

텔레프레즌스 서비스를 위한 몰입도 분석 기술

Engagement Analysis Technology for Tele-presence Services

윤현진 [H.J. Yoon, hgyoon73@etri.re.kr]

한미경 [M.K. Han, mkhan@etri.re.kr]

장중현 [J.H. Jang, jangjh@etri.re.kr]

실감감성플랫폼연구실 선임연구원

실감감성플랫폼연구실 책임연구원/실장

5G 기가서비스연구부 책임연구원/부장

* 본고는 2017년도 정부(과학기술정보통신부)의 재원으로 ‘범부처 Giga KOREA 사업’의 지원을 받아 수행된 연구임[No. GK17P0100, Giga Media 기반 Tele-Experience 서비스 SW플랫폼 기술 개발].

A Telepresence service is an advanced video conferencing service at aimed providing remote users with the feeling of being present together at a particular location for a face-to-face group meeting. The effectiveness in this type of meeting can be further increased by automatically recognizing the audiovisual behaviors of the video conferencing users, accurately inferring their level of engagement from the recognized reactions, and providing proper feedback on their engagement state. In this paper, we review the recent developments of such engagement analysis techniques being utilized in various applications, such as human-robot interaction, content evaluation, telematics, and online collaboration services. In addition, we introduce a real-time engagement analysis framework employed in our telepresence service platform for an increased participation in online group collaboration settings.



본 저작물은 공공누리 제4유형
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

2017
Electronics and
Telecommunications
Trends

텔레프레즌스 & 홀로그래피
특집

- I. 머리말
- II. 몰입도 분석 기술 개요
- III. 몰입도 분석 기술 동향
- IV. 영상회의 사용자 몰입도
분석 기술
- V. 맺음말

I. 머리말

텔레프레즌스 서비스(Telepresence service)는 원격의 사용자들에게 마치 동일한 장소에 모여 함께 대면 회의를 하는 것과 같은 경험을 제공하는 것을 목표로 하는 실감형 영상회의의 서비스를 의미한다[1].

초기의 텔레프레즌스 서비스는 원격 사용자와 실제로 대면하고 있는 듯한 실재감을 높이기 위하여 대형 스크린과 고품질의 비디오 및 오디오를 지원하는 것을 주요 특징으로 하였다. 최근에는 자료 공유, 주화자 인식, 시선 맞춤, 몰입도 평가 기술 등을 도입하여 온라인 협업의 효율성을 극대화할 수 있는 몰입형 영상회의의 서비스 형태로 발전하고 있다.

몰입도 분석 기술은 서비스 사용자의 몰입 수준을 정확하게 판단하고 몰입 저하에 따라 서비스의 구성 및 흐름을 제어하여 몰입의 선순환 구조를 형성하는 기술로 정의할 수 있다. 따라서, 사용자의 몰입 상태를 정확하게 판단하는 몰입도 평가 기술뿐만 아니라 몰입도 평가 결과를 서비스에 반영하여 사용자의 몰입 경험을 증강시키는 몰입도 피드백 기술을 모두 포함하고 있다.

몰입도 평가 기술은 크게 설문지를 이용한 정성적 평가와 심리 생리학적 또는 시각적 반응을 측정된 데이터 분석을 이용한 데이터 주도형 평가로 구분할 수 있다. 전자의 설문지 기반 평가는 사용자의 몰입 상태를 실시간으로 판단하기 어렵고 사용자의 주관적 답변에 의존할 수밖에 없다는 단점을 갖는다. 후자의 데이터 주도형 몰입도 평가의 경우 데이터 수집과 평가 모델 학습에 추가 비용이 필요하지만, 눈 깜박임, 시선, 자세, 표정, 음성 등의 객관적인 몰입지표를 데이터로부터 자동 인식하여 몰입 수준을 실시간 평가할 수 있다는 장점 때문에 최근 디지털 콘텐츠, HRI(Human robot interaction), 텔레매틱스, 온라인 협업 등 다양한 서비스 분야에서 활용되고 있다.

텔레프레즌스 서비스 분야에서도 영상회의 사용자의

회의에 대한 집중도 및 그룹 상호작용에 대한 참여도를 평가하여 영상회의 서비스에 반영함으로써 온라인 협업의 효율성 개선에 기여할 수 있는 요소 기술로 관심을 받고 있다.

본고에서는 몰입도 분석 기술의 개요와 최근의 기술 동향을 살펴보고, 한국전자통신연구원에서 개발하고 있는 텔레프레즌스 서비스를 위한 영상회의 사용자 몰입도 분석 기술을 소개하고자 한다.

II. 몰입도 분석 기술 개요

몰입(Engagement)이란 어떤 활동이나 상황에 완전히 빠져들어 집중하고 있는 심리적 상태를 의미한다. 몰입은 중독 현상과 같은 부정적인 형태와 사용자에게 최적의 경험을 느끼게 해주는 긍정적인 형태로 구분되기 때문에, 사용자의 몰입 상태를 정확하게 판단하는 것이 긍정적인 몰입을 유도하는 시작이 된다[2].

몰입도 분석 기술은 서비스를 사용하는 사용자의 몰입 상태를 정확하게 판단하는 몰입도 평가 기술뿐만 아니라 몰입도 평가 결과를 서비스에 반영하여 사용자의 몰입 경험을 증강시키는 몰입도 피드백 기술까지 포함한다. 단순히 사용자의 몰입 수준을 판단하는 것에 그치지 않고 그 결과를 서비스 구성 및 흐름에 활용하여 몰입의 선순환 구조를 구축하는 것이 몰입도 분석의 주요 목적이 되기 때문이다.

1. 몰입도 평가 기술

몰입도 분석 기술 중 몰입도 평가 기술은 사용자의 서비스에 대한 집중도 및 서비스를 통한 상호작용에 대한 참여도를 판단하는 기술로 정의할 수 있다. 전통적으로 교육 및 조직의 성과관리 분야에서 시작하여 현재는 UI(User interface)/UX(User experience), HRI, 디지털 콘텐츠, 텔레매틱스, 의료, 온라인 협업 등 다양한 분야에서 연구 및 활용되고 있다.

초기 몰입도 평가 기술은 설문지를 통해 사후에 사용자의 몰입 경험을 조사하는 주관적 평가 방법을 사용하였다. 이러한 설문지 기반의 몰입도 평가는 사용자의 몰입 상태를 실시간으로 판단하기 어렵고 사용자의 주관적 답변에 의존할 수밖에 없다는 단점을 갖는다. 최근에는 사용자의 몰입 상태에 따른 눈 깜박임, 동공 크기, 자세, 표정, 발원 빈도 등 사용자의 다양한 시청각적 반응을 실시간 측정하고 분석하여 몰입 여부를 객관적으로 판단하는 데이터 주도형의 객관적 평가 방법이 시도되고 있다. 이러한 데이터 주도형의 몰입도 평가 기술은 몰입 수준을 객관적으로 판단할 수 있다는 장점을 갖지만, 몰입 수준을 판단하는 심리 생리학적 또는 시청각적 지표가 측정하는 데이터에 종속적이라는 제약을 갖기도 한다. 또한, 방대한 양의 데이터 수집과 객관적 몰입지표를 정확하게 인식할 수 있는 모델 학습이 선행되어야 하므로 이에 따른 추가 비용이 예상된다.

설문지 기반의 몰입도 평가는 평가 항목에 대한 사용자의 답변에 따라 일정 점수를 매기는 형태로 몰입도를 점수로 정량화할 수 있다. 반면, 데이터 주도형의 몰입도 평가 기술은 사용자의 몰입 수준을 몰입과 비몰입 단계로 구분하거나, 몰입을 여러 단계로 세분화하여 판단한다. 집중도를 비몰입 상태, 수동적 몰입 상태, 의도적 몰입 상태, 적극적 몰입 상태로 구분한 4단계 몰입 모델이 가장 일반적이며, 비몰입 상태부터 완전 몰입 상태까지 6단계로 세분화한 몰입 모델도 찾아볼 수 있다.

데이터 주도형 몰입도 평가의 경우 측정된 데이터로부터 관찰된 몰입지표를 기반으로 사용자의 몰입 상태를 도출하는 방법이 필요한데, 대부분 직관적 몰입도 판단 지식을 기반으로 하는 휴리스틱(Heuristic) 규칙을 이용하고 있다. 이러한 규칙 기반의 몰입도 판단은 일종의 경험적 판단에 근거를 둔 정성적인 평가 방법으로 앞으로 이를 보완하거나 대체할 수 있는 정량적 평가 기술에 대한 연구가 필요하다.

2. 몰입도 피드백 기술

몰입도 피드백 기술은 몰입도 평가 결과를 서비스에 반영하여 사용자의 몰입 경험을 높일 수 있는 기술을 의미한다. 몰입도 평가 결과를 사용자들에게 실시간 제공하거나 몰입도 개선이 필요한 시점에 서비스 구성을 동적으로 변경하거나 서비스 흐름을 제어하는 전략을 통해 사용자의 몰입을 직간접적으로 유도할 수 있다. 특히 사용자의 몰입도 평가 결과를 실시간으로 확인하거나 협업하는 다른 사용자들과 공유하는 것만으로도 몰입 증강에 기여할 수 있는 것으로 보고되고 있다.

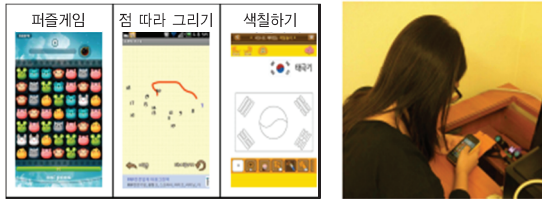
몰입도 피드백 기술은 몰입 수준에 따른 사용자 맞춤형 추천 기술, 보조 로봇 및 에이전트 제어 기술, 운전자 보조 기술과 결합하여 차별화된 서비스 경험을 제공하고 있다. 다만, 몰입도 피드백 기술이 사용자의 몰입도 증강에 얼마나 효과적으로 작용하는지에 대한 정량적인 연구는 부족한 상태로 앞으로 추가 연구가 필요한 분야이다.

III. 몰입도 분석 기술 동향

몰입도 분석 기술은 디지털 콘텐츠, HRI, 텔레매틱스, 온라인 협업 등 다양한 서비스 분야에서 사용자의 집중도, 참여도, 관심도를 평가하기 위해 활용되고 있다. 최근에는 사용자의 몰입 경험을 객관적으로 판단할 수 있는 데이터 주도형 몰입도 분석 기술을 위주로 연구 개발이 이루어지고 있다.

국내에서는 몰입수준이 다른 게임 콘텐츠를 사용할 때 사용자의 동공 크기, 눈 깜빡임 빈도, 눈 깜빡임 지속 시간에 차이가 있음을 확인하고, 사용자가 경험하는 몰입의 상태를 이러한 시각적 반응을 측정하여 정량적으로 평가하여, 몰입의 수준을 민감하게 반영할 수 있는 시각적 몰입지표를 탐색하는 연구가 시도되었다(그림 1) 참고].

구글은 코넬대학교와의 공동연구를 통해 사용자의 시



(그림 1) 게임 콘텐츠 사용자 몰입도 평가 연구[2]

[출처] 정주현 외, 대한인간공학회 학술대회논문집, 2012, pp. 347-351.



(그림 3) HRI 서비스를 위한 몰입도 분석 기술[4]

[출처] <http://dx.doi.org/10.1145/3025453.3025469>



(그림 2) 시선기반 콘텐츠 몰입도 평가 기술[3]

[출처] Y. Li et al., *Proc. Int. Conf. World Wide Web Companion*, 2017, pp. 525-533, CC BY, 2.0.

선을 추적하여 구글 플레이 스토어(Google play store)의 콘텐츠 아이템에 대한 사용자의 관심도 및 몰입도를 평가하는 기술을 개발하였다. 또한, 시선 기반 몰입도 평가 기술을 활용하여 제로 인터페이스 기반의 사용자 맞춤형 콘텐츠 추천 서비스를 제시하였다(그림 2) 참고.

몰입도 분석 기술이 가장 많이 연구되고 있는 분야 중 하나가 바로 인간로봇 상호작용(HRI) 분야이다. 인간과 로봇의 보다 자연스럽게 효율적인 상호작용을 지원하기 위해 로봇은 사용자의 의도를 정확하게 파악하고 사용자의 의도에 맞게 적절하게 반응하여야 한다.

(그림 3)과 같이 로봇과 상호작용하는 사용자의 시선, 머리 자세, 신체 자세 등 사용자의 행동을 로봇이 촬영한 2D 영상으로부터 인식하여 사용자가 로봇과 상호작용에 얼마나 몰입하고 있는지를 판단하였다. 또한, 판단 결과에 따라 로봇의 응답을 재구성하여 대화의 흐름을

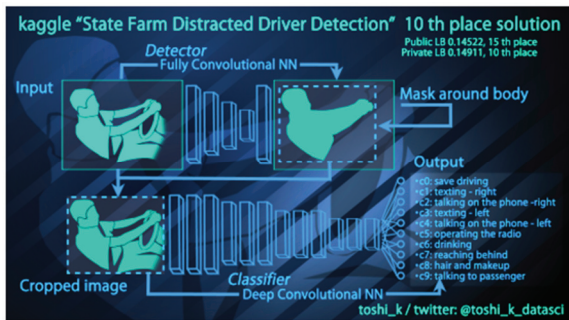
제어하는 서비스를 제공하고 있다.

로봇이 획득한 사용자 영상에서 사용자의 시선 방향, 자세, 발음 등의 몰입지표를 인식하여 사용자의 몰입 상태를 4단계 중 하나로 판단한다. 주변 장치 사용으로 인해 사용자의 몰입도가 높지 않은 상태에서는 대화형 로봇의 대화 속도를 낮추어 사용자의 컴퓨터 작업을 방해하지 않도록 하였다. 또한, 사용자의 몰입도가 완전히 저하된 상황에서는 대화를 중단하고 사용자가 다시 몰입 상태로 되돌아오면 대화를 재계하도록 하는 사용자 몰입 상태 기반 로봇제어 기술이 시도되었다.

텔레매틱스 서비스 분야에서도 자동차 실내에 설치한 카메라 영상에서 운전자의 휴대전화 사용 행위, 음료수를 마시는 행위 등 운전을 방해하는 9종류의 행위를 인식하여 운전자의 집중 여부를 판단하는 운전자 몰입도 평가 기술이 시도되었다[5], [6].

미국 보험회사인 스테이트팜은 2016년도 데이터 분석 플랫폼 캐글 사이트(<https://www.kaggle.com>)에 자동차 실내 대시보드 위에 설치한 카메라로 촬영한 운전자 영상으로부터 운전 몰입도가 낮은 상태의 운전자 영상을 판단하는 경진 문제를 게시하여 전 세계 데이터과학자들의 높은 관심을 모았다.

경진대회에 제출된 솔루션들 대부분 최근 영상분석 분야에서 우수한 성능을 보여주고 있는 딥러닝 모델을 판단에 활용하였다. 운전자 전체 영상뿐만 아니라 특정



(그림 4) 딥러닝 모델 기반 운전자 집중도 분류 기술

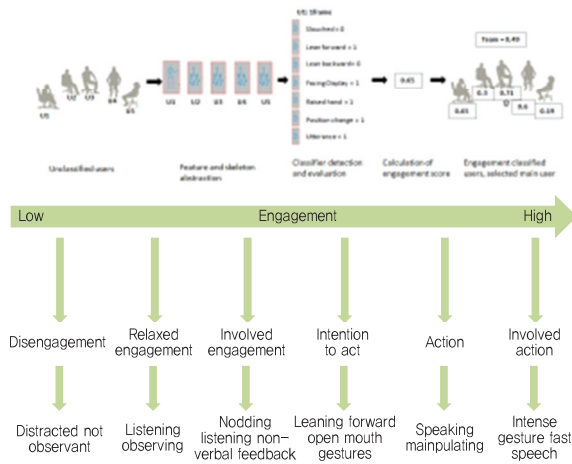
[출처] <https://github.com/toshi-k/kaggle-distracted-driver-detection>

영상 영역을 처리하는 다양한 CNN(Convolutional neural network) 모델을 이용하여 운전자의 몰입 여부나 구체적인 비몰입 유형을 판단하였다[그림 4] 참고]. 기존 영상기반 몰입도 평가 기술들이 영상에서 몰입도 판단을 위한 척도가 되는 몰입지표를 먼저 인식하고 인식한 몰입지표를 기반으로 몰입 상태를 도출한 반면, 딥러닝 모델 기반 몰입도 분석은 몰입지표 추출 단계 없이 영상을 바로 몰입 상태로 매핑한다는 차이가 있다.

텔레매틱스 분야에서는 이러한 운전자 몰입도 평가 결과를 기반으로 자동차 브레이크를 제어하여 사고를 예방하거나 라디오의 볼륨을 제어하여 운전자의 집중을 유도하는 몰입도 피드백 서비스도 제시되고 있다.

영상회의의 서비스 분야에서도 영상회의의 참석자들의 행위를 인식하여 몰입도를 판단하고 참석자들간 몰입도 정보를 공유하는 기술이 시도되었다[7], [8].

온라인 회의의 참여자 자세를 측정하는 3D 센서를 통하여 획득한 골격데이터와 마이크로부터 수집한 음성신호로부터 7종류의 몰입지표를 인식하고 참석자 개인의 몰입 수준을 6단계 중 하나로 판단하였다. 개인별 몰입도 평가 결과 중 가장 빈도가 높은 몰입 단계로 그룹의 몰입 수준을 결정하여 참석자들에게 제공하였다[그림 5] 참고]. 해당 실험 결과에서는 몰입도 평가 결과를 실시간 공유하는 것만으로도 온라인 회의의 효율을 개선하는데 기여한다고 하였다.



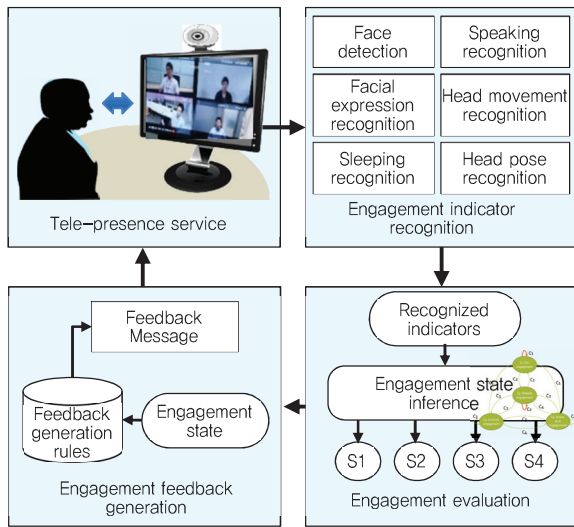
(그림 5) 회의 참석자 몰입도 분석 기술[7]

[출처] <https://arxiv.org/pdf/1608.08711.pdf>

데이터 주도형의 몰입도 분석 기술의 성능은 데이터로부터 얼마나 정확하게 몰입 지표를 자동 인식하느냐에 달려 있다. 기존 연구에서는 대부분 자체 학습한 기계학습 분류 모델을 사용하고 있으며, 일부 딥러닝 모델을 사용하여 인식 성능을 고도화하고 있다[6], [7]. 인식한 몰입지표를 기반으로 사용자의 몰입 상태를 판단하는 기술은 대부분 경험적 규칙을 기반으로 하고 있다. 이러한 규칙기반의 몰입도 판단은 일종의 정성적인 평가 방법으로 앞으로 이를 보완하거나 대체할 수 있는 정량적 평가 기술에 대한 연구가 필요하다.

IV. 영상회의의 사용자 몰입도 분석 기술

한국전자통신연구원은 고품질의 UHD(Ultra high definition) 대면 영상과 무안경 입체 영상인 초다시점 콘텐츠를 공유할 수 있는 실감형 텔레프레즌스 서비스 플랫폼 기술을 개발하고 있다. 특히 영상회의의 서비스에 대한 사용자의 몰입 경험 및 만족도를 극대화하기 위해 영상회의의 사용자의 행위를 인식하여 몰입 상태를 판단하고 몰입도 저하가 지속되는 상황을 인지하여 적절한 몰입도 피드백 메시지를 제공하는 몰입도 분석 기술을 개발하고 있다.



(그림 6) 영상회의의 사용자 몰입도 분석 기술 구성도

(그림 6)과 같이 영상회의의 사용자 몰입도 분석 기술은 사용자의 대면 영상을 수신하여 얼굴 영상으로부터 몰입 상태를 직간접적으로 판단할 수 있는 척도가 되는 표정, 머리 동작, 머리 자세, 발언 여부, 줄음 여부를 인식하는 몰입지표 인식 기술과 자동 인식한 몰입지표를 기반으로 몰입 상태를 판단하는 몰입도 평가 기술, 몰입 상태 판단 결과 몰입도가 지속적으로 낮은 상황을 인지하여 몰입도 피드백 메시지를 생성하고 제공하는 몰입도 피드백 생성 기술로 구성된다.

1. 몰입지표 인식 기술

몰입지표는 어떤 활동이나 상황에 집중하고 있는 심리적 상태를 직간접적으로 판단할 수 있는 심리 생리학적 또는 시청각적 척도를 의미한다. 영상회의의 사용자의 몰입 수준을 판단하기 위하여 사용자의 대면 얼굴 영상에서 인식할 수 있는 시각적 반응 중 표정 4종(기쁨, 불쾌, 놀람, 평상), 머리 동작 2종(끄덕임, 좌우 흔들), 발언 여부, 줄음 여부, 머리/시선 방향을 몰입지표로 선정하였다. 특히, 사용자의 얼굴 영상은 영상회의의 서비스를 위해 필수적으로 획득해야 하는 데이터로 이를 활용하여 추가적인 장비 없이도 몰입도 분석이 가능하도록 하

였다.

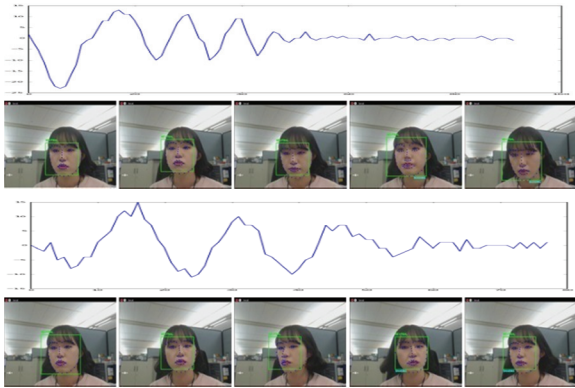
몰입지표 중 표정을 자동으로 인식하기 위하여, 수신한 사용자 얼굴 영상에서 얼굴 영역을 검출하고, 검출한 얼굴 영역에서 눈썹, 눈, 코, 입, 턱 등 얼굴 윤곽에 해당하는 총 68개의 얼굴 랜드마크(Facial landmarks) 위치를 검출하였다. 검출한 68개 점은 정면을 바라보는 표준 얼굴에서 검출한 표준 랜드마크 중 좌우 눈꼬리에 위치한 2개 점과 콧등 끝에 위치한 1개의 점을 기준으로 변환(Affine transform)하였다. 랜드마크 위치를 변환하는 이유는 표정 외에 얼굴의 방향, 회전, 크기에 따른 랜드마크 위치의 차이를 최소화하여 보다 정확한 표정 인식 결과를 얻기 위함이다. 변환된 68개의 얼굴 랜드마크의 위치 값을 4종류의 표정을 판단하는 표정 분류기에 입력하여 현재 프레임에서 검출한 얼굴에 나타난 표정을 기쁨(Happiness), 불쾌(Disgust), 놀람(Surprise), 평상(Neutral) 중 하나로 매핑하였다.

표정 분류기로 RBF(Radial basis function)-서포트벡터머신(SVM: Support vector machine) 모델을 사용하였고, 모델 파라미터는 오픈 데이터인 CK+¹⁾에서 선별한 이미지와 자체적으로 수집한 표정 이미지를 이용하여 학습하였다. 초파라미터(Hyper-parameter)는 교차 검증기반의 그리드 검색(Grid search)을 이용하여 선택하였다.

머리 동작 중 고개 끄덕임(Head nodding)과 좌우 흔들(Head shaking) 동작은 영상회의의 사용자가 현재 논의하고 있는 의제에 대해서 동의하거나 혹은 거부할 때 자연스럽게 나타나는 행동으로 몰입 수준을 가늠하는 주요 척도가 될 수 있다.

머리 동작을 인식하기 위해 표정 인식 과정에서 사용한 68개의 얼굴 랜드마크 중 콧등 끝에 위치한 점의 위치를 일련의 연속된 프레임에서 추적하여 은닉마르코프 모델(HMM: Hidden markov model)기반의 머리 동작

1) http://www.consortium_ri_cmu.edu/ckagree/

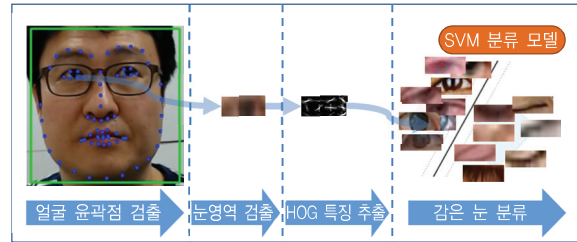


(그림 7) 머리동작 데이터 및 인식 결과

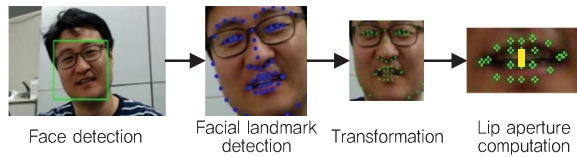
인식기의 입력으로 사용하였다. 68개의 얼굴 랜드마크 중 콧등 끝점은 머리를 상하 또는 좌우로 끄덕일 때 위치 변화가 가장 큰 점이기 때문에 콧등 끝점의 y 변위량과 x 변위량 시퀀스를 고개 끄덕임 인식기와 좌우 흔들 인식기에 각각 입력하여 머리동작을 인식하였다(그림 7) 참조]. 은닉마르코프모델은 3개의 은닉상태와 가우시안분포(Gaussian distribution)를 갖는 출력변수로 구성하였으며, 모델 파라미터는 자체 수집한 머리 동작 영상을 이용하여 학습하였다.

영상회의 사용자가 회의 중 즐기고 있다는 것은 회의에 전혀 몰입하고 있는 못한 상태로 볼 수 있기 때문에 졸음 여부를 몰입지표로 선정하였다. 졸음 인식은 표정 인식 과정에서 검출한 68개의 얼굴 랜드마크를 이용하여 눈 영역의 이미지 패치를 획득하고, 눈 영역에서 HOG(Histogram of oriented gradient) 특징 벡터를 추출하여 이를 분류기에 입력하여 눈감음(Closed eye) 여부를 판단하였다(그림 8) 참조].

눈감음 분류기로 SVM 모델을 사용하였고, 모델 파라미터는 자체 수집한 감은 눈과 뜬 눈 이미지를 이용하여 학습하였다. 최종 졸음 여부는 감은 눈이 인식된 영상 프레임의 누적 수를 미리 설정한 임계 값과 비교하여 판단하였다. 즉 최근 100개의 영상 프레임에서 85개 이상의 프레임에서 감은 눈이 인식되면 사용자가 즐기고 있다고 판단하였다.



(그림 8) 졸음 인식을 위한 감은 눈 분류 과정



(그림 9) 발언 여부 인식 과정

영상회의에서 사용자가 발언한다는 것은 적극적으로 회의에 참여하고 있는 것으로 볼 수 있기 때문에 발언 여부를 몰입지표 중 하나로 선정하였다. 발언(Speaking) 인식은 얼굴 영상에서 입술 구경(Lip aperture)을 추적하여 판단하였다. 입술 구경은 상하 입술의 중간에 위치한 2개의 랜드마크 사이의 거리로 정의하였다. 입술 구경 값이 미리 설정한 임계 값 보다 크면 현재 프레임에서 사용자가 발언한 것으로 판단하였다(그림 9) 참조].

졸임 인식에서와같이 최근 복수의 연속된 영상 프레임에서 발언한 것으로 판단된 프레임의 수를 미리 설정한 기준값과 비교하여 발언 여부를 최종 판단하였다.

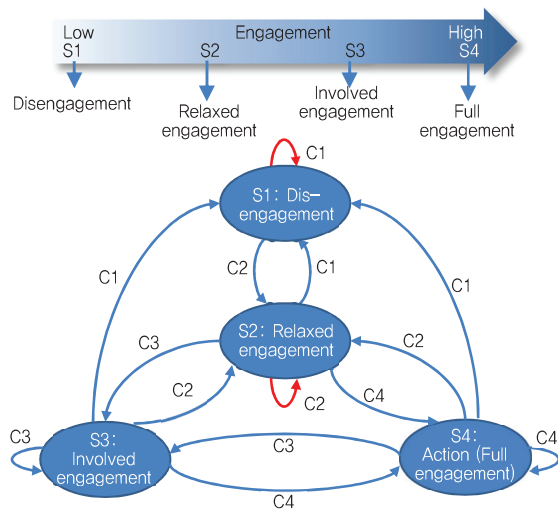
마지막으로 사용자가 영상회의의 대면화면이 제공되는 스크린을 정면으로 바라보지 않는다는 것은 회의에 몰입하지 않는 것으로 판단할 수 있기 때문에 사용자의 시선을 몰입지표로 선정하였다. 순수 RGB 영상에서 사용자의 시선을 정확하게 검출하기 어렵기 때문에 사용자의 얼굴 방향(pitch, roll, yaw)을 먼저 검출하고 검출한 얼굴 방향 값으로부터 시선의 방향을 간접적으로 추정하였다. 시선 인식 과정은 앞서 표정 인식을 위해 검출한 68개의 얼굴 랜드마크 위치와 대응하는 표준 얼굴의 랜드마크 위치를 이용하여 얼굴의 방향(pitch, roll, yaw)을 추정하고, 얼굴의 방향이 미리 설정한 임계 범

위 내에 있으면 사용자가 정면을 바라보고 있는 상태(Gaze direct)로 판단하고 임계 범위를 벗어나면 사용자의 시선이 정면을 벗어난 상태(Gaze away)로 최종 판단하였다.

2. 몰입도 평가 기술

영상회의 사용자의 몰입 상태를 판단하기 위해 몰입의 수준을 비몰입, 수동형 몰입, 참여형 몰입, 행동형 몰입의 4단계로 구분하였다(그림 10) 참조.

첫 번째 단계인 비몰입(Disengagement) 상태는 사용자가 웹캠 앞의 자리를 이탈하였거나, 졸고 있거나 사용자가 제공되는 정면의 대면 영상을 응시하고 있지 않은 상태로 이는 사용자가 회의에 전혀 몰입하고 있지 않다는 것을 의미한다. 두 번째 몰입 단계인 수동형 몰입(Passive/relaxed engagement) 상태는 사용자가 대면 영상이 제공되는 화면을 평상의 표정, 즉 무표정으로 응시하고 있는 상태로 주로 상대방의 말을 경청하고 있는 수동적인 몰입 상태를 의미한다. 세 번째 몰입 단계는 참여형 몰입(Involved engagement) 상태로 사용자가 정면을 응시하며 웃는 표정을 짓거나 고개를 끄덕이거나 짧은 발언을 통해 동의 또는 긍정의 반응을 보이며 상호 작용에 참여하고 있는 상태로 정의하였다. 마지막으로



(그림 10) 몰입 상태 및 상태간 전이 관계

〈표 1〉 몰입 상태 별 몰입지표 조건

Transition	Engagement indicators
C1	- Absence of face ∨ gaze away ∨ sleeping
C2	- Direct gaze ∧ neutral expression
C3	- Direct gaze ∧ (nodding ∨ smile exp. ∨ short speaking)
C4	- Direct gaze ∧ (shaking ∨ long speaking ∨ disgust exp. ∨ surprise exp.)

몰입의 가장 높은 단계인 행동형 몰입(Full engagement/action) 상태는 불쾌한 표정이나 놀란 표정을 짓거나 머리를 좌우로 흔드는 행동을 통해 부정의 의사를 강하게 표시하거나 발언을 주도하며 회의에 적극적으로 참여하고 있는 상태로 정의하였다. 마지막으로 4단계 몰입 상태 간의 전이를 기존 몰입도 평가 연구 결과 [4]와 [7]을 참고하여 (그림 10)과 같이 정의하였다.

사용자의 몰입 상태는 사용자 얼굴 영상에서 인식한 몰입지표로 판단하는데 〈표 1〉에 정의한 몰입지표 조건을 기반으로 몰입 상태의 전이를 판단하였다.

만약 사용자의 이전 몰입 상태가 비몰입 상태 S1으로 판단된 경우, 현재 얼굴 영상에서 정면 시선 또는 평상 표정이 몰입지표로 인식되었다면 사용자의 몰입 상태가 비몰입 상태 S1에서 수동형 몰입 상태 S2로 전이한 것으로 판단할 수 있다. 만약 사용자 얼굴 영상에서 시선 이탈, 졸음, 얼굴부재의 몰입지표가 지속적으로 관찰되면 사용자는 계속해서 비몰입 상태 S1에 머물러 있는 것으로 판단하게 된다.

사용자의 몰입 상태가 참여형 몰입 상태 S3으로 전이하기 위해서는 정면 시선, 고개 끄덕임, 웃는 표정, 짧은 발언과 같은 몰입지표가 관찰되어야 하며, 행동형 몰입 상태 S4로 전이하기 위해서는 정면 시선과 고개 좌우 흔들, 발언주도, 불쾌 또는 놀란 표정과 같은 몰입지표가 인식되어야 한다.

3. 몰입도 피드백 기술

몰입도 피드백 기술은 사용자의 몰입 상태가 낮게 평

가된 경우 회의 집중을 유도하는 알림 메시지를 제공하는 기능을 담당한다. 몰입도 피드백을 제공하는 상황은 사용자의 몰입 상태가 비몰입 상태로 계속해서 판단되는 지속적 비몰입 경우와 수동형 몰입 상태에 오랜 시간 머물러 있는 지속적 수동형 몰입 경우로 정의하였다. 지속적 비몰입 상태의 사용자에게는 '회의에 집중해 주세요.'라는 알림 메시지를 생성하여 제공하였고, 지속적 수동형 몰입 상태의 사용자에게는 '회의에 적극적으로 참여해 보세요.'라는 알림 메시지를 제공하였다. 또한, 지속적 비몰입 상황 판단에 사용된 몰입지표에 따라 얼굴 부재가 주로 인식된 경우 '자리를 이탈한 것 같습니다.', 졸음이 주로 인식된 경우 '일어나세요.'의 메시지를 추가로 제공하였다[[표 2](#) 참조].

생성한 몰입도 피드백 메시지는 (그림 11)과 같이 영상회의 플랫폼의 메시지 전달 기능을 통해 영상회의 클라이언트 화면으로 전송되어 사용자에게 제공된다. 현재 몰입도 피드백 메시지는 해당 사용자에게만 제공되고 있으나, 차후 영상회의에 참여하고 있는 모든 사용자

가 몰입도 정보를 공유할 수 있는 기능을 추가로 개발할 예정이다.

V. 맺음말

본고에서는 서비스 효율의 극대화를 위해 활용되고 있는 몰입도 분석 기술의 개요와 최근의 기술 동향을 살펴보고 있다. 몰입도 분석 기술은 서비스를 사용하는 사용자의 몰입 상태를 정확하게 판단하는 몰입도 평가 기술 뿐만 아니라 몰입도 평가 결과를 서비스 구성 및 흐름에 반영하여 사용자의 몰입 경험을 증강시키는 몰입도 피드백 기술까지 포함하는 것을 특징으로 한다.

몰입도 분석 기술 중 사용자의 심리 생리학적 또는 시청각적 반응을 실시간 측정하고 분석하여 사용자의 몰입 수준을 객관적으로 판단하는 데이터 주도형의 몰입도 평가 기술이 현재 디지털 콘텐츠, HRI, 텔레메틱스, 온라인 협업 등 다양한 서비스 분야에서 활용되고 있다.

한국전자통신연구원에서도 차별화된 텔레프레즌스 서비스를 위하여 영상회의 사용자의 얼굴 영상으로부터 표정, 머리 동작, 발음 여부, 졸음 여부, 머리/시선 방향 등 총 12개의 몰입지표를 자동 인식하고, 실시간 몰입 지표 인식 결과를 기반으로 사용자의 몰입 상태를 4단계 중 하나로 판단하는 몰입도 평가기술을 개발하였다. 또한, 몰입도 평가 결과 몰입도가 계속해서 저하되는 상황을 지속해서 모니터링하여 몰입도 피드백 메시지를 사용자에게 제공하는 몰입도 피드백 기술도 개발하였다.

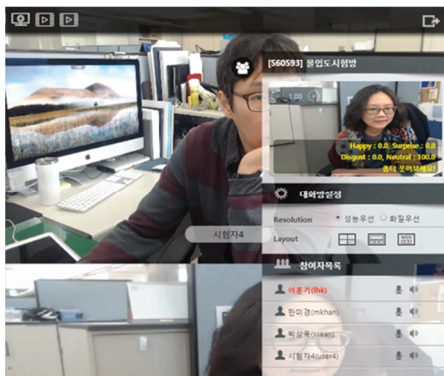
향후, 몰입도 분석 기술의 고도화를 위하여 몰입지표 인식 모델의 정확도를 개선하는 연구와 경험적 규칙을 기반으로 사용자의 몰입 상태를 도출하는 현재의 규칙 기반 정성적 몰입도 평가 방법을 보완하거나 대체할 수 있는 정량적 몰입도 평가 기술에 대한 연구가 필요하다.

약어 정리

CNN Convolutional Neural Networks

〈표 2〉 몰입도 피드백 메시지 유형

몰입 상태	몰입도 피드백 메시지	
지속적 비몰입 상태	회의에 집중해 주세요.	
	Absence of face	자리를 이탈한 것 같습니다.
	Sleeping	일어나세요.
지속적 수동형 몰입 상태	회의에 적극적으로 참여해 주세요.	



(그림 11) 몰입도 피드백 메시지 제공 화면

HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients
HRI	Human Robot Interaction
RBF	Radial Basis Function
SVM	Support Vector Machine
UHD	Ultra High Definition
UI	User Interface
UX	User eXperience

참고문헌

- [1] 이미숙 외, “Tele-experience 실감 스마트워크 서비스를 위한 텔레프레즌스 기술,” 한국통신학회지: 정보와 통신, 제31권 제3호, 2014, pp. 3-11.
- [2] 정주현 외, “시각적 반응을 이용한 사용자 몰입 경험 (UX) 평가,” 대한인간공학회 학술대회논문집, 2012, pp. 347-351.
- [3] Y. Li et al., “Towards Measuring and Inferring User Interest from Gaze,” In *Proc. Int. Conf. World Wide Web Companion*, Perth, Australia, Apr. 3-7, 2017, pp. 525-533.
- [4] M. Sun, Z. Zhao, and X. Ma, “Sensing and Handling Engagement Dynamics in Human-Robot Interaction Involving Peripheral Computing Devices,” In *Proc. Conf. Human Factors Comput. Syst.*, Denver, CO, USA, May 6-11, 2017, pp. 556-567.
- [5] M. Kuttila et al., “Driver Distraction Detection with a Camera Vision System,” In *Proc. IEEE Int. Conf. Image Process.*, San Antonio, TX, USA, Sept. 16-19, 2007, pp. VI-201-VI-204.
- [6] Y. Abouelnaga et al., “Real-Time Distracted Driver Posture Classification,” arXiv preprint, arXiv:1706.09498, 2017.
- [7] M. Frank et al., “Engagement Detection in Meetings,” arXiv preprint, arXiv:1608.08711, 2016.
- [8] G. Tofighi, H. Gu, and K. Raahemifar, “Vision-Based Engagement Detection in Virtual Reality,” *Digital Media Ind. Academic Forum*, Santorini, Greece, July 4-6, 2016, pp. 202-206.