



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년09월06일
(11) 등록번호 10-2574273
(24) 등록일자 2023년08월30일

(51) 국제특허분류(Int. Cl.)
G06V 40/20 (2022.01) G06N 3/04 (2023.01)
G06N 3/08 (2023.01) G06V 10/82 (2022.01)
G06V 20/40 (2022.01)
(52) CPC특허분류
G06V 40/20 (2022.01)
G06N 3/047 (2023.01)
(21) 출원번호 10-2022-0142983
(22) 출원일자 2022년10월31일
심사청구일자 2022년10월31일
(56) 선행기술조사문헌
KR1020210055532 A*
KR102395040 B1*
US20190080176 A1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
국방과학연구소
대전광역시 유성구 북유성대로488번길 160 (수남동)
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
김영휘
서울특별시 서대문구 연세로 50 연세대학교
남성현
서울특별시 서대문구 연세로 50 연세대학교
김선주
서울특별시 서대문구 연세로 50 연세대학교
(74) 대리인
제일특허법인(유)

전체 청구항 수 : 총 10 항

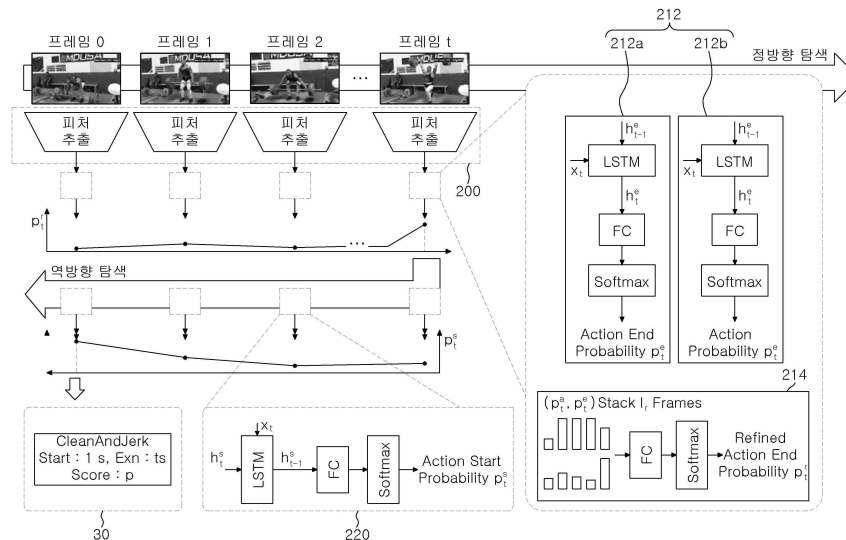
심사관 : 노용완

(54) 발명의 명칭 비디오 동작 검출 장치 및 방법

(57) 요약

본 발명은 비디오 동작 검출에 관한 것으로, 비디오 프레임을 스트리밍 방식으로 입력하는 프레임 입력부; 상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 저장부; 및 상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 액션 인스턴스 생성부;를 포함할 수 있다.

대표도



(52) CPC특허분류

G06N 3/049 (2023.01)

G06N 3/084 (2023.01)

G06V 10/82 (2022.01)

G06V 20/46 (2022.01)

명세서

청구범위

청구항 1

비디오 프레임을 스트리밍 방식으로 입력하는 프레임 입력부;

상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 저장부; 및

상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 액션 인스턴스 생성부;를 포함하되,

상기 저장부는,

상기 비디오 프레임을 인코딩하여 피처를 추출하는 피처 추출부;

상기 피처를 정방향(forward pass)으로 탐색하여 액션의 종료 지점을 검출하는 액션 종료 검출부; 및

상기 피처를 역방향(backward pass)으로 탐색하여 액션의 시작 지점을 검출하는 액션 시작 검출부;를 포함하고,

상기 액션 종료 검출부는,

상기 피처를 2종의 확률값으로 출력하는 멀티 헤드 검출부; 및

상기 2종의 확률값을 기초로, 최종 액션 종료 확률값을 출력하는 액션 종료 감지 개선부;를 포함하며,

상기 액션 시작 검출부는 상기 액션의 시작 지점이 존재할 액션 시작 확률값을 출력하고, 상기 액션 시작 확률값이 1회 상승한 후 하락하는 시점까지 상기 역방향으로 탐색하는 과정을 유지하며, 상기 하락하는 시점에서의 이전 프레임의 액션 시작 확률값을 최종 액션 시작 확률값으로 결정하고,

상기 액션 인스턴스 생성부는 상기 액션의 시작 지점과 상기 액션의 종료 지점을 기초로 상기 액션 인스턴스를 생성하며, 상기 액션 종료 확률값이 기 설정된 임계값을 초과하는 경우에 상기 액션 시작 검출부를 호출함으로써 상기 액션 종료 확률값에 종속되어 상기 액션 시작 검출부의 호출을 명령하고, 상기 액션 인스턴스는 상기 최종 액션 종료 확률값, 상기 최종 액션 시작 확률값, 액션의 클래스 및 클래스 확률값을 포함하는

비디오 동작 검출 장치.

청구항 2

삭제

청구항 3

삭제

청구항 4

제 1 항에 있어서,

상기 멀티 헤드 검출부는,

액션이 존재할 확률값을 출력하는 액션 감지 헤드; 및

액션의 종료 지점이 존재할 액션 종료 확률값을 출력하는 액션 종료 감지 헤드;를 포함하는

비디오 동작 검출 장치.

청구항 5

제 4 항에 있어서,
상기 액션 감지 헤드와 상기 액션 종료 감지 헤드 각각은,
LSTM(long short-term memory) 네트워크, FC(fully-connected) 레이어 및 소프트맥스(softmax) 레이어를 포함하는
비디오 동작 검출 장치.

청구항 6

제 1 항에 있어서,
상기 액션 종료 감지 개선부는,
FC 레이어 및 소프트맥스 레이어를 포함하는
비디오 동작 검출 장치.

청구항 7

제 1 항에 있어서,
상기 액션 시작 검출부는,
LSTM 네트워크, FC 레이어 및 소프트맥스 레이어를 포함하는
비디오 동작 검출 장치.

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

제 1 항에 있어서,
상기 피처 추출부는,
상기 피처를 시퀀스 형태로 보관하는
비디오 동작 검출 장치.

청구항 14

비디오 동작 검출 장치의 비디오 동작 검출 방법에 있어서,
비디오 프레임을 스트리밍 방식으로 입력하는 단계;
상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 단계; 및
상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계;를 포함하되,
상기 확률값을 출력하는 단계는,
상기 비디오 프레임을 인코딩하여 피처를 추출하는 단계;
상기 피처를 정방향으로 탐색하여 액션의 종료 지점을 검출하는 단계; 및
상기 피처를 역방향으로 탐색하여 액션의 시작 지점을 검출하는 단계;를 포함하고,
상기 액션의 종료 지점을 검출하는 단계는,
상기 피처를 2종의 확률값으로 출력하는 단계; 및
상기 2종의 확률값을 기초로, 최종 액션 종료 확률값을 출력하는 단계;를 포함하며,
상기 액션의 시작 지점을 검출하는 단계는 상기 액션의 시작 지점이 존재할 액션 시작 확률값이 1회 상승한 후 하락하는 시점까지 상기 역방향으로 탐색하는 단계를 포함하고,
상기 시작 지점을 검출하는 단계는 상기 하락하는 시점에서의 이전 프레임의 액션 시작 확률값을 최종 액션 시작 확률값으로 결정하는 단계를 포함하며,
상기 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계는, 상기 액션의 시작 지점과 상기 액션의 종료 지점을 기초로 상기 액션 인스턴스를 생성하는 단계와, 상기 액션 종료 확률값이 기 설정된 임계값을 초과하는 경우에 상기 액션의 시작 지점을 검출하는 단계를 포함함으로써, 상기 액션 종료 확률값에 종속되어 상기 액션 시작 지점을 검출하도록 호출되고,
상기 액션 인스턴스는 상기 최종 액션 종료 확률값, 상기 최종 액션 시작 확률값, 액션의 클래스 및 클래스 확률값을 포함하는
비디오 동작 검출 방법.

청구항 15

삭제

청구항 16

삭제

청구항 17

제 14 항에 있어서,
상기 2종의 확률값으로 출력하는 단계는,
액션이 존재할 확률값을 출력하는 단계; 및
액션의 종료 지점이 존재할 액션 종료 확률값을 출력하는 단계;를 포함하는
비디오 동작 검출 방법.

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

컴퓨터 프로그램을 저장하고 있는 컴퓨터 판독 가능 기록매체로서,

상기 컴퓨터 프로그램은,

비디오 동작 검출 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고,

상기 방법은,

비디오 프레임을 스트리밍 방식으로 입력하는 단계;

상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 단계; 및

상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계;를 포함하되,

상기 확률값을 출력하는 단계는,

상기 비디오 프레임을 인코딩하여 피처를 추출하는 단계;

상기 피처를 정방향으로 탐색하여 액션의 종료 지점을 검출하는 단계; 및

상기 피처를 역방향으로 탐색하여 액션의 시작 지점을 검출하는 단계;를 포함하고,

상기 액션의 종료 지점을 검출하는 단계는,

상기 피처를 2종의 확률값으로 출력하는 단계; 및

상기 2종의 확률값을 기초로, 최종 액션 종료 확률값을 출력하는 단계;를 포함하며,

상기 액션의 시작 지점을 검출하는 단계는 상기 액션의 시작 지점이 존재할 액션 시작 확률값이 1회 상승한 후 하락하는 시점까지 상기 역방향으로 탐색하는 단계를 포함하고,

상기 시작 지점을 검출하는 단계는 상기 하락하는 시점에서의 이전 프레임의 액션 시작 확률값을 최종 액션 시작 확률값으로 결정하는 단계를 포함하며,

상기 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계는, 상기 액션의 시작 지점과 상기 액션의 종료 지점을 기초로 상기 액션 인스턴스를 생성하는 단계와, 상기 액션 종료 확률값이 기 설정된 임계값을 초과하는 경우에 상기 액션의 시작 지점을 검출하는 단계를 포함함으로써, 상기 액션 종료 확률값에 종속되어 상기 액션의 시작 지점을 검출하도록 호출되고,

상기 액션 인스턴스는 상기 최종 액션 종료 확률값, 상기 최종 액션 시작 확률값, 액션의 클래스 및 클래스 확률값을 포함하는

컴퓨터 판독 가능한 기록매체.

청구항 22

컴퓨터 판독 가능 기록매체에 저장된 컴퓨터 프로그램으로서,

상기 컴퓨터 프로그램은,

비디오 동작 검출 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고,

상기 방법은,

비디오 프레임을 스트리밍 방식으로 입력하는 단계;

상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 단계; 및

상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계;를 포함하되,

상기 확률값을 출력하는 단계는,

상기 비디오 프레임을 인코딩하여 피처를 추출하는 단계;

상기 피처를 정방향으로 탐색하여 액션의 종료 지점을 검출하는 단계; 및

상기 피처를 역방향으로 탐색하여 액션의 시작 지점을 검출하는 단계;를 포함하고,

상기 액션의 종료 지점을 검출하는 단계는,

상기 피처를 2종의 확률값으로 출력하는 단계; 및

상기 2종의 확률값을 기초로, 최종 액션 종료 확률값을 출력하는 단계;를 포함하며,

상기 액션의 시작 지점을 검출하는 단계는 상기 액션의 시작 지점이 존재할 액션 시작 확률값이 1회 상승한 후 하락하는 시점까지 상기 역방향으로 탐색하는 단계를 포함하고,

상기 시작 지점을 검출하는 단계는 상기 하락하는 시점에서의 이전 프레임의 액션 시작 확률값을 최종 액션 시작 확률값으로 결정하는 단계를 포함하며,

상기 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계는, 상기 액션의 시작 지점과 상기 액션의 종료 지점을 기초로 상기 액션 인스턴스를 생성하는 단계와, 상기 액션 종료 확률값이 기 설정된 임계값을 초과하는 경우에 상기 액션의 시작 지점을 검출하는 단계를 포함함으로써, 상기 액션 종료 확률값에 종속되어 상기 액션의 시작 지점을 검출하도록 호출되고,

상기 액션 인스턴스는 상기 최종 액션 종료 확률값, 상기 최종 액션 시작 확률값, 액션의 클래스 및 클래스 확률값을 포함하는

기록매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 비디오의 동작을 실시간으로 검출하기 위한 기술에 관한 것이다.

배경 기술

[0003] 비디오에서 일어나고 있는 일을 이해하는 것은 오랜 연구 문제이며, 컴퓨터 비전과 기계 학습에서 여전히 어려운 작업에 속한다.

[0004] 최근, 딥러닝의 발전은 해당 분야의 많은 연구자들의 관심을 끌었고, 이는 다양한 비디오 이해 작업(video understanding task)에서 상당한 진전을 가져왔다. 예컨대, 액션 분류(action classification), 액션 시간 추정(temporal action localization), 비디오 질답(video question and answering), 비디오 요약(video summarization)과 같은 다양한 작업에서 비디오 이해의 많은 성공적인 진전을 이루었다. 이러한 작업의 대부분은 추론할 때 전체 비디오에 액세스할 수 있는 상태인 '오프라인 환경(offline setting)'으로 비디오 이해 작업을 다룬다.

[0005] 라이브 스트리밍 서비스 및 보안 감시 카메라의 스트리밍 비디오 포맷이 증가함에 따라 비디오 스트림을 처리하는 방법에 대한 연구 또한 계속 증가하고 있다. 특히, 요즘은 강의, 화상회의, 유튜브 채널 등 원격 소통의 도구로 스트리밍 영상이 성행하고 있다.

- [0006] 이러한 스트리밍 비디오를 이해하는 작업을 수행할 때는 알 수 없는 길이의 비디오를 입력으로 삼고, 비디오를 이해하는 데 전체 비디오를 액세스 할 수 없으며, 과거 프레임부터 현재 스트리밍 받은 프레임까지만 사용할 수 있다.
- [0007] 이러한 제약을 극복하기 위해 프레임 단위로 비디오 스트리밍을 분석하는 기술이나 비디오 액션을 추정하는 기술이 연구되고는 있으나, 정확한 동작 예측이 힘들고 대부분 오프라인 환경에 최적화가 되어 있다.
- [0008] 따라서, 온라인 환경에서 스트리밍 비디오를 실시간으로 분석할 수 있는 새로운 방식의 비디오 동작 검출 기술이 필요한 실정이다.

선행기술문헌

특허문헌

- [0010] (특허문헌 0001) 등록특허공보 제10-2420887호 (2022년07월15일 등록공고)
- (특허문헌 0002) 등록특허공보 제10-0420747호 (2004년03월02일 등록공고)

발명의 내용

해결하려는 과제

- [0011] 본 발명의 실시예에서는, 액션 인스턴스(action instance) 단위로 비디오 스트리밍을 실시간으로 분석하여 온라인 기반의 액션 시간 추정(online temporal action localization) 문제를 해결할 수 있는 비디오 동작 검출 장치 및 방법을 제안하고자 한다.
- [0012] 본 발명의 실시예에서는, 온라인 비디오의 액션 전체의 컨텍스트를 취하여 높은 정확성을 확보할 수 있는 비디오 동작 검출 장치 및 방법을 제안하고자 한다.
- [0013] 본 발명의 실시예에서는, 액션의 컨텍스트와 액션 컨텍스트의 변화를 모니터링하여 저지연 및 고성능의 검출 결과를 제공할 수 있는 비디오 동작 검출 장치 및 방법을 제안하고자 한다.
- [0014] 본 발명이 해결하고자 하는 과제는 상기에서 언급한 것으로 제한되지 않으며, 언급되지 않은 또 다른 해결하고자 하는 과제는 아래의 기재들로부터 본 발명이 속하는 통상의 지식을 가진 자에 의해 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0016] 본 발명의 실시예에 따르면, 비디오 프레임을 스트리밍 방식으로 입력하는 프레임 입력부; 상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 저장부; 및 상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 액션 인스턴스 생성부;를 포함하는 비디오 동작 검출 장치를 제공할 수 있다.
- [0017] 여기서, 상기 저장부는, 상기 비디오 프레임을 인코딩하여 피처를 추출하는 피처 추출부; 상기 피처를 정방향(forward pass)으로 탐색하여 액션의 종료 지점을 검출하는 액션 종료 검출부; 및 상기 피처를 역방향(backward pass)으로 탐색하여 액션의 시작 지점을 검출하는 액션 시작 검출부;를 포함할 수 있다.
- [0018] 또한, 상기 액션 종료 검출부는, 상기 피처를 2종의 확률값으로 출력하는 멀티 헤드 검출부; 및 상기 2종의 확률값을 기초로, 최종 액션 종료 확률값을 출력하는 액션 종료 감지 개선부;를 포함할 수 있다.
- [0019] 또한, 상기 멀티 헤드 검출부는, 액션이 존재할 확률값을 출력하는 액션 감지 헤드; 및 액션의 종료 지점이 존재할 액션 종료 확률값을 출력하는 액션 종료 감지 헤드;를 포함할 수 있다.
- [0020] 또한, 상기 액션 감지 헤드와 상기 액션 종료 감지 헤드 각각은, LSTM(long short-term memory) 네트워크, FC(fully-connected) 레이어 및 소프트맥스(softmax) 레이어를 포함할 수 있다.
- [0021] 또한, 상기 액션 종료 감지 개선부는, FC 레이어 및 소프트맥스 레이어를 포함할 수 있다.
- [0022] 또한, 상기 액션 시작 검출부는, LSTM 네트워크, FC 레이어 및 소프트맥스 레이어를 포함할 수 있다.

- [0023] 또한, 상기 액션 시작 검출부는, 액션의 시작 지점이 존재할 액션 시작 확률값을 출력할 수 있다.
- [0024] 또한, 상기 액션 시작 검출부는, 상기 액션 시작 확률값이 1회 상승한 후 하락하는 시점까지 상기 역방향으로 탐색하는 과정을 유지할 수 있다.
- [0025] 또한, 상기 액션 시작 검출부는, 상기 하락하는 시점에서의 이전 프레임의 액션 시작 확률값을 최종 액션 시작 확률값으로 결정할 수 있다.
- [0026] 또한, 상기 액션 인스턴스 생성부는, 상기 액션 종료 확률값이 기 설정된 임계값을 초과하는 경우에 상기 액션 시작 검출부를 호출할 수 있다.
- [0027] 또한, 상기 액션 인스턴스 생성부는 상기 액션의 시작 지점과 상기 액션의 종료 지점을 기초로 상기 액션 인스턴스를 생성하며, 상기 액션 인스턴스는 상기 최종 액션 종료 확률값, 상기 최종 액션 시작 확률값 및 액션의 클래스를 포함할 수 있다.
- [0028] 또한, 상기 피쳐 추출부는, 상기 피쳐를 시퀀스 형태로 보관할 수 있다.
- [0029] 본 발명의 실시예에 따른 비디오 동작 검출 장치의 비디오 동작 검출 방법에 있어서, 비디오 프레임을 스트리밍 방식으로 입력하는 단계; 상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 단계; 및 상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계;를 포함하는 비디오 동작 검출 방법을 제공할 수 있다.
- [0030] 여기서, 상기 확률값을 출력하는 단계는, 상기 비디오 프레임을 인코딩하여 피쳐를 추출하는 단계; 상기 피쳐를 정방향으로 탐색하여 액션의 종료 지점을 검출하는 단계; 및 상기 피쳐를 역방향으로 탐색하여 액션의 시작 지점을 검출하는 단계;를 포함할 수 있다.
- [0031] 또한, 상기 종료 지점을 검출하는 단계는, 상기 피쳐를 2종의 확률값으로 출력하는 단계; 및 상기 2종의 확률값을 기초로, 최종 액션 종료 확률값을 출력하는 단계;를 포함할 수 있다.
- [0032] 또한, 상기 2종의 확률값으로 출력하는 단계는, 액션이 존재할 확률값을 출력하는 단계; 및 액션의 종료 지점이 존재할 액션 종료 확률값을 출력하는 단계;를 포함할 수 있다.
- [0033] 또한, 상기 시작 지점을 검출하는 단계는, 상기 액션 시작 확률값이 1회 상승한 후 하락하는 시점까지 상기 역방향으로 탐색하는 단계를 포함할 수 있다.
- [0034] 또한, 상기 시작 지점을 검출하는 단계는, 상기 하락하는 시점에서의 이전 프레임의 액션 시작 확률값을 최종 액션 시작 확률값으로 결정하는 단계를 포함할 수 있다.
- [0035] 또한, 상기 액션 종료 확률값이 기 설정된 임계값을 초과하는 경우에 상기 액션의 시작 지점을 검출하는 단계를 수행하는 것을 특징으로 할 수 있다.
- [0036] 본 발명의 실시예에 따르면, 컴퓨터 프로그램을 저장하고 있는 컴퓨터 판독 가능 기록매체로서, 상기 컴퓨터 프로그램은, 비디오 동작 검출 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고, 상기 방법은, 비디오 프레임을 스트리밍 방식으로 입력하는 단계; 상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 단계; 및 상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계;를 포함할 수 있다.
- [0037] 본 발명의 실시예에 따르면, 컴퓨터 판독 가능 기록매체에 저장된 컴퓨터 프로그램으로서, 상기 컴퓨터 프로그램은, 비디오 동작 검출 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고, 상기 방법은, 비디오 프레임을 스트리밍 방식으로 입력하는 단계; 상기 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력하는 단계; 및 상기 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력하는 단계;를 포함할 수 있다.

발명의 효과

- [0039] 본 발명의 실시예에 의하면, 온라인 액션 시간 추정 문제를 해결하여 스트리밍 비디오에서 액션의 시작점, 끝점 및 클래스를 생성하여 인스턴스 단위의 실시간 동작 탐지를 수행할 수 있다. 또한, 본 발명의 실시예에 의하면, 정방향 패스에서 액션의 종료점을 먼저 찾아 온라인 작업에서 필요한 높은 응답성을 해결하면서, 역방향 패스를 통해 액션의 시작점을 찾아 액션 전체의 컨텍스트를 취하여 높은 정확성을 확보할 수 있다. 또한, 본 발명의 실시예에 의하면, 액션 종료 검출 시 멀티 헤드 검출기로 액션의 컨텍스트를 파악하고, 액션 종료 감지 개선부를

통해 컨텍스트의 변화 양상을 관찰하여 단일 액션 종료 검출기를 사용하는 것보다 저지연, 고성능 검출을 수행할 수 있다.

도면의 간단한 설명

[0041] 도 1은 본 발명의 실시예에 따른 비디오 동작 검출 장치(1)를 설명하는 블록도이다.

도 2는 도 1의 저장부(20)의 구체적인 구성을 나타낸 블록도이다.

도 3은 도 2의 액션 종료 검출부(210)의 구체적인 구성을 나타낸 블록도이다.

도 4는 도 3의 멀티 헤드 검출부(212)의 구체적인 구성을 나타낸 블록도이다.

도 5는 본 발명의 실시예에 따른 비디오 동작 검출 장치(1)의 저장부(20)에 대한 세부 구성 및 동작 과정을 예시적으로 설명하는 개념도이다.

발명을 실시하기 위한 구체적인 내용

[0042] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명의 범주는 청구항에 의해 정의될 뿐이다.

[0043] 본 발명의 실시예들을 설명함에 있어서 공지 기능 또는 구성에 대한 구체적인 설명은 본 발명의 실시예들을 설명함에 있어 실제로 필요한 경우 외에는 생략될 것이다. 그리고 후술되는 용어들은 본 발명의 실시예에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

[0044] 스트리밍 비디오를 분석하는 기술로서, 온라인 액션 감지(online action detection)와 액션 시작의 온라인 감지(online detection of action start)와 같이 프레임 단위로 분할되어 스트리밍 비디오를 분석하는 연구 방식이 존재한다.

[0045] 이러한 온라인 액션 감지 및 액션 시작의 온라인 감지는, 액션 클래스 또는 액션 시작점을 프레임 단위로 예측을 하나, 그것이 반드시 액션 인스턴스 단위의 이해와 연결되지는 않는다. 예를 들어, 3개의 연속된 점프 행동을 포함하는 30프레임 비디오 클립의 경우, 온라인 액션 감지의 접근 방식은 30개의 프레임에 30개의 동작 예측을 출력하고, 세 개의 점프 행동이 있다고 예측할 수 없다.

[0046] 따라서, 본 발명의 실시예에서는, 온라인 액션 감지의 접근 방식으로 온라인 액션 시간 추정 작업을 수행하기 위해, 프레임을 액션 인스턴스로 그룹화하는 과정을 적용하고자 한다.

[0047] 본 발명의 실시예에서는 액션 시작의 온라인 감지를 액션 끝점의 온라인 탐지(online detection of action end)로 확장하고, 시작점과 끝점을 탐지하여 온라인 액션 시간 추정 작업에 대해 적용하기 위한 기술을 제안하고자 한다.

[0048] 본 발명의 실시예에서는, 액션의 시작점과 끝점을 일관된 한 쌍으로 감지하고자, 액션 시작과 액션 인스턴스 사이의 프레임을 그룹화하는 데 필요한 액션의 가장 신뢰할 수 있는 액션의 끝점 후보를 결정하기 위한 기술을 제안하고자 한다.

[0049] 본 발명의 실시예에서는, 액션 인스턴스 단위로 비디오 스트리밍을 실시간으로 분석하여 온라인 기반의 액션 시간 추정 문제를 해결할 수 있으며, 온라인 비디오의 액션 전체의 컨텍스트를 취하여 높은 정확성을 확보할 수 있고, 액션의 컨텍스트와 액션 컨텍스트의 변화를 모니터링하여 저지연 및 고성능의 검출 결과를 제공할 수 있는 비디오 동작 검출 장치 및 방법을 제안하고자 한다.

[0051] 이하, 첨부된 도면을 참조하여 본 발명의 실시예에 대해 상세히 설명하기로 한다.

[0052] 도 1은 본 발명의 실시예에 따른 비디오 동작 검출 장치(1)를 설명하는 블록도이다.

[0053] 도 1에 도시한 바와 같이, 본 발명의 실시예에 따른 비디오 동작 검출 장치(1)는, 프레임 입력부(10), 저장부(20) 및 액션 인스턴스 생성부(30)를 포함할 수 있다.

- [0054] 프레임 입력부(10)는 비디오 프레임을 스트리밍 방식으로 실시간 입력 받을 수 있다.
- [0055] 저장부(20)는 프레임 입력부(10)를 통해 입력되는 비디오 프레임으로부터 액션 시점에 대한 확률값을 출력할 수 있다.
- [0056] 액션 인스턴스 생성부(30)는 저장부(20)의 확률값을 기초로 액션 인스턴스를 생성하여 인스턴스 단위의 비디오 동작 검출 결과를 출력할 수 있다.
- [0058] 도 2는 이러한 저장부(20)의 구체적인 구성을 나타낸 블록도이다.
- [0059] 도 2에 도시한 바와 같이, 저장부(20)는 피처 추출부(feature extractor, 200), 액션 종료 검출부(action end detector, 210) 및 액션 시작 검출부(action start detector, 220)를 포함할 수 있다.
- [0060] 피처 추출부(200)는 프레임 입력부(10)으로부터 비디오 프레임을 실시간으로 입력 받아 피처를 추출할 수 있다.
- [0061] 액션 종료 검출부(210)는 피처 추출부(200)에서 추출된 피처를 정방향(forward pass)으로 탐색하여 액션의 종료 지점을 검출할 수 있다.
- [0063] 이러한 액션 종료 검출부(210)는 도 3에 도시한 바와 같이, 멀티 헤드 검출부(multi-head detector, 212)와 액션 종료 감지 개선부(end detection refinement, 214)를 포함할 수 있다.
- [0064] 멀티 헤드 검출부(212)는 피처 추출부(200)에서 추출된 피처를, 서로 다른 2종의 확률값으로 출력하는 역할을 할 수 있다.
- [0065] 액션 종료 감지 개선부(214)는 멀티 헤드 검출부(212)로부터 출력되는 2종의 확률값을 기초로, 최종 액션 종료 확률값을 출력하는 역할을 할 수 있다.
- [0067] 도 4는 도 3의 액션 종료 검출부(210)의 구성 중 멀티 헤드 검출부(212)의 구체적인 구성을 나타낸 블록도로서, 본 발명의 실시예에 따른 멀티 헤드 검출부(212)는 액션 감지 헤드(action detection head, 212a)와 액션 종료 감지 헤드(action end head, 212b)를 포함할 수 있다.
- [0068] 액션 감지 헤드(212a)는 액션이 존재할 확률값을 출력하는 역할을 할 수 있다.
- [0069] 액션 종료 감지 헤드(212b)는 액션의 종료 지점이 존재할 액션 종료 확률값을 출력하는 역할을 할 수 있다.
- [0071] 다시 도 1 및 도 2를 참조하면, 액션 시작 검출부(220)는 액션의 시작 지점이 존재할 액션 시작 확률값을 출력하는 수단으로서, 피처 추출부(200)에서 추출된 피처를 역방향(backward pass)으로 탐색하여 액션의 시작 지점을 검출하는 역할을 수행할 수 있다.
- [0072] 이러한 액션 시작 검출부(220)는 액션 시작 확률값이 1회 상승한 후 하락하는 시점까지 역방향 탐색을 유지할 수 있으며, 하락하는 시점에서의 이전 프레임의 액션 시작 확률값을 최종 액션 시작 확률값으로 결정할 수 있다.
- [0073] 이때, 액션 시작 검출부(220)는 도 4의 액션 종료 감지 헤드(212b)에서 출력된 액션 종료 확률값이 기 설정된 임계값을 초과하는 경우에, 액션 인스턴스 생성부(30)에 의해 호출될 수 있다. 즉, 액션 인스턴스 생성부(30)는 액션 종료 검출부(210)의 액션 종료 확률값을 모니터링하고, 이러한 액션 종료 확률값에 종속되어 액션 시작 검출부(220)의 호출을 명령할 수 있다.
- [0074] 여기서, 액션 인스턴스 생성부(30)는 액션의 시작 지점과 액션의 종료 지점을 기초로 상술한 액션 인스턴스를 생성할 수 있다.
- [0075] 액션 인스턴스 생성부(30)에 의해 생성되는 액션 인스턴스는 액션 종료 검출부(210)로부터 생성된 최종 액션 종료 확률값과 액션 시작 검출부(220)로부터 생성된 최종 액션 시작 확률값을 포함할 수 있다. 또한, 액션 인스턴스는 액션의 클래스를 더 포함할 수 있다.
- [0076] 따라서, 액션 인스턴스 생성부(30)는 최종 액션 종료 확률값과 최종 액션 시작 확률값, 그리고 액션의 클래스를 조합하여 하나의 액션 인스턴스를 생성할 수 있다.
- [0078] 도 5는 본 발명의 실시예에 따른 비디오 동작 검출 장치(1)의 저장부(20)에 대한 세부 구성 및 동작 과정을 예시적으로 설명하는 개념도이다.
- [0079] 도 5에 도시한 바와 같이, 피처 추출부(200)는 프레임 입력부(10)를 통해 실시간으로 입력되는 비디오 프레임을

기 설정된 개수만큼 취합하여 인코딩하는 과정을 거칠 수 있다. 피쳐 추출부(200)는 이러한 기 설정 개수의 비디오 프레임을 인코딩하는 과정을 통해 피쳐를 생성할 수 있으며, 생성된 피쳐를 시퀀스 형태로 저장할 수 있다.

[0080] 액션 종료 검출부(210)는 피쳐 추출부(200)로부터 추출된 피쳐를 정방향으로 탐색하여 액션의 종료 지점을 검출하는 수단으로서, 멀티 헤드 검출부(212)와 액션 종료 감지 개선부(214)를 포함할 수 있다.

[0081] 멀티 헤드 검출부(212)는 다시 액션 감지 헤드(212a)와 액션 종료 감지 헤드(212b)를 포함할 수 있으며, 이러한 액션 감지 헤드(212a)와 액션 종료 감지 헤드(212b) 각각은, LSTM(long short-term memory) 네트워크, FC(fully-connected) 레이어 및 소프트맥스(softmax) 레이어를 포함하는 신경망 구조로 이루어질 수 있다. 따라서, 액션 감지 헤드(212a)를 통해 액션이 존재할 확률값을 출력할 수 있으며, 액션 종료 감지 헤드(212b)를 통해 액션의 종료 지점이 존재할 액션 종료 확률값을 출력할 수 있다.

[0082] 액션 종료 감지 개선부(214)는 멀티 헤드 검출부(212)의 액션 감지 헤드(212a)와 액션 종료 감지 헤드(212b) 각각으로부터 출력되는 2종의 확률값을 기초로 최종 액션 종료 확률값을 출력하는 수단으로서, FC 레이어 및 소프트맥스 레이어를 포함하는 신경망 구조로 이루어질 수 있다.

[0083] 액션 시작 검출부(220)는 피쳐 추출부(200)로부터 추출된 피쳐를 역방향으로 탐색하여 액션의 시작 지점을 검출하는 수단으로서, LSTM 네트워크, FC 레이어 및 소프트맥스 레이어를 포함하는 신경망 구조로 이루어질 수 있다.

[0084] 액션 인스턴스 생성부(30)는 감지된 액션 종료 지점(최종 액션 종료 확률값)과 액션 시작 지점(최종 액션 시작 확률값)을 조합하여 하나의 액션 인스턴스를 생성할 수 있다.

[0085] 도 5에서 상단 그래프는 최종 액션 종료 확률값을, 하단 그래프는 최종 액션 시작 확률값을 각각 표현하고 있다. 각각의 그래프에서 x축은 비디오 프레임을, y축은 확률값을 나타낸다.

[0086] 여기서, 액션 인스턴스는 액션의 클래스, 클래스 확률값, 시작 지점 프레임 및 종료 지점 프레임의 벡터를 포함할 수 있다. 따라서, 액션 인스턴스 생성부(30)로부터 생성된 액션 인스턴스에 의해 비디오 동작 검출 장치(1)에서 제공되는 출력 결과는 액션 범위, 액션 종류 등으로 표현될 수 있다. 액션 범위라 함은 상술한 액션 시작 지점과 액션 종료 지점을 의미할 수 있으며, 액션 종류는, 예를 들어 배구, 농구와 같은 특정 스포츠 종목을 의미할 수 있다.

[0088] 이상 설명한 바와 같은 본 발명의 실시예에 의하면, 온라인 액션 시간 추정 문제를 해결하여 스트리밍 비디오에서 액션의 시작점, 끝점 및 클래스를 생성하여 인스턴스 단위의 실시간 동작 탐지를 수행할 수 있으며, 정방향 패스에서 액션의 종료점을 먼저 찾아 온라인 작업에서 필요한 높은 응답성을 해결하면서 역방향 패스를 통해 액션의 시작점을 찾아 액션 전체의 컨텍스트를 취하여 높은 정확성을 확보할 수 있을 뿐만 아니라, 액션 종료 검출 시 멀티 헤드 검출기로 액션의 컨텍스트를 파악하고, 액션 종료 감지 개선부를 통해 컨텍스트의 변화 양상을 관찰하여 단일 액션 종료 검출기를 사용하는 것보다 저지연, 고성능 검출을 수행하도록 구현한 것이다.

[0089] 한편, 첨부된 블록도의 각 블록과 흐름도의 각 단계의 조합들은 컴퓨터 프로그램 인스트럭션들에 의해 수행될 수도 있다. 이들 컴퓨터 프로그램 인스트럭션들은 범용 컴퓨터, 특수용 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서에 탑재될 수 있으므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서를 통해 수행되는 그 인스트럭션들이 블록도의 각 블록에서 설명된 기능들을 수행하는 수단을 생성하게 된다.

[0090] 이들 컴퓨터 프로그램 인스트럭션들은 특정 방식으로 기능을 구현하기 위해 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 지향할 수 있는 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 기록매체(또는 메모리) 등에 저장되는 것도 가능하므로, 그 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 기록매체(또는 메모리)에 저장된 인스트럭션들은 블록도의 각 블록에서 설명된 기능을 수행하는 인스트럭션 수단을 내포하는 제조 품목을 생산하는 것도 가능하다.

[0091] 그리고, 컴퓨터 프로그램 인스트럭션들은 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에 탑재되는 것도 가능하므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에서 일련의 동작 단계들이 수행되어 컴퓨터로 실행되는 프로세스를 생성해서 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 수행하는 인스트럭션들은 블록도의 각 블록에서 설명된 기능들을 실행하기 위한 단계들을 제공하는 것도 가능하다.

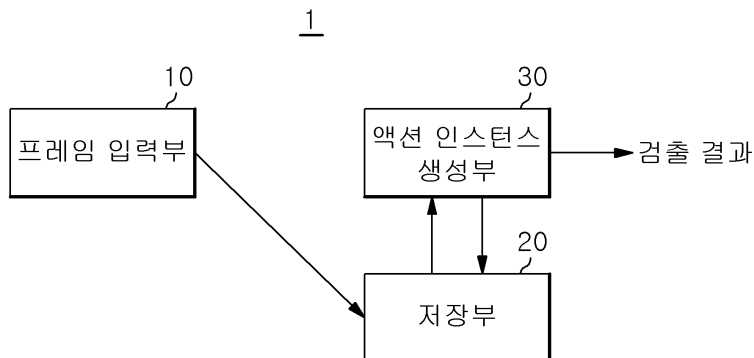
[0092] 또한, 각 블록은 특정된 논리적 기능(들)을 실행하기 위한 적어도 하나 이상의 실행 가능한 인스트럭션들을 포함하는 모듈, 세그먼트 또는 코드의 일부를 나타낼 수 있다. 또, 몇 가지 대체 실시 예들에서는 블록들에서 언급된 기능들이 순서를 벗어나서 발생하는 것도 가능함을 주목해야 한다. 예컨대, 잇달아 도시되어 있는 두 개의 블록들은 사실 실질적으로 동시에 수행되는 것도 가능하고 또는 그 블록들이 때때로 해당하는 기능에 따라 역순으로 수행되는 것도 가능하다.

부호의 설명

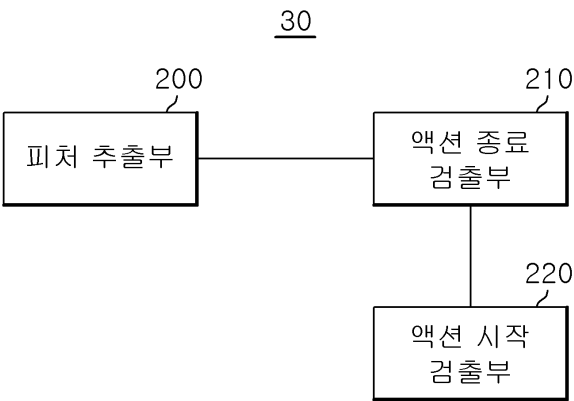
[0094] 1: 비디오 동작 검출 장치
10: 프레임 입력부
20: 저장부
30: 액션 인스턴스 생성부
200: 피처 추출부
210: 액션 종료 검출부
212: 멀티 헤드 검출부
212a: 액션 감지 헤드
212b: 액션 종료 감지 헤드
214: 액션 종료 감지 개선부
220: 액션 시작 검출부

도면

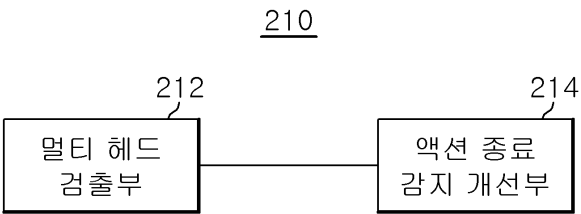
도면1



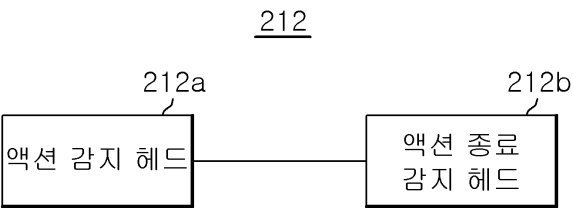
도면2



도면3



도면4



도면5

