

얼굴 랜드마크 거리 특징을 이용한 표정 분류에 대한 연구

Study for Classification of Facial Expression using Distance Features of Facial Landmarks

배진희*, 왕보현*, 임준식*

Jin Hee Bae*, Bo Hyeon Wang*, Joon S. Lim*

Abstract

Facial expression recognition has long been established as a subject of continuous research in various fields. In this paper, the relationship between each landmark is analyzed using the features obtained by calculating the distance between the facial landmarks in the image, and five facial expressions are classified. We increased data and label reliability based on our labeling work with multiple observers. In addition, faces were recognized from the original data and landmark coordinates were extracted and used as features. A genetic algorithm was used to select features that are relatively more helpful for classification. We performed facial recognition classification and analysis with the method proposed in this paper, which shows the validity and effectiveness of the proposed method.

요약

표정 인식은 다양한 분야에서 지속적인 연구의 주제로서 자리 잡아 왔다. 본 논문에서는 얼굴 이미지 랜드마크 간의 거리를 계산하여 추출된 특징을 사용해 각 랜드마크들의 관계를 분석하고 5가지의 표정을 분류한다. 다수의 관측자들에 의해 수행된 라벨링 작업을 기반으로 데이터와 라벨 신뢰도를 높였다. 또한 원본 데이터에서 얼굴을 인식하고 랜드마크 좌표를 추출해 특징으로 사용하였으며 유전 알고리즘을 이용해 상대적으로 분류에 더 도움이 되는 특징을 선택하였다. 본 논문에서 제안한 방법을 이용하여 표정 인식 분류를 수행하였으며 제안된 방법을 이용하였을 때가 CNN을 이용하여 분류를 수행하였을 때보다 성능이 향상됨을 볼 수 있었다.

Key words : facial recognition, emotion recognition, classification, coordinate distance correlation, feature selection

1. 서론

다윈은 표정 연구를 통해 분노, 기쁨 등의 정서

는 동서양을 막론하고 인간이 공통으로 가지고 있는 감정이라는 것을 밝혔다[1]. 표정은 그러한 인간의 감정을 두드러지게 혹은 미세하게 표현하는 요

* Dept. of Computer Engineering, Gachon University

★ Corresponding author

E-mail : jslim@gachon.ac.kr, Tel : +82-31-750-5330

※ Acknowledgment

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(2020R1I1A1A01066599)

* This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-2017-0-01630) supervised by the IITP(Institute for Information & communications Technology Promotion)

Manuscript received Nov. 8, 2021; revised Nov. 30, 2021; accepted Dec. 15, 2021.

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.

소 중 하나이다. 인간은 상대방의 표정을 인식하기 위해 수학적 계산이나 오랜 시간의 분석을 필요로 하지 않는다. 눈, 코, 입 등의 위치나 목소리 등 여러 가지 요소를 통해 바로 파악할 수 있다. 하지만 컴퓨터는 즉각적인 인식이 아닌 누적된 학습을 기반으로 표정을 인식한다. 높은 분류 정확도를 위해 분류에 도움이 되고 의미 있는 특징들을 선택하는 것은 많은 분석과 시간을 필요로 한다.

특징을 추출하고 선택하는 것은 인공지능을 활용한 분류 작업에서 중요하다. 따라서 많은 분류 혹은 예측 연구에서 데이터의 분석에 따른 특징 추출과 선택이 진행된다[2]. 새로운 데이터를 적용해도 예측하는 능력이 향상되기 위해서는 적절한 특징을 추출하는 것이 중요한 역할을 한다. 하지만 추출된 특징 또한 분류에 악영향을 미칠 수 있기 때문에 분석 및 평가를 통해 유의미한 특징을 선별하는 것은 중요하다.

본 연구에서는 명확한 감정을 알 수 있는 표정을 분석하며 입 가장자리 끝의 위치, 눈의 면적, 눈 가장자리 끝의 위치, 광대의 위치 등이 표정 분류에 정확성을 높이는데 기여한다고 분석했다. 이 과정이 임의의 경로를 통해 획득된 모든 얼굴 이미지에 적용될 수 있도록 실험 데이터는 비교적 쉽게 구할 수 있는 일반적인 다양한 인종, 나이, 성별을 가진 얼굴 이미지를 사용하여 실험을 수행하였다.

II. 본론

1. 관련 연구

표정 인식에 대한 분석 및 연구는 인간과 컴퓨터의 상호작용(Human Computer Interaction)뿐만 아니라 컴퓨터 과학, 심리학, 산업 공학, 인간 공학, 로봇 비전 등 다양한 곳에서 주요한 연구 주제로 활발히 연구되고 있다[3, 4, 5, 6, 7]. 또한 표정에 대한 연구는 이미 많은 산업 필드에 적용되고 있다[8, 9].

텍스트, 이미지, 영상 등 얼굴을 이용한 실험은 다양하다. 최근까지도 본 논문의 실험과 유사한 형태로써 이미지를 이용하여 표정을 분류한 실험이 수행되었다[10]. [10]에서 수행된 실험에서는 이미지 분류 성능에서 우수한 Convolutional Neural Network(CNN)을 사용하였다. 데이터를 7가지의 표정으로 분류 했으며 Local Binary Pattern(LBP) feature와

attention 메커니즘을 결합하여 여러 데이터 셋을 사용해 분류 정확성을 입증하였다. 또 다른 연구에서는 연속적인 데이터로 즉각적인 표정을 인식하여 실험을 수행하였으며 새로운 multi-level hidden Markov models(HMMs)을 제안하였다[11].

2. 데이터

모바일 3D graphics 소프트웨어 개발 회사로부터 획득된 이미지를 이용하였다. 모든 데이터는 네이버, 구글, 플릭커, 카카오, Unsplash, Qwant에서 얻었으며 총 48260개의 이미지가 실험에 사용되었다. 라벨링 과정은 대학생 혹은 대학을 졸업한 40명의 성인이 참여했다. 관측자들은 각 이미지를 보고 분노, 행복, 중립, 슬픔, 놀람, 얼굴 아님 중 하나를 선택하였다.

데이터의 신뢰도는 [12]의 논문의 GLAD(Generative model of Labels, Abilities and Difficulties)를 사용해 각 관측자의 전문성, 이미지의 난이도, 각 이미지의 레이블에 대한 신뢰도를 추출했다. 관측된 레이블 결과는 L_{ij} (i 는 표정 인덱스, j 는 이미지 인덱스) 이고, 이미지의 라벨 신뢰도는 Z_j , 관측자의 전문성 값은 α_i , 이미지의 난이도 값은 $1 / \beta_j$ 이다. 신뢰도 도출을 위해 Expectation-Maximization(EM)을 사용하여 매개 변수의 최대 가능성 추정치를 도출했다. EM알고리즘의 실행에서 이미지의 라벨 신뢰도 Z_j 의 초기 값은 5개의 라벨에 각각 0.125로 설정하고, 관측자의 전문성과 이미지 난이도에서 사용하는 α_i 와 β_j 의 초기 값은 1로 설정했다.

EM알고리즘은 Expectation 단계인 E-Step과 Maximization 단계인 M-Step으로 구성된다. E-Step에서는 이미지 j 에 대해 주어진 모든 레이블 값과 이전 단계의 α_i, β_j 값으로 이미지 라벨의 신뢰도 $Z_j \in \{0, 1\}$ 를 계산한다. M-Step은 unconstrained minimization을 사용하여 E-Step에서 계산된 Z_j 값으로 α_i, β_j 의 값을 산출한다. α_i, β_j, Z_j 값이 수렴할 때까지 E-Step과 M-Step을 반복했다. 관측자 전문성 결과값 α 값은 양의 무한대(+ ∞)일 경우 관측자가 항상 올바르게 레이블을 지정한다는 의미이고, 음의 무한대(- ∞)일 경우 관측자가 항상 잘못된 레이블을 지정한다는 의미이다. 신뢰도 결과값이 0인 경우 라벨러가 클래스를 구별 할 수 없음을 의미한다. 관측 영상 신뢰도 결과 $1 / \beta_j$ 값은 양의 무한대(+ ∞)일 경우 이미지를 판단하기 매우 모

호함을 의미하고, 0일 경우 이미지를 판단하기 매우 쉬운 값을 의미한다. 관측 레이블 신뢰도 결과 값은 관측자들이 각 이미지에 대해 분노, 행복, 중립, 슬픔, 놀람, 얼굴 아님에 따라 입력한 값에 대해서 0에서 1사이로 정규화된 신뢰도 값을 의미한다. 따라서 각 이미지의 통계값의 최댓값이 해당 이미지의 카테고리가 되었다. 얼굴 아님 카테고리로 분류된 이미지는 없었다. 그리고 통계의 정규화된 신뢰도가 너무 낮다면 그 데이터는 분류의 명확성을 제시할 수 없기 때문에 실험에 사용하지 않았다. 이러한 이유들로 우리는 10128개의 분노, 13320개의 행복, 13105개의 중립, 6493개의 슬픔, 5214개의 놀람, 총 5개로 라벨링된 결과를 얻을 수 있었다.

3. 제안 방법

(1) 얼굴 탐지

표정을 인식하려면 먼저 사람의 얼굴을 감지해야 한다. 얼굴 감지를 위해 C++ 라이브러리인 dlib를 사용했다. 그림 1처럼 얼굴이 인지된 부분을 원본 이미지 위에 직사각형으로 표시해 성공적으로 얼굴을 발견했다는 것을 알 수 있었다.



Fig. 1. Upper picture is original and lower picture shows detected face.

그림 1. 위쪽은 원본 사진, 아래쪽은 얼굴이 감지된 사진의 예시

(2) 랜드마크 추출

표정을 표현하는 부위는 크게 눈, 코, 입이다. 따라서 Dlib를 사용해 눈, 코, 입, 얼굴 윤곽에 찍히는

68개의 랜드마크가 획득 되었다. 감정에 따라 달라지는 표정을 통해 랜드마크들의 거리 관계를 특징으로 사용하기 위해 모든 랜드마크들을 좌표로 변환하였다.



Fig. 2. Example of landmarked picture.

그림 2. 원본에 랜드마크가 찍힌 사진 예시

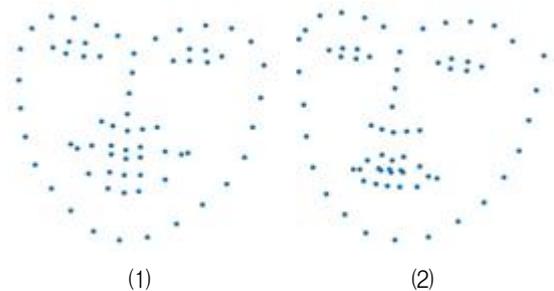


Fig. 3. (1) landmark extracted from Fig 1, (2) landmark of a picture landmark in Sad Category.

그림 3. (1) 그림 1로부터 추출한 랜드마크, (2) 실험에 사용한 슬픔 카테고리 중 하나의 랜드마크 이미지

(3) 특징 추출

각 이미지 마다 추출된 68개의 좌표를 사용해 거리를 계산했다. 랜드마크 사이의 거리는 유클리디안 거리로 계산되었다[14]. 유클리디안 거리 계산식은 식 (1)과 같다.

from 1 st to 2 nd	From 1 st to 3 rd	From 1 st to 4 th	...	From 1 st To 68 th
Already calculated	From 2 nd to 3 rd	From 2 nd to 4 th	...	From 2 nd to 68 th
Already calculated	Already calculated	From 3 rd to 4 th	...	From 3 rd to 68 th
⋮	⋮	⋮	...	⋮
Already calculated	Already calculated	Already calculated	Already calculated	From 67 th to 68 th

Fig. 4. 2278 features : distance between landmark numbers.

그림 4. 2278개의 특징 : 랜드마크의 번호들 간의 거리를 의미

$$d(x,y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad \text{식(1)}$$

모든 랜드마크들은 그림 4처럼 하나의 랜드마크를 기준으로 기준이 되는 랜드마크와 이미 계산한 랜드마크를 제외한 다른 랜드마크와 거리를 계산할 수 있다. 계산된 모든 거리들은 표정 분류에 기반이 되는 특징이며 총 2278개의 특징이 추출되었다.

(4) 특징 선택

2278개의 특징은 랜드마크의 거리의 정보를 가지고 있다. 분류의 정확도를 높이고자 모든 특징들을 사용하지 않고 유의미한 특징을 선별했다. 특징 선택은 유전알고리즘을 사용했다. 유전 알고리즘에서 gene은 0과 1 중 하나이다. 0은 해당 인덱스의 특징이 선택되지 않는다는 뜻이고 1은 해당 인덱스의 특징이 선택된다는 뜻이다. 각각의 데이터 세트에 대해서 분류기로 사용되는 Artificial Neural Network (ANN) 테스트를 시행하여 fitness를 구한다. 하나의 generation 과정은 다음과 같다. 두 개의 데이터 세트를 one-point crossover로 연산한다. 분류에 효과적인 특징들로 구성돼 있을 때 fitness가 높기 때문에 적합도가 높은 chromosome은 크로스오버의 비율을 낮춘다. 데이터 세트에 대한 큰 변동을 주지 않기 위해서, Roulette Wheel 방식은 사용하지 않았다. mutation은 0과 1을 바꾸어 진행한다. 이를 100번 반복하여 최적의 특징 선택과 가장 좋은 적합도를 구한다.

많은 연구들에서 광범위하게 사용된 ANN을 분류기로 사용하였다. 다양한 실험을 통해 증명된 ANN의 성능이 본 실험의 효율적인 학습을 위해

사용되었다[15].

3. 실험 및 결과

학습 데이터와 테스트 데이터의 비율은 각 카테고리마다 랜덤으로 7:3으로 나누어 실험에 사용하였다. 사용된 데이터 세트를 토대로 CNN을 적용했을 때 74.4%의 분류 정확도를 얻었다. 그러나 본 논문에서 제안한 좌표 거리 기반 특징을 사용한 분류 방식은 87.6%의 정확도를 얻었다. 그림 5는 68개의 랜드마크의 위치를 나타낸다.

더 높은 정확도에 기여한 특징들은 총 90개 이다. (45to47, 39to41, 38to42, 44to48, 21to41, 24to48, 20to42, 25to47, 21to40, 24to43, 22to40, 20to41, 25to48, 23to43, 62to63, 10to55, 67to68, 21to42, 62to64, 8to49, 51to53, 66to68, 45to48, 53to64, 51to62, 63to64, 61to65, 39to42, 66to67, 24to47, 44to47, 38to41, 22to28, 65to68, 61to66, 51to52, 23to28, 10to65, 61to67, 9to65, 65to67, 9to61, 8to61, 9to55, 9to49, 34to63, 58to59, 52to63, 52to53, 9to56, 62to65, 57to59, 9to60, 65to66, 49to65, 55to61, 56to61, 60to65, 61to68, 59to65, 57to58, 54to62, 10to56, 22to41, 57to61, 21to39, 8to60, 50to65, 49to66, 61to64, 55to68, 56to59, 58to65, 50to64, 20to37, 54to61, 24to44, 58to61, 31to63, 34to52, 23to48, 49to67, 57to60, 56to60, 55to67, 19to42, 63to65, 50to54, 26to47, 49to56) 유전자 알고리즘을 통해 분류에 기여가 상대적으로 낮은 특징을 줄임으로써 실행 속도와 정확도 향상의 성과를 보였다. 본 알고리즘과 CNN을 비교하여 본 알고리즘이 CNN의 분류 정확도보다 향상됨을 보였다.

III. 결론

본 논문에서는 표정 분류를 위해 얼굴 랜드마크의 거리 계산을 기반으로 특징을 추출하였고 유전자 알고리즘을 이용하여 학습 정확도를 향상시킬 수 있는 특징들을 선택하여 학습을 수행하였다. 본 논문에서 제안한 방식은 CNN과 비교했을 때 13.2%의 정확도 향상을 보여주었다. 주로 눈과 입에 위치한 거리들이 특징으로 선택되었으며 이는 다섯가지의 표정을 분류하는데 가장 효과적이었던 특징이라는 것을 의미한다.

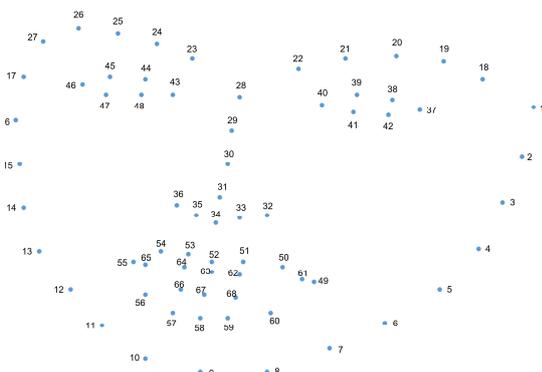


Fig. 5. location of each landmark.

그림 5. 각 랜드마크의 위치

이 실험의 결과는 컴퓨터 과학을 전공하지 않는 심리학, 사회학, 인문학 등 여러 분야에도 함께 융합적 결합 연구가 될 수 있는 가능성을 가지고 있다. 본 논문의 제안 기법은 이미지 뿐만 아니라 컴퓨터가 충분히 인지가 가능한 얼굴의 영상 혹은 real-time data가 주어져도 적용 가능할 것이라 기대한다.

References

- [1] Byun, In-Kyung, and Lee, Jae-Ho. "Facial Expression Research according to Arbitrary Changes in Emotions through Visual Analytic Method," *The Journal of the Korea Contents Association*, vol.13, no.10, pp.71-81, 2013. DOI: 10.5392/JKCA.2013.13.10.071.
- [2] Dash, Manoranjan, and Huan Liu. "Feature selection for classification," *Intelligent data analysis* Vol.1, No.1-4 pp.131-156, 1997.
- [3] Gajarla, V., & Gupta, A. "Emotion detection and sentiment analysis of images," *Georgia Institute of Technology*, pp.1-4, 2015. DOI: 10.1109/ICCDW45521.2020.9318713
- [4] Cohen, Michelle E., and W. J. Carr. "Facial recognition and the von Restorff effect," *Bulletin of the Psychonomic Society*, Vol.6, No.4, pp.383-384, 1975. DOI: 10.3758/BF03333209
- [5] Li, Jiaying, et al. "Facial expression recognition with faster R-CNN," *Procedia Computer Science*, Vol.107, pp.135-140, 2017. DOI: 10.1016/j.procs.2017.03.069
- [6] Bartneck, Christoph, and Michael J. Lyons. "HCI and the face: Towards an art of the soluble." *International Conference on Human-computer Interaction*. Springer, 2007. DOI: 10.1007/978-3-540-73105-4_3
- [7] Martino, L. D.; Preciozzi, J.; Lecumberry, F. "Face matching with an a-contrario false detection control," *Neurocomputing*, Vol.173, pp.64-71, 2016. DOI: 10.1016/j.neucom.2015.02.093
- [8] Di Martino, Luis, et al. "Face matching with an a contrario false detection control," *Neurocomputing*, Vol.173, pp.64-71, 2016.
- [9] Napoléon, Thibault, and Ayman Alfalou.

"Pose invariant face recognition: 3D model from single photo," *Optics and Lasers in Engineering*, Vol.89, pp.150-161, 2017.

DOI: 10.1016/j.optlaseng.2016.06.019

[10] Bendjillali, Ridha Ilyas, et al. "Improved facial expression recognition based on DWT feature for deep CNN," *Electronics*, Vol.8, No.3, pp.324, 2019. DOI: 10.3390/electronics8030324

[11] Cohen, Ira, et al. "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and image understanding*, Vol.91, No.1-2, pp.160-187, 2003.

DOI: 10.1016/S1077-3142(03)00081-X

[12] Whitehill, Jacob, et al. "Whose vote should count more: Optimal integration of labels from labelers of unknown expertise," *Advances in neural information processing systems*, Vol.22, pp.2035-2043, 2009.

[13] Davis E. King. "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research*, Vol.10, pp.1755-1758, 2009.

DOI: 10.5555/1577069.1755843

[14] Danielsson, Per-Erik. "Euclidean distance mapping," *Computer Graphics and image processing*, Vol.14, No.3 pp.227-248, 1980.

DOI: 10.1016/0146-664X(80)90054-4

[15] Ren, Jinchang. "ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging," *Knowledge-Based Systems*, Vol.26, pp.144-153, 2012.

DOI: 10.1016/j.knosys.2011.07.016

BIOGRAPHY

Jin Hee Bae (Member)

2020 : BS degree in Computer Engineering, Gachon University.



Wang Bo Hyun (Member)



1994 : BS degree in Computer Science, Gachon University.
1996 : MS degree in Computer Science, Gachon University.
2008 : PhD degree in Computer Engineering, Gachon University.

Joon Shik Lim (Member)



2000 : BS degree in Computer Science, Inha University.
2005 : MS degree in Computer Science, Alabama University.
2006 : PhD degree in Computer Science, Louisiana State University.