



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2020년11월05일

(11) 등록번호 10-2174658

(24) 등록일자 2020년10월30일

(51) 국제특허분류(Int. Cl.)  
*G06K 9/00* (2006.01) *G06N 3/08* (2006.01)  
*G11B 27/36* (2006.01) *H04N 21/845* (2011.01)  
 (52) CPC특허분류  
*G06K 9/00711* (2013.01)  
*G06K 9/42* (2013.01)  
 (21) 출원번호 10-2019-0034832  
 (22) 출원일자 2019년03월27일  
 심사청구일자 2019년03월27일  
 (65) 공개번호 10-2020-0119391  
 (43) 공개일자 2020년10월20일  
 (56) 선행기술조사문헌  
 KR101731461 B1  
 KR1020150065370 A  
 KR1020170070298 A

(73) 특허권자  
 연세대학교 산학협력단  
 서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)  
 (72) 발명자  
 손광훈  
 서울특별시 서대문구 연세로 50, 제3공학관 C129호(신촌동, 연세대학교)  
 박정인  
 서울특별시 서대문구 연세로 50, 제3공학관 C129호(신촌동, 연세대학교)  
 (74) 대리인  
 민영준

전체 청구항 수 : 총 12 항

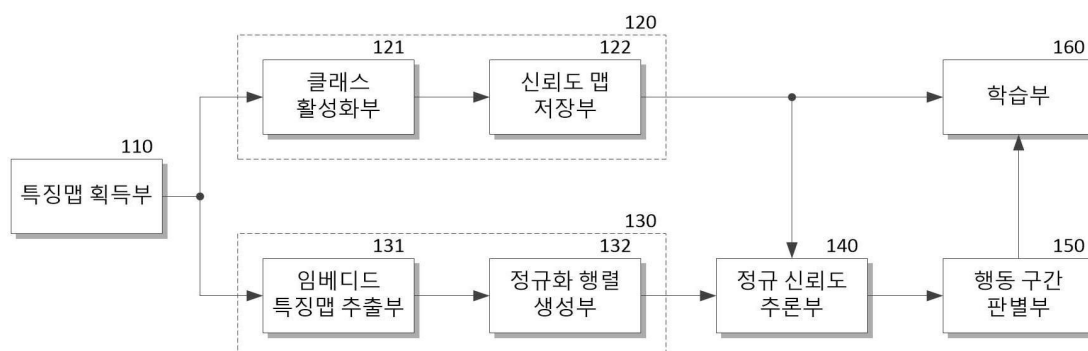
심사관 : 황승희

(54) 발명의 명칭 비디오 행동 인식 및 행동 구간 탐지 장치 및 방법

## (57) 요약

본 발명은 간단한 행동 레이블만이 주석된 학습용 비디오를 이용하여 학습되어 비디오에 대한 시간적 행동 로컬라이제이션을 수행할 수 있어, 학습용 비디오를 획득하기 위한 시간적 비용적 부담을 경감할 수 있으며, 비디오에서 세그먼트별 특징맵을 추출하여 세그먼트별 행동 신뢰도와 동일 행동에 대한 세그먼트 간 시멘틱 유사성을 분석하여 세그먼트별 행동 신뢰도에 세그먼트 간 시멘틱 유사성을 기반으로 가중치를 적용함으로써, 시간적 일관성을 갖고 행동의 정확한 시간적 위치를 인식할 수 있는 비디오 행동 인식 및 행동 구간 탐지 장치 및 방법을 제공할 수 있다.

## 대표도



(52) CPC특허분류

*G06N 3/08* (2013.01)

*G11B 27/36* (2013.01)

*H04N 21/8456* (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 2017M3C4A7069370

부처명 과학기술정보통신부

과제관리(전문)기관명

연구사업명 원천기술개발사업

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

기 여 율 1/1

과제수행기관명 연세대학교 산학협력단

연구기간 2018.06.01 ~ 2019.05.31

---

## 명세서

### 청구범위

#### 청구항 1

비디오의 다수 프레임을 기설정된 개수씩 프레임을 그룹화하여 다수의 세그먼트로 구분하고, 미리 학습된 제1 패턴 추정 방식에 따라 상기 다수의 세그먼트 각각의 특징을 추출하여 다수의 세그먼트 특징맵을 획득하는 특징맵 획득부;

약지도 학습 방식으로 제2 패턴 추정 방식이 미리 학습되어 상기 다수의 세그먼트 특징맵 각각이 기설정된 다수의 행동 클래스 각각에 대응하는 수준을 나타내는 클래스 신뢰도 점수를 추정하여 다수의 클래스 신뢰도 맵을 획득하는 클래스 신뢰도 맵 획득부;

미리 학습된 제3 패턴 추정 방식에 따라 상기 다수의 세그먼트 특징맵 각각에 대한 특징을 추출하여 다수의 임베디드 특징맵을 추출하고, 상기 다수의 임베디드 특징맵으로부터 그래프 라플라시안 정규화 기법에 따라 정규화 행렬을 생성하는 정규화부;

상기 정규화 행렬을 기반으로 상기 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수에 대한 정규 클래스 신뢰도 점수를 계산하여 다수의 정규 클래스 신뢰도 맵을 획득하는 정규 신뢰도 추론부; 및

상기 다수의 정규 클래스 신뢰도 맵의 정규 클래스 신뢰도 점수들로부터 기설정된 기준 신뢰도 점수 이상인 정규 클래스 신뢰도 점수에 대응하는 행동 및 시간 구간을 검출하는 비디오 행동 인식 및 행동 구간 탐지 장치.

#### 청구항 2

제1항에 있어서, 상기 정규화부는

상기 제3 패턴 추정 방식에 따라 상기 다수의 세그먼트 특징맵 각각에 대한 특징을 추출하여 다수의 임베디드 특징맵을 추출하는 임베디드 특징맵 추출부; 및

상기 임베디드 특징맵으로부터 다수의 세그먼트 사이의 시멘틱 유사성을 나타내는 유사도 행렬(affinity matrix)(A)과 대각 행렬인 정도 행렬(degree matrix)(D)을 기설정된 방식으로 획득하고, 상기 정도 행렬(D)과 상기 유사도 행렬(A)의 차인 그래프 라플라시안 행렬( $L = D - A$ )을 상기 정규화 행렬로 획득하는 정규화 행렬 생성부; 를 포함하는 비디오 행동 인식 및 행동 구간 탐지 장치.

#### 청구항 3

제2항에 있어서, 상기 정규화 행렬 생성부는

상기 유사도 행렬의 원소를 상기 다수의 임베디드 특징맵(e)에서 임베디드 특징맵( $e_i, e_j$ )들 사이의 에지 가중치( $w_{ij}$ )로서 수학식

$$w_{ij} = \exp(-\|e_i - e_j\|^2 / 2\epsilon^2)$$

(여기서  $\epsilon$ 는 특징 유사도 공간에서 임베디드 특징맵( $e_i, e_j$ )들 사이의 거리에 대한 감도를 조절하기 위해 기설정되는 상수)

에 따라 획득하고,

상기 정도 행렬의 i번째 대각 원소를 수학식

$$\sum_{j=1}^N w_{ij}$$

(여기서 N은 세그먼트의 개수이다)

에 따라 획득하는 비디오 행동 인식 및 행동 구간 탐지 장치.

#### 청구항 4

제3항에 있어서, 상기 정규 신뢰도 추론부는

상기 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수( $s^c$ )와 상기 정규화 행렬(L)로부터 수학적식

$$\mathbf{s}^{*c} = (\mathbf{I} + \mu \mathbf{L})^{-1} \mathbf{s}^c$$

(여기서 I는 항등 행렬이고,  $\mu$ 는 기지정된 하이퍼 파라미터이다.)

에 따라 정규 클래스 신뢰도 점수( $s^{*c}$ )를 계산하여, 상기 다수의 정규 클래스 신뢰도 맵을 획득하는 비디오 행동 인식 및 행동 구간 탐지 장치.

#### 청구항 5

삭제

#### 청구항 6

삭제

#### 청구항 7

제3항에 있어서, 상기 비디오 행동 인식 및 행동 구간 탐지 장치는

행동 레이블만이 주석된 학습용 비디오를 기반으로 상기 클래스 신뢰도 맵 획득부 및 상기 정규화부를 약지도 학습시키는 학습부; 를 더 포함하고,

상기 학습부는

상기 학습용 비디오에 응답하여, 상기 클래스 신뢰도 맵 획득부에서 획득되는 다수의 클래스 신뢰도 맵 각각의 클래스 신뢰도 점수의 평균을 기지정된 학습 기준 신뢰도 점수와 비교하여 학습 오차를 획득하고, 획득된 학습 오차를 역전파하여 상기 클래스 신뢰도 맵 획득부를 약지도 학습시키고,

이후 약지도 학습된 상기 클래스 신뢰도 맵 획득부와 상기 정규화부의 총 손실( $L_{total}$ )을 수학적식

$$\mathcal{L}_{total} = \mathcal{L}_{aff} + \lambda \cdot \mathcal{L}_{cls}$$

(여기서  $L_{aff}$ 는 정규화부의 유사도 행렬(A)의 유사도 손실을 나타내고,  $L_{cls}$ 는 학습용 비디오에 주석된 행동 레이블과 다수의 정규 클래스 신뢰도 맵 사이의 클래스 손실을 나타내며,  $\lambda$ 는 기지정된 손실 가중치를 나타낸다.)

에 따라 획득하고 역전파하여 상기 클래스 신뢰도 맵 획득부와 상기 정규화부를 함께 학습시키는 비디오 행동 인식 및 행동 구간 탐지 장치.

#### 청구항 8

제7항에 있어서, 상기 학습부는

약지도 학습된 상기 클래스 신뢰도 맵 획득부에서 다시 학습용 비디오에 응답하여 상기 다수의 클래스 신뢰도 맵이 획득되면, 상기 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수를 기지정된 방식에 따라 세그먼트 단위로 선택하고, 선택된 다수의 클래스 신뢰도 점수 각각의 클래스 신뢰도 맵에 대응하는 임시 클래스 레이블을 할당하여 학습용 신뢰도 맵을 생성하며,

상기 학습용 신뢰도 맵의 다수의 클래스 신뢰도 점수 중 기지정된 상위 기준 신뢰도 점수( $\tau_h$ ) 이상 또는 하위

기준 신뢰도 점수( $\tau_i$ ) 이하인 클래스 신뢰도 점수들 사이의 유사도를 나타내는 의사 유사도 행렬( $A^*$ )을 수학식

$$A_{ij}^* = \begin{cases} 1 & \text{if } z_i = z_j \\ 0 & \text{otherwise} \end{cases}$$

에 따라 생성하고,

상기 정규화부에서 획득되는 유사도 행렬(A)과 의사 유사도 행렬( $A^*$ ) 사이의 유사도 손실( $L_{aff}$ )을 수학식

$$\mathcal{L}_{aff} = \frac{1}{N} \sum_{i=1}^N \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \|A_{ij} - A_{ij}^*\|^2$$

(여기서 N은 세그먼트의 개수이고,  $\mathcal{N}(i)$ 는 세그먼트( $v_i$ )에 대한 시멘틱 이웃(semantic neighborhood)을 의미한다.)

에 따라 획득하는 비디오 행동 인식 및 행동 구간 탐지 장치.

#### 청구항 9

제7항에 있어서, 상기 학습부는

상기 정규 신뢰도 추론부에서 획득되는 정규 클래스 신뢰도 점수( $s^c$ )를 합산하여 비디오 레벨 점수 벡터( $\hat{y}^c$ )를 획득하고, 획득된 비디오 레벨 점수 벡터( $\hat{y}^c$ )와 학습용 비디오에 주석된 행동 레이블( $y^c$ ) 사이의 표준 다중 레이블 교차 엔트로피 손실(standard multi-label cross-entropy loss)을 수학식

$$\mathcal{L}_{cls} = - \sum_{c=1}^K [y^c \log(\hat{y}^c) + (1 - y^c) \log(1 - \hat{y}^c)]$$

에 따라 계산하여 상기 클래스 손실( $L_{cls}$ )을 획득하는 비디오 행동 인식 및 행동 구간 탐지 장치.

#### 청구항 10

비디오의 다수 프레임을 기설정된 개수씩 프레임을 그룹화하여 다수의 세그먼트로 구분하고, 미리 학습된 제1 패턴 추정 방식에 따라 상기 다수의 세그먼트 각각의 특징을 추출하여 다수의 세그먼트 특징맵을 획득하는 단계;

약지도 학습 방식으로 제2 패턴 추정 방식이 미리 학습되어 상기 다수의 세그먼트 특징맵 각각이 기설정된 다수의 행동 클래스 각각에 대응하는 수준을 나타내는 클래스 신뢰도 점수를 추정하여 다수의 클래스 신뢰도 맵을 획득하는 단계;

미리 학습된 제3 패턴 추정 방식에 따라 상기 다수의 세그먼트 특징맵 각각에 대한 특징을 추출하여 다수의 임베디드 특징맵을 추출하고, 상기 다수의 임베디드 특징맵으로부터 그래프 라플라시안 정규화 기법에 따라 정규화 행렬을 생성하는 단계;

상기 정규화 행렬을 기반으로 상기 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수에 대한 정규 클래스 신뢰도 점수를 계산하여 다수의 정규 클래스 신뢰도 맵을 획득하는 단계; 및

상기 다수의 정규 클래스 신뢰도 맵의 정규 클래스 신뢰도 점수들로부터 기설정된 기준 신뢰도 점수 이상인 정규 클래스 신뢰도 점수에 대응하는 행동 및 시간 구간을 판별하는 단계를 포함하는 비디오 행동 인식 및 행동 구간 탐지 방법.

#### 청구항 11

제10항에 있어서, 상기 정규화 행렬을 생성하는 단계는

상기 제3 패턴 추정 방식에 따라 상기 다수의 세그먼트 특징맵 각각에 대한 특징을 추출하여 다수의 임베디드 특징맵을 추출하는 단계; 및

상기 임베디드 특징맵으로부터 다수의 세그먼트 사이의 시멘틱 유사성을 나타내는 유사도 행렬(affinity matrix)(A)과 대각 행렬인 정도 행렬(degree matrix)(D)을 기지정된 방식으로 획득하고, 상기 정도 행렬(D)과 상기 유사도 행렬(A)의 차인 그래프 라플라시안 행렬( $L = D - A$ )을 상기 정규화 행렬로 획득하는 단계; 를 포함하는 비디오 행동 인식 및 행동 구간 탐지 방법.

## 청구항 12

제11항에 있어서, 상기 정규화 행렬로 획득하는 단계는

상기 유사도 행렬의 원소를 상기 다수의 임베디드 특징맵( $e$ )에서 임베디드 특징맵( $e_i, e_j$ )들 사이의 에지 가중치( $w_{ij}$ )로서 수학식

$$w_{ij} = \exp(-\|e_i - e_j\|^2 / 2\epsilon^2)$$

(여기서  $\epsilon$ 는 특징 유사도 공간에서 임베디드 특징맵( $e_i, e_j$ )들 사이의 거리에 대한 감도를 조절하기 위해 기지정되는 상수)

에 따라 획득하는 단계; 및

상기 정도 행렬의  $i$ 번째 대각 원소를 수학식

$$\sum_{j=1}^N w_{ij}$$

(여기서  $N$ 은 세그먼트의 개수이다)

에 따라 획득하는 단계; 를 포함하는 비디오 행동 인식 및 행동 구간 탐지 방법.

## 청구항 13

제12항에 있어서, 상기 다수의 정규 클래스 신뢰도 맵을 획득하는 단계는

상기 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수( $s^c$ )와 상기 정규화 행렬( $L$ )로부터 수학식

$$s^{*c} = (I + \mu L)^{-1} s^c$$

(여기서  $I$ 는 항등 행렬이고,  $\mu$ 는 기지정된 하이퍼 파라미터이다.)

에 따라 정규 클래스 신뢰도 점수( $s^{*c}$ )를 계산하여, 상기 다수의 정규 클래스 신뢰도 맵을 획득하는 비디오 행동 인식 및 행동 구간 탐지 방법.

## 청구항 14

삭제

## 청구항 15

제10항에 있어서, 상기 비디오 행동 인식 및 행동 구간 탐지 방법은

행동 레이블만이 주석된 학습용 비디오에 응답하여, 상기 다수의 클래스 신뢰도 맵 각각의 클래스 신뢰도 점수의 평균을 기지정된 학습 기준 신뢰도 점수와의 비교하여 학습 오차를 획득하고, 획득된 학습 오차를 역전파하여 약지도 학습되는 단계; 및

상기 약지도 학습 된 이후, 다시 상기 학습용 비디오를 인가받아 총 손실( $L_{total}$ )을 수학식

$$\mathcal{L}_{total} = \mathcal{L}_{aff} + \lambda \cdot \mathcal{L}_{cls}$$

(여기서  $L_{aff}$  는 상기 유사도 행렬(A)의 유사도 손실을 나타내고,  $L_{cls}$  는 학습용 비디오에 주석된 행동 레이블과 다수의 정규 클래스 신뢰도 맵 사이의 클래스 손실을 나타내며,  $\lambda$  는 기지정된 손실 가중치를 나타낸다.)

에 따라 획득하고 역전파하여 학습되는 단계; 를 더 포함하는 비디오 행동 인식 및 행동 구간 탐지 방법.

## 발명의 설명

### 기술 분야

- [0001] 본 발명은 비디오 행동 인식 및 행동 구간 탐지 장치 및 방법에 관한 것으로, 특히 비디오에서 객체의 행동을 인식하고 행동의 구간을 추출할 수 있는 행동 인식 및 행동 구간 탐지 장치 및 방법에 관한 것이다.

### 배경 기술

- [0002] 비디오로부터 이벤트 탐지, 요약, 시각적 질의 응답을 포함한 다양한 작업을 수행하기 위해서는 비디오에 나타나는 각종 행동을 인식하고, 인식된 행동이 나타나는 시간 구간을 구분하는 시간적 행동 로컬라이제이션(Temporal action localization)이 필수적이다.
- [0003] 한편 최근에는 딥 러닝(Deep learning) 기법으로 학습된 인공 신경망(artificial neural network)을 이용하여 비디오에서 자동으로 시간적 행동 로컬라이제이션을 수행하기 위한 다양한 연구가 진행되었다. 딥 러닝 기법을 이용함에 의해 비디오에 대한 시간적 행동 로컬라이제이션 작업이 인공 신경망에 의해 자동으로 수행될 뿐만 아니라 성능이 크게 향상되었다.
- [0004] 기존의 딥러닝 기법에서 인공 신경망은 완전 지도(fully supervised) 학습 방식으로 학습되었으며, 학습 시에 학습용 비디오 내의 행동별 시간적 경계 각각에 대해 검증 자료 레이블(ground truth label)이 완전하게 주석(full annotation)될 것이 요구되었다.
- [0005] 그러나 비디오에서 각 행동 각각에 대한 시간적 경계를 수작업으로 주석 처리하는 것은 시간적으로나 비용적으로 매우 비효율적이다. 뿐만 아니라, 각 행동의 시간적 경계는 작업자에 따라 주관적으로 판단될 수 있어, 인공 신경망을 부정확하게 학습시킬 수 있다는 문제가 있다.

## 선행기술문헌

### 특허문헌

- [0006] (특허문헌 0001) 한국 등록 특허 제10-1936947호 (2019.01.03 등록)

## 발명의 내용

### 해결하려는 과제

- [0007] 본 발명의 목적은 획득이 용이한 간단한 행동 레이블만이 주석된 학습용 비디오를 이용하는 약지도 학습(weakly-supervised learning) 방식을 기반으로 학습시킬 수 있는 비디오 행동 인식 및 행동 구간 탐지 장치 및 방법을 제공하는데 있다.
- [0008] 본 발명의 다른 목적은 약지도 학습으로 학습되어 비디오에 대한 시간적 행동 로컬라이제이션을 수행할 수 있는 비디오 행동 인식 및 행동 구간 탐지 장치 및 방법을 제공하는데 있다.
- [0009] 본 발명의 또 다른 목적은 비디오로부터 각 세그먼트 별로 정확한 행동을 인식할 수 있는 비디오 행동 인식 및 행동 구간 탐지 장치 및 방법을 제공하는데 있다.

### 과제의 해결 수단

[0010] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치는 비디오의 다수 프레임을 기설정된 개수씩 프레임을 그룹화하여 다수의 세그먼트로 구분하고, 미리 학습된 패턴 추정 방식에 따라 상기 다수의 세그먼트 각각의 특징을 추출하여 다수의 세그먼트 특징맵을 획득하는 특징맵 획득부; 약지도 학습 방식으로 패턴 추정 방식이 미리 학습되어 상기 다수의 세그먼트 특징맵 각각이 기지정된 다수의 행동 클래스 각각에 대응하는 수준을 나타내는 클래스 신뢰도 점수를 추정하여 다수의 클래스 신뢰도 맵을 획득하는 클래스 신뢰도 맵 획득부; 미리 학습된 패턴 추정 방식에 따라 상기 다수의 세그먼트 특징맵 각각에 대한 특징을 추출하여 다수의 임베디드 특징맵을 추출하고, 상기 다수의 임베디드 특징맵으로부터 그래프 라플라시안 정규화 기법에 따라 정규화 행렬을 생성하는 정규화부; 상기 정규화 행렬을 기반으로 상기 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수에 대한 정규 클래스 신뢰도 점수를 계산하여 다수의 정규 클래스 신뢰도 맵을 획득하는 정규 신뢰도 추론부; 및 상기 다수의 정규 클래스 신뢰도 맵의 정규 클래스 신뢰도 점수들을 분석하여 세그먼트 단위의 행동과 행동별 시간 구간을 판별하는 행동 구간 판별부; 를 포함한다.

[0011] 상기 정규화부는 미리 학습된 패턴 추정 방식에 따라 상기 다수의 세그먼트 특징맵 각각에 대한 특징을 추출하여 다수의 임베디드 특징맵을 추출하는 임베디드 특징맵 추출부; 및 상기 임베디드 특징맵으로부터 다수의 세그먼트 사이의 시멘틱 유사성을 나타내는 유사도 행렬(affinity matrix)(A)과 대각 행렬인 정도 행렬(degree matrix)(D)을 기지정된 방식으로 획득하고, 상기 정도 행렬(D)과 상기 유사도 행렬(A)의 차인 그래프 라플라시안 행렬( $L = D - A$ )을 상기 정규화 행렬로 획득하는 정규화 행렬 생성부; 를 포함할 수 있다.

[0012] 상기 정규화 행렬 생성부는 상기 유사도 행렬의 원소를 상기 다수의 임베디드 특징맵( $e_i, e_j$ )들 사이의 에지 가중치( $w_{ij}$ )로서 수학식

$$w_{ij} = \exp(-\|e_i - e_j\|^2 / 2\epsilon^2)$$

[0014] (여기서  $\epsilon$ 는 특징 유사도 공간에서 임베디드 특징맵( $e_i, e_j$ )들 사이의 거리에 대한 감도를 조절하기 위해 기지정되는 상수)에 따라 획득하고, 상기 정도 행렬의  $i$ 번째 대각 원소를 수학식

$$\sum_{j=1}^N w_{ij}$$

[0016] (여기서  $N$ 은 세그먼트의 개수이다)에 따라 획득할 수 있다.

[0017] 상기 정규 신뢰도 추론부는 상기 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수( $s^c$ )와 상기 정규화 행렬( $L$ )로부터 수학식

$$s^{*c} = (I + \mu L)^{-1} s^c$$

[0019] (여기서  $I$ 는 항등 행렬이고,  $\mu$ 는 기지정된 하이퍼 파라미터이다.)에 따라 정규 클래스 신뢰도 점수( $s^{*c}$ )를 계산하여, 상기 다수의 정규 클래스 신뢰도 맵을 획득할 수 있다.

[0020] 상기 행동 구간 판별부는 상기 다수의 정규 클래스 신뢰도 맵 각각의 정규 클래스 신뢰도 점수의 시간 평균인 신뢰도 평균값을 획득하고, 상기 신뢰도 평균값이 기지정된 기준 평균값 이상인 신뢰도 평균값을 갖는 정규 클래스 신뢰도 맵을 선택하며, 선택된 정규 클래스 신뢰도 맵에서 기준 신뢰도 점수 이상인 정규 클래스 신뢰도 점수를 분석하여, 세그먼트 단위의 행동 및 시간 구간을 검출할 수 있다.

[0021] 상기 행동 구간 판별부는 상기 다수의 정규 클래스 신뢰도 맵의 다수의 정규 클래스 신뢰도 점수를 세그먼트 구간별로 분석하여, 기지정된 기준 신뢰도 점수 이상인 정규 클래스 신뢰도 점수에 대응하는 행동 및 시간 구간을 검출할 수 있다.

[0022] 상기 비디오 행동 인식 및 행동 구간 탐지 장치는

[0023] 행동 레이블만이 주석된 학습용 비디오를 기반으로 상기 클래스 신뢰도 맵 획득부 및 상기 정규화부를 약지도 학습시키는 학습부; 를 더 포함하고, 상기 학습부는 상기 학습용 비디오에 응답하여, 상기 클래스 신뢰도 맵 획득부에서 획득되는 다수의 클래스 신뢰도 맵 각각의 클래스 신뢰도 점수의 평균을 기지정된 학습 기준 신뢰도 점수와 비교하여 학습 오차를 획득하고, 획득된 학습 오차를 역전파하여 상기 클래스 신뢰도 맵 획득부를 약



지도 학습시키고, 이후 약지도 학습된 상기 클래스 신뢰도 맵 획득부와 상기 정규화부의 총 손실( $L_{total}$ )을 수학식

$$\mathcal{L}_{total} = \mathcal{L}_{aff} + \lambda \cdot \mathcal{L}_{cls}$$

(여기서  $L_{aff}$  는 정규화부의 유사도 행렬(A)의 유사도 손실을 나타내고,  $L_{cls}$  는 학습용 비디오에 주석된 행동 레이블과 다수의 정규 클래스 신뢰도 맵 사이의 클래스 손실을 나타내며,  $\lambda$ 는 기지정된 손실 가중치를 나타낸다.)에 따라 획득하고 역전파하여 상기 클래스 신뢰도 맵 획득부와 상기 정규화부를 함께 학습시킬 수 있다.

상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 방법은 비디오의 다수 프레임을 기설정된 개수씩 프레임을 그룹화하여 다수의 세그먼트로 구분하고, 미리 학습된 패턴 추정 방식에 따라 상기 다수의 세그먼트 각각의 특징을 추출하여 다수의 세그먼트 특징맵을 획득하는 단계;

약지도 학습 방식으로 패턴 추정 방식이 미리 학습되어 상기 다수의 세그먼트 특징맵 각각이 기지정된 다수의 행동 클래스 각각에 대응하는 수준을 나타내는 클래스 신뢰도 점수를 추정하여 다수의 클래스 신뢰도 맵을 획득하는 단계; 미리 학습된 패턴 추정 방식에 따라 상기 다수의 세그먼트 특징맵 각각에 대한 특징을 추출하여 다수의 임베디드 특징맵을 추출하고, 상기 다수의 임베디드 특징맵으로부터 그래프 라플라시안 정규화 기법에 따라 정규화 행렬을 생성하는 단계; 상기 정규화 행렬을 기반으로 상기 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수에 대한 정규 클래스 신뢰도 점수를 계산하여 다수의 정규 클래스 신뢰도 맵을 획득하는 단계; 및 상기 다수의 정규 클래스 신뢰도 맵의 정규 클래스 신뢰도 점수들을 분석하여 세그먼트 단위의 행동과 행동별 시간 구간을 판별하는 단계;를 포함한다.

### 발명의 효과

따라서, 본 발명의 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치 및 방법은 간단한 행동 레이블만이 주석된 학습용 비디오를 이용하여 학습되어 비디오에 대한 시간적 행동 로컬라이제이션을 수행할 수 있어, 학습용 비디오를 획득하기 위한 시간적 비용적 부담을 경감할 수 있다. 또한 비디오에서 세그먼트별 특징맵을 추출하여 세그먼트별 행동 신뢰도와 동일 행동에 대한 세그먼트 간 시멘틱 유사성을 분석하여 세그먼트별 행동 신뢰도에 세그먼트 간 시멘틱 유사성을 기반으로 가중치를 적용함으로써, 시간적 일관성을 갖고 행동의 정확한 시간적 위치를 인식할 수 있다.

### 도면의 간단한 설명

도 1은 본 발명의 일 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치의 개략적 구조를 나타낸다.

도 2는 도 1의 특징맵 획득부의 상세 구성을 나타낸다.

도 3은 클래스 신뢰도 맵 획득부에서 획득되는 클래스 신뢰도 맵의 구조를 나타낸다.

도 4는 본 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치가 비디오에 대한 시간적 행동 로컬라이제이션을 수행한 결과의 일예를 나타낸다.

도 5는 약지도 학습을 위한 학습용 비디오의 일예를 나타낸다.

도 6은 유사도 손실을 계산하기 위한 학습용 클래스 신뢰도 맵의 일예를 나타낸다.

도 7은 본 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치의 성능을 시뮬레이션한 결과를 나타낸다.

도 8은 본 발명의 일 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 방법을 나타낸다.

### 발명을 실시하기 위한 구체적인 내용

본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재를 나타낸다.

- [0032] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0033] 도 1은 본 발명의 일 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치의 개략적 구조를 나타내고, 도 2는 도 1의 특징맵 획득부의 상세 구성을 나타내며, 도 3은 클래스 신뢰도 맵 획득부에서 획득되는 클래스 신뢰도 맵의 구조를 나타낸다. 그리고 도 4는 본 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치가 비디오에 대한 시간적 행동 로컬라이제이션을 수행한 결과의 일예를 나타낸다.
- [0034] 도 1을 참조하면, 본 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치는 특징맵 획득부(110), 클래스 신뢰도 맵 획득부(120), 정규화부(130), 정규 신뢰도 추론부(140) 및 행동 구간 판별부(150)를 포함할 수 있다.
- [0035] 특징맵 획득부(110)는 비디오의 다수 프레임을 그룹화하여 다수의 세그먼트로 구분하고, 구분된 다수의 세그먼트 각각에 대해 미리 학습된 패턴 추정 방식에 따라 다수의 세그먼트별 특징을 추출하여 다수의 세그먼트 특징맵을 획득한다.
- [0036] 도 2를 참조하면, 특징맵 획득부(110)는 비디오 획득부(111), 세그먼트 구분부(112), 특징 추출부(113) 및 특징맵 버퍼부(114)를 포함할 수 있다.
- [0037] 비디오 획득부(111)는 시간적 행동 로컬라이제이션이 수행되어야 하는 다수 프레임의 비디오를 획득한다. 여기서 시간적 행동 로컬라이제이션은 비디오 내에 포함된 적어도 하나의 객체의 시간별 행동을 구분하는 것으로서, 객체의 행동을 인지하고, 인지된 행동의 수행되는 시간 구간을 판별하는 것을 의미한다.
- [0038] 세그먼트 구분부(112)는 비디오 획득부(111)에서 획득된 비디오를 인가받고, 비디오를 구성하는 다수의 프레임을 시간적 순서에 따라 기설정된 개수씩 프레임을 그룹화하여 다수(여기서는 일예로 N개)의 세그먼트( $v = \{v_1, v_2, \dots, v_N\}$ )로 구분한다. 이때 다수의 세그먼트 각각에 포함되는 프레임의 개수는 다양하게 설정될 수 있으나 여기서는 일예로 세그먼트 구분부(112)가 16개의 프레임을 하나의 세그먼트로 그룹화하는 것으로 가정한다.
- [0039] 특징 추출부(113)는 미리 학습된 패턴 추정 방식에 따라 학습된 인공 신경망으로 구현되어 세그먼트 구분부(112)에서 구분된 다수의 세그먼트( $v$ ) 각각의 특징을 추출하여 다수의 세그먼트 특징맵( $f = \{f_1, f_2, \dots, f_N\}$ )을 획득한다. 특징 추출부(113)는 일예로 컨볼루션 신경망(Convolutional Neural Networks)으로 구현될 수 있으며, 다수의 프레임이 포함된 세그먼트에 대해 용이하게 특징을 추출할 수 있도록 3차원 컨볼루션 신경망으로 구현될 수도 있다.
- [0040] 특징 추출부(113)에서 획득되는 다수의 세그먼트 특징맵( $f$ )은 인공 신경망의 구성에 따라 다양한 차원의 행렬 형태로 획득될 수 있으나, 일예로 1차원 행렬 형태로 획득할 수 있다.
- [0041] 특징맵 버퍼부(114)는 다수의 세그먼트 각각에 대해 특징 추출부(113)에서 획득된 다수의 세그먼트 특징맵( $f$ )을 저장하고, 저장된 다수의 세그먼트 특징맵( $f$ )을 클래스 활성화부(121) 및 정규화부(130)로 전달한다.
- [0042] 클래스 신뢰도 맵 획득부(120)는 기지정된 다수의 행동 클래스 각각에 대해 다수의 세그먼트 특징맵( $f$ ) 각각이 대응 수준을 추정하여 다수(여기서는 일예로 C개)의 클래스 신뢰도 맵( $c$ )을 획득하여 저장한다.
- [0043] 한편 클래스 신뢰도 맵 획득부(120)는 클래스 활성화부(121) 및 신뢰도 맵 저장부(122)를 포함한다.
- [0044] 클래스 활성화부(121)는 특징맵 획득부(110)로부터 다수의 세그먼트 특징맵( $f$ )을 인가받고, 다수의 세그먼트 특징맵( $f$ )으로부터 다수의 클래스 신뢰도 맵을 획득한다. 여기서 다수의 클래스 신뢰도 맵 각각은 비디오에 레이블되는 다수의 행동 클래스( $c = \{c_1, c_2, \dots, c_C\}$ )에 대응하며, 다수의 세그먼트 특징맵( $f$ )이 다수의 행동 클래스 각각에 대응하는 수준을 나타내는 클래스 신뢰도 점수( $s^c$ )가 포함된 맵을 의미한다. 여기서 다수의 클래스 신뢰도 맵 각각은 학습 시에 비디오에 레이블되는 행동 중 하나에 대응한다. 예를 들어, 학습 시에 비디오에 레이블되는 행동이 다이빙, 마라톤, 사이클, 축구의 4개인 경우, 4개의 클래스 신뢰도 맵 이 각각 다이빙, 마라톤, 사이클, 축구에 대응할 수 있다. 그리고 다수의 클래스 신뢰도 맵 각각에는 다수의 세그먼트 특징맵( $f$ ) 각각이 해당 행동에 대응할 확률을 나타내는 신뢰도 점수가 포함된다.
- [0045] 즉 클래스 활성화부(121)는 도 3에 도시된 바와 같이, N개의 세그먼트( $v_1, v_2, \dots, v_N$ ) 각각이 C개의 행동 클레

스 레이블( $c = \{c_1, c_2, \dots, c_c\}$ ) 각각에 대응하는 수준을 신뢰도 점수로서 획득하는 것으로 볼 수 있다.

- [0046] 여기서 클래스 활성화부(121)는 약지도 학습에 의해 패턴 추정 방식이 미리 학습된 인공 신경망으로 구현되어 세그먼트 특징맵(f)의 행동 클래스 각각에 대한 클래스 신뢰도 점수( $s^c$ )를 추출할 수 있다. 일례로 클래스 활성화부(121)는 다수의 완전 연결 레이어(Fully connected layer)와 적어도 하나의 활성화 함수 레이어(Activation function layer)를 포함하는 인공 신경망으로 구현될 수 있으며, 활성화 함수 레이어는 일례로 시그모이드(sigmoid) 함수로 구현될 수 있다. 다수의 완전 연결 레이어는 입력된 세그먼트 특징맵(f)의 패턴을 추정하여 C개의 행동 클래스(c) 각각에 대한 클래스 신뢰도 점수( $s^c$ )를 추출하고, 시그모이드 함수는 다수의 완전 연결 레이어에서 추출된 신뢰도가 기지정된 범위(예를 들면, 0 ~ 1) 이내의 값이 되도록 전환하여 출력한다.
- [0047] 클래스 활성화부(121)가 2개의 완전 연결 레이어(FC1, FC2)와 시그모이드(sigmoid) 함수를 포함하는 경우, 클래스 활성화부(121)는 수학적 식 1에 따라 행동클래스(c)에서 t번째 세그먼트( $v_t$ )의 행동 클래스 신뢰도 점수( $s_t^c$ )를 획득할 수 있다.

### 수학적 식 1

$$s_t^c = \sigma\left(\sum_k w^c(k) f_{fc1,t}(k)\right)$$

- [0048]
- [0049] 여기서  $w^c(k)$ 는 행동 클래스(c)에 대한 최종 완전 연결 레이어(제2 완전 연결 레이어(FC2))의 k번째 가중치(w)를 나타내고,  $\sigma(\circ)$ 는 활성화 함수 레이어의 시그모이드 함수를 나타낸다.
- [0050] 수학적 식 1에 따르면, 클래스 활성화부(121)의 제2 완전 연결 레이어(FC2)는 t번째 세그먼트 특징맵( $f_t$ )에 대한 제1 완전 연결 레이어(FC1)의 출력( $f_{fc1,t}$ )을 인가받아 C개의 행동 클래스(c) 각각에 대한 클래스 신뢰도를 추출한다. 그리고 활성화 함수 레이어의 시그모이드 함수는 제2 완전 연결 레이어(FC2)에서 추출된 행동 클래스 신뢰도 점수( $s_t^c$ )를 기지정된 범위 이내의 값으로 전환하여 출력한다.
- [0051] 한편, 정규화부(130)는 특징맵 획득부(110)로부터 다수의 세그먼트 특징맵(f)을 인가받아 다수의 세그먼트 특징맵(f) 각각에 대한 특징을 추출하고, 추출된 다수의 세그먼트 특징맵(f)의 특징인 임베디드 특징맵으로부터 다수의 세그먼트(v) 사이의 시멘틱 유사성(semantic affinity)을 나타내는 유사도 행렬을 생성한다.
- [0052] 일반적으로 비디오에서 객체의 행동은 시간적 연속성을 갖고 수행되므로, 연속되는 다수의 세그먼트(v)에서 객체가 동일한 행동을 수행할 가능성이 높다. 따라서 다수의 클래스 신뢰도 맵 각각에서 동일한 행동이 수행되고 있는 기간에 대응하는 세그먼트들에 대한 클래스 신뢰도 점수는 유사하게 나타나야 한다. 즉 특정 행동이 수행되는 시간 구간에서의 클래스 신뢰도 점수 서로 유사하게 나타나야 하는 반면, 다른 구간에서의 클래스 신뢰도 점수는 상이하게 나타나야 한다.
- [0053] 그러나 클래스 활성화부(121)에서 획득되는 클래스 신뢰도 맵은 단지 다수의 행동 클래스(c)에 대해 각 세그먼트가 대응하는 수준만을 판별한다. 즉 시간적 변화에 따른 행동의 연속성에 따른 클래스 신뢰도 점수의 유사도를 고려하지 않아, 비디오에 포함된 객체의 행동에 시간 구간을 식별하기에 최적화되어 있지 않다.
- [0054] 이에 정규화부(130)는 세그먼트 특징맵(f)들이 서로 유사하면, 클래스 신뢰도 맵의 클래스 신뢰도 또한 서로 유사하게 추출되어야 한다는 전제로부터 세그먼트 특징맵(f)의 특징을 추출하여 특징 유사도 공간(feature affinity space)에 투영하고, 특징 유사도 공간에 투영된 세그먼트 특징맵(f)의 특징 사이의 거리를 기반으로 세그먼트 특징맵(f) 사이의 유사도를 판별하여 유사도 행렬을 생성하고, 유사도 행렬을 기반으로 정규화 행렬을 생성한다.
- [0055] 정규화부(130)는 임베디드 특징맵 추출부(131) 및 정규화 행렬 생성부(132)를 포함할 수 있다.
- [0056] 임베디드 특징맵 추출부(131)는 특징맵 획득부(110)로부터 다수의 세그먼트 특징맵(f)을 인가받고, 인가된 다수의 세그먼트 특징맵(f)에 대해 미리 학습된 패턴 추출 방식에 따라 특징을 다시 추출하여 임베디드 특징맵(e)을 추출한다.

[0057] 여기서 임베디드 특징맵 추출부(131)는 미리 학습된 인공 신경망으로 구현될 수 있으며, 일례로 특징 추출부(113)와 유사하게 다수의 컨볼루션 레이어를 포함하는 컨볼루션 신경망으로 구현될 수 있다.

[0058] 임베디드 특징맵 추출부(131)는 다수의 세그먼트( $v$ ) 사이의 시맨틱 유사성을 고려하기 위해, 수학식 2에 따라 다수의 세그먼트 특징맵( $f$ )에서 특징을 추출하여 임베디드 특징맵( $e$ )을 획득한다.

### 수학식 2

[0059] 
$$\mathbf{e} = \mathcal{F}(\mathbf{f}; \mathbf{w})$$

[0060] 수학식 2에서  $w$ 는 입력되는 세그먼트 특징맵( $f$ )의 특징을 특징 유사도 공간에 투영하기 위한 매개변수로 인공 신경망( $F$ )의 가중치이다.

[0061] 수학식 2에 의해 다수의 세그먼트 특징맵( $f$ )이 특징 유사도 공간에 투영되어 임베디드 특징맵( $e$ )이 획득된다. 이때 임베디드 특징맵 추출부(131)는 세그먼트 특징맵( $f$ )과 동일 차원에서 원소의 개수가 줄어 특징을 추출할 수 있다. 일례로, 다수의 세그먼트 특징맵( $f$ ) 각각이 1024개의 원소를 갖는 1차원 행렬 형태로 획득된 경우, 다수의 임베디드 특징맵( $e$ ) 각각은 128개의 원소를 갖는 1차원 행렬 형태로 획득될 수 있다. 즉 1024개의 세부적인 특징을 갖는 다수의 세그먼트 특징맵( $f$ ) 각각을 128개의 개략적 특징을 갖는 다수의 임베디드 특징맵( $e$ )으로 획득될 수 있다.

[0062] 임베디드 특징맵 추출부(131)는 적어도 하나의 컨볼루션 레이어와 적어도 하나의 활성화 함수 레이어를 포함하는 컨볼루션 신경망으로 구현될 수 있으며, 여기서 활성화 함수 레이어는 일례로 ReLU 레이어로 구현될 수 있다.

[0063] 정규화 행렬 생성부(132)는 그래프 라플라시안 정규화 기법에 따라 다수의 임베디드 특징맵( $e$ )으로부터 정규화 행렬인 그래프 라플라시안 행렬( $L$ )을 획득한다.

[0064] 그래프 라플라시안 정규화 기법은 다수의 절절점(vertex)이 있는 적절한 인접 그래프(neighborhood graph)( $G$ )가 주어지면, 높은 친화성으로 연결된 최적의 클래스 신뢰도 점수는 인접 그래프( $G$ )에 대해 매끄럽게 변화하는 값을 갖는다는 가정에 따라 클래스 신뢰도 점수를 정규화하기 위한 그래프 라플라시안 행렬( $L$ )을 획득하는 기법이다.

[0065] 정규화 행렬 생성부(132)는 우선 임베디드 특징맵 추출부(131)에서 추출된 다수의 임베디드 특징맵( $e$ )으로부터 에지 가중치( $w$ )를 원소로 포함하는 유사도 행렬(affinity matrix)( $\mathbf{A} \in \mathbb{R}^{T \times T}$ )을 획득한다. 여기서 정규화 행렬 생성부(132)는 다수의 임베디드 특징맵( $e$ )에서 2개의 임베디드 특징맵( $e_i, e_j$ )들 사이의 에지 가중치( $w_{ij}$ )를 수학식 3에 따라 획득한다.

### 수학식 3

[0066] 
$$w_{ij} = \exp(-\|\mathbf{e}_i - \mathbf{e}_j\|^2 / 2\epsilon^2)$$

[0067] 여기서  $\epsilon$ 는 특징 유사도 공간에서 임베디드 특징맵( $e_i, e_j$ )들 사이의 거리에 대한 감도를 조절하기 위한 상수로 미리 지정된다.

[0068] 수학식 3에 따르면, 특징 유사도 공간에서 2개의 임베디드 특징맵( $e_i, e_j$ ) 사이의 거리가 가까우면, 즉 2개의 세그먼트 특징맵( $f_i, f_j$ )이 유사하면, 에지 가중치( $w_{ij}$ )가 증가하고, 2개의 임베디드 특징맵( $e_i, e_j$ ) 사이의 거리가 멀면, 즉 2개의 세그먼트 특징맵( $f_i, f_j$ )이 서로 상이하면, 에지 가중치( $w_{ij}$ )가 감소한다.

[0069] 또한 정규화 행렬 생성부(132)는 대각 행렬(diagonal matrix)로서  $i$ 번째 대각 원소(diagonal entry)가

$\sum_{j=1}^N w_{ij}$ 로 계산되는 정도 행렬(degree matrix)(D)을 획득한다. 그리고 정규화 행렬 생성부(132)는 정도 행렬(D)에서 유사도 행렬(A)을 차감하여, 정규화 행렬인 그래프 라플라시안 행렬( $L = D - A$ )을 획득한다.

[0070] 결과적으로 그래프 라플라시안 행렬(L)은 다수의 세그먼트 특징맵(f) 각각에 대해 다른 세그먼트 특징맵과의 유사도를 원소로 갖게 되며, 세그먼트 특징맵(f)이 N개인 경우, 그래프 라플라시안 행렬(L)은  $N \times N$  행렬로 획득된다.

[0071] 정규 신뢰도 추론부(140)는 클래스 신뢰도 맵 획득부(120)로부터 기지정된 다수의 행동 클래스 각각에 대응하여 다수의 세그먼트 특징맵(f)의 클래스 신뢰도 점수( $s^c$ )가 포함된 다수의 클래스 신뢰도 맵을 인가받고, 정규화부(130)로부터 다수의 세그먼트 특징맵(f) 사이의 유사도를 나타내는 정규화 행렬인 그래프 라플라시안 행렬(L)을 인가받는다.

[0072] 그리고 클래스 신뢰도 점수( $s^c$ )와 그래프 라플라시안 행렬(L)을 이용하여 정규 클래스 신뢰도 점수( $s^{*c}$ )를 획득한다.

[0073] 정규 신뢰도 추론부(140)는 클래스 신뢰도 점수( $s^c$ )로부터 그래프 라플라시안 행렬(L)에 대한 최적의 클래스 신뢰도 점수( $\{\hat{s}^{*c}\}_{c=1}^K \in \mathbb{R}^N$ )를 탐색해야 한다. 그리고 최적 클래스 신뢰도 점수( $\hat{s}^{*c}$ )를 획득하기 위해서는 MAP(a maximum a posteriori) 문제를 해결해야 한다. MAP 문제는 수학식 4와 같이 공식화 될 수 있다.

#### 수학식 4

$$\hat{s}^{*c} = \arg \min_{\hat{s}^c} (\|s^c - \hat{s}^c\|_2^2 + \mu \cdot \hat{s}^{cT} L \hat{s}^c)$$

[0074]

[0075] 여기서  $\|s^c - \hat{s}^c\|_2^2$ 는 클래스(c)에 대한 클래스 신뢰도 점수(벡터)( $s^c$ )와 리파인 클래스 신뢰도 점수(벡터)( $\hat{s}^c$ ) 사이의 차를 계산하는  $l_2$ -놈( $l_2$ -norm) 함수이고, 나머지는 하이퍼 파라미터( $\mu$ )를 갖는 그래프 라플라시안 정규 표현이다.

[0076] 다만 수학식 4로 공식화되는 MAP 문제에서 최적 클래스 신뢰도 점수( $\hat{s}^{*c}$ )를 획득하는 것은 용이하지 않으므로, 정규 신뢰도 추론부(133)는 수학식 4를 수학식 5와 같이 선형 방정식의 역 시스템 문제로 재구성하여, 정규 클래스 신뢰도 점수( $s^{*c}$ )를 획득한다.

#### 수학식 5

$$s^{*c} = (I + \mu L)^{-1} s^c$$

[0077]

[0078] (여기서 I는 항등 행렬이고,  $\mu$ 는 기지정된 하이퍼 파라미터이다.)

[0079] 수학식 5에서 그래프 라플라시안 행렬(L)이 클래스에 무관하기 때문에 정규 클래스 신뢰도 점수( $s^{*c}$ )는 다수의 세그먼트 특징맵(f) 사이의 시멘틱 유사도에 의해서 획득된다. 즉 정규 신뢰도 추론부(140)는 다수의 클래스 신뢰도 맵 각각의 클래스 신뢰도 점수( $s^c$ )를 그래프 라플라시안 행렬(L)에 따라 정규화함으로써, 유사성이 높은 세그먼트 특징맵(f)들의 클래스 신뢰도 점수( $s^c$ )가 서로 유사하게 되도록 조절하여 정규 클래스 신뢰도 점수( $s^{*c}$ )가 포함된 다수의 정규 클래스 신뢰도 맵을 출력한다.

[0080] 행동 구간 판별부(150)는 정규 신뢰도 추론부(140)에서 정규화된 다수의 정규 클래스 신뢰도 맵을 인가받고, 다수의 정규 클래스 신뢰도 맵의 정규 클래스 신뢰도 점수( $s^{*c}$ )들을 분석한다. 그리고 분석 결과로 비디오의 세그



먼트별 대표 행동 클래스를 선택하여, 각 세그먼트에서의 대표 행동을 판별할 수 있다. 즉 비디오의 특정 시간에서 객체가 수행하는 행동을 판별할 수 있다. 또한 비디오에서 특정 행동이 수행되는 시간 구간을 세그먼트 단위로 정확하게 검출할 수 있다.

[0081] 여기서 행동 구간 판별부(150)는 다수의 세그먼트 각각에서 정규 클래스 신뢰도 점수( $s_i^{*c}$ )가 가장 높은 하나의 대표 행동을 선택할 수도 있으나, 경우에 따라서는 기지정된 기준 신뢰도 점수 이상인 모든 정규 클래스 신뢰도 점수( $s_i^{*c}$ )에 대응하는 행동을 선택할 수도 있다.

[0082] 또한 행동 구간 판별부(150)는 다수의 정규 클래스 신뢰도 맵 각각의 정규 클래스 신뢰도 점수( $s^{*c}$ )에 대한 시간 평균 풀링(time average pooling)을 수행하여 신뢰도 평균값을 획득할 수 있다. 그리고 각각의 다수의 정규 클래스 신뢰도 맵의 신뢰도 평균값이 획득되면, 기지정된 기준 평균값 이상인 신뢰도 평균값을 갖는 정규 클래스 신뢰도 맵을 선택하고, 선택된 정규 클래스 신뢰도 맵에서 기준 신뢰도 점수 이상인 정규 클래스 신뢰도 점수( $s^{*c}$ )를 다시 분석하여 각 시간 구간별 행동을 검출할 수도 있다. 즉 비디오에서 전체적으로 해당 행동이 포함된 가능성을 먼저 분석하여 해당 행동이 포함된 것으로 판별되면, 세그먼트 단위의 시간 구간을 다시 검출할 수 있다.

[0083] 상기와 같이 신뢰도 평균값을 획득하여 정규 클래스 신뢰도 맵을 우선 선택한 후, 다시 선택된 정규 클래스 신뢰도 맵에서 각 시간 구간별 행동을 검출하는 경우, 비디오의 세그먼트에 의도하지 않은 행동이 검출되는 것을 방지할 수 있다.

[0084] 특히 행동 구간 판별부(150)는 도 4에 도시된 바와 같이, 비디오에서 선택된 적어도 하나의 행동 클래스가 나타나는 시간 구간을 함께 출력할 수 있다. 즉 비디오를 시간적 행동 로컬라이제이션하여 각 행동의 시작 시간( $s_1^c, s_2^c, s_3^c$ )과 종료 시간( $e_1^c, e_2^c, e_3^c$ )을 출력할 수 있다.

[0085] 한편, 비디오 행동 인식 및 행동 구간 탐지 장치는 학습부(160)를 더 포함할 수 있다. 학습부(160)는 인공 신경망이 포함되는 클래스 신뢰도 맵 획득부(120) 및 정규화부(130)를 학습시키기 위한 구성으로, 비디오 행동 인식 및 행동 구간 탐지 장치의 학습 과정에서 이용되며, 학습이 완료된 이후에는 제거 될 수 있다.

[0086] 본 실시예에서 특징맵 획득부(110) 또한 인공 신경망을 포함하여 구현되지만, 특징맵 획득부(110)가 세그먼트( $v$ )로부터 세그먼트 특징맵( $f$ )을 획득하기 위한 학습 기법은 공지된 기술이므로, 특징맵 획득부(110)는 기존의 패턴 추정 방식에 따라 미리 학습되어 다수의 세그먼트 특징맵( $f$ )을 획득할 수 있다. 따라서 여기서는 특징맵 획득부(110)의 학습 방식은 별도로 설명하지 않는다.

[0087] 학습부(160)는 단순한 행동 레이블만이 주석된 학습용 비디오를 기반으로 클래스 신뢰도 맵 획득부(120) 및 정규화부(130)를 약지도 학습(weakly-supervised learning) 방식으로 학습시킨다.

[0088] 도 5는 약지도 학습을 위한 학습용 비디오의 일예를 나타내고, 도 6은 유사도 손실을 계산하기 위한 학습용 클래스 신뢰도 맵의 일예를 나타낸다.

[0089] 본 실시예에서 단순 행동 레이블만이 주석된 학습용 비디오는 시간 구간이나 행동 영역에 대한 별도의 주석 없이 이 비디오 전체에 대해 행동 레이블만이 제공되는 비디오를 의미한다. 일예로, 본 실시예에서 학습용 비디오에는 도 5에 도시된 바와 같이 다이빙, 테니스 스윙 또는 사이클링, 워킹, 러닝, 골프 등과 같이 단순히 객체의 행동만이 주석으로 제공되며, 객체의 위치 영역이나, 객체가 행동을 수행하는 시간 구간 등에 대해서는 별도의 주석이 제공되지 않는다.

[0090] 이러한 비디오에 주석된 간단한 행동 레이블을 기반으로 학습을 수행하는 방식을 비디오의 객체 영역에 대한 주석과 각 행동에 대한 시간 구간의 주석이 모두 함께 제공되어 학습을 수행하는 완전 지도 학습 방식과 구분하여 약지도 학습이라 한다.

[0091] 약지도 학습의 경우, 완전 지도 학습과 달리 행동 각각에 대한 시간적 경계를 수작업으로 주석 처리할 필요가 없으므로, 대량의 학습용 비디오를 저비용으로 용이하게 제작할 수 있다는 장점이 있다. 또한 학습 시에 작업자의 주관에 따라 상이할 수 있는 행동의 시간적 경계를 제공하지 않고 인공 신경망이 스스로 학습할 수 있도록 함으로써, 비디오 행동 인식 및 행동 구간 탐지 장치가 부정확하게 학습되는 것을 방지할 수 있다.

[0092] 본 실시예에서 학습부(160)는 학습 시에 우선 클래스 신뢰도 맵 획득부(120)를 약지도 학습시킨다. 이때 학습

부(160)는 클래스 신뢰도 맵 획득부(120)를 우선 학습시키기 위해, 정규화부(130)를 비활성화 할 수 있다. 그리고 클래스 신뢰도 맵 획득부(120)의 학습 시에 비디오 행동 인식 및 행동 구간 탐지 장치에는 도 5에 도시된 바와 같이, 객체의 단일 행동이 포함되고 행동 레이블만이 주석된 학습용 비디오가 입력된다.

[0093] 학습부(160)는 학습용 비디오에 대해 클래스 신뢰도 맵 획득부(120)가 획득한 다수의 클래스 신뢰도 맵 각각의 클래스 신뢰도 점수( $s^c$ )를 인가받아 분석하여 학습 오차를 역전파함으로써, 클래스 신뢰도 맵 획득부(120)를 학습 시킬 수 있다. 일례로 학습부(160)는 다수의 클래스 신뢰도 맵 각각의 클래스 신뢰도 점수( $s^c$ )를 기지정된 학습 기준 신뢰도 점수와의 비교하여 학습 오차를 획득할 수 있다.

[0094] 도 5에 도시된 바와 같이, 학습용 비디오는 단일 행동이 포함된 비디오로 구성되므로, 클래스 신뢰도 맵 획득부(120)에서 비디오에 포함된 행동에 대응하는 클래스 신뢰도 맵의 클래스 신뢰도 점수( $s^c$ )는 기지정된 학습 기준 신뢰도 점수 이상으로 획득되어야 하는 반면, 나머지 클래스 신뢰도 맵의 클래스 신뢰도 점수( $s^c$ )는 학습 기준 신뢰도 점수 미만으로 획득되어야 한다.

[0095] 따라서 학습부(160)는 각 클래스 신뢰도 맵의 클래스 신뢰도 점수( $s^c$ )와 학습 기준 신뢰도 점수를 비교하여 용이하게 학습 오차를 용이하게 획득하고, 획득된 학습 오차를 역전파하여 클래스 신뢰도 맵 획득부(120)를 학습 시킬 수 있다.

[0096] 또한 학습부(160)는 각 클래스 신뢰도 맵의 클래스 신뢰도 점수( $s^c$ )에 대해 시간 평균 풀링을 수행하여, 평균 신뢰도 점수를 획득하고, 획득된 평균 신뢰도 점수가 기지정된 평균 학습 기준 신뢰도 점수와 비교하여 학습 오차를 획득할 수도 있다.

[0097] 학습부(160)는 클래스 신뢰도 맵 획득부(120)가 약지도 학습 방식으로 우선 학습되면, 우선 학습된 클래스 신뢰도 맵 획득부(120)에서 획득되는 다수의 클래스 신뢰도 맵을 기반으로 클래스 신뢰도 맵 획득부(120)와 정규화부(130)를 학습시킨다.

[0098] 클래스 신뢰도 맵 획득부(120)와 정규화부(130)를 모두 학습시키는 경우, 학습부(160)는 정규화부(130)를 활성화한다. 그리고 학습부(160)는 수학식 6의 손실 함수에 따라 총 손실( $L_{total}$ )을 계산하고, 계산된 총 손실( $L_{total}$ )을 역전파함으로써, 클래스 신뢰도 맵 획득부(120)와 정규화부(130)를 학습시킨다.

## 수학식 6

[0099] 
$$\mathcal{L}_{total} = \mathcal{L}_{aff} + \lambda \cdot \mathcal{L}_{cls}$$

[0100] 여기서  $L_{total}$  은 비디오 행동 인식 및 행동 구간 탐지 장치의 전체 손실을 나타내고,  $L_{aff}$  는 유사도 행렬(A)의 유사도 손실을 나타내며,  $L_{cls}$  는 학습용 비디오에 주석된 행동 레이블과 다수의 정규 클래스 신뢰도 맵 사이의 클래스 손실을 나타낸다. 그리고  $\lambda$ 는 기지정된 손실 가중치이다.

[0101] 수학식 6에서 유사도 손실( $L_{aff}$ )은 약지도 학습된 클래스 신뢰도 맵 획득부(120)에서 획득되는 다수의 클래스 신뢰도 맵을 기반으로 획득될 수 있다.

[0102] 학습부(160)는 유사도 손실( $L_{aff}$ )을 획득하기 위해, 클래스 신뢰도 맵 획득부(120)에서 획득된 다수의 클래스 신뢰도 맵을 인가받아 도 6과 같은 학습용 클래스 신뢰도 맵을 생성한다.

[0103] 학습부(160)는 다수의 클래스 신뢰도 맵에서 세그먼트 단위로 가장 큰 클래스 신뢰도 점수를 획득하고, 획득된 클래스 신뢰도 점수들을 조합하여 학습용 클래스 신뢰도 맵을 생성한다. 즉 다수의 클래스 신뢰도 맵에서 특정 시간의 세그먼트에 대한 클래스 신뢰도 점수들 중 가장 큰 값을 갖는 클래스 신뢰도 점수를 추출하여, 학습용 클래스 신뢰도 맵의 클래스 신뢰도 점수로 포함시킨다. 이때, 학습부(160)는 수학식 7에 따라 획득된 클래스 신뢰도 점수에 대응하는 클래스 신뢰도 맵의 레이블을 학습용 클래스 신뢰도 맵의 임시 클래스 레이블( $z_t$ )로 할당한다.

### 수학식 7

$$z_t = \arg \max_c s_t^c$$

[0104]

[0105] 수학식 7에 따르면, 학습부(160)는 다수의 클래스 신뢰도 맵에서 특정 시간의 세그먼트들 중 최대 클래스 신뢰도 점수를 갖는 세그먼트가 포함된 클래스 신뢰도 맵(c)의 레이블을 임시 클래스 레이블( $z_t$ )로 획득하여 할당한다.

[0106] 따라서 학습용 클래스 신뢰도 맵에서는 세그먼트 단위로 임시 클래스 레이블이 가변될 수 있다.

[0107] 그리고 획득된 학습용 클래스 신뢰도 맵에 대해 기지정된 방식으로 상위 기준 신뢰도 점수( $\tau_h$ )와 하위 기준 신뢰도 점수( $\tau_l$ )를 설정한다. 여기서 상위 기준 신뢰도 점수( $\tau_h$ )는 해당 세그먼트에 대응하는 행동이 포함되었는지 여부를 판별하는 기준 신뢰도 점수로 설정될 수 있으나, 기준 신뢰도 점수와 별도로 설정될 수도 있다. 하위 기준 신뢰도 점수( $\tau_l$ )는 해당 세그먼트에 대응하는 행동이 포함되지 않았음을 판별하는 기준이 될 수 있다.

[0108] 상위 및 하위 기준 신뢰도 점수( $\tau_h, \tau_l$ )가 설정되면, 학습부(160)는 학습용 클래스 신뢰도 맵의 다수의 세그먼트들 중 상위 기준 신뢰도 점수( $\tau_h$ ) 이상 또는 하위 기준 신뢰도 점수( $\tau_l$ ) 이하인 클래스 신뢰도 점수들 사이의 유사도를 나타내는 의사 유사도 행렬( $A^*$ )을 수학식 8에 따라 생성한다.

### 수학식 8

$$A_{ij}^* = \begin{cases} 1 & \text{if } z_i = z_j \\ 0 & \text{otherwise} \end{cases}$$

[0109]

[0110] 도 6을 참조하면, 학습부(160)는 학습용 클래스 신뢰도 맵에서 상위 기준 신뢰도 점수( $\tau_h$ ) 이상이거나, 하위 기준 신뢰도 점수( $\tau_l$ ) 이하인 클래스 신뢰도 점수들이 서로 동일한 임시 클래스 레이블( $z_t$ )을 가지면 1을 할당하고, 서로 다른 임시 클래스 레이블( $z_t$ )을 가지면 0을 할당하여, 의사 유사도 행렬( $A^*$ )을 생성한다. 이때, 학습부(160)는 상위 기준 신뢰도 점수( $\tau_h$ )와 하위 기준 신뢰도 점수( $\tau_l$ ) 사이의 클래스 신뢰도 점수들에 대한 유사도는 판별하지 않는다. 즉 유사도 판별 대상이 되는 2개의 클래스 신뢰도 점수 중 적어도 하나가 상위 기준 신뢰도 점수( $\tau_h$ )와 하위 기준 신뢰도 점수( $\tau_l$ ) 사이의 클래스 신뢰도 점수이면, 유사도를 판별하지 않는다.

[0111] 수학식 8에 따라 의사 유사도 행렬( $A^*$ )이 획득되면, 학습부(160)는 정규화부(130)에서 획득된 유사도 행렬(A)과 의사 유사도 행렬( $A^*$ )을 이용하여 수학식 9에 따라 유사도 손실( $L_{aff}$ )을 획득한다.

### 수학식 9

$$\mathcal{L}_{aff} = \frac{1}{N} \sum_{i=1}^N \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \|A_{ij} - A_{ij}^*\|^2$$

[0112]

[0113] 여기서  $\mathcal{N}(i)$ 는 세그먼트( $v_i$ )에 대한 시멘틱 이웃(semantic neighborhood)을 의미한다.



[0114] 한편 학습부(160)는 정규 신뢰도 추론부(140)에서 획득되는 정규 클래스 신뢰도 점수( $s^{*c}$ )를 합산하여 비디오 레벨 점수 벡터( $\hat{y}^c$ )를 획득하고, 획득된 비디오 레벨 점수 벡터( $\hat{y}^c$ )와 학습용 비디오에 주석되는 행동 레이블( $y^c$ ) 사이의 표준 다중 레이블 교차 엔트로피 손실(standard multi-label cross-entropy loss)을 수학적 식 10에 따라 계산하여 클래스 손실( $L_{cls}$ )을 획득할 수 있다.

### 수학적 식 10

$$L_{cls} = - \sum_{c=1}^K [y^c \log(\hat{y}^c) + (1 - y^c) \log(1 - \hat{y}^c)]$$

[0115]

[0116] 학습부(160)는 유사도 손실( $L_{aff}$ )과 클래스 손실( $L_{cls}$ )이 획득되면, 수학적 식 6에 따라 총 손실( $L_{total}$ )을 획득하고, 획득된 총 손실( $L_{total}$ )을 역전파함으로써, 클래스 신뢰도 맵 획득부(120)와 정규화부(130)를 학습시킨다.

[0117] 도 7은 본 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치의 성능을 시뮬레이션한 결과를 나타낸다.

[0118] 도 7에서 w/o GR, FE는 정규화부(130)를 비활성화한 하여 시간 행동 로컬라이제이션을 수행한 경우를 나타내고, w/o FE는 정규화부(130)에서 임베디드 특징맵 추출부(131)를 제외하여 임베디드 특징맵(e)을 추출하지 않고, 정규화를 수행하여 시간 행동 로컬라이제이션을 수행한 경우를 나타낸다. 그리고 score는 본 실시예에 따른 시간 행동 로컬라이제이션을 수행한 경우를 나타낸다.

[0119] 그리고 Detection은 본 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치가 행동을 검출한 시간 구간을 나타내며, Ground Truth는 비디오에 실제 포함된 행동들의 시간 구간을 나타낸다.

[0120] 도 7에 도시된 바와 같이, 본 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 장치는 임베디드 특징맵(e)을 추출하고 정규화함으로써, 상대적으로 정확하게 시간 행동 로컬라이제이션을 수행할 수 있음을 확인할 수 있다.

[0121] 도 8은 본 발명의 일 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 방법을 나타낸다.

[0122] 도 8을 참조하면, 본 실시예에 따른 비디오 행동 인식 및 행동 구간 탐지 방법은 우선 학습부(160)가 클래스 신뢰도 맵 획득부(120) 및 정규화부(130)를 학습시킨다.

[0123] 먼저 클래스 신뢰도 맵 획득부(120)가 약지도 학습된다(S11). 학습부(160)는 클래스 신뢰도 맵 획득부(120)를 약지도 학습시키기 위해, 정규화부(130)를 비활성화시킬 수 있다. 그리고 학습부(160)는 객체의 단일 행동이 포함되고 행동 레이블만이 주석된 학습용 비디오에 대해 클래스 신뢰도 맵 획득부(120)로부터 획득된 다수의 클래스 신뢰도 맵 각각의 클래스 신뢰도 점수( $s^c$ )를 기지정된 학습 기준 신뢰도 점수와 비교하여 학습 오차를 획득하고, 획득된 학습 오차를 클래스 신뢰도 맵 획득부(120)로 역전파하여 클래스 신뢰도 맵 획득부(120)를 약지도 학습시킬 수 있다.

[0124] 이때, 학습부(160)는 다수의 클래스 신뢰도 맵 각각의 클래스 신뢰도 점수( $s^c$ )의 시간 평균인 평균 신뢰도 점수를 획득하고, 획득된 평균 신뢰도 점수가 기지정된 평균 학습 기준 신뢰도 점수와 비교하여 학습 오차를 획득할 수도 있다.

[0125] 학습부(160)는 클래스 신뢰도 맵 획득부(120)가 약지도 학습되면, 클래스 신뢰도 맵 획득부(120)와 함께 정규화부(130)를 학습시킨다(S12).

[0126] 학습부(160)는 유사도 손실( $L_{aff}$ )과 클래스 손실( $L_{cls}$ )을 획득하고, 수학적 식 6에 따라 총 손실( $L_{total}$ )을 계산하여 역전파함으로써 클래스 신뢰도 맵 획득부(120)와 정규화부(130)를 함께 학습시킬 수 있다. 여기서 유사도 손실( $L_{aff}$ )은 정규화부(130)에 대한 손실이고, 클래스 손실( $L_{cls}$ )은 클래스 신뢰도 맵 획득부(120)와 정규화부(130)의 공동 손실이다.

[0127] 유사도 손실( $L_{aff}$ )을 획득하기 위해, 학습부(160)는 약지도 학습된 클래스 신뢰도 맵 획득부(120)에서 획득되는 다수의 클래스 신뢰도 맵에서 세그먼트 단위로 최대값을 갖는 클래스 신뢰도 점수를 선택하여 조합함으로써, 학

습용 클래스 신뢰도 맵을 생성하고, 학습용 클래스 신뢰도 맵의 각 클래스 신뢰도 점수에 대해 대응하는 임시 클래스 레이블을 수학적식 7과 같이 할당한다.

[0128] 학습용 클래스 신뢰도 맵에 대해 상위 기준 신뢰도 점수( $\tau_h$ )와 하위 기준 신뢰도 점수( $\tau_l$ )를 설정하여, 상위 기준 신뢰도 점수( $\tau_h$ ) 이상 또는 하위 기준 신뢰도 점수( $\tau_l$ ) 이하인 클래스 신뢰도 점수들 사이의 유사도를 나타내는 의사 유사도 행렬( $A^*$ )을 수학적식 8에 따라 생성한다.

[0129] 의사 유사도 행렬( $A^*$ )이 획득되면, 학습부(160)는 유사도 행렬(A)과 의사 유사도 행렬( $A^*$ )을 이용하여 수학적식 9에 따라 유사도 손실( $L_{aff}$ )을 획득한다.

[0130] 한편, 학습부(160)는 정규 신뢰도 추론부(140)에서 획득되는 정규 클래스 신뢰도 점수( $s^{*c}$ )를 합산하여 비디오 레벨 점수 벡터( $\hat{y}^c$ )를 획득하고, 획득된 비디오 레벨 점수 벡터( $\hat{y}^c$ )와 학습용 비디오에 주석되는 행동 레이블( $y^c$ ) 사이의 표준 다중 레이블 교차 엔트로피 손실을 수학적식 10에 따라 계산하여 클래스 손실( $L_{cls}$ )을 획득한다.

[0131] 학습부(160)는 유사도 손실( $L_{aff}$ )과 클래스 손실( $L_{cls}$ )을 획득하고, 수학적식 6에 따라 총 손실( $L_{total}$ )을 계산하여 역전파함으로써 클래스 신뢰도 맵 획득부(120)와 정규화부(130)를 함께 학습시킬 수 있다.

[0132] 유사도 손실( $L_{aff}$ )과 클래스 손실( $L_{cls}$ )이 획득되면, 학습부(160)는 총 손실( $L_{total}$ )을 계산하여 역전파한다.

[0133] 클래스 신뢰도 맵 획득부(120)와 정규화부(130)에 대한 학습이 되면, 비디오 행동 인식 및 행동 구간 탐지 장치는 비디오를 인가받아 시간적 행동 로컬라이제이션을 수행한다.

[0134] 시간적 행동 로컬라이제이션을 수행하는 경우, 비디오 행동 인식 및 행동 구간 탐지 장치의 특징맵 획득부(110)가 우선 인가된 비디오의 다수 프레임을 시간적 순서에 따라 기설정된 개수씩 프레임을 그룹화하여 다수의 세그먼트로 구분하고, 미리 학습된 패턴 추정 방식에 따라 구분된 다수의 세그먼트 각각의 특징을 추출하여 다수의 세그먼트 특징맵을 획득한다(S21).

[0135] 그리고 클래스 신뢰도 맵 획득부(120)는 미리 학습된 패턴 추정 방식에 따라 다수의 세그먼트 특징맵 각각이 기 지정된 다수의 행동 클래스 각각에 대응하는 수준을 나타내는 클래스 신뢰도 점수( $s^c$ )를 추정하여 다수의 클래스 신뢰도 맵을 획득한다(S22).

[0136] 한편, 정규화부(130)는 미리 학습된 패턴 추정 방식에 따라 다수의 세그먼트 특징맵 각각에 대한 특징을 추출하여 다수의 임베디드 특징맵(e)을 추출한다(S23). 그리고 추출된 다수의 임베디드 특징맵(e)으로부터 그래프 라플라시안 정규화 기법에 따라 그래프 라플라시안 행렬인 정규화 행렬(L)을 생성한다(S24).

[0137] 정규화부(130)는 세그먼트 특징맵 사이의 유사도를 나타내는 유사도 행렬(A)을 수학적식 3에 따라 획득하고, 대각 행렬인 정도 행렬(D)을 획득하여 정도 행렬(D)에서 유사도 행렬(A)을 차감함으로써, 정규화 행렬(L)을 생성할 수 있다.

[0138] 정규 신뢰도 추론부(140)는 정규화 행렬(L)을 기반으로 수학적식 5에 따라 다수의 클래스 신뢰도 맵의 클래스 신뢰도 점수( $s^c$ )에 대한 정규 클래스 신뢰도 점수( $s^{*c}$ )를 계산하여 다수의 정규 클래스 신뢰도 맵을 획득한다(S25).

[0139] 다수의 정규 클래스 신뢰도 맵이 획득되면, 행동 구간 판별부(150)는 다수의 정규 클래스 신뢰도 맵의 정규 클래스 신뢰도 점수( $s^{*c}$ )들을 분석하여 세그먼트 단위의 행동과 행동별 시간 구간을 판별한다. 즉 입력된 비디오에 대한 시간적 행동 로컬라이제이션을 수행한다.

[0140] 이때, 행동 구간 판별부(150)는 세그먼트 단위로 정규 클래스 신뢰도 점수( $s_i^{*c}$ )가 가장 높은 하나의 대표 행동을 선택하거나, 기지정된 기준 신뢰도 점수 이상인 다수의 행동을 선택할 수 있다.

[0141] 또한 다수의 정규 클래스 신뢰도 맵의 신뢰도 평균값을 획득하고, 기지정된 기준 평균값 이상인 신뢰도 평균값을 갖는 정규 클래스 신뢰도 맵을 선택하여 선택된 정규 클래스 신뢰도 맵에서 기준 신뢰도 점수 이상인 정규

클래스 신뢰도 점수( $s^c$ )를 다시 분석하여 각 시간 구간별 행동을 검출할 수도 있다.

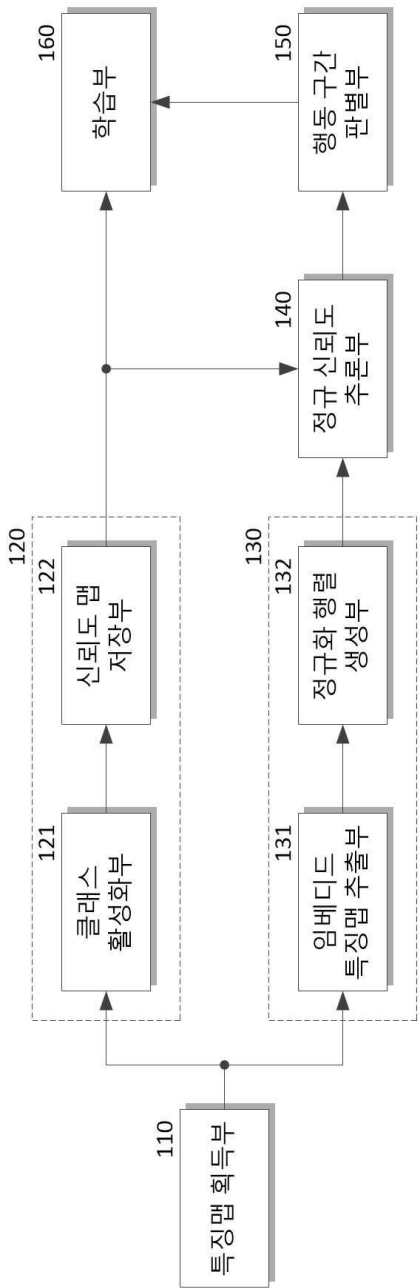
- [0142] 본 발명에 따른 방법은 컴퓨터에서 실행 시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.
- [0143] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.
- [0144] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

### 부호의 설명

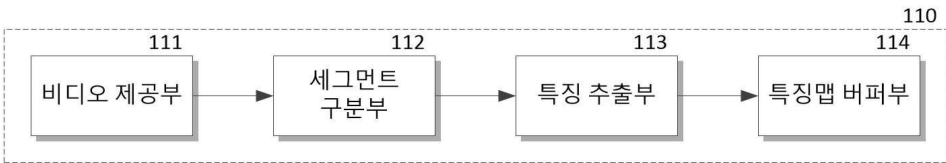
- |        |                 |                    |
|--------|-----------------|--------------------|
| [0145] | 110: 특징맵 획득부    | 120: 클래스 신뢰도 맵 획득부 |
|        | 121: 클래스 활성화부   | 122: 신뢰도 맵 저장부     |
|        | 130: 정규화부       | 131: 임베디드 특징맵 추출부  |
|        | 132: 정규화 행렬 생성부 | 140: 정규 신뢰도 추론부    |
|        | 150: 행동 구간 판별부  | 160: 학습부           |

도면

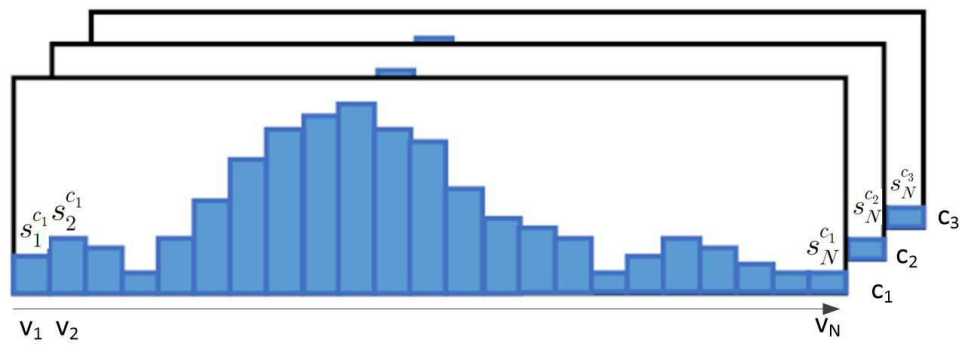
도면1



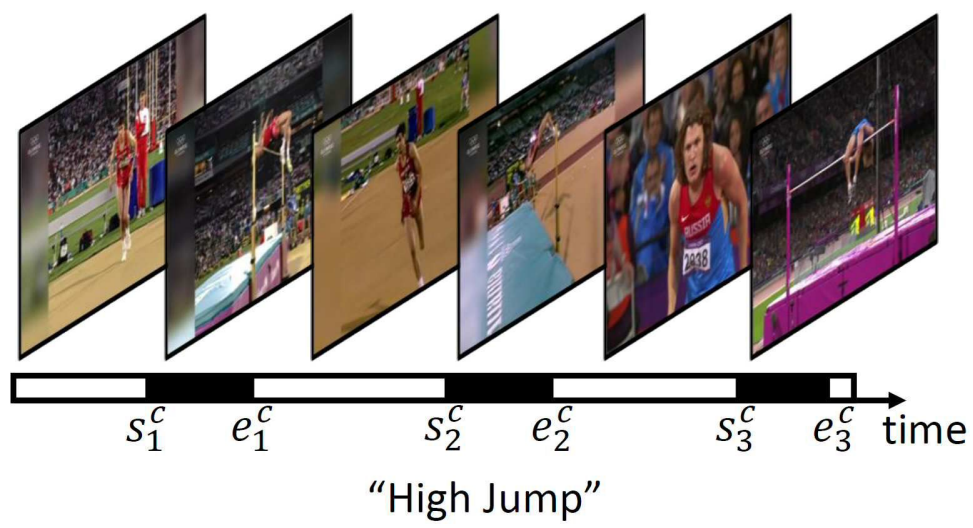
도면2



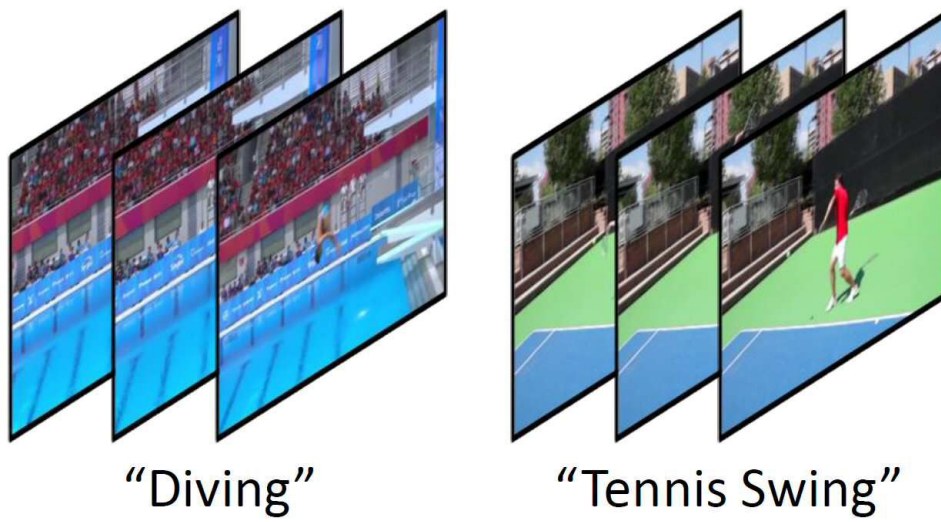
도면3



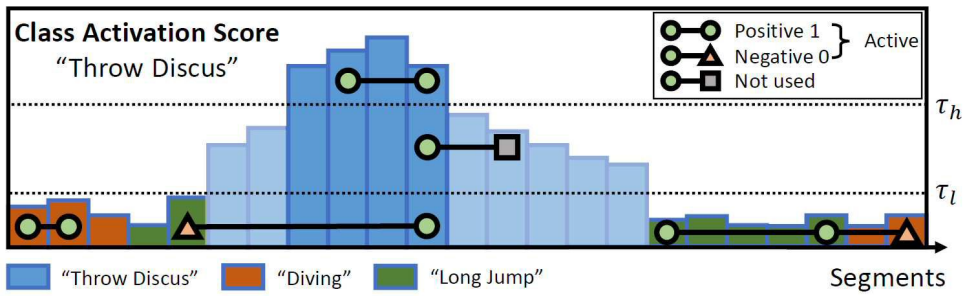
도면4



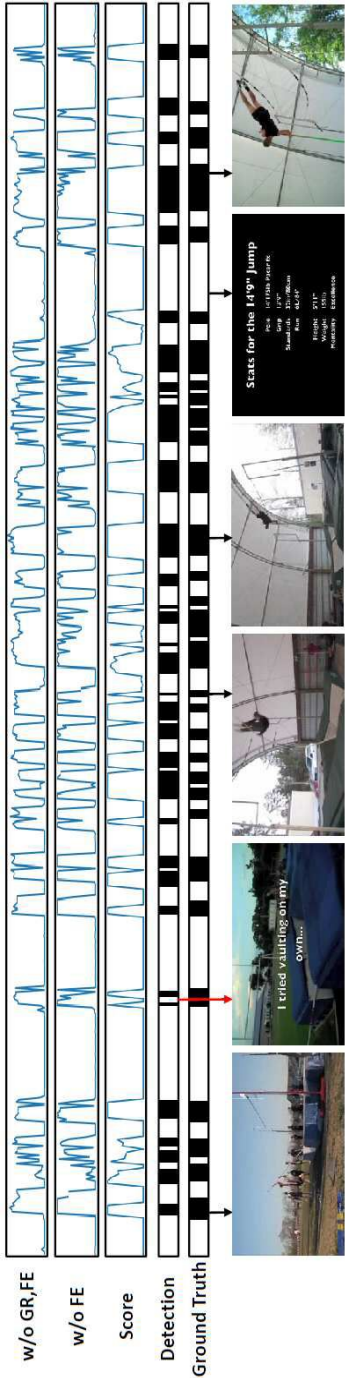
도면5



도면6



도면7



도면8

