

동공크기 변화신호의 STFT와 CNN을 이용한 2차원 감성분류

이희재[†], 이다빛^{**}, 이상국^{***}

2D Emotion Classification using Short-Time Fourier Transform of Pupil Size Variation Signals and Convolutional Neural Network

Hee-Jae Lee[†], David Lee^{**}, Sang-Goog Lee^{***}

ABSTRACT

Pupil size variation can not be controlled intentionally by the user and includes various features such as the blinking frequency and the duration of a blink, so it is suitable for understanding the user's emotional state. In addition, an ocular feature based emotion classification method should be studied for virtual and augmented reality, which is expected to be applied to various fields. In this paper, we propose a novel emotion classification based on CNN with pupil size variation signals which include not only various ocular feature information but also time information. As a result, compared to previous studies using the same database, the proposed method showed improved results of 5.99% and 12.98% respectively from arousal and valence emotion classification.

Key words: Emotion Classification, Short-Time Fourier Transform, Convolutional Neural Network, Arousal-Valence, Pupil Size Variation Signal

1. 서 론

최근, 인간-컴퓨터 상호작용(HCI) 분야에서 사용자의 욕구를 파악하고 반응하는 인터페이스 연구들이 진행되고 있으며, 대표적으로 사용자 감성분류 기술이 있다. 감성을 측정하는 방법으로는 크게 얼굴표정[1], 제스처[2], 음성신호[3], 생리신호[4], 눈특징[5]을 사용하는 방법과 다양한 생리신호들을 결합하여 특징으로 사용하는 방법[6] 등이 있다. 센서기술이 발전함에 따라 생리신호 기반 감성분류 연구들이 진행되고 있지만, 피부 표면에 센서를 부착하는 센서

기반 감성분류 방법들은 착용이 불편하고, 움직임에 따른 노이즈로 인해 감성분류 성능이 저하되는 문제가 있다. 이러한 문제를 해결하기 위해 카메라 기반 감성분류 방법들이 많이 연구되고 있으며, 그 중 눈특징(Ocular feature)은 피험자가 의도적으로 제어할 수 없고, 무의식적으로 나타나는 특징까지 파악할 수 있기 때문에 감성분류에 사용하기 적합하다[7]. 또한, 다양한 분야에 적용될 것으로 기대되는 가상 및 증강현실(Virtual Reality & Augmented Reality)을 위해 눈특징 기반의 감성분류 기술은 필수적으로 연구되어야 한다.

※ Corresponding Author : Sang-Goog Lee, Address: (14662), 43 Jibong-ro, Bucheon-si, Gyeonggi-do, Korea, TEL : +82-2-2164-4909, FAX : +82-2-2164-4945, E-mail : sg.lee@catholic.ac.kr

Receipt date : Jun. 15, 2017, Revision date : Sep. 7, 2017
Approval date : Sep. 18, 2017

[†] Dept. of Digital Media., Graduate School, The Catholic University of Korea
(E-mail : jay_0624@catholic.ac.kr)

^{**} Dept. of Digital Media., Graduate School, The Catholic University of Korea
(E-mail : leedabid@catholic.ac.kr)

^{***} Dept. of Media Technology & Media Contents, The Catholic University of Korea

눈특징과 감성 사이의 관계성은 1960년대 초반부터 진행되었다[7]. Hess[8]는 불쾌한(Unpleasant) 영상에 의해 감성이 유발될 때 동공크기는 작아지며, 쾌한(Pleasant) 영상에 의해 감성이 유발될 때 동공크기가 확장됨을 보였다. Bradley[9]는 다양한 조명 세기와 쾌-불쾌 시각 자극 영상에 대한 동공크기 변화, 피부 전도도(Skin conductance) 그리고 심박율(Heart rate)의 변화에 대해 연구하였다. Ren[10]는 스트레스(Stress)와 완화(Relaxation) 시각 자극 영상이 제시되었을 때 Galvanic Skin Response(GSR)과 동공크기 변화에 따른 감성분류 연구를 진행하였다. Aracena[5]는 시각 자극 영상이 제시되었을 때 동공크기와 응시 위치(Gaze position)의 특징을 Neural Network와 Decision Tree에 적용하여 감성분류 정확도 연구를 진행하였다.

Krizhevsky[11]이 Convolutional Neural Networks(CNN)모델을 이용하여 영상분류 대회에서 큰 성공을 보이면서 CNN과 다양한 모달리티(Modality)를 이용한 감성분류 연구들이 제안되고 있다[12, 13, 14]. 하지만, Deep Neural Networks(DNNs)기반 감성분류[15] 연구에서 눈특징과 함께 다른 모달리티들을 함께 사용한 연구들이 제안되었을 뿐 눈특징만을 이용한 CNN기반의 감성분류 연구는 미비한 편이다.

본 논문은 시간에 따라 다양한 눈특징의 변화 정보를 포함하는 동공크기 변화신호를 훈련데이터로 사용하는 CNN기반 감성분류 방법을 제안하며, 그 과정은 다음과 같다. 첫째, 동공크기 변화, 눈 깜빡임 빈도수, 눈을 감은 지속시간 등 다양한 눈특징들을 포함하는 동공크기 변화신호를 데이터베이스로부터 획득한다. 이때, 시선 추적기의 기술적인 결함으로 인해 동공크기 변화신호를 획득하지 못한 구간은 눈 깜빡임에 대한 생물학적 정신의학 연구에 근거하여 보간(Interpolation)하여 신호를 생성하고, 수집된 신호들의 길이를 정규화(Normalization)한다. 둘째, 동공크기 변화신호를 시간과 주파수 정보로 시각화하여 파악할 수 있도록 Short-Time Fourier Transform(STFT)를 수행한다. 마지막으로 동공크기 변화신호에 적합한 CNN모델을 구성하여 2차원 감성분류를 실시한다.

본 논문의 구성은 다음과 같다. 2장에서는 시각화된 동공크기 변화신호와 CNN을 이용한 2차원 감성분류 방법을 제안한다. 3장에서는 실험을 통해 제안

하는 방법의 성능을 선행연구와 비교하여 분석하며, 4장에서는 결론 및 향후 과제를 기술한다.

2. 2차원 감성 분류

2.1 동공크기 변화신호의 보간 및 수집 신호 길이의 정규화

본 논문에서 사용한 데이터베이스는 HCI Tagging Database[16]로써 감성을 유발하는 20개의 비디오 영상을 1~2분 간 시청하는 동안 피실험자들의 뇌파(EEG), 심전도(Electrocardiogram; ECG), 피부 전도도(Galvanic Skin Resistance; GSR)와 같은 다양한 생리적인 신호들과 피부 온도, 호흡 그리고 눈특징 등을 수집하여 제공한다. 특히, 눈특징 데이터의 경우 좌·우 동공의 크기변화 뿐만 아니라 눈 깜빡임, 모니터 상 응시위치, 시선 추적기(Gaze Tracker)로부터 피실험자 간 거리 등을 포함한다. Soleymani는 [17]의 논문에서 30명의 피실험자로부터 데이터를 수집하였으나 기술적인 문제 등으로 인해 6명의 피실험자 데이터를 제외하고 총 24명의 데이터가 사용 가능하다고 기술하였다. 하지만, 다수의 데이터에서 시선추적기의 기술적인 결함으로 인해 눈 깜빡임 노이즈 등으로 동공을 제대로 추적하지 못하는 것을 확인하였다. 그러므로 본 논문에서는 동공크기 변화신호들을 눈 깜빡임은 약 300msec동안 발생된다는 Oh[18]의 자료에 근거하여 보간을 수행한다. HCI Tagging Database는 1분 동안 60Hz씩 동공신호를 수집하였기 때문에 신호가 없는 구간이 18 프레임 미만인 경우 시선 추적기의 기술적인 결함으로 인해 동공크기 변화신호를 획득하지 못한 구간으로 판단한다. 이러한 신호 미획득 구간에 신호를 채워 넣기 위해 각 미획득 구간에 이웃한 동공크기 변화신호의 지역 최대점(Local Maxima)과 지역 최소점(Local Minima)을 3개씩 추출하고 선형 보간법(Linear Interpolation)을 수행한다.

또한, 데이터베이스 수집에 사용된 비디오의 길이는 짧게는 35초 길게는 117초로 일정하지 않은 문제가 있다. 그러므로 본 논문에서는 단일 감성을 유발하는 비디오 시청 시 1분에서 10분이 필요하다는 Schaefer[19]의 연구 결과에 따라, 1분 보다 긴 동공크기 변화신호의 경우 가장 동공 변화가 뚜렷한 1분 동안의 신호를 추출하며, 54초 이상의 데이터는 1분

으로 신호의 길이를 정규화할 수 있도록 일정간격마다 공간을 할당하고 선형 보간법을 통해 보간한다. 이외에 54초 미만의 107.avi(35초), 79.avi(43초) 데이터는 보간법만으로 신호를 나타내는데 한계가 있어 제외한다. Fig. 1(a)는 총 길이가 1분 이상인 신호의 경우 동공크기 변화신호의 보간 및 수집 신호 길이의 정규화 과정을 보여주기 위해 선택한 데이터이다(첫 번째 피실험자의 26번째 실험으로부터 획득한 동공크기 변화신호). Fig. 1(b)는 기술적인 결함 등으로 발생된 문제를 보간을 통해 해결 후 Fig. 1(a)의 빗금영역을 삭제하여 신호 길이를 1분으로 정규화한 신호이다.

2.2 STFT를 이용한 동공크기 변화신호의 시간-주파수 시각화

동공크기 변화신호는 신호를 획득한 시점의 동공크기, 눈 깜빡임 빈도수뿐 만 아니라 동공크기의 시간적인 변화, 눈을 감은 지속시간 등 시간의 흐름에 따른 감성변화의 중요정보들을 포함한다. 따라서, CNN기반의 감성분류를 수행하기 전 동공크기 변화신호를 시간과 주파수 정보로 시각화할 수 있도록 STFT를 수행한다. 1초당 동공크기의 변화정보를 나타내기 위해 윈도우(Window)의 크기는 60으로 하였으며, 이를 해밍(Hamming) 함수를 이용하여 벡터로 구성하였다. 이산 푸리에 변환(Discrete Fourier Transform; DFT)은 2⁸ 포인트로 약 0.234 크기간격으로 주파수를 추출한다. 또한, 이전과 이후 시간의 동공변화 정보를 포함하기 위해 0.5초씩 오버랩(Overlap)한다. Pamplona[20]에 따르면 동공은 조명이나 그 크기 조절에는 관계없이 비정상적으로 수축

과 확대가 일어나는 동공동요(Hippus effect)가 0.05~0.3Hz에서 발생하므로 STFT 후 0.46Hz 미만의 특징은 제거한다. 그리고 제안하는 파라미터를 통해 생성된 STFT 특징을 살펴보면 Soleymani[17]이 주장한 바와 같이 2Hz 이상에서 두드러진 정보를 포함하지 않아 0.5Hz이상 2.0Hz 미만의 특징들만 추출하므로 최종적으로 6×119 크기의 특징행렬을 획득한다. Fig. 2는 STFT 변환 후 0.5~2.0Hz 사이의 특징들을 시각화 한 것으로 주파수 크기 값이 낮은 파란색부터 주파수 크기값이 높은 빨간색으로 나타낸다. Fig. 2에서 볼 수 있듯이 주파수 크기값이 큰 두드러진 특징들이 1.6Hz 미만에 나타남을 볼 수 있다.

아울러, 피실험자들 간 추출된 특징들의 차이를 줄이기 위해 정규화를 수행한다. 분산과 표준편차를 통해 정규화를 실시하는 경우 이상치(Outlier)에 영향을 많이 받는 문제가 발생할 수 있기 때문에 상대적으로 이상치의 영향을 덜 받으며, 강건성(Robustness)이 뛰어난 Median Absolute Deviation(MAD)를 통해 다음 식과 같이 정규화를 실시한다.

$$X'_i = (X_i - median(X))/median(|X_i - median(X)|)$$

이때, X는 각 피실험자의 모든 실험(Trial)에서 획득한 전체 특징이며, X_i는 i번째 실험에서 획득한 특징이다.

2.3 감성분류를 위한 CNN 모델 구성

영상 분류를 위해 제안된 CNN들은 일반적으로 합성곱(Convolution)과 서브 샘플링(Subsampling)을 반복적으로 수행함으로써 국부적 특징(Local feature)들로부터 전역적 특징(Global feature)을 획득

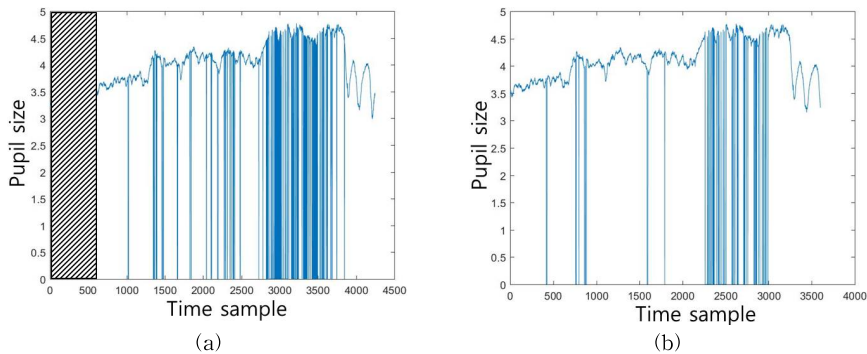


Fig. 1. Signal interpolation & normalization (Subject 1th, Trial 26th). (a) Original signal, (b) Interpolated & Normalized signal.

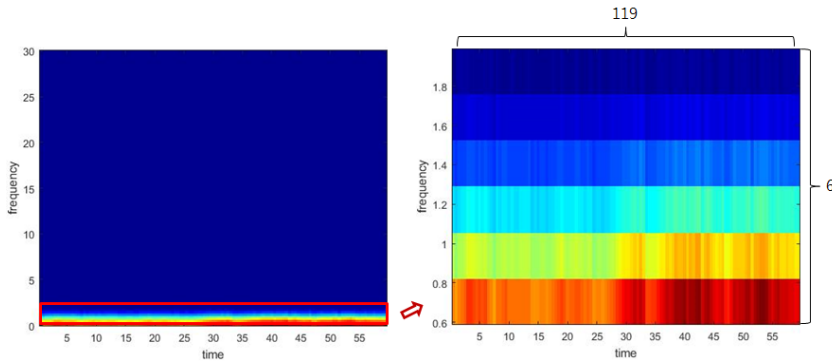


Fig. 2. STFT & Feature extraction(between 0.5~2.0Hz).

한다. 하지만, 본 논문에서는 시간의 흐름에 따른 동공크기의 주파수 정보가 감성분류에 중요한 요소이므로 시간의 국부 특징을 유지할 수 있도록 Fig. 3과 같이 2번의 합성곱과 1번의 서브 샘플링을 수행한다.

앞서 언급한 바와 같이 입력으로 사용될 특징은 STFT 후 0.5~2.0Hz 사이에서 추출한 것으로써 6×119 크기의 행렬이며 제안하는 CNN 모델의 구성은 다음과 같다. 첫째, 이웃한 시간의 동공변화 정보를 포함하기 위해 0.5초씩 오버랩하여 생성한 입력 특징의 시간축 국부 특징을 유지할 수 있도록 초단위의 특징맵(Feature map)을 구성하기 위해 1×3 크기의 커널 32개와 합성곱을 수행한다. 이때, 시간축으로 2칸씩 전진(Stride)하면서 합성곱을 실시하며 시간축의 시작과 끝의 경계정보를 획득하기 위해 입력 특징의 시작과 끝 열에 제로 패딩(Zero padding)을 수행한다. 이를 통해 Fig. 3의 C1과 같이 6×60 크기의 특징맵 32개가 생성된다. 둘째, 주파수의 국부 특징들 0.5Hz(0.4686Hz) 간격으로 특징맵을 구성하기 위해 3×1 크기의 커널 32개와 합성곱을 수행한다. 이때, 시간축과 주파수축으로 각각 1칸씩 전진하며 합성곱을 수행하여 시간축 정보는 그대로 유지하도록 한다.

이를 통해 Fig. 3의 C2와 같이 4×60 크기의 특징맵 32개가 생성된다. 셋째, 매 초당 주파수축의 전역 특징을 획득하기 위해 주파수축의 특징에 4×1 크기의 커널로 Max-pooling 방식의 서브 샘플링을 수행한다. 이를 통해 Fig. 3의 S1과 같이 1×60 크기의 특징맵 32개가 생성된다. C1 특징맵에서 6개의 특징에 대해 서브 샘플링을 수행하는 경우 S1과 동일한 크기의 특징맵을 생성할 수 있으나 가장 큰 주파수 크기(Magnitude)를 갖는 0.7031Hz의 특징만이 남는 문제가 있다. 이를 해결하기 위해 두번째 합성곱 단계에서 생성한 C2 특징맵에 서브 샘플링을 수행함으로써 더 많은 국부 특징 정보를 포함하는 전역 특징을 획득하도록 CNN 모델을 구성한다. 넷째, 신경망 학습을 위해 역전파(Backpropagation) 수행 시 발생하는 그래디언트 소실(Gradient vanishing)문제를 해결하기 위해 Rectified Linear Unit(ReLU) 활성화 함수를 적용한 후 분류할 감성의 개수와 동일한 수의 뉴런으로 구성된 완전 연결층(Fully connected layer)과 S1 특징맵을 연결한다. 분류할 감성의 수가 적기 때문에 층의 수가 많을 경우 과적합(Overfitting) 문제가 발생할 수 있으므로 완전 연결층은 하나로

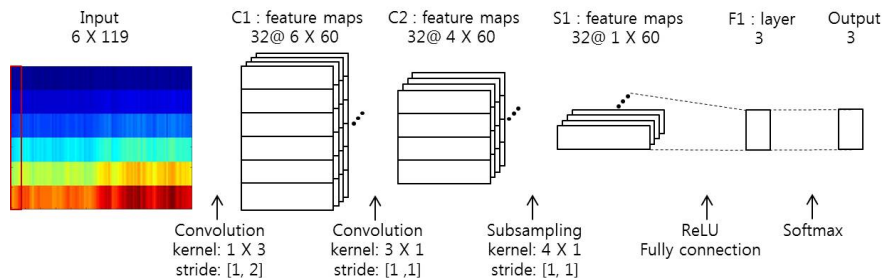


Fig. 3. Proposed CNN model.

구성한다. 마지막으로 소프트맥스(SoftMax) 함수를 통해 F1 내 뉴런들간의 상대적인 비교 후 최종 출력값을 결정한다. 뉴런의 활성화 함수로써 시그모이드(Sigmoid)를 사용하면 출력값을 공정하게 평가하기가 어렵기 때문에 소프트맥스함수를 통해 뉴런의 출력값을 정규화하고, 최종적으로 분류층(Classification layer)을 거쳐 감성을 분류한다. CNN 모델 훈련에 사용된 에폭(Epoch)은 1000이며, 실험을 통해 배치 크기(Mini-batch size)와 학습속도에 대한 평가를 실시하여 배치 크기를 50으로 결정하였다. 배치 크기의 결정을 위한 실험은 3.2절에 자세히 기술한다.

3. 실험 및 고찰

제안하는 CNN모델의 학습 성능을 평가하기 위해 손실함수(Loss function)를 이용하여 훈련 과정에서 배치(batch)의 수와 학습 속도의 영향을 살펴보고, HCI Tagging Database로 2차원 감성분류 실험을 실시한 Soleymani[17]의 결과와 비교를 통해 2차원 감성분류 성능을 평가한다. 실험은 Intel(R) Core(TM) i7-2600k CPU 20.00GB RAM PC환경에서 Matlab 2016a를 이용하여 수행하였다.

3.1 실험 데이터

HCI Tagging Database는 감성을 유발하는 20개의 비디오를 30명의 피실험자들이 시청하는 동안 슬픔(Sadness), 기쁨(Joy), 역겨움(Disgust), 평온(Neu-

tral), 즐거움(Amusement), 화남(Anger), 공포(Fear), 놀람(Surprise), 불안(Anxiety)과 같이 총 9개의 감성을 수집한 데이터베이스이며, 기술적인 문제로 인해 6명(P3, P9, P12, P15, P16, P26)의 피실험자 데이터를 제외한 총 24명의 데이터가 사용가능하다. HCI Tagging Database 수집에 참여하였던 M. Soleymani는 감성은 다른 언어로 정확하게 번역될 수 없음을 지적하였으며, Arousal-Valence 감성모델을 기반으로 Table 1과 같이 Arousal축을 차분함(Calm), 중간 각성(Medium arousal), 활성화(Activated), Valence축을 불쾌함(Unpleasant), 평온(Neutral valence), 쾌함(Pleasant)으로 재분류하였다[16].

Table 2는 Table 1의 감성 레이블(Emotion label)에 따라 HCI Tagging Database의 데이터 수집 시 사용한 시각자극 영상을 통해 재구성하여 작성한 것이다. Table 2에 따라 2차원 감성모델을 학습하고, Soleymani[17]의 실험결과와 비교하여 제안하는 CNN 모델의 성능을 평가한다.

3.2 배치(batch)의 수와 학습속도 결과

CNN 모델 구성 중 고려해야할 다양한 초-매개변수(Hyper-parameter) 중 하나인 미니 배치 크기를 통해 손실을 계산하고, 이를 손실 그래프(Loss graph)로 나타냄으로써 훈련과정 중 학습속도의 영향에 대해 평가하였다. 학습 속도가 낮은 경우 그래프가 선형적인 형태로 나타나며, 학습 속도가 좋은 경우 완만한 곡선 형태로 최저값에 빠르게 수렴한다. 학습

Table 1. The emotional keywords are mapped into three classes on Arousal and Valence[16]

		Emotional keywords
Arousal	Calm	sadness, disgust, neutral
	Medium arousal	joy and happiness, amusement
	Activated	surprise, fear, anger, anxiety
Valence	Unpleasant	fear, anger, disgust, sadness, anxiety
	Neutral valence	surprise, neutral
	Pleasant	joy and happiness, amusement

Table 2. Emotion label and video clips sources

Valence \ Arousal	Calm	Medium arousal	Activated
Unpleasant		146, 138	69, 55, 107, 30, 111
Neutral valence	newyork_f, detroit_f, dallas_f	53, 73	earworm_f
Pleasant	79	58, 80, 52, 90	cats_f, funny_f

속도가 높은 경우 손실의 감소가 가속되어 지수적인 형태로 나타나는데, 더 나쁜 손실값에 빠질 가능성이 있다. Fig. 4(a), (b)는 각각 Arousal과 Valence 감성에 대해 훈련 중 미니 배치 크기가 25(검정색), 50(빨간색), 100(녹색), 200(파란색) 일 때 에폭의 증가에 따른 손실의 변화를 나타낸 그래프이다. Fig. 4(a)를 보면 미니 배치 크기가 50인 경우 완만한 곡선 형태로 감소하여 최저값에 빠르게 수렴하며, 25인 경우 지수적인 형태로 최저값에 도달하지 못함을 볼 수 있다. Fig. 4(b)를 보면 미니 배치 크기가 25와 50인 경우 지수적인 형태를 보이지만 지속적으로 최저값으로 수렴하고 있으며, 50 일 때 좀 더 빠르게 최저값으로 수렴하는 것을 볼 수 있다. 손실 그래프를 통해 동공크기 변화신호 훈련 시 미니 배치 크기가 50일 때 가장 적합함을 확인하였다.

3.3 2차원 감성분류 결과

본 논문에서 제안하는 방법의 2차원 감성분류 정

확도를 평가하기 위해 Soleymani[17]의 실험과 동일하게 한명의 피실험자 데이터를 테스트 데이터로 사용하고 나머지 피실험자의 데이터를 훈련으로 사용하는 One-participant-out cross validation 방법을 채택하였다. Fig. 5(a), (b)는 각각 모든 피실험자 24명에 대한 Arousal과 Valence 감성분류 정확도를 나타내는 그래프이다. Arousal의 경우 피실험자 1이 97.22%로 분류 정확도가 가장 높았으며, 피실험자 6이 53.47%로 가장 낮았다. Valence의 경우 피실험자 10이 100%로 분류 정확도가 가장 높았으며, 피실험자 29가 57.78%로 가장 낮았다. 두 감성 모두에서 70% 미만의 분류율을 보인 피실험자 11, 19, 24, 25, 29의 경우, 동공크기 변화신호들 중 80번 이상씩 눈 깜빡임으로 인식된 신호들을 적어도 2개 이상 포함하고 있다. 이러한 신호들은 데이터양이 적기 때문에 신호를 획득하지 못한 구간을 보간하는 도중 잘못된 값들로 신호들을 채우는 문제가 발생하였고, STFT 후 1.5~2.0Hz 구간의 신호가 다른 실험의 신호들에

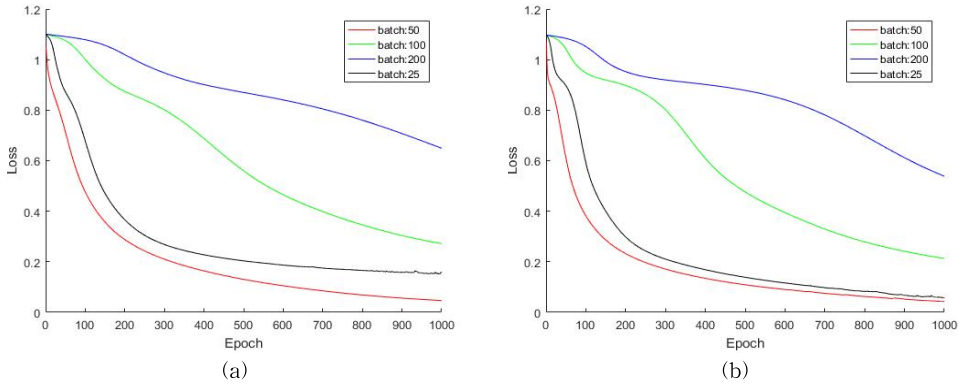


Fig. 4. Loss–Epoch graph with mini–batch size 25, 50, 100 and 200. (a) Arousal, (b) Valence.

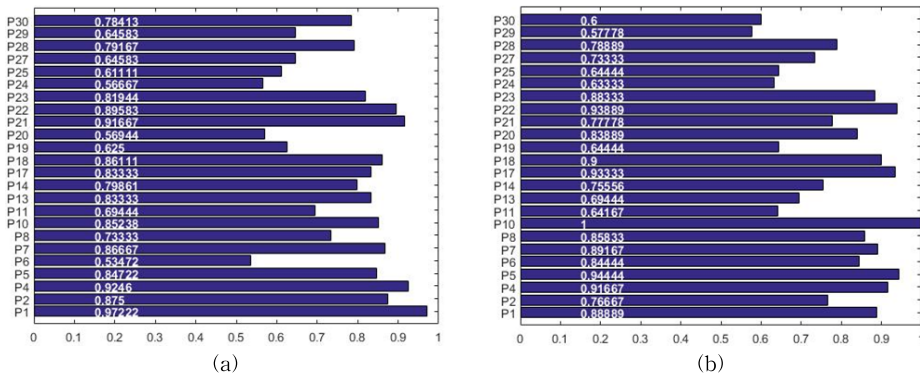


Fig. 5. Emotion classification accuracy of all subjects. (a) Arousal, (b) Valence.

Table 3. Arousal & Valence emotion classification average accuracy

		Arousal	Valence
Soleymani	Eye gaze	71.09%	66.59%
[17]	Eye gaze + EEG	76.44%	68.52%
Proposed method		77.08%	79.57%

비해 평균이 높게 나타나 감성 분류 정확도가 낮아지는 결과를 보였다.

Soleymani[17]는 다양한 실험을 통해 단일 모달리티(Single modality)기반 감성분류보다 다중 모달리티(Multiple modalities)기반 감성분류가 효과적임을 주장하였다. 단일 모달리티 기반 감성분류는 시선특징(총 12가지)과 Support Vector Machine(SVM)을 통해 수행했고, 다중 모달리티 기반 감성분류는 시선특징과 EEG 신호를 Decision Level Fusion(DLF)로 융합하여 SVM을 통해 수행했다. 그 결과 Table 3에서 볼 수 있듯이 다중 모달리티 기반 감성분류가 Arousal과 Valence 축에서 각각 5.35%, 1.93%씩 정확도가 향상되었음을 확인하였다. 하지만, 제안한 방법의 경우 동공크기 변화신호만을 이용하여도 다중 모달리티를 사용한 결과보다 Arousal과 Valence 축에서 각각 0.64%, 11.05%씩 정확도가 향상되었다.

4. 결론 및 향후과제

본 논문은 눈특징 데이터 평균과 표준편차, 눈 깜빡임 횟수 등 전체 데이터를 대표하는 특정 값들을 특징으로 사용하는 선행 연구들과 달리 시간정보를 포함하는 동공크기 변화신호를 감성분류를 위한 특징으로 사용하였으며, STFT를 통해 시간과 주파수 정보로 시각화 한 후 CNN으로 감성을 분류하는 새로운 방법을 제안하였다. 그 결과 동일한 데이터베이스에서 동일한 신호를 사용한 선행연구보다 Arousal과 Valence 축에서 각각 5.99%와 12.98%의 향상된 결과를 보였다. 이를 통해 얼굴 표정, 음성신호, EEG 등 다양한 모달리티를 특징으로 사용하는 CNN 기반의 감성분류 연구들과 같이 눈특징 또한 CNN을 적용하여 감성분류의 성능을 향상시킬 수 있음을 확인하였다. 앞으로 다양한 분야에 적용될 것으로 기대되는 가상 및 증강현실 기술 발전을 위해 본 연구와 같은 눈특징 기반 감성분류 방법은 더 많은 연구가 필요하다. 향후 CNN 수행 시 특징 추출단계 초기에

임의로 주어져 분류 성능에 큰 영향을 끼치는 필터들을 대체할 수 있는 특징 추출 방법에 대해 연구를 진행할 것이다.

REFERENCE

- [1] C. Won, "Recognition of Facial Emotion Using Multi-scale LBP," *Journal of Korea Multimedia Society*, Vol. 17. No. 12. pp. 1383-1392, 2014.
- [2] S. Chen, Y.L. Tian, Q. Liu, and D.N. Metaxas, "Recognizing Expressions from Face and Body Gesture by Temporal Normalized Motion and Appearance Features," *Image and Vision Computing*, Vol. 31, pp. 175-185, 2013.
- [3] G. Trigeorgis, F. Ringeval, R. Bruckner, E. Marchi, M. Nicolaou, and B. Schulleret, et al., "Adieu Features? End-To-End Speech Emotion Recognition Using A Deep Convolutional Recurrent Network," *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5200-5204, 2016.
- [4] Y. Gao, H.J. Lee, and R.M. Mehmood, "Deep Learning of EEG Signals for Emotion Recognition," *Proceeding of IEEE International Conference on Multimedia & Expo Workshops*, pp. 1-5, 2015.
- [5] C. Aracena, S. Basterrech, V. Snášel, and J. Velásquez, "Neural Networks for Emotion Recognition Based on Eye Tracking Data," *Proceeding of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2632-2637, 2015.
- [6] W. Zheng, B. Dong, and B. Lu, "Multimodal Emotion Recognition Using EEG and Eye Tracking Data," *Proceeding of Annual International Conference on IEEE Engineering in Medicine and Biology Society*, pp. 5040-5043, 2014.
- [7] P. Ekman, R.W. Levenson, and W.V. Friesen, "Autonomic Nervous System Activity Distinguishes between Emotions," *Science*, Vol. 221,

- pp. 1208-1210, 1983.
- [8] E.H. Hess and J.M. Polt, "Pupil Size as Related to Interest Value of Visual Stimuli," *Science*, pp. 349-35, 1960.
- [9] M.M. Bradley, L. Miccoli, M.A. Escrig, and P.J. Lang, "The Pupil as A Measure of Emotional Arousal and Autonomic Activation," *Psychophysiology* 45, pp. 602-607, 2008.
- [10] P. Ren, A. Barreto, Y. Gao, and M. Adjouadi, "Affective Assessment by Digital Processing of The Pupil Diameter," *IEEE Transactions on Affective Computing*, Vol. 4, pp. 2-14, 2013.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.
- [12] S. Tripathi, S. Acharya, R.D. Sharma, S. Mittal, and S. Bhattacharya, "Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset," *Innovative Applications of Artificial Intelligence*, pp. 4746-4752, 2017.
- [13] A. TeixeiraLopes, E.d. Aguiar, A. DeSouza, and T.O. Santos, "Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and The Training Sample Order," *Pattern Recognition*, Vol. 61, pp. 610-628, 2017.
- [14] W. Lim, D. Jang, and T. Lee, "Speech Emotion Recognition using Convolutional and Recurrent Neural Networks," *Proceedings of the Signal and Information Processing Association Annual Summit and Conference*, pp. 1-4, December 2016.
- [15] Wei Liu, W.L. Zheng, and B.L. Lu, "Multimodal Emotion Recognition Using Multimodal Deep Learning," *Proceeding of IEEE Winter Conference on Applications of Computer Vision*, 2016.
- [16] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," *IEEE Transactions on Affective Computing*, pp. 42-55, 2012.
- [17] M. Soleymani, M. Pantic, and T. Pun, "Multimodal Emotion Recognition in Response to Videos," *IEEE Transactions on Affective Computing*, Vol. 3, pp. 211-223, 2012.
- [18] J. Oh and J. Jeong, "Potential Significance of Eyeblinks as a Behavior Marker of Neuropsychiatric Disorders," *Korean Society of Biological Psychiatry*, Vol. 19, No. 1, pp. 9-20, 2012.
- [19] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the Effectiveness of a Large Database of Emotion-Eliciting Films: A New Tool for Emotion Researchers," *Cognition & Emotion*, Vol. 24, No. 7, pp. 1153-1172, 2010.
- [20] V.F. Pamplona, M.M. Oliveira, and G.V.G. Baranoski, "Photorealistic Models for Pupil Light Reflex and Iridal Pattern Deformation," *ACM Transactions on Graphics*, Vol. 28, No. 4, pp. 1-12, 2009.



이 희 재

2009년 가톨릭대학교 미디어공학과 학사
2011년 가톨릭대학교 미디어공학과 석사
2013년 가톨릭대학교 미디어공학과 박사수료

관심분야 : Pattern Recognition, Object recognition, Face recognition, AR, HCI



이 다 빛

2009년 8월 가톨릭대학교 미디어공학과(공학사)
2011년 8월 가톨릭대학교 미디어공학과(공학석사)
2012년~현재 가톨릭대학교 미디어공학과 박사과정

관심분야 : Biomedical Signal Processing, Pattern Recognition, HCI



이 상 국

1988년 인하대학교 전기공학과 학사
1990년 인하대학교 전기공학과 석사
1994년 프랑스 국립응용과학원 (Institut National des Sciences Appliquees)

전자공학과 공학박사
1995년~1999년 프랑스크립루앙(ROUEN Univ.) 대학교 전기전자공학부 조교수.
1990년~1999년 프랑스크립응용과학원 PSI(Perception System & Information)연구소연구원.
1999년~2001년 삼성전자 중앙연구소 Wearable Computer Project 팀장, 수석연구원.
2001년~2006년 삼성종합기술원 Master(기술명인), Mobile Augmented Reality Project 팀장, Ubicomp Interactive Device 그룹장.
2006년~현재 가톨릭대학교 미디어기술콘텐츠학과 교수
관심분야 : Smart Sensor & System, HCI, Ubiquitous/Wearable Computing, Augmented Reality