN 네트워크로 학습한 후, 이를 소프트 보팅 방법을 이용하여 얼굴 표정을 분류한다.

그림 4에서 특징 정보로 사용한 랜드마크는 추출된 68개의 랜드마크 중 평온함, 기쁨, 슬픔, 두려움의 표정 분류를 하기 위해, 1970년대 이후 수십 년 동안 사회 및 심리학 연구에서 감정의 7가지 주요 표정으로 보편성을 확립한 폴 에크만이 제안한 표 1의 얼굴 표정에 따른 특징[14]에 따라서 눈, 눈썹, 코 그리고 입에 관련된 랜드마크만 이용한다.

표 1의 얼굴 표정 특징에 따라서 눈썹이 올라가는지 눈이 커지는지 입이 벌어지는지 등의 얼굴 표정 특징 정보를 고려하기 위해서, 그림 5와 같이 랜드마크들의 거리 정보를 추출한 후 특징 벡터로 구성한다.

얼굴 전체 이미지와 그림 5에서 추출한 랜드마크 특징 정보들을 각각의 CNN 네트워크를 이용하여 학습하고 얼굴 표정을 분류한 후, 앙상블 학습을 통해서 가장 높은 비율을 차지하고 있는 얼굴 표정을 분류해낸다. 여기서 사용한 앙상블 학습 방법은 소프트 보팅 방법이다. 소프트 보팅 방법은 그림 6에서와 같이 훈련 데이터를 샘플 데이터로 분류한 후 각각의 샘플 데이터를 각각의 네트워크(분류기 1~분류기 4)에 학습시킨다.

표 1. 얼굴 표정 특징[14]

Table 1. Facial expression characteristics

Facial Expression	Characteristics
Happiness	<ul style="list-style-type: none"> Lip corners pulled down Slight frown Outer corners of eyebrows pulled down Smaller eyes
Sadness	<ul style="list-style-type: none"> Lip corners raised Mouth slightly stretched Slightly smaller eyes
Fear	<ul style="list-style-type: none"> Mouth slightly stretched Eyebrows pulled down Eyebrows pulled to middle together

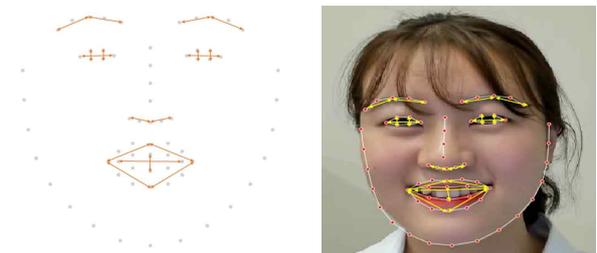


그림 5. 랜드마크 거리 정보 추출 예
 Fig. 5. Example of extracting landmarks distance information

샘플 데이터는 얼굴 전체 이미지와 얼굴의 특징 중 눈, 눈썹, 코 그리고 입에 관련된 랜드마크간 거리 정보를 이용하였다. 그리고 네트워크는 CNN을 이용하여 학습한다. 그런 다음

학습 결과 중 첫 번째와 두 번째로 높은 비율을 가지고 있는 클래스를 추출한 후, 각각의 학습 결과 비율 값을 이용하여 평균을 구하고 그 결과로 가장 높은 비율을 가지고 있는 클래스로 분류해낸다[15].

그림 6으로 예를 들어 설명하면, 분류기 1, 3, 4는 Class 1을 출력 결과로 나타내고 분류기 2는 Class 2를 출력 결과로 나타낸다. 하지만 최종 출력 결과를 1순위 Class로 바로 분류하지 않고, 분류기 1에서 Class 1의 확률이 0.7, 분류기 2에서 Class 1의 확률이 0.2, 분류기 3에서 Class 1의 확률이 0.8, 분류기 4에서 Class 1의 확률이 0.9가 나오므로 이 4개의 분류기가 나타낸 Class 1에 대한 확률을 모두 더해서 평균을 구한 0.65가 최종 Class 1의 출력값이 되고, 이 평균 확률값이 가장 높은 Class가 최종 출력값이 된다. 이렇게 소프트 보팅 방법을 이용해서 얼굴 표정 이미지와 랜드마크 특징 정보를 CNN 네트워크에 학습하여 나온 결과로 최종 얼굴 표정을 분류해낸다.

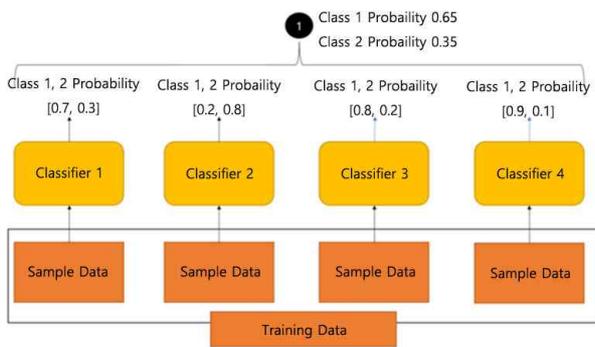


그림 6. 소프트 보팅 앙상블 학습 방법
Fig. 6. Soft voting ensemble's prediction

IV. 실험 및 분석

자체 구축한 얼굴 표정 이미지 데이터베이스는 건강한 피험자 50명을 대상으로 조명과 배경이 한정된 공간에서 평온함, 기쁨, 슬픔, 두려움의 4가지 감정 상태별로 각각 30초간 취득한다. 이 영상은 해상도 HD(1280×720), 초당 30프레임, MPEG-4 파일 형식으로 저장이 되고, 이 영상에서 얼굴 표정 이미지를 추출하여 표 2와 같이 데이터셋을 구성한다. 그림 7은 얼굴 이미지 데이터베이스 중 피험자 별 4가지 감정 상태를 나타낸 샘플이다. 이렇게 자체 구축한 얼굴 표정 이미지 데이터베이스를 MATLAB을 이용하여 실험한다.

표 2. 얼굴 이미지 데이터셋
Table 2. Facial image dataset

Emotion	Neutral	Happiness	Fear	Sadness
Number of data	8,132	7,705	7,678	7,595



그림 7. 얼굴 이미지 데이터베이스 샘플
Fig. 7. Face image database sample

표 2의 데이터셋을 이용하여 본 논문에서 제안한 랜드마크 기반 앙상블 네트워크를 이용한 기법과 기존 얼굴 이미지만 이용하여 얼굴 표정을 분류하는 기법[7], 랜드마크 특징 정보만을 이용하여 얼굴 표정을 분류하는 기법[8]을 비교 분석한다.

그 결과, 표 3과 같이 제안한 알고리즘의 성능이 얼굴 이미지만 이용하여 얼굴 표정 분류를 한 알고리즘보다는 1.14%, 랜드마크를 이용하여 얼굴 표정 분류를 한 알고리즘 보다는 0.57% 더 높은 것을 확인할 수 있다.

그리고 그림 8과 같이 테스트 결과 클래스별로 올바르게 검색된 결과와 잘못 검색된 결과를 비교해 보면, 제안한 알고리즘이 기존 두 알고리즘 보다 육안으로 봤을 때 비슷한 표정을 나타내는 기법과 평온함 감정에 대해서 오분류율이 더 낮은 것을 확인할 수 있다.

표 3. 성능 측정 결과
Table 3. Accuracy results

Algorithm	Accuracy(%)
Classifying facial expressions using only facial images[7]	95.86
Classification of facial expressions using landmarks[8]	96.43
Classification of facial expressions using rearrangement landmarks	97.00

표 4. 그룹 테스트 결과
Table 4. Accuracy results of group dataset

Train set	Test set	Accuracy	F1-Score
A, B, C, D	E	97.02	97.08
A, C, D, E	B	96.87	96.97
A, B, D, E	C	97.08	97.15
B, C, D, E	A	97.10	97.20
A, B, C, E	D	96.94	97.03
Average		97.02	97.09

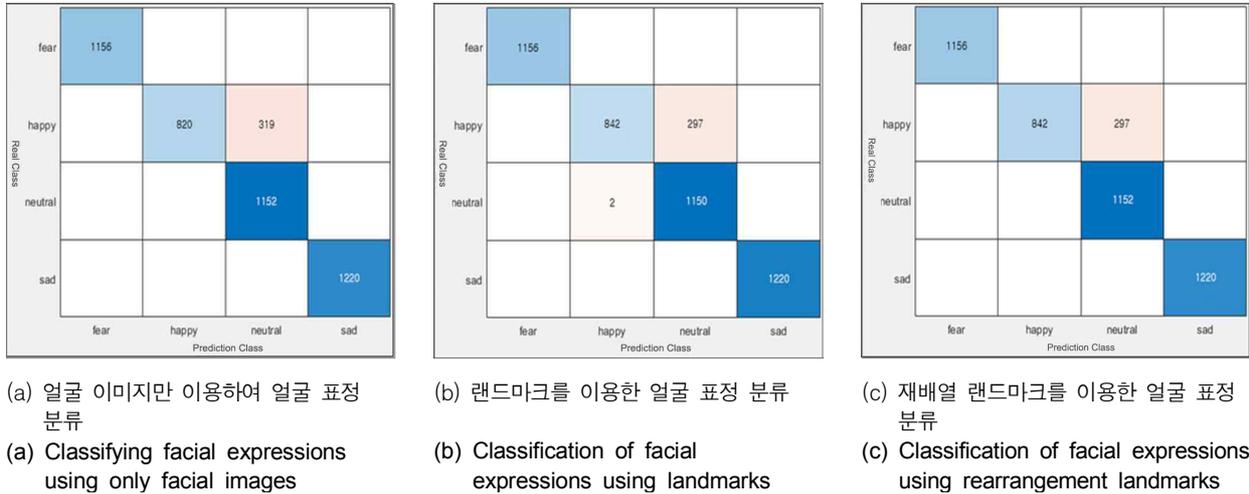


그림 8. 소프트 보팅 앙상블 학습 방법
Fig. 8. Soft voting ensemble's prediction

또한 자체 구축한 데이터베이스를 10명씩 총 5개의 그룹으로 나눈 후, 학습 데이터셋과 테스트 데이터셋으로 구성해서 5-fold cross 검증을 수행한 결과는 표 4와 같다. 그 결과, 가장 높은 성능을 보인 그룹은 97.10%, 가장 낮은 성능을 보인 그룹은 96.87%였고, 평균 97.02%의 성능을 나타내는 것을 확인하였고, 이는 전체 데이터를 이용하여 학습한 것과 비슷한 결과를 보이는 것을 확인하였다.

V. 결론

본 논문에서는 얼굴 표정 분류를 위해 랜드마크 기반 앙상블 네트워크를 이용하는 기법을 제안한다. 먼저, 상반신 이미지에서 전처리를 통해 얼굴만 검출하여 특징 벡터를 구성한다. 얼굴 이미지에서 랜드마크를 추출하여 랜드마크 간 거리 정보로 새로운 랜드마크 기반 특징 벡터를 구성한다. 이렇게 추출된 특징 벡터를 각각의 CNN 네트워크에 학습시킨 후, 앙상블 학습을 이용하여 가장 높은 비율을 차지하고 있는 얼굴 표정을 분류해낸다.

그 결과, 기존 얼굴 이미지만 이용하여 표정 분류를 한 기술보다는 1.14%, 랜드마크 특징 정보만 이용하여 표정 분류를 한 기술보다는 0.57% 더 좋은 분류 성능을 나타내고, 기존 알고리즘 보다 제안한 알고리즘의 오분류율이 더 낮게 나타나므로 본 논문에서 제안한 랜드마크 기반 앙상블 네트워크를 이용한 얼굴 표정 분류 기술을 기초로 더 다양한 특징 정보와 다양한 데이터베이스를 이용한다면, 심리적 건강 상태 예측 뿐만 아니라, 인간의 감정 및 감성 상태에 대한 데이터와 빅데이터 인공지능 알고리즘을 활용해 인간의 감정 및 감성을 반영한 물품추천 서비스, 고객 수요량 예측 등 새로운 비즈니스 시장 창출이 가능할 것이라 생각한다.

감사의 글

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신산업진흥원의 지원을 받아 수행된 헬스케어 AI 융합 연구개발 사업(No. S1601-20-1041)과 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No.NRF-2017R1A6A1A03015496)의 지원을 받아 수행된 연구이다.

참고문헌

- [1] A. Mehrabian, Communication theory 2nd ed. New York, Routledge, 2008.
- [2] K. Kaulard, D. W. Cunningham, H. H. Bülthoff and C. Wallraven, "The MPI facial expression database—A validated database of emotional and conversational facial expressions", *PLOS ONE*, Vol. 7, No. 3, pp. 1-18, March 2012. <https://doi.org/10.1371/journal.pone.0032321>
- [3] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion", *PNAS*, Vol. 111, No. 15, pp. E1454-E1462, April 2014. <https://doi.org/10.1073/pnas.1322355111>
- [4] A. Martinez and S. Du, "A model of the perception of facial expressions of emotion by humans: Research overview and perspectives", *Journal of Machine Learning Research*, Vol. 13, No. 1, pp. 1589-1608, May 2012.
- [5] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild", *2016 IEEE Conference on Computer Vision and*

Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 5562–5570, June 2016.

<https://doi.org/10.1109/CVPR.2016.600>

[6] J. Kumark, R. Rajesh and K.M. Pooja, “Facial Expression Recognition: A Survey”, *Procedia Computer Science*, Vol. 58, pp. 486-491, 2015.,
<https://doi.org/10.1016/j.procs.2015.08.011>

[7] M. J. Lyons, J. Budynek and S. Akamatsu, “Automatic classification of single facial images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 12, pp. 1357-1362, December 1999.
<https://doi.org/10.1109/34.817413>

[8] B. T. Nguyen, M. H. Trinh, T. V. Phan and H. D. Nguyen, “An Efficient Real-Time Emotion Detection Using Camera and Facial Landmarks”, *Seventh International Conference on Information Science and Technology*, Da Nang, Vietnam, pp. 251-255, April 2017.
<https://doi.org/10.1109/ICIST.2017.7926765>

[9] Q. T. Ngoc, S. Lee and B. C. Song, “Facial Landmark-Based Emotion Recognition via Directed Graph Neural Network”, *Electronics*, Vol. 9, No. 5, May 2020.
<https://doi.org/10.3390/electronics9050764>

[10] S. Yang, P. Luo, C. C. Loy and X. Tang, “Faceness-Net: Face Detection through Deep Facial Part Responses”, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, Vol. 40, No. 8, pp. 1845-1859, August 2018.
<https://doi.ieeecomputersociety.org/10.1109/TPAMI.2017.2738644>

[11] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features”, *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR)*, 2001, Kauai, HI, USA, pp. 511-518, April 2003.
<https://doi.org/10.1109/CVPR.2001.990517>

[12] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees”, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1867–1874, September 2014. <https://doi.org/10.1109/CVPR.2014.241>

[13] J. H. Kim and C. S. Won, “Emotion Enhancement for Facial Images Using GAN”, *2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, eoul, Korea (South), December 2020.
<https://doi.org/10.1109/ICCE-Asia49877.2020.9277349>

[14] P. Ekman, “Facial Action Coding System: FACS”, Available:
<https://www.paulekman.com/facial-action-coding-system/>

[15] S. W. A. Sherazi, J. W. Bae and J. Y. Lee, “A soft voting

ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome”, *PLOS ONE*, Vol. 16, No. 6, pp. 1-20, June 2021.

<https://doi.org/10.1371/journal.pone.0249338>



안영은(Young-Eun An)

2004년 : 조선대학교 이학사
2006년 : 조선대학교 대학원 (공학석사)
2010년 : 조선대학교 대학원 (공학박사)

2010년~2014년: 조선이공대학 메카트로닉스공학과 조교수
2014년~2020년: 조선대학교 자유전공학부 조교수
2020년~현재: 조선대학교 IT연구소 연구교수
※관심분야 : 멀티미디어 영상처리, 인공지능, 패턴인식



이지민(Ji-Min Lee)

2002년 : 전남대학교 (이학사)
2011년 : 조선대학교 대학원 (공학석사)
2015년 : 조선대학교 대학원 (공학박사-멀티미디어 신호처리)

2014년~2017년: 아주통신(주) 기업부설연구소 연구소장
2018년~현재: 조선대학교 초빙객원교수
※관심분야 : 멀티미디어 영상처리, 인공지능, 패턴인식 등



김민구(Min-Gu Kim)

2010년 : 조선대학교 공학사
2013년 : 조선대학교 대학원 (공학석사)
2019년 : 조선대학교 대학원 (공학박사)

2019년~현재: 조선대학교 IT연구소 연구교수
※관심분야 : 생체인식, 영상처리, 인공지능, 패턴인식



반성범(Sung-Bum Pan)

1999년 : 서강대학교 대학원 (공학박사)

1999년~2005년: 한국전자통신연구원 정보보호연구단
2005년~현재: 조선대학교 전자공학과 교수
※관심분야 : 바이오인식, 영상처리