(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0144383
(43) 공개일자 2023년10월16일

(51) 국제특허분류(Int. Cl.)
G06V 40/20 (2022.01) *G06N 5/04* (2023.01)
G06V 10/25 (2022.01) *G06V 10/764* (2022.01)
G06V 10/774 (2022.01)
(52) CPC특허분류
G06V 40/28 (2022.01)
G06N 5/04 (2023.01)
(21) 출원번호 10-2022-0043575
(22) 출원일자 2022년04월07일
심사청구일자 2022년04월07일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
김시호
인천광역시 연수구 송도과학로 85, Faculty House D동 912호(송도동)
김진혁
인천광역시 미추홀구 주안로 27, 4동 204호(도화동, 동원아파트)
(74) 대리인
특허법인우인

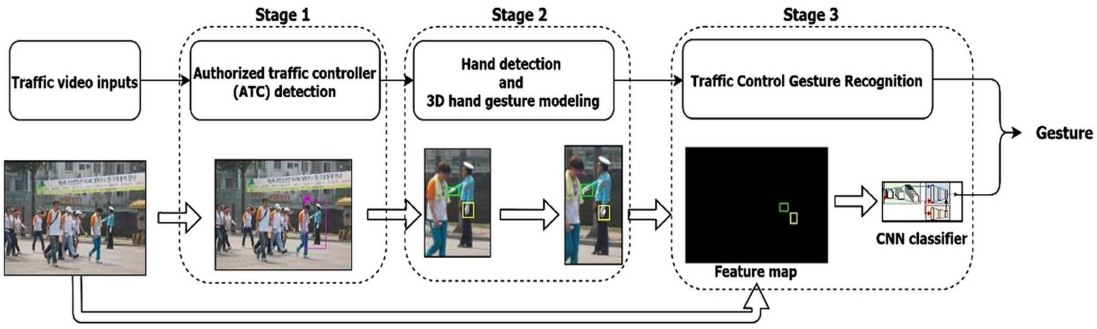
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 **교통 관제사의 제스처 인식 방법 및 그 방법을 구현하기 위한 장치**

(57) 요 약

본 실시예들은 복수의 프레임으로부터 교통 관제사 영역을 추출하고, 교통 관제사 영역으로부터 손 영역을 추출하고 손 모델링 데이터를 생성하고, 좌우 구분 가능한 손 모델링 데이터를 이용하여 복수의 프레임으로부터 교통 수신호 유형을 인식하는 제스처 인식 방법 및 그 방법을 구현하기 위한 장치, 장치가 탑재된 이동체를 제공한다.

대 표 도



(52) CPC특허분류
G06V 10/25 (2023.08)
G06V 10/764 (2023.08)
G06V 10/774 (2023.08)
G06T 2207/20044 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711144648
과제번호	2019H1D3A1A01071115
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	교육인력양성사업
연구과제명	자율주행 자동차의 인공지능 기반 비정규 상황인식과 대응 시스템(3/5)
기 여 율	1/1
과제수행기관명	연세대학교
연구기간	2021.01.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

제스처 인식 방법에 있어서,

대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하는 단계;

손 인식 및 모델링 네트워크를 통해 상기 교통 관제사 영역으로부터 손 영역을 추출하고, 상기 손 영역으로부터 손 모델링 데이터를 생성하는 단계;

수신호 분류 네트워크를 통해 상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 교통 수신호를 분류하는 단계를 포함하는 제스처 인식 방법.

청구항 2

제1항에 있어서,

상기 교통 수신호는,

(i) 전방 정지 유형, (ii) 좌우측방 동시정지 유형, (iii) 좌측에서 우측으로 진행 유형, (iv) 전방에서 좌측으로 회전 유형, (v) 우측에서 좌측으로 진행 유형, (vi) 전방에서 우측으로 회전 유형을 포함하는 것을 특징으로 하는 제스처 인식 방법.

청구항 3

제1항에 있어서,

상기 손 모델링 데이터를 생성하는 단계는,

상기 손 모델링 데이터의 골격 포인트를 이용하여 왼손과 오른손을 구분하고, 손바닥과 손등을 구분하는 것을 특징으로 하는 제스처 인식 방법.

청구항 4

제1항에 있어서,

상기 손 모델링 네트워크의 손 모델링 손실 함수는 손 포즈 손실 함수, 3차원 키포인트 손실 함수, 2차원 키포인트 손실 함수, 정규화 손실 함수에 각각 가중치가 적용된 합으로 정의되며, 상기 손 모델링 손실 함수가 최소화되도록 학습되는 것을 특징으로 하는 제스처 인식 방법.

청구항 5

제1항에 있어서,

상기 수신호 분류 네트워크는 특징 추출 네트워크, 상기 특징 추출 네트워크에 연결된 대조 임베딩 네트워크, 상기 특징 추출 네트워크에 연결된 단일 추론 네트워크를 포함하며, 상기 대조 임베딩 네트워크를 이용하여 제스처 변화가 기준치보다 큰 유형에 대해서 상기 교통 수신호를 분류하고, 상기 단일 추론 네트워크를 이용하여 상기 제스처 변화가 상기 기준치보다 작은 유형에 대해서 상기 교통 수신호를 분류하는 것을 특징으로 하는 제스처 인식 방법.

청구항 6

제5항에 있어서,

상기 특징 추출 네트워크는 상기 특징 추출 네트워크에 추가로 연결된 특징 디코딩 네트워크를 적용하여 오토인코더 구조로 학습되는 것을 특징으로 하는 제스처 인식 방법.

청구항 7

제5항에 있어서,

상기 교통 수신호를 분류하는 단계는,

상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 특징 추출 네트워크를 통해 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터를 추출하는 것을 특징으로 하는 제스처 인식 방법.

청구항 8

제7항에 있어서,

상기 교통 수신호를 분류하는 단계는,

상기 대조 임베딩 네트워크를 통해 상기 제1 특징 벡터 및 상기 제2 특징 벡터의 차이로부터 상기 교통 수신호를 분류하는 것을 특징으로 하는 제스처 인식 방법.

청구항 9

제7항에 있어서,

상기 교통 수신호를 분류하는 단계는,

상기 단일 추론 네트워크를 통해 상기 제1 특징 벡터로부터 상기 교통 수신호를 분류하는 것을 특징으로 하는 제스처 인식 방법.

청구항 10

제7항에 있어서,

상기 대조 임베딩 네트워크의 대조 임베딩 손실 함수는 (i) 상기 제1 특징 벡터 및 상기 제2 특징 벡터의 차이, (ii) 임베딩 공간의 마진, (iii) 상기 포즈 프레임 및 상기 이전 프레임의 유사도를 기반으로 정의되며, 상기 대조 임베딩 손실 함수가 최소화되도록 학습되는 것을 특징으로 하는 제스처 인식 방법.

청구항 11

제10항에 있어서,

상기 교통 수신호를 분류하는 단계는,

상기 제1 특징 벡터 및 상기 제2 특징 벡터 중에서 이미지 쌍의 유사하지 않은 특징 벡터를 상기 마진의 범위 내에 상기 특징 벡터의 거리가 존재하면 상기 대조 임베딩 손실 함수에서 적용하는 것을 특징으로 하는 제스처 인식 방법.

청구항 12

프로세서를 포함하는 제스처 인식 장치에 있어서,

상기 프로세서는,

대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하고,

손 인식 및 모델링 네트워크를 통해 상기 교통 관제사 영역으로부터 손 영역을 추출하고, 상기 손 영역으로부터 손 모델링 데이터를 생성하고,

수신호 분류 네트워크를 통해 상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 교통 수신호를 분류하는 것을 특징으로 하는 제스처 인식 장치.

청구항 13

제12항에 있어서,

상기 손 모델링 네트워크의 손 모델링 손실 함수는 손 포즈 손실 함수, 3차원 키포인트 손실 함수, 2차원 키포

인트 손실 함수, 정규화 손실 함수에 각각 가중치가 적용된 합으로 정의되며, 상기 손 모델링 손실 함수가 최소화되도록 학습되는 것을 특징으로 하는 제스처 인식 장치.

청구항 14

제12항에 있어서,

상기 수신호 분류 네트워크는 특징 추출 네트워크, 상기 특징 추출 네트워크에 연결된 대조 임베딩 네트워크, 상기 특징 추출 네트워크에 연결된 단일 추론 네트워크를 포함하며, 상기 대조 임베딩 네트워크를 이용하여 제스처 변화가 기준치보다 큰 유형에 대해서 상기 교통 수신호를 분류하고, 상기 단일 추론 네트워크를 이용하여 상기 제스처 변화가 상기 기준치보다 작은 유형에 대해서 상기 교통 수신호를 분류하는 것을 특징으로 하는 제스처 인식 장치.

청구항 15

제14항에 있어서,

상기 프로세서는,

상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 특징 추출 네트워크를 통해 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터를 추출하는 것을 특징으로 하는 제스처 인식 장치.

청구항 16

제15항에 있어서,

상기 프로세서는,

상기 대조 임베딩 네트워크를 통해 상기 제1 특징 벡터 및 상기 제2 특징 벡터의 차이로부터 상기 교통 수신호를 분류하는 것을 특징으로 하는 제스처 인식 장치.

청구항 17

제15항에 있어서,

상기 프로세서는,

상기 단일 추론 네트워크를 통해 상기 제1 특징 벡터로부터 상기 교통 수신호를 분류하는 것을 특징으로 하는 제스처 인식 장치.

청구항 18

제15항에 있어서,

상기 대조 임베딩 네트워크의 대조 임베딩 손실 함수는 (i) 상기 제1 특징 벡터 및 상기 제2 특징 벡터의 차이, (ii) 임베딩 공간의 마진, (iii) 상기 포즈 프레임 및 상기 이전 프레임의 유사도를 기반으로 정의되며, 상기 대조 임베딩 손실 함수가 최소화되도록 학습되는 것을 특징으로 하는 제스처 인식 장치.

청구항 19

이동체에 있어서,

상기 제스처 인식 장치; 및

상기 제스처 인식 장치의 출력 결과를 기반으로 이동 계획을 갱신하고 이동을 수행하는 주행 수단을 포함하며,

상기 제스처 인식 장치는,

대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하고,

손 인식 및 모델링 네트워크를 통해 상기 교통 관제사 영역으로부터 손 영역을 추출하고, 상기 손 영역으로부터 손 모델링 데이터를 생성하고,

수신호 분류 네트워크를 통해 상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 교통 수신호를 분류하는

것을 특징으로 하는 이동체.

발명의 설명

기 술 분 야

[0001] 본 발명이 속하는 기술 분야는 교통 수신호를 인식하는 방법 및 그 방법을 구현하는 장치에 관한 것이다.

배 경 기 술

[0002] 이 부분에 기술된 내용은 단순히 본 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 도 1에 도시된 교통 관제사를 참조하면, 교통 관제사는 사람의 안전을 위해 도로에서 차량과 교통의 흐름을 제어하는 교통 수신호를 지시한다. 인가된 교통 관제사로는 교통경찰, 민간교통관제사, 모범운전자(국내만 해당), 건설근로자, 헌병 등이 있다.

[0004] 운전자는 상황 인식 능력을 갖고, 예상치 못한 불규칙한 상황에서 교통 관제사의 지시를 따른다. 자율 주행 시스템은 상황 인식 교통 제스처 인식에 대해서 인간 수준의 이해도 포함되어야 한다. 레벨 3 이상의 자율 주행차는 원활한 운송을 위해 교통 수신호에 대한 이해가 필요하다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국등록특허공보 제10-1494959호 (2015.02.12)

(특허문헌 0002) 한국등록특허공보 제10-0883521호 (2009.02.06)

발명의 내용

해결하려는 과제

[0006] 본 발명의 실시예들은 복수의 프레임으로부터 교통 관제사 영역을 추출하고, 교통 관제사 영역으로부터 손 영역을 추출하고 손 모델링 데이터를 생성하고, 좌우 구분 가능한 손 모델링 데이터를 이용하여 복수의 프레임으로부터 교통 수신호 유형을 인식하는데 주된 목적이 있다.

[0007] 본 발명의 명시되지 않은 또 다른 목적들은 하기의 상세한 설명 및 그 효과로부터 용이하게 추론할 수 있는 범위 내에서 추가적으로 고려될 수 있다.

과제의 해결 수단

[0008] 본 실시예의 일 측면에 의하면 제스처 인식 방법에 있어서, 대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하는 단계; 손 인식 및 모델링 네트워크를 통해 상기 교통 관제사 영역으로부터 손 영역을 추출하고, 상기 손 영역으로부터 손 모델링 데이터를 생성하는 단계; 수신호 분류 네트워크를 통해 상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 교통 수신호를 분류하는 단계를 포함하는 제스처 인식 방법을 제공한다.

[0009] 상기 교통 수신호는, (i) 전방 정지 유형, (ii) 좌우측방 동시정지 유형, (iii) 좌측에서 우측으로 진행 유형, (iv) 전방에서 좌측으로 회전 유형, (v) 우측에서 좌측으로 진행 유형, (vi) 전방에서 우측으로 회전 유형을 포함할 수 있다.

[0010] 상기 손 모델링 데이터를 생성하는 단계는, 상기 손 모델링 데이터의 골격 포인트를 이용하여 왼손과 오른손을 구분하고, 손바닥과 손등을 구분할 수 있다.

[0011] 상기 손 모델링 네트워크의 손 모델링 손실 함수는 손 포즈 손실 함수, 3차원 키포인트 손실 함수, 2차원 키포인트 손실 함수, 정규화 손실 함수에 각각 가중치가 적용된 합으로 정의되며, 상기 손 모델링 손실 함수가 최소화되도록 학습될 수 있다.

[0012] 상기 수신호 분류 네트워크는 특징 추출 네트워크, 상기 특징 추출 네트워크에 연결된 대조 임베딩 네트워크,

상기 특징 추출 네트워크에 연결된 단일 추론 네트워크를 포함하며, 상기 대조 임베딩 네트워크를 이용하여 제스처 변화가 기준치보다 큰 유형에 대해서 상기 교통 수신호를 분류하고, 상기 단일 추론 네트워크를 이용하여 상기 제스처 변화가 상기 기준치보다 작은 유형에 대해서 상기 교통 수신호를 분류할 수 있다.

[0013] 상기 특징 추출 네트워크는 상기 특징 추출 네트워크에 추가로 연결된 특징 디코딩 네트워크를 적용하여 오토인코더 구조로 학습될 수 있다.

[0014] 상기 교통 수신호를 분류하는 단계는, 상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 특징 추출 네트워크를 통해 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터를 추출할 수 있다.

[0015] 상기 교통 수신호를 분류하는 단계는, 상기 대조 임베딩 네트워크를 통해 상기 제1 특징 벡터 및 상기 제2 특징 벡터의 차이로부터 상기 교통 수신호를 분류할 수 있다.

[0016] 상기 교통 수신호를 분류하는 단계는, 상기 단일 추론 네트워크를 통해 상기 제1 특징 벡터로부터 상기 교통 수신호를 분류할 수 있다.

[0017] 상기 대조 임베딩 네트워크의 대조 임베딩 손실 함수는 (i) 상기 제1 특징 벡터 및 상기 제2 특징 벡터의 차이, (ii) 임베딩 공간의 마진, (iii) 상기 포즈 프레임 및 상기 이전 프레임의 유사도를 기반으로 정의되며, 상기 대조 임베딩 손실 함수가 최소화되도록 학습될 수 있다.

[0018] 상기 교통 수신호를 분류하는 단계는, 상기 제1 특징 벡터 및 상기 제2 특징 벡터 중에서 이미지 쌍의 유사하지 않은 특징 벡터를 상기 마진의 범위 내에 상기 특징 벡터의 거리가 존재하면 상기 대조 임베딩 손실 함수에서 적용할 수 있다.

[0019] 본 실시예의 다른 측면에 의하면 프로세서를 포함하는 제스처 인식 장치에 있어서, 상기 프로세서는, 대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하고, 손 인식 및 모델링 네트워크를 통해 상기 교통 관제사 영역으로부터 손 영역을 추출하고, 상기 손 영역으로부터 손 모델링 데이터를 생성하고, 수신호 분류 네트워크를 통해 상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 교통 수신호를 분류하는 것을 특징으로 하는 제스처 인식 장치를 제공한다.

[0020] 본 실시예의 또 다른 측면에 의하면 이동체에 있어서, 상기 제스처 인식 장치; 및 상기 제스처 인식 장치의 출력 결과를 기반으로 이동 계획을 갱신하고 이동을 수행하는 주행 수단을 포함하며, 상기 제스처 인식 장치는, 대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하고, 손 인식 및 모델링 네트워크를 통해 상기 교통 관제사 영역으로부터 손 영역을 추출하고, 상기 손 영역으로부터 손 모델링 데이터를 생성하고, 수신호 분류 네트워크를 통해 상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 교통 수신호를 분류하는 것을 특징으로 하는 이동체를 제공한다.

발명의 효과

[0021] 이상에서 설명한 바와 같이 본 발명의 실시예들에 의하면, 복수의 프레임으로부터 교통 관제사 영역을 추출하고, 교통 관제사 영역으로부터 손 영역을 추출하고 좌우 구분 가능한 손 모델링 데이터를 생성하고, 손 모델링 데이터를 기반으로 복수의 프레임에 담긴 제스처 변화가 큰 제스처 유형에 대해서 대조 임베딩 네트워크를 통해 교통 수신호를 분류하고, 제스처 변화가 작은 제스처 유형에 대해서 단일 추론 네트워크를 통해 교통 수신호를 분류할 수 있는 효과가 있다.

[0022] 여기에서 명시적으로 언급되지 않은 효과라 하더라도, 본 발명의 기술적 특징에 의해 기대되는 이하의 명세서에서 기재된 효과 및 그 잠정적인 효과는 본 발명의 명세서에 기재된 것과 같이 취급된다.

도면의 간단한 설명

[0023] 도 1은 교통 관제사를 예시한 도면이다.

도 2는 실제 교통 관제 상황에서 발생할 수 있는 시나리오를 예시한 도면이다.

도 3은 교통 수신호를 예시한 도면이다.

도 4는 본 발명의 일 실시예에 따른 제스처 인식 방법을 예시한 흐름도이다.

도 5는 본 발명의 일 실시예에 따른 제스처 인식 방법의 전체 동작을 예시한 도면이다.

도 6은 본 발명의 일 실시예에 따른 제스처 인식 방법의 손 모델링 네트워크의 동작을 예시한 도면이다.

도 7은 본 발명의 일 실시예에 따른 제스처 인식 방법의 손 모델링 네트워크가 처리하는 손 모델링 데이터의 골격 포인트를 예시한 도면이다.

도 8은 본 발명의 일 실시예에 따른 제스처 인식 방법의 손 모델링 네트워크가 처리하는 손 모델링 데이터를 예시한 도면이다.

도 9는 본 발명의 일 실시예에 따른 제스처 인식 방법의 수신호 분류 네트워크를 예시한 도면이다.

도 10은 본 발명의 일 실시예에 따른 제스처 인식 방법의 수신호 분류 네트워크의 학습 동작을 예시한 도면이다.

도 11은 본 발명의 일 실시예에 따른 제스처 인식 방법의 수신호 분류 네트워크의 인식 동작을 예시한 도면이다.

도 12는 본 발명의 일 실시예에 따른 제스처 인식 방법의 인식률을 예시한 도면이다.

도 13 및 도 14는 본 발명의 일 실시예에 따른 제스처 인식 방법의 프레임별 제스처 인식여부를 예시한 도면이다.

도 15는 본 발명의 다른 실시예에 따른 제스처 인식 장치를 예시한 블록도이다.

도 16은 본 발명의 다른 실시예에 따른 이동체를 예시한 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0024] 이하, 본 발명을 설명함에 있어서 관련된 공지기능에 대하여 이 분야의 기술자에게 자명한 사항으로서 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하고, 본 발명의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다.

[0025] 도 2는 실제 교통 관제 상황에서 발생할 수 있는 시나리오를 예시한 도면이다.

[0026] 도 2를 참조하면, 도로에서 패인 영역을 우회시키는 교통 관제사를 확인할 수 있다. 교통 관제사는 정지 수신호를 지시하고 있으나 교통 신호등이 녹색인 상황으로 서로 다른 메시지를 나타낸다. 교통 관제사의 수신호를 인지하지 못하는 일반적인 자율 주행차는 교통 신호등을 기준으로 교차로를 통과하는게 맞다고 판단하고 직진 통과할 수 있다.

[0027] 수신호를 인지할 수 있더라도 교차로 주변을 살펴보면, 자전거 탑승자도 수신호로 알리고, 횡단보도 근처에 위치하는 보행자는 손 제스처를 취하고 있고, 다른 보행자는 핸드폰을 조작하는 손 제스처를 취하는 상황이다. 이와 같이 도로에서 발생 가능한 다양한 손 제스처 중에서 자율 주행차는 주행 판단의 기준이 되는 교통 관제사의 수신호를 정확하게 인식할 필요가 있다.

[0028] 자율 주행차뿐만 아니라 일반 자동차의 주행 보조 장치 등에서도 주변 영상을 획득 및 분석하여 수신호의 의미를 판단할 필요가 있다.

[0029] 도 3은 교통 수신호를 예시한 도면이다.

[0030] 도 3을 참조하면 교통 수신호는 (i) 전방 정지 유형, (ii) 좌우측방 동시정지 유형, (iii) 좌측에서 우측으로 진행 유형, (iv) 전방에서 좌측으로 회전 유형, (v) 우측에서 좌측으로 진행 유형, (vi) 전방에서 우측으로 회전 유형을 포함할 수 있다.

[0031] 본 실시예에 따른 제스처 인식 방법은 표 1과 같이 알고리즘으로 구현되어 동작할 수 있다.

표 1

Algorithm 1: ATC Hand Gesture Recognition

Functions: $\mathcal{F}$ = object detector; $\mathcal{H}$ = hand detector; $\mathcal{M}$ = 3D hand model; $\mathcal{E}$ = feature extractor; $\mathcal{C}$ = contrastive embedding; $\mathfrak{T}$ = single inference.

Input: *videoframes* (*onboard device*)

$A = \mathcal{F}(Input)$ // ATC detection

for each hand in *f1* & *f2*:

$H = \mathcal{H}(A)$ // Hand detection

$M = \mathcal{M}(H)$ // 3D hand modeling

$E = \mathcal{E}(M)$ // Hand feature extraction

$E = \{E_{f1}, E_{f2}\}$ // Stores hand features (of E_{f1}, and E_{f2})

$C = \mathcal{C}(E)$ // Contrastive embedding calculation for E_{f1}, and E_{f2}

$I = \mathfrak{T}(E_{f2})$ // Inference for E_{f2}

$$G = \begin{cases} C + I, \; if \; poses \; have \; difference \\ I, \; if \; poses \; have \; no \; difference \end{cases}$$

return (G)

[0032]

[0033] 제스처 인식 방법은 교통 관제사(Authorized Traffic Controller, ATC)를 영상에서 탐지하고, 교통 관제사의 6가지 교통 수신호(Traffic Gesture)를 인식한다. YOLO 기반의 교통 관제사 인식 네트워크 구축하고, 교통 관제사의 손 감지 및 3D 손 모델링을 수행하고, 손의 위치 및 방향, 움직임을 바탕으로 교통 수신호 인식하고, 이러한 과정을 통해 영상에서 교통 관제사를 인식하고, 그의 교통 수신호를 정확하게 검출할 수 있다.

[0034] 도 4는 본 발명의 일 실시예에 따른 제스처 인식 방법을 예시한 흐름도이다. 제스처 인식 방법은 제스처 인식 장치에 의해 수행될 수 있다.

[0035] 단계 S10에서는 대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하는 단계를 수행할 수 있다.

[0036] 단계 S20에서는 손 인식 및 모델링 네트워크를 통해 교통 관제사 영역으로부터 손 영역을 추출하고, 손 영역으로부터 손 모델링 데이터를 생성하는 단계를 수행할 수 있다.

[0037] 손 모델링 데이터를 생성하는 단계(S20)는 손 모델링 데이터의 골격 포인트를 이용하여 왼손과 오른손을 구분하고, 손바닥과 손등을 구분할 수 있다.

[0038] 제스처 인식 방법이 사용하는 손 모델링 네트워크의 손 모델링 손실 함수는 손 포즈 손실 함수, 3차원 키포인트 손실 함수, 2차원 키포인트 손실 함수, 정규화 손실 함수에 각각 가중치가 적용된 합으로 정의되며, 손 모델링 손실 함수가 최소화되도록 학습된다.

[0039] 단계 S30에서는 수신호 분류 네트워크를 통해 손 모델링 데이터 및 복수의 프레임으로부터 교통 수신호를 분류하는 단계를 수행할 수 있다.

[0040] 제스처 인식 방법이 인식하는 교통 수신호는 (i) 전방 정지 유형, (ii) 좌우측방 동시정지 유형, (iii) 좌측에서 우측으로 진행 유형, (iv) 전방에서 좌측으로 회전 유형, (v) 우측에서 좌측으로 진행 유형, (vi) 전방에서 우측으로 회전 유형을 포함할 수 있다.

[0041] 제스처 인식 방법이 사용하는 수신호 분류 네트워크는 특징 추출 네트워크, 특징 추출 네트워크에 연결된 대조 임베딩 네트워크, 특징 추출 네트워크에 연결된 단일 추론 네트워크를 포함할 수 있다. 특징 추출 네트워크는 특징 추출 네트워크에 추가로 연결된 특징 디코딩 네트워크를 적용하여 오토인코더 구조로 학습될 수 있다.

[0042] 제스처 인식 방법은 대조 임베딩 네트워크를 이용하여 제스처 변화가 기준치보다 큰 유형에 대해서 교통 수신호를 분류하고, 단일 추론 네트워크를 이용하여 제스처 변화가 기준치보다 작은 유형에 대해서 교통 수신호를 분류할 수 있다.

[0043] 교통 수신호를 분류하는 단계(S30)는 손 모델링 데이터 및 복수의 프레임으로부터 특징 추출 네트워크를 통해 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터를 추출할 수 있다.

[0044] 교통 수신호를 분류하는 단계(S30)는 단일 추론 네트워크를 통해 제1 특징 벡터로부터 교통 수신호를 분류할 수 있다.

[0045] 교통 수신호를 분류하는 단계(S30)는 대조 임베딩 네트워크를 통해 제1 특징 벡터 및 제2 특징 벡터의 차이로부터 교통 수신호를 분류할 수 있다.

[0046] 대조 임베딩 네트워크의 대조 임베딩 손실 함수는 (i) 제1 특징 벡터 및 제2 특징 벡터의 차이, (ii) 임베딩 공간의 마진, (iii) 포즈 프레임 및 이전 프레임의 유사도를 기반으로 정의되며, 대조 임베딩 손실 함수가 최소화되도록 학습될 수 있다.

[0047] 교통 수신호를 분류하는 단계(S30)는 제1 특징 벡터 및 제2 특징 벡터 중에서 이미지 쌍의 유사하지 않은 특징 벡터를 마진의 범위 내에 특징 벡터의 거리가 존재하면 대조 임베딩 손실 함수에서 적용할 수 있다.

[0048] 도 5는 본 발명의 일 실시예에 따른 제스처 인식 방법의 전체 동작을 예시한 도면이다.

[0049] 제스처 인식 방법은 교통 관제사(Authorized Traffic Controller, ATC)를 영상에서 탐지하고, 6가지 교통 수신호(Traffic Gesture)를 인식한다.

[0050] 먼저, 교통 관제사 인식 네트워크를 구축하여 교통 관제사를 검출한다. 교통 관제사의 손 감지 및 3차원 손 모델링을 수행한다. 손의 위치 및 방향, 움직임을 바탕으로 교통 수신호를 인식한다. 이러한 단계를 통해 영상에서 교통 관제사를 인식하고, 그의 교통 수신호를 검출할 수 있다.

[0051] 인식하고자 하는 교통 관제사는 교통경찰, 모범 운전수, 교통 마네킹 등이 있다.

[0052] 기존 사물 인식 네트워크(예컨대, YOLO 네트워크)를 재학습시켜 ATC 인식용 네트워크를 구축할 수 있다. YOLO 네트워크의 최종 출력단에서 바운더리 박스 위치 찾기와 클래스 분류가 동시에 이뤄진다. 출력 데이터가 바운더리 박스 위치와 클래스에 관한 두 개의 유형이다. 크기가 상이한 바운더리 박스를 여러 개 생성하고 객체가 있다고 판단되는 박스를 선별한다. 복수의 그리드로 분할된 영역에서 클래스를 분류한다. 바운더리 박스에 존재하는 객체의 클래스를 매칭한다.

[0053] 교통 관제사 인식을 성공한 경우, 다음 단계인 손 인식 및 3차원 모델링을 진행한다.

[0054] 도 6은 본 발명의 일 실시예에 따른 제스처 인식 방법의 손 모델링 네트워크의 동작을 예시한 도면이고, 도 7은 본 발명의 일 실시예에 따른 제스처 인식 방법의 손 모델링 네트워크가 처리하는 손 모델링 데이터의 골격 포인트를 예시한 도면이고, 도 8은 본 발명의 일 실시예에 따른 제스처 인식 방법의 손 모델링 네트워크가 처리하는 손 모델링 데이터를 예시한 도면이다.

[0055] 교통 관제사 이미지로부터 양 손의 위치를 인식하고, 이를 3차원 손 모델로 재구성하는 네트워크를 구축한다. 사람의 몸, 얼굴, 손을 인식하는 모수 모델(parametric model)로서 SMPL-X(Skinned Multi-Person Linear model eXpressive) 모델을 적용할 수 있다.

[0056] 교통 관제사는 대부분 흰 장갑을 착용하고 있으므로, SMPL-X 모델의 손 파트를 흰 장갑을 인식할 수 있도록 따로 분리하고 재구성한다.

[0057] 일반적인 손 인식 네트워크의 경우 장갑을 낀 손의 좌우를 구분하는데 어려움이 있다.

[0058] 구성된 3차원 손 모델링 네트워크는 교통 관제사 영역으로부터 왼손과 오른손을 추출하고, 손 관절(Hand Skeleton) 구성을 위한 특징점(Hand parameter)으로 변환한다. 변환한 손 파라미터를 바탕으로 손 관절, 즉 3차원 손 모델을 생성한다. 3차원 손 모델은 특징점(키포인트)을 바탕으로 구성되므로 장갑을 낀 손이어도 좌우를 구분할 수 있다.

[0059] 제스처 인식 장치는 손 모델링 데이터의 골격 포인트를 이용하여 왼손과 오른손을 구분하고, 손바닥과 손등을 구분할 수 있다.

[0060] 손 영역(I_h)은 손 인식 및 모델링 네트워크에 입력된다. 손 인식 네트워크와 손 모델링 네트워크는 하나의 손 인식 및 모델링 네트워크 모델로 구현될 수도 있고 분리된 모델로 각각 구현될 수 있다.

[0061] 제스처 인식 방법은 네트워크 모델을 적용한다. 네트워크 모델은 특징을 추출하고 특징을 데이터 가공

처리한다. 네트워크 모델은 다수의 레이어가 네트워크로 연결되며 히든 레이어를 포함한다. 레이어는 파라미터를 포함할 수 있고, 레이어의 파라미터는 학습가능한 필터 집합을 포함한다. 파라미터는 노드 간의 가중치 및/또는 바이어스를 포함한다. 모델은 손실 함수를 최소화하도록 파라미터를 학습한다.

[0062] 손 인식 및 모델링 네트워크(M_H)는 수학식 1과 같이 표현된다.

수학식 1

[0063] $$M_H(I_h) = [\varnothing_h, \theta_h, \beta_h, C_h]$$

[0064] 수학식 1에서 $\varnothing_h$는 손의 방향($\varnothing_h \in \mathbb{R}^3$)이고, θ_h는 손 포즈 파라미터($\theta_h \in \mathbb{R}^{3\times15}$)이고, β_h는 형상 파라미터($\beta_h \in \mathbb{R}^{10}$)이고, $C_h = (t_h, s_h)$는 스케일 요소($s_h \in \mathbb{R}$), 손 영상 평면의 2차원 변환($t_h \in \mathbb{R}^2$)으로 정의된 손 영역이다.

[0065] 손 인식 및 모델링 네트워크는 손 메시 정점($V_h \in \mathbb{R}^{778\times3}$)을 갖는다. 손의 3차원 조인트 회귀 함수는 $J_h^{3D} \in \mathbb{R}^{21\times3}$이고, 손의 회귀 행렬 R_h에 의해 정의된다. J_h^{3D}는 수학식 2와 같이 표현된다.

수학식 2

[0066] $$J_h^{3D} = R_h(V_h)$$

[0067] 수학식 2에서 i번째 3차원 손 조인트를 2차원 손 조인트로 변환하는 것은 스케일 요소 (s_h)를 이용한 2차원 변환(t_h)의 정투영(orthographic projection, Π)을 통해 획득된다.

수학식 3

[0068] $$J_{h,i}^{2D} = s_h \prod\left(J_{h,i}^{3D}\right) + t_h$$

[0069] 수학식 3에서 손 모델링 네트워크의 손 모델링 손실 함수는 손 포즈 손실 함수($\mathcal{L}_\theta$), 3차원 키포인트 손실 함수($\mathcal{L}_{3D}$), 2차원 키포인트 손실 함수($\mathcal{L}_{2D}$), 정규화 손실 함수($\mathcal{L}_R$)에 각각 가중치가 적용된 합으로 정의되며, 손 모델링 손실 함수가 최소화되도록 학습될 수 있다.

수학식 4

[0070]
$$\mathcal{L}_{Overall} = \omega_\theta \mathcal{L}_\theta + \omega_{2D}\mathcal{L}_{2D} + \omega_{3D}\mathcal{L}_{3D} + \omega_R \mathcal{L}_R$$
$$\mathcal{L}_\theta = \|\varnothing_h - \hat{\varnothing}_h\|_2^2$$
$$\mathcal{L}_{2D} = \|J_h^{2D} - \hat{J}_h^{2D}\|$$
$$\mathcal{L}_{3D} = \|J_h^{3D} - \hat{J}_h^{3D}\|_2^2$$
$$\mathcal{L}_R = \|\beta_h\|_2^2$$

[0071] 수학식 4에서 손실은 제곱 오차를 이용하여 획득되며, $\hat{\varnothing}_h$, $\hat{J}_h^{2D}$, $\hat{J}_h^{3D}$ 는 손 포즈, 2차원 키포인트, 3차원 키포인트의 검증자료(ground truth)를 나타낸다.

[0072] 도 9는 본 발명의 일 실시예에 따른 제스처 인식 방법의 수신호 분류 네트워크를 예시한 도면이고, 도 10은 본 발명의 일 실시예에 따른 제스처 인식 방법의 수신호 분류 네트워크의 학습 동작을 예시한 도면이고, 도 11은 본 발명의 일 실시예에 따른 제스처 인식 방법의 수신호 분류 네트워크의 인식 동작을 예시한 도면이다.

[0073] 제스처 인식 방법은 6가지 교통 수신호를 구분하기 위한 CNN(Convolutional Neural Network) 분류기를 적용한다. 제스처 인식 방법은 네트워크 모델을 적용하며, 네트워크 모델은 특징을 추출하고 특징을 데이터 가공 처리한다. 네트워크 모델은 다수의 레이어가 네트워크로 연결되며 히든 레이어를 포함한다. 레이어는 파라미터를 포함할 수 있고, 레이어의 파라미터는 학습가능한 필터 집합을 포함한다. 파라미터는 노드 간의 가중치 및/또는 바이어스를 포함한다. 모델은 손실 함수를 최소화하도록 파라미터를 학습한다.

[0074] 제스처 인식 방법에 적용되는 수신호 분류 네트워크는 특징 추출 네트워크(Feature Extraction Network, FEN), 특징 추출 네트워크에 연결된 대조 임베딩 네트워크(Contrastive Embedding Network, CEN), 특징 추출 네트워크에 연결된 단일 추론 네트워크(Single Inference Network, SIN)를 포함한다.

[0075] 제스처 인식 방법에 적용되는 수신호 분류 네트워크는 대조 임베딩 네트워크(CEN)를 이용하여 제스처 변화가 기준치보다 큰 유형에 대해서 교통 수신호를 분류하고, 단일 추론 네트워크(SIN)를 이용하여 제스처 변화가 기준치보다 작은 유형에 대해서 교통 수신호를 분류할 수 있다.

[0076] 제스처 인식 방법에 적용되는 특징 추출 네트워크는 특징 추출 네트워크에 추가로 연결된 특징 디코딩 네트워크(Feature Decode Network, FDN)를 적용하여 오토인코더(Autoencoder) 구조로 학습될 수 있다.

[0077] 제스처 인식 방법은 손 모델링 데이터 및 복수의 프레임으로부터 특징 추출 네트워크를 통해 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터를 추출할 수 있다. 제스처 인식 방법은 대조 임베딩 네트워크를 통해 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터 간의 차이로부터 교통 수신호를 분류할 수 있다. 한편 제스처 인식 방법은 단일 추론 네트워크를 통해 포즈 프레임의 제1 특징 벡터로부터 교통 수신호를 분류할 수 있다.

[0078] 제스처 인식 방법은 현재 프레임과 이전 프레임의 2개 프레임을 사용하여 제스처를 인식한다. 예컨대, 이전 프레임은 15frame 이전(0.625초 간격)으로 설정될 수 있다.

[0079] FEN(Feature Extraction Network)은 프레임으로부터 특징을 추출하여 CEN과 SIN으로 전달한다. FEN의 Feature Extracted Layer는 16-32-64차원으로 이루어진 3층 Hidden CNN으로 구현될 수 있다.

[0080] CEN(Contrastive Embedding Network)은 현재 프레임과 이전 프레임의 특징 차이를 학습하여 제스처를 분석한다. CEN은 이전 프레임과의 변화가 큰 제스처(예컨대, 좌측에서 우측으로 진행 등 대부분의 제스처)에 대해 높은 인식률을 보인다. CEN의 Contrastive Embedding Layer는 64-100-6차원으로 이루어진 3층 FC(Fully Connected)로 구현될 수 있다.

[0081] SIN(Single Inference Network)은 현재 프레임만으로 학습하여 제스처를 분석한다. SIN은 이전 프레임과의 변화

가 크지 않은 제스처(예컨대, 전방 정지, 측방 정지)에 대해 높은 인식률을 보인다. SIN의 Single Inference Layer는 32-50-6차원으로 이루어진 3층 FC로 구현될 수 있다.

[0082] FDN의 Feature Decode Layer는 64-32-16차원으로 이루어진 3층 Hidden Transpose CNN으로 구현될 수 있다. 학습에만 사용되는 레이어이다.

[0083] 대조 임베딩 네트워크의 대조 임베딩 손실 함수는 (i) 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터의 차이, (ii) 임베딩 공간의 마진, (iii) 포즈 프레임 및 이전 프레임의 유사도를 기반으로 정의되며, 대조 임베딩 손실 함수가 최소화되도록 학습될 수 있다.

수학식 5

[0084]
$$L_{Contrastive}(I_1, I_2) = (1-S)\frac{1}{2}(ED)^2 + (S)\frac{1}{2}\{max(0,(m-ED))\}^2$$

[0085] 수학식 5에서 ED는 이미지 쌍의 특징 벡터 간의 유클리드 거리를 의미한다. m은 임베딩 공간 주변의 마진을 제공하는 양수값이다.

[0086] 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터 중에서 이미지 쌍의 유사하지 않은 특징 벡터에 대해서 마진의 범위 내에 특징 벡터의 거리가 존재하면 대조 임베딩 손실 함수에서 적용될 수 있다.

[0087] 제스처 인식 방법에 적용 가능한 유사도는 수학식 6과 같이 표현된다.

수학식 6

[0088]
$$S = \begin{cases} 1 \; if \; I_1 \cong I_2 \\ 0 \; otherwise \end{cases}$$

[0089] 제스처 인식 방법은 학습 과정에서 이전 데이터(이전 프레임 데이터)와 포즈 추론을 위한 데이터(대상 프레임 데이터)를 동시에 입력한다.

[0090] 제스처 인식 방법은 두 데이터에 대해 추출(Extract)-디코딩(Decode)으로 이루어진 오토인코더 구조를 통과시켜 Feature Extracted Layer를 학습한다. (Error: MSE)

[0091] 제스처 인식 방법은 두 데이터에 대해 Extracted Layer의 출력 (64차원)을 계산하여, 포즈 추론을 위한 데이터(Pose Inference Data)의 출력(A)과 이전 데이터-포즈 추론을 위한 데이터 간 대비 출력(Contrastive Output)(B)을 출력한다.

[0092] 제스처 인식 방법은 A를 Single Inference Layer에 넣어 학습시키고 (Error: Cross Entropy), B를 Contrastive Embedding Layer에 넣어 학습시킨다. (Error: Cross Entropy)

[0093] 제스처 인식 방법은 추론 과정에서 이전 데이터(이전 프레임 데이터)와 포즈 추론을 위한 데이터(대상 프레임 데이터)를 입력으로 집어넣는다.

[0094] 두 데이터에 대해 Extracted Layer의 출력(64차원)을 계산하여, 포즈 추론을 위한 데이터(Pose Inference Data)의 출력(A)과 이전 데이터-포즈 추론을 위한 데이터 간 대비 출력(Contrastive Output)(B)을 출력한다.

[0095] 제스처 인식 방법은 A를 Single Inference Layer에, B를 Contrastive Embedding Layer에 넣은 뒤 결합(Concatenate)시켜 추론(Inference)한다.

[0096] 제스처 인식 방법은 분리된 모델 구조를 통해 다양한 수신호를 정확하게 추론할 수 있다.

[0097] 도 12는 본 발명의 일 실시예에 따른 제스처 인식 방법의 인식률을 예시한 도면이고, 도 13 및 도 14는 본 발명

의 일 실시예에 따른 제스처 인식 방법의 프레임별 제스처 인식여부를 예시한 도면이다.

[0098] 해당 정확도는 프레임 단위로 측정한 정확도이며, 프레임 단위 제스처 인식에 대해 평균 96.70%의 인식 정확도를 보인다. 제스처에 따라 수행되는 시간이 다르고, 수행하는 사람마다 개인차가 있으므로 프레임 단위의 정확도보다 하나의 제스처를 수행함에 있어서 어느 지점에서 에러가 발생하는지의 분석이 필요하다.

[0099] 제스처 인식에서 발생한 에러는 대부분 제스처의 시작 혹은 끝에서 발생한다. 일부 프레임에서 손의 위치를 놓쳐 에러가 발생하는 경우가 대부분이다.

[0100] 도 13 및 도 14의 10 frame(0.416초) 단위로 샘플링된 제스처 인식 결과를 살펴보면, 일부 프레임에서 잘못된 제스처를 인식하는 경우는 있지만, 이는 하나의 제스처의 전체 과정을 놓고 볼 때 극히 일부에 불과하다. 실시예에서 발명자의 실험 조건에서는 제스처 하나당 평균 수행 시간은 3.49초(84프레임)이므로, 일부 에러가 발생하여도 전체 제스처 인식에는 문제가 없다. 완전한 제스처(Complete gesture)의 인식률은 100%를 달성하는 것을 확인할 수 있다.

[0101] 자율 주행 자동차 실험에 사용되는 임베디드 보드(embedded board)에서 3가지 딥모델(deep-model)을 수행했을 때의 총 수행시간이 0.625초 이내여야 한다. 네트워크가 제스처 인식에 2개의 프레임을 필요로 하는데, 15 frame(0.625초) 단위로 프레임을 사용하기 때문이다.

[0102] 실시예에서 발명자의 실험 조건에서는 ATC 인식 시간은 0.138초이고, ATC 손 인식 및 3D 모델링 소요 시간은 0.250초이고, 제스처 인식 시간은 0.044초이므로, 총 수행 시간은 0.432초이다. 총 수행시간이 0.625초 이내를 만족하므로 실제 자율 주행차에 적용 가능한 것을 확인할 수 있다.

[0103] 제스처 인식 방법은 제스처 인식 장치에 의해 수행될 수 있다.

[0104] 도 15는 본 발명의 다른 실시예에 따른 제스처 인식 장치를 예시한 블록도이다.

[0105] 제스처 인식 장치(100)는 적어도 하나의 프로세서(120), 컴퓨터 판독 가능한 저장매체(130) 및 통신 버스(170)를 포함한다.

[0106] 프로세서(120)는 제스처 인식 장치(110)로 동작하도록 제어할 수 있다. 예컨대, 프로세서(120)는 컴퓨터 판독 가능한 저장 매체(130)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 컴퓨터 실행 가능 명령어는 프로세서(120)에 의해 실행되는 경우 제스처 인식 장치(100)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.

[0107] 컴퓨터 판독 가능한 저장 매체(130)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보는 입출력 인터페이스(150)나 통신 인터페이스(160)를 통해서도 주어질 수 있다. 컴퓨터 판독 가능한 저장 매체(130)에 저장된 프로그램(140)은 프로세서(120)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독 가능한 저장 매체(130)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 제스처 인식 장치(110)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.

[0108] 통신 버스(170)는 프로세서(120), 컴퓨터 판독 가능한 저장 매체(130)를 포함하여 제스처 인식 장치(100)의 다른 다양한 컴포넌트들을 상호 연결한다.

[0109] 제스처 인식 장치(100)는 또한 하나 이상의 입출력 장치를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(150) 및 하나 이상의 통신 인터페이스(160)를 포함할 수 있다. 입출력 인터페이스(150) 및 통신 인터페이스(160)는 통신 버스(170)에 연결된다. 입출력 장치(미도시)는 입출력 인터페이스(150)를 통해 제스처 인식 장치(100)의 다른 컴포넌트들에 연결될 수 있다.

[0110] 프로세서는 대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하고, 손 인식 네트워크를 통해 교통 관제사 영역으로부터 손 영역을 추출하고, 손 모델링 네트워크를 통해 손 영역으로부터 손 모델링 데이터를 생성하고, 수신호 분류 네트워크를 통해 손 모델링 데이터 및 복수의 프레임으로부터 교통 수신호를 분류한다.

[0111] 제스처 인식 장치가 분류하는 교통 수신호는, (i) 전방 정지 유형, (ii) 좌우측방 동시정지 유형, (iii) 좌측에서 우측으로 진행 유형, (iv) 전방에서 좌측으로 회전 유형, (v) 우측에서 좌측으로 진행 유형, (vi) 전방에서

우측으로 회전 유형을 포함할 수 있다.

[0112] 프로세서는 손 모델링 데이터의 골격 포인트를 이용하여 왼손과 오른손을 구분하고, 손바닥과 손등을 구분할 수 있다.

[0113] 손 모델링 네트워크의 손 모델링 손실 함수는 손 포즈 손실 함수, 3차원 키포인트 손실 함수, 2차원 키포인트 손실 함수, 정규화 손실 함수에 각각 가중치가 적용된 합으로 정의되며, 손 모델링 손실 함수가 최소화되도록 학습될 수 있다.

[0114] 수신호 분류 네트워크는 특징 추출 네트워크, 특징 추출 네트워크에 연결된 대조 임베딩 네트워크, 특징 추출 네트워크에 연결된 단일 추론 네트워크를 포함하며, 대조 임베딩 네트워크를 이용하여 제스처 변화가 기준치보다 큰 유형에 대해서 상기 교통 수신호를 분류하고, 단일 추론 네트워크를 이용하여 제스처 변화가 상기 기준치보다 작은 유형에 대해서 교통 수신호를 분류할 수 있다.

[0115] 특징 추출 네트워크는 특징 추출 네트워크에 추가로 연결된 특징 디코딩 네트워크를 적용하여 오토인코더 구조로 학습될 수 있다.

[0116] 프로세서는 손 모델링 데이터 및 복수의 프레임으로부터 특징 추출 네트워크를 통해 포즈 프레임의 제1 특징 벡터 및 이전 프레임의 제2 특징 벡터를 추출할 수 있다. 프로세서는 단일 추론 네트워크를 통해 제1 특징 벡터로부터 교통 수신호를 분류할 수 있다. 프로세서는 대조 임베딩 네트워크를 통해 제1 특징 벡터 및 제2 특징 벡터의 차이로부터 교통 수신호를 분류할 수 있다.

[0117] 대조 임베딩 네트워크의 대조 임베딩 손실 함수는 (i) 제1 특징 벡터 및 제2 특징 벡터의 차이, (ii) 임베딩 공간의 마진, (iii) 포즈 프레임 및 이전 프레임의 유사도를 기반으로 정의되며, 대조 임베딩 손실 함수가 최소화되도록 학습될 수 있다.

[0118] 제1 특징 벡터 및 제2 특징 벡터 중에서 이미지 쌍의 유사하지 않은 특징 벡터에 대해서 마진의 범위 내에 특징 벡터의 거리가 존재하면 대조 임베딩 손실 함수에서 적용될 수 있다.

[0119] 제스처 인식 장치는 이동체에 탑재될 수 있다.

[0120] 도 16은 본 발명의 다른 실시예에 따른 이동체를 예시한 블록도이다.

[0121] 이동체(10)는 제스처 인식 장치(100) 및 주행 수단(200)을 포함한다. 이동체(10)는 일반 자동차, 자율 주행차, 이동 로봇 등을 포함하는 개념이다.

[0122] 제스처 인식 장치(100)는 대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하고, 손 인식 네트워크를 통해 교통 관제사 영역으로부터 손 영역을 추출하고, 손 모델링 네트워크를 통해 손 영역으로부터 손 모델링 데이터를 생성하고, 수신호 분류 네트워크를 통해 상기 손 모델링 데이터 및 상기 복수의 프레임으로부터 교통 수신호를 분류한다.

[0123] 주행 수단(200)은 제스처 인식 장치의 출력 결과를 기반으로 이동 계획을 갱신하고 이동을 수행한다. 주행 수단(200)은 바퀴, 로봇 다리, 체인, 선로 등을 통해 이동 가능하도록 구현된 장치이다. 산출된 주행 경로를 따라 자동으로 이동시키거나 조작 등을 통해 수동으로 이동시킬 수도 있다.

[0124] 제스처 인식 장치는 하드웨어, 펌웨어, 소프트웨어 또는 이들의 조합에 의해 로직회로 내에서 구현될 수 있고, 범용 또는 특정 목적 컴퓨터를 이용하여 구현될 수도 있다. 장치는 고정배선형(Hardwired) 기기, 필드 프로그램 가능한 게이트 어레이(Field Programmable Gate Array, FPGA), 주문형 반도체(Application Specific Integrated Circuit, ASIC) 등을 이용하여 구현될 수 있다. 또한, 장치는 하나 이상의 프로세서 및 컨트롤러를 포함한 시스템온칩(System on Chip, SoC)으로 구현될 수 있다.

[0125] 제스처 인식 장치는 하드웨어적 요소가 마련된 컴퓨팅 디바이스 또는 서버에 소프트웨어, 하드웨어, 또는 이들의 조합하는 형태로 탑재될 수 있다. 컴퓨팅 디바이스 또는 서버는 각종 기기 또는 유무선 통신망과 통신을 수행하기 위한 통신 모뎀 등의 통신장치, 프로그램을 실행하기 위한 데이터를 저장하는 메모리, 프로그램을 실행하여 연산 및 명령하기 위한 마이크로프로세서 등을 전부 또는 일부 포함한 다양한 장치를 의미할 수 있다.

[0126] 도 4에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나 이는 예시적으로 설명한 것에 불과하고, 이 분야의 기술자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 4에 기재된 순서를 변경하여 실행하거나 또는 하나 이상의 과정을 병렬적으로 실행하거나 다른 과정을 추가하는 것으로 다

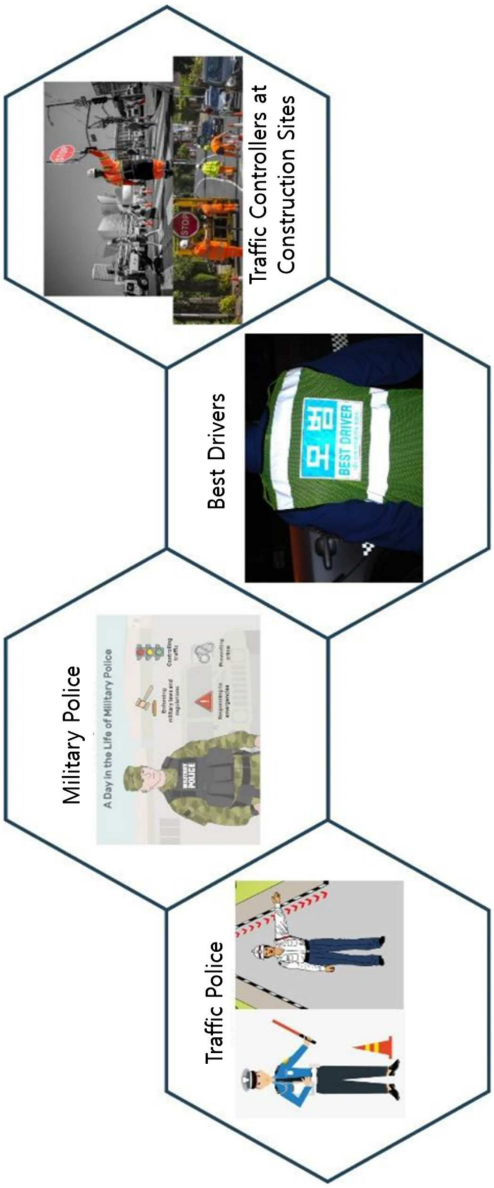
양하게 수정 및 변형하여 적용 가능할 것이다.

[0127] 본 실시예들에 따른 동작은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능한 매체에 기록될 수 있다. 컴퓨터 판독 가능한 매체는 실행을 위해 프로세서에 명령어를 제공하는 데 참여한 임의의 매체를 나타낸다. 컴퓨터 판독 가능한 매체는 프로그램 명령, 데이터 파일, 데이터 구조 또는 이들의 조합을 포함할 수 있다. 예를 들면, 자기 매체, 광기록 매체, 메모리 등이 있을 수 있다. 컴퓨터 프로그램은 네트워크로 연결된 컴퓨터 시스템 상에 분산되어 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 본 실시예를 구현하기 위한 기능적인(Functional) 프로그램, 코드, 및 코드 세그먼트들은 본 실시예가 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있을 것이다.

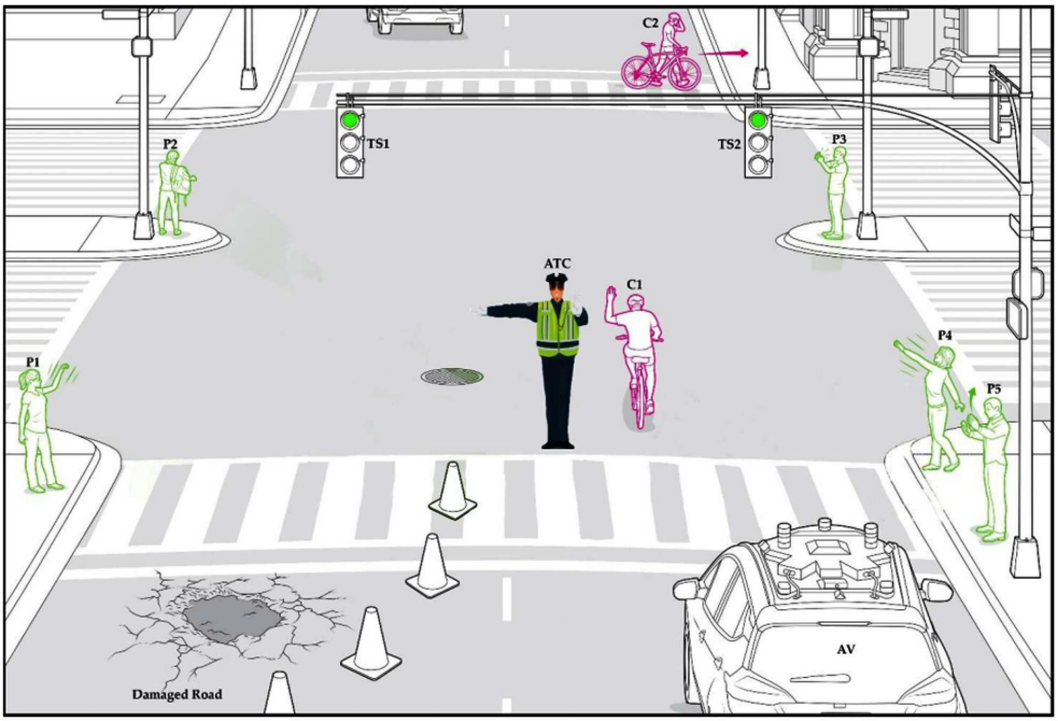
[0128] 본 실시예들은 본 실시예의 기술 사상을 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

도면1



도면2

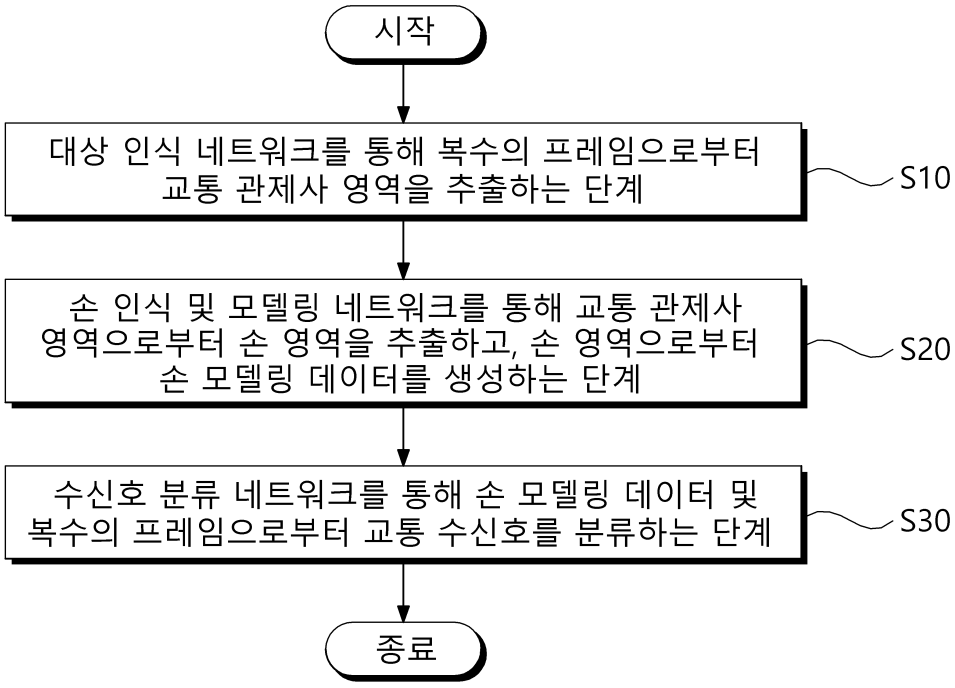


도면3

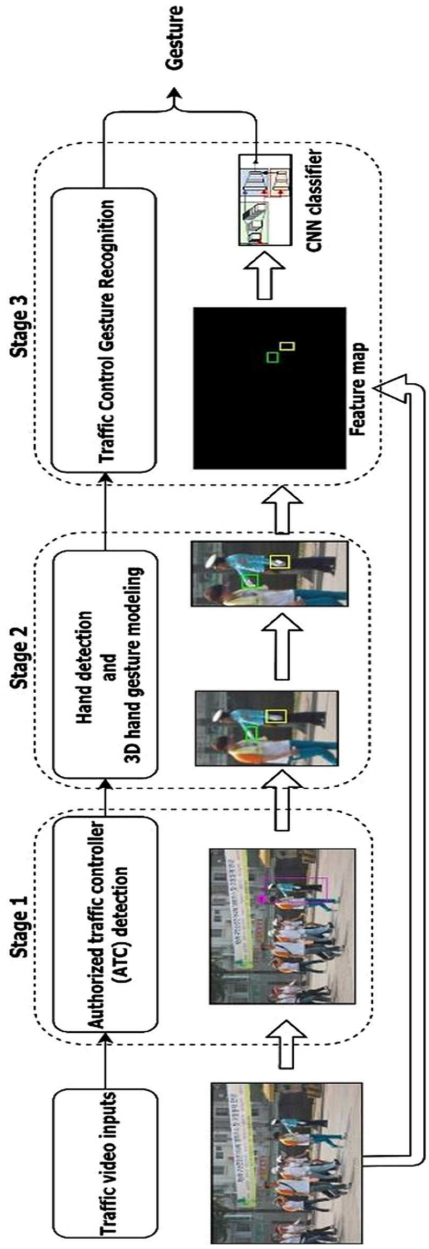
Gestures (Important frame sequences)	Commands
① Stop traffic approaching from the front	Stop (Front)
② Stop traffic approaching from the side	Stop (Side)
③Vehicles on the left proceed to right	Go (L → R)
④ Vehicles on the front proceed to left	Go (F → L)
⑤ Vehicles on the right proceed to left	Go (R → L)
⑥ Vehicles on the front proceed to right	Go (F → R)

Here, L: left; R: right; F: front

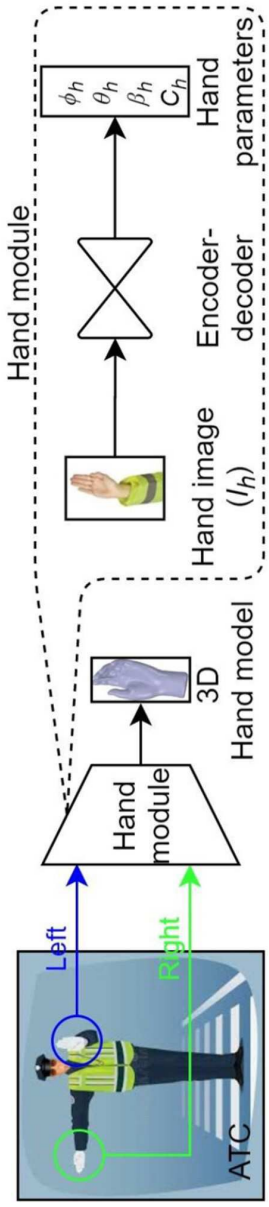
도면4


시작
대상 인식 네트워크를 통해 복수의 프레임으로부터 교통 관제사 영역을 추출하는 단계
S10
손 인식 및 모델링 네트워크를 통해 교통 관제사 영역으로부터 손 영역을 추출하고, 손 영역으로부터 손 모델링 데이터를 생성하는 단계
S20
수신호 분류 네트워크를 통해 손 모델링 데이터 및 복수의 프레임으로부터 교통 수신호를 분류하는 단계
S30
종료

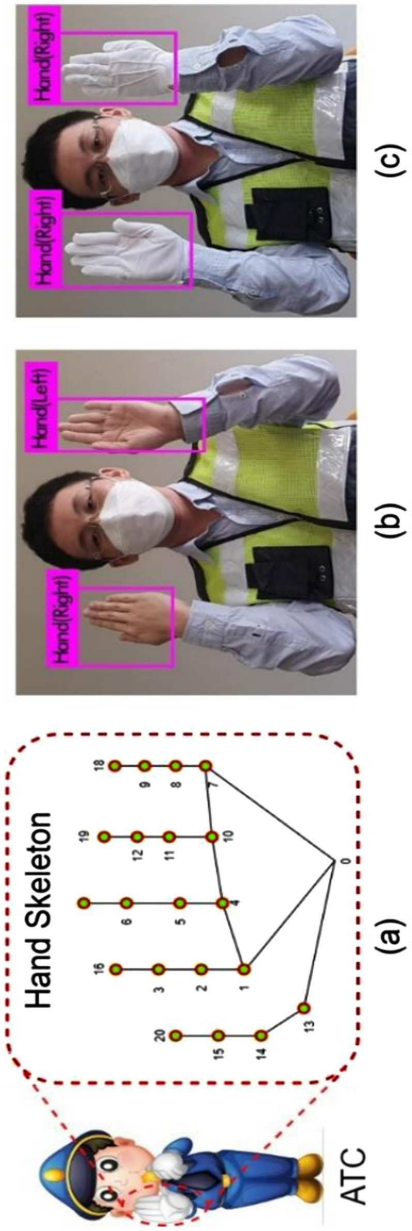
도면5


Gesture
Stage 3
Traffic Control Gesture Recognition
CNN classifier
Feature map
Stage 2
Hand detection
and
3D hand gesture modeling
Stage 1
Authorized traffic controller
(ATC) detection
Traffic video inputs

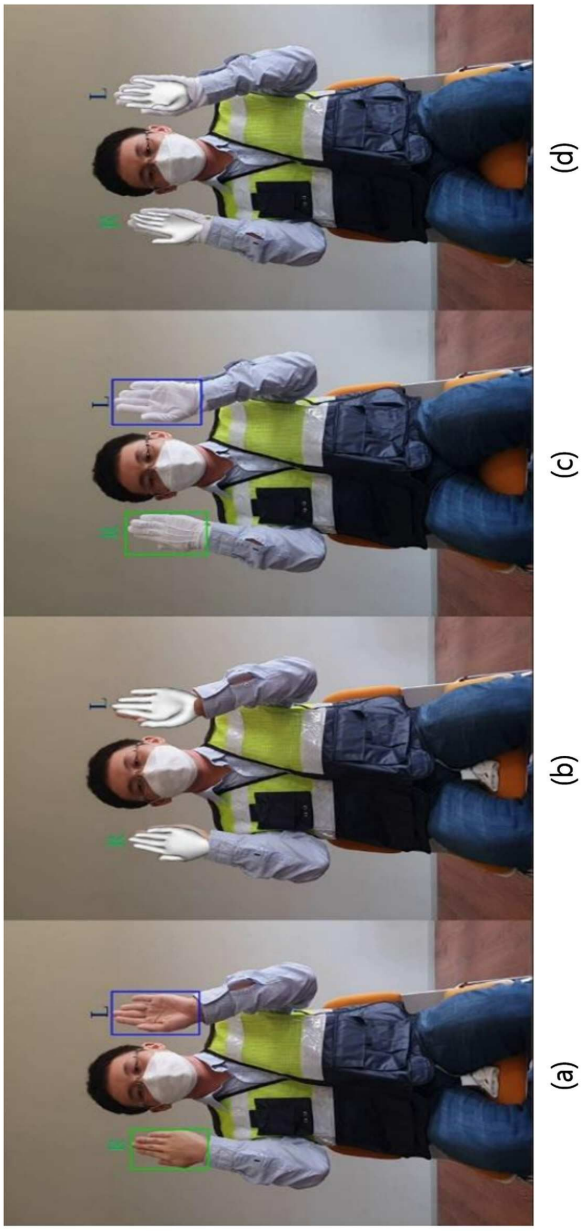
도면6


Hand module
Hand parameters
Encoder-decoder
Hand image (I_h)
3D Hand model
Hand module
Left
Right
ATC

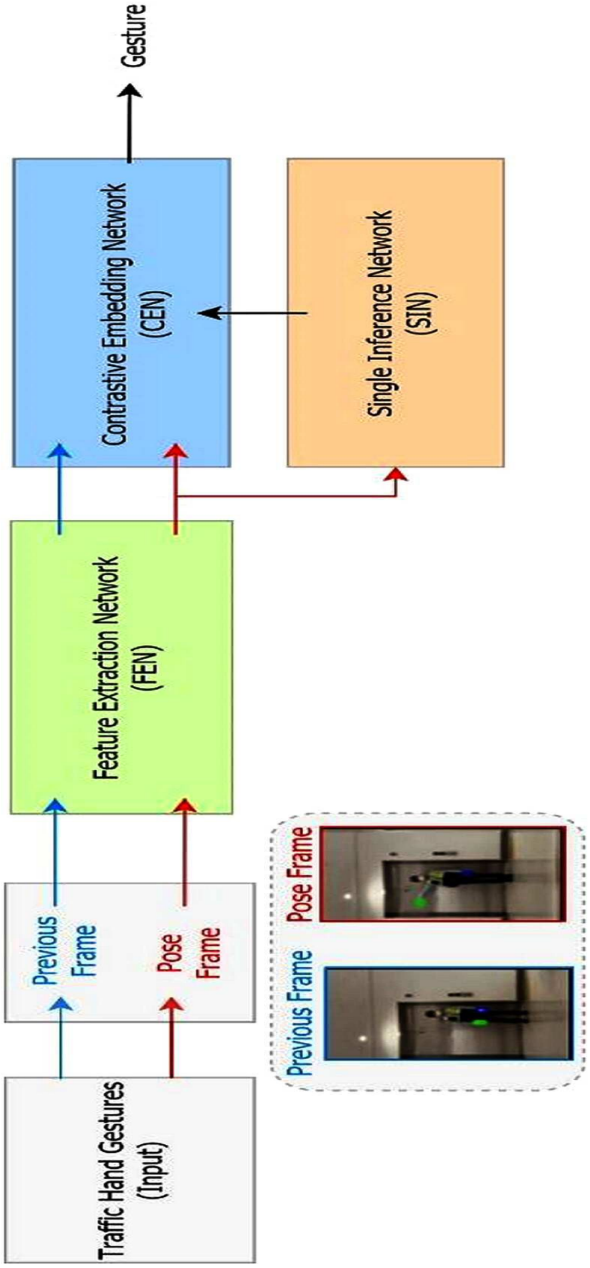
도면7


Hand(Right)
Hand(Right)
(c)
Hand(Left)
Hand(Right)
(b)
Hand Skeleton
ATC
(a)

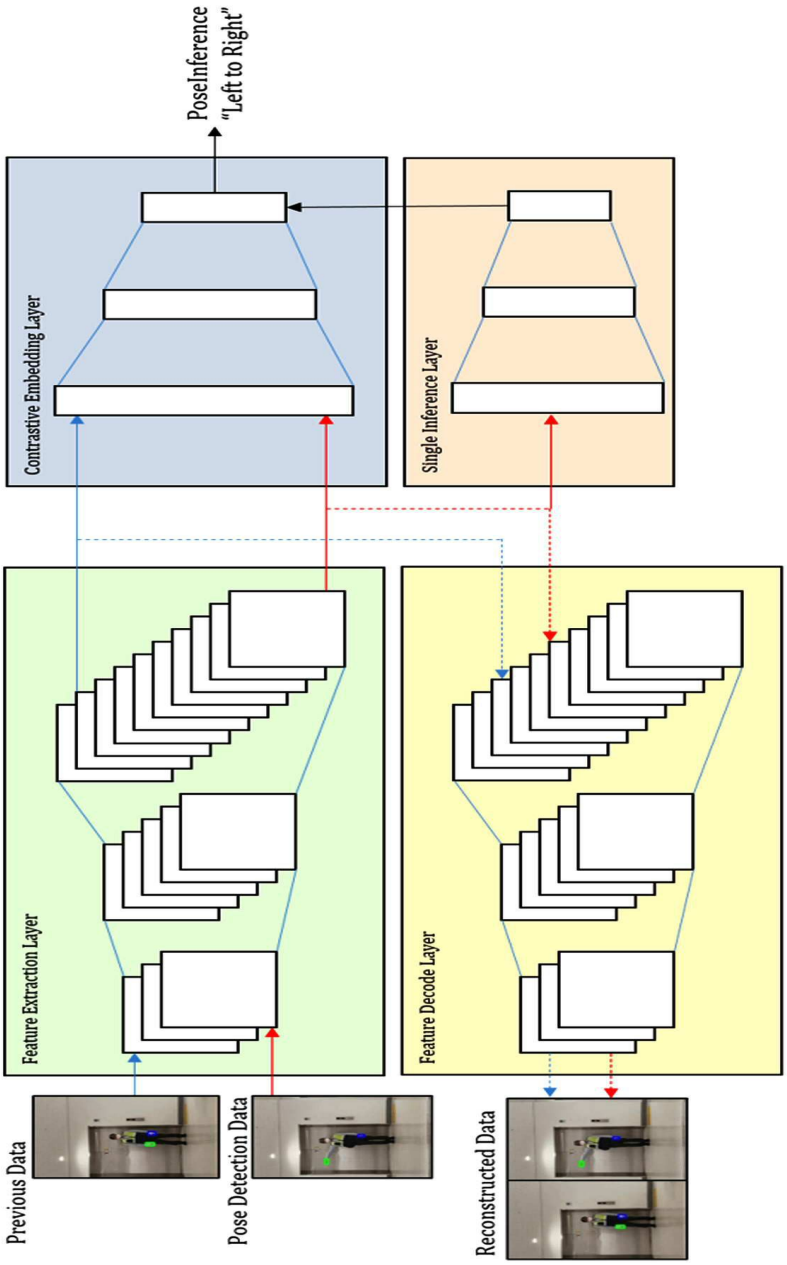
도면8



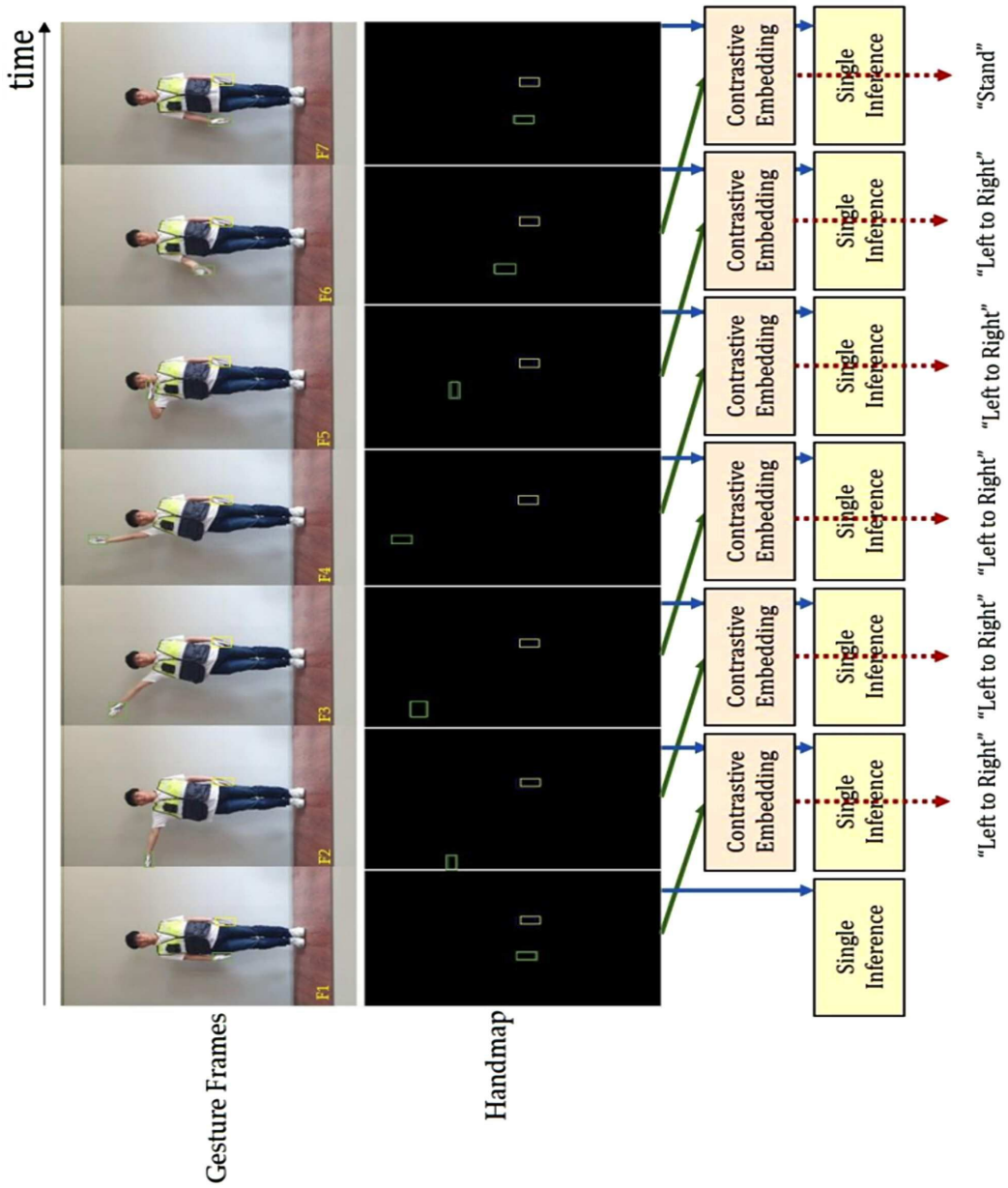
도면9


Gesture
Contrastive Embedding Network (CEN)
Single Inference Network (SIN)
Feature Extraction Network (FEN)
Previous Frame
Pose Frame
Traffic Hand Gestures (Input)
Pose Frame
Previous Frame

도면10


PoseInference
"Left to Right"
Contrastive Embedding Layer
Single Inference Layer
Feature Extraction Layer
Feature Decode Layer
Previous Data
Pose Detection Data
Reconstructed Data

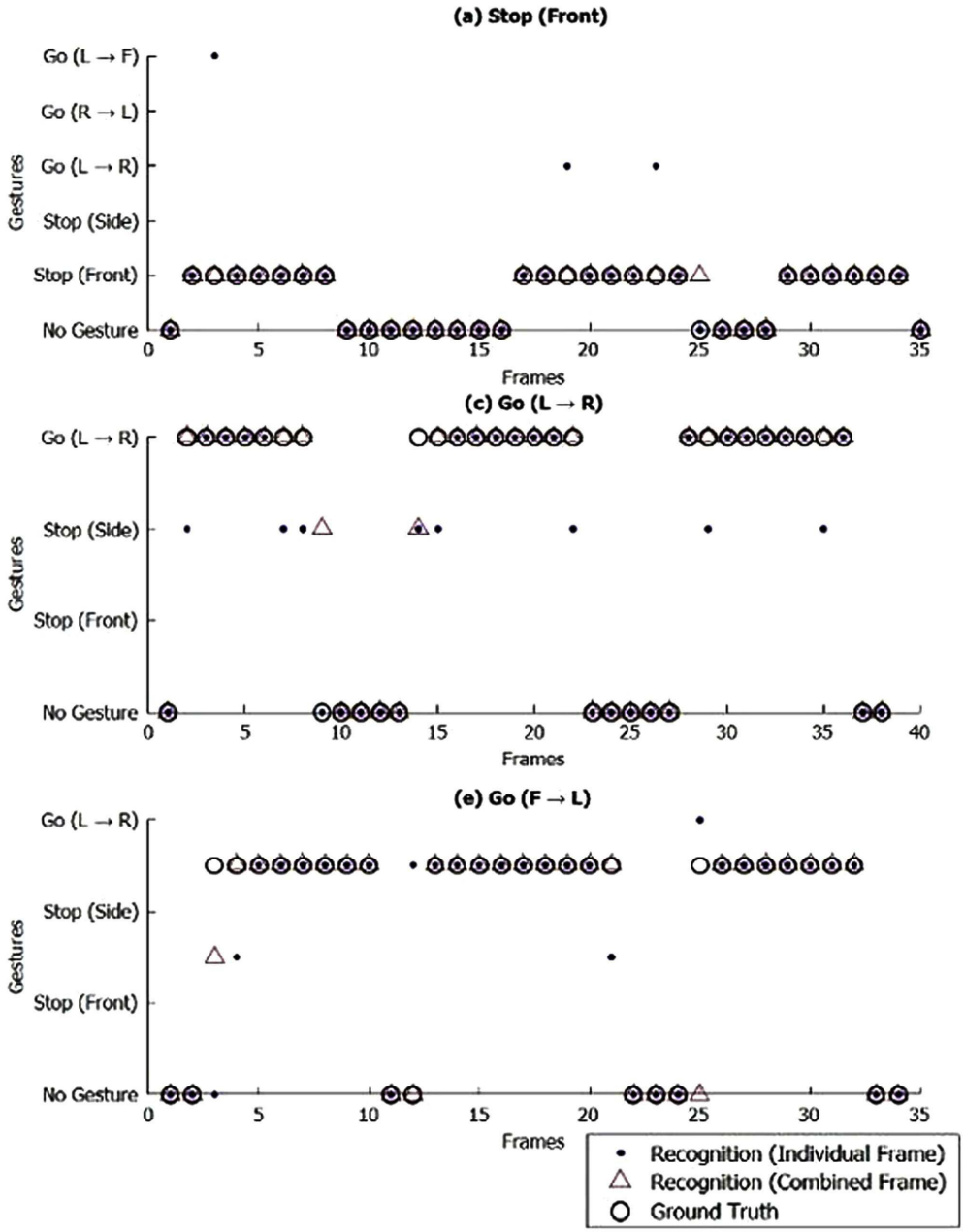
도면11



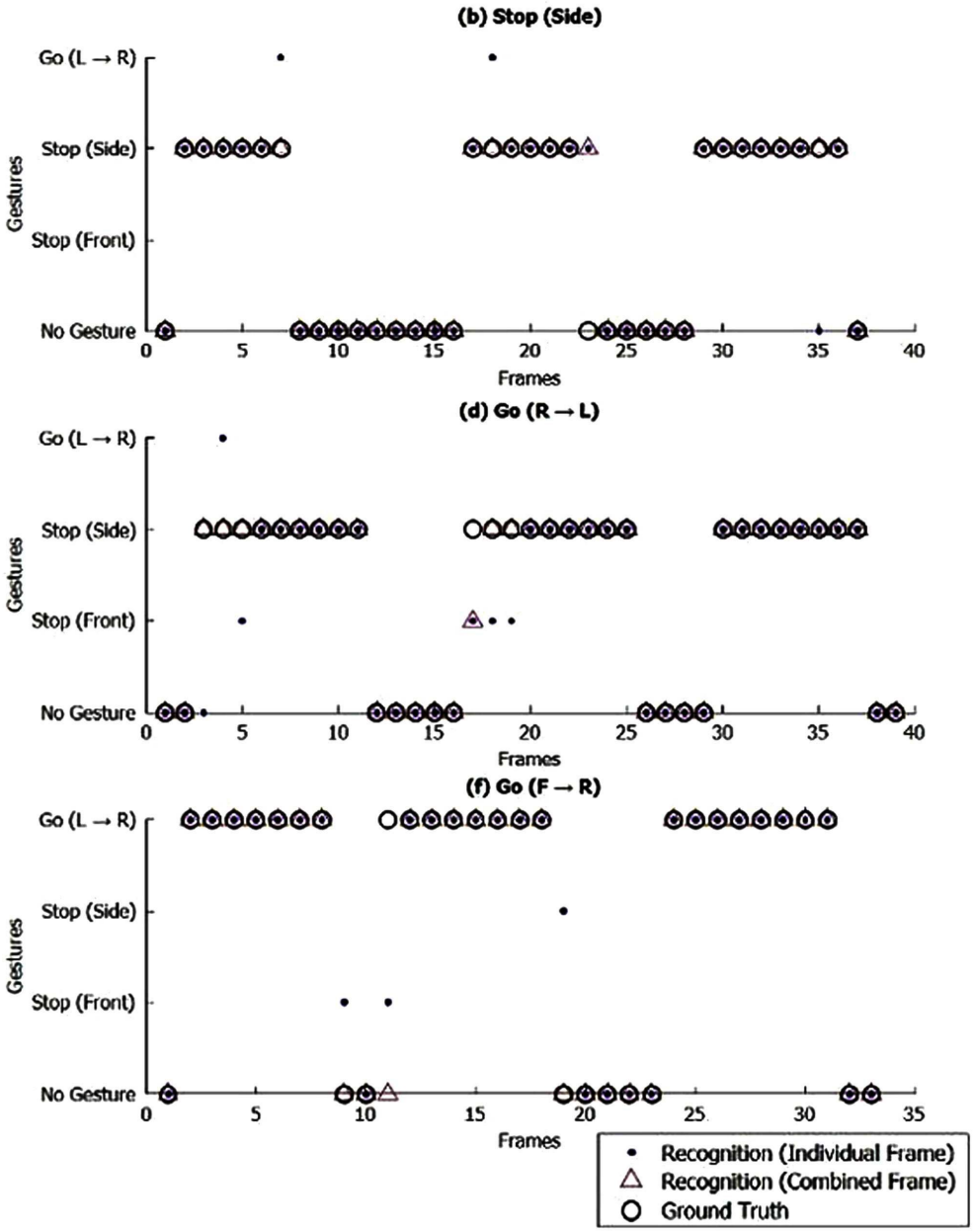
도면12

Gestures	Stop (Front)	Stop (Side)	L → R	R → L	F → L	F → R
Input Frames	9966	3740	4841	4121	6528	6095
Error	72	103	182	231	219	356
Recognition (%)	99.27	97.25	96.24	94.39	96.65	94.16

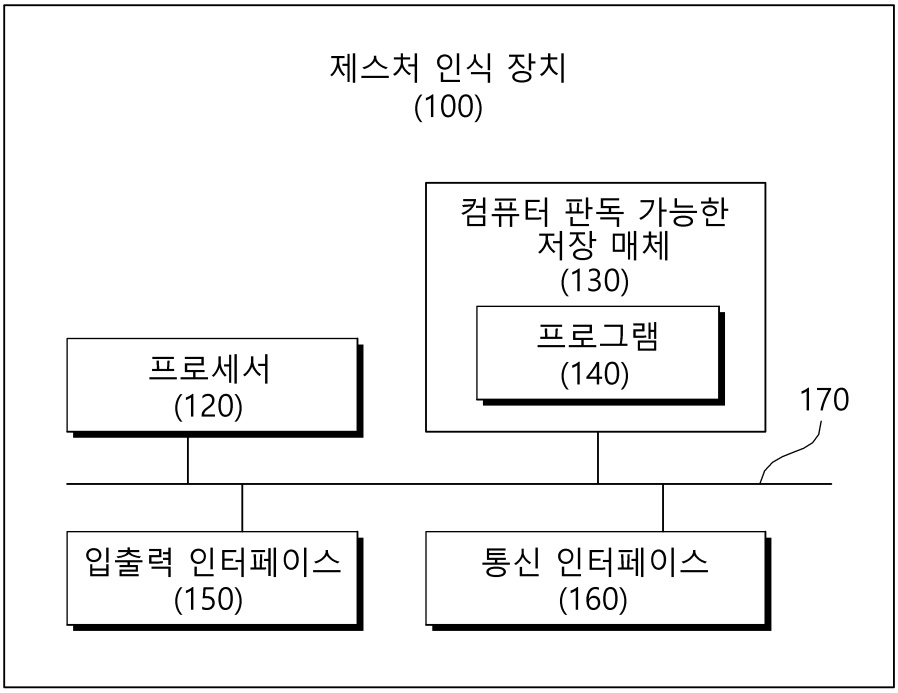
도면13


(a) Stop (Front)
Go (L → F)
Go (R → L)
Go (L → R)
Stop (Side)
Stop (Front)
No Gesture
Gestures
0 5 10 15 20 25 30 35
Frames
(c) Go (L → R)
Go (L → R)
Stop (Side)
Stop (Front)
No Gesture
Gestures
0 5 10 15 20 25 30 35 40
Frames
(e) Go (F → L)
Go (L → R)
Stop (Side)
Stop (Front)
No Gesture
Gestures
0 5 10 15 20 25 30 35
Frames
Recognition (Individual Frame)
Recognition (Combined Frame)
Ground Truth

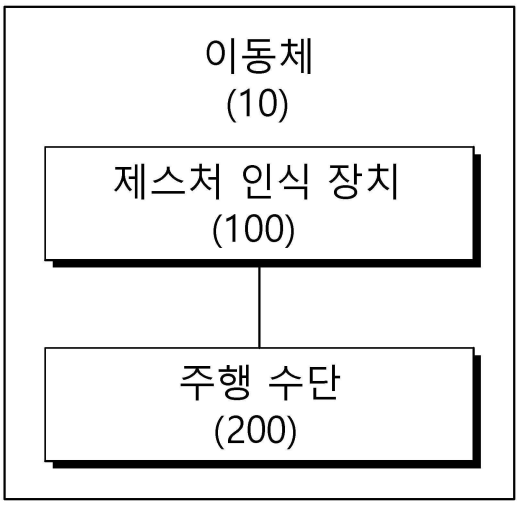
도면14


(b) Stop (Side)
Go (L → R)
Stop (Side)
Stop (Front)
No Gesture
Gestures
0
5
10
15
20
25
30
35
40
Frames
(d) Go (R → L)
Go (L → R)
Stop (Side)
Stop (Front)
No Gesture
Gestures
0
5
10
15
20
25
30
35
40
Frames
(f) Go (F → R)
Go (L → R)
Stop (Side)
Stop (Front)
No Gesture
Gestures
0
5
10
15
20
25
30
35
Frames
Recognition (Individual Frame)
Recognition (Combined Frame)
Ground Truth

도면15


제스처 인식 장치
(100)
컴퓨터 판독 가능한
저장 매체
(130)
프로그램
(140)
프로세서
(120)
170
입출력 인터페이스
(150)
통신 인터페이스
(160)

도면16


이동체
(10)
제스처 인식 장치
(100)
주행 수단
(200)