



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2020년06월08일  
(11) 등록번호 10-2110478  
(24) 등록일자 2020년05월07일

(51) 국제특허분류(Int. Cl.)  
G06F 3/01 (2006.01) G06K 9/46 (2006.01)  
G06N 3/08 (2006.01)  
(52) CPC특허분류  
G06F 3/015 (2013.01)  
G06K 9/00885 (2013.01)  
(21) 출원번호 10-2019-0150558  
(22) 출원일자 2019년11월21일  
심사청구일자 2019년11월21일  
(56) 선행기술조사문헌  
KR101728635 B1  
(뒷면에 계속)

(73) 특허권자  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)  
(72) 발명자  
김시호  
인천광역시 연수구 송도과학로 85 Faculty House  
D동 912호  
차재광  
경기도 용인시 수지구 용구대로2753번길 27, 113  
동 401호(죽전동, 한솔 노블빌리지)  
(74) 대리인  
오위환, 나성곤, 정기택

전체 청구항 수 : 총 6 항

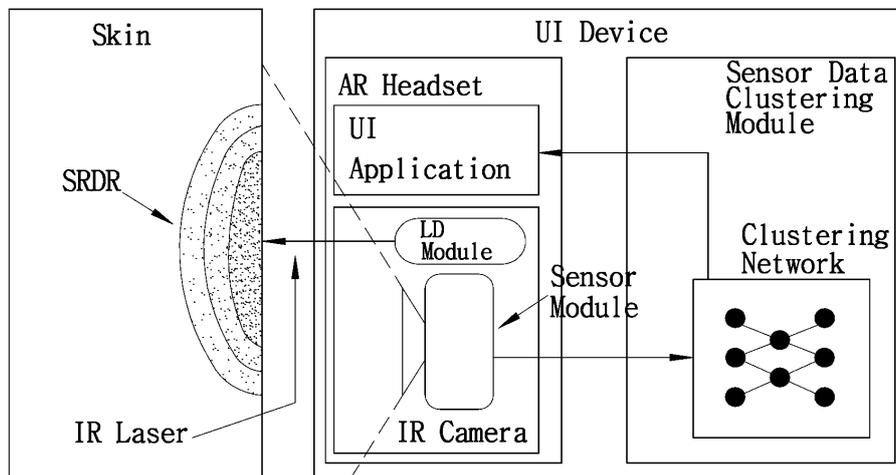
심사관 : 임지환

(54) 발명의 명칭 헤드셋 형태의 사용자 인터페이스 시스템 및 이를 위한 사용자 특징정보 분류 방법

(57) 요약

본 발명은 일정한 거리를 두고 근적외선을 피부에 방출하는 근적외선 방출부, 상기 방출된 근적외선에 의해 상기 피부로 입사되는 근적외선 중 상기 피부 내에서 산란되어 상기 피부 밖으로 방출되는 근적외선의 신호가 포함된 적어도 하나 이상의 신호 프레임 수신하는 신호 프레임 수신부, 상기 적어도 하나 이상의 신호 프레임을 수신된 순서로 전처리하는 신호 프레임 전처리부, 상기 전처리된 적어도 하나 이상의 신호 프레임으로부터 상기 방출된 근적외선에 의해 변형된 상기 피부의 특징정보를 추출하는 특징정보 추출부, 상기 추출된 특징정보를 기 설정된 사용자의 의도에 분류하는 특징정보 분류부를 포함하여 AR/VR용 핸드프리 NUI(Natural User Interface) 기술을 구현할 수 있는 효과를 가진다.

대표도 - 도3



(52) CPC특허분류

G06K 9/46 (2013.01)  
 G06N 3/08 (2013.01)  
 G06F 2203/011 (2013.01)

(56) 선행기술조사문헌

KR101759444 B1\*  
 M.Taini et. al., "Facial Expression Recognition from Near-Infrared Video Sequences"  
 공성곤, "얼굴영상과 예측한 열 적외선 텍스처의 융합에 의한 얼굴 인식," JKIIIS, 페이지 437-443, 2015.

J.Cha et. al., "Noninvasive determination of fiber orientation and tracking 2-dimensional deformation of human skin utilizing spatially resolved reflectance of infrared light measurement in vivo," in\*

S.Nickell et. al., "Anisotropy of light propagation in human skin," In the Jour. of Phys. Med. & Biol., vol. 45, no. 10, pp.2873-2886, 2000.\*

\*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호	1711081194
부처명	과학기술정보통신부
연구관리전문기관	정보통신기획평가원
연구사업명	디지털콘텐츠원천기술개발(R&D)
연구과제명	HMD 표정 인식 센서와 사이버 인터랙션 인터페이스 기술
기 여 율	1/1
주관기관	연세대학교 산학협력단
연구기간	2019.01.01 ~ 2019.12.31

공지예외적용 : 있음

**명세서**

**청구범위**

**청구항 1**

헤드셋 형태의 사용자 인터페이스 시스템에 있어서,

일정한 거리를 두고 근적외선을 피부에 방출하는 근적외선 방출부;

상기 방출된 근적외선에 의해 상기 피부로 입사되는 근적외선 중 상기 피부 내에서 산란되어 상기 피부 밖으로 방출되는 근적외선의 신호가 포함된 적어도 하나 이상의 신호 프레임 수신하는 신호 프레임 수신부;

상기 적어도 하나 이상의 신호 프레임을 수신된 순서로 전처리하는 신호 프레임 전처리부;

상기 전처리된 적어도 하나 이상의 신호 프레임으로부터 상기 방출된 근적외선에 의해 변형된 상기 피부의 특징 정보를 추출하는 특징정보 추출부; 및

상기 추출된 특징정보를 기 설정된 사용자의 의도에 분류하는 특징정보 분류부;를 포함하고,

상기 신호 프레임 수신부에서 상기 수신하는 신호 프레임은 상기 피부에 방출된 근적외선에 의한 산란현상으로부터 측정된 공간 분해 확산 반사율(SRDR: Spatially Resolved Diffuse Reflectance)이 포함된 근적외선 이미지 형태이며,

상기 신호 프레임 전처리부는 상기 근적외선 이미지를 수신 순서대로 전처리하고, 상기 특징정보 추출부의 데이터 입력 형태에 따라 데이터 형태를 재구성하며,

상기 특징정보 추출부는 상기 전처리된 근적외선 이미지로부터 상기 사용자 의도에 분류하기 위한 적어도 하나 이상의 특징을 추출하고,

상기 특징정보 분류부는 상기 추출된 적어도 하나 이상의 특징을 상기 사용자 의도에 분류하여 매핑하며,

상기 특징정보 추출부는 심층 신경망(DNN: Deep Neural Network)을 기반으로 사전 학습된 시공간 오토인코더(STAE: Spatiotemporal Autoencoder)의 인코더를 이용하여 상기 특징정보를 추출하고,

상기 특징정보 분류부는 DEC(Deep Embedded Clustering) 기법을 기반으로 사전 학습된 분류기를 이용하여 상기 추출된 적어도 하나 이상의 특징을 분류하여 기 설정된 사용자 의도에 매핑하는 헤드셋 형태의 사용자 인터페이스 시스템.

**청구항 2**

제 1 항에 있어서, 상기 근적외선 방출부는,

상기 피부에 방출된 근적외선에 의한 산란현상에 따른 신호 프레임의 수신에 상기 피부에 방출된 근적외선에 의한 반사현상의 영향을 최소화하도록 시준된(Collimated) 근적외선을 방출하는 헤드셋 형태의 사용자 인터페이스 시스템.

**청구항 3**

삭제

**청구항 4**

제 1 항에 있어서, 상기 신호 프레임 수신부는,

상기 근적외선 이미지에 목표 대역의 근적외선 신호 외에 다른 대역의 노이즈 신호가 포함되지 않도록 근적외선 필터를 포함하는 헤드셋 형태의 사용자 인터페이스 시스템.

**청구항 5**

삭제

**청구항 6**

삭제

**청구항 7**

삭제

**청구항 8**

사용자 인터페이스를 위한 사용자 특징정보 분류 방법에 있어서,

근적외선 방출부에서 일정한 거리를 두고 근적외선을 피부에 방출하는 근적외선 방출단계;

신호 프레임 수신부에서 상기 근적외선 방출단계로부터 방출된 근적외선에 의해 상기 피부로 입사되는 근적외선 중 상기 피부 내에서 산란되어 상기 피부 밖으로 방출되는 근적외선의 신호가 포함된 적어도 하나 이상의 신호 프레임을 수신하는 신호 프레임 수신단계;

신호 프레임 전처리부에서 상기 신호 프레임 수신단계로부터 수신된 적어도 하나 이상의 신호 프레임을 수신된 순서로 전처리하는 신호 프레임 전처리단계;

특징정보 추출부에서 상기 신호 프레임 전처리단계로부터 전처리된 적어도 하나 이상의 신호 프레임으로부터 상기 방출된 근적외선에 의해 변형된 상기 피부의 특징정보를 추출하는 특징정보 추출단계; 및

특징정보 분류부에서 상기 특징정보 추출단계로부터 추출된 특징정보를 기 설정된 사용자의 의도에 분류하는 특징정보 분류단계;를 포함하고,

상기 신호 프레임 수신단계에서 상기 수신하는 신호 프레임은 상기 피부에 방출된 근적외선에 의한 산란현상으로부터 측정된 공간 분해 확산 반사율(SRDR: Spatially Resolved Diffuse Reflectance)이 포함된 근적외선 이미지 형태이며,

상기 신호 프레임 전처리단계는 상기 근적외선 이미지를 수신 순서대로 전처리하고, 상기 특징정보 추출단계의 데이터 입력 형태에 따라 데이터 형태를 재구성하며,

상기 특징정보 추출단계는 상기 신호 프레임 전처리단계로부터 전처리된 근적외선 이미지로부터 상기 사용자 의도에 분류하기 위한 적어도 하나 이상의 특징을 추출하고,

상기 특징정보 분류단계는 상기 특징정보 추출단계로부터 추출된 적어도 하나 이상의 특징을 상기 사용자 의도에 분류하여 매핑하며,

상기 특징정보 추출단계는 심층 신경망(DNN: Deep Neural Network)을 기반으로 사전 학습된 시공간 오토인코더(STAE: Spatiotemporal Autoencoder)의 인코더를 이용하여 상기 특징정보를 추출하고,

상기 특징정보 분류단계는 DEC(Deep Embedded Clustering) 기법을 기반으로 사전 학습된 분류기를 이용하여 상기 추출된 적어도 하나 이상의 특징을 분류하여 기 설정된 사용자 의도에 매핑하는 사용자 인터페이스를 위한 사용자 특징정보 분류 방법.

**청구항 9**

제 8 항에 있어서, 상기 근적외선 방출단계는,

상기 피부에 방출된 근적외선에 의한 산란현상에 따른 신호 프레임의 수신에 상기 피부에 방출된 근적외선에 의한 반사현상의 영향을 최소화하도록 시준된(Collimated) 근적외선을 방출하는 사용자 인터페이스를 위한 사용자 특징정보 분류 방법.

**청구항 10**

삭제

**청구항 11**

제 8 항에 있어서, 상기 신호 프레임 수신부는,

상기 근적외선 이미지에 목표 대역의 근적외선 신호 외에 다른 대역의 노이즈 신호가 포함되지 않도록 근적외선 필터를 포함하는 사용자 인터페이스를 위한 사용자 특징정보 분류 방법.

**청구항 12**

삭제

**청구항 13**

삭제

**청구항 14**

삭제

**발명의 설명**

**기술 분야**

[0001] 본 발명은 사용자 인터페이스에 관한 것으로, 더욱 상세하게는 근적외선에 의한 산란현상에 따른 공간 분해 확산 반사율(SRDR: Spatially Resolved Diffuse Reflectance)을 통해 사용자의 의도를 파악함으로써 인터페이스 기술을 구현하는 헤드셋 형태의 사용자 인터페이스 시스템 및 이를 위한 사용자 특징정보 분류 방법에 관한 것이다.

**배경 기술**

[0002] 증강현실(AR)은 정보통신기술(ICT) 업계에서 가장 뜨거운 이슈 중 하나가 됐다. AR 분야에서는 AR 헤드셋에 최적화된 UI(사용자 인터페이스) 기기를 개발하는 것이 주요 과제 중 하나이다. 개인용 컴퓨터나 휴대전화와 크게 다른 헤드셋 환경 때문에 기존의 UI(키보드, 마우스, 터치스크린 등)를 사용하기 어렵기 때문이다. 헤드셋 측면의 버튼을 누르거나 리모컨을 잡고 있는 것이 현재 헤드셋 UI에 채택된 가장 대표적인 해결책이다.

[0003] 다만 이러한 UI는 수술이나 수작업 등 복잡하고 위험한 수술을 지원하는 데 한계가 있을 뿐만 아니라 일반 사용에도 충분히 편리하지 않다. 음성인식 기법 및 EOG(Electrooculography), EEG(Electrencephalography), EMG(Electromyography) 등의 생리학적 센서는 AR 핸즈프리 UI 개발에 유망한 기술들이지만 이를 이용한 최적의 헤드셋 UI 구현 방법은 아직 개발되지 않았다.

[0004] 학계나 산업계에서는 이미 헤드셋 타입으로 사용자 인터랙션을 가능하게 하는 기술을 개발한 바 있다. 일반적인 인터페이스 기술은 시선 인식, 손 동작 인식, EMG 센서, 터치패드, 리모콘 등을 이용하는 등 다양한 방식이 존재하는데, 그 중 손 제스처 감지 또는 시선 추적 기술이 가장 많이 사용되고 있다.

[0005] 다만 대부분의 상용 AR 헤드셋의 인터랙션 기술은 여전히 효과적인 UI에 필요한 충분한 기술이 개발되지 않아, 헤드셋의 주변 장치에 해당하는 핸드-헤드 컨트롤러(Hand-held controller) 또는 버튼 기술에 의존하고 있다.

[0006] 이러한 기술은 수술, 공장 현장 등과 같이 사용자가 직접 손을 사용해야만 하는 특수한 상황에 적용되기에는 한계점이 존재한다.

[0007] 이에 UI 기술에 있어서, 핸즈프리 방식으로 사용자의 의도를 읽어내고, 이에 따른 명령을 수행할 수 있는 기술을 개발할 필요성이 있다.

**선행기술문헌**

**특허문헌**

[0008] (특허문헌 0001) 한국공개특허공보 제2019-0056833호

**발명의 내용**

**해결하려는 과제**

- [0009] 이에 본 발명은 상기와 같은 제반 사항을 고려하여 제안된 것으로, 얼굴 제스처의 피부 변형을 감지한 후, 이를 UI 입력 명령으로 이용함으로써 AR/VR용 핸드프리 NUI(Natural User Interface) 기술을 구현하는 것을 목적으로 한다.
- [0010] 또한, 본 발명은 얼굴 피부의 움직임을 비침습적으로 쉽게 감지하고, 감지된 얼굴 동작을 입력 명령에 매핑하여 헤드셋 사용자와 AR 시스템 간의 핸드프리 상호작용을 지원하는 것을 목적으로 한다.
- [0011] 또한, 본 발명은 헤드셋 형태의 사용자 인터페이스 기술을 구현함으로써, 특수한 상황 또는 환경에서도 적용이 가능하도록 하는 것을 목적으로 한다.
- [0012] 또한, 본 발명은 피부 변형에 관한 데이터를 2차원 형태로 수신함으로써, 양적 및 질적으로 향상된 데이터를 수신할 수 있도록 하는 것을 목적으로 한다.
- [0013] 또한, 본 발명은 연속적으로 수신된 데이터의 변화량 산출을 통해 사용자 동작 변화에 관한 데이터를 생성함으로써, 데이터 분류가 용이하도록 하는 것을 목적으로 한다.
- [0014] 또한, 본 발명은 특징 추출 및 분류를 비지도 학습방법으로 구현함으로써, 태깅 정보 없이도 학습용 데이터셋을 분류할 수 있도록 하는 것으로 목적으로 한다.
- [0015] 또한, 본 발명은 DEC(Deep Embedded Clustering) 기법의 적용을 통해, 비지도 학습방법의 한계점인 정확도 문제점을 개선하는 것을 목적으로 한다.
- [0016] 본 발명의 목적들은 이상에서 언급한 목적들로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 당업자에게 명확하게 이해할 수 있을 것이다.

**과제의 해결 수단**

- [0017] 상기와 같은 목적을 달성하기 위하여 본 발명의 기술적 사상에 의한 헤드셋 형태의 사용자 인터페이스 시스템은 일정한 거리를 두고 근적외선을 피부에 방출하는 근적외선 방출부, 상기 방출된 근적외선에 의해 상기 피부로 입사되는 근적외선 중 상기 피부 내에서 산란되어 상기 피부 밖으로 방출되는 근적외선의 신호가 포함된 적어도 하나 이상의 신호 프레임 수신하는 신호 프레임 수신부, 상기 적어도 하나 이상의 신호 프레임을 수신된 순서로 전처리하는 신호 프레임 전처리부, 상기 전처리된 적어도 하나 이상의 신호 프레임으로부터 상기 방출된 근적외선에 의해 변형된 상기 피부의 특징정보를 추출하는 특징정보 추출부, 상기 추출된 특징정보를 기 설정된 사용자의 의도에 분류하는 특징정보 분류부를 포함할 수 있다.
- [0018] 상기 근적외선 방출부는 상기 피부에 방출된 근적외선에 의한 산란현상에 따른 신호 프레임의 수신에 상기 피부에 방출된 근적외선에 의한 반사현상의 영향을 최소화하도록 시준된(Collimated) 근적외선을 방출할 수 있다.
- [0019] 상기 신호 프레임 수신부에서 상기 수신하는 신호 프레임은 상기 피부에 방출된 근적외선에 의한 산란현상으로부터 측정된 공간 분해 확산 반사율(SRDR: Spatially Resolved Diffuse Reflectance)이 포함된 근적외선 이미지 형태라 할 수 있다.
- [0020] 상기 신호 프레임 수신부는 상기 근적외선 이미지에 목표 대역의 근적외선 신호 외에 다른 대역의 노이즈 신호가 포함되지 않도록 근적외선 필터를 포함할 수 있다.
- [0021] 상기 신호 프레임 전처리부는 상기 근적외선 이미지를 수신 순서대로 전처리하고, 상기 특징 추출부의 데이터 입력 형태에 따라 데이터 형태를 재구성할 수 있다.
- [0022] 상기 특징정보 추출부는 상기 전처리된 근적외선 이미지로부터 상기 사용자 의도에 분류하기 위한 적어도 하나 이상의 특징을 추출하고, 상기 특징정보 분류부는 상기 추출된 적어도 하나 이상의 특징을 상기 사용자 의도에 분류하여 매핑할 수 있다.
- [0023] 상기 특징정보 추출부는 심층 신경망(DNN: Deep Neural Network)을 기반으로 사전 학습된 시공간 오토인코더(STAE: Spatiotemporal Autoencoder)의 인코더를 이용하여 상기 특징정보를 추출하고, 상기 특징정보 분류부는 DEC(Deep Embedded Clustering) 기법을 기반으로 사전 학습된 분류기를 이용하여 상기 추출된 적어도 하나 이상의 특징을 분류하여 기 설정된 사용자 의도에 매핑할 수 있다.
- [0024] 상기와 같은 목적을 달성하기 위하여 본 발명의 기술적 사상에 의한 사용자 인터페이스를 위한 사용자 특징정보 분류 방법은 근적외선 방출부에서 일정한 거리를 두고 근적외선을 피부에 방출하는 근적외선 방출단계, 신호 프레임 수신부에서 상기 방출된 근적외선에 의해 상기 피부로 입사되는 근적외선 중 상기 피부 내에서 산란되어

상기 피부 밖으로 방출되는 근적외선의 신호가 포함된 적어도 하나 이상의 신호 프레임을 수신하는 신호 프레임 수신단계, 신호 프레임 전처리부에서 상기 적어도 하나 이상의 신호 프레임을 수신된 순서로 전처리하는 신호 프레임 전처리단계, 특징정보 추출부에서 상기 전처리된 적어도 하나 이상의 신호 프레임으로부터 상기 방출된 근적외선에 의해 변형된 상기 피부의 특징정보를 추출하는 특징정보 추출단계, 특징정보 분류부에서 상기 추출된 특징정보를 기 설정된 사용자의 의도에 분류하는 특징정보 분류단계를 포함할 수 있다.

- [0025] 상기 근적외선 방출단계는 상기 피부에 방출된 근적외선에 의한 산란현상에 따른 신호 프레임의 수신에 상기 피부에 방출된 근적외선에 의한 반사현상의 영향을 최소화하도록 시준된(Collimated) 근적외선을 방출할 수 있다.
- [0026] 상기 신호 프레임 수신단계에서 상기 수신하는 신호 프레임은 상기 피부에 방출된 근적외선에 의한 산란현상으로부터 측정된 공간 분해 확산 반사율(SRDR: Spatially Resolved Diffuse Reflectance)이 포함된 근적외선 이미지 형태라 할 수 있다.
- [0027] 상기 신호 프레임 수신부는 상기 근적외선 이미지에 목표 대역의 근적외선 신호 외에 다른 대역의 노이즈 신호가 포함되지 않도록 근적외선 필터를 포함할 수 있다.
- [0028] 상기 신호 프레임 전처리단계는 상기 근적외선 이미지를 수신 순서대로 전처리하고, 상기 특징 추출단계의 데이터 입력 형태에 따라 데이터 형태를 재구성할 수 있다.
- [0029] 상기 특징정보 추출단계는 상기 전처리된 근적외선 이미지로부터 상기 사용자 의도에 분류하기 위한 적어도 하나 이상의 특징을 추출하고, 상기 특징정보 분류단계는 상기 추출된 적어도 하나 이상의 특징을 상기 사용자 의도에 분류하여 매핑할 수 있다.
- [0030] 상기 특징정보 추출단계는 심층 신경망(DNN: Deep Neural Network)을 기반으로 사전 학습된 시공간 오토인코더(STAE: Spatiotemporal Autoencoder)의 인코더를 이용하여 상기 특징정보를 추출하고, 상기 특징정보 분류단계는 DEC(Deep Embedded Clustering) 기법을 기반으로 사전 학습된 분류기를 이용하여 상기 추출된 적어도 하나 이상의 특징을 분류하여 기 설정된 사용자 의도에 매핑할 수 있다.

**발명의 효과**

- [0031] 이상에서 설명한 바와 같은 헤드셋 형태의 사용자 인터페이스 시스템 및 이를 위한 사용자 특징정보 분류 방법에 따르면,
- [0032] 첫째, 얼굴 제스처의 피부 변형을 감지한 후, 이를 UI 입력 명령으로 이용함으로써 AR/VR용 핸드프리 NUI(Natural User Interface) 기술을 구현할 수 있는 효과를 가진다.
- [0033] 둘째, 얼굴 피부의 움직임을 비침습적으로 쉽게 감지하고, 감지된 얼굴 동작을 입력 명령에 매핑하여 헤드셋 사용자와 AR 시스템 간의 핸드프리 상호작용을 지원할 수 있는 효과를 가진다.
- [0034] 셋째, 헤드셋 형태의 사용자 인터페이스 기술을 구현함으로써, 특수한 상황 또는 환경에서도 적용이 가능한 효과를 가진다.
- [0035] 넷째, 피부 변형에 관한 데이터를 2차원 형태로 수신함으로써, 양적 및 질적으로 향상된 데이터를 수신할 수 있는 효과를 가진다.
- [0036] 다섯째, 연속적으로 수신된 데이터의 변화량 산출을 통해 사용자 동작 변화에 관한 데이터를 생성함으로써, 데이터 분류가 용이한 효과를 가진다.
- [0037] 여섯째, 특징 추출 및 분류를 비지도 학습방법으로 구현함으로써, 태깅 정보 없이도 학습용 데이터셋을 분류할 수 있는 효과를 가진다.
- [0038] 일곱째, DEC(Deep Embedded Clustering) 기법의 적용을 통해, 비지도 학습방법의 한계점인 정확도 문제점을 개선할 수 있는 효과를 가진다.

**도면의 간단한 설명**

- [0039] 도 1은 본 발명의 실시예에 따른 헤드셋 형태의 사용자 인터페이스 시스템을 나타낸 구성도.
- 도 2는 본 발명의 실시예에 따른 사용자 인터페이스를 위한 사용자 특징정보 분류 방법을 나타낸 순서도.
- 도 3은 본 발명의 일 실시예에 따른 핸드프리 UI 시스템을 나타낸 도면.

도 4는 도 3에 따른 적외선 카메라의 촬영 원리는 나타낸 도면.

도 5는 본 발명의 일 실시예에 따라 구현된 것으로, AR 헤드셋에 적용된 센서 모듈을 나타낸 도면.

도 6은 본 발명의 일 실시예로서, 사용자가 AR 헤드셋을 착용하고 있는 동안 센서 모듈에 의해 기록된 일부 IR 확산 패턴을 보여주는 도면.

도 7은 본 발명의 일 실시예로서, 클러스터링 네트워크에 대한 전처리 과정을 나타낸 도면.

도 8은 본 발명의 일 실시예로서, 센서 데이터 특징을 추출하는 네트워크 구조를 나타낸 도면.

도 9는 본 발명의 일 실시예로서, 특징 추출을 위해 사용된 STAE의 네트워크 구조에 대한 세부사항을 나타낸 도면.

도 10은 본 발명의 일 실시예로서, STAE에 기반한 특징 추출기와 DEC에 기반한 특징 분류기로 구성된 분류기 네트워크를 나타낸 도면.

도 11은 발명의 일 실시예로서, 미세 조정 기법(DEC)을 적용하여 클러스터링 한 결과를 나타낸 도면.

도 12는 본 발명의 일 실시예로서, 맞춤형 응용 어플리케이션을 사용하여 시연한 스크린샷을 보여주는 도면.

**발명을 실시하기 위한 구체적인 내용**

- [0040] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다. 본 발명의 특징 및 이점들은 첨부 도면에 의거한 다음의 상세한 설명으로 더욱 명백해질 것이다. 이에 앞서, 본 명세서 및 청구범위에 사용된 용어나 단어는 발명자가 그 자신의 발명의 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야 할 것이다. 또한 본 발명과 관련된 공지 기능 및 그 구성에 대한 구체적인 설명은 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우, 그 구체적인 설명을 생략하였음에 유의해야 할 것이다.
- [0041] 도 1은 본 발명의 실시예에 따른 헤드셋 형태의 사용자 인터페이스 시스템을 나타낸 구성도이다.
- [0042] 도 1을 참조하면, 본 발명의 실시예에 따른 헤드셋 형태의 사용자 인터페이스 시스템은 크게 근적외선 방출부(100), 신호 프레임 수신부(200), 신호 프레임 전처리부(300), 특징정보 추출부(400), 특징정보 분류부(500)를 포함할 수 있다.
- [0043] 근적외선 방출부(100)는 일정한 거리를 두고 근적외선을 피부에 방출할 수 있다. 이는 비접촉 방식으로 피부를 센싱하기 위한 구성요소라 할 수 있다. 이 때 근적외선 방출부(100)는 LED가 아닌, 레이저 다이오드라 할 수 있다. 그 이유는 LED의 경우 피부에 접촉하여 사용함으로써 신호 인식이 가능한 반면, 레이저 다이오드는 피부에 접촉하지 않고도 신호 인식이 가능하기 때문이다.
- [0044] 이 때 근적외선 방출부(100)는 상기 피부에 방출된 근적외선에 의한 산란현상에 따른 신호 프레임의 수신이 상기 피부에 방출된 근적외선에 의한 반사현상의 영향을 최소화하도록 시준된(Collimated) 근적외선을 방출할 수 있다. 그 이유는 상기 피부에 방출된 근적외선으로 인해 발생한 적외선 산란현상이 상기 피부에 방출된 근적외선으로 인해 발생한 적외선 반사현상에 가려져 신호 인식 또는 측정이 불가능할 수도 있기 때문이다. 이에 근적외선 방출부(100)는 충분히 시준된(Collimated) 근적외선을 피부에 방출할 수 있다.
- [0045] 신호 프레임 수신부(200)는 상기 근적외선 방출부(100)로부터 방출된 근적외선에 의해 상기 피부로 입사되는 근적외선 중 상기 피부 내에서 산란되어 상기 피부 밖으로 방출되는 근적외선의 신호가 포함된 적어도 하나 이상의 신호 프레임을 수신할 수 있다. 이는 피부 변형에 관한 신호를 수신하기 위한 구성요소라 할 수 있다.
- [0046] 이 때 신호 프레임 수신부(200)는 적외선 포토다이오드(PD: photodiode)가 아닌, 적외선 카메라라 할 수 있다. 그 이유는 적외선 포토다이오드는 1차원 포인트 데이터(point data)를 획득하는 반면, 적외선 카메라는 카메라 영상 데이터를 이용한 2차원 데이터를 획득할 수 있기 때문이다. 이는 양적 및 질적으로 향상된 데이터라 할 수 있다.
- [0047] 신호 프레임 수신부(200)에서 수신하는 신호 프레임은 피부에 방출된 근적외선에 의한 산란현상으로부터 측정된 공간 분해 확산 반사율(SRDR: Spatially Resolved Diffuse Reflectance)이 포함된 근적외선 이미지 형태라 할 수 있다. 이는 근적외선에 의한 산란현상에 따른 공간 분해 확산 반사율(SRDR: Spatially Resolved Diffuse

Reflectance)을 통해 사용자의 의도를 파악하기 위한 특징이라 할 수 있다.

- [0048] 한편, 신호 프레임 수신부(200)는 근적외선 이미지에 목표 대역의 근적외선 신호 외에 다른 대역의 노이즈 신호가 포함되지 않도록 근적외선 필터를 포함할 수 있다.
- [0049] 신호 프레임 전처리부(300)는 적어도 하나 이상의 신호 프레임을 수신된 순서로 전처리할 수 있다. 이는 수신된 신호 프레임으로부터 특징정보 추출 및 분류의 성능을 높이기 위한 구성요소라 할 수 있다. 이 때 전처리 순서를 수신된 순서로 하는 이유는 방출된 근적외선에 의해 피부 내에서 확산되는 변화를 확인할 수 있기 때문이다.
- [0050] 이 때 신호 프레임 전처리부(300)는 하기 특징정보 추출부(400)의 데이터 입력 형태에 따라 데이터 형태를 재구성할 수 있다.
- [0051] 특징정보 추출부(400)는 상기 신호 프레임 전처리부(300)로부터 전처리된 적어도 하나 이상의 신호 프레임으로부터 상기 방출된 근적외선에 의해 변형된 상기 피부의 특징정보를 추출할 수 있다.
- [0052] 보다 상세하게, 특징정보 추출부(400)는 상기 신호 프레임 전처리부(300)로부터 전처리된 근적외선 이미지로부터 상기 사용자 의도에 분류하기 위한 적어도 하나 이상의 특징을 추출할 수 있다.
- [0053] 이 때 특징정보 추출부(400)는 심층 신경망(DNN: Deep Neural Network)을 기반으로 사전 학습된 시공간 오토인코더(STAE: Spatiotemporal Autoencoder)의 인코더를 이용하여 상기 특징정보를 추출할 수 있다.
- [0054] 특징정보 분류부(500)는 상기 특징정보 추출부(400)로부터 추출된 특징정보를 기 설정된 사용자의 의도에 분류할 수 있다.
- [0055] 보다 상세하게, 특징정보 분류부(500)는 상기 특징정보 추출부(400)로부터 추출된 적어도 하나 이상의 특징을 사용자 의도에 분류하여 매핑할 수 있다.
- [0056] 이 때 특징정보 분류부(500)는 DEC(Deep Embedded Clustering) 기법을 기반으로 사전 학습된 분류기를 이용하여 상기 추출된 적어도 하나 이상의 특징을 분류하여 기 설정된 사용자 의도에 매핑할 수 있다.
- [0057] 이러한 특징정보 추출부(400) 및 특징정보 분류부(500)는 비지도 학습 방법을 기반으로 하는데, 이는 태깅 정보 없이도 학습용 데이터셋을 분류할 수 있다는 특징을 가질 수 있다. 그 중 DEC(Deep Embedded Clustering) 기법은 비지도 학습 방법의 한계인 정확도 문제점을 개선하기 위한 기법이라 할 수 있다.
- [0058] 도 2는 본 발명의 실시예에 따른 사용자 인터페이스를 위한 사용자 특징정보 분류 방법을 나타낸 순서도이다.
- [0059] 도 2를 참조하면, 본 발명의 실시예에 따른 사용자 인터페이스를 위한 사용자 특징정보 분류 방법은 크게 근적외선 방출단계(S100), 신호 프레임 수신단계(S200), 신호 프레임 전처리단계(S300), 특징정보 추출단계(S400), 특징정보 분류단계(S500)를 포함할 수 있다.
- [0060] 근적외선 방출단계는 근적외선 방출부(100)에서 일정한 거리를 두고 근적외선을 피부에 방출할 수 있다(S100). 이는 비접촉 방식으로 피부를 센싱하기 위한 단계라 할 수 있다. 이 때 근적외선 방출단계(S100)에서 근적외선 방출부(100)는 LED가 아닌, 레이저 다이오드라 할 수 있다. 그 이유는 LED의 경우 피부에 접촉하여 사용함으로써 신호 인식이 가능한 반면, 레이저 다이오드는 피부에 접촉하지 않고도 신호 인식이 가능하기 때문이다.
- [0061] 이 때 근적외선 방출단계(S100)는 상기 피부에 방출된 근적외선에 의한 산란현상에 따른 신호 프레임의 수신에 상기 피부에 방출된 근적외선에 의한 반사현상의 영향을 최소화하도록 시준된(Collimated) 근적외선을 방출할 수 있다. 그 이유는 상기 피부에 방출된 근적외선으로 인해 발생한 적외선 산란현상이 상기 피부에 방출된 근적외선으로 인해 발생한 적외선 반사현상에 가려져 신호 인식 또는 측정이 불가능할 수도 있기 때문이다. 이에 근적외선 방출단계(S100)는 충분히 시준된(Collimated) 근적외선을 피부에 방출할 수 있다.
- [0062] 신호 프레임 수신단계는 상기 근적외선 방출단계(S100)로부터 방출된 근적외선에 의해 상기 피부로 입사되는 근적외선 중 상기 피부 내에서 산란되어 상기 피부 밖으로 방출되는 근적외선의 신호가 포함된 적어도 하나 이상의 신호 프레임을 수신할 수 있다(S200). 이는 피부 변형에 관한 신호를 수신하기 위한 단계라 할 수 있다.
- [0063] 이 때 신호 프레임 수신단계(S200)에서 신호 프레임 수신부(200)는 적외선 포토다이오드(PD: photodiode)가 아닌, 적외선 카메라라 할 수 있다. 그 이유는 적외선 포토다이오드는 1차원 포인트 데이터(point data)를 획득하는 반면, 적외선 카메라는 카메라 영상 데이터를 이용한 2차원 데이터를 획득할 수 있기 때문이다. 이는 양적 및 질적으로 향상된 데이터라 할 수 있다.
- [0064] 신호 프레임 수신단계(S200)에서 수신하는 신호 프레임은 피부에 방출된 근적외선에 의한 산란현상으로부터 측

정된 공간 분해 확산 반사율(SRDR: Satially Resolved Diffuse Relectance)이 포함된 근적외선 이미지 형태라 할 수 있다. 이는 근적외선에 의한 산란현상에 따른 공간 분해 확산 반사율(SRDR: Spatially Resolved Diffuse Reflectance)을 통해 사용자의 의도를 파악하기 위한 특징이라 할 수 있다.

- [0065] 한편, 신호 프레임 수신부(200)는 근적외선 이미지에 목표 대역의 근적외선 신호 외에 다른 대역의 노이즈 신호가 포함되지 않도록 근적외선 필터를 포함할 수 있다.
- [0066] 신호 프레임 전처리단계는 적어도 하나 이상의 신호 프레임을 수신된 순서로 전처리할 수 있다(S300). 이는 수신된 신호 프레임으로부터 특징정보 추출 및 분류의 성능을 높이기 위한 단계라 할 수 있다. 이 때 전처리 순서를 수신된 순서로 하는 이유는 방출된 근적외선에 의해 피부 내에서 확산되는 변화를 확인할 수 있기 때문이다.
- [0067] 이 때 신호 프레임 전처리단계(S300)는 하기 특징정보 추출단계(S400)의 데이터 입력 형태에 따라 데이터 형태를 재구성할 수 있다.
- [0068] 특징정보 추출단계는 상기 신호 프레임 전처리단계(S300)로부터 전처리된 적어도 하나 이상의 신호 프레임으로부터 상기 방출된 근적외선에 의해 변형된 상기 피부의 특징정보를 추출할 수 있다(S400).
- [0069] 보다 상세하게, 특징정보 추출단계(S400)는 상기 신호 프레임 전처리단계(S300)로부터 전처리된 근적외선 이미지로부터 상기 사용자 의도에 분류하기 위한 적어도 하나 이상의 특징을 추출할 수 있다.
- [0070] 이 때 특징정보 추출단계(S400)는 심층 신경망(DNN: Deep Neural Network)을 기반으로 사전 학습된 시공간 오토인코더(STAE: Spatiotemporal Autoencoder)의 인코더를 이용하여 상기 특징정보를 추출할 수 있다.
- [0071] 특징정보 분류단계는 상기 특징정보 추출단계(S400)로부터 추출된 특징정보를 기 설정된 사용자의 의도에 분류할 수 있다(S500).
- [0072] 보다 상세하게, 특징정보 분류단계(S500)는 상기 특징정보 추출단계(S400)로부터 추출된 적어도 하나 이상의 특징을 사용자 의도에 분류하여 매핑할 수 있다.
- [0073] 이 때 특징정보 분류단계(S500)는 DEC(Deep Embedded Clustering) 기법을 기반으로 사전 학습된 분류기를 이용하여 상기 추출된 적어도 하나 이상의 특징을 분류하여 기 설정된 사용자 의도에 매핑할 수 있다.
- [0074] 이러한 특징정보 추출단계(S400) 및 특징정보 분류단계(S500)는 비지도 학습 방법을 기반으로 하는데, 이는 태그 정보 없이도 학습용 데이터셋을 분류할 수 있다는 특징을 가질 수 있다. 그 중 DEC(Deep Embedded Clustering) 기법은 비지도 학습 방법의 한계인 정확도 문제점을 개선하기 위한 기법이라 할 수 있다.
- [0075] 보다 상세하게 본 발명을 설명하자면, 도 3 내지 도 12를 참조할 수 있다.
- [0076] 도 3은 본 발명의 일 실시예에 따른 핸드프리 UI 시스템을 나타낸 도면이고, 도 4는 도 3에 따른 적외선 카메라의 촬영 원리는 나타낸 도면이다.
- [0077] 도 5는 본 발명의 일 실시예에 따라 구현된 것으로, AR 헤드셋에 적용된 센서 모듈을 나타낸 도면이라 할 수 있다.
- [0078] 먼저 도 3 및 도 4를 참조하면, 본 발명의 일 실시예에 따른 핸드프리 UI 시스템에서 센서 모듈(Sensor module)은 IR LD와 IR 카메라를 포함할 수 있다. IR 카메라는 IR 확산 패턴에 대한 이미지를 촬영(캡처)하고, IR LD은 IR 광을 피부에 방출하는 구성요소라 할 수 있다.
- [0079] 도 5에서 (a)은 센서 모듈(Sensor module)로서 USB 카메라와 NIR 레이저 다이오드를 포함할 수 있다. 이는 Epson BT-350 AR 안경의 왼쪽에 설치한 것이라 할 수 있다. (b)은 사용자가 헤드셋을 착용한 모습으로서, 레이저 다이오드(Laser diode)는 왼쪽 뺨 근처 피부를 타겟으로 할 수 있다. 왼쪽 뺨 근처 피부는 사용자가 링크 제스처를 했을 때 변형될 수 있다.
- [0080] 사람 피부의 변형은 피부 밑(피부 내)에서 전달되는 IR SRDR의 변화를 측정하여 검출할 수 있다. 이 때 SR(Spatially resolved)은 광의 입사지점으로부터 방사상 거리(Radial distance)에 따라 측정된 확산반사이며, DR(Diffuse reflectance)은 상기 확산반사에 의해 재방사되는 플럭스(Flux) 양의 비율인 산란반사라 할 수 있다. DR의 강도는 피부 밑에서 분포되는 콜라겐 섬유의 미세구조에 의해 영향을 받을 수 있다.
- [0081] 본 발명의 일 실시예에서는 사람 피부에서 콜라겐 섬유 분포를 나타내는 할선(Landger's line)이 피부의 SRDR 측정 방향을 따른다는 것에 기초하여, IR 확산 반사체를 이용하여 사용자의 표정을 감지하는 실시 예를 구현할

수 있다.

- [0082] 이를 기초로 도 3 내지 도 5의 실시 예를 보다 상세하게 설명하면 다음과 같다.
- [0083] 도 3을 참조하면, 본 발명의 일 실시예에 따른 핸드프리 UI 시스템에서는 크게 센서 착용자의 피부 표면으로부터 IR 확산 패턴을 검출하는 센서 모듈(Sensor module)과, DNN 구조에 기반한 센서 데이터 분류기(Sensor Data Clustering Module)를 포함할 수 있다.
- [0084] 센서 모듈은 웨어러블 AR 헤드셋에 장착할 수 있다. 센서 모듈은 LD와 USB 카메라를 한 쌍으로 구성할 수 있다. 이는 각각 IR 방출기와 수신기라 할 수 있다. 사용자가 AR 안경을 쓰고 있는 동안, LD로부터 방출되는 IR 광은 사용자의 피부를 통해 확산될 수 있다. 카메라는 확산된 IR 광의 순차적 이미지를 촬영할 수 있고, 이를 분류기로 전송하여 사용자의 피부 변형 상태를 인터페이스 명령으로 변환할 수 있다.
- [0085] 이 때 일 실시예를 통해 제작된 AR 헤드셋은 상용 AR 헤드셋(Epson BT-350)과, 얼굴 제스처 추적 센서 모듈을 이용할 수 있다. 그리고 상기 얼굴 제스처 추적 센서 모듈은 도 5에서와 같이 AR 안경의 왼쪽에 장착되었고, ROI 센서는 사용자가 윈크 제스처를 할 때 피부가 국소적으로 퍼지거나 압축되는 사용자의 왼쪽 눈과 귀 사이의 지점을 목표로 한 것이라 할 수 있다.
- [0086] NIR LD(즉, IR 방출기)는 평균 출력 전압이 0.4mW인 850nm 중심 파장을 가졌는데, 이는 레벨 1 수준의 레이저로 분류될 수 있다. 이러한 NIR 조명은 사용자의 시야를 방해하지 않았으며, 사용자의 눈에 대한 잠재적 유해 영향을 방지하고자 NIR 선량을 안전한 수준으로 제한한 것이라 할 수 있다.
- [0087] 그리고 상기 NIR의 밴드는 무해한 광 스펙트럼 중 가장 깊게 피부로 침투할 수 있다. 이는 센서 구현을 위한 광원을 선택하는 데에 있어서, 낮은 전력 소비량과 함께 중요한 특성이라 할 수 있다. 발광 다이오드(LED) 대신 LD를 선택한 이유는 LED로부터 나오는 IR 광이 실제보다 내구성이 더 좋지만, LED로부터 나오는 IR 광은 LD 빛과는 대조적으로 피부에 직접 접촉하지 않을 때 사용자 피부를 통해 전달될 수 있는 시준(Collimation)이 불충분하기 때문이다.
- [0088] IR 방출기는 LD, 보호회로, 시준 렌즈(Collimation lens)를 포함할 수 있다. 보호회로는 LD의 수명을 연장할 수 있고, 시준 렌즈는 방출의 분산을 감소시키기 위해 IR 레이저 빔을 시준할 수 있다. IR 수신기는 OV2710 칩셋 카메라를 포함하는 센서 모듈이라 할 수 있다. 이 때 센서 모듈은 주변 가시광선으로부터 발생하는 방해 없이, IR 광을 감지할 수 있는 IR 밴드 패스 필터를 내장할 수 있다. 카메라의 시야(FOV: Field of view)는 100 각도이고, 제조사 사양에 따라 320 X 240 해상도를 가진 120fps로 기록할 수 있다. 본 발명의 일 실시예에서의 카메라라는 데이터 처리 지연시간 때문에, (평균적으로) 20fps로 실행될 수 있다. 실제로 입력 데이터에 320 X 240 해상도 그레이스케일(Grayscale) jpeg 이미지를 사용했다.
- [0089] 도 6은 본 발명의 일 실시예로서, 사용자가 AR 헤드셋을 착용하고 있는 동안 센서 모듈에 의해 기록된 일부 IR 확산 패턴을 보여주는 도면이라 할 수 있다.
- [0090] 보다 상세하게, 도 6에서 (a)은 얼굴 제스처 없이 촬영된 IR SRDR이라 할 수 있다. (b)은 사용자가 윈크 제스처를 취하는 동안 촬영된 IR SRDR이라 할 수 있다.
- [0091] 도 6을 참조하면, 카메라에 설치된 NIR 대역 밴드 패스 필터는 850nm(NIR)을 제외한 모든 광의 파장대를 절단할 수 있다. 패턴의 모양은 사용자의 얼굴 제스처와 관련해서 바뀔 수 있다.
- [0092] 센서 데이터 분류기(Sensor Data Clustering Module)로 입력되는 입력 데이터는 사용자 피부를 촬영한 NIR SRDR 이미지 형태에 포함될 수 있다. 카메라로 촬영한 SRDR 모양은 사용자의 얼굴 제스처가 동일함에도 불구하고, 얼굴 제스처의 모델링이 거의 되지 않는 다양한 요인에 의해, 피투피(Person-to-person) 변형이 존재할 수도 있다. 본 발명의 일 실시예에서의 센서 데이터 분류기는 사용자의 제스처를 인식하기 위해 촬영된 SRDR 모양의 변형을 감지해야한다.
- [0093] 이에 본 발명에서는 실시예로서, 신경망 네트워크 센서 데이터 분류기에 대해 감독되지 않은 학습, 즉 비지도 학습 방법을 활용할 수 있다. 비지도 학습 방법은 피투피(Person-to-person) 변형에 대응할 필요가 있을 경우 쉽게 보정할 수 있는 방안이라 할 수 있다. 그 이유는 비지도 학습 방법은 훈련 데이터 셋을 생성하기 위한 데이터 주석 작업을 하지 않을 수 있기 때문이다.
- [0094] 센서 데이터 분류기를 훈련하는 과정은 다음과 같은 2단계로 구성할 수 있다.
- [0095] (1단계) 오토인코더(Autoencoder)를 사용한 비지도 학습을 통해 데이터 셋으로부터 특징을 추출 할 수 있다.

- [0096] (2단계) 추출된 특징을 클러스터링(Clustering) 및 미세 조정(Fine-tuning)하여 분류기의 성능을 향상시킬 수 있다.
- [0097] 상기 도 5에 도시된 바와 같이, 본 발명의 실시예에서는 내장 IR 센서 모듈을 사용하여 분류기 신경망 네트워크에 대한 훈련 데이터를 수집할 수 있다. 보다 상세하게, 상기 분류기 네트워크에 대한 훈련 데이터 셋에 대한 데이터 주석 작업 없이, 1인당 5,000개 이상의 순차적인 SRDR 이미지를 수집할 수 있다. 여기서, 분류기 네트워크를 훈련시키기 위해 10명의 다른 사용자로부터 수집한 81,758 개의 SRDR 이미지를 이용할 수 있다.
- [0098] 분류기 네트워크로의 데이터 입력은 4개의 320 X 240 pixel IR 확산 스냅 샷 이미지 시리즈로서, 이는 카메라에 의해 순차적으로 촬영되어 입력 이미지 큐(Queue)에 저장될 수 있다. 상기 분류기 네트워크는 상기 이미지를 네트워크의 특징 추출 부분으로 보내기 전에 전처리를 수행할 수 있다.
- [0099] 도 7은 본 발명의 일 실시예로서, 클러스터링 네트워크에 대한 전처리 과정을 나타낸 도면이라 할 수 있다.
- [0100] 보다 상세하게, 도 7에서의 전처리 유닛은 두 개의 이미지 사이의 차를 계산할 수 있다. 계산 이후, 전처리 유닛은 입력 이미지를 임계값(Thresholds)으로 설정할 수 있고, 상기 두 개의 이미지 사이의 픽셀 별 뺄셈을 수행할 수 있으며, 이를 클러스터링 네트워크 입력 사이즈에 맞추기 위해 28 X 28pixel 크기로 조정하여 클러스터링 네트워크에서 요구하는 입력 길이 만큼 한 그룹으로 묶을 수 있다.
- [0101] 도 7에 도시된 바와 같이, 전처리 유닛은 SRDR 컨투어의 형태 변동에 초점을 맞추어 분류기 네트워크를 만들기 위해, SRDR 컨투어 형태의 변형을 강조한 다음, STAE의 RNN 창 크기에 맞추어 4개의 순차 이미지를 그룹화할 수 있다.
- [0102] 또한, 전처리 유닛은 먼저 각 SRDR의 컨투어형을 강조하기 위해 입력 이미지를 기 설정된 임계값을 가진 흑백 이진 이미지로 변환할 수 있다. 그런 다음, 전처리 유닛은 SRDR 컨투어형의 변형을 강조하기 위해, 일정한 시간 간격 사이에서의 SRDR 이진 이미지의 픽셀 단위 차를 계산할 수 있다. 이 때 일정한 시간 간격은 현재 샘플링 시간[n]과 이전 샘플링 시간[n-2] 사이인 2개의 샘플링 간격이라 할 수 있다. 이후 계산 과정의 축소를 위해, SRDR 이진 이미지를 28 X 28pixel 이미지 크기로 줄일 수 있다. 줄이기 전의 이미지는 이미지 큐에 저장할 수 있다. 마지막으로, 전처리 유닛을 특징 추출 부분에 입력하기 위해, 입력 이미지 시퀀스에 대한 세트로서 4개의 순차 이미지를 합성할 수 있다.
- [0103] 입력 시퀀스 끝에 있는 이미지([n]th)는 카메라로부터 촬영한 가장 최신의 이미지라 할 수 있다.
- [0104] 도 8은 본 발명의 일 실시예로서, 센서 데이터 특징을 추출하는 네트워크 구조를 나타낸 도면이다.
- [0105] 보다 상세하게, 도 8에서 STAE은 인코더(Encoder)와 디코더(Decoder)로 구성되며, STAE의 학습 이후 특징 추출을 위해서는 인코더 부분만이 사용될 수 있다.
- [0106] 도 8에 도시된 바와 같이, 본 발명의 일 실시예의 센서 카메라는 사용자의 얼굴 चे스트처를 순차적인 NIR SRDR 이미지로 변환할 수 있다. 변환된 이미지들은 분류기 네트워크에 입력된 데이터였고, 이는 윙크 चे스트처를 암시하는지 여부를 결정할 수 있는 데이터라 할 수 있다. 이미지 클러스터링은 비지도 학습 방법에서 이미지 분류를 위한 최적의 방법이라 할 수 있다. 다만, 이러한 방법은 여전히 아래 2가지 이유로 한계점을 안고 있다.
- [0107] 첫 번째 이유는 입력 이미지 데이터의 차원이다. 이미지는 종종 매우 높은 차원성을 가지며, 이는 K-평균 알고리즘(K-means algorithm)의 성능에 큰 영향을 미치는 차원성의 저주(Curse of dimensionality)를 피하기 위해 이를 줄여야 한다.
- [0108] 두 번째 이유는 이미지에서의 특징이 주로 2D 또는 3D 로컬 구조를 가지는데, 이는 클러스터링을 하는 동안 무시해서는 안 된다. 이러한 특징은 로컬 구조의 형태에 최적화된 커널(Kernels)을 필요로 할 수 있다. 나아가, 센서로 입력되는 데이터는 시간적 특징을 가지는 순차적인 데이터이기 때문에, 본 발명의 클러스터링 네트워크는 입력으로부터 시간적 특징과 공간 정보를 추출할 수 있어야 한다.
- [0109] 본 발명의 실시예에서는 사용자의 감독 없이 입력의 시공간적 특성을 보존하면서 입력 차수를 줄이는 STAE을 채택할 수 있다. 오토인코더의 인코더 부분은 클러스터링 방법으로 특징 추출부를 만들 수 있다. 인코더는 오토인코더를 훈련시키는 동안 압축 손실을 최소화하면서 입력 데이터를 압축하는 방법을 학습할 수 있기 때문이다.
- [0110] STAE은 입력 데이터로부터 각각 시간적 및 공간적 특징을 추출하기 위한 RNN(Recurrent neural network)와 CNN(Convolutional neural network)으로 구성된 오토인코더의 타입 중 하나라 할 수 있다. n개의 이미지 시퀀

스  $\{x_i \in X\}_{i=1}^n$  를 중심  $\mu_j, j=1, \dots, k,$  STAE으로 대표되는 k 클러스터에 클러스터링한 경우, 비선형 매핑  $f_\theta : X \rightarrow Z,$  을 가진 각  $x_i$  의 차원 수를 줄일 수 있다. 여기서  $\theta$  은 커널의 중량이고, Z은 잠재된 특징 공간이라 할 수 있다. 커널은 차원 감소를 지원할 뿐만 아니라, 이미지 시퀀스에서 특징을 감지할 수 있는 용량을 가질 수 있다. STAE 학습이 끝나면, 학습  $\theta$  은 완료되고(이는 분류기 파라미터의 초기화를 의미할 수 있다.), STAE의 디코더 부분은 더 이상 필요하지 않을 수 있다. 그런 다음 후속 부스팅 단계에서 폐기될 수 있다.

- [0111] 오토인코더는 비지도 학습 방법으로 데이터 표현을 학습하는 인공 신경망의 일종이며, 일반적으로 인코더-디코더 쌍의 일부분이라 할 수 있다. 인코더는 입력 데이터를 압축하기 위해 훈련되고, 디코더는 압축된 데이터를 최소의 손실이 있는 원래 입력 데이터로 복원하도록 훈련될 수 있다. 여기서, 인코더는 입력 데이터에서 불필요한 중복(Redundancy)을 줄이기 위한 데이터 표현을 학습하도록 유도될 수 있다.
- [0112] 도 8과 같이, 본 발명의 일 실시예에서의 STAE은 전처리 유닛으로부터 입력 데이터로 4개의 연속적인 이미지를 수신할 수 있다.
- [0113] 도 9는 본 발명의 일 실시예로서, 특징 추출을 위해 사용된 STAE의 네트워크 구조에 대한 세부사항을 나타낸 도면이다.
- [0114] 도 9에 도시된 바와 같이, 공간적 특징을 찾고 입력 이미지 크기를 줄이기 위해, STAE 네트워크에서 2개의 3D CNN 레이어가 최우선적으로 배치될 수 있다. 입력 이미지의 차원 수를 줄이기 위해, 인코더의 각 컨볼루션 레이어는 non-padding mode에서 작동할 수 있다. 커널의 가중치를 초기화하기 위해, 분산 스케일을 사용할 수 있다. 각 CNN 레이어는 이미지 크기를 절반으로 줄인 최대 풀링 레이어가 뒤따를 수 있다. 이 때 입력 데이터는 3D(폭, 높이, 샘플링 시간)이기 때문에, 네트워크는 일반적인 2D CNN이 아니라 3D CNN 레이어를 사용하여 입력 데이터를 처리할 필요가 있다.
- [0115] 그런 다음, 연속적인 2개의 컨볼루션 LSTM(ConvLSTM: Convolutional long-short-term memory) 레이어는 압축된 입력 데이터로부터 일시적 특징을 추출할 수 있다.
- [0116] 기존의 RNN은 숨겨진 상태 매개변수를 통해 일시적 특징을 학습할 수 있지만, 이는 기울기 소실 문제(Vanishing gradients problem)를 가지고 있다. LSTM은 망각 게이트(forget gate)라고 불리는 반복적인 게이트를 도입함으로써 기울기 소실 문제를 완화할 수 있다. ConvLSTM은 영상 데이터와 같은 이미지 프레임의 시퀀스로부터 시간 정보를 찾도록 설계된 LSTM의 일종으로서, 2D 형상 가중치를 가질 수 있다. 입력과 가중치 사이의 곱셈 연산을 사용하는 대신, ConvLSTM은 입력 간 컨볼루션 연산을 사용할 수 있다. 상기 LSTM 레이어는 각 LSTM 상태를 통해 시공간적 특징을 전파할 수 있는 기능을 가질 수 있다.
- [0117] 마지막으로, 본 발명의 일 실시예에서는 압축된 입력 이미지와 동일한 커널 크기를 가지도록 설계된 ConvLSTM 레이어 뒤에 또 다른 3D CNN 레이어를 배열하여, 4개의 2D 이미지를 추출된 특징을 암시할 수 있는 4개의 2D 벡터로 변환할 수 있다.
- [0118] 상기 배열된 3D CNN 레이어는 클러스터링을 위해 이전 레이어의 출력을 2D 공간에 매핑할 수 있다. 본 발명에서는 의도한 클러스터의 개수와 일치하도록 특징 공간의 차원을 기 설정할 수 있다. 2개의 상태(피부 변형 상태 및 변형이 없는 상태)만 포함하고자 하는 클러스터의 개수 때문에, 본 발명에서는 특징 공간의 차원을 2D로 설정할 수 있다. 4개의 추출된 벡터 중 가장 최근에 촬영된 데이터의 특징을 포함하고 있는 마지막 벡터만 센서 데이터 클러스터링 과정에서 특징 추출의 최종 결과로 이용될 수 있다.
- [0119] 디코더는 네트워크의 구현의 편의를 위해, 최대 언폴링 레이어를 사용하는 대신 평균 언폴링 레이어를 채택하는 것을 제외하고 인코더의 구성과 완전히 대칭적으로 구성할 수 있다. 오토인코더에 대한 학습 이후, 디코더는 더 이상 필요하지 않기 때문에 분류기 네트워크에서 폐기될 수 있다. 특징 추출기에서는 인코더 부분만 사용할 수 있다.
- [0120] 도 10은 본 발명의 일 실시예로서, STAE에 기반한 특징 추출기와 DEC에 기반한 특징 분류기로 구성된 분류기 네트워크를 나타낸 도면이다.
- [0121] 도 10에 도시된 바와 같이, 클러스터링은 비지도 학습 방법 중 하나로, 이는 유사한 특성을 가진 데이터를 그룹화하는 데에 사용할 수 있다. 지도 학습 방법에 기반한 분류기와 비교하여, 비지도 학습 방법에 기반한 분류기는 애매한 분류 대상 데이터에 대해 배경 답변(background answer)을 할 필요가 없다는 장점을 가질 수 있다.

- [0122] 한편, 본 발명에서는 일 실시예로서 클러스터링 구현을 위해 k-평균 알고리즘을 채택할 수 있다. 그 이유는 k-평균 알고리즘은 다른 클러스터링 알고리즘에 비해 간단하고, 비교적 효율적이기 때문이다. 다만, 비지도적 클러스터링에 기반한 k-평균 알고리즘은 일반적으로 지도적 분류 방법보다 정확성이 다소 낮다.
- [0123] 본 발명의 실시예에서의 센서는 클러스터링 네트워크 성능을 강화하기 위해 DEC(Deep embedded clustering) 방법을 채택할 수 있다. 본 발명에서의 클러스터링 방법은 클러스터링 품질을 향상시키기 위해, 다음과 같은 2단계의 단계를 포함할 수 있다.
- [0124] (1단계) 첫 번째로는 센서 데이터 특징 공간에 K-평균 클러스터링 알고리즘을 적용함으로써 초기 클러스터 중심을 찾는 것이다.
- [0125] (2단계) 두 번째로는 더 많은 효과를 제공할 때까지 전체 분류기 네트워크에 미세 조정 기법(DEC)을 반복적으로 적용하여 정확도를 높이는 것이다.
- [0126] 미세 조정 기법(DEC)은 특징 공간  $Z$ 에서 사전에 훈련된  $k$  클러스터 중심 세트  $\{\mu_j \in Z\}_{j=1}^k$  와,  $Z$ 에 입력을 위한 인코드 데이터에 대한 STAE의 가중치  $\theta$  를 동시에 조정함으로써 데이터를 클러스터 할 수 있다.
- [0127] 중심  $\mu_j$  와 인코더  $\theta$  의 가중치는 미세 조정 기법(DEC) 과정 동안 파라미터가 최적의 클러스터링 결과를 보여줄 때까지 반복적으로 계산될 수 있다. Deep embedded clustering(DEC)의 첫 번째 단계로서, K-평균 클러스터링 알고리즘을 이용하여 원래의 분포  $Q$ 를 구하고  $Q$ 의 클러스터링 결과가 보다 잘 구별되게 하는 목표 분포  $P$ 의 소프트 할당(soft assignment)을 계산할 수 있다. 여기에서 목표 분포  $P$ 의 클러스터간 분산을 기존 분포  $Q$ 의 클러스터간 분산보다 크게 하도록 스튜던트 T분포(Student's T-distribution)를 소프트 할당의 커널 함수로써 사용할 수 있다. 그런 다음, KL 분산을 이용하여 상기 두 분포의 간의 차이를 구할 수 있다. 여기서 KL 분산은 보통 상기 두 분포 간의 차이를 구하기 위해 사용될 수 있다. 상기 계산된 차이는 네트워크 훈련에 대한 Deep embedded clustering(DEC)의 손실함수로 사용될 수 있다. 이는 손실을 역전파(backpropagation)하고,  $\theta$  및  $\mu$  을 업데이트함으로써 분류기의 성능을 향상시키기 위한 것이라 할 수 있다.
- [0128] 이러한 과정은  $Q$  의 하드 할당(hard assignment)이 기 설정된 임계값 내에서 변하지 않을 때까지 반복될 수 있다. 미세 조정 기법 과정 이후, 클러스터의 인코더 가중치와 중심 좌표는 입력 이미지 스트림의 분류에 대해 최적화될 수 있다.
- [0129] 도 11은 발명의 일 실시예로서, 미세 조정 기법(DEC)을 적용하여 클러스터링 한 결과를 나타낸 도면이다.
- [0130] 보다 상세하게, 도 11에서 (a)은 81,758개의 훈련 데이터 셋 이미지에 대한 클러스터링 결과를 나타낸 도면이고, (b)은 사용자(on-line validation)로부터 실시간으로 감지한 클러스터링 결과를 나타낸 도면이다.
- [0131] 도 11에 도시된 바와 같이, 분류기 네트워크는 클러스터링 프로세스의 성능을 확인하기 위해 81,758개의 훈련 세트 입력 이미지의 시퀀스를 클러스터링했다. 클러스터링 결과는 t-SNE 방법을 사용하여 2D 특징 공간으로 나타낼 수 있다. 점의 색상과 모양은 뉴트럴(neutral) 또는 링크 상태의 각 클러스터를 나타내며, 클러스터의 이름은 산점도(scatter plots)의 범례에 나타낼 수 있다. 미세 조정 기법(DEC) 적용 전에 획득한 결과와 비교하여, 미세 조정 기법(DEC) 적용 후의 결과는 학습이 시작됨에 따라 클러스터 경계의 분해능이 크게 향상되었다.
- [0132] 다만, 본 발명에서는 미세 조정 기법(DEC) 기반 분류기와 특징 추출에 있어서 주석되지 않은 데이터를 사용했기 때문에, 클러스터링 네트워크의 정확성을 판단할 수 없다. 본 발명은 분류기 네트워크에 비지도 학습 방법을 적용했기 때문에, 사용자 상태가 링크인지 또는 뉴트럴(neutral)인지 나타내는 정보는 없었다.
- [0133] 본 발명에서는 일 실시예로서 분류기 네트워크의 정확성을 평가하기 위해, AR 헤드셋을 착용한 사용자들을 대상으로 추가적인 실험(두 번째 실험 데이터 셋에 관한 실험)을 수행할 수 있다.
- [0134] 이 때 검증 및 테스트를 위한 데이터 셋은 수집하지 않았기 때문에, 개인의 온라인 센서 데이터를 검증 데이터 셋으로 활용할 수 있다. 실험 유효성 검사 단계에서는 학습의 횟수에 따라 네트워크에 훈련된 가중치를 적용한 후, 분류 정확도를 측정할 수 있다.
- [0135] 도 11(b)은 추가 실험을 통해 획득된 온라인 센서 데이터에 대한 실시간 클러스터링 결과를 보여주는 도면이다.

각각의 실험 참가자는 각 측정동안 30번의 윙크 제스처를 취했으며, 본 발명에서는 이러한 의도적인 윙크 제스처의 성공적인 탐지 횟수를 기록할 수 있다. 도 11(b)에 도시된 바와 같이, 학습 단계가 증가함에 따라, 클러스터링 네트워크의 성능이 향상된 것을 확인할 수 있다. 미세 조정 기법(DEC) 학습을 500회 반복한 후, 분류기 네트워크는 윙크 제스처 검출에 있어서 30/30의 정확도를 달성할 수 있다.

[0136] 본 발명에서의 제스처 감지 시스템은 30fps의 속도로 실시간 모드로 운영될 수 있다. 이러한 분류기 네트워크는 데이터 수집에 있어서 자원하여 참여하지 않은 다른 사람들에게도 적용될 수 있는지를 테스트하기 위해, 본 발명에서는 맞춤형 응용 프로그램을 사용한 시연으로부터 스크린샷에 나타난 세 번째 추가 실험을 수행할 수 있다.

[0137] 사용자는 배경 (c,d)을 변경하기 위해 풍선 (a,b)을 터뜨리거나 버튼을 선택할 수 있다. 사용자는 특정 물체에 빨간색 중앙 점을 목표로 물체를 선택한 다음, 윙크 제스처를 실행할 수 있다.

[0138] 세 번째 실험에서는 참가자들이 얼굴 동작에 대한 반응으로 제스처 검출 결과를 실시간으로 확인할 수 있었고, 윙크 제스처가 센서에 의해 검출될 때 1점을 받을 수 있었다. 각각의 참가자는 100점을 받아야 실험을 완료할 수 있었다. 본 발명에서는 각 실험을 완료할 때까지 각 참가자가 의도한 윙크 제스처의 수를 카운트할 수 있다. 이러한 실험은 모두 새로운 참가자(즉, 훈련 데이터 셋 데이터 수집에 참여하지 않은 사람) 10명을 대상으로 수행되었다. 이는 아래 표 1와 같이 95.4%의 평균 정확도를 달성할 수 있다. 표 1은 본 발명의 AR 헤드셋을 착용한 사용자의 제스처 감지 정확도에 대한 실시간 측정 결과를 나타낸 표라 할 수 있다.

[0139] <표 1>

Participant IDs	1	2	3	4	5	6	7	8	9	10
The Number of detected gestures	100	100	100	100	100	100	100	100	100	100
The Number of intended gestures	104	110	102	107	111	101	100	103	101	109
Accuracy of Gesture-recognition (%)	96.2	91.0	98.0	93.5	90.1	99.0	100	97.1	99.0	91.7

[0140]

[0141] 이러한 본 발명의 인터페이스 기술은 IR 카메라를 입력 장치로 활용하였으나, 일반적인 이미지 기반 표정 분류와는 완전히 다르다고 할 수 있다. 본 발명에서의 센서는 IR 확산 반사체를 기반으로 피부 내 콜라겐 섬유의 왜곡된 얼라이먼트(alignment)를 추적하여 피부 표면 아래 근육의 움직임을 감지할 수 있다.

[0142] 특히 본 발명에서는 피부와의 접촉 없이 ROI의 피부 변형을 감지하기 때문에, 얼굴 전체 이미지가 필요하지 않다는 특징이 있다. 일반적으로 인간은 서로 다른 안면 근육을 사용하여 눈을 깜박이고 의도적으로 윙크를 하기 때문에, 본 발명에서의 SRDR 센서는 상기 두 가지 동작을 구별할 수 있는 장점을 가지고 있다. 또한, 이러한 센싱 방법의 장점은 본 발명에서의 센서가 카메라로 어떤 특징도 찾을 수 없는 평범한 얼굴 피부 표면으로부터 분류 특징을 생성할 수 있다는 점이다.

[0143] 본 발명에서는 맞춤형 얼굴 제스처 감지 센서와 분류기 네트워크를 통해 AR 헤드셋을 착용한 사용자의 윙크로 생성된 IR 확산 이미지의 클러스터링을 달성할 수 있다. 즉, 본 발명에서의 센서는 AR 헤드셋용 UI를 구현하는 효율적인 방법을 제공할 수 있다.

[0144] 많은 명령어는 얼굴 제스처와 머리 회전 데이터를 조합하여 적용할 수 있는데, 후자는 대부분의 AR 헤드셋에 통합된 자이로 센서에 의해 캡처될 수 있다. 예를 들어, 사용자는 윙크를 통해 버튼을 선택하거나, 데스크톱 환경에서 흔히 사용되는 끌어서 놓기(drag-and-drop) 명령과 유사한 윙크 제스처로 머리를 돌려 가상공간에 있는 일부 객체를 움직일 수 있다.

[0145] 도 12는 본 발명의 일 실시예로서, 맞춤형 응용 어플리케이션을 사용하여 시연한 스크린샷을 보여주는 도면이다. 보다 상세하게, 도 12에서 사용자는 배경 (c,d)을 변경하기 위해, 풍선 (a,b)을 터뜨리거나 버튼을 선택할 수 있으며, 특정 물체에 빨간색 중심점을 목표로 물체를 선택한 후 윙크 동작으로 이를 실행할 수 있다.

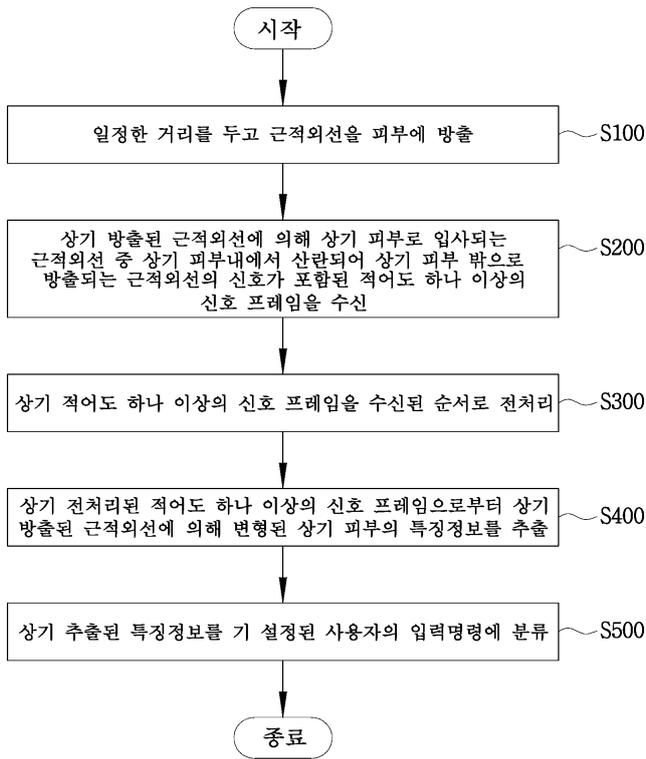
[0146] 도 12를 참조하면, 사용자는 윙크 제스처를 통해 AR 환경에서 풍선을 터뜨리거나 버튼을 선택할 수 있다. 이는 간단한 핸드프리 제어와 저비용 구현의 장점을 가진 사용자 친화적인 UI를 제공할 수 있다.

[0147] 본 발명에서의 센서는 적절한 위치가 필요하며, 얼굴 피부 변형을 감지하기 위해 인체 내 특정 위치로 제한될 수 있다. IR SRDR의 변화량은 피부 표면 아래의 근육이 충분히 움직여 피부의 내부 구조에 영향을 미치는 경우로도 충분하기 때문이다.

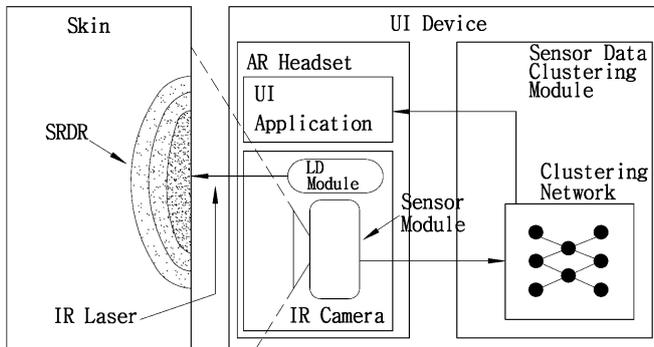
[0148] 최상의 결과를 얻기 위해서는, 센서의 적절한 ROI을 설정하는 것이 중요할 수 있다. 센서를 부적절하게 배치할



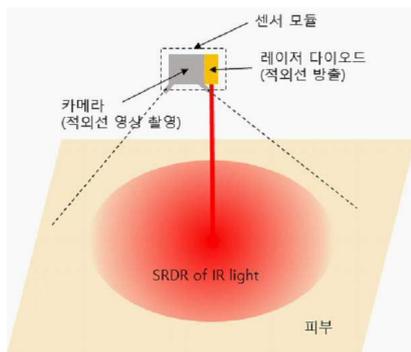
도면2



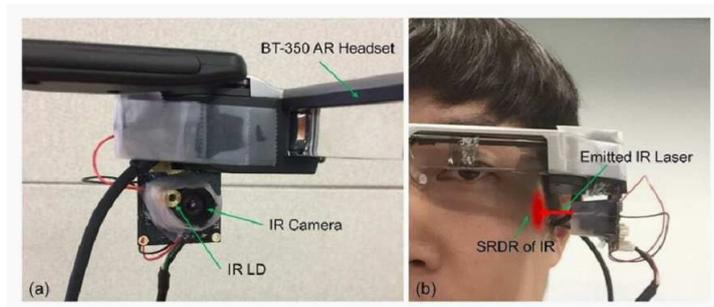
도면3



도면4



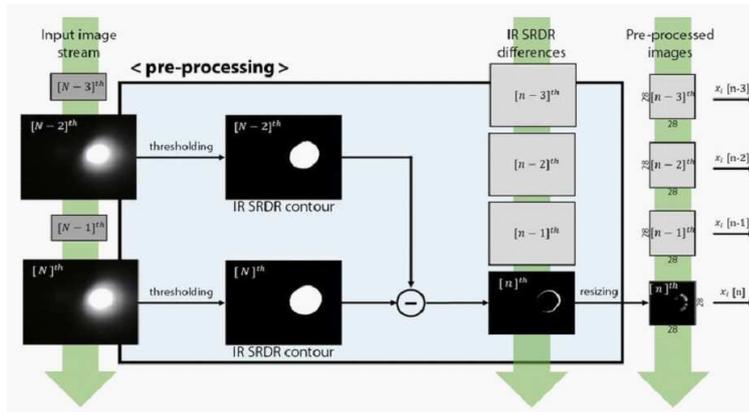
도면5



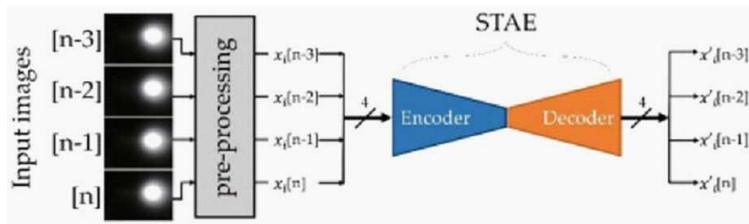
도면6



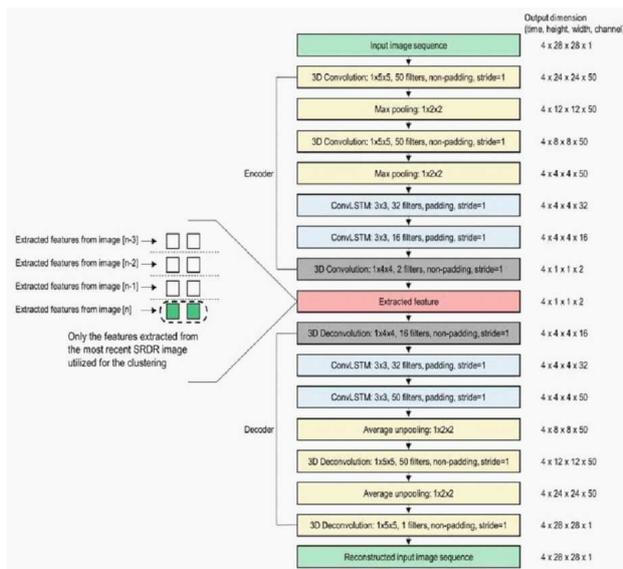
도면7



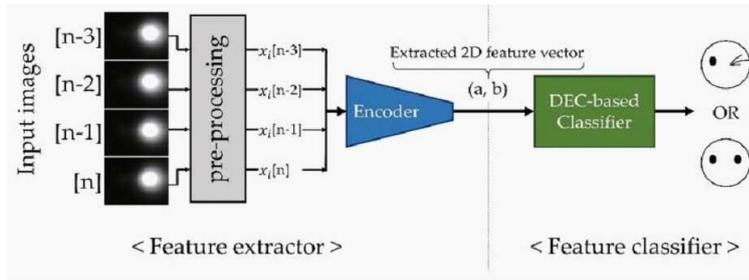
도면8



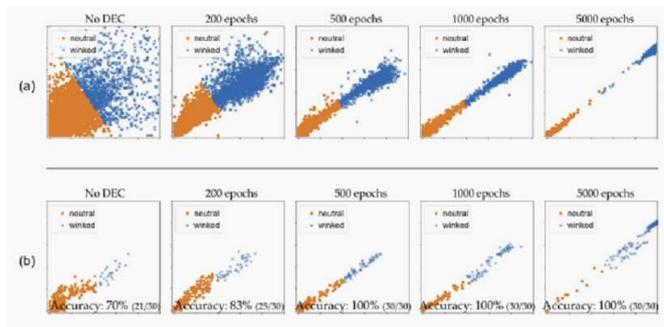
도면9



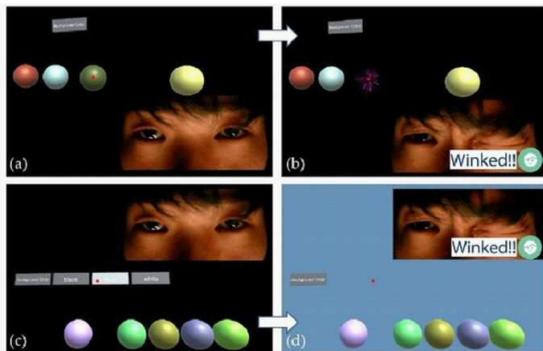
도면10



도면11



도면12



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 제8항

【변경전】

상기 특징 추출단계의 데이터 입력 형태에 따라

**【변경후】**

상기 특징정보 추출단계의 데이터 입력 형태에 따라

**【식권보정 2】**

**【보정항목】** 청구범위

**【보정세부항목】** 청구항 제1항

**【변경전】**

상기 특징 추출부의 데이터 입력 형태에 따라

**【변경후】**

상기 특징정보 추출부의 데이터 입력 형태에 따라