

감정 인식 기술 동향

송병철 김대하* 최동윤* 이민규*

인하대학교 교수

인하대학교 대학원생 *

I. 서론

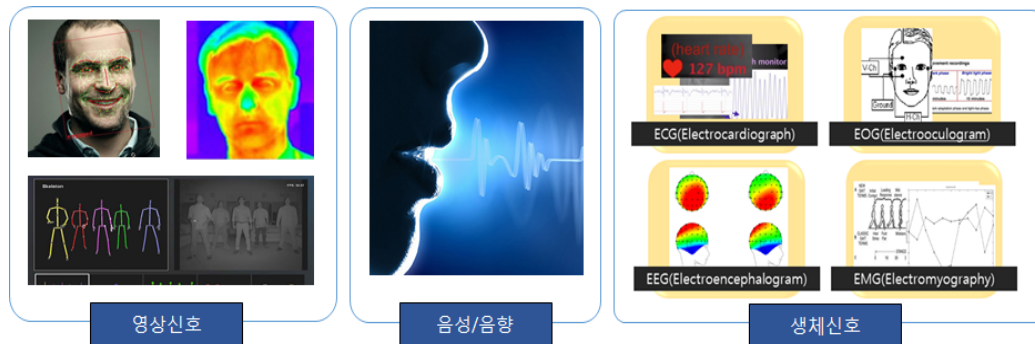
최근 소피아나 페퍼 같은 소셜 로봇이 화제가 되고 있다. 소셜 로봇이란 과거 기계적 움직임을 통해 인간의 육체적 노동을 대신하는 기존 로봇의 개념과 달리 인간과 커뮤니케이션을 수행할 수 있는 능력을 갖추고 자율적으로 동작하여 사회적 행동을 하는 감성중심의 로봇을 말한다. 소셜 로봇의 주요 기능 중 하나가 사람과의 자연스러운 대화를 통해 사람의 감정 상태를 파악하고 로봇 자신의 감정을 전달하는 정서적인 상호작용이다.

센싱 기술과 기계학습 기술의 발달로 인해 영상 및 음성, 생체신호를 통해 인간의 감정 및 내면 상태의 파악이 가능한 기술들이 개발되고 있으며 일부는 실제 적용되고 있다. 먼저 영상 센싱 기반 감정 인식 기술을 살펴보면, 영상 센서 정보를 이용한 얼굴 인식 및 개인 ID 식별 등의 연구가 지속적으로 연구되어 왔으며 각종 스마트 기기 및 SNS 등에서 기능들이 구현되어 사용되고 있다. 또한, 최근에는 얼굴의 특징점을 이용하여 인간의 감정을 판단하는 기술이 개발되고 있으며 인간의 동작인식을 통한 행동 분류 기술 등이 연구되고 있다. 이와 같은 컴퓨터 비전 기술을 이용한 사용자 식별과 인간 감정 및 행동 인식에 대한 기술은 소셜 로봇이 인간의 내면을 인식하는데 있어서 필수적인 기술이다.

한편, 음성 정보 기반 감정 인식과 관련해서는 인간의 목소리의 떨림 등의 패턴 분석을 통해 감정 상태를 파악하는 기술들이 개발되고 실제 현장에서 응용되고 있다. 예를 들어, 미국 AT&T의 콜센터에서는 음성인식 기술을 이용하여 실시간으로 속기록을 생성하고 고객의 감정 상태를 파악하여 대응하는데 이용하고 있다. 또한, 자폐증과 같이 표정으로 감정이 드러나지 않는 질환에서는 소셜 로봇과의 대화에서 얻을 수 있는 환자의 음성정보를 이용하여 내면상태를 파악할 수 있다.

* 본 내용은 송병철 교수(☎ 032-860-7413, bcsong@inha.ac.kr)에게 문의하시기 바랍니다.

** 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.



〈자료〉 Google

[그림 1] 영상/음성/생체의 복합적 신호 시스템

마지막으로 생체 신호 정보 기반 감정 인식의 경우, EEG, EMG, ECG 등의 생체 신호가 인간의 심리 상태를 파악하는데 상관관계가 매우 큰 정보가 될 수 있다. 최근 수요가 증가하고 있는 스마트 워치 등의 웨어러블 디바이스를 이용한 생체 신호 및 감정 상태 모니터링 기술이 개발되고 있다. 마이크로소프트의 경우 심전도 센서가 장착된 스포츠브라를 통해 사용자의 신체 및 감정상태를 파악하는 제품을 개발 중이다.

딥러닝은 신호 센싱과 함께 정확한 감정 인식을 위해 필수적인 요소이다. 딥러닝이란 컴퓨터가 데이터베이스를 이용하여 스스로 데이터를 분류하는 기준을 학습하도록 하는 기계학습을 말한다. 일반적인 딥러닝은 인간의 뇌를 모델링한 신경망 구조(Neural Network)를 기반으로 하며 여러 단계에 걸쳐 입력 데이터의 가치 판단 및 분류과정을 수행한다. 이론적 개념은 70~80년대에 완성되었으나 학습 과정에서 필요한 방대한 데이터와 연산량 때문에 한동안 실용화되지 못하고 있었다. 그러나 최근 빅데이터를 통한 충분한 데이터베이스와 하드웨어 처리속도의 발달로 인해 딥러닝 기술의 적용이 한층 용이해졌다. 딥러닝 기술의 발달로 인해 최근 인공지능(AI) 기술의 발전도 급격하게 이루어지고 있는데 이에 대한 사례는 우리 사회에서 쉽게 접할 수 있다. 세계적으로 이슈가 되었던 구글 딥마인드가 개발한 알파고는 딥러닝 기술을 통해 바둑에서 인공지능 성능을 확인하였고 이는 인공지능 시대를 여는데 큰 이바지를 하였다. 컴퓨터 비전 기술 분야에서는 CNN(Convolutional Neural Network)을 이용하여 영상 데이터의 객체 검출 및 인식 기술이 확연하게 향상되었고, 음성 분야에서는 RNN(Recurrent Neural Network)을 이용하여 음성 인식 및 음악 장르 분류 등에 우수한 성능을 보이고 있다.

앞서 언급한 센싱 기술과 딥러닝 기술, 그리고 데이터베이스 기술이 융합되어 최근 인간의 감정 인식 기술이 발전하는 추세에 있다. 인간의 표면적인 감정은 얼굴 표정이나 행동 그리고 음성으로

인식이 가능하다. 하지만, 내면적인 감정 인식은 쉽지 않은데, 생체 신호를 이용하면 인식이 가능할 수 있다. 따라서 단일 신호에만 의존하지 않고, 영상, 음성, 생체 신호 등 멀티모달 신호를 사용하는 것은 감정 인식의 정확도를 향상시키는데 기여할 수 있다. 그러므로 멀티모달 신호를 이용한 딥러닝 알고리즘을 효과적으로 설계하고 학습시킬 수 있다면 소셜 로봇 같은 기계와 인간의 상호작용이 좀 더 원활해질 수 있을 것이다.

II. 감정 인식 기술 개요

1. 동영상 기반 감정 인식 기술 동향

동영상 기반 감정 인식은 단일 영상에서 얼굴의 특징을 분석하는 것에서부터 시작한다. 초기에는 고전적인 기계학습 및 컴퓨터 비전을 사용하여 얼굴의 특징을 분류하는 연구가 진행되었다[1]. 예를 들면, 얼굴 영상에서 추출된 그래디언트를 기반으로 얼굴의 특징을 추출하고 SVM(Singular Vector Machine)이나 랜덤 포레스트 같은 알고리즘으로 그 특징들을 분석함으로써 얼굴의 표정을 파악한다. 그러나 이와 같은 기술은 주변 배경 또는 영상의 조도에 크게 영향을 받는다는 단점이 있다. 또한, 얼굴의 각도에 영향을 받기 때문에 높은 성능을 달성하기 매우 힘들다. [그림 2]는 영상의 그래디언트 기반 분석과 이를 응용한 HOG(Histogram of Gradients) 분석의 예시를 보여준다[2].



〈자료〉 Dadi, H. S., and Pillutla, G. K. M.(2016). Improved face recognition rate using HOG features and SVM classifier. IOSR J Electron. Commun. Eng(IOSR-JECE)

[그림 2] 그래디언트 기반 얼굴 특징 분석 및 해당 영상의 HOG 특징 맵

초기 연구에 사용되었던 데이터셋은 [그림 3]의 CK+와 같이 주로 실험실 및 제한된 환경에서 확보된 것들이었다. 이후에는 [그림 3]의 AFEW와 같이 점차 일상적인 상황을 담은 동영상들을 담은 데이터셋들이 주로 이용되고 있다.



(a) CK+

(b) AFEW

〈자료〉 <http://www.pitt.edu/~emotion/ck-spread.htm>

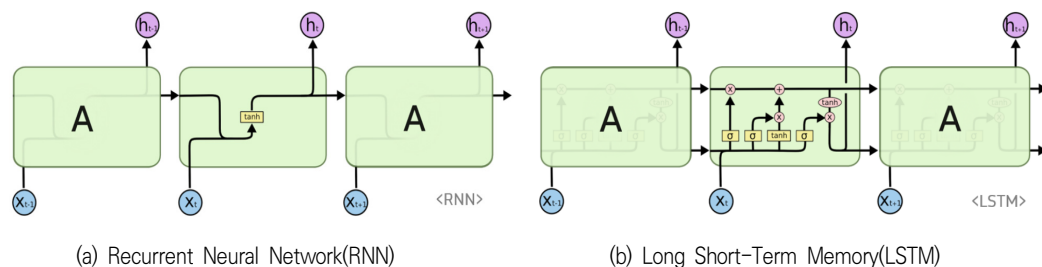
〈자료〉 <https://cs.anu.edu.au/few/AFEW.html>

[그림 3] 데이터베이스의 예

앞서 언급한 바와 같이 감정 인식 연구는 갈수록 판단하기 힘든 상황에서의 인물 감정 인식쪽으로 연구가 확장되고 있다. 예를 들면, AFEW 데이터셋은 점차 리얼리티 쇼나 시트콤과 같이 좀 더 즉각적인 감정의 분석에 초점을 두고 있다.

한편, 딥러닝 기술의 급속한 발전에 따라 영상 내 얼굴의 이차원적인 특징을 추출하고 분석하기 위해 CNN과 같은 딥 네트워크가 널리 사용되고 있다. 또한, 동영상 내 프레임들의 시간적 연속성을 추출/분석하기 위한 LSTM(Long Short-Term Memory)이라는 기법도 자주 사용되고 있다. 예를 들면, [그림 4]와 같이 LSTM 네트워크 내부에서 프레임별로 가중치를 달리하여 연속적인 프레임에서 좀 더 의미 있는 프레임을 찾기 위한 연구가 진행되고 있다[3].

마지막으로 사람의 감정을 단순히 화남, 역겨움, 두려움, 행복함, 중립, 슬픔, 놀라움 등 7가지로 분류하는데서 벗어나 좀 더 세밀한 감정 분석을 위해 연속적인 Arousal-Valence(AV) 영역에서의 감정 인식 연구도 활발해지고 있다. 이는 사람의 감정을 특정한 라벨로 분류하는 것보다 흥분 정도와 positive/negative 정도를 파악하여 그 수치로 사람의 감정을 연속적으로 분류하는 것이 더 효율적



(a) Recurrent Neural Network(RNN)

(b) Long Short-Term Memory(LSTM)

〈자료〉 <http://ratsgo.github.io/>

[그림 4] RNN 및 LSTM 구조

일 수 있기 때문이다. 한편 AV 영역에서의 감정 인식 분석을 위한 AffectNet 데이터셋이 만들어져 관련 연구가 활발히 진행되고 있다[4].

2. 음성 기반 감정 인식 기술 동향

상기 동영상 기반 감정 인식 기술 동향에서 알 수 있듯이 음성은 영상과 함께 감정을 인식하기 위한 또다른 중요한 실마리이다. 그러나 음성만으로 감정 인식을 하는 응용 사례는 실제로 많지 않다. 이는 동영상 속에서 음성 정보는 산발적으로 존재하기 때문이다. 일부 음성정보만을 이용한 감정인식 연구를 살펴보면 다음과 같다. 한동안 음성 기반 감정 인식 연구는 작은 시간 영역 단위의 실시간 감정 인식에 집중되어 있었다[11]. 실시간 감정 인식 연구는 크게 두 가지로서 새로운 특징을 추출하거나 분류 방법론을 달리 하여 정확도를 개선하는 연구들로 나뉜다. 특징 추출 연구로는 개인마다 발성의 특징이 다르기 때문에 이를 반영하는 특징을 찾는 것을 목표로 한다. 예를 들면, 구간이 아닌 순간 감정을 인식하는 기술이 있다[12]. 분류 방법론 중 하나인 계층적 분류 방법론은 여러 개의 분류기들을 사용하여 음성에서 비슷한 감정의 인자를 나누어 분류하는 방식이다[13]. 비교적 높은 정확도를 보이지만 짧은 음성만 인지가 가능한 단점이 있다. 한편 남성과 여성의 학습 모델을 각각 생성한 다음, 입력 음성을 남성 혹은 여성으로 선별한 다음 성별에 맞는 학습 모델과 비교하는 방식으로 인지하는 방법도 있다[14]. 최근에는 딥러닝을 이용한 음성 기반 감정 인식 기술들이 주를 이루고 있다. 예를 들면, 이지원 등은 일반화 오류를 보완할 수 있는 다중 작업 기반 합성곱 신경망을 이용한 음성 감정 인식 시스템을 제안하였다[15]. 즉, 감정 분류만을 수행하던 기존 신경망을 확장시켜 성별, 감정 활성화도, 긍정도 정보를 활용한 다중 작업 기반 신경망 학습을 통해 감정 인식의 성능을 높이하고자 하였다. 한편, 강소연 등은 베이지안 로지스틱 회귀(Bayesian logistic regression)를 랜덤 포레스트(random forest)로 대체하여 종래 베이지안 기법들보다 정량적인 감정 인식 성능을 향상시킨 바 있다[16].

3. 생체신호 기반 감정 인식 기술동향

한편, 맥박이나 EEG같은 생체 신호를 이용한 감정 인식 연구도 오랫동안 연구되어 왔다. 먼저 해외 연구사례를 살펴보면, Haag 등은 여러 바이오 센서들로부터 얻은 다중 신호를 사용하여 감정을 인식하도록 컴퓨터를 학습시키는 방법을 소개했다[20]. 논문에 따르면 자체 데이터베이스를 이용 하기는 했지만 제안하는 신경망 기반 시스템을 사용하였을 때 90% 내외의 높은 정확도를 보였다.

참고문헌 [21]은 감정 인식을 위한 뇌파에 대한 포괄적인 설명과 함께 딥러닝을 포함한 여러 가지 특징 추출 기법들을 소개하였다. 우리나라에서도 최근 생체 신호를 이용한 감정 인식 및 그 응용 연구들이 이루어지고 있다. 송병호 등은 사용자와 로봇 간 감성적인 교감을 통해 LED 조명을 제어하는 시스템으로써 사용자의 생체 신호(맥박, 혈압, 혈당)를 측정하여 분석한 후 감성을 분류하는 연구를 수행하였다[17]. 또한, 이현수 등은 사용자의 생체 신호를 기반으로 감정을 판별하여 이를 바탕으로 음악을 추천하는 시스템을 제안하였다[18]. 류기민 등은 인공신경망을 이용하여 감정에 따른 EEG 신호를 AV 영역에서 분류하는 기법을 선보였다[19]. 제안 기법은 Valence 축의 경우 종래 기법보다 약간 낮은 결과를 얻었지만, Arousal 축에 대해서는 5% 향상된 결과를 보였다.

III. EmotiW의 소개

EmotiW(Emotion Recognition in the Wild Challenge)는 ACM의 ICMI(International Conference on Multimodal Interaction)라는 학술대회와 함께 개최되는 감정 인식 기술 경진 대회이다. 세부적으로는 세 가지 부문으로 나뉘며, 그룹-단위 감정 인식(group-level emotion recognition), 참여 강도 인식(engagement in the wild), 그리고 동영상 인식(audio-video sub-challenge) 등이다.

먼저 그룹-단위 감정인식 부문을 살펴보면 다음과 같다. 주요 작업은 사진 속사람들로부터 인지된 감정을 긍정적, 중립적, 부정적 감정 중 하나로 분류하는 것이다. 예를 들면, SNS 사용자는 다양한 사고 행사 중 촬영한 영상들을 인터넷을 통해 시스템에 업로드한다. 영상은 회의, 결혼, 파티와 같은 긍정적 상황을 담을 수 있고, 회사 내 회의와 같은 중립적인 상황을 표현할 수도 있다. 때로는 장례식 같은 부정적인 상황의 영상일 수도 있다. 이 부문 경쟁에서 사용하는 데이터셋은 Group Affect Database 2.0[1]으로서 상기와 같은 다양한 상황을 담고 있다. 상기 데이터셋은 학습, 검증 및 테스트의 세 가지 서브-세트로 구성된다. 두 번째로 2018년 신설된 “Engagement in the wild” 경쟁 부문에서는 동영상에서 피사체의 참여 강도를 예측하는 것이 목표이다. 녹화 세션 중 피실험자는 교육용 동영상(MOOC)을 시청한다. 데이터셋은 다양한 조건과 다양한 환경에서 기록되었다. 마지막으로 동영상 내 인물의 일반적인 감정 인식을 평가하는 부문으로 “Audio-video Emotion Recognition”이 있다. 다양한 동영상에 나오는 인물들의 표정 및 분위기를 통해 감정을 예측해 가장 높은 정확도를 보이는 팀이 우승하는 대회이다. 이 대회에서 사용하는 기본 데이터셋은 앞서 언급한 AFEW 데이터셋이며, 인위적인 환경에서의 데이터셋들보다 훨씬 난이도가

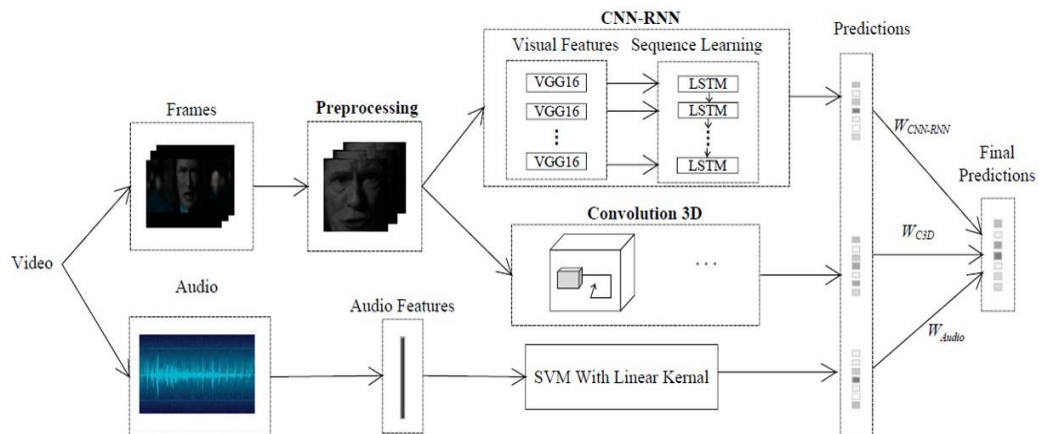
높다. AFEW 데이터셋도 학습, 평가, 그리고 테스트 데이터 셋들로 이루어져 있다. 학습 및 평가 데이터는 각 동영상 내 인물이 어떤 감정을 가지는지에 대한 정보가 주어져서 성능에 대한 지표로 활용이 가능하다.

IV. 멀티모달 딥러닝 기반 감정 인식

1. 기술 발전 추이 및 고찰

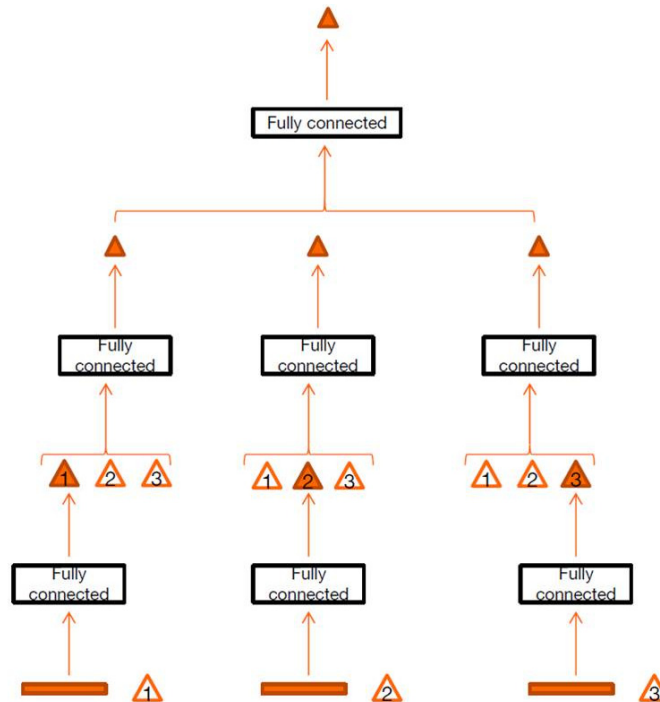
2015년 이전까지는 영상에서 핸드 크래프트 특징을 추출하는 방법이 많이 사용되었으나, 그 이후에는 딥러닝을 이용한 다양한 방법이 나오고, 더 나아가 영상 및 음성의 멀티모달을 통해 꾸준히 성능이 발전되고 있다([그림 5] 참조).

딥러닝을 이용한 감정 인식 방법에는 CNN-RNN과 Convolutional 3D 기법을 적절히 융합한 사례가 있다[5]. CNN은 영상의 공간적인 특징을 잘 추출해주며, 이는 인간의 표정에 대한 적절한 특징을 잘 추출하도록 도와준다. 또한, RNN에서는 시간적인 특징을 고려해줄 수 있기 때문에 영상의 맥락 및 분위기를 적절히 학습할 수 있도록 도와준다. 또한, CNN은 한 장의 영상을 입력으로 받지만 convolutional 3D 네트워크는 동영상을 입력으로 받을 수 있으며, 이를 통해 공간 영역에서의 컨벌루션(convolution) 연산을 적용할 수 있을 뿐만 아니라 시간 영역에 대해서도 컨벌루션



〈자료〉 Fan, Yin, et al. "Video-based emotion recognition using CNN-RNN and C3D hybrid networks." Proceedings of the 18th ACM International Conference on Multimodal Interaction. ACM, 2016.

[그림 5] CNN-LSTM과 Convolutional 3D 기법을 이용한 감정 인식 네트워크



〈자료〉 Vielzeuf, Valentin, Stephane Pateux, and Frederic Jurie. "Temporal multimodal fusion for video emotion classification in the wild." Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017.

[그림 6] 딥러닝 기반 멀티모달 융합 기법

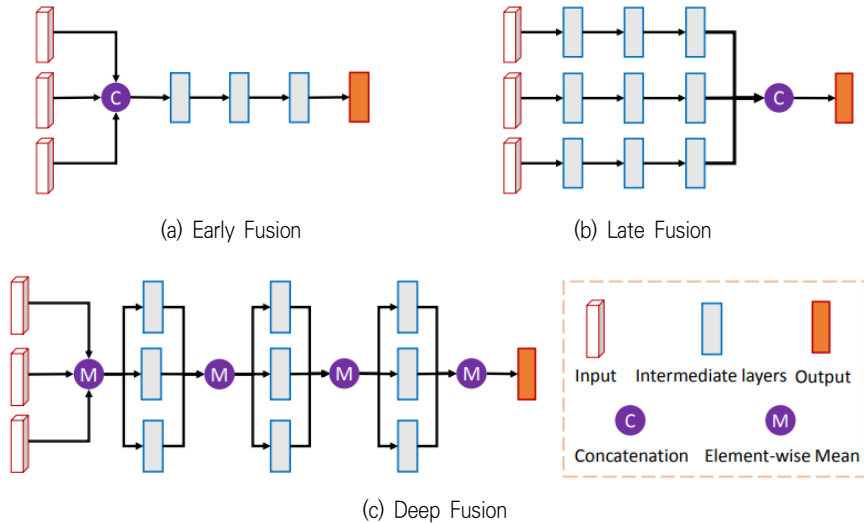
연산을 수행할 수 있다. 따라서 학습 성능이 향상될 수 있다.

한편, 딥러닝 네트워크 개선뿐만 아니라 다양한 신호에 대한 융합을 통해 감정 인식 성능을 높이는 접근 방법들이 있다. [그림 6]은 딥러닝 네트워크 기반으로 추출한 세 가지 신호(영상, 음성, 동영상)를 MLP(Multi-Layer Perceptron)를 통해 늦은 융합(Late Fusion)하는 방식을 보인다[6].

프레임 단위로는 공간적인 특징, 동영상 단위로는 움직임 등 시간축 특징, 음성에서 분위기에 대한 요소를 각각 추출하여 특징 벡터를 구성한다. [그림 6]과 같이 총 세 단계에 걸친 MLP를 통해 상기 세 가지 신호가 적절한 가중치를 부여받으면서 합쳐진다.

2. 멀티모달 딥러닝

멀티모달 딥러닝은 성질이 다른 복수의 입력을 처리하는 딥러닝 알고리즘으로 각각의 신호 모달리티에서 부족한 정보를 상호 보완하여 인식 성능을 향상하고자 하는데 목적이 있다. 감정 인식



〈자료〉 © IEEE CVPR 2017.

[그림 7] 멀티모달 딥러닝 융합 방식 예시

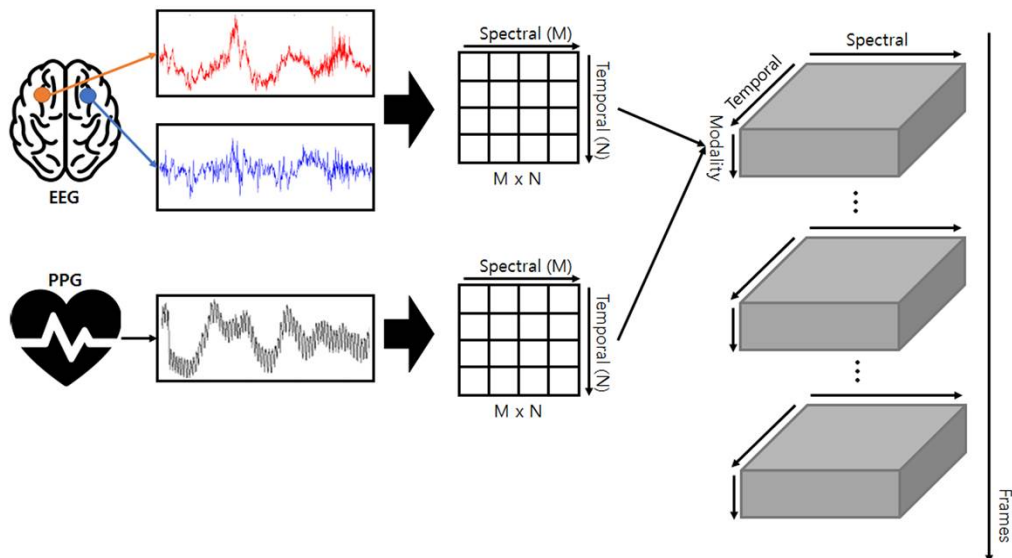
분야에서도 널리 사용되고 있으며 일반적으로 앞에서 언급된 영상정보를 이용한 감정 인식에서 영상 내부의 표정으로 드러나지 않는 감정 정보를 인물의 대화 톤이나 의성어 같은 음성 신호나 EEG와 맥박 같은 생체 신호를 분석하여 보완할 수 있다.

멀티모달 딥러닝 과정에서는 서로 다른 모달리티의 정보를 융합하는 과정이 핵심이다. 융합 과정은 [그림 7]과 같이 크게 조기 융합과 늦은 융합으로 나누어진다. 조기 융합의 경우 서로 다른 모달리티의 입력에 대해 정규화 등의 전처리를 수행하고 두 입력을 결합하여 하나의 알고리즘에 입력하는 방식이다. 전처리의 복잡도가 있을 수는 있지만, 하나의 딥 네트워크로 처리하기 때문에 상대적으로 연산량이 적은 장점이 있다. 반면, 데이터 레벨에서 융합이 먼저 이루어지기 때문에 각 모달리티가 가진 고유의 특징을 추출하는 데는 한계가 존재한다. 늦은 융합의 경우 서로 다른 모달리티가 각각의 네트워크를 통해 처리되어 출력 직전에 결과를 융합하여 최종 결과를 구하는 방식이다. 이는 각 모달리티 고유의 특징을 추출하는데 효과적이다. 반면, 모달리티별로 네트워크를 사용하기 때문에 연산량 측면에서는 부담이 될 수 있다. [그림 7] 하단은 하이브리드 융합 혹은 깊은 융합을 보여주고 있으며, 조기 융합과 늦은 융합을 합친 방식이라고 할 수 있다. 즉, 모달리티 별 네트워크가 존재하면서 중간 중간 특징 맵을 융합하는 방식이다[7]. 잘 활용하면 조기 융합과 늦은 융합의 장점을 살릴 수 있다. 그러나 설계의 난이도가 높고, 복잡도 또한 만만치 않다.

3. 최근 동향

멀티모달 딥러닝 기법은 EmotiW에서도 대부분의 팀들에 의해 활용되고 있다. EmotiW에서 제공하는 데이터셋은 인물의 표정과 음성이 존재하는 비디오와 오디오의 멀티모달 정보를 제공한다. 최근 EmotiW에서 발표되는 기술은 영상정보를 처리하는 네트워크와 음성정보를 처리하는 네트워크를 별도로 설계하여 처리하고 감정에 대한 출력 확률 정보를 융합하는 낮은 융합 과정이 주를 이루고 있다[6]. 영상 정보를 처리하는 네트워크는 영상 정보를 CNN과 LSTM이 융합된 구조를 이용하는 것이 대표적이고, 음성 정보의 경우 스펙트로그램 및 MFCC(Mel-Frequency Cepstrum Coefficient) 등 주파수 도메인의 정보로 변환하여 영상과 마찬가지로 CNN 등의 딥러닝 기법을 활용하고 있다[8].

생체 신호에 대한 멀티모달 딥러닝 감정 인식으로 [그림 8]과 같이 뇌파정보인 EEG 신호와 맥박 정보인 PPG 신호를 동시에 이용한 사례가 있다[9]. 이 기술의 경우 EEG 신호와 PPG 신호를 주파수 도메인의 신호로 변환하여 행렬의 형태로 표현한 다음 두 신호를 결합하여 3차원 텐서의 형태로 구성하여 딥러닝 네트워크의 입력으로 사용하는 조기 융합 방식을 이용하고 있다. 즉, EEG 신호와 PPG 신호가 결합된 입력을 영상 시퀀스로 처리하는 Convolutional LSTM[10]을 통해 감정 인식을 수행한다.



〈자료〉 © IEEE TAFFC 2018.

[그림 8] EEG신호의 PPG신호의 융합

최근 감정 인식 분야에서의 멀티모달 딥러닝의 활용으로는 영상 정보와 음성 정보를 동시에 활용하거나 여러 종류의 생체 신호를 융합하여 사용하는 사례가 대부분이다. 현 시점에서는 실험 데이터의 취득 및 동기화 문제 등으로 인해 영상/음성/생체신호 등의 복합적인 멀티모달 감정 인식에 대한 감정 인식 기술은 많이 등장하지 않고 있으며 영상 정보와 생체 신호 간의 분석 정도가 발표되고 있다. 향후 영상/생체신호 등을 이용한 멀티모달 딥러닝 기반의 감정 인식 기법의 개발이 이루어질 것으로 전망되며 기존의 영상/음성 또는 생체신호들 간의 멀티모달 감정 인식 기술에 비해 성능이 크게 향상될 것으로 예상된다.

V. 결론 및 시사점

감정 인식 기술은 진정한 인간-컴퓨터(혹은 로봇)간 상호작용을 위한 핵심 요소기술이다. 감정 인식의 재료가 되는 얼굴표정(영상), 목소리(음성), EEG나 맥박(생체신호) 등이 정확한 감정 인식을 위해서는 함께 이용되는 것이 바람직하다. 그러기 위해서는 소위 멀티모달 센싱 및 딥러닝 기술이 지속적으로 발전되어야 한다. 아울러 본 고에서는 자세히 언급되지 않았지만 데이터베이스의 구축도 매우 중요하다. 일반적인 물체인식 데이터베이스에 비해 감정 인식 데이터베이스는 턱없이 부족하다. 기술발전을 위해서는 이에 대한 투자와 연구가 절실하다. 마지막으로 감정 인식 기술은 정상인뿐만 아니라 자폐 같은 장애가 있는 분들에게 매우 유용하게 사용될 것으로 보인다. 이미 몇몇 연구기관에서 수행된 연구 결과에 따르면 감정 인식 기술은 진찰은 물론 치료 목적으로 매우 중요하게 사용될 것으로 보인다.

[참고문헌]

- [1] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. "Face recognition: A literature survey," ACM computing surveys(CSUR), 2003, 35(4), 399-458.
- [2] Dadi, H. S., & Pillutla, G. K. M., "Improved face recognition rate using HOG features and SVM classifier," IOSR J Electron Commun Eng(IOSR-JECE), 2016, 11(4), 34-44.
- [3] TAO, Fei; LIU, Gang. "Advanced LSTM: A Study about Better Time Dependency Modeling in Emotion Recognition," arXiv preprint arXiv:1710.10197, 2017.
- [4] Mollahosseini, A., Hasani, B., & Mahoor, M. H. "Affectnet: A database for facial expression, valence, and arousal computing in the wild," arXiv preprint arXiv:1708.03985. 2017.
- [5] Fan, Yin, et al. "Video-based emotion recognition using CNN-RNN and C3D hybrid networks," Proceedings of the 18th ACM International Conference on Multimodal Interaction. ACM, 2016.

- [6] Vielzeuf, Valentin, Stephane Pateux, and Frederic Jurie. "Temporal multimodal fusion for video emotion classification in the wild," Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017.
- [7] Baltrušaitis, T., Ahuja, C., & Morency, L. P. "Multimodal machine learning: A survey and taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence(TPAMI), 2018, early access
- [8] Wang, Shuai, et al. "Emotion recognition with multimodal features and temporal models," Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017.
- [9] Kim, B. H., & Jo, S. "Deep Physiological Affect Network for the Recognition of Human Emotions," IEEE Transactions on Affective Computing(TAFFC), 2018, early access
- [10] Xingjian, S. H. I., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C., "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," In Advances in neural information processing systems(NIPS), 2015, pp.802-810.
- [11] 방재훈, 이승룡, "감성기반 서비스를 위한 통화 음성 감정인식 기법", 정보과학회논문지, 제41권, 제3호, 2014, pp.208-213.
- [12] A. B. Kandali, A. Routray, T. K. Basu, "Emotion recognition from Assamese speeches using MFCC features and GMM classifier," IEEE Region 10 Conference(TENCON), Nov, 2008, pp.1-5, 19-21.
- [13] Z. Xiao, Dellandrea, L. Chen, W. Dou, "Recognition of emotions in speech by a hierarchical approach," ACII 2009, 2009, pp.401-408.
- [14] Y. Cho, K. S. Park, "A Study on The Improvement of Emotion Recognition by Gender Discrimination," Journal of IEEK, vol.45, 2008, pp.401-408.
- [15] 이지원 외, "다중 작업 기반의 합성곱 신경망을 이용한 음성 감정인식", 2017년 한국통신학회 하계종합학술대회, 2017. 6.
- [16] 강소연, 최욱, "Random forest를 이용한 음성신호 기반 감정인식", 2017년 한국통신학회 동계 종합학술발표회, 2017년.
- [17] 송병호 외, "사용자 생체신호 인식 기반 감성 소셜 로봇 시스템 설계", 제어로봇시스템학회 하계학술대회, 2018.
- [18] 이현수 외, "생체 신호를 이용한 사용자 감정 기반의 음악 추천 시스템 제안", 대한전자공학회 하계종합학술대회, 2017.
- [19] 류기민, 차형태, "EEG 신호 기반 인공신경망을 통한 감정 인식에 대한 연구," 한국통신학회 동계종합학술발표회, 2018.
- [20] Haag, Andreas, et al. "Emotion recognition using bio-sensors: First steps towards an automatic system," Tutorial and research workshop on affective dialogue systems. Springer, Berlin, Heidelberg, 2004.
- [21] Jenke, Robert, Angelika Peer, and Martin Buss. "Feature extraction and selection for emotion recognition from EEG," IEEE Transactions on Affective Computing 5.3(2014): 327-339.