



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2020년07월28일

(11) 등록번호 10-2138680

(24) 등록일자 2020년07월22일

(51) 국제특허분류(Int. Cl.)  
*G06K 9/00* (2006.01) *G06T 7/10* (2017.01)  
*G06T 7/73* (2017.01)  
(52) CPC특허분류  
*G06K 9/00335* (2013.01)  
*G06K 9/00744* (2013.01)  
(21) 출원번호 10-2018-0043730  
(22) 출원일자 2018년04월16일  
심사청구일자 2018년04월16일  
(65) 공개번호 10-2019-0120489  
(43) 공개일자 2019년10월24일  
(56) 선행기술조사문헌  
JP2011100175 A\*  
KR1020150103443 A\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
**연세대학교 산학협력단**  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)  
(72) 발명자  
**변혜란**  
서울특별시 서대문구 연세로 50, (제4공학관) 810호 (신촌동)  
**조보라**  
서울특별시 서대문구 연세로 50, (제4공학관) 810호 (신촌동)  
(뒷면에 계속)  
(74) 대리인  
**특허법인우인**

전체 청구항 수 : 총 18 항

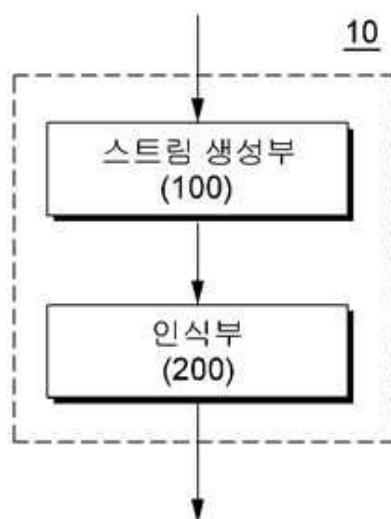
심사관 : 노용완

(54) 발명의 명칭 **영상 인식 장치 및 방법**

### (57) 요약

본 발명은 영상 인식 장치로서, 특히, 동영상 분석이 가능한 합성곱 신경망을 이용하여 영상 내 객체의 행동을 인식할 수 있는 영상 인식 장치를 개시한다. 본 발명의 영상 인식 장치는 적어도 하나의 객체들을 포함하는 입력 영상에서 상기 객체별 행동에 관한 모션 정보를 포함하는 액션 스트림을 생성하는 스트림 생성부; 및 상기 생성된 액션 스트림 또는 상기 입력 영상에서 액션 스트림의 위치 관계를 나타내는 위치 정보를 입력 받아, 상기 객체의 행동을 분류하기 위한 지표로서 적어도 하나의 클래스 벡터를 출력하는 제1 인식기를 이용하여 상기 객체들의 행동을 인식하는 인식부;를 포함한다.

**대표도** - 도1



(52) CPC특허분류

**G06T 7/10** (2017.01)

**G06T 7/73** (2017.01)

(72) 발명자

**홍기범**

서울특별시 서대문구 연세로 50, (제4공학관) 810호 (신촌동)

**홍중광**

서울특별시 서대문구 연세로 50, (제4공학관) 810호 (신촌동)

**김호성**

서울특별시 서대문구 연세로 50, (제4공학관) 810호 (신촌동)

**황선희**

서울특별시 서대문구 연세로 50, (제4공학관) 810호 (신촌동)

**기민승**

서울특별시 서대문구 연세로 50, (제4공학관) 810호 (신촌동)

**김태형**

서울특별시 서대문구 연세로 50, (제4공학관) 810호 (신촌동)

이 발명을 지원한 국가연구개발사업

과제고유번호 2019070646

부처명 과학기술정보통신부

연구관리전문기관 한국연구재단

연구사업명 원천기술개발사업

연구과제명 (2세부)딥러닝 기반 의미론적 상황 이해 원천기술 연구 (2단계)(1/2)

기 여 율 1/1

주관기관 연세대학교

연구기간 2019.06.01 ~ 2020.03.31

공지예외적용 : 있음

## 명세서

### 청구범위

#### 청구항 1

행동하는 주체가 되는 제1 객체와 또 다른 상기 행동하는 주체의 주변에 위치하는 적어도 하나의 제2 객체들을 포함하는 인접하는 서로 다른 프레임 영상들에서 상기 행동하는 주체가 되는 제1 객체의 행동에 관한 모션 정보를 포함하는 제1 액션 스트림을 생성하고, 상기 제2 객체의 변화에 따른 제2 액션 스트림을 생성하여, 상기 제1 액션 스트림과 제2 액션 스트림을 조합하여 조합 스트림을 생성하는 스트림 생성부; 및

상기 제1 액션 스트림, 제2 액션 스트림 및 조합 스트림을 입력 받아, 상기 행동하는 주체가 되는 객체의 행동을 분류하기 위한 지표로서 각각의 클래스 벡터를 출력하는 제1 인식기;와 상기 제1 인식기에서 출력된 상기 클래스 벡터들을 미리 설정된 방법에 따라 합산하는 합산부;를 포함하여, 상기 합산된 클래스 벡터를 이용하여 상기 객체들의 행동을 분석하여 인식하는 인식부;를 포함하며,

상기 조합 스트림은 상기 제1 객체와 제2 객체 별로 생성된 제1 액션 스트림과 제2 액션 스트림의 객체 별로 검출된 관심 영역들의 상호 위치에 대한 위치 정보를 기반으로 하며,

상기 클래스 벡터는 상기 제1 객체의 행동을 분류하는 행동 목록과 입력 영상 내의 제1 객체와 제2 객체들의 행동이 상기 행동 목록에 해당할 확률을 나타내는 확률 정보를 포함하며, 상기 인식부는 상기 합산된 클래스 벡터의 상기 행동 목록별 상기 확률 정보에 따라 상기 제1 객체와 상기 제2 객체의 연관성을 고려하여 상기 객체들의 행동을 인식하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 2

제1항에 있어서, 상기 스트림 생성부는

상기 입력 영상을 시간의 흐름에 따라 분할하여 복수개의 프레임 영상들을 생성하며, 상기 생성된 프레임 영상들에서 상기 제1 객체 및 제2 객체 별로 상기 제1 객체 및 제2 객체를 적어도 일부 포함하는 관심 영역을 검출하는 관심 영역 검출부;를 더 포함하고,

상기 검출된 관심 영역을 이용하여 상기 액션 스트림을 생성하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 3

삭제

#### 청구항 4

제2항에 있어서, 상기 스트림 생성부는

상기 인접하는 서로 다른 프레임 영상들에서 검출된 상기 관심 영역간 연결 점수를 산출하며, 상기 산출된 연결 점수를 고려하여 상기 서로 다른 프레임 영상들에서 검출된 상기 관심 영역을 연결하는 관심 영역 연결부;를 더 포함하고,

상기 연결된 관심 영역들을 이용하여 상기 제1 및 제2 객체 별로 상기 액션 스트림을 생성하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 5

제4항에 있어서, 상기 스트림 생성부는

상기 프레임 영상들에서 검출된 상기 관심 영역을 프레임 영상 단위로 재배열하고, 상기 재배열된 관심 영역을 연결하여 생성된 상기 조합 스트림으로부터 상기 액션 스트림의 상호간 위치에 관한 위치 정보를 산출하는 위치 정보 산출부;를 더 포함하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 6

제4항에 있어서, 상기 관심 영역 연결부는

인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 특징 정보를 입력으로 하는 유사도 함수, 상기 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 오버랩 비율을 출력으로 하는 교차비 함수 및 상기 관심 영역들의 클래스 정보의 유사도 중 적어도 하나를 고려하여 연결 점수를 산출하는 연결 점수 산출부; 를 더 포함하고,

상기 산출된 연결 점수를 이용하여 상기 관심 영역을 연결하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 7

제6항에 있어서, 상기 관심 영역 연결부는

상기 관심 영역이 검출되지 않은 프레임 영상이 존재하는 경우,

상기 관심 영역이 검출되지 않은 프레임 영상에 인접한 프레임 영상들에서 검출된 관심 영역들의 특징 정보를 기반으로 상기 관심 영역이 검출되지 않은 프레임 영상 내의 관심 영역의 특징 정보를 추정하는 특징 정보 추정부; 를 더 포함하고,

상기 추정된 특징 정보를 이용하여 상기 관심 영역을 연결하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 8

제2항에 있어서, 상기 관심 영역 검출부는

상기 생성된 프레임 영상 내 상기 관심 영역이 위치하는 좌표를 나타내는 특징 정보를 산출하는 특징 정보 산출부; 를 더 포함하고,

상기 산출된 특징 정보를 이용하여 상기 관심 영역을 검출하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 9

제5항에 있어서, 상기 위치 정보 산출부는

상기 프레임 영상들에서 상기 제1 및 제2 객체 별로 검출된 상기 관심 영역 및 상기 관심 영역이 검출되지 않은 부분을 구분하는 영역 구분부; 및

상기 구분된 관심 영역을 소정의 조합 방법으로 재배열하는 재배열부; 를 더 포함하고,

상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림을 기반으로 상기 위치 정보를 산출하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 10

제8항에 있어서, 상기 특징 정보 산출부는

상기 생성된 프레임 영상들을 각각 분할하여 미리 결정된 크기의 격자셀을 생성하는 전처리부; 및

상기 생성된 프레임 영상들을 입력으로 하고, 상기 격자셀 내부에 중심을 가지고 상기 제1 객체 또는 제2 객체가 존재하는 확률을 나타내는 바운더리 셀의 중심 좌표 또는 상기 바운더리 셀 내 상기 객체가 존재하는 확률을 출력으로 하는 제2 인식기(Neural Network)을 이용하여 상기 바운더리 셀 각각의 중심 좌표 및 상기 바운더리 셀 내의 상기 객체가 존재하는 확률을 계산하는 계산부; 를 더 포함하고,

상기 확률 및 상기 중심 좌표가 계산된 바운더리 셀을 이용하여 상기 특징 정보를 산출하는 것을 특징으로 하는 영상 인식 장치.

#### 청구항 11

제10항에 있어서, 상기 관심 영역 검출부는

상기 프레임 영상들 내에 상기 제1 객체를 중복하여 포함하는 바운더리 셀 중 상기 바운더리 셀 내에 상기 제1 객체가 존재하는 확률이 기 설정된 임계치 이상인지 여부를 고려하여 상기 제1 객체를 중복하여 포함하는 바운더리 셀의 일부를 제거하는 바운더리 셀 제거부; 를 더 포함하고,

제거되고 남은 바운더리 셀을 이용하여 상기 관심 영역을 검출하는 것을 특징으로 하는 영상 인식 장치.

## 청구항 12

삭제

## 청구항 13

스트림 생성부가 행동하는 주체가 되는 제1 객체와 또 다른 상기 행동하는 주체의 주변에 위치하는 적어도 하나의 제2 객체들을 포함하는 인접하는 서로 다른 프레임 영상들에서 상기 행동하는 주체가 되는 제1 객체의 행동에 관한 모션 정보를 포함하는 제1 액션 스트림을 생성하고, 상기 제2 객체의 변화에 따른 제2 액션 스트림을 생성하여, 상기 제1 액션 스트림과 제2 액션 스트림을 조합하여 조합 스트림을 생성하는 단계; 및

인식부가 합산된 클래스 벡터를 이용하여 객체들의 행동을 분석하여 인식하는 단계; 를 포함하며,

상기 합산된 클래스 벡터를 이용하여 객체들의 행동을 분석하여 인식하는 단계는,

제1 인식기가 상기 제1 액션 스트림, 제2 액션 스트림 및 조합 스트림을 입력 받아, 상기 행동하는 주체가 되는 객체의 행동을 분류하기 위한 지표로서 각각의 클래스 벡터를 출력하는 단계;와 합산부가 상기 제1 인식기에서 출력된 상기 클래스 벡터들을 미리 설정된 방법에 따라 합산하는 단계;를 포함하고,

상기 조합 스트림은 상기 제1 객체와 제2 객체 별로 생성된 제1 액션 스트림과 제2 액션 스트림의 객체 별로 검출된 관심 영역들의 상호 위치에 대한 위치 정보를 기반으로 하며,

상기 클래스 벡터는 상기 제1 객체의 행동을 분류하는 행동 목록과 입력 영상 내의 제1 객체와 제2 객체들의 행동이 상기 행동 목록에 해당할 확률을 나타내는 확률 정보를 포함하며, 상기 인식부는 상기 합산된 클래스 벡터의 상기 행동 목록별 상기 확률 정보에 따라 상기 제1 객체와 상기 제2 객체의 연관성을 고려하여 상기 객체들의 행동을 인식하는 것을 특징으로 하는 영상 인식 방법.

## 청구항 14

제13항에 있어서, 상기 생성하는 단계는

상기 입력 영상을 시간의 흐름에 따라 분할하여 복수개의 프레임 영상들을 생성하며, 상기 생성된 프레임 영상들에서 상기 제1 객체 및 제2 객체 별로 상기 제1 객체 및 제2 객체를 적어도 일부 포함하는 관심 영역을 검출하는 단계; 를 더 포함하고,

상기 검출된 관심 영역을 이용하여 상기 액션 스트림을 생성하는 것을 특징으로 하는 영상 인식 방법.

## 청구항 15

제14항에 있어서, 상기 생성하는 단계는

인접하는 서로 다른 프레임 영상들에서 검출된 상기 관심 영역간 연결 점수를 산출하며, 상기 산출된 연결 점수를 고려하여 상기 서로 다른 프레임 영상들에서 검출된 상기 관심 영역을 연결하는 단계; 를 더 포함하고,

상기 연결된 관심 영역들을 이용하여 상기 제1 객체 및 제2 객체 별로 상기 액션 스트림을 생성하는 것을 특징으로 하는 영상 인식 방법.

## 청구항 16

제15항에 있어서, 상기 생성하는 단계는

상기 프레임 영상들에서 검출된 상기 관심 영역을 프레임 영상 단위로 재배열하고, 상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림으로부터 상기 액션 스트림의 상호간 위치에 관한 위치 정보를 산출하는 단계; 를 더 포함하는 것을 특징으로 하는 영상 인식 방법.

## 청구항 17

제15항에 있어서, 상기 연결하는 단계는

인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 특징 정보를 입력으로 하는 유사도 함수, 상기 인접

한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 오버랩 비율을 출력으로 하는 교차비 함수 및 상기 관심 영역들의 클래스 정보의 유사도 중 적어도 하나를 고려하여 연결 점수를 산출하는 단계; 를 더 포함하고,  
상기 산출된 연결 점수를 이용하여 상기 관심 영역을 연결하는 것을 특징으로 하는 영상 인식 방법.

#### 청구항 18

제17항에 있어서, 상기 연결하는 단계는

상기 관심 영역이 검출되지 않은 프레임 영상이 존재하는 경우,

상기 관심 영역이 검출되지 않은 프레임 영상에 인접한 프레임 영상들에서 검출된 관심 영역들의 특징 정보를 기반으로 상기 관심 영역이 검출되지 않은 프레임 영상 내의 관심 영역의 특징 정보를 추정하는 단계; 를 더 포함하고,

상기 추정된 특징 정보를 이용하여 상기 관심 영역을 연결하는 것을 특징으로 하는 영상 인식 방법.

#### 청구항 19

제16항에 있어서, 상기 위치 정보를 산출하는 단계는

상기 프레임 영상들에서 상기 제1 객체 및 제2 객체 별로 검출된 상기 관심 영역 및 상기 관심 영역이 검출되지 않은 부분을 구분하는 단계; 및

상기 구분된 관심 영역을 소정의 조합 방법으로 재배열하는 단계; 를 더 포함하고,

상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림을 기반으로 상기 위치 정보를 산출하는 것을 특징으로 하는 영상 인식 방법.

#### 청구항 20

프로세서에 의해 실행되는 것을 통하여 제13항 내지 제19항 중 어느 한 항에 기재된 영상 인식 방법을 실현하는 컴퓨터에서 판독 가능한 기록매체에 저장된 프로그램.

### 발명의 설명

#### 기술 분야

[0001] 본 발명은 영상 인식 장치에 관한 것이다. 보다 상세하게는 신경망을 이용한 영상 내 객체의 행동을 인식하는 영상 인식 장치에 관한 것이다.

#### 배경 기술

[0002] 행동 인식은 시간 및 공간적 정보를 모두 포함하고 있는 비디오 영상에서 사람의 행동을 분류하는데 사용될 수 있다. 영상 내 주체와 그 주변의 사람, 주체와 주체의 행동 대상이 되는 객체, 객체와 또 다른 객체간의 관계는 영상 내 특정 객체의 행동을 분류하기 위한 요소로 사용될 수 있다.

[0003] 인공 신경망(Neural Network)은 기계 학습과 인지 과학에서 생물학의 신경망에서 비롯된 통계학적 학습 알고리즘인데, 최근 컴퓨터 비전 등의 분야에서 기계적으로 영상을 인식하기 위한 기술이 활발히 연구되고 있다. 컨벌루션 신경망(Convolutional Neural Network, CNN)은 하나 또는 여러 개의 컨벌루션 레이어와 그 위에 올려진 인공 신경망 레이어들로 이루어진 신경망으로 2차원 입력 데이터인 영상과 음성 분석에 주로 사용되는 인공 신경망이다.

[0004] 종래의, 인공 신경망을 이용한 영상 인식 기술은 주로 정지 영상에서만 이루어져왔고, 또한, 정지 영상 내 주체-객체와의 관계를 분석하는 한계가 있었다. 또한, 종래의 영상 인식 기술은 사람, 사물을 포함한 객체간의 관계를 분석함에 있어서 시간의 흐름을 고려하지 않는 문제점이 있었다.

[0005] 따라서, 동영상 단위의 영상 분석을 수행할 수 있는 인공 신경망을 기반으로, 영상 내 객체와, 상기 객체들의 문맥적인 상호 관계를 고려하여 영상을 인식할 수 있는 기술의 개발이 요구되고 있다.

### 선행기술문헌

## 특허문헌

[0006] (특허문헌 0001) 한국 공개 특허 제 10-2017-0034226 (공고)

## 발명의 내용

### 해결하려는 과제

[0007] 본 발명은 상기한 문제점을 해결하기 위하여 안출된 것으로서, 영상 인식 장치를 개시한다. 특히, 신경망을 이용하여 동영상 내 객체의 행동을 인식할 수 있는 영상 인식 장치를 개시한다.

### 과제의 해결 수단

[0008] 본 발명은 상기한 목적을 달성하기 위해 안출된 것으로서, 본 발명의 영상 인식 장치는 적어도 하나의 객체들을 포함하는 입력 영상에서 상기 객체별 행동에 관한 모션 정보를 포함하는 액션 스트림을 생성하는 스트림 생성부; 및 상기 생성된 액션 스트림 또는 상기 입력 영상에서 액션 스트림의 위치 관계를 나타내는 위치 정보를 입력 받아, 상기 객체의 행동을 분류하기 위한 지표로서 적어도 하나의 클래스 벡터를 출력하는 제1 인식기를 이용하여 상기 객체들의 행동을 인식하는 인식부;를 포함한다.

[0009] 본 발명에서 상기 스트림 생성부는 상기 입력 영상을 시간의 흐름에 따라 분할하여 복수개의 프레임 영상들을 생성하며, 상기 생성된 프레임 영상들에서 상기 객체 별로 상기 객체를 적어도 일부 포함하는 관심 영역을 검출하는 관심 영역 검출부;를 더 포함하고, 상기 검출된 관심 영역을 이용하여 상기 액션 스트림을 생성할 수 있다.

[0010] 본 발명에서 상기 인식부는 상기 제1 인식기를 이용하여 상기 객체 별로 생성되는 액션 스트림의 클래스 벡터를 미리 설정된 방법에 따라 합산하는 합산부;를 더 포함하고, 상기 합산된 클래스 벡터를 이용하여 상기 객체들의 행동을 인식할 수 있다.

[0011] 본 발명에서 상기 스트림 생성부는 인접하는 서로 다른 프레임 영상들에서 검출된 상기 관심 영역간 연결 점수를 산출하며, 상기 산출된 연결 점수를 고려하여 상기 서로 다른 프레임 영상들에서 검출된 상기 관심 영역을 연결하는 관심 영역 연결부;를 더 포함하고, 상기 연결된 관심 영역들을 이용하여 상기 객체 별로 상기 액션 스트림을 생성할 수 있다.

[0012] 본 발명에서 상기 스트림 생성부는 상기 프레임 영상들에서 검출된 상기 관심 영역을 프레임 영상 단위로 재배열하고, 상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림으로부터 상기 액션 스트림의 상호간 위치에 관한 위치 정보를 산출하는 위치 정보 산출부;를 더 포함할 수 있다.

[0013] 본 발명에서 상기 관심 영역 연결부는 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 특징 정보를 입력으로 하는 유사도 함수, 상기 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 오버랩 비율을 출력으로 하는 교차비 함수 및 상기 관심 영역들의 클래스 정보의 유사도 중 적어도 하나를 고려하여 연결 점수를 산출하는 연결 점수 산출부;를 더 포함하고, 상기 산출된 연결 점수를 이용하여 상기 관심 영역을 연결할 수 있다.

[0014] 본 발명에서 상기 관심 영역 연결부는 상기 관심 영역이 검출되지 않은 프레임 영상이 존재하는 경우, 상기 관심 영역이 검출되지 않은 프레임 영상에 인접한 프레임 영상들에서 검출된 관심 영역들의 특징 정보를 기반으로 상기 관심 영역이 검출되지 않은 프레임 영상 내의 관심 영역의 특징 정보를 추정하는 특징 정보 추정부;를 더 포함하고, 상기 추정된 특징 정보를 이용하여 상기 관심 영역을 연결할 수 있다.

[0015] 본 발명에서 상기 관심 영역 검출부는 상기 생성된 프레임 영상 내 상기 관심 영역이 위치하는 좌표를 나타내는 특징 정보를 산출하는 특징 정보 산출부;를 더 포함하고, 상기 산출된 특징 정보를 이용하여 상기 관심 영역을 검출할 수 있다.

[0016] 본 발명에서 상기 위치 정보 산출부는 상기 프레임 영상들에서 상기 객체 별로 검출된 상기 관심 영역 및 상기 관심 영역이 검출되지 않은 부분을 구분하여 구분하는 영역 구분부; 및 상기 구분된 관심 영역을 소정의 조합 방법으로 재배열하는 재배열부;를 더 포함하고, 상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림을 기반으로 상기 위치 정보를 산출할 수 있다.

- [0017] 본 발명에서 상기 특징 정보 산출부는 상기 생성된 프레임 영상들을 각각 분할하여 미리 결정된 크기의 격자셀을 생성하는 전처리부; 및 상기 생성된 프레임 영상들을 입력으로 하고, 상기 격자셀 내부에 중심을 가지고 상기 객체가 존재하는 확률을 나타내는 바운더리 셀의 중심 좌표 또는 상기 바운더리 셀 내 상기 객체가 존재하는 확률을 출력으로 하는 제2 인식기(Neural Network)를 이용하여 상기 바운더리 셀 각각의 중심 좌표 및 상기 바운더리 셀 내의 상기 객체가 존재하는 확률을 계산하는 계산부; 를 더 포함하고, 상기 확률 및 상기 중심 좌표가 계산된 바운더리 셀을 이용하여 상기 특징 정보를 산출할 수 있다.
- [0018] 본 발명에서 상기 관심 영역 검출부는 상기 프레임 영상들 내에 상기 객체를 중복하여 포함하는 바운더리 셀 중 상기 바운더리 셀 내에 상기 객체가 존재하는 확률이 기 설정된 임계치 이상인지 여부를 고려하여 상기 객체를 중복하여 포함하는 바운더리 셀의 일부를 제거하는 바운더리 셀 제거부; 를 더 포함하고, 제거되고 남은 바운더리 셀을 이용하여 상기 관심 영역을 검출할 수 있다.
- [0019] 본 발명에서 상기 클래스 벡터는 상기 객체의 행동을 분류하는 행동 목록과 상기 입력 영상 내 객체들의 행동이 상기 행동 목록에 해당할 확률을 나타내는 확률 정보를 포함하고, 상기 인식부는 상기 합산된 클래스 벡터의 상기 행동 목록별 상기 확률 정보를 이용하여 상기 객체들의 행동을 인식할 수 있다.
- [0020] 또한 상기한 목적을 달성하기 위하여 본 발명의 영상 인식 방법은 적어도 하나의 객체들을 포함하는 입력 영상에서 상기 객체별 행동에 관한 모션 정보를 포함하는 액션 스트림을 생성하는 단계; 및 상기 생성된 액션 스트림 또는 상기 입력 영상에서 상기 액션 스트림의 위치 관계를 나타내는 위치 정보를 입력 받아, 상기 객체의 행동을 분류하기 위한 지표로서 적어도 하나의 클래스 벡터를 출력으로 하는 제1 인식기를 이용하여 상기 객체들의 행동을 인식하는 단계; 를 포함한다.
- [0021] 본 발명에서 상기 생성하는 단계는 상기 입력 영상을 시간의 흐름에 따라 분할하여 복수개의 프레임 영상들을 생성하며, 상기 생성된 프레임 영상들에서 상기 객체 별로 상기 객체를 적어도 일부 포함하는 관심 영역을 검출하는 단계; 를 더 포함하고, 상기 검출된 관심 영역을 이용하여 상기 액션 스트림을 생성할 수 있다.
- [0022] 본 발명에서 상기 인식하는 단계는 상기 제1 인식기를 이용하여 상기 객체 별로 생성되는 액션 스트림의 클래스 벡터를 미리 설정된 방법에 따라 합산하는 단계; 를 더 포함하고, 상기 합산된 클래스 벡터를 이용하여 상기 객체들의 행동을 인식할 수 있다.
- [0023] 본 발명에서 상기 생성하는 단계는 인접하는 서로 다른 프레임 영상들에서 검출된 상기 관심 영역간 연결 점수를 산출하며, 상기 산출된 연결 점수를 고려하여 상기 서로 다른 프레임 영상들에서 검출된 상기 관심 영역을 연결하는 단계; 를 더 포함하고, 상기 연결된 관심 영역들을 이용하여 상기 객체 별로 상기 액션 스트림을 생성할 수 있다.
- [0024] 본 발명에서 상기 생성하는 단계는 상기 프레임 영상들에서 검출된 상기 관심 영역을 프레임 영상 단위로 재배열하고, 상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림으로부터 상기 액션 스트림의 상호간 위치에 관한 위치 정보를 산출하는 단계; 를 더 포함할 수 있다.
- [0025] 본 발명에서 상기 연결하는 단계는 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 특징 정보를 입력으로 하는 유사도 함수, 상기 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 오버랩 비율을 출력으로 하는 교차비 함수 및 상기 관심 영역들의 클래스 정보의 유사도 중 적어도 하나를 고려하여 연결 점수를 산출하는 단계; 를 더 포함하고, 상기 산출된 연결 점수를 이용하여 상기 관심 영역을 연결할 수 있다.
- [0026] 본 발명에서 상기 연결하는 단계는 상기 관심 영역이 검출되지 않은 프레임 영상이 존재하는 경우, 상기 관심 영역이 검출되지 않은 프레임 영상에 인접한 프레임 영상들에서 검출된 관심 영역들의 특징 정보를 기반으로 상기 관심 영역이 검출되지 않은 프레임 영상 내의 관심 영역의 특징 정보를 추정하는 단계; 를 더 포함하고, 상기 추정된 특징 정보를 이용하여 상기 관심 영역을 연결할 수 있다.
- [0027] 본 발명에서 상기 검출하는 단계는 상기 생성된 프레임 영상 내 상기 관심 영역이 위치하는 좌표를 나타내는 특징 정보를 산출하는 단계; 를 더 포함하고, 상기 산출된 특징 정보를 이용하여 상기 관심 영역을 검출할 수 있다.
- [0028] 본 발명에서 상기 위치 정보를 산출하는 단계는 상기 프레임 영상들에서 상기 객체 별로 검출된 상기 관심 영역 및 상기 관심 영역이 검출되지 않은 부분을 구분하는 단계; 및 상기 구분된 관심 영역을 소정의 조합 방법으로 재배열하는 단계; 를 더 포함하고, 상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림을 기반으로 상기 위치 정보를 산출할 수 있다.

[0029] 또한 본 발명은 컴퓨터에서 상기한 영상 인식 방법을 실행시키기 위한 컴퓨터에서 판독 가능한 기록매체에 저장된 컴퓨터 프로그램을 개시한다.

### 발명의 효과

[0030] 본 발명에 따르면, 영상 내 객체의 행동을 인식할 수 있다.

[0031] 특히, 동영상 단위의 영상 분석이 가능한 합성곱 신경망을 이용하여 영상을 인식할 수 있다.

### 도면의 간단한 설명

[0032] 도 1은 본 발명의 일 실시 예에 따른 영상 인식 장치의 블록도이다.

도 2는 도 1의 실시 예에서 스트림 생성부의 확대 블록도이다.

도 3은 도 2의 실시 예에서 관심 영역 검출부의 확대 블록도이다.

도 4는 도 3의 실시 예에서 특징 정보 산출부의 확대 블록도이다.

도 5는 관심 영역 검출부가 관심 영역을 검출하는 과정을 나타내는 참고도이다.

도 6은 관심 영역 검출부가 검출한 관심 영역들을 나타내는 예시도이다.

도 7은 도 2의 실시 예에서 관심 영역 연결부의 확대 블록도이다.

도 8은 관심 영역 연결부가 관심 영역들을 연결하는 과정을 나타내는 참고도이다.

도 9는 특징 정보 추정부가 관심 영역의 특징 정보를 추정하는 과정을 나타내는 예시도이다.

도 10은 도 2의 실시 예에서 위치 정보 산출부의 확대 블록도이다.

도 11은 본 발명의 영상 인식 장치가 수행하는 영상 인식 과정을 나타낸다.

도 12는 도 1의 실시 예에서 인식부의 확대 블록도이다.

도 13은 본 발명의 일 실시 예에 따른 영상 인식 방법의 흐름도이다.

도 14는 도 13의 실시 예에서 생성하는 단계의 확대 흐름도이다.

도 15는 도 14의 실시 예에서 검출하는 단계의 확대 흐름도이다.

도 16은 도 14의 실시 예에서 연결하는 단계의 확대 흐름도이다.

도 17은 도 14의 실시 예에서 위치 정보를 산출하는 단계의 확대 흐름도이다.

도 18은 도 13의 실시 예에서 인식하는 단계의 확대 흐름도이다.

### 발명을 실시하기 위한 구체적인 내용

[0033] 이하, 본 발명의 일 실시예를 첨부된 도면들을 참조하여 상세히 설명한다.

[0034] 첨부 도면을 참조하여 설명함에 있어, 동일하거나 대응하는 구성 요소는 동일한 도면번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.

[0035] 또한 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략할 수 있다.

[0036] 본 출원에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 용어를 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 이하에서 설명하는 각 단계는 하나 또는 여러 개의 소프트웨어 모듈로도 구비가 되거나 또는 각 기능을 담당하는 하드웨어로도 구현이 가능하며, 소프트웨어와 하드웨어가 복합된 형태로도 가능하다.

[0037] 각 용어의 구체적인 의미와 예시는 각 도면의 순서에 따라 이하 설명 한다.

[0038] 이하에서는 본 발명의 실시예에 따른 영상 인식 장치(10)의 구성을 관련된 도면을 상세히 설명한다.

[0039] 도 1은 본 발명의 일 실시 예에 따른 영상 인식 장치(10)의 블록도이다.

- [0040] 영상 인식 장치(10)는 스트림 생성부(100) 및 인식부(200)를 포함한다. 예를 들어, 본 발명의 영상 인식 장치(10)는 적어도 하나의 객체들을 포함하는 영상을 입력 받고, 입력 영상에서 액션 스트림을 생성하며, 생성된 액션 스트림을 입력으로 하는 제1 인식기를 이용하여 영상 내 객체의 행동을 인식할 수 있다.
- [0041] 본 발명의 영상 인식 장치(10)가 인식하는 영상은 정지 영상 외에도 시간의 흐름에 따라 영상 정보가 변화하는 비디오 영상(동영상)을 포함한다. 본 발명의 영상 인식 장치(10)는 종래의 영상 인식 기술과는 달리, 동영상 단위의 영상 분석을 수행하여 영상을 인식할 수 있다. 보다 상세하게는, 본 발명의 영상 인식 장치(10)는 영상 내 객체의 행동을 분류함으로써 영상을 인식할 수 있다.
- [0042] 본 발명의 영상 인식 장치(10)는 자율 주행 로봇과 같은 자율 주행 장치에 사용되어 자율 주행을 제어하는데 사용될 수 있고, CCTV와 같은 감시 장치에 사용되어 입력 영상의 자동 분석에 사용될 수 있다. 또한, 본 발명의 영상 인식 장치(10)는 홈 IOT(Internet Of Things)의 제어장치로 마련되어 사람의 행동을 자동으로 인식함으로써 가전 기기의 동작을 제어하는데 사용될 수 있다.
- [0043] 본 발명에서 영상 인식 장치(10)가 사용하는 제1 인식기 및 제2 인식기는 여러 계층을 가진 깊은 신경망 구조를 가지는 딥러닝 알고리즘을 기반으로 지도 학습(Supervised Learning)되는 신경망(Neural Network)으로 마련될 수 있다. 바람직하게는, 본 발명의 영상 인식 장치(10)가 사용하는 제1 인식기 및 제2 인식기는 지도 학습을 기반으로 학습되는 신경망 구조로서, 적어도 하나의 컨벌루션 레이어 및 풀링 레이어를 포함하는 합성곱 신경망(Convolutional Neural Network)으로 마련될 수 있다. 역전파 알고리즘을 기반으로 고안된 인공 신경망은 신경망의 계층이 많아지게 될 경우 과도한 깊이로 인한 학습 시간 지연과 등의 문제점이 밝혀지게 됨으로써 한동안 연구가 정체되었으나, 오버 피팅 문제가 dropout 등의 방법을 통해 해결됨으로써 알고리즘의 성능이 비약적으로 향상되었다.
- [0044] 영상 인식 장치(10)가 이용하는 대표적 딥러닝 구조인 합성곱 신경망(Convolutional Neural Network)은 사람의 시각 처리 과정을 모방하여 고안되었기 때문에 영상 및 이미지 처리에 적합한 딥러닝 알고리즘으로 평가 받고 있으며, 이미지를 추상화하여 표현할 수 있는 특징(feature)을 추출함으로써 영상 인식 분야에서 높은 성능을 나타내고 있다.
- [0045] 구체적으로, 합성곱 신경망은 입력 이미지에서 특징(feature)들을 계층적으로 추출하고, 추출된 특징(feature)들을 이용하여 특징맵을 형성할 수 있다. 특징맵상 분포되는 특징들은 재배치된 이미지 정보들을 포함하고, 이를 통하여 합성곱 신경망은 효과적으로 이미지를 분류할 수 있다. 합성곱 신경망은 컨벌루션 레이어를 통하여 추출된 특징(feature)에 활성화 함수(Activation Function)를 적용하고, 다시 풀링 레이어를 반복 배치함으로써, 컨벌루션 레이어에서 추출된 특징 값들의 사이즈를 재조정 할 수 있다.
- [0046] 본 발명의 영상 인식 장치(10)가 이용하는 합성곱 신경망은 기본적으로 합성곱 신경망 구조로 마련되고, 입력 영상의 형태, 크기 및 화소값 정보에 따라 신경망 내의 구조가 다르게 마련될 수 있다. 즉, 영상 인식 장치(10)가 이용하는 합성곱 신경망의 입력 레이어, 히든 레이어, 및 출력 레이어의 수, 각 레이어에 포함된 노드의 수, 각 노드들의 엣지에 적용되는 가중치, 러닝 레이트(Learning Rate)등은 영상 인식 목적에 따라 서로 다르게 설정될 수 있다.
- [0047] 또한, 영상 인식 장치(10)가 이용하는 합성곱 신경망은 신경망은 VGG-Net, GoogLeNet 및 ResNet 와 같은 구조가 적용된 신경망을 이용할 수 있다.
- [0048] 또한, 본 발명의 영상 인식 장치(10)가 사용하는 신경망들은 컨벌루션 레이어와 풀링 레이어(pooling layer)를 포함하는 CNN 구조에 풀리 커넥티드 레이어(fully-connected)가 연결된 구조로서, 과적합 문제가 발생하지 않는다면 계층의 깊이가 깊을수록 이미지 인식 정확도가 높아질 수 있다. 도 2를 참조하여 설명한다.
- [0049] 스트림 생성부(100)는 관심 영역 검출부(120), 관심 영역 연결부(140) 및 위치 정보 산출부(160)를 포함한다. 예를 들어, 스트림 생성부(100)는 적어도 하나의 객체들을 포함하는 입력 영상에서 상기 객체 별 행동에 관한 모션 정보를 포함하는 액션 스트림을 생성한다. 본 발명의 액션 스트림(Action Stream)은 영상 내 존재하는 객체 별로 생성되는 액션 튜브(Action Tube)를 신경망에 입력되기 위한 형태로 전처리하여 생성할 수 있다.
- [0050] 본 발명의 액션 튜브(Action Tube)는 입력 영상에서 객체의 적어도 일부를 포함하는 3차원(X, Y, Z) 영상 데이터의 집합으로 객체 별로 생성될 수 있다. 본 발명의 액션 튜브는 시간의 흐름에 따라 변하는 관심 영역들 내 화소값 변화량을 포함하여 객체의 행동 변화에 관한 모션 정보를 포함할 수 있다. 본 발명의 모션 정보는 시간의 흐름에 따라 변하는 관심 영역내의 화소값들의 변화량의 일종으로, 단위 시간에서 픽셀 별 화소값 변화량의

미분값으로 마련될 수도 있다.

- [0051] 예를 들어, 스트림 생성부(100)는 입력 영상을 분할하여 소정의 시간 간격을 가지는 프레임 영상들을 생성할 수 있고, 생성된 프레임 영상들에서 객체 별로 관심 영역을 검출하고, 검출된 관심 영역을 연결함으로써 액션 튜브를 생성하며, 생성된 액션 튜브를 전처리하여 액션 스트림을 생성할 수 있다.
- [0052] 소정의 시간 간격을 가지는 프레임 영상들에서 검출된 관심 영역을 연결하여 생성된 액션 튜브를 전처리하여 생성되는 액션 스트림은 시간의 흐름에 따라 변하는 화소값 변화량을 포함하는 액션 튜브를 전처리하여 생성되므로 시간의 흐름에 따른 객체의 행동 변화를 나타낼 수 있다.
- [0053] 본 발명의 스트림 생성부(100)에서 객체 별로 생성된 액션 스트림은 56\*56\*3\*16의 사이즈를 가질 수 있다. 스트림 생성부(100)에서 생성된 액션 스트림은 가로\*세로 사이즈가 56\*56인 관심 영역들을 연결하여 생성되고, RGB 표색계의 화소값(픽셀값)을 포함하며, 하나의 액션 스트림은 16개의 영상 프레임들을 포함할 수 있다. 스트림 생성부(100)에서 생성된 액션 스트림의 사이즈는 제1 인식기의 설정 방법에 따라 달라질 수 있다. 도 3을 참조하여 설명한다.
- [0054] 관심영역 검출부(120)는 특징 정보 산출부(122) 및 바운더리 셀 제거부(126)를 포함한다. 예를 들어, 관심 영역 검출부(120)는 입력 영상을 시간의 흐름에 따라 분할하여 복수개의 프레임 영상들을 생성하며, 상기 생성된 프레임 영상들에서 상기 객체 별로 구분되어 상기 객체를 적어도 일부 포함하는 관심 영역을 검출할 수 있다. 도 4를 참조하여 설명한다.
- [0055] 특징 정보 산출부(122)는 전처리부(123) 및 계산부(124)를 포함한다. 예를 들어, 특징 정보 산출부(122)는 생성된 프레임 영상 내 상기 관심 영역이 위치하는 좌표를 나타내는 특징 정보를 산출할 수 있다. 본 발명에서 특징 정보는 프레임 영상 내 검출된 관심 영역의 위치에 관한 좌표로서, 프레임 영상에서 관심 영역의 중심 좌표 또는 관심 영역의 경계에 위치하는 좌표로 마련될 수 있다. 도 5를 참조하여 설명한다.
- [0056] 전처리부(123)는 생성된 프레임 영상들을 각각 분할하여 도 5에 도시된 바와 같이, 미리 결정된 크기의 격자셀을 생성할 수 있다. 예를 들어, 전처리부(123)는 9\*6 또는 7\*7 크기의 격자셀들을 생성할 수 있다. 전처리부(123)에서 생성된 격자 셀들은 각각의 격자 셀들에 종속되는 바운더리 셀들을 포함할 수 있다.
- [0057] 계산부(124)는 생성된 프레임 영상들을 입력으로 하고, 상기 격자셀 내부에 중심을 가지고 상기 객체가 존재하는 확률을 나타내는 바운더리 셀의 중심 좌표 또는 상기 바운더리 셀 내 상기 객체가 존재하는 확률을 출력으로 하는 제2 인식기(Neural Network)를 이용하여 상기 바운더리 셀 각각의 중심 좌표 및 상기 바운더리 셀 내의 상기 객체가 존재하는 확률을 계산한다.
- [0058] 본 발명의 제2 인식기는 지도 학습을 기반으로 학습되는 신경망 구조로서, 적어도 하나의 컨벌루션 레이어 및 풀링 레이어를 포함하는 합성곱 신경망(Convolutional Neural Network)으로 마련될 수 있다. 바람직하게는, 계산부(124)가 이용하는 제2 인식기는 R-CNN(Regions with Convolutional Neural Network) 또는 Faster R-CNN으로 마련될 수 있다.
- [0059] 특징 정보 산출부(122)는 바운더리 셀 각각의 중심 좌표 및 상기 바운더리 셀 내의 상기 객체가 존재하는 확률이 계산된 바운더리 셀들을 이용하여 특징 정보를 산출할 수 있다. 예를 들어, 계산부(124)가 이용하는 제2 인식기는 특징(feature)을 추출하는 적어도 하나의 컨벌루션 레이어 및 상기 컨벌루션 레이어의 일단에 연결되어 상기 바운더리 셀의 중심 좌표와 상기 바운더리 셀 내의 상기 객체가 존재하는 확률을 계산하는 적어도 하나의 풀리 커넥티드 레이어를 포함할 수 있다. 또한, 본 발명의 계산부(140)가 이용하는 제2 인식기는 컨벌루션 레이어 및 풀리 커넥티드 레이어에 더하여 컨벌루션 레이어와 교대로 반복 배치되는 풀링 레이어를 더 포함할 수 있다.
- [0060] 본 발명의 일 실시 예에 따른 특징 정보 산출부(122)가 특징 정보를 산출하는 과정을 설명하면 다음과 같다. 특징 정보 산출부(122)는 적어도 하나의 객체들이 포함된 입력 영상을 분할하여 프레임 영상을 생성하고, 상기 프레임 영상에서 기 설정된 간격의 격자 셀들을 생성한다. 계산부(124)는 제2 인식기를 이용하여 격자셀 내부에 중심 좌표를 가지고, 상기 격자셀에 종속되는 임의의 수의 바운더리 셀들(127, 128, 129, 132, 133)을 생성함과 동시에, 바운더리 셀 내부에 포함되는 객체들이 존재하는 확률을 계산할 수 있다.
- [0061] 바운더리 셀 제거부(126)는 프레임 영상들 내에 상기 객체를 중복하여 포함하는 바운더리 셀 중 상기 바운더리 셀 내에 상기 객체가 존재하는 확률이 기 설정된 임계치 이상인지 여부를 고려하여 상기 객체를 중복하여 포함하는 바운더리 셀의 일부를 제거한다. 예를 들어, 바운더리 셀 제거부(126)는 계산부(124)가 생성한 바운더리

셀 내부에 객체가 존재할 확률이 모두 다른 바운더리 셀들(127, 128, 129, 132, 133, 134)중에서 내부에 객체가 존재할 확률이 가장 높은 바운더리 셀(132, 133)을 관심 영역으로 검출할 수 있다.

[0062] 이를 위하여 바운더리 셀 제거부(126)는 프레임 영상에서 존재하는 임의의 객체를 중복하여 포함하는 바운더리 셀 들 중에서 객체가 존재하는 확률이 기 설정된 임계치 이상인 바운더리 셀(132, 133)들만을 남기고 나머지 바운더리 셀들(127, 128, 129, 134)을 제거할 수 있다. 또한, 바운더리 셀 제거부(126)는 셀 내부에 병변이 존재할 확률이 기 설정된 임계치 이하인 바운더리 셀을 제거하기 위하여 NMS(Nom-maximal Suppression 비-최대값 억제) 알고리즘을 사용하여 오직 하나의 바운더리 셀만을 남길 수 있다. 도 6을 참조하여 설명한다.

[0063] 관심 영역 검출부(120)는 전술한 바와 같이, 특징 정보 산출부(122)에서 산출된 특징 정보들을 이용하여 입력 영상이 분할되어 생성된 프레임 영상들에서 사람 또는 사물을 포함하는 적어도 하나의 객체에 대한 이미지가 포함된 관심 영역을 검출할 수 있다. 도 7을 참조하여 설명한다.

[0064] 관심 영역 연결부(140)는 연결 점수 산출부(142) 및 특징 정부 추정부(144)를 포함한다. 예를 들어, 관심 영역 연결부(140)는 인접하는 서로 다른 프레임 영상들에서 검출된 상기 관심 영역간 연결 점수를 산출하며, 상기 산출된 연결 점수를 고려하여 상기 서로 다른 프레임 영상들에서 검출된 상기 관심 영역을 연결한다. 도 8을 참조하여 설명한다.

[0065] 예를 들어, 관심 영역 연결부(140)는 입력 영상이 분할되어 생성된 복수의 프레임 영상들(302, 304, 306, 308)에서 각각 검출된 제1 객체에 대한 관심 영역들(331, 333, 335, 337, 339, 341, 343)과 제2 객체에 대한 관심 영역들(332, 334, 336, 338, 340, 342, 344) 각각을 연결하여 제1 객체에 대한 제1 액션 튜브(361)과 제2 객체에 대한 제2 액션 튜브(362)를 생성할 수 있다.

[0066] 또한, 관심 영역 연결부(140)는 생성된 제1 액션 튜브(361)과 제2 객체에 대한 제2 액션 튜브(362)를 전처리하여 제1 객체 및 제2 객체에 대한 액션 스트림을 각각 생성할 수 있다. 즉, 본 발명의 액션 스트림들은 시간의 흐름에 따른 객체 별 행동에 대한 모션 정보들을 포함할 수 있음은 전술한 바와 같다. 본 발명에서 관심 영역 연결부(140)는 프레임 영상들 내에서 생성된 관심 영역들을 연결함에 있어, 객체 별로 연결하여 액션 튜브를 생성하고, 생성된 액션 튜브들을 전처리하여 객체 별 액션 스트림을 생성할 수 있다.

[0067] 연결 점수 산출부(142)는 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 특징 정보를 입력으로 하는 유사도 함수, 상기 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 오버랩 비율을 출력으로 하는 교차비 함수 및 상기 관심 영역들의 클래스 정보의 유사도 중 적어도 하나를 고려하여 연결 점수를 산출한다.

## 수학적식 1

$$sim(\Phi(r^t), \Phi(r^{t+1})) + \lambda ov(r^t, r^{t+1}) + \alpha$$

[0068]

[0069] 여기에서,  $sim()$ 은 프레임 영상들 각각에서 검출된 관심 영역들의 특징 정보를 입력으로 하여 0~1사이의 출력값을 가지는 유사도 함수,  $ov()$ 는 인접한 프레임 영상들에서 각각 검출된 관심 영역들의 오버랩 비율을 출력으로 하는 교차비 함수,  $\alpha$ 는 인접한 프레임 영상들에서 각각 검출된 관심 영역들의 행동 분류를 위한 클래스 정보

의 유사도,  $r^t$  및  $r^{t+1}$ 는 인접한 프레임 영상들의 인덱스 번호,  $\lambda$ 는 임의의 스칼라 값,  $\Phi(r^t)$ 는  $r^t$ 인덱스 번호의 프레임 영상에서 검출된 관심 영역의 특징 정보,  $\Phi(r^{t+1})$ 는  $r^{t+1}$ 인덱스 번호의 프레임 영상에 검출된 관심 영역의 특징 정보를 나타낸다. 본 발명의 프레임 영상에서 검출된 관심 영역의 특징 정보는 관심 영역이 해당 프레임 영상에서 위치하는 좌표 정보를 포함함은 전술한 바와 같다. 상기 클래스 정보는 관심 영역에서 검출된 객체의 행동을 분류하는 카테고리(목록)를 의미한다.

[0070] 또한, 상기 수학적식 1에서  $ov()$  함수는 IOU(Intersection Over Union) 값을 출력으로 하는 교차비 함수로서, 인접한 프레임 영상들 내에 각각 검출된 관심 영역들의 교집합 영역 및 합집합 영역을 입력으로 한다. 구체적으로는, 상기 교차비 함수는 상기 교집합 영역을 합집합 영역으로 나눈값을 출력으로 할 수 있다.

[0071] 본 발명의 일 실시 예에 따른 연결 점수 산출부(142)가 연결 점수(Link Score)를 산출하는 과정을 영상 프레임(302, 304)에서 설명하면 다음과 같다. 연결 점수 산출부(142)는 제1 영상 프레임(302)에서 검출된 관심 영역들

(331, 332)중 제1 객체에 관한 제1 관심 영역(331)과 제2 영상 프레임(304)내의 제1 관심 영역(333)의 연결 점수를 산출한다.

[0072] 또한, 연결 점수 산출부(142)는 제1 영상 프레임(302)내의 제1 관심 영역(331)과 제2 영상 프레임(304)내의 제2 관심 영역(334)의 연결 점수를 산출한다. 즉, 본 발명의 연결 점수 산출부(142)는 하나의 영상 프레임 내의 관심 영역이 인접한 다른 영상 프레임 내의 관심 영역과 연결될 수 있는 모든 경우의 연결(Link)에 대하여 연결 점수(Link Score)를 산출할 수 있다. 본 발명의 관심 영역 연결부(140)는 연결 점수가 가장 높게 나타나게 영상 프레임 내의 관심 영역들을 연결할 수 있다. 도 9를 참조하여 설명한다.

[0073] 특징 정보 추정부(144)는 관심 영역이 검출되지 않은 프레임 영상이 존재하는 경우, 상기 관심 영역이 검출되지 않은 프레임 영상(372)에 인접한 프레임 영상들(375)에서 검출된 관심 영역들(335, 336, 337, 338, 341, 342, 343, 344)의 특징 정보를 기반으로 상기 관심 영역이 검출되지 않은 프레임 영상(372) 내의 관심 영역의 특징 정보를 추정한다.

[0074] 예를 들어, 본 발명의 관심 영역 검출부(120)가 입력 영상을 분할하여 프레임 영상에서 각각 관심 영역을 검출함에 있어서, 영상 내 화질 등의 문제로 인하여 관심 영역을 제대로 추출하지 못하는 경우가 발생할 수 있다. 즉, 도 9에 도시된 바와 같이, 프레임 영상들(375)중에서 관심 영역이 검출되지 않은 프레임 영상(372)이 존재할 수 있다.

[0075] 이러한 경우, 특징 정보 추정부(144)는 관심 영역이 검출되지 않은 프레임 영상들(374, 376, 378)에 인접한 프레임 영상들(371, 373, 379)내의 관심 영역들의 특징 정보를 함수 입력으로 가지는 선형 함수를 생성하고, 생성된 선형함수를 이용하여 관심 영역이 검출되지 않은 프레임 영상(374, 376, 378) 내 관심 영역의 특징 정보를 추정할 수 있다.

[0076] 본 발명의 특징 정보는 프레임 영상 내에서 관심 영역의 좌표를 의미하고, 특징 정보 추정부(144)는 관심 영역이 검출되지 않은 프레임 영상들(374, 376, 378)에 인접한 프레임 영상들(371, 373, 379)내의 관심 영역들의 위치를 기반으로 관심 영역이 검출되지 않은 프레임 영상들(274, 276, 278)에서 검출된 관심 영역의 위치를 추정할 수 있다.

[0077] 즉, 특징 정보 추정부(144)는 관심 영역이 검출된 인접한 프레임 영상들(371, 373, 379)의 특징 정보들을 기반으로 보간 과정(Interpolation)을 수행함으로써 관심 영역이 검출되지 않은 프레임 영상들의 관심 영역의 좌표 정보인 특징 정보를 추정한다. 관심 영역 연결부(140)는 추정된 특징 정보를 이용하여 관심 영역이 검출되지 않은 프레임 영상 내 추정된 관심 영역을 생성하고, 추정된 관심 영역들을 인접한 프레임 영상 내의 관심 영역들과 연결할 수 있다.

[0078] 위치 정보 산출부(160)는 영역 구분부(162) 및 재배열 부(164)를 포함한다. 예를 들어, 위치 정보 산출부(160)는 프레임 영상들에서 검출된 상기 관심 영역을 프레임 영상 단위로 재배열하고, 상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림으로부터 상기 액션 스트림의 상호간 위치에 관한 위치 정보를 산출한다.

[0079] 예를 들어, 위치 정보 산출부(160)는 프레임 영상들에서 객체 별로 검출된 관심 영역들을 소정의 조합 방법으로 재배열하고, 재배열된 관심 영역들을 이용하여 조합 스트림을 생성하며, 생성된 조합 스트림을 이용하여 객체 별로 생성된 액션 스트림의 상호간 위치에 관한 위치 정보를 산출한다. 본 발명에서 위치 정보는 외형 정보가 아닌 프레임 영상 내에서 객체 별로 검출된 관심 영역들의 상호 위치에 대한 관계를 의미하며, 위치 정보를 산출하는 구체적인 방법은 후술한다.

[0080] 예를 들어, 위치 정보 산출부(160)는 제1 객체(사람), 제2 객체(사물)을 포함하는 입력 영상에서, 제1 객체에 관한 관심 영역들과 제2 객체에 관한 관심 영역들을 교차 배열 하거나 또는 밀집 배열 함으로서 조합 스트림을 생성하고, 생성된 조합 스트림을 이용하여 제1 객체에 관한 액션 스트림과 제2 객체에 관한 액션 스트림의 상호간 위치에 관한 위치 정보를 산출할 수 있다. 본 발명의 위치 정보 산출부(160)가 객체별로 생성된 액션 스트림의 위치 관계를 산출하기 위한 방법은 마스킹 또는 바이너리 처리에 제한되는 것은 아니며, 관심 영역들의 위치 관계를 입력하기 위한 기타 공지의 기술을 포함한다. 도 11을 참조하여 설명한다.

[0081] 영역 구분부(162)는 프레임 영상들에서 상기 객체 별로 검출된 상기 관심 영역 및 상기 관심 영역이 검출되지 않은 부분을 구분하여 마스킹 처리한다. 관심 영역 검출부(120)가 제1 객체(사람) 및 제2 객체(사물)을 포함하는 입력 영상에서, 제1 객체 및 제2 객체에 관한 관심 영역들을 검출한 경우를 예로 설명한다.

[0082] 이 경우, 본 실시예에서 영역 구분부(162)는 관심 영역이 검출되지 않은 부분을 구분하여 마스킹 처리하는 마스

킹부(미도시)를 포함할 수 있다.

- [0083] 예를 들어, 마스킹부(미도시)는 프레임 영상들에서 제1 객체에 관한 관심 영역들이 검출된 부분과 제1 객체에 관한 관심 영역이 아닌 부분을 2가지 화소값 정보로 구분하여 마스킹할 수 있다. 또한, 제2 객체에 관한 관심 영역들이 검출된 부분과 제2 객체에 관한 관심 영역이 아닌 부분을 2가지 화소값 정보로 구분하여 마스킹할 수 있다.
- [0084] 본 발명에서 영역 구분부(162)가 프레임 영상들에서 관심 영역들이 검출된 부분과 관심 영역들이 검출되지 않은 부분을 마스킹 하는 것은 프레임 영상을 바이너리(Binary) 처리하는 것으로 마련될 수 있다. 본 발명의 위치 정보 산출부(160)는 영역 구분부(162)에서 수행되는 마스킹 과정을 기반으로, 각 프레임 영상들에서 검출된 관심 영역들간의 위치 관계를 나타내는 위치 정보를 산출할 수 있다.
- [0085] 재배열 부(164)는 구분된 관심 영역을 소정의 조합 방법으로 재배열한다. 예를 들어, 재배열 부(164)는 제1 객체에 관한 마스킹된 관심 영역들(277)과 제2 객체에 관한 마스킹된 관심 영역들(279)을 소정의 조합 방법으로 조합할 수 있다. 재배열 부(164)가 구분된 관심 영역들을 조합하는 방법은 프레임별 정보를 고려하여 제1 객체에 관한 마스킹된 관심 영역-제2 객체에 관한 마스킹된 관심 영역-제1객체에 관한 마스킹된 관심 영역과 같이 객체 별 프레임 영상들을 교차 조합하는 방법을 포함한다.
- [0086] 또한, 재배열 부(164)가 구분된 관심 영역들을 조합하는 방법은 제1 객체 및 제2 객체 별 정보를 고려하여, 제1 객체에 관한 마스킹된 관심 영역들을 우선 배열 후, 제2 객체에 관한 마스킹된 관심 영역들을 후 배열하는 방법을 포함한다.
- [0087] 위치 정보 산출부(160)는 소정의 조합 방법으로 재배열된 관심 영역들을 이용하여 조합 스트림을 생성하며, 생성된 조합 스트림을 이용하여 객체 별로 생성된 액션 스트림의 상호간 위치에 관한 위치 정보를 산출할 수 있다. 본 발명에서 위치 정보 산출부(160)가 생성하는 조합 스트림은 제1 객체에 관한 마스킹된 관심 영역들이 연결된 스트림과 제2 객체에 관한 마스킹된 관심 영역들이 연결된 스트림 쌍을 포함하는 페어와이즈 스트림(Pairwise Stream)으로 마련될 수 있다.
- [0088] 본 발명의 페어와이즈 스트림은 영상 내 객체별 액션 스트림에 포함된 외형 정보를 제외한 객체별 액션 스트림의 위치 관계를 나타내는 위치 정보를 포함한다. 즉, 객체위치 정보 산출부(160)는 전술한 조합 스트림을 이용하여 제1 객체에 관한 액션 스트림과 제2 객체에 관한 액션 스트림의 위치 관계를 나타내는 위치 정보를 산출할 수 있으며, 이를 통하여 본 발명의 영상 인식 장치(10)는 영상 내 객체 별 위치 관계를 기반으로 객체의 행동을 인식 할 수 있다.
- [0089] 인식부(200)는 합산부(220) 및 비교부(240)를 포함한다. 예를 들어 인식부(200)는 생성된 액션 스트림 또는 상기 액션 스트림의 위치 정보를 입력으로 하고, 상기 객체의 행동을 분류하기 위한 지표로서 적어도 하나의 클래스 벡터를 출력으로 하는 제1 인식기를 이용하여 상기 객체들의 행동을 인식할 수 있다. 본 발명의 인식부(200)가 이용하는 클래스 벡터는 상기 객체의 행동을 분류하는 행동 목록과 상기 입력 영상 내 객체들의 행동이 상기 행동 목록에 해당할 확률을 나타내는 확률 정보를 포함한다. 인식부(200)는 상기 합산된 클래스 벡터의 상기 행동 목록별 상기 확률 정보를 이용하여 상기 객체들의 행동을 인식할 수 있다.
- [0090] 합산부(220)는 제1 인식기를 이용하여 상기 객체 별로 구분되는 액션 스트림 각각의 클래스 벡터를 미리 설정된 방법에 따라 합산하는 합산한다. 예를 들어, 합산부(220)는 스트림 생성부(100)에서 객체 별로 생성된 액션 스트림 및 객체 별 생성된 액션 스트림의 위치 관계를 포함하는 조합 스트림을 입력으로 하는 제1 인식기의 출력값을 행동 목록 별로 미리 설정된 방법에 따라 합산할 수 있다.
- [0091] 예를 들어, 합산부(220)는 제1 인식기에 제1 객체(사람)에 관한 액션 스트림을 입력 시 출력되는 제1 객체(사람)의 행동을 분류하기 위한 지표인 제1 클래스 벡터, 제1 인식기에 제2 객체(사물)에 관한 액션 스트림을 입력 시 출력되는 제2 객체(사물)의 행동을 분류하기 위한 지표인 제2 클래스 벡터 및 제1 인식기에 사람 또는 사물 별로 생성된 액션 스트림의 위치 관계를 포함하는 조합 스트림을 입력 시에 출력되는 제3 클래스 벡터에 서로 다른 가중치를 적용하여 합산하거나, 1/3과 같이 동일한 가중치를 적용하여 합산할 수 있다. 합산부(220)가 제1 클래스 벡터, 제2 클래스 벡터 및 제3 클래스 벡터에 동일한 가중치인 1/3을 적용하여 합산하는 것은 제1 클래스 벡터, 제2 클래스 벡터 및 제3 클래스 벡터의 평균을 구하는 과정으로 마련될 수 있다.
- [0092] 비교부(240)는 합산된 클래스 벡터 내 행동을 분류하는 행동 목록 별 확률 정보들을 미리 마련된 정답 데이터 셋(UCF-101 Detection Data Set)에 저장된 객체의 행동 목록 별 확률 정보와 비교한다. 예를 들어, 비교부(240)는 이미 알려진 객체의 행동 목록별 확률 정보들이 저장된 데이터 셋을 이용하여 합산부(220)에서 합산된

클래스 벡터가 나타내는 객체의 행동이 일치하는지 여부를 비교함으로써, 영상 내 객체의 행동을 정확하게 인식할 수 있다.

[0093] 도 2는 도 1의 실시 예에서 스트림 생성부의 확대 블록도이다.

[0094] 스트림 생성부(100)는 관심 영역 검출부(120), 관심 영역 연결부(140) 및 위치 정보 산출부(160)를 포함한다. 예를 들어, 스트림 생성부(100)는 입력 영상에 포함된 객체의 행동에 관한 모션 정보를 객체 별로 구분하여 포함하는 액션 스트림을 생성한다. 액션 스트림은 시간의 흐름에 따라 각 영상 프레임들에서 검출된 관심 영역들을 연결하여 생성된 것이다.

[0095] 객체의 행동에 관한 모션 정보는 시간의 흐름에 따른 관심 영역들의 위치 변화량 및 관심 영역 내의 화소값 변화량을 모두 포함할 수 있지만, 프레임 영상 내 관심 영역들의 위치 변화량을 제외한 관심 영역 내의 화소값 변화량만을 포함할 수 있다.

[0096] 예를 들어, 영상 프레임 내에 위치하는 관심 영역들의 위치와 화소값 정보들은 시간의 흐름에 따라 달라지므로, 시간의 흐름에 따라 영상 프레임 내에서 변하는 관심 영역들의 위치 변화량, 시간의 흐름에 따라 변하는 관심 영역내의 화소값 변화량등은 그 자체로 객체의 행동 변화에 관한 모션 정보를 구성할 수 있다. 다만, 바람직하게는, 본 발명의 모션 정보는 시간의 흐름에 따라 변하는 관심 영역내의 화소값 변화량만을 포함할 수 있음은 전술한 바와 같다.

[0097] 다만, 본 발명의 영상 인식 장치(10)는 페어 와이즈 스트림을 이용하기 때문에, 프레임 영상내 관심 영역들의 위치 변화량까지 모두 고려하여 영상 내 객체들의 행동을 인식할 수 있음은 후술한다.

[0098] 관심 영역 검출부(120)는 입력 영상을 시간의 흐름에 따라 분할하여 복수개의 프레임 영상들을 생성하며, 상기 생성된 프레임 영상들에서 상기 객체 별로 구분되어 상기 객체를 적어도 일부 포함하는 관심 영역을 검출할 수 있다. 관심 영역 검출부(120)는 입력 영상들을 16개 프레임 단위로 분할할 수 있고, 각각의 분할된 프레임 영상에서 객체에 관한 이미지를 주로 포함하는 관심 영역들을 검출할 수 있다.

[0099] 관심 영역 연결부(140)는 시간의 흐름에 따라 분할된 프레임 영상들에서 검출된 관심 영역들을 서로 연결하여 객체 별 액션 튜브들을 생성한다. 관심 영역 연결부(140)는 생성된 액션 튜브들을 제1 인식기의 입력 형태에 맞게 전처리하여 액션 스트림을 생성할 수 있음은 전술한 바와 같다.

[0100] 위치 정보 산출부(160)는 위치 정보 산출부(160)는 프레임 영상들에서 검출된 상기 관심 영역을 프레임 영상 단위로 재배열하고, 상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림으로부터 상기 액션 스트림의 상호간 위치에 관한 위치 정보를 산출한다. 본 발명의 영상 인식 장치(10)는 위치 정보 산출부(160)가 산출한 위치 정보를 이용하여 영상 내 객체 별로 생성된 액션 스트림의 위치 관계를 이용하여 영상 내 객체의 행동을 인식할 수 있다.

[0101] 도 3은 도 2의 실시 예에서 관심 영역 검출부의 확대 블록도이다.

[0102] 관심 영역 검출부(120)는 특징 정보 산출부(122) 및 바운더리 셀 제거부(126)를 포함한다. 특징 정보 산출부(122)는 프레임 영상 내 검출될 관심 영역의 위치에 관한 좌표로서, 특징 정보를 산출한다. 특징 정보 산출부(122)는 영상 내에서 관심 영역을 검출하기 위하여 합성곱 신경망을 기반으로 하는 YOLO(You Only Look Once), Fast R-CNN등을 포함하는 비전 인식 알고리즘을 사용할 수 있다.

[0103] 바운더리 셀 제거부(126)는 프레임 영상들 내에 상기 객체를 중복하여 포함하는 바운더리 셀 중 상기 바운더리 셀 내에 상기 객체가 존재하는 확률이 기 설정된 임계치 이상인지 여부를 고려하여 상기 객체를 중복하여 포함하는 바운더리 셀의 일부를 제거한다. 바운더리 셀 제거부(126)가 바운더리 셀을 제거 하고 남은 바운더리 셀을 이용하여 관심 영역을 검출하는 방법은 전술한 바와 같다.

[0104] 도 4는 도 3의 실시 예에서 특징 정보 산출부의 확대 블록도이다.

[0105] 특징 정보 산출부(122)는 전처리부(123) 및 계산부(124)를 포함한다. 전처리부(123)는 생성된 프레임 영상들을 각각 분할하여 미리 결정된 크기의 격자셀을 생성할 수 있고, 생성된 격자 셀들을 각각의 격자셀들에 종속되는 바운더리 셀들을 포함함은 전술한 바와 같다.

[0106] 계산부(124)는 생성된 프레임 영상들을 입력으로 하고, 상기 격자셀 내부에 중심을 가지고 상기 객체가 존재하는 확률을 나타내는 바운더리 셀의 중심 좌표 또는 상기 바운더리 셀 내 상기 객체가 존재하는 확률을 출력으로 하는 제2 인식기(Neural Network)을 이용하여 상기 바운더리 셀 각각의 중심 좌표 및 상기 바운더리 셀 내의 상

기 객체가 존재하는 확률을 계산한다.

- [0107] 본 발명의 제2 인식기는 컨벌루션 특징을 추출하는 적어도 하나의 컨벌루션 레이어 및 컨벌루션 레이어의 일단에 연결되어 바운더리 셀의 중심 좌표와 바운더리 셀 내의 객체가 존재하는 확률을 계산하는 폴리 커넥티드 레이어를 포함할 수 있다.
- [0108] 도 5는 관심 영역 검출부가 관심 영역을 검출하는 과정을 나타내는 참고도이다.
- [0109] 전처리부(123)에서 생성된 격자 셀들에 종속되는 바운더리 셀들(127, 128, 129, 132, 133, 134)은 격자 셀들에 종속되어 내부에 객체에 관한 부분이 포함될 확률을 나타낸다. 관심 영역 검출부(120)는 생성된 바운더리 셀들 중 객체가 존재하는 확률이 기 설정된 임계치 이상이 아닌 바운더리 셀들을 모두 제거하고, 제거하고 남은 바운더리 셀들을 이용하여 관심 영역을 검출할 수 있음은 전술한 바와 같다.
- [0110] 도 5에 도시된 바와 같이, 관심 영역 검출부(120)는 바운더리 셀들이 나타내는 확률을 고려하여 99%의 확률을 가지는 제1 객체(사람)에 관한 관심 영역 및 67%의 확률을 가지는 제2 객체(사물)에 관한 관심 영역을 검출할 수 있다. 관심 영역 검출부(120)는 관심 영역을 검출함에 이용하는 관심 영역들의 확률 정보는 전술한 IOU(Intersection Over Union) 를 포함한다.
- [0111] 도 6은 관심 영역 검출부가 검출한 관심 영역들을 나타내는 예시도이다.
- [0112] 관심 영역 검출부(120)는 도 6에 도시된 바와 같이, 영상 내 포함되는 객체 별 관심 영역들을 검출할 수 있다. 예를 들어, 입력 영상 내 사람들이 존재하는 경우, 제1 객체(사람) 및 제2 객체(사람)에 관한 관심 영역들을 검출할 수 있고, 입력 영상 내 사람 및 사물이 존재하는 경우, 제1 객체(사람) 및 제2 객체(사물)에 관한 관심 영역들을 검출할 수 있다.
- [0113] 도 7은 도 2의 실시 예에서 관심 영역 연결부의 확대 블록도이다.
- [0114] 관심 영역 연결부(140)는 연결 점수 산출부(142) 및 특징 정보 추정부(144)를 포함한다. 예를 들어, 연결 점수 산출부(142)는 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 특징 정보를 입력으로 하는 유사도 함수, 상기 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 오버랩 비율을 출력으로 하는 교차비 함수 및 상기 관심 영역들의 클래스 정보의 유사도 중 적어도 하나를 고려하여 연결 점수를 산출할 수 있다. 본 발명의 연결 점수 산출부(142)가 수학식 1을 이용하여 연결 점수를 산출하는 구체적인 방법은 전술한 바와 같으므로 생략한다.
- [0115] 특징 정보 추정부(144)는 관심 영역이 검출되지 않은 프레임 영상이 존재하는 경우, 상기 관심 영역이 검출되지 않은 프레임 영상에 인접한 프레임 영상들에서 검출된 관심 영역들의 특징 정보를 기반으로 상기 관심 영역이 검출되지 않은 프레임 영상 내의 관심 영역의 특징 정보를 추정한다. 본 발명의 특징 정보 추정부(144)가 보간 방법을 사용하여 관심 영역이 검출되지 않은 프레임 영상 내의 관심영역의 위치를 추정하고, 추정된 관심 영역의 위치를 이용하여 관심 영역들을 연결하는 구체적인 방법은 전술한 바와 같다.
- [0116] 도 8은 관심 영역 연결부가 관심 영역들을 연결하는 과정을 나타내는 참고도이다.
- [0117] 예를 들어, 관심 영역 연결부(140)는 입력 영상이 분할되어 생성된 복수의 프레임 영상들(302, 304, 306, 308)에서 각각 검출된 제1 객체에 대한 관심 영역들(331, 333, 335, 337, 339, 341, 343)과 제2 객체에 대한 관심 영역들(332, 334, 336, 338, 340, 342, 344) 각각을 연결하여 제1 객체에 대한 제1 액션 튜브(361)과 제2 객체에 대한 제2 액션 튜브(362)를 생성할 수 있다.
- [0118] 또한, 관심 영역 연결부(140)는 생성된 제1 액션 튜브(361)과 제2 객체에 대한 제2 액션 튜브(362)를 전처리하여 제1 객체 및 제2 객체에 대한 액션 스트림을 각각 생성할 수 있음은 전술한 바와 같다.
- [0119] 도 9는 특징 정보 추정부가 관심 영역의 특징 정보를 추정하는 과정을 나타내는 예시도이다.
- [0120] 특징 정보 추정부(144)는 관심 영역이 검출되지 않은 프레임 영상이 존재하는 경우, 상기 관심 영역이 검출되지 않은 프레임 영상에 인접한 프레임 영상들에서 검출된 관심 영역들의 특징 정보를 기반으로 상기 관심 영역이 검출되지 않은 프레임 영상 내의 관심 영역의 특징 정보를 추정한다. 특징 정보 추정부(144)가 보간 과정을 이용하여 관심 영역이 검출되지 않은 프레임 영상 내에서 특징 정보를 추정하는 과정은 전술한 바와 같으므로 생략한다.
- [0121] 도 10은 도 2의 실시 예에서 위치 정보 산출부의 확대 블록도이다.

- [0122] 위치 정보 산출부(160)는 영역 구분부(162) 및 재배열 부(164)를 포함한다. 예를 들어, 위치 정보 산출부(160)는 프레임 영상들에서 검출된 상기 관심 영역을 프레임 영상 단위로 재배열하고, 상기 재배열된 관심 영역을 연결하여 생성된 조합 스트림으로부터 상기 액션 스트림의 상호간 위치에 관한 위치 정보를 산출한다. 위치 정보 산출부(160)과 객체 별 생성된 액션 스트림의 위치 관계를 산출하는 구체적인 과정은 전술한 바와 같으므로 생략한다.
- [0123] 도 11은 본 발명의 영상 인식 장치가 수행하는 영상 인식 과정을 나타낸다.
- [0124] 본 발명의 영상 인식 장치(10)는 적어도 하나의 동적 객체들을 포함하는 입력 영상에서 객체 별 액션 스트림(272, 274)를 생성하고, 동시에 객체 별로 생성된 액션 스트림을 조합하여 조합 스트림(279)를 생성한다. 본 발명의 조합 스트림은 객체 별 생성된 액션 스트림(272, 274)의 위치 관계를 나타내는 전술한 바와 같다.
- [0125] 영상 인식 장치(10)는 객체 별 생성된 액션 스트림들(272, 274) 및 조합 스트림(페어와이즈 스트림, 279)를 제1 인식기에 입력하고, 출력되는 각각의 클래스 벡터들을 미리 설정된 방법으로 합산하며, 합산된 클래스 벡터들을 이용하여 영상 내 객체들의 행동을 인식할 수 있다.
- [0126] 도 12는 도 1의 실시 예에서 인식부의 확대 블록도이다.
- [0127] 인식부(200)는 합산부(220) 및 비교부(240)를 포함한다. 예를 들어 인식부(200)는 생성된 액션 스트림 또는 상기 액션 스트림의 위치 정보를 입력으로 하고, 상기 객체의 행동을 분류하기 위한 지표로서 적어도 하나의 클래스 벡터를 출력으로 하는 제1 인식기를 이용하여 상기 객체들의 행동을 인식할 수 있다.
- [0128] 도 13은 본 발명의 일 실시 예에 따른 영상 인식 방법의 흐름도이다.
- [0129] 영상 인식 장치(10)가 수행하는 영상 인식 방법(10)은 시계열적으로 수행되는 하기의 단계들을 포함한다.
- [0130] S100에서, 스트림 생성부(100)는 적어도 하나의 객체들을 포함하는 입력 영상에서 상기 객체별 행동에 관한 모션 정보를 포함하는 액션 스트림을 생성한다. 스트림 생성부(100)는 입력 영상 내에 복수의 객체들이 포함되는 경우, 객체 별로 액션 스트림을 생성할 수 있다. 또한 스트림 생성부(100)는 객체 별로 생성된 액션 스트림의 위치 관계를 나타내는 위치 정보를 포함하는 조합 스트림을 생성함으로써, 객체 별로 생성된 액션 스트림의 위치 관계를 이용하여 영상 내 객체의 행동을 인식할 수 있음은 전술한 바와 같다.
- [0131] S200에서, 인식부(200)는 생성된 액션 스트림 또는 상기 액션 스트림의 위치 정보를 입력으로 하고, 상기 객체의 행동을 분류하기 위한 지표로서 적어도 하나의 클래스 벡터를 출력으로 하는 제1 인식기를 이용하여 상기 객체들의 행동을 인식할 수 있다. 예를 들어, 인식부(200)가 객체 별로 생성된 액션 스트림 및 객체 별로 생성된 액션 스트림의 위치 관계를 나타내는 조합 스트림을 각각 입력으로 하는 제1 인식기의 출력을 합산하여 영상 내 객체들의 행동을 인식하는 구체적인 방법은 전술한 바와 같다.
- [0132] 도 14는 도 13의 실시 예에서 생성하는 단계의 확대 흐름도이다.
- [0133] S120에서, 관심 영역 검출부(120)는 입력 영상을 시간의 흐름에 따라 분할하여 복수개의 프레임 영상들을 생성하며, 상기 생성된 프레임 영상들에서 상기 객체 별로 구분되어 상기 객체를 적어도 일부 포함하는 관심 영역을 검출한다. 관심 영역 검출부(120)가 복수의 프레임 영상 들에서 관심 영역을 검출하는 구체적인 방법은 전술한 바와 같으므로 생략한다.
- [0134] S140에서, 관심 영역 연결부(140)는 서로 다른 프레임 영상들에서 검출된 상기 관심 영역간 연결 점수를 산출하며, 상기 산출된 연결 점수를 고려하여 상기 서로 다른 프레임 영상들에서 검출된 상기 관심 영역을 연결한다. 관심 영역 연결부(140)는 임의 프레임 영상 내 검출된 하나의 관심 영역과 다른 프레임 영상 내에 검출된 모든 관심 영역들과의 연결 점수를 고려하여 관심 영역을 연결할 수 있다.
- [0135] S160에서, 위치 정보 산출부(160)는 프레임 영상 들에서 검출된 객체 별 관심 영역을 프레임 영상 단위로 재배열하고, 재배열된 관심 영역을 연결하여 생성된 조합 스트림으로부터 액션 스트림의 상호간 위치 관계를 나타내는 위치 정보를 산출할 수 있다. 위치 정보 산출부(160)가 프레임 영상들에서 검출된 관심 영역들을 재배열 하는 방법은 전술한 바와 같으므로 생략한다.
- [0136] 도 15는 도 14의 실시 예에서 검출하는 단계의 확대 흐름도이다.
- [0137] S122에서, 특징 정보 산출부(122)는 생성된 프레임 영상 내 상기 관심 영역이 위치하는 좌표를 나타내는 특징 정보를 산출할 수 있다. 본 발명에서 특징 정보는 프레임 영상 내 검출될 관심 영역의 위치에 관한 좌표로서,

프레임 영상에서 관심 영역의 중심 좌표로 마련될 수 있음은 전술한 바와 같다.

- [0138] S126에서, 바운더리 셀 제거부(126)는 프레임 영상들 내에 상기 객체를 중복하여 포함하는 바운더리 셀 중 상기 바운더리 셀 내에 상기 객체가 존재하는 확률이 기 설정된 임계치 이상인지 여부를 고려하여 상기 객체를 중복하여 포함하는 바운더리 셀의 일부를 제거한다. 관심 영역 검출부(120)는 바운더리 셀 제거부(126)에서 제거하고 남은 바운더리 셀들을 이용하여 관심 영역들을 검출한다.
- [0139] 도 16은 도 14의 실시 예에서 연결하는 단계의 확대 흐름도이다.
- [0140] S142에서, 연결 점수 산출부(142)는 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 특징 정보를 입력으로 하는 유사도 함수, 상기 인접한 프레임 영상들 각각에서 검출된 상기 관심 영역들의 오버랩 비율을 출력으로 하는 교차비 함수 및 상기 관심 영역들의 클래스 정보의 유사도 중 적어도 하나를 고려하여 연결 점수를 산출한다. 연결 점수 산출부(142)가 인접한 프레임 영상들 각각에서 검출된 관심 영역들의 연결 점수를 산출하는 구체적인 방법은 전술한 바와 같으므로 생략한다.
- [0141] S144에서, 특징 정부 추정부(144)는 입력 영상이 분할되어 생성된 프레임 영상들 중에서 관심 영역이 검출되지 않은 프레임 영상이 존재하는 경우, 관심 영역이 검출되지 않은 프레임 영상에 인접한 프레임 영상 내에서 검출된 관심 영역들의 좌표를 이용하여 관심 영역이 검출되지 않은 프레임 영상 내 관심 영역의 위치를 추정할 수 있다.
- [0142] 도 17은 도 14의 실시 예에서 위치 정보를 산출하는 단계의 확대 흐름도이다.
- [0143] S162에서, 영역 구분부(162)는 프레임 영상들에서 상기 객체 별로 검출된 상기 관심 영역 및 상기 관심 영역이 검출되지 않은 부분을 구분한다. 예를 들어, 영역 구분부(162)는 프레임 영상 내 검출된 관심 영역들의 위치 정보를 알기 위해서 2가지 화소값 정보를 이용하여 관심 영역과 관심 영역이 아닌 부분을 구분하여 마스킹 할 수 있다.
- [0144] S164에서, 재배열 부(164)는 재배열 부(164)는 구분된 관심 영역을 소정의 조합 방법으로 재배열한다. 예를 들어, 재배열 부(164)는 제1 객체에 관한 마스킹된 관심 영역들(277)과 제2 객체에 관한 마스킹된 관심 영역들(279)을 소정의 조합 방법으로 조합할 수 있다. 재배열 부(164)가 마스킹된 관심 영역을 조합하는 방법은 제1 객체에 관한 관심 영역과 제2 객체에 관한 관심 영역을 교차로 조합하는 방법, 제1 객체에 관한 관심 영역들을 선배치하고, 제2 객체에 관한 관심 영역들을 후배치 하는 방법을 포함할 수 있다.
- [0145] 영상 인식 장치(10)는 객체 별로 검출된 관심 영역들을 구분하고, 구분된 관심 영역들을 재배열하여 조합 스트림을 생성함으로써, 영상 내 객체 들의 위치 관계를 이용하여 객체들의 행동을 인식할 수 있다.
- [0146] 도 18은 도 13의 실시 예에서 인식하는 단계의 확대 흐름도이다.
- [0147] S220에서, 합산부(220)는 제1 인식기를 이용하여 상기 객체 별로 구분되는 액션 스트림 각각의 클래스 벡터를 미리 설정된 방법에 따라 행동 목록 별로 합산한다. 예를 들어, 합산부(220)는 객체 별 액션 스트림 및 객체 별로 생성된 액션 스트림의 위치 관계를 나타내는 조합 스트림을 입력으로 하는 제1 인식기의 출력인 클래스 벡터들에 가중치를 적용하여 합산할 수 있다.
- [0148] S240에서, 비교부(240)는 비교부(240)는 합산된 클래스 벡터 내 행동을 분류하는 행동 목록 별 확률 정보들을 미리 마련된 정답 데이터 셋(UCF-101 Detection Data Set)에 저장된 객체의 행동 목록 별 확률 정보와 비교한다.
- [0149] 상기 설명된 본 발명의 일 실시예의 방법의 전체 또는 일부는, 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 기록 매체의 형태(또는 컴퓨터 프로그램 제품)로 구현될 수 있다. 여기에서, 컴퓨터 판독 가능 매체는 컴퓨터 저장 매체(예를 들어, 메모리, 하드디스크, 자기/광학 매체 또는 SSD(Solid-State Drive) 등)를 포함할 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다.
- [0150] 또한, 본 발명의 일 실시예에 따르는 방법의 전체 또는 일부는 컴퓨터에 의해 실행 가능한 명령어를 포함하며, 컴퓨터 프로그램은 프로세서에 의해 처리되는 프로그래밍 가능한 기계 명령어를 포함하고, 고레벨 프로그래밍 언어(High-level Programming Language), 객체 지향 프로그래밍 언어(Object-oriented Programming Language), 어셈블리 언어 또는 기계 언어 등으로 구현될 수 있다.
- [0151] 본 명세서에서의 부(means) 또는 모듈(Module)은 본 명세서에서 설명되는 각 명칭에 따른 기능과 동작을 수행할

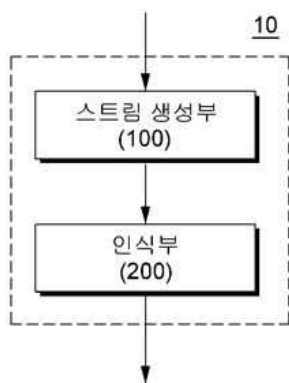
수 있는 하드웨어를 의미할 수도 있고, 특정 기능과 동작을 수행할 수 있는 컴퓨터 프로그램 코드를 의미할 수도 있고, 또는 특정 기능과 동작을 수행시킬 수 있는 컴퓨터 프로그램 코드가 탑재된 전자적 기록 매체, 예를 들어 프로세서 또는 마이크로 프로세서를 의미할 수 있다. 다시 말해, 부(means) 또는 모듈(Module)은 본 발명의 기술적 사상을 수행하기 위한 하드웨어 및/또는 상기 하드웨어를 구동하기 위한 소프트웨어의 기능적 및/또는 구조적 결합을 의미할 수 있다.

[0152] 따라서 본 발명의 일 실시예에 따르는 방법은 상술한 바와 같은 컴퓨터 프로그램이 컴퓨팅 장치에 의해 실행됨으로써 구현될 수 있다. 컴퓨팅 장치는 프로세서와, 메모리와, 저장 장치와, 메모리 및 고속 확장포트에 접속하고 있는 고속 인터페이스와, 저속 버스와 저장 장치에 접속하고 있는 저속 인터페이스 중 적어도 일부를 포함할 수 있다. 이러한 성분들 각각은 다양한 버스를 이용하여 서로 접속되어 있으며, 공통 머더보드에 탑재되거나 다른 적절한 방식으로 장착될 수 있다.

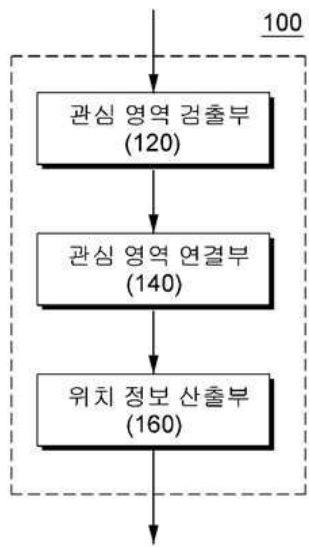
[0153] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다. 따라서, 본 발명에 개시된 실시예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구 범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

## 도면

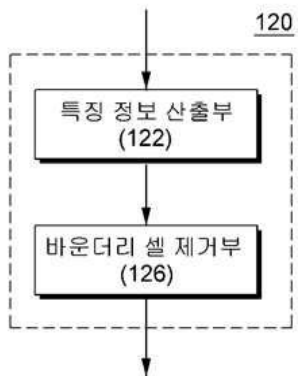
### 도면1



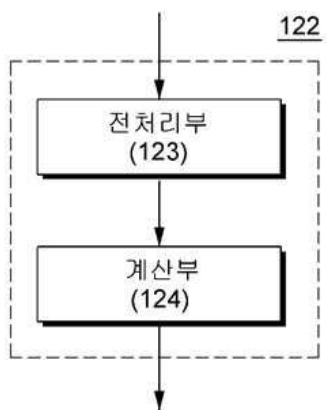
도면2



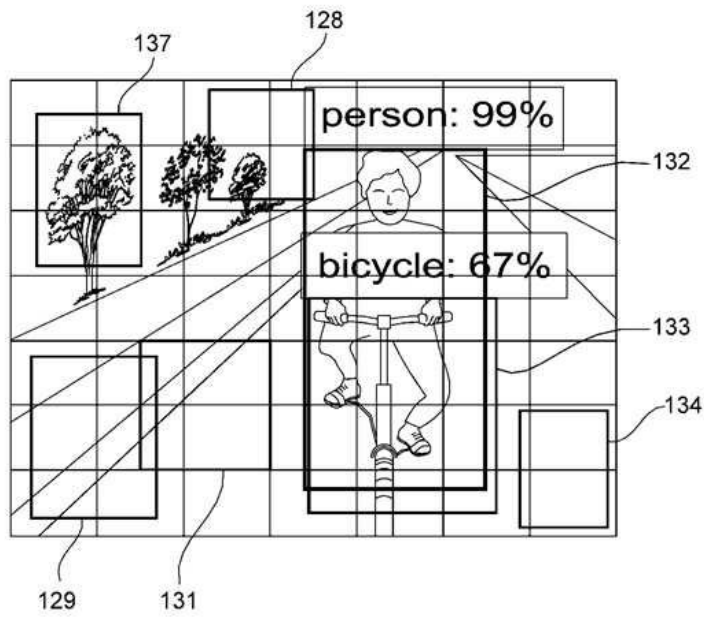
도면3



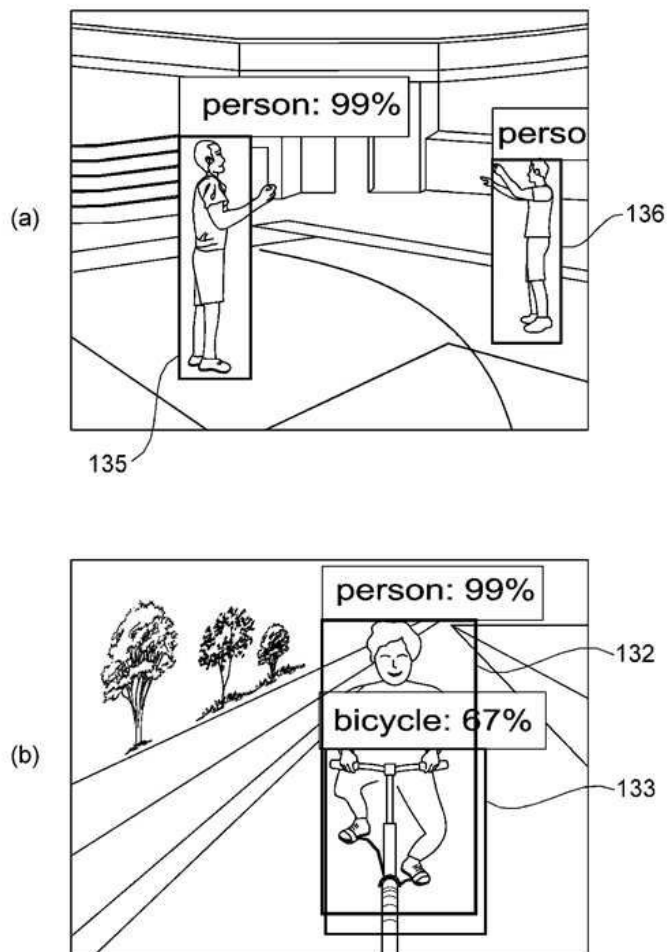
도면4



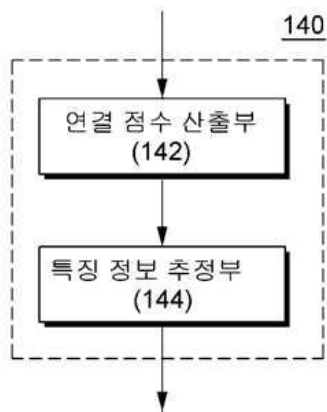
도면5



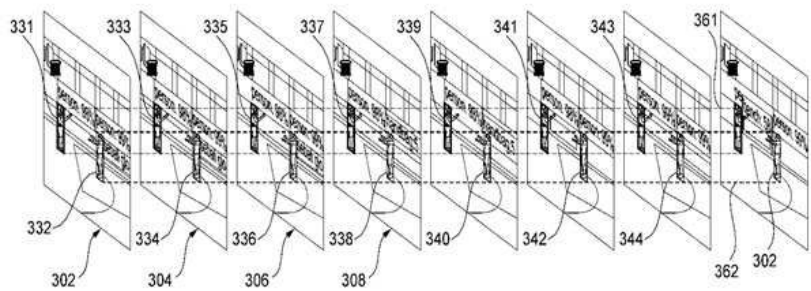
도면6



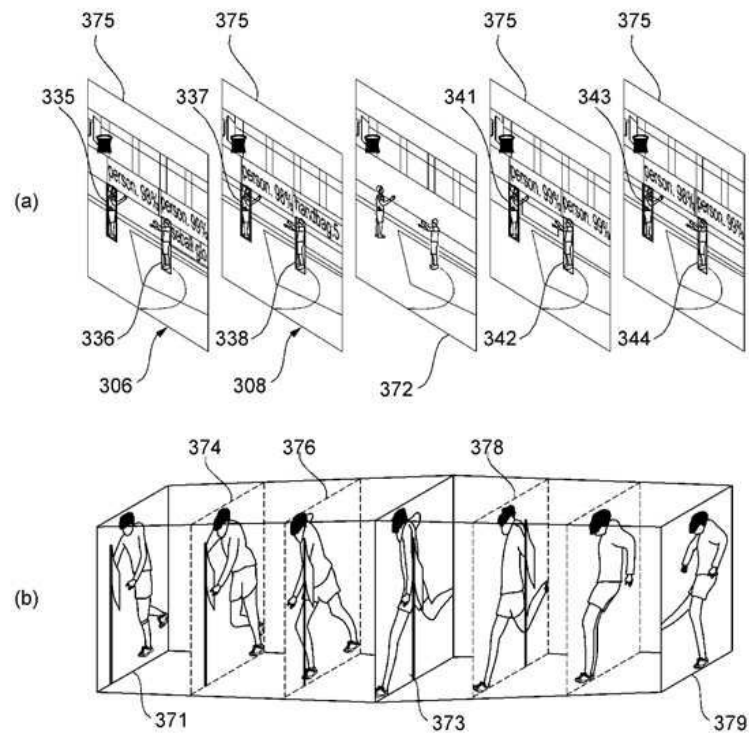
도면7



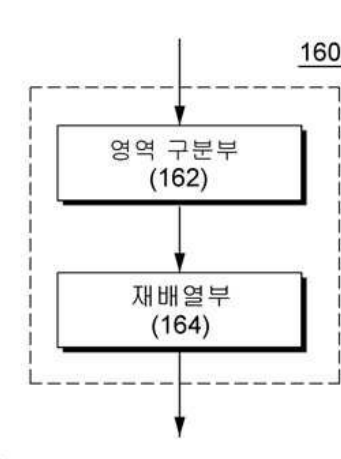
도면8



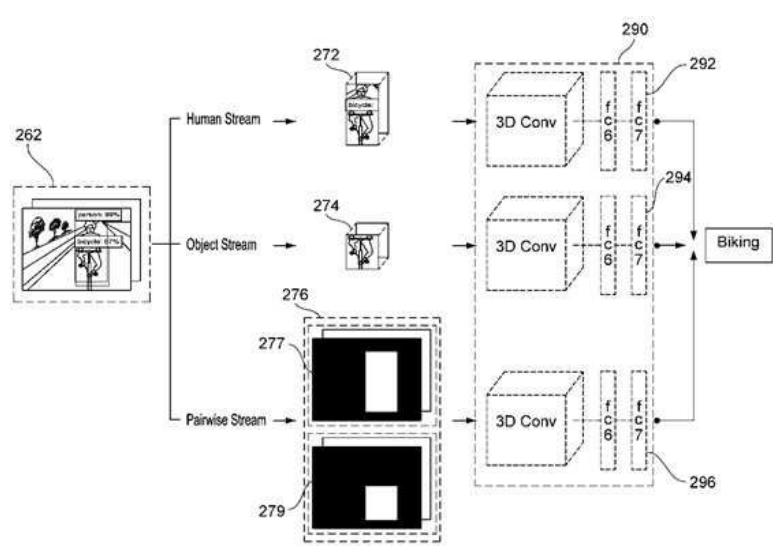
도면9



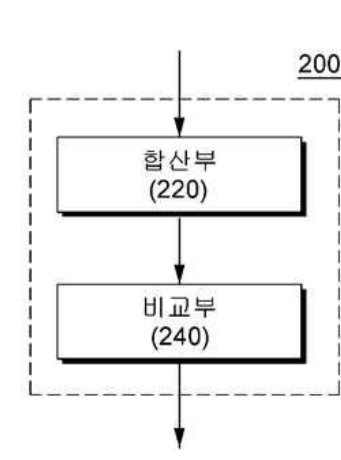
도면10



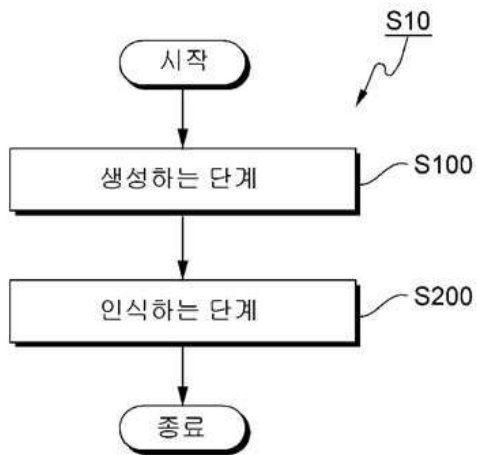
도면11



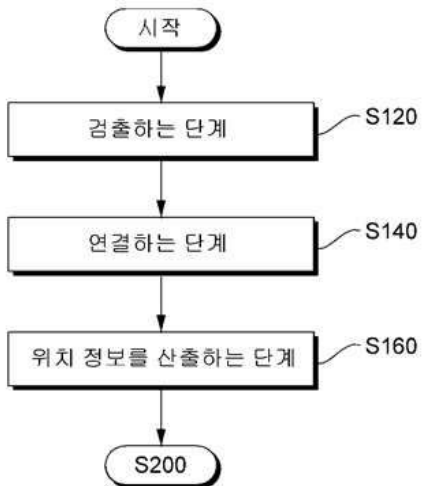
도면12



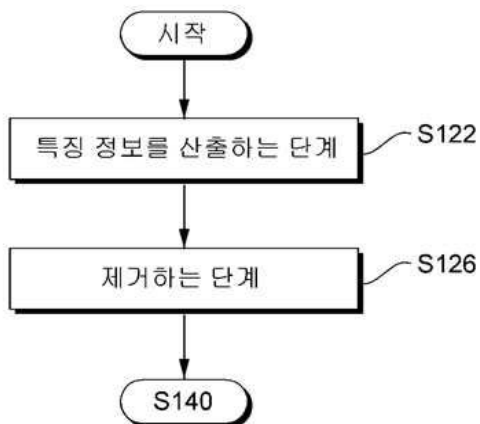
도면13



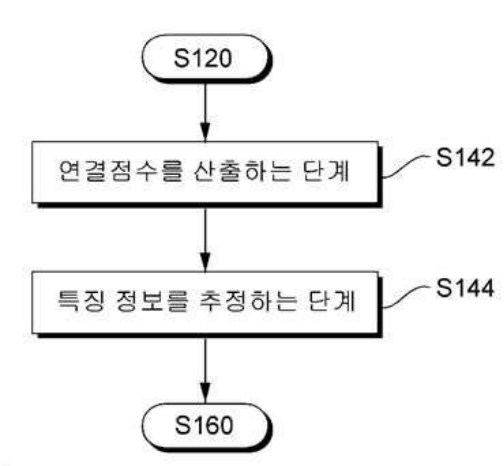
도면14



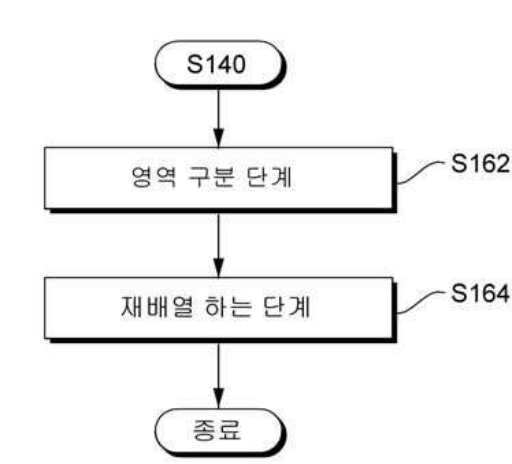
도면15



도면16



도면17



도면18

