



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년01월08일

(11) 등록번호 10-2201353

(24) 등록일자 2021년01월05일

(51) 국제특허분류(Int. Cl.)
H04N 5/14 (2006.01) G06N 20/00 (2019.01)
G06T 7/90 (2017.01)

(52) CPC특허분류
H04N 5/144 (2013.01)
G06N 20/00 (2019.01)

(21) 출원번호 10-2019-0151551

(22) 출원일자 2019년11월22일

심사청구일자 2019년11월22일

(56) 선행기술조사문헌

Haisheng Su ET AL: "Weakly Supervised Temporal Action Detection with Shot-Based Temporal Pooling Network", ICONIP 2018_Neural Information Processing, pp 426-436, 17 November 2018(2018.11.17.) 1부.*

KR1020190127261 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, 연세대학교)

이필현

서울특별시 서대문구 연세로 50, 제4공학관 810호(신촌동, 연세대학교)

(74) 대리인

특허법인우인

전체 청구항 수 : 총 10 항

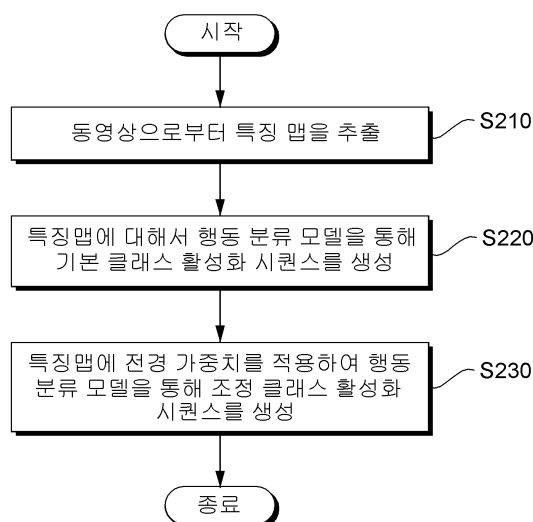
심사관 : 김광식

(54) 발명의 명칭 배경 프레임 억제를 통한 약한 지도 학습 기반의 행동 프레임 검출 방법 및 장치

(57) 요약

본 실시예들은 동영상으로부터 추출한 특징 맵에서 배경 프레임을 억제한 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하여, 약한 지도 학습 기반으로 행동 프레임을 정확하게 검출하는 장치 및 방법을 제공한다.

대표도 - 도2



(52) CPC특허분류
G06T 7/90 (2017.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	2019-11-1088
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	원천기술개발사업
연구과제명	(2세부)딥러닝 기반 의미론적 상황 이해 원천기술 연구 (2단계)(1/2)
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2019.06.01 ~ 2020.03.31

명세서

청구범위

청구항 1

컴퓨팅 디바이스에 의한 약한 지도 학습 기반의 행동 프레임 검출 방법에 있어서,

동영상으로부터 특징 맵을 추출하는 단계; 및

상기 특징 맵에 전경 가중치를 적용하여 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하는 단계를 포함하며,

상기 조정 클래스 활성화 시퀀스를 생성하는 단계는, 상기 특징 맵에 대해서 필터링 모델을 통해 상기 조정 클래스 활성화 시퀀스에서 배경 프레임이 활성화되지 않도록 상기 배경 프레임을 여과하여 상기 전경 가중치를 조정하는 것을 특징으로 하는 약한 지도 학습 기반의 행동 프레임 검출 방법.

청구항 2

제1항에 있어서,

상기 특징 맵을 추출하는 단계는,

상기 동영상을 복수의 색상 프레임으로 변환하고 상기 복수의 색상 프레임에서 색상 특징을 추출하고,

상기 동영상을 광류(Optical Flow) 프레임으로 변환하고 상기 광류 프레임에서 광류 특징을 추출하고,

상기 색상 특징과 상기 광류 특징을 결합하여 상기 특징 맵을 생성하는 것을 특징으로 하는 약한 지도 학습 기반의 행동 프레임 검출 방법.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 조정 클래스 활성화 시퀀스를 생성하는 단계는,

상기 조정 클래스 활성화 시퀀스에 대해서 조정 클래스 점수를 산출하여 행동 클래스에 대해 포지티브 학습과 배경 클래스에 대해 네거티브 학습을 수행하는 것을 특징으로 하는 약한 지도 학습 기반의 행동 프레임 검출 방법.

청구항 5

제1항에 있어서,

상기 조정 클래스 활성화 시퀀스를 생성하면서, 상기 특징 맵에 대해서 상기 행동 분류 모델을 통해 기본 클래스 활성화 시퀀스를 생성하는 단계를 포함하는 것을 특징으로 하는 약한 지도 학습 기반의 행동 프레임 검출 방법.

청구항 6

제5항에 있어서,

상기 기본 클래스 활성화 시퀀스를 생성하는 단계는,

상기 기본 클래스 활성화 시퀀스에 대해서 기본 클래스 점수를 산출하여 행동 클래스와 배경 클래스에 대해 포지티브 학습을 수행하는 것을 특징으로 하는 약한 지도 학습 기반의 행동 프레임 검출 방법.

청구항 7

제5항에 있어서,

상기 기본 클래스 활성화 시퀀스를 생성하는 단계와 상기 조정 클래스 활성화 시퀀스를 생성하는 단계는 상기 행동 분류 모델의 가중치를 공유하여 함께 학습하는 것을 특징으로 하는 약한 지도 학습 기반의 행동 프레임 검출 방법.

청구항 8

하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 행동 프레임 검출 장치에 있어서,

상기 프로세서는 동영상으로부터 특징 맵을 추출하고,

상기 프로세서는 상기 특징 맵에 전경 가중치를 적용하여 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하며,

상기 프로세서는, 상기 특징 맵에 대해서 필터링 모델을 통해 상기 조정 클래스 활성화 시퀀스에서 배경 프레임이 활성화되지 않도록 상기 배경 프레임을 여과하여 상기 전경 가중치를 조정하여, 상기 조정 클래스 활성화 시퀀스를 생성하는 것을 특징으로 하는 행동 프레임 검출 장치..

청구항 9

제8항에 있어서,

상기 프로세서는,

상기 동영상을 복수의 색상 프레임으로 변환하고 상기 복수의 색상 프레임에서 색상 특징을 추출하고,

상기 동영상을 광류(Optical Flow) 프레임으로 변환하고 상기 광류 프레임에서 광류 특징을 추출하고,

상기 색상 특징과 상기 광류 특징을 결합하여 상기 특징 맵을 생성하는 것을 특징으로 하는 행동 프레임 검출 장치.

청구항 10

삭제

청구항 11

제8항에 있어서,

상기 프로세서는,

상기 조정 클래스 활성화 시퀀스를 생성하면서, 상기 특징 맵에 대해서 상기 행동 분류 모델을 통해 기본 클래스 활성화 시퀀스를 생성하는 것을 특징으로 하는 행동 프레임 검출 장치.

청구항 12

제11항에 있어서,

상기 프로세서는,

상기 기본 클래스 활성화 시퀀스에 대해서 기본 클래스 점수를 산출하여 행동 클래스와 배경 클래스에 대해 포지티브 학습을 수행하고,

상기 조정 클래스 활성화 시퀀스에 대해서 조정 클래스 점수를 산출하여 상기 행동 클래스에 대해 포지티브 학습과 상기 배경 클래스에 대해 네거티브 학습을 수행하는 것을 특징으로 하는 행동 프레임 검출 장치.

발명의 설명

기술 분야

[0001] 본 발명이 속하는 기술 분야는 비정제 동영상 내에서 배경 프레임 억제를 통한 약한 지도 학습 기반의 행동 프레임 검출 방법 및 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] 유튜브(Youtube) 등의 동영상 공유 플랫폼을 통하여, 사용자들은 직접 만든 콘텐츠를 업로드하고 전 세계 사람들과 콘텐츠를 공유한다. 유튜브에 업로드된 20분 이상의 원본 스포츠 동영상의 개수는 약 11,000,000로 추산된다. 바쁜 현대인 입장에서 원본 스포츠 동영상에 대해 전체 경기 시간 동안 시청하는게 쉽지 않다. 축구 경기를 예로 들면, 골 장면과 같은 하이라이트만을 보고 싶은 시청자의 수요는 계속 증가하고 있다.

[0004] 실제 세계에 존재하는 동영상의 대부분은 우리가 관심 있는 행동 프레임뿐만 아니라, 행동이 존재하지 않는 배경 프레임을 포함하고 있다. 이러한 동영상을 비정제 동영상이라 한다.

[0005] 행동 프레임만을 검출하는 기술이 요구되며, 행동 프레임을 검출하여 정제된 동영상은 사용자뿐만 아니라 다른 딥 러닝 모델에서 학습 데이터로 사용하는데 매우 용이하다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국공개특허공보 제10-2015-0127684호 (2015.11.17)

(특허문헌 0002) 한국등록특허공보 제10-0785076호 (2017.12.05)

발명의 내용

해결하려는 과제

[0007] 본 발명의 실시예들은 동영상으로부터 추출한 특징 맵에서 배경 프레임을 억제한 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하여, 약한 지도 학습 기반으로 행동 프레임을 정확하게 검출하는데 주된 목적이 있다.

[0008] 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

[0009] 본 실시예의 일 측면에 의하면, 컴퓨팅 디바이스에 의한 약한 지도 학습 기반의 행동 프레임 검출 방법에 있어서, 동영상으로부터 특징 맵을 추출하는 단계, 및 상기 특징 맵에 전경 가중치를 적용하여 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하는 단계를 포함하는 약한 지도 학습 기반의 행동 프레임 검출 방법을 제공한다.

[0010] 상기 특징 맵을 추출하는 단계는, 상기 동영상을 복수의 색상 프레임으로 변환하고 상기 복수의 색상 프레임에서 색상 특징을 추출하고, 상기 동영상을 광류(Optical Flow) 프레임으로 변환하고 상기 광류 프레임에서 광류 특징을 추출하고, 상기 색상 특징과 상기 광류 특징을 결합하여 상기 특징 맵을 생성할 수 있다.

[0011] 상기 조정 클래스 활성화 시퀀스를 생성하는 단계는, 상기 특징 맵에 대해서 필터링 모델을 통해 상기 조정 클래스 활성화 시퀀스에서 배경 프레임이 활성화되지 않도록 상기 배경 프레임을 여과하여 상기 전경 가중치를 조정할 수 있다.

[0012] 상기 조정 클래스 활성화 시퀀스를 생성하는 단계는, 상기 조정 클래스 활성화 시퀀스에 대해서 조정 클래스 점수를 산출하여 행동 클래스에 대해 포지티브 학습과 배경 클래스에 대해 네거티브 학습을 수행할 수 있다.

[0013] 상기 조정 클래스 활성화 시퀀스를 생성하면서, 상기 특징 맵에 대해서 상기 행동 분류 모델을 통해 기본 클래스 활성화 시퀀스를 생성하는 단계를 포함할 수 있다.

[0014] 상기 기본 클래스 활성화 시퀀스를 생성하는 단계는, 상기 기본 클래스 활성화 시퀀스에 대해서 기본 클래스 점

수를 산출하여 행동 클래스와 배경 클래스에 대해 포지티브 학습을 수행할 수 있다.

- [0015] 상기 기본 클래스 활성화 시퀀스를 생성하는 단계와 상기 조정 클래스 활성화 시퀀스를 생성하는 단계는 상기 행동 분류 모델의 가중치를 공유하여 함께 학습할 수 있다.
- [0016] 본 실시예의 다른 측면에 의하면, 하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 행동 프레임 검출 장치에 있어서, 상기 프로세서는 동영상으로부터 특징 맵을 추출하고, 상기 프로세서는 상기 특징 맵에 전경 가중치를 적용하여 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하는 것을 특징으로 하는 행동 프레임 검출 장치를 제공할 수 있다.
- [0017] 상기 프로세서는, 상기 동영상을 복수의 색상 프레임으로 변환하고 상기 복수의 색상 프레임에서 색상 특징을 추출하고, 상기 동영상을 광류(Optical Flow) 프레임으로 변환하고 상기 광류 프레임에서 광류 특징을 추출하고, 상기 색상 특징과 상기 광류 특징을 결합하여 상기 특징 맵을 생성할 수 있다.
- [0018] 상기 프로세서는, 상기 특징 맵에 대해서 필터링 모델을 통해 상기 조정 클래스 활성화 시퀀스에서 배경 프레임이 활성화되지 않도록 상기 배경 프레임을 여과하여 상기 전경 가중치를 조정하여, 상기 조정 클래스 활성화 시퀀스를 생성할 수 있다.
- [0019] 상기 프로세서는, 상기 조정 클래스 활성화 시퀀스를 생성하면서, 상기 특징 맵에 대해서 상기 행동 분류 모델을 통해 기본 클래스 활성화 시퀀스를 생성할 수 있다.
- [0020] 상기 프로세서는, 상기 기본 클래스 활성화 시퀀스에 대해서 기본 클래스 점수를 산출하여 행동 클래스와 배경 클래스에 대해 포지티브 학습을 수행하고, 상기 조정 클래스 활성화 시퀀스에 대해서 조정 클래스 점수를 산출하여 상기 행동 클래스에 대해 포지티브 학습과 상기 배경 클래스에 대해 네거티브 학습을 수행할 수 있다.

발명의 효과

- [0021] 이상에서 설명한 바와 같이 본 발명의 실시예들에 의하면, 동영상으로부터 추출한 특징 맵에서 배경 프레임을 억제한 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하여, 약한 지도 학습 기반으로 행동 프레임을 정확하게 검출할 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 본 발명의 일 실시예에 따른 행동 프레임 검출 장치를 예시한 블록도이다.
- 도 2는 본 발명의 다른 실시예에 따른 행동 프레임 검출 방법을 예시한 흐름도이다.
- 도 3은 본 발명의 실시예들에 따른 행동 프레임 검출 장치의 행동 분류 모델을 예시한 도면이다.
- 도 4 내지 도 6은 본 발명의 실시예들에 따라 수행된 모의실험 결과를 도시한 것이다.

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] 지도 학습(Supervised Learning)은 정답이 주어지는 학습 전략으로, 어떤 입력(Input)에 대한 올바른 출력(Output)이 무엇인지 알 수 있다는 전제를 가진다. 지도 학습을 위해서는 데이터 세트에 대하여 학습할 때, 해당 데이터 세트를 구성하는 데이터 각각에 대한 정답이 필요하다.
- [0026] 약한 지도 학습(Weakly Supervised Learning) 환경에서는, 주어지는 정답에 대한 정보가 제한된다. 비지도 학습(Unsupervised Learning)과 같이 아무 정보도 주어지지 않는 것은 아니지만, 일부에 대한 정보만 제공받아 학습하고, 학습을 통해 제공받지 않은 정보를 예측해야 한다. 영상 처리 기계 학습을 위하여 영상 또는 동영상 등을 포함한 데이터 세트에 대하여 라벨링(Labeling)이 필요하다. 라벨링은 데이터 세트에 대한 정보를 제공한다.
- [0027] 프레임 단위의 정답을 갖는 강한 학습과 달리 동영상 단위의 정답을 갖는 약한 학습 환경에서는 정보가 제한되므로, 축구 동영상에서 키를 수행하거나 농구 동영상에서 슛을 수행하는 등의 행동 프레임을 학습하고 검출하는

것은 쉽지 않다.

- [0028] 본 발명은 동영상으로부터 추출한 특징 맵에서 배경 프레임을 억제한 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하여, 약한 지도 학습 기반으로 행동 프레임을 정확하게 검출한다.
- [0029] 도 1은 본 발명의 일 실시예에 따른 행동 프레임 검출 장치를 예시한 블록도이다.
- [0030] 행동 프레임 검출 장치(110)는 적어도 하나의 프로세서(120), 컴퓨터 판독 가능한 저장매체(130) 및 통신 버스(170)를 포함한다.
- [0031] 프로세서(120)는 행동 프레임 검출 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 행동 프레임 검출 장치(110)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0032] 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 판독 가능한 저장 매체(130)에 저장된 프로그램(140)은 프로세서(120)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독한 가능 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 행동 프레임 검출 장치(110)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.
- [0033] 통신 버스(170)는 프로세서(120), 컴퓨터 판독 가능한 저장 매체(140)를 포함하여 행동 프레임 검출 장치(110)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0034] 행동 프레임 검출 장치(110)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 장치는 입출력 인터페이스(150)를 통해 행동 프레임 검출 장치(110)의 다른 컴포넌트들에 연결될 수 있다.
- [0035] 본 실시예에 따른 행동 프레임 검출 장치는 동영상으로부터 추출한 특징 맵에서 배경 프레임을 억제한 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성하여, 약한 지도 학습 기반으로 행동 프레임을 정확하게 검출한다.
- [0036] 행동 프레임 검출 장치는 특징 맵에 대해서 필터링 모델을 통해 조정 클래스 활성화 시퀀스에서 배경 프레임이 활성화되지 않도록 배경 프레임을 여과하여 전경 가중치를 조정하여, 조정 클래스 활성화 시퀀스를 생성한다.
- [0037] 행동 프레임 검출 장치는 조정 클래스 활성화 시퀀스를 생성하면서, 특징 맵에 대해서 행동 분류 모델을 통해 기본 클래스 활성화 시퀀스를 생성한다.
- [0038] 행동 프레임 검출 장치는 기본 클래스 활성화 시퀀스에 대해서 기본 클래스 점수를 산출하여 행동 클래스와 배경 클래스에 대해 포지티브 학습을 수행하고, 조정 클래스 활성화 시퀀스에 대해서 조정 클래스 점수를 산출하여 행동 클래스에 대해 포지티브 학습과 배경 클래스에 대해 네거티브 학습을 수행한다.
- [0039] 도 2는 본 발명의 다른 실시예에 따른 행동 프레임 검출 방법을 예시한 흐름도이다. 행동 프레임 검출 방법은 행동 프레임 검출 장치 또는 컴퓨팅 디바이스 등에 의해 수행될 수 있다.
- [0040] 단계 S210에서 프로세서는 동영상으로부터 특징 맵을 추출한다.
- [0041] 특징 맵을 추출하는 단계(S210)는 동영상을 복수의 색상 프레임으로 변환하고 복수의 색상 프레임에서 색상 특징을 추출한다. 특징 맵을 추출하는 단계(S210)는 동영상을 광류(Optical Flow) 프레임으로 변환하고 광류 프레임에서 광류 특징을 추출한다. 특징 맵을 추출하는 단계(S210)는 색상 특징과 광류 특징을 결합하여 특징 맵을 생성한다.
- [0042] 단계 S220에서 프로세서는 특징 맵에 대해서 행동 분류 모델을 통해 기본 클래스 활성화 시퀀스를 생성한다.
- [0043] 기본 클래스 활성화 시퀀스를 생성하는 단계(S220)는, 기본 클래스 활성화 시퀀스에 대해서 기본 클래스 점수를 산출하여 행동 클래스와 배경 클래스에 대해 포지티브 학습을 수행한다.

- [0044] 단계 S230에서 프로세서는 특징 맵에 전경 가중치를 적용하여 행동 분류 모델을 통해 조정 클래스 활성화 시퀀스를 생성한다.
- [0045] 조정 클래스 활성화 시퀀스를 생성하는 단계(S230)는, 특징 맵에 대해서 필터링 모델을 통해 조정 클래스 활성화 시퀀스에서 배경 프레임이 활성화되지 않도록 배경 프레임을 여과하여 전경 가중치를 조정한다.
- [0046] 조정 클래스 활성화 시퀀스를 생성하는 단계(S230)는, 조정 클래스 활성화 시퀀스에 대해서 클래스 점수를 산출하여 행동 클래스에 대해 포지티브 학습과 배경 클래스에 대해 네거티브 학습을 수행한다.
- [0047] 기본 클래스 활성화 시퀀스를 생성하는 단계(S220)와 조정 클래스 활성화 시퀀스를 생성하는 단계(S230)는 행동 분류 모델의 가중치를 공유하여 함께 학습한다.
- [0048] 도 3은 본 발명의 실시예들에 따른 행동 프레임 검출 장치의 행동 분류 모델을 예시한 도면이다.
- [0049] 행동 프레임 검출 장치는 특징 추출 모델과 행동 분류 모델을 포함한다.
- [0050] 특징 추출 모델은 레이어가 연결된 네트워크이며 가중치 및 바이어스를 학습하는 모델이다. 특징 추출 모델은 CNN(Convolutional Neural Network) 등의 신경 네트워크로 구현될 수 있다.
- [0051] 특징 추출 모델은 입력 동영상으로부터 RGB 프레임과 광류(Optical Flow) 프레임을 추출한다. 추출된 프레임들을 16 프레임 단위의 세그먼트로 분할한 후, 각 세그먼트로부터 각 1024 차원의 RGB 특징 정보와 광류 특징 정보를 획득한다. RGB 특징 정보와 광류 특징 정보를 연결하여 2048 차원의 특징 맵을 생성한다.
- [0052] 행동 분류 모델은 데이터에 대한 클래스를 예측하고 해당하는 라벨을 부여한다. 행동 분류 모델은 신경 네트워크 등으로 구현된 다양한 분류 모델이 적용될 수 있다.
- [0053] 행동 프레임 검출 장치의 기본 처리 동작(Base Branch)을 설명하면, 특징 맵을 컨볼루션 네트워크에 입력하고 각 시간에서의 행동 및 배경 클래스 점수를 생성한다. 클래스 점수를 갖는 클래스 활성화 시퀀스(Class Activation Sequence)를 생성한다. 생성된 클래스 활성화 시퀀스는 행동 클래스와 배경 클래스로 분류하도록 학습된다. 행동 클래스와 배경 클래스의 개수는 C+1로 표현되며, C는 행동 클래스의 개수이다. 클래스 활성화 시퀀스가 행동 부분에 활성화되도록 유도한다.
- [0054] 행동 프레임 검출 장치는 세그먼트 레벨의 클래스 점수를 예측하기 위해서, 기본 클래스 활성화 시퀀스(A_n)를 생성한다.

수학식 1

$$A_n = f_{conv}(X_n; \phi)$$

ϕ 는 컨볼루션 레이어의 학습 파라미터이다.

수학식 2

$$a_{n;c} = \text{aggregate}(A_n) = \frac{1}{k} \max_{\substack{A \subset A_n[:,c] \\ |A|=k}} \sum_{\forall a \in A} a$$

동영상 단위의 점수를 예측하기 위해서, 세그먼트 레벨의 클래스 점수를 종합한다. 동영상 단위의 점수는 검증 자료(Ground Truth)와 비교된다. 종합 점수를 산출하는데 상위 K 개의 평균을 적용한다.

동영상 단위의 클래스 점수는 각 클래스의 포지티브 샘플의 확률(p_n)을 예측하는데 사용된다.

수학식 3

$$p_{n;c} = \text{softmax}(a_{n;c})$$

기본 처리 동작의 손실 함수는 수학식 4와 같이 표현된다.

수학식 4

$$\mathcal{L}_{base} = \frac{1}{N} \sum_{n=1}^N \sum_{c=1}^{C+1} -y_{n;c}^{base} \log(p_{n;c})$$

y_n 은 n 번째 동영상에서 동영상 레벨의 라벨이다.

행동 프레임 검출 장치의 억제 처리 동작(Suppression Branch)을 설명하면, 특징 맵을 필터링 필터링 모듈에 넣어서 전경 가중치(Foreground Weight)를 계산한 후, 전경 가중치를 특징 맵에 곱한다. 배경이 억제된 특징 맵을 획득할 수 있다. 기본 처리 동작과 마찬가지로 컨볼루션 네트워크에 입력하여 클래스 활성화 시퀀스를 생성하고 클래스를 분류한다. 억제 처리 동작은 배경 프레임을 억제하기 위한 목적으로 설계되었기 때문에 배경 클래스에 대해서 네거티브로 학습한다. 그 결과 필터링 모듈이 전경 가중치를 정확하게 계산할 수 있다. 배경 클래스에 대한 포지티브 학습은 배경 있음을 정답으로 학습하고, 배경 클래스에 대한 네거티브 학습은 배경 없음을 정답으로 학습한다.

전경 가중치를 곱한 특징 맵(X'_n)은 수학식 5와 같이 표현된다.

수학식 5

$$X'_n = X_n \otimes W_n$$

\otimes 는 시간 차원의 요소별 곱이다.

조정 클래스 활성화 시퀀스(A'_n)는 수학식 6과 같이 표현된다.

수학식 6

$$A'_n = f_{conv}(X'_n; \phi)$$

억제 처리 동작의 손실 함수는 수학식 7와 같이 표현된다.

수학식 7

$$\mathcal{L}_{supp} = \frac{1}{N} \sum_{n=1}^N \sum_{c=1}^{C+1} -y_{n;c}^{supp} \log(p'_{n;c})$$

[0072] 기본 처리 동작에서 동영상 레벨의 라벨은 $\mathbf{y}_n^{base} = [y_{n;1}, \dots, y_{n;C}, 1]^T \in \mathbb{R}^{C+1}$ 이다. 배경 클래스에 대한 라벨이 1로 설정된다.

[0073] 억제 처리 동작에서 동영상 레벨의 라벨은 $\mathbf{y}_n^{supp} = [y_{n;1}, \dots, y_{n;C}, 0]^T \in \mathbb{R}^{C+1}$ 이다. 배경 클래스에 대한 라벨이 0으로 설정된다.

수학식 8

[0074] $\mathcal{L}_{overall} = \alpha \mathcal{L}_{base} + \beta \mathcal{L}_{supp} + \gamma \mathcal{L}_{norm}$

[0075] α , β , γ 는 최적화 파라미터이고, 어텐션 가중치에 대한 L1 정규화 $\mathcal{L}_{norm} = \frac{1}{N} \sum_{n=1}^N |\mathcal{W}_n|$ 를 적용할 수 있다.

[0076] 행동 프레임 검출 장치는 실제 동영상을 테스트할 때, 억제 처리 동작에서 배경 프레임이 억제된 결과를 사용하여 행동 프레임을 검출한다.

[0077] 도 4 내지 도 6은 본 발명의 실시예들에 따라 수행된 모의실험 결과를 도시한 것이다. 도 4는 배구 동영상의 스파이크 장면이고, 도 5는 투포환 동영상이고, 도 6은 축구 동영상의 페널티 킥 장면이다. 도 4 내지 도 6을 참조하면, 본 실시예에 따른 배경 프레임이 억제된 행동 분류 모델이 GT(Ground Truth)에 대응하는 행동 프레임을 정확하게 예측할 수 있음을 파악할 수 있다.

[0078] 행동 프레임 검출 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.

[0079] 행동 프레임 검출 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 모듈 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.

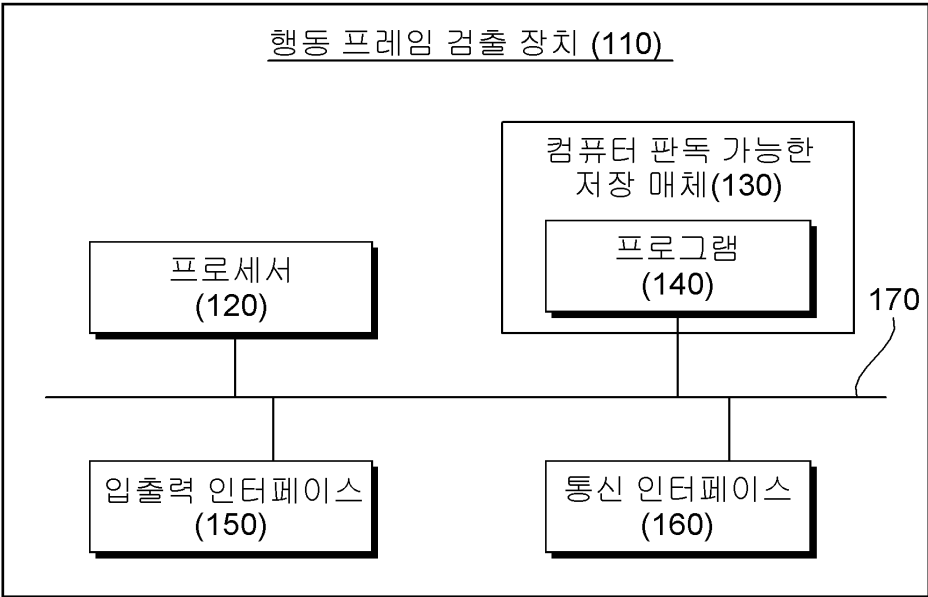
[0080] 도 2에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나 이는 예시적으로 설명한 것에 불과하고, 이 분야의 기술자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 2에 기재된 순서를 변경하여 실행하거나 또는 하나 이상의 과정을 병렬적으로 실행하거나 다른 과정을 추가하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이다.

[0081] 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.

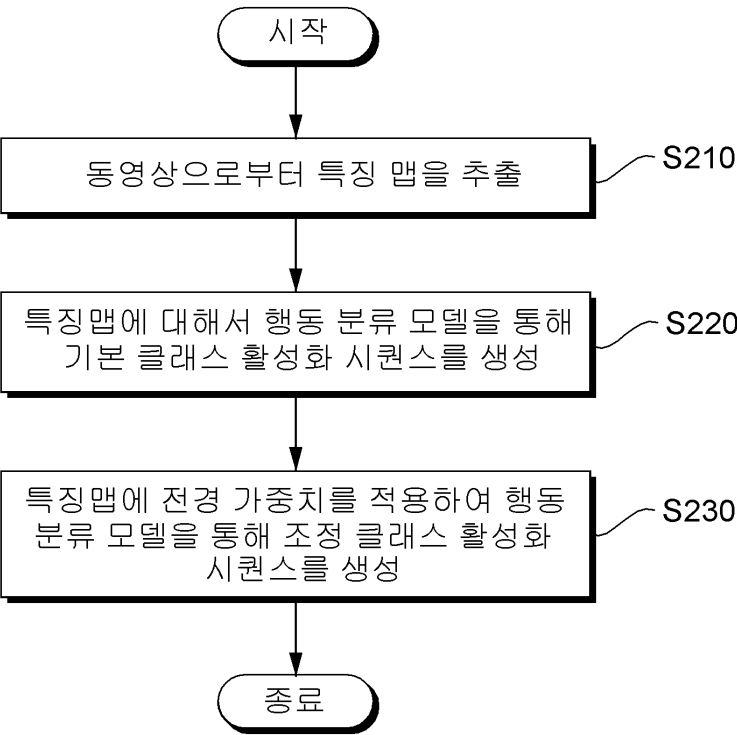
[0082] 본 실시예들은 본 실시예의 기술 사상을 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

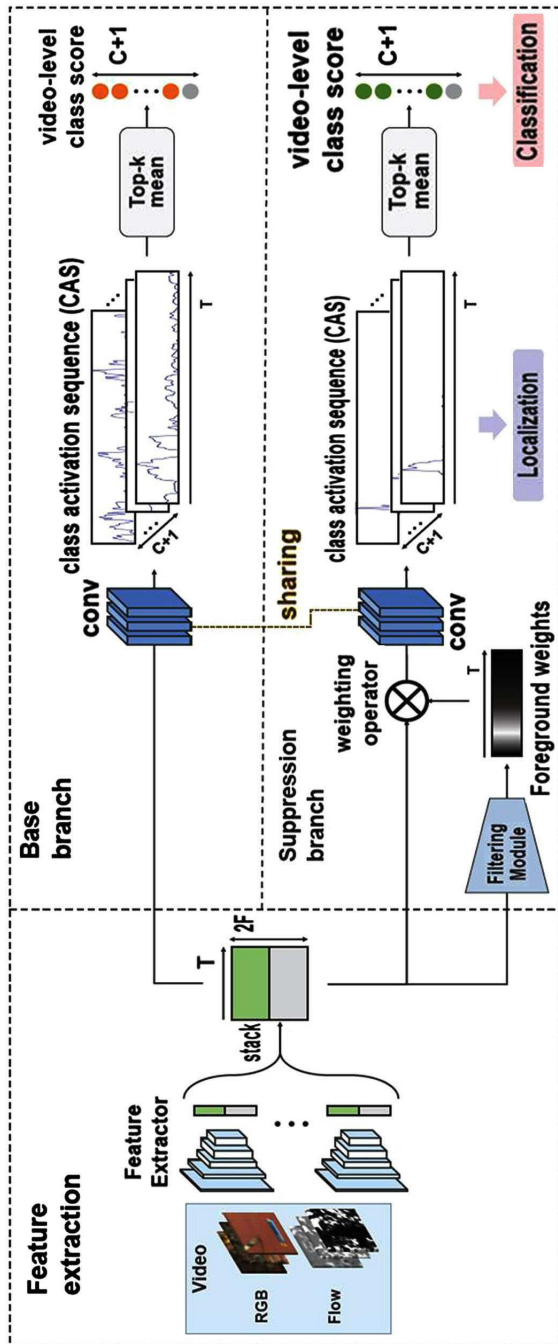
도면1



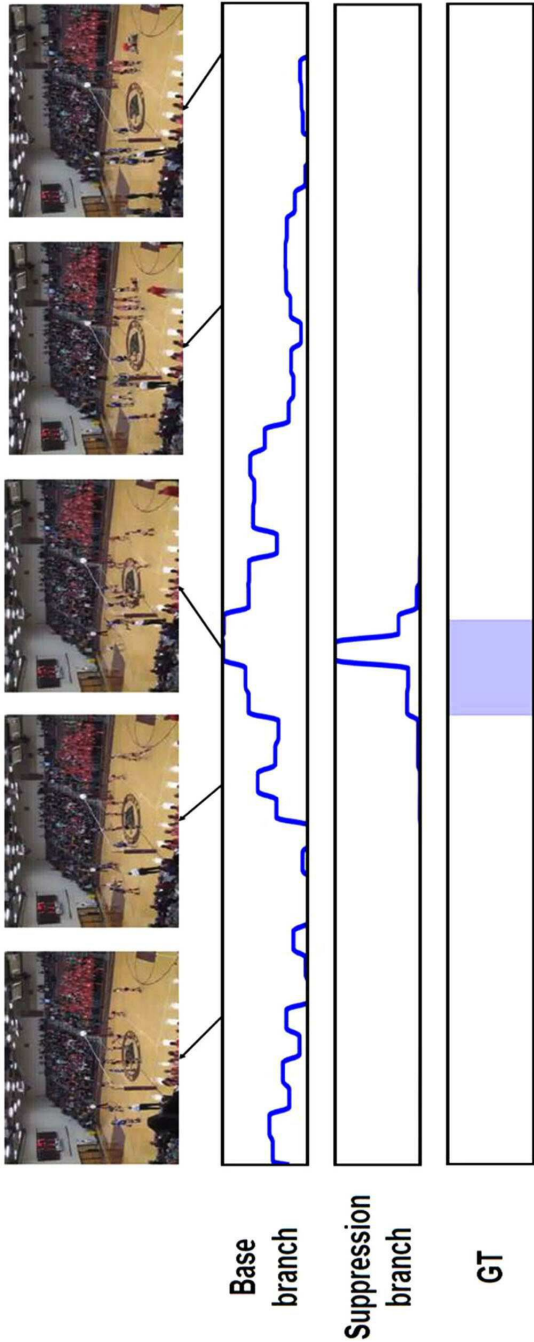
도면2



도면3



도면4



도면5



도면6

