

가상과 현실의 공존공간에서 인간 감정정보와 신체정보를 이용한 인공지능 몰입증강 기술

이 영 재*

Artificial Intelligence Immersion Enhancement Technology using Human Emotion Information and Body Information in the Coexistence Space of Virtual and Reality

Young Jae Lee*

요약

IoT 기반의 스마트 환경하에서 사용자 맞춤형 서비스를 제공하기 위해 동적으로 변화하는 상황 정보를 이용한 인공지능 몰입증강 기술을 제안한다. 이를 검증하기 위해 3차원 공간상의 사용자의 얼굴검출과 손, 손가락의 움직임 정보를 KNN 기법으로 검출하고, 검출된 정보를 활용하여 가상공간 객체와 상호작용을 한다. 이 과정에 사용자의 콘텐츠 몰입과 만족도 정도를 사용자 얼굴에 나타난 감정 정보를 인공지능 기법을 사용해 검출하고, 동시에 눈의 깜빡임, 하품 정보 등을 분석해 신뢰성 있는 상황 정보를 만들어 사용자의 감정과 신체 상황에 기반한 맞춤형 서비스를 제공하는 방법이다. 제안한 알고리즘 실험결과 사용자의 섬세한 감정인식뿐 아니라 눈의 깜빡임, 하품 등의 신체 변화 분석을 통한 인공지능 기반 몰입 증강 기술의 유의미한 성능을 확인할 수 있었다. 제안한 방법은 사람을 더 잘 이해할 수 있는 기술을 만드는 기본 자료로 활용될 수 있다.

Abstract

We propose an artificial intelligence immersion enhancement technology using dynamically changing context information to provide customized services in an IoT-based smart environment. To verify this, the user's face detection and hand and finger movement information in 3D space are detected using the KNN technique, and the detected information is used to interact with the virtual space object. In this process, the user's content immersion and satisfaction level is detected using artificial intelligence techniques to detect the emotional information displayed on the user's face, and at the same time, it analyzes eye blinking and yawning information to create reliable situational information. And provide customized services based on the user's emotions and physical conditions. As a result of the proposed algorithm experiment it was possible to confirm the meaningful performance of AI-based immersion enhancement technology through analysis of body changes such as blinking eyes and yawning as well as sensitive emotional recognition of users. The proposed method can be used as a basic data to create a technology that can better understand people.

Keywords

artificial intelligence, immersion enhancement, customized services, blinking eyes and yawning, emotional recognition

* 전주대학교 문화융합대학 스마트미디어학과 교수
- ORCID: <https://orcid.org/0000-0002-8812-3843>

• Received: Aug. 17, 2022, Revised: Sep. 19, 2022, Accepted: Sep. 22, 2022
• Corresponding Author: Young Jae Lee
Dept. of Smartmedia, Jeonju University, 303 Cheonjam-ro, Wansan-gu,
Jeonju, Korea
Tel.: +82-63-220-2936, Email: leeyj@jj.ac.kr

I. 서 론

4차산업혁명과 ICT 기술의 발전은 모든 산업 분야에 확산되어 다양한 기업 및 산업의 서비스에 융합되어 부가가치 창출을 하고 있다. 이중 IoT 기술은 4차 산업혁명의 중요한 기술로 기술의 변화에 빠르게 융합하여 다양한 서비스를 통한 새로운 시대를 열고 있다[1].

IoT는 인간과 사물, 서비스로 분산된 환경요소에 대해 인간의 명시적 개입 없이 상호 협력적으로 센싱, 네트워킹, 정보 처리 등 지능적 관계를 형성하는 사물 공간 연결망으로 정의되고 있으며 또한 IoT는 인터넷을 통해 인간이 생성하고 가공한 데이터, 정보 및 지식뿐만 아니라, 우리 주변의 센싱과 구동체 기능을 갖는 일상 사물로부터 정보를 획득할 수 있다. 이를 위해 IoT는 자체 식별자를 갖고 컴퓨팅과 통신 기능을 통해 인간과 사물, 서비스로 분산된 환경요소에 대해 인간의 명시적 상호 협력적으로 센싱, 네트워킹, 정보 처리 등 지능적 관계를 형성할 수 있다[2].

최근 IoT 기반의 다양한 스마트 환경에서 상황(Context)에 적합한 고도화된 맞춤형 서비스를 제공하기 위해서는 동적으로 변화하는 상황 정보가 필요하다. 특히 IoT 기반의 지능적인 서비스를 제공하기 위해서는 사용자와 환경의 상황을 정확하게 파악할 필요가 있으며, 이를 위해서는 상황 인식 기술이 필수적이다. 여기서 상황은 공간에서 설치된 센서로부터 가져올 수 있는 사용자 정보와 관련된 환경정보를 의미하며 이러한 상황정보를 얻어내는 과정을 상황인식(Context awareness)이라고 한다[3][4].

상황인식 기술을 활용할 수 있는 대표적인 융합기술로 증강현실이 있다. 증강현실은 다양한 콘텐츠와 융합 가능한 주목 받고 있는 4차산업혁명의 융합기술의 핵심으로 부상하고 있다. 특히 실감형 인터페이스 기술은 시각, 청각, 동작을 스마트미디어와 정서적 교감을 나누며 현실과 가상의 경계를 넘나드는 다채로운 경험을 선사할 뿐 아니라 사용자의 오감을 자극하는 몰입도 높은 증강현실 환경을 제공을 통해 고객 참여를 새로운 차원으로 끌어 올릴 수 있다[5].

증강현실(AR, Augmented Reality)은 실제 공간에

가상정보를 실시간으로 증강하여 사용자가 증강된 가상정보와 상호작용함으로써 작업 효율성을 향상하는 기술이다. 이와 유사한 가상현실은 자신(객체)과 배경, 환경 모두 현실이 아닌 가상의 이미지를 사용하고, 현실과 단절된 ‘가상 세계에서의 몰입과 상호작용’을 강조하는 것과 달리 증강현실은 현실의 이미지나 배경에 3차원 가상 이미지를 겹쳐 하나의 영상으로 보여주는 기술로 현실과 유기적으로 결합한 ‘확장세계에서의 지능적 증강과 직접적 상호작용’을 강조한다. 현실공간을 직접 다양한 형태의 미디어로 활용하는 증강현실은 가상세계 구성을 위한 3차원 모델링의 부담을 줄인 반면, 사용자 이동과 주변 환경 변화에 대응하여 실시간 정보나 콘텐츠, 몰입을 위한 인간 친화적인 인터페이스를 제공해야 하므로 기술적 어려움이 있다[6].

특히 컴퓨터, 비디오와 오디오 등의 첨단 기술들이 결합된 스마트 미디어의 단순한 기능적인 내용뿐만 아니라 생활의 편리성과 감성을 충분히 충족시키 위한 구체적인 연구가 필요하다[7].

II. 인간 친화적 인터페이스

인간 친화적인 콘텐츠는 우리의 삶과 생활에 있어 중요한 역할을 담당하고 있다. 또한 감성에 기반한 인터페이스는 제품의 가치 부여와 역할을 결정하는 주요 핵심 기능으로 큰 관심을 가지며 자리를 잡고 있다. 특히 게임과 같은 대화형 미디어 콘텐츠의 감정 인터페이스는 디자인의 기능성은 물론 심리적 만족감까지 부여하여 서비스를 사용자에게 제공해 줄 수 있다[6][7].

사용자를 위한 콘텐츠의 감정적 접근과 솔루션은 많은 공감을 일으키며 본래의 기능을 넘어선 다양한 경험을 선사해 정신적인 만족까지 시켜주고 있으며 특히 스마트 미디어와 사용자 사이의 감성적 소통을 통해 깊은 유대감과 신뢰감을 만들어 낼 수 있다. 따라서 인간의 신체의 움직임 즉, 자세에 대한 구체적인 연구와 정보가 필요하다.

여기서 인체의 자세를 추정하는 연구는 얼굴, 어깨, 팔, 손, 손가락 등을 주 대상으로 하며 본 연구에서는 사람의 몸에 장비를 장착하지 않고 영상으로 사람의 얼굴 검출과 그 응용 및 손과 손가락에

대한 위치를 딥러닝 기반의 접근 방식을 사용해 실생활 및 응용분야에 적용 가능한 시스템과 환경을 만들어 다양한 분야에서 생산성을 향상시킬 수 있는 방법을 제공하고자 한다.

얼굴 검출과 인식 기술의 경우 보안, 타깃마케팅, 광고 뿐 아니라 출입 관리 시스템, 사회 안전망 강화 등 그 사용 용도가 광범위하다. 이를 위해 얼굴의 검출, 위치와 방향 추정, 표정 감정 인식, 얼굴인식 등의 다양한 응용이 필요하다[8]-[13].

손과 손가락을 검출 및 인식 기술은 자연스럽고 편안한 인터페이스로 활용 할 수 있는 기술로 다양한 기기와 상호연동 할 수 있는 향상된 사용자 상호작용 인터페이스 시스템 구성을 가능하게 하는 필요한 기술이다.

2.1 손자세 인식 및 KNN(K-Nearest Neighbors)

손 자세 추정 경우는 원손, 오른손의 구분과 손가락을 사용하여 다양한 정보 표현이 가능하나 이를 정확하게 인식하고 활용하기는 어려운 편이다. 따라서 이에 대한 많은 연구가 진행 중이며 특히 CNN(Convolutional Neural Network), RNN(Recurrent Neural Network)과 같은 AI 기반의 다양한 연구가 된 이후에야 일반적인 손 자세 추정이 가능해졌다 [8][9].

손 표현 인식을 위해서는 손 자세 인식을 위한 정적인 모델과 손 동작 인식을 위한 동적인 모델이 필요하다. 정적인 모델은 손 관절의 위치와 상태를 지속적인 움직임 아닌 정지상태가 있는 모델로 예를 들면 손가락을 모두 편 정지 상태의 모습이 있는 상태로 이때의 손가락의 각 특징점 별 위치 좌표를 이용해 인식하는 방법이다. 동적인 모델은 손의 지속적인 움직임을 나타내는 모델로 예를 들어 손가락의 검지를 사용해 회전하는 모양을 나타내며 이 움직임을 지속적으로 모니터링해 손 특징점 별 움직임을 계산해 인식하는 모델이다[9].

손 동작 인식은 손이 움직이는 방향과 거리를 통해 특정 움직임으로 인식 할 수 있다. 즉, 손가락 특징점인 관절 위치 추정과 원손, 오른손, 엄지, 검지, 중지, 약지 등을 구분해 두 손과 손가락의 구분

을 통해 각각의 거리와 방향을 표현할 수 있다. 이 기능을 활용하여 효과적인 손가락 인식 통한 다양한 인터페이스가 가능하다. 구글의 미디어파이프는 [10] 파이프라인 처리를 통해 다양한 데이터를 처리 할 수 있도록 만들어진 크로스 플랫폼 프레임워크로 이미지에서 객체를 감지하고 객체 인식 데이터 세트에 대해 훈련된 머신러닝(ML) 모델을 통해 객체의 위치 및 자세 추정에 사용할 수 있도록 학습된 얼굴, 신체, 손 등의 모델을 제공하고 있다. 손 관절과 관련하여 21개 손가락 특징점(그림 1참조)을 제공한다. 본 연구에서는 정적 모델과 미디어파이프 [10][11], KNN[12]을 사용해 구성한다.

그림 1은 손 랜드마크 모델로 이미지에 대한 손 바닥 감지 후 직접 좌표 예측을 통해 감지된 손 영역 내부의 21개의 특징점과 특징점 이름이다. 손가락 인식을 위해 사용한 KNN(K-Nearest Neighbors)은 머신러닝에서 사용되는 분류(Classification) 알고리즘이다. 유사한 특성을 가진 데이터는 유사한 범주에 속하는 경향이 있다는 가정하에 사용한다.

n개의 특성(Feature)을 가진 데이터는 n차원의 공간에 점으로 개념화 할 수 있으므로 유사한 특성을 가진 데이터들끼리는 거리를 구해 분류할 수 있다.

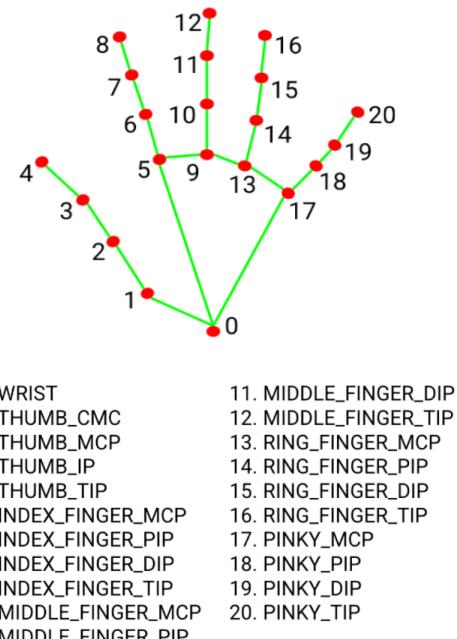


그림 1. 미디어파이프의 손 랜드마크

Fig. 1. Hand landmarks in the mediapipe

이때 분류를 알 수 없는 데이터의 경우 가장 가까운 이웃 k 개의 분류를 확인하여 결정할 수 있으며 분류기의 효과를 높이기 위해 파라미터를 조정할 수 있다[12][13].

미디어파이프 사용해 구한 21개의 특징점을 벡터 정규화하고 정규화 된 조인트 점들의 벡터를 내적 공식을 활용해 두 벡터의 각도를 구해 데이터 파일을 만든다. 이를 위해 입력의 다양한 손가락 표현 영상에서 21개의 조인트 점을 미디어파이프를 사용해 구하고 레이블에 따른 영상에서 마우스를 클릭해 조인트 별 두 벡터의 각도를 구해 csv 파일에 기록한다. 이를 위해 본 연구에서는 편리하게 마우스를 사용하여 클릭할 때마다 각각의 레이블에 대한 데이터셋을 자동으로 만드는 방법으로 데이터셋을 작성하였다.

작성한 데이터의 성능평가를 위해 실험에서는 학습데이터: 70%, 실험데이터: 30%로 구성해 실험하였고 그 결과 0.9484(SCORE 값)로 비교적 우수하였다.

그림 2는 본 연구에서 사용한 KNN 기법에서 최적의 K 를 구하기 위하여 데이터를 사용하여 k 의 변화(1-100까지)에 따른 검증 정확도(Validation accuracy)의 변화를 측정해 출력한 그래프 결과이며 좌상향에서 우하향의 기울기를 가지고 있으며 특히 k 가 0-6까지는 비교적 완만한 기울기를 보이고, 이 후 비교적 급한 경사와 완만한 경사를 반복하고 있다. 본연구에서는 0.94의 값을 갖는 $k=3$ 로 설정해 알고리즘을 구현한다.

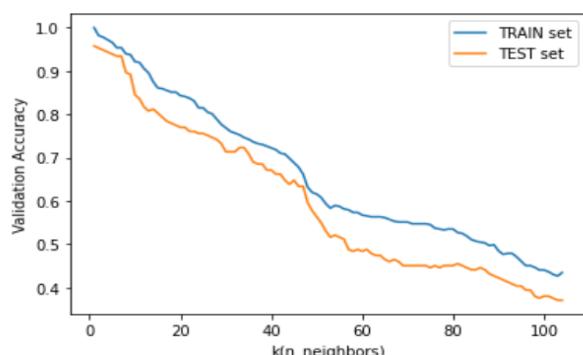


그림 2. 거리 K 가 1-100 사이로 변경될 때 Validation Accuracy 변화 그래프

Fig. 2. Validation accuracy change graph when distance K is changed between 1-100

2.2 감정 기반 인터랙티브 미디어 평가

게임과 같은 양방향 미디어 콘텐츠 사용자들의 평가를 수집하기 위하여 사람들의 리뷰나 게시판, 설문지, 이메일, SNS 등을 통한 평가가 전의 방식이었다. 그러나 이 방법은 단순 평가로 조작 등의 가능성이 있어 신뢰도가 떨어지게 된다. 이 같은 문제점을 해결하기 위해 사용자의 미디어 사용 시 나타나는 긍정적 경험 혹은 부정적인 경험과 관련된 감정 정보를 활용해 사용하고 있는 콘텐츠에 대한 사용자 만족 여부를 확인한다. 얼굴은 감정이 표현되는 주요 채널[14]로 표정으로 그 여부와 강도를 확인 할 수 있다. 표정의 감정표현은 콘텐츠를 접할 때 나타나는 비언어적 표현 수단이며 특히 얼굴 표정은 감정이 표현되는 주요채널로 사용자의 얼굴 표정 정보 추론을 통해 사용자의 감정 정보를 이해하고[15] 이에 기반한 몰입 정도와 만족도 등의 관계를 유의미하게 분석하고 활용한다.

본 연구에서 제작한 게임 콘텐츠를 사용자가 즐기면서 느끼는 감정의 변화에 대해서 입력한 얼굴 검출영상에서 감정정보를 추출해 실시간 표정 분석을 통해 사용자의 콘텐츠 사용 소비 욕구에 대한 만족도를 분석하고 활용한 최적의 대응 방법에 대한 연구와 실험을 진행한다. 실험을 위해 머신러닝 라이브러리 dlib과 학습 데이터셋으로는 FER2013 [16]을 사용한다. 이 공공 데이터셋을 평상심, 행복, 슬픔, 놀람, 화남, 혐오감, 두려움(neutral, happy, sad, surprised, angry, disgust, fear) 총 7가지의 감정 데이터를 사용하고 본 연구의 취지에 맞게 불만족(화남, 혐오감, 두려운, 슬픔), 평상심, 행복, 놀람으로 재구성해 사용한다.

2.3 감정 기반 인터랙티브 미디어 평가를 위한 눈, 입 정보 검출

콘텐츠 소비자의 콘텐츠의 사용경험과 소비욕구에 대한 객관적인 평가를 위해 여러 연구들이 다양하게 진행중이다[17]. 그러나 소비자의 표정에 따른 시각 정보를 활용하여 만족도를 알기 위해 영상처리 기법을 이용하여 인식하는 일은 쉽지 않다. 따라서 소비자의 추가적인 행동 분석과 뿐 아니라 종합

적인 실시간 감정인식 기술 개발이 필요하며 실시간 인식을 위한 알고리즘 개발과 데이터 처리 및 대응 방법 등의 활용방법과 환경구축에 대한 연구가 필요하다. 이를 위해 본 연구에서는 눈과 입의 움직임 정보 즉 물입에 따른 눈의 깜빡임의 빈도 및 지루함을 나타내는 하품 빈도 등을 사용자 신체 정보를 검출해[18][19] 객관적인 콘텐츠 사용에 대한 사용자 감정 정보를 구하고자 한다. 눈 깜박임(Eye blinking)은 일상에서 여러 환경적인 요인으로부터 눈을 보호하기 위하여 일어나기도 하고 이와는 달리 개인의 감정이나 인지 상태가 반영되어 눈을 깜빡이기도 한다. 인간의 눈깜빡임은 평온한 상태에서 보통 분당 약 15회 정도 발생하는데 개인의 인지나 감정상태가 반영되면 눈깜빡임 속도가 변화한다고 알려져 있다[20][21].

눈 깜빡임 시간과 집중력의 변화 사이에 밀접한 연관이 있고 이는 기억력이라는 매개체를 통해 측정, 평가 할 수 있다 즉 자연스럽고 연속적인 상황에서 시시각각 변하는 눈 깜빡임 시간의 관찰을 통해 눈 깜빡임 시간이유의미하게 증가한 지점, 즉 우리가 집중하고 있을 것으로 예측되는 지점이다[21].

그러나 주의력이 감소 된 경우 사람들은 눈을 감거나 아주 작게 뜨며, 하품 등의 얼굴 특징으로 쉽게 구별할 수 있다. 하품을 인식하는 기준 방법의 경우 얼굴과 입의 모양 및 색상 정보를 이용하여 눈 개폐 상태와 하품 여부를 확인 하는 알고리즘, 입의 위치 및 색상 정보를 이용하여 하품 여부를 확인하는 알고리즘[22]-[25] 등 다양한 연구가 진행 중이다. 본 논문에서 웹캠을 사용한 실시간 검출을 위하여 머신러닝 라이브러리 dlib과 shape_predictor_68_face_landmarks.dat를 사용하여 얼굴과 입의 위치 정보를 확인해 효과적인 입의 개폐 여부 및 입벌림의 크기를 판단하여 하품 여부를 검증한다. 눈의 깜빡임, 하품과 같은 현상발생을 통하여 사용자가 현재 사용하고 있는 콘텐츠에 대한 물입의 수준을 민감하게 반영할 수 있다. 입 검출과 눈검출을 알고리즘을 사용하여 68개의 얼굴 특징의 x,y 좌표를 추출 할 수 있으며 이중 눈 가장자리, 입술 바깥쪽, 코끝, 턱끝의 24개 특징으로 축소해 눈과 입의 움직임 변화 정보를 하나의 알고리즘[25]을 통해 거의 동시에 추출해 낼 수 있다.

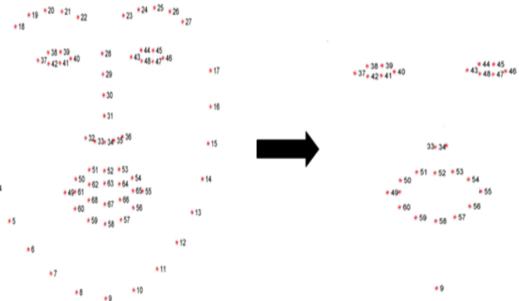


그림 3. 입과 눈의 움직임을 확인하기 위해 68개의 특징점 중 사용하는 24개의 특징점 영상

Fig. 3. 24 feature points images used out of 68 feature points for check the movement of the mouth and eyes

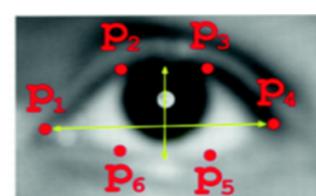


그림 4. 눈 특징점과 EAR 수식

Fig. 4. Eye feature points and EAR formula

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \quad (1)$$

눈 특징점 추출을 이용한 눈 깜빡임 검출 방법 예로 EAR(Eye Aspect Ratio)[25]을 사용한다. EAR는 눈의 너비에 대한 길이의 비율이다. 눈의 길이는 아래 그림과 같이 눈을 가로지르는 두 개의 뚜렷한 수직선을 평균화하여 계산한다(그림 4, 식 (1) 참조).

이때 사용한 가설은 연속 프레임에서 사용자가 눈 종횡비가 감소가 시작하고 사용자가 관심이 없이 졸린 것으로 예상할 수 있다. 그러나 EAR 값이 작게 표현되기 때문에 필요에 따라 식 (1)의 역수 수평값분에 수직값으로 나타내기도 한다.

입의 개폐 여부와 열린 정도를 확인하기 위하여 식 (2)를 사용한다(그림 5, 식 (2) 참조)[25].

MAR(Mouth Aspect Ratio)은 입의 너비와 입의 길이의 비율을 측정한 것으로 값이 커지면 콘텐츠가 재미가 없거나 관심이 줄어든 가능성이 있으며 식 1과 같이 역수 수평값분에 수직값으로 나타내기도 한다. 본 연구에서는 역수를 사용해 눈, 입 열린 정보를 표현한다.

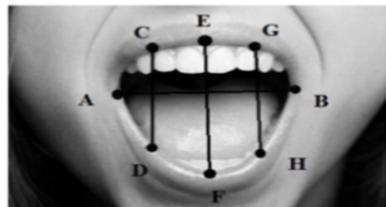


그림 5. 입 특징점과 MAR 수식

Fig. 5. Mouth feature points and MAR formula

$$MAR = \frac{|EF|}{|AB|} \quad (2)$$

III. 제안한 알고리즘과 게임 구성

그림 6은 제안한 알고리즘 구성에 관한 순서도이다. 프로그램이 시작되면 게임과 시스템과 관련된 웹캠, 아두이노, 센서등의 프로그램 초기화가 진행된다. 이때 웹캠, 아두이노, 프로그램 초기화에 문제가 발생한 경우 에러 내용이 출력되어 에러 내용에 따른 처리가 가능하게 했다. 문제가 없는 경우 게임이 시작되면서 배경(그림 7 참조) 안에서 적 캐릭터, 메인 캐릭터가 등장하며 메인 캐릭터인 차는 사용자의 손과 손가락을 인터페이스로 직진, 정지, 후진, 왼쪽 회전, 오른쪽 회전 등의 움직임을 통해 게임이 레이싱 게임이 진행된다.

게임 진행중 사용자의 감정정보를 인공지능을 사용해 검출하고 사용자의 7가지의 감정 데이터를 사용해 연구의 취지에 맞게 불만족, 평상심, 만족(행복, 놀람) 등으로 재구성해 사용자 감정 정보를 읽는다. 만일 감정지수가 낮은 경우엔 게임 난이도, 배경음악 등 여러 요인들을 조정하여 사용자가 쉽게 몰입해 게임을 진행하도록 유도 한다. 다음 루틴으로 사용자의 눈 깜빡임과 하품의 유무를 판단을 위해 입크기 검출 등을 통하여 관심에 따른 몰입도, 지루함 여부 및 정도를 판단한다. 제어가 필요한 경우엔 배경음악, 조명, 난이도 등을 추가적으로 조정 한다. 마지막 단계로 눈 깜빡임 횟수의 감소로 인한 눈건조증, 이산화탄소 과다로 인한 집중력 감소 등을 방지하기 위하여 아두이노와 온습도 센서, 이산화탄소 센서를 사용해 사용자 주변의 물리적인 환경요소를 검출해 가습기 구동이나 통풍 제어가 가능하도록 하는 최적의 환경 구축 루틴과 앞서 나온 정보 데이터 결과를 추가로 분석, 추론하여 전반적

인 미세조정을 추가해 가상과 현실 공간에서 사용자 감정과 신체정보를 활용한 몰입 증강을 위한 맞춤형 기술을 구현한다.

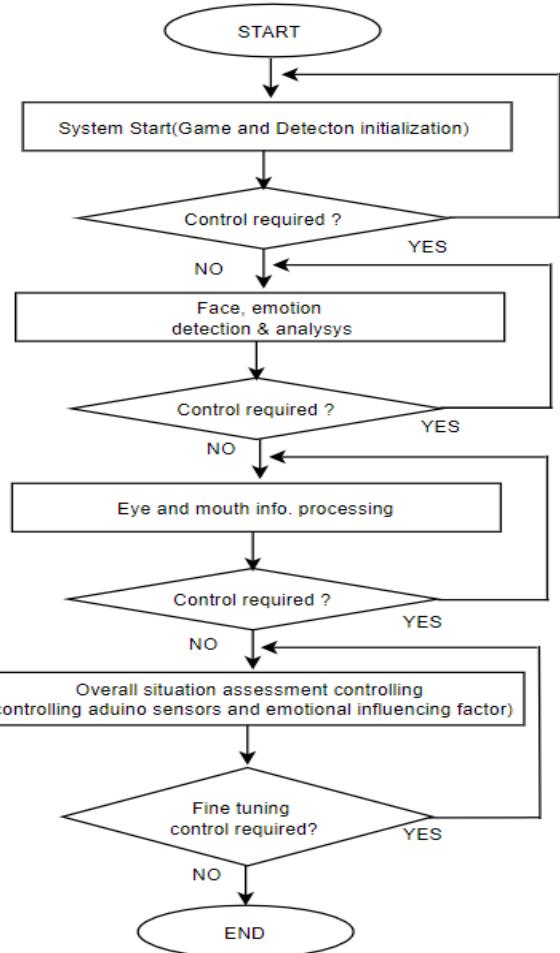


그림 6. 제안한 알고리즘 순서도

Fig. 6. Proposed algorithm flowchart



그림 7. 제안한 알고리즘 실험을 위해 제작한 3인칭 탑뷰 레이싱 게임

Fig. 7. Third-person top-view racing game created for the proposed algorithm experiment

3.1 게임 구성

그림 7은 제안한 알고리즘 성능을 실험하기 위해 제작한 3인칭 탑뷰 레이싱 게임이다. 배경을 기반으로 3개의 움직이는 객체로 구성되어 있다. 두 개의 공중 이동 캐릭터와 메인 캐릭터인 자동차로 구성되어 있으며, 두 개의 공중 이동 캐릭터는 적 캐릭터로 메인 캐릭터인 자동차를 공격할 수 있다.

이때 메인 캐릭터인 자동차는 이들을 피해 전후 좌우로 이동할 수 있다. 메인 캐릭터의 이동은 사용자의 왼손과 오른손 혹은 양손으로 제어할 수 있는데 이와 더불어 양 손가락으로 여러 모양의 입력을 만들어 메인 캐릭터의 다양한 행동을 구현할 수 있다. 예를 들어 정지(주먹), 앞으로 이동(보), 뒤로 이동(락) 등등이 가능하며 왼손가락 엄지와 중지를 사용하여 왼쪽과 오른쪽으로 회전 이동할 수 있다. 또한 아래 공중 이동 캐릭터는 미사일 공격을 할 수 있는데 왼손의 중지('1')을 표시하면 미사일을 발사한다. 이때 사용자 인터페이스는 왼손, 오른손, 양손 손가락으로 미디어파이프와 KNN 기법을 사용해 손, 손가락표현 제스처를 빠른 실시간 인식이 가능하다.

IV. 실험

4.1 실험 1

인터랙션의 방식은 점차 감성적이고 직관적인 방향으로 발전해 가고 있다. 언어를 통한 인터랙션은 언어의 표현 능력으로 감정표현에 한계가 있으며 텍스트를 통한 인터랙션의 경우 제한적인 인터랙션을 제공한다. 이 같은 한계를 극복하기 위하여 얼굴 표정, 손, 손가락, 머리의 움직임 등 비언어적인 인터랙션 접근 방식이 새로운 대안으로 부각되고 있다[26].

손의 모양과 움직임을 비접촉식으로 인식하는 기술은 여러 기술 영역과 플랫폼에서 서비스 이용자에게 별도의 센서 디바이스나 접촉이 필요없는 편안하고 사용하기 좋은 인터페이스이다.

특히 게임과 같은 실시간 상호작용이 중요한 콘텐츠의 경우 손, 손가락과 같은 신체 인터페이스는

개인화, 단순화, 감성화를 추구하는 새로운 기술 트랜드로 자리 매김 할 수 있다. 손 자세 정보 추정을 위하여 왼손, 오른손의 구분을 기본으로 이동과 회전, 손가락 관절의 위치 분석에 따른 특정 자세 추론 및 자세가 의미하는 표현의 인식이 필요하다. 이를 위해 구글의 미디어파이프를 사용한다[27].

실험 1에서는 미디어파이프를 사용해 찾은 21개의 특징점(랜드마크)의 좌표와 제작한 데이터 세트(Data set)을 사용해 적용한 KNN 알고리즘을 사용해 손가락을 인식하였다.(그림 8 참조)

그림 8은 왼손과 오른손을 인식하고 손가락 정보를 인식하여 그 결과를 출력한 영상이다. 그림 8(a)는 왼손의 경우 왼손의 각 특징점 부분 표시와 과 엄지와 중지를 끝 위치를 인식해 연결한 결과이며 오른손은 각 조인트 표시와 숫자 '1'을 표시한 것을 인식해 출력한 그림이다.

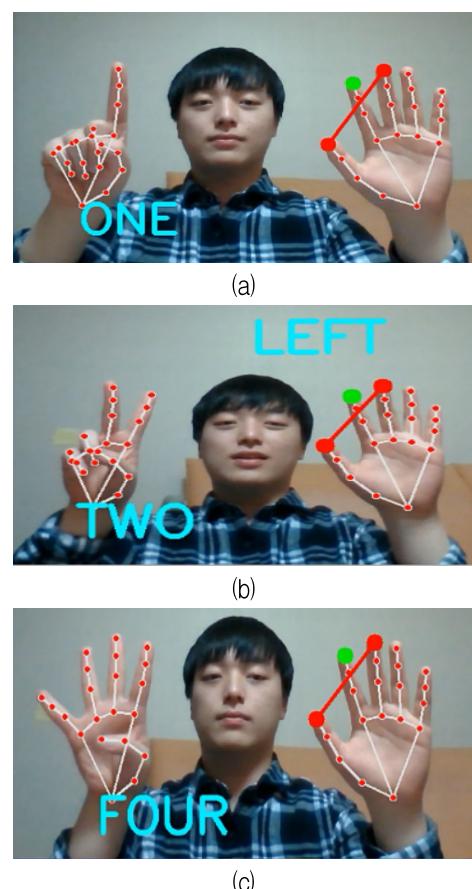


그림 8. KNN과 미디어파이프 랜드마크를 사용하여 손가락 정보를 인식한 결과 영상

Fig. 8. Result image of fingers information expression recognition using KNN and Mediapipe landmark

그림 8(b), (c)는 숫자 ‘2’, ‘3’을 인식해 출력한 결과 영상이다. 실험에서 사용한 ‘1, 2, 4’, 외에 3, 주먹, 락(Rock), 스파이더맨(Spiderman) 등 다양한 입력을 실험한 결과 제작한 데이터셋을 사용한 KNN 알고리즘을 통해 잘 인식 됨을 확인할 수 있다.

4.2 실험 2

FER-2013 얼굴데이터베이스를 활용해 얻을 수 있는 감정 데이터는 얼굴 검출을 통해 표정으로 7 가지의 감정 데이터에서 3가지로 분류하고 각각의 문턱값(Threshold) 크기에 따라 평상심 상태와 행복 상태로 구분해 적절한 클래스로 분류하였다. 콘텐츠에서의 감정 데이터의 활용은 실시간 데이터 수집 및 응용이 중요하므로 빠른 처리가 필요하며 정확한 감정인식으로 인한 시간 지연을 최소화하는 기법이 필요하다. 또한 사용자 얼굴이 자세 특히 경사진 얼굴의 인식도 가능해야 한다. 이를 위해 얼굴인식은 쉽게 이미지 처리 및 인공지능 기능을 실행할 수 있도록 컴퓨터 비전 패키지(Cvzone, 케라스, dlib, opencv) 등을 사용한다[26].

그림 9는 얼굴인식을 통해 사용자의 감정 정보를 손가락 인터페이스를 사용해 콘텐츠를 제어하는 중 익은 결과이다. 현재 감정 데이터 중 만족도와 관련 있는 행복:0.56로 가장 크며, 평상심:0.36, 슬픔:0.04, 화남:0.03, 두려움:0.01, 놀람-역겨움:0.0이다. 그림(a)에서 보면 게임의 메인 캐릭터인 차가 왼쪽 벽에 충돌하기 직전의 상태와 감정 정보를 보면 사용자가 일상적인 상태로 콘텐츠에 관심이 시작되는 초입 단계(행복-평상심)의 상태로 약간의 긴장이(슬픔-화남-두려움 등) 있음을 예측할 수 있다. 이런 경우 게임에 대한 몰입도를 높이기 위해 게임 난이도, 배경음악, 차 제어 자유도 확장, 조명 등 여러 다양한 방법으로 사용자 감정정보에 대한 변화를 유도할 수 있다.

입력영상으로부터 제안한 방법을 통해 실시간으로 7가지의 감정 정보 값이 잘 검출되었으며, 이를 근거로 사용자의 만족, 불만족, 평상심 등의 감정 상태를 구분할 수 있음을 확인 할 수 있다.

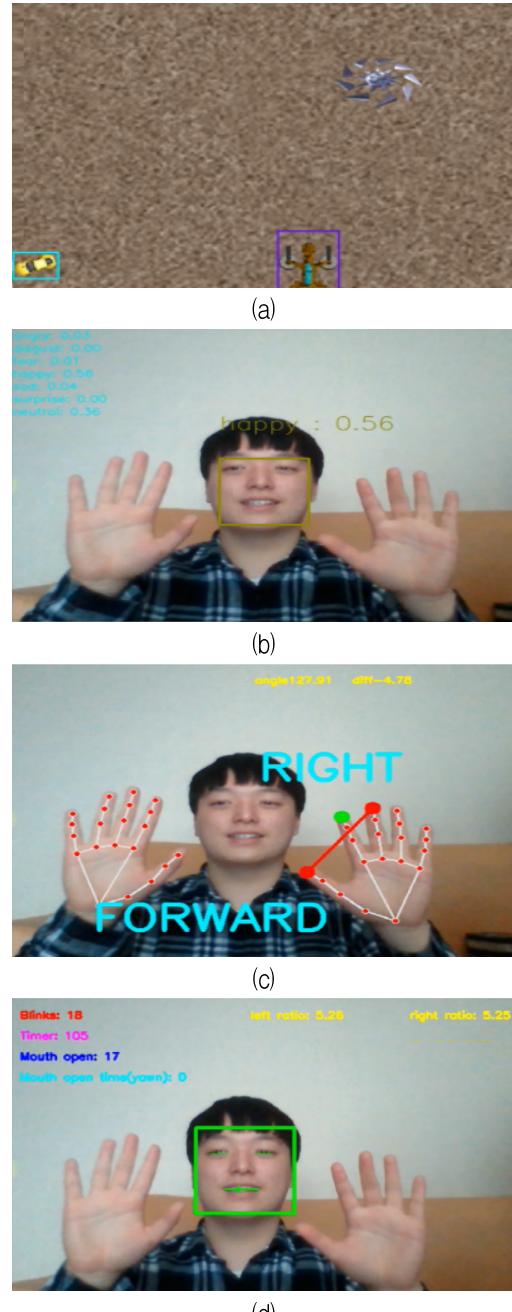


그림 9. 실험 2 결과 영상
Fig. 9. Experiment 2 result image

4.3 실험 3

실험 2에서의 사용자 만족도와 관련 감정정보 행복:0.56, 놀람: 0.0으로 낮은 편이다. 이 같은 상황을 개선하고자 실험3에서는 게임의 난이도를 게임2의 상황보다 약간 낮추어 변경하고 사용자에게 게임을 하도록 했다. 그 결과 그림 10(c), (d)와 같은 만족도 관련 요소가 다소 올라간 만족도 관련으로 감정

정보를 확인 할 수 있다. 실험 3은 제작한 레이싱 게임에서 사용자와의 공격 상호작용 루틴으로 오른 손가락 중 검지를 사용해 메인 캐릭터인 카를 향해 공격하는 신호를 주면 KNN을 사용해 오른손-검지 손가락을 인식하고 게임루틴에서 미사일을 발사시킨다(그림 10(a), (b) 참조). 미사일 영역과 car 영역의 사각형이 중첩되는 경우엔 폭발 이벤트를 게임 루틴에서 실행한다.

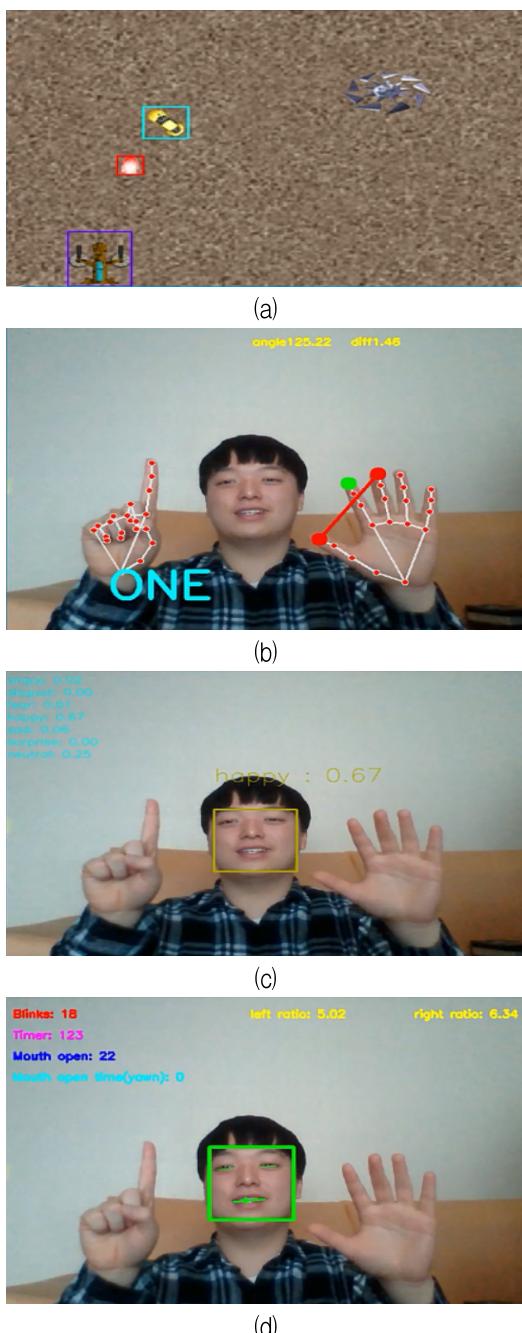


그림 10. 실험3 결과 영상
Fig. 10. Experiment 3 result image

그림 10(c)는 사용자의 감정상태로(화남:0.02, 역겨움:0.0, 두려움:0.01, 행복:0.67, 슬픔:0.06, 놀람:0.0 평상심:0.25)로 만족도와 관련있는 행복:0.67로 실험 2보다는 행복데이터가 상승되었음을 확인 할 수 있다. 또한 평상심의 경우도 0.36에서 0.25로 데이터가 낮아졌음을 확인 할 수 있다.

이는 실험 3이 실험 2보다 만족도에 관련된 감정 정보에 유의미적인 변화가 있음을 확인할 수 있다. 눈 깜빡임은 7.1초, 입은 22정도 열린 상태로 하품은 하지 않았고, 왼쪽 눈은 5.02 오른쪽 눈은 6.34 상태이다. 실험1의 눈 깜빡임은 5.8초, 입은 17정도 열린 상태로 하품은 하지 않았고, 왼쪽 눈은 5.25 오른쪽 눈은 5.26 상태로이다. 눈 깜빡임의 경우 정상적인 경우로 대략 4초 정도로 유추한다면 실험2는 실험1보다 현재 상황에 관심을 가지고 응시하고 몰입 상태 입을 알 수 있으며 하품은 하지 않은 상태며 개안의 크기는 서로 유사한 경우이다.

즉 개인차는 있지만 상대적으로 나이도가 쉬워져 사용자가 비교적 쉽게 미사일 공격 패턴을 사용하므로 공격에 대한 몰입과 이를 관심 있게 지켜보는 상황으로 예측할 수 있다. 또한 실험에서 게임중 오른손 검지 손가락을 잘 인식했으며, 눈의 깜빡임과 하품, 개안 크기 등을 제안한 알고리즘을 통하여 실시간으로 잘 동작 됨을 확인 할 수 있다.

4.4 실험 4

사용자가 게임에 점점 몰입할수록 눈깜빡임이 점차 느려짐을 알 수 있다. 이는 시각 정보 소실을 막기 위한 행동으로 이로 인해 눈건조증과 같은 질병을 유발 시킬 수 있다.

실험 4에서는 온도와 습도 센서를 추가하여 현재 환경에 대한 온, 습도 데이터를 모아 상황을 판단해 가습기, 에어콘 가동이나 신선한 공기 유입을 한 환기를 사용자가 인식하지 못하더라도 자동으로 하게끔하는 최적의 환경을 제공 할 수 있도록 환경구축 실험이다. 안조건조증을 예방하기 위해서는 실내 온도 22도, 습도는 40-60% 정도를 유지해야하며 의식적으로 눈을 자주 깜박이는 것이 좋다[28][29].

또한 신선한 공기 유입을 위하여 이산화탄소 (CO_2)양을 검출해 제어할 필요가 있다. 실내 이산

화탄소 농도가 1,000 피피엠(ppm)을 초과하면, 사람들이 집중력 저하와 졸음을 느끼고, 심한 경우 두통을 겪을 수 있기 때문에, 실내공기질 관리법은 이산화탄소의 실내 기준 농도를 1,000 피피엠으로 권고하고 있다[30]. 그림 11(a)는 사용자가 그림 11(b) 와 같은 제어 즉 앞으로 왼쪽 회전 상태 제어를 하면서 게임을 하고 있다.

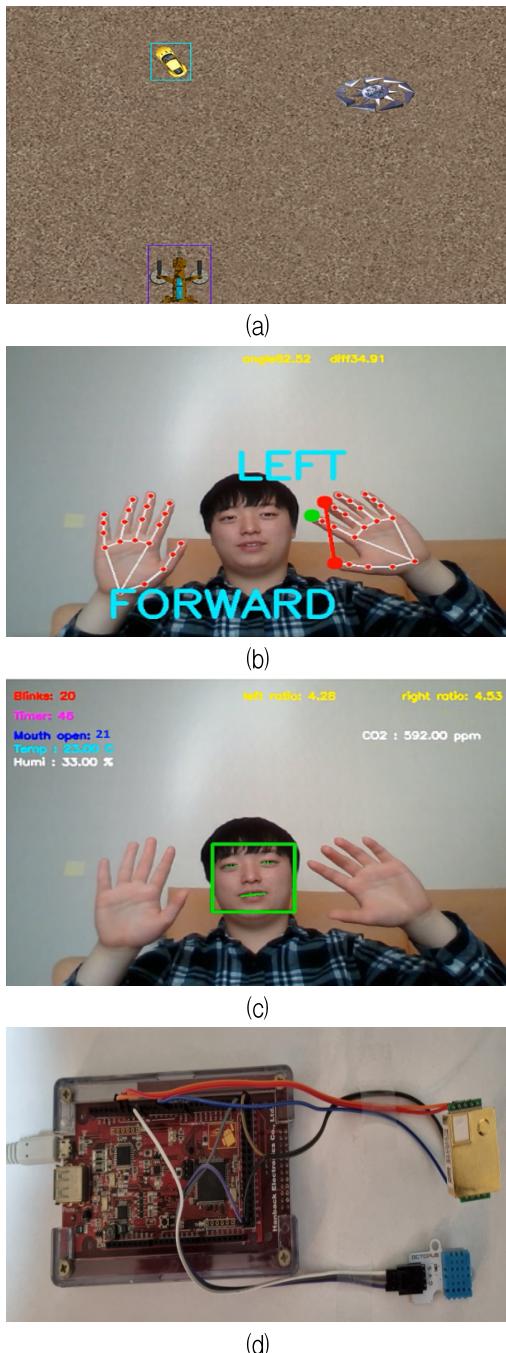


그림 11. 실험4의 결과 영상
Fig. 11. Experiment 4 result image

그림 11(a)는 사용자가 그림 11(b) 와 같은 제어 즉 앞으로 왼쪽 회전 상태 제어를 하면서 게임을 하고 있다. 이때 사용자의 주변 환경변수인 온도, 습도, 이산화탄소 값을 측정해 출력하고 있다. 온도 습도를 측정해 온도나 습도가 기준보다 낮은 경우나 게임의 몰입으로 인하여 눈 깜빡임이 현저하게 줄어든 경우 안조 건조증과 같은 불편한 상황이 생기지 않게 환경 개선을 할 수 있다. 습도가 낮은 경우엔 가습기를 구동하고 온도가 낮거나 높은 경우 냉, 나방기를 가동할 수 있다.

또한, 이산화탄소 과다로 인한 집중력 감소 등 몰입을 방해하는 요소들을 측정해 이산화탄소 농도가 높은 경우 환기를 통해 온도와 이산화탄소 농도를 적절하게 조정할 수 있다. 그림 11(d)는 현재의 온도, 습도, 이산화탄소 측정을 위해 사용한 아두이노와 온습도센서, 이산화탄소 센서이다. 측정결과 온도 23도, 습도 33%, 이산화탄소: 592피피엠으로 온도, 습도는 약간 높거나 낮은 편이며 온도와 이산화탄소는 양호한 편이다. 실험 4를 통하여 제안한 방법을 통하여 사용자가 보다 쾌적한 환경에서 콘텐츠를 사용할 수 있는 환경을 구축할 수 있는 가능성을 확인할 수 있다.

V. 결 론

사용자와의 가상공간 사이의 교감 강화를 위하여 확장된 인터페이스를 제안하고 이를 통한 사용자의 감정반응을 인공지능을 사용해 검출, 추론하여 반응하는 인공지능 몰입증강 기술을 제안하였다. 제안한 인터페이스는 기존의 인터페이스들에 비해서 사용자에게 직관적이고 자연스러운 인터페이스를 제공할 수 있으며, 사용자의 움직임이라는 단편적인 정보만 활용하는 것이 아니라 사용자의 감정 정보가 함께 고려된 형태의 개인화된 사용자 인터페이스 특징을 구현한 것이다.

레이싱 게임을 제작하여 메인 캐릭터인 카의 움직임을 사용자의 손과 손가락의 신체 인터페이스의 움직임 정보를 가상 공간의 메인 캐릭터에 적용하여 움직임을 제어하며, 이 과정에 사용자의 몰입 정도를 사용자의 얼굴에 나타난 감정 정보와 눈의 깜

빡임과 하품 정보를 제안한 인식 알고리즘을 사용해 분석하고 이 정보를 피드백하여 사용자의 감성을 만족하는 콘텐츠를 제공할 수 있는 환경 구축방법 제안과 응용 방법도 구현했다.

특히 사용자의 몰입 및 만족감 증가로 인하여 눈 깜빡임 횟수의 감소로 인한 눈건조증, 이산화탄소 과다로 인한 집중력 감소 등을 방지하기 위하여 아두이노와 온습도 센서, 이산화탄소 센서를 사용해 사용자 주변의 물리적인 환경요소를 검출해 가습기 구동이나 통풍 제어가 가능하도록 환경 구축을 하였고 실험을 통하여 그 가능성을 확인하였다. 입력 영상을 사용해 수행한 얼굴인식, 원손, 오른 손인식, 업지, 검지, 중지 등의 손가락 인식, 눈 깜빡임, 입 개폐 인식을 위한 비접촉식 알고리즘을 레이싱 게임에 적용해 실시간 상호작용을 하며, 아두이노를 사용한 환경정보 획득 및 적용과 지속적인 데이터 피드백을 통한 게임 난이도, 음악 등 감성적인 요인의 지속적인 실시간 처리 구현을 위해 게임과 제어 루틴을 스레드로 분리해 코딩하였다.

실험결과 감정인식 정보와 눈 깜빡임, 입의 개폐 정보 등을 제안한 알고리즘을 통해 종합적으로 분석할 수 있는 환경구축이 가능함을 확인하였으며, 또한 사용자의 복잡하고 섬세한 감정인식과 신체 변화에 대한 미세한 인식이 가능하였다. 또한 이 정보를 활용해 게임 난이도 조정과 공격 등의 이벤트 통한 재미 요소를 통한 상호작용을 하였고 이를 통해 제안한 알고리즘의 성능의 유의미한 가능성을 확인할 수 있었다.

향후 다양한 사용자에 대한 경험적 빅데이터 구축을 통하여 보다 세밀하고 정확한 사용자 맞춤형 감성 서비스에 대한 연구와 관심이 필요하다.

References

- [1] Young Jae Lee, "Mobile AR Contents Based on IoT(Internet of Things) Using Shading Language over Smartphone and Arduino Bluetooth Sensors", Journal of KIIT, Vol. 17, No. 2, pp. 109-116, Feb. 2019. <http://doi.org/10.14801/jkiit.2019.17.2.109>
- [2] Sachin Kumar, Prayag Tiwari, and Mikhail Zymbler, "Internet of Things is a revolutionary approach for future technology enhancement:a review", Journal of Big Data, Vol. 6, No. 1, Article number, 111, Dec. 2019. <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0268-2>.
- [3] Kim Do-hyeon, "IoT Context Awareness Technology Trend", Korea Electromagnetic Society, Electromagnetic Wave Technology, Vol. 24, No. 4, pp. 20-26, Jul. 2013.
- [4] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos, "Context aware computing for the internet of things : A survey", IEEE Communications Surveys & Tutorials, Vol. 16, No. 1, pp. 414-454, First Quarter, May 2013. <https://doi.org/10.1109/SURV.2013.042313.00197>.
- [5] <https://business.adobe.com/kr/resources/5-realworld-examples-of-augmented-reality-innovation.html> [acessed: Jun. 04, 2022]
- [6] <https://www.etri.re.kr/webzine/20170512/sub04.html> [acessed:Jun. 04, 2022]
- [7] Koo-Hyeong Lee, "Sentiment Interface of Smart Terminals", Journal of Information Proc. Society, Vol. 20, No. 1, pp. 23-30, Jan. 2013.
- [8] T. Chatzis, A. Stergioulas, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "A Comprehensive Study on Deep Learning-Based 3D Hand Pose Estimation Methods", Applied Sciences, Vol. 10, No. 19, pp. 1-27, Sep. 2020. <https://doi.org/10.3390/app10196850>.
- [9] Gyeongyang Heo, Bok-Deuk Song, and JiHong Kim, "Hierarchical Hand Pose Model for Hand Expression Recognition", JKIIICE, Vol. 25, No. 10, pp. 1323-1329, Oct. 2021. <https://doi.org/10.6109/jkiice.2021.25.10.1323>.
- [10] <https://mediapipe.dev> [acessed: Feb, 10, 2022]
- [11] <https://puleugo.tistory.com/11> [acessed: Feb. 10, 2022]
- [12] <https://hleecaster.com/ml-knn-concept> [acessed: Mar. 14, 2022]
- [13] <https://github.com/ntu-rris/google-mediapipe> [acessed: Feb. 10, 2022]

- [14] Ekman, P., "Differential communication of affect by head and body cues", *J. Pers. Soc. Psychol.* 2, pp. 726-735, 1965. <https://doi.org/10.1037/h0022736>.
- [15] R. Lowe and T. Ziemke, "The feeling of action tendencies: on the emotional regulation of goal-directed behavior", *Front. Psychol.* 2:346, 2011. <https://doi.org/10.3389/fpsyg.2011.00346>.
- [16] Arriaga, O, Valdenegro-Toro, and M. & Plöger, P., "Real-time convolutional neural networks for emotion and gender classification.", arXiv preprint arXiv:1710.07557. 2017.
- [17] HyeokJin Yun and Jaechoon Jo, "Research of Media Evaluation Index Generation Model based on Emotion ", Vol. 23, No. 5, pp. 71-79, 2020.
- [18] L. F. Bacher and W. P. Smotherman, "Spontaneous eye blinking in human infants: a review", *Dev. Psychobiology*, Vol. 44, No. 2, pp. 95-102, 2004.
- [19] M. J. Doughty, "Consideration of three types of spontaneous eyeblink activity in normal humans: during reading and video display terminal use, in primary gaze, and while in conversation", *Optometry and vision science*, Vol. 78, No. 10, pp. 712-725, 2001.
- [20] E. Ponder and W.P. Kennedy, "On the act of blinking", *Exp physiol*, Vol. 18, pp. 89-119, 1928.
- [21] Young Seok Shin, et al., "The Investigation of the Relationship between Eye Blink and Visual Attention with Video Clip", Vol. 35, No. 4, pp. 99-104, 2014.
- [22] J. A. Stern, L. C. Walrath, and R. Goldstein, "The enogenous eyeblink", *Psychophysiology*, Vol. 21, pp. 22-33, 1984.
- [23] T. Nakano, Y. Yamamoto, K. Kitajo, T. Takahashi, and S. Kitazawa, "Synchronization of spontaneous eyeblinks while viewing video stories", *Proc. R. Soc. B.*, Vol. 276, No. 1673, pp. 3635-3644, 2009.
- [24] M. Omidyeganeh, A. Javadtalab, and S. Shirmohammadi, "Intelligent Driver Drowsiness Detection through Fusion of Yawning and Eye Closure", Proc. of IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems, pp. 1-6, 2011.
- [25] [https://towardsdatascience.com/drowsiness detection with machine learning 765a16ca208a0](https://towardsdatascience.com/drowsiness-detection-with-machine-learning-765a16ca208a0). [acessed: May 02, 2022]
- [26] Octavio A., Matias V., and Paul P., "Real-time Convolution Neural Networks for emotion and Gender Classification", Cornell University, Oct 2017. <https://doi.org/10.48550/arXiv.1710.07557>.
- [27] <https://google.github.io/mediapipe/solutions/hands.html> [acessed: May 10, 2022]
- [28] <https://star8353.tistory.com/entry/%EA%B3%84%EC%A0%88%EB%B3%84-%EC%8B%A4%EB%82%B4-%EC%A0%81%EC%A0%95-%EC%8A%B5%EB%8F%84> [acessed: May 10, 2022]
- [29] E. Ponder and W.P. Kennedy, "On the act of blinking", *Exp physiol*, Vol. 18, pp. 89-119, 1928.
- [30] Donghae Lee and Youngbo Choi, "Measurements and CFD Analysis for Release Rate of CO₂ and Characteristics of Natural Ventilation in Lecture Room", *Journal of the Korean Society of Safety*, Vol. 36, No. 1, pp. 86-94, Feb. 2021. <https://doi.org/10.14346/JKOSOS.2021.36.1.86>.

저자소개

이영재 (Young Jae Lee)



1984년 2월: 충남대학교
전자교육공학과(공학사)
1994년 8월 : 연세대학교
전자공학과(공학석사)
2000년 8월 : 경희대학교
전자공학과(공학박사)
2002년 3월 ~ 현재 : 전주대학교

스마트미디어학과 교수

관심분야 : 컴퓨터비전, App. 개발, Game, AI, GUI