

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2022-0109622
(43) 공개일자 2022년08월05일

(51) 국제특허분류(Int. Cl.)

G06K 9/00 (2022.01) G06N 3/04 (2006.01)

G06N 3/08 (2006.01) G06T 5/50 (2006.01)

(52) CPC특허분류

G06V 20/40 (2022.01)

G06N 3/0454 (2013.01)

(21) 출원번호 10-2021-0012863

(22) 출원일자 2021년01월29일

심사청구일자 없음

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

박형민

서울특별시 강남구 삼성로 212, 3동 1002호 (대치동, 은마아파트)

강민석

서울특별시 마포구 백범로 13, 506호(대흥동, 신촌르메이에르타운II)

박래홍

경기도 고양시 일산동구 강촌로 191, 407동 601호 (마두동, 백마마을)

(74) 대리인

이준영

전체 청구항 수 : 총 17 항

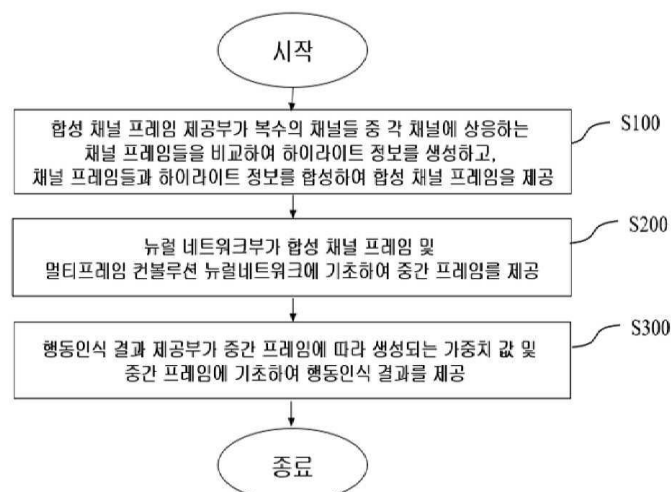
(54) 발명의 명칭 동영상 기반 행동인식 장치 및 이의 동작방법

(57) 요약

본 발명의 실시예에 따른 동영상 기반 행동인식 장치의 동작방법에서는, 합성 채널 프레임 제공부가 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 채널 프레임들과 하이라이트 정보를 합성하여 합성 채널 프레임을 제공할 수 있다. 뉴럴 네트워크부가 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공할 수 있다. 행동인식 결과 제공부가 중간 프레임에 따라 생성되는 가중치 값 및 중간 프레임에 기초하여 행동인식 결과를 제공할 수 있다.

본 발명에 따른 동영상 기반 행동인식 장치의 동작방법에서는, 채널별로 제공되는 채널 프레임들로부터 합성되는 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 행동인식 결과를 제공함으로써 동영상 내의 이벤트 발생을 보다 효과적으로 검출할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06N 3/08 (2013.01)

G06T 5/50 (2013.01)

G06T 2207/10016 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1615011433

과제번호 153358

부처명 국토교통부

과제관리(전문)기관명 국토교통과학기술진흥원

연구사업명 스마트도로조명플랫폼개발및실증연구(R&D)(국토부)

연구과제명 스마트 도로조명 활용 도시재난안전관리 연계 기술 개발

기 여 율 1/1

과제수행기관명 서강대학교산학협력단

연구기간 2020.01.01 ~ 2020.12.31

명세서

청구범위

청구항 1

합성 채널 프레임 제공부가 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 합성 채널 프레임을 제공하는 단계;

뉴럴 네트워크부가 상기 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공하는 단계; 및

행동인식 결과 제공부가 상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공하는 단계를 포함하는 동영상 기반 행동인식 장치의 동작방법.

청구항 2

제1항에 있어서,

상기 뉴럴 네트워크부는,

상기 복수의 채널들의 각각에 상응하는 상기 합성 채널 프레임을 합성하여 그레이 합성 프레임을 생성하는 그레이 합성 프레임 생성부; 및

상기 그레이 합성 프레임에 기초하여 상기 중간 프레임을 생성하는 멀티프레임 컨볼루션 뉴럴네트워크를 포함하는 것을 특징으로 하는 동영상 기반 행동인식 장치의 동작방법.

청구항 3

제2항에 있어서,

상기 복수의 채널들은 제1 채널, 제2 채널 및 제3 채널을 포함하고,

상기 제1 채널은 레드(Red) 채널이고, 상기 제2 채널은 그린(Green)이고, 상기 제3 채널은 블루(Blue) 채널인 것을 특징으로 하는 동영상 기반 행동인식 장치의 동작방법.

청구항 4

제3항에 있어서,

상기 그레이 합성 프레임 생성부는,

상기 레드 채널에 상응하는 레드 합성 채널 프레임, 상기 그린 채널에 상응하는 그린 합성 채널 프레임 및 상기 블루 채널에 상응하는 블루 합성 채널 프레임을 합성하여 상기 그레이 합성 프레임을 생성하는 것을 특징으로 하는 동영상 기반 행동인식 장치의 동작방법.

청구항 5

제4항에 있어서,

상기 그레이 합성 프레임 생성부는,

복수의 시간들의 각각에 상응하는 상기 레드 합성 채널 프레임, 상기 그린 합성 채널 프레임 및 상기 블루 합성 채널 프레임을 합성하여 상기 그레이 합성 프레임을 생성하고,

상기 복수의 시간들 중 연속되는 3개의 시간에 상응하는 상기 그레이 합성 프레임을 하나의 입력단위로 구성하여 상기 멀티프레임 컨볼루션 뉴럴네트워크에 제공하는 것을 특징으로 하는 동영상 기반 행동인식 장치의 동작방법.

청구항 6

제5항에 있어서,

상기 행동인식 결과 제공부는,

상기 중간 프레임을 에버리지 폴링하여 폴링 데이터를 생성하고, 상기 폴링 데이터에 기초하여 상기 가중치 값을 생성하고,

상기 가중치 값에 기초하여 상기 행동인식 결과를 제공하는 것을 특징으로 하는 동영상 기반 행동인식 장치의 동작방법.

청구항 7

제6항에 있어서,

상기 행동인식 결과 제공부는,

상기 중간 프레임을 각각 에버리지 폴링하여 중간 폴링 데이터를 생성하는 제1 폴링부; 및

상기 중간 폴링 데이터를 에버리지 폴링하여 폴링 데이터를 제공하는 제2 폴링부를 포함하는 것을 특징으로 하는 동영상 기반 행동인식 장치의 동작방법.

청구항 8

제7항에 있어서,

상기 행동인식 결과 제공부는,

상기 중간 폴링 데이터 및 상기 가중치 값에 기초하여 상기 행동인식 결과를 제공하는 것을 특징으로 하는 동영상 기반 행동인식 장치의 동작방법.

청구항 9

복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 합성 채널 프레임을 제공하는 합성 채널 프레임 제공부;

상기 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공하는 뉴럴 네트워크부; 및

상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공하는 행동인식 결과 제공부를 포함하는 동영상 기반 행동인식 장치.

청구항 10

제9항에 있어서,

상기 뉴럴 네트워크부는 상기 복수의 채널들의 각각에 상응하는 상기 합성 채널 프레임을 합성하여 그레이 합성 프레임을 생성하는 것을 특징으로 하는 동영상 기반 행동인식 장치.

청구항 11

합성 채널 프레임 제공부가 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 합성 채널 프레임을 제공하는 단계;

뉴럴 네트워크부가 상기 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공하는 단계;

행동인식 결과 제공부가 상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공하는 단계; 및

디스플레이부가 상기 행동인식 결과에 따라 생성되는 알림 정보를 디스플레이하는 단계를 포함하는 동영상 기반 행동인식 시스템의 동작방법.

청구항 12

제11항에 있어서,

상기 뉴럴 네트워크부는,

상기 복수의 채널들의 각각에 상응하는 상기 합성 채널 프레임을 합성하여 그레이 합성 프레임을 생성하는 그레이 합성 프레임 생성부; 및

상기 그레이 합성 프레임에 기초하여 상기 중간 프레임을 생성하는 멀티프레임 컨볼루션 뉴럴네트워크를 포함하는 것을 특징으로 하는 동영상 기반 행동인식 시스템의 동작방법.

청구항 13

복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 합성 채널 프레임을 제공하는 합성 채널 프레임 제공부;

상기 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공하는 뉴럴 네트워크부;

상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공하는 행동인식 결과 제공부; 및

상기 행동인식 결과에 따라 생성되는 알림 정보를 디스플레이하는 디스플레이부를 포함하는 동영상 기반 행동인식 시스템.

청구항 14

제13항에 있어서,

상기 복수의 채널들은 제1 채널, 제2 채널 및 제3 채널을 포함하고,

상기 제1 채널은 레드(Red) 채널이고, 상기 제2 채널은 그린(Green)이고, 상기 제3 채널은 블루(Blue) 채널인 것을 특징으로 하는 동영상 기반 행동인식 시스템.

청구항 15

합성 채널 프레임 제공부가 복수의 채널들 중 각 채널에 상응하는 채널 프레임들에 기초하여 합성 채널 프레임을 제공하는 단계;

뉴럴 네트워크부가 상기 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공하는 단계; 및

행동인식 결과 제공부가 상기 중간 프레임에 기초하여 행동인식 결과를 제공하는 단계를 포함하는 동영상 기반 행동인식 장치의 동작방법.

청구항 16

프레임 제공부가 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 프레임을 제공하는 단계;

뉴럴 네트워크부가 상기 프레임에 기초하여 중간 프레임을 제공하는 단계; 및

행동인식 결과 제공부가 상기 중간 프레임에 기초하여 행동인식 결과를 제공하는 단계를 포함하는 동영상 기반 행동인식 장치의 동작방법.

청구항 17

입력 프레임 제공부가 3 채널 프레임을 제공하는 단계;

뉴럴 네트워크부가 상기 3채널 프레임에 기초하여 중간 프레임을 제공하는 단계; 및

행동인식 결과 제공부가 상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공하는 단계를 포함하는 동영상 기반 행동인식 장치의 동작방법.

발명의 설명

기술 분야

[0001] 본 발명은 동영상 기반 행동인식 장치 및 이의 동작방법에 관한 것이다.

배경 기술

[0002] 최근, 주거지나 건물 주변의 방법을 위하여 CCTV를 사용하는 경우가 급속하게 증가하고 있다. 그러나, 사람이 CCTV로부터 제공되는 동영상상을 항상 감시하는 것은 비용과 시간적 낭비를 발생시킬 수 있다. 이와 같은 문제점을 해결하기 위하여 다양한 연구가 진행되고 있다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) (한국등록특허) 제10-1023951호 (등록일자, 2011.3.15)

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 채널별로 제공되는 채널 프레임들로부터 합성되는 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 행동인식 결과를 제공함으로써 동영상 내의 이벤트 발생을 보다 효과적으로 검출할 수 있는 동영상 기반 행동인식 장치의 동작방법을 제공하는 것이다.

과제의 해결 수단

[0005] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 동영상 기반 행동인식 장치의 동작방법에서는, 합성 채널 프레임 제공부가 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 합성 채널 프레임을 제공할 수 있다. 뉴럴 네트워크부가 상기 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공할 수 있다. 행동인식 결과 제공부가 상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공할 수 있다.

[0006] 일 실시예에 있어서, 상기 뉴럴 네트워크부는 그레이 합성 프레임 생성부 및 멀티프레임 컨볼루션 뉴럴네트워크를 포함할 수 있다. 그레이 합성 프레임 생성부는 상기 복수의 채널들의 각각에 상응하는 상기 합성 채널 프레임을 합성하여 그레이 합성 프레임을 생성할 수 있다. 멀티프레임 컨볼루션 뉴럴네트워크는 상기 그레이 합성 프레임에 기초하여 상기 중간 프레임을 생성할 수 있다.

[0007] 일 실시예에 있어서, 상기 복수의 채널들은 제1 채널, 제2 채널 및 제3 채널을 포함할 수 있다. 상기 제1 채널은 레드(Red) 채널이고, 상기 제2 채널은 그린(Green)이고, 상기 제3 채널은 블루(Blue) 채널일 수 있다.

[0008] 일 실시예에 있어서, 상기 그레이 합성 프레임 생성부는 상기 레드 채널에 상응하는 레드 합성 채널 프레임, 상기 그린 채널에 상응하는 그린 합성 채널 프레임 및 상기 블루 채널에 상응하는 블루 합성 채널 프레임을 합성하여 상기 그레이 합성 프레임을 생성할 수 있다.

[0009] 일 실시예에 있어서, 상기 그레이 합성 프레임 생성부는 복수의 시간들의 각각에 상응하는 상기 레드 합성 채널 프레임, 상기 그린 합성 채널 프레임 및 상기 블루 합성 채널 프레임을 합성하여 상기 그레이 합성 프레임을 생성할 수 있다. 상기 복수의 시간들 중 연속되는 3개의 시간에 상응하는 상기 그레이 합성 프레임을 하나의 입력 단위로 구성하여 상기 멀티프레임 컨볼루션 뉴럴네트워크에 제공할 수 있다.

[0010] 일 실시예에 있어서, 상기 행동인식 결과 제공부는 상기 중간 프레임을 에버리지 풀링하여 풀링 데이터를 생성하고, 상기 풀링 데이터에 기초하여 상기 가중치 값을 생성할 수 있다. 상기 행동인식 결과 제공부는 상기 가중치 값에 기초하여 상기 행동인식 결과를 제공할 수 있다.

[0011] 일 실시예에 있어서, 상기 행동인식 결과 제공부는 제1 풀링부 및 제2 풀링부를 포함할 수 있다. 제1 풀링부는

상기 중간 프레임을 각각 공간 축으로 에버리지 폴링하여 중간 폴링 데이터를 생성할 수 있다. 제2 폴링부는 상기 중간 폴링 데이터를 채널 축으로 에버리지 폴링하여 폴링데이터를 제공할 수 있다.

[0012] 일 실시예에 있어서, 상기 행동인식 결과 제공부는 상기 중간 폴링 데이터 및 폴리 커넥티드 레이어를 이용해서 생성되는 상기 가중치 값에 기초하여 상기 행동인식 결과를 제공할 수 있다.

[0013] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 동영상 기반 행동인식 장치는 합성 채널 프레임 제공부, 뉴럴 네트워크부 및 행동인식 결과 제공부를 포함할 수 있다. 합성 채널 프레임 제공부는 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 합성 채널 프레임을 제공할 수 있다. 뉴럴 네트워크부는 상기 합성 채널 프레임 및 멀티 프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공할 수 있다. 행동인식 결과 제공부는 상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공할 수 있다.

[0014] 일 실시예에 있어서, 상기 뉴럴 네트워크부는 상기 복수의 채널들의 각각에 상응하는 상기 합성 채널 프레임을 합성하여 그레이 합성 프레임을 생성할 수 있다.

[0015] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 동영상 기반 행동인식 시스템의 동작방법에서는, 합성 채널 프레임 제공부가 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 합성 채널 프레임을 제공할 수 있다. 뉴럴 네트워크부가 상기 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공할 수 있다. 행동인식 결과 제공부가 상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공할 수 있다. 디스플레이부가 상기 행동인식 결과에 따라 생성되는 알림 정보를 디스플레이할 수 있다.

[0016] 일 실시예에 있어서, 상기 뉴럴 네트워크부는 그레이 합성 프레임 생성부 및 멀티프레임 컨볼루션 뉴럴네트워크를 포함할 수 있다. 그레이 합성 프레임 생성부는 상기 복수의 채널들의 각각에 상응하는 상기 합성 채널 프레임을 합성하여 그레이 합성 프레임을 생성할 수 있다. 멀티프레임 컨볼루션 뉴럴네트워크는 상기 그레이 합성 프레임에 기초하여 상기 중간 프레임을 생성할 수 있다.

[0017] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 동영상 기반 행동인식 시스템은 합성 채널 프레임 제공부, 뉴럴 네트워크부, 결과 제공부 및 디스플레이부를 포함할 수 있다. 합성 채널 프레임 제공부는 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보를 생성하고, 상기 채널 프레임들과 상기 하이라이트 정보를 합성하여 합성 채널 프레임을 제공할 수 있다. 뉴럴 네트워크부는 상기 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임을 제공할 수 있다. 행동인식 결과 제공부는 상기 중간 프레임에 따라 생성되는 가중치 값 및 상기 중간 프레임에 기초하여 행동인식 결과를 제공할 수 있다. 디스플레이부는 상기 행동인식 결과에 따라 생성되는 알림 정보를 디스플레이할 수 있다.

[0018] 일 실시예에 있어서, 상기 복수의 채널들은 제1 채널, 제2 채널 및 제3 채널을 포함할 수 있다. 상기 제1 채널은 레드(Red) 채널이고, 상기 제2 채널은 그린(Green)이고, 상기 제3 채널은 블루(Blue) 채널일 수 있다.

[0019] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

발명의 효과

[0020] 이상과 같은 본 발명에 따르면 다음과 같은 효과가 있다.

[0021] 본 발명에 따른 동영상 기반 행동인식 장치의 동작방법에서는, 채널별로 제공되는 채널 프레임들로부터 합성되는 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 행동인식 결과를 제공함으로써 동영상 내의 이벤트 발생을 보다 효과적으로 검출할 수 있다.

[0022] 또한, 본 발명에 따른 동영상 기반 행동인식 장치의 동작방법에서는, 3채널 이미지를 1채널 이미지로 변환하여 멀티프레임 컨볼루션 뉴럴네트워크(2D CNN)에 입력하기 때문에 기존의 2D CNN보다 실시간 인식을 위한 연산량을 약 1/3 배 줄일 수 있어 행동인식의 성능을 높일 수 있다.

[0023] 이 밖에도, 본 발명의 실시 예들을 통해 본 발명의 또 다른 특징 및 이점들이 새롭게 파악될 수도 있을 것이다.

도면의 간단한 설명

- [0024] 도 1은 본 발명의 실시예들에 따른 동영상 기반 행동인식 장치의 동작방법을 나타내는 순서도이다.
- 도 2는 본 발명의 실시예들에 따른 동영상 기반 행동인식 장치를 나타내는 도면이다.
- 도 3은 도 2의 동영상 기반 행동인식 장치에 포함되는 합성 채널 프레임 제공부의 동작을 설명하기 위한 도면이다.
- 도 4는 도 1의 동영상 기반 행동인식장치의 일 동작예를 설명하기 위한 도면이다.
- 도 5는 도 2의 동영상 기반 행동인식 장치에 포함되는 뉴럴 네트워크부를 나타내는 도면이다.
- 도 6은 도 5의 뉴럴 네트워크부에 포함되는 그레이 합성 프레임 생성부의 동작을 설명하기 위한 도면이다.
- 도 7은 도 5의 뉴럴 네트워크부에 포함되는 멀티프레임 컨볼루션 뉴럴네트워크를 설명하기 위한 도면이다.
- 도 8은 도 2의 동영상 기반 행동인식 장치에 포함되는 행동인식 결과 제공부를 나타내는 도면이다.
- 도 9는 도 8의 행동인식 결과 제공부를 설명하기 위한 도면이다.
- 도 10은 본 발명의 실시예들에 따른 동영상 기반 행동인식 시스템의 동작방법을 나타내는 순서도이다.
- 도 11은 본 발명의 실시예들에 따른 동영상 기반 행동인식 시스템을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0025] 본 명세서에서 각 도면의 구성 요소들에 참조번호를 부가함에 있어서 동일한 구성 요소들에 한해서는 비록 다른 도면상에 표시되더라도 가능한한 동일한 번호를 가지도록 하고 있음에 유의하여야 한다.
- [0026] 한편, 본 명세서에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0027] 단수의 표현은 문맥상 명백하게 다르게 정의하지 않는 한, 복수의 표현을 포함하는 것으로 이해되어야 하는 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다.
- [0028] "포함하다" 또는 "가지다" 등의 용어는 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0029] 이하, 첨부되는 도면을 참고하여 상기 문제점을 해결하기 위해 고안된 본 발명의 바람직한 실시예들에 대해 상세히 설명한다.
- [0030] 도 1은 본 발명의 실시예들에 따른 동영상 기반 행동인식 장치의 동작방법을 나타내는 순서도이고, 도 2는 본 발명의 실시예들에 따른 동영상 기반 행동인식 장치를 나타내는 도면이고, 도 3은 도 2의 동영상 기반 행동인식 장치에 포함되는 합성 채널 프레임 제공부의 동작을 설명하기 위한 도면이고, 도 4는 도 1의 동영상 기반 행동인식장치의 일 동작예를 설명하기 위한 도면이다.
- [0031] 도 1 내지 4를 참조하면, 본 발명의 실시예에 따른 동영상 기반 행동인식 장치(10)는 합성 채널 프레임 제공부(100), 뉴럴 네트워크부(200) 및 행동인식 결과 제공부(300)를 포함할 수 있다. 본 발명의 실시예에 따른 동영상 기반 행동인식 장치(10)의 동작방법에서는, 합성 채널 프레임 제공부(100)가 복수의 채널들 중 각 채널에 상응하는 채널 프레임(CHF)들을 비교하여 하이라이트 정보(HIN)를 생성하고, 채널 프레임(CHF)들과 하이라이트 정보(HIN)를 합성하여 합성 채널 프레임(HCF)을 제공할 수 있다(S100). 예를 들어, 복수의 시간들은 제1 시간(T1), 제2 시간(T2), 제3 시간(T3) 및 제4 시간(T4)을 포함할 수 있고, 복수의 채널들은 제1 채널(C1), 제2 채널(C2) 및 제3 채널(C3)을 포함할 수 있다. 제1 채널(C1)은 레드(Red) 채널이고, 제2 채널(C2)은 그린(Green) 채널이고, 제3 채널(C3)은 블루(Blue) 채널일 수 있다. 제1 채널(C1)에 상응하는 제1 채널 프레임은 RGB 영상 중 레드 영상일 수 있고, 제2 채널(C2)에 상응하는 제2 채널 프레임은 RGB 영상 중 그린 영상일 수 있고, 제3 채널(C3)에 상응하는 제3 채널 프레임은 RGB 영상 중 블루 영상일 수 있다.
- [0032] 예를 들어, 동영상 기반 행동인식 장치(10)에 포함되는 합성 채널 프레임 제공부(100)는 복수의 시간들 중 제1 시간(T1)의 제1 채널 프레임(CHF1_1)을 제2 시간(T2)의 제1 채널 프레임(CHF2_1)과 비교하여 하이라이트 정보(HIN)를 생성할 수 있고, 제1 시간(T1)의 제2 채널 프레임(CHF1_2)을 제2 시간(T2)의 제2 채널 프레임(CHF2_2)과 비교하여 하이라이트 정보(HIN)를 생성할 수도 있다. 또한, 합성 채널 프레임 제공부(100)는 복수의 시간

들 중 제1 시간(T1)의 제3 채널 프레임(CHF1_3)을 제2 시간(T2)의 제3 채널 프레임(CHF2_3)과 비교하여 하이라이트 정보(HIN)를 생성할 수도 있다. 채널 프레임(CHF)들 간의 유클리디안 거리를 계산한 후, 컨볼루션 및 풀링 레이어를 이용해 움직이는 물체에 집중할 수 있는 하이라이트 정보(HIN)를 생성할 수 있다. 이 경우, 합성 채널 프레임 제공부(100)는 채널 프레임(CHF)들과 하이라이트 정보(HIN)를 합성하여 합성 채널 프레임(HCF)을 제공할 수 있다. 예를 들어, 합성 채널 프레임 제공부(100)는 제1 시간(T1)의 제1 채널 프레임(CHF1_1)에 하이라이트 정보(HIN)를 합성하여 제1 시간(T1)의 제1 합성 채널 프레임(HCF1_C1)을 제공할 수 있고, 제1 시간(T1)의 제2 채널 프레임(CHF1_2)에 하이라이트 정보(HIN)를 합성하여 제1 시간(T1)의 제2 합성 채널 프레임(HCF1_C2)을 제공할 수 있다. 또한, 합성 채널 프레임 제공부(100)는 제1 시간(T1)의 제3 채널 프레임(CHF1_3)에 하이라이트 정보(HIN)를 합성하여 제1 시간(T1)의 제3 합성 채널 프레임(HCF1_C3)을 제공할 수 있다. 동일한 방식으로 제2 시간(T2) 내지 제N 시간의 합성 채널 프레임(HCF)이 생성될 수 있다.

[0033] 뉴럴 네트워크부(200)가 합성 채널 프레임(HCF) 및 멀티프레임 컨볼루션 뉴럴네트워크(220)에 기초하여 중간 프레임(MF)을 제공할 수 있다(S200). 행동인식 결과 제공부(300)가 중간 프레임(MF)에 따라 생성되는 가중치 값(WT) 및 중간 프레임(MF)에 기초하여 행동인식 결과(ARR)를 제공할 수 있다(S300).

[0034] 본 발명에 따른 동영상 기반 행동인식 장치(10)의 동작방법에서는, 채널별로 제공되는 채널 프레임(CHF)들로부터 합성되는 합성 채널 프레임(HCF) 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 행동인식 결과(ARR)를 제공함으로써 동영상 내의 이벤트 발생을 보다 효과적으로 검출할 수 있다.

[0035] 또한, 본 발명에 따른 동영상 기반 행동인식 장치의 동작방법에서는, 3채널 이미지를 1채널 이미지로 변환하여 멀티프레임 컨볼루션 뉴럴네트워크(2D CNN)에 입력하기 때문에 기존의 2D CNN보다 실시간 인식을 위한 연산량을 약 1/3 배 줄일 수 있어 행동인식의 성능을 높일 수 있다.

[0036] 도 4는 도 1의 동영상 기반 행동인식장치의 일 동작예를 설명하기 위한 도면이고, 5는 도 2의 동영상 기반 행동인식 장치에 포함되는 뉴럴 네트워크부를 나타내는 도면이고, 도 6은 도 5의 뉴럴 네트워크부에 포함되는 그레이 합성 프레임 생성부의 동작을 설명하기 위한 도면이고, 도 7은 도 5의 뉴럴 네트워크부에 포함되는 멀티프레임 컨볼루션 뉴럴네트워크를 설명하기 위한 도면이다.

[0037] 도 1 내지 7을 참조하면, 본 발명의 실시예에 따른 동영상 기반 행동인식 장치(10)는 합성 채널 프레임 제공부(100), 뉴럴 네트워크부(200) 및 행동인식 결과 제공부(300)를 포함할 수 있다. 뉴럴 네트워크부(200)는 그레이 합성 프레임 생성부(210) 및 멀티프레임 컨볼루션 뉴럴네트워크(220)를 포함할 수 있다. 그레이 합성 프레임 생성부(210)는 복수의 채널들의 각각에 상응하는 합성 채널 프레임(HCF)을 합성하여 그레이 합성 프레임(GHF)을 생성할 수 있다. 예를 들어, 합성 채널 프레임(HCF)은 제1 시간(T1)의 제1 합성 채널 프레임(HCF1_C1), 제1 시간(T1)의 제2 합성 채널 프레임(HCF1_C2) 및 제1 시간(T1)의 제3 합성 채널 프레임(HCF1_C3)을 포함할 수 있다. 제2 시간(T2) 내지 제N 시간의 경우에도 동일한 방식으로 합성 채널 프레임(HCF)이 구성될 수 있다.

[0038] 또한, 위에서는 합성 채널 프레임 제공부(100)가 연속되는 시간들에 대해서 합성 채널 프레임(HCF)을 제공하는 것으로 개시되어 있으나, 합성 채널 프레임 제공부(100)는 카메라가 받아들이는 프레임의 일부를 건너 뛸 수 있다. 예를 들어, 제1 시간(T1)이 카메라가 받아들이는 프레임의 첫 번째 프레임이라면 제2 시간(T2)는 다섯 번째 프레임 후인 여섯 번째 프레임이 될 수 있다.

[0039] 이 경우, 그레이 합성 프레임 생성부(210)는 제1 시간(T1)의 제1 합성 채널 프레임(HCF1_C1), 제1 시간(T1)의 제2 합성 채널 프레임(HCF1_C2) 및 제1 시간(T1)의 제3 합성 채널 프레임(HCF1_C3)을 합성하여 제1 시간(T1)의 그레이 합성 프레임(GHF1)을 생성할 수 있다. 제2 시간(T2) 내지 제N 시간의 그레이 합성 프레임(GHF)도 동일한 방식으로 생성될 수 있다. 여기서, 합성은 채널 별 합성 프레임들을 합산하는 방식으로 이루어질 수도 있고, 채널별로 가중치를 부가하여 합산하는 방식일 수도 있다.

[0040] 일 실시예에 있어서, 그레이 합성 프레임 생성부(210)는 레드 채널에 상응하는 레드 합성 채널 프레임(HCF), 그린 채널에 상응하는 그린 합성 채널 프레임(HCF) 및 블루 채널에 상응하는 블루 합성 채널 프레임(HCF)을 합성하여 그레이 합성 프레임(GHF)을 생성할 수 있다. 예를 들어, 제1 합성 채널 프레임(HCF_C1)은 레드 합성 채널 프레임일 수 있고, 제2 합성 채널 프레임(HCF_C2)은 그린 합성 채널 프레임일 수 있다. 또한, 제3 합성 채널 프레임(HCF_C3)은 블루 합성 채널 프레임일 수 있다. 그레이 합성 프레임 생성부(210)는 제1 합성 채널 프레임(HCF_C1), 제2 합성 채널 프레임(HCF_C2) 및 제3 합성 채널 프레임(HCF_C3)을 합성하여 그레이 합성 프레임(GHF)을 생성할 수 있다.

[0041] 일 실시예에 있어서, 그레이 합성 프레임 생성부(210)는 복수의 시간들의 각각에 상응하는 레드 합성 채널 프레

임, 그런 합성 채널 프레임 및 블루 합성 채널 프레임을 합성하여 그레이 합성 프레임(GHF)을 생성할 수 있다. 예를 들어, 복수의 시간들은 제1 시간(T1), 제2 시간(T2), 제3 시간(T3) 및 제4 시간 내지 제N 시간을 포함할 수 있다. 그레이 합성 프레임(GHF)은 제1 시간(T1)의 제1 그레이 합성 프레임(GHF1), 제2 시간(T2)의 제2 그레이 합성 프레임(GHF2), 제3 시간(T3)의 제3 그레이 합성 프레임(GHF3) 및 제4 시간의 제4 그레이 합성 프레임 내지 제N 시간의 제N 그레이 합성 프레임을 포함할 수 있다.

[0042] 복수의 시간들 중 연속되는 3개의 시간에 상응하는 그레이 합성 프레임(GHF)을 하나의 입력단위로 구성하여 멀티프레임 컨볼루션 뉴럴네트워크(220)에 제공할 수 있다. 예를 들어, 복수의 시간들 중 연속되는 3개의 시간은 제1 시간(T1), 제2 시간(T2) 및 제3 시간(T3)일 수 있다. 멀티프레임 컨볼루션 뉴럴 네트워크(220)의 제1 입력(INPUT1)에는 제1 그레이 합성 프레임(GHF1), 제2 그레이 합성 프레임(GHF2), 제3 그레이 합성 프레임(GHF3)이 입력될 수 있고, 멀티프레임 컨볼루션 뉴럴 네트워크의 제2 입력(INPUT2)에는 제4 그레이 합성 프레임(GHF4), 제5 그레이 합성 프레임(GHF5), 제6 그레이 합성 프레임(GHF6)이 입력될 수 있다.

[0043] 멀티프레임 컨볼루션 뉴럴네트워크(220)는 그레이 합성 프레임(GHF)에 기초하여 중간 프레임(MF)을 생성할 수 있다. 예를 들어, 제1 그레이 합성 프레임(GHF1) 내지 제6 그레이 합성 프레임에 기초하여 멀티프레임 컨볼루션 뉴럴네트워크(220)의 제1 출력(OUTPUT1)으로 출력되는 중간 프레임(MF)은 제1 중간 프레임(MF1), 제2 중간 프레임(MF2) 및 제3 중간 프레임(MF3)일 수 있고, 멀티프레임 컨볼루션 뉴럴네트워크(220)의 제2 출력(OUTPUT2)으로 출력되는 중간 프레임(MF)은 제4 중간 프레임(MF4), 제5 중간 프레임(MF5) 및 제6 중간 프레임(MF6)일 수 있다.

[0044] 도 8은 도 2의 동영상 기반 행동인식 장치에 포함되는 행동인식 결과 제공부를 나타내는 도면이고, 도 9는 도 8의 행동인식 결과 제공부를 설명하기 위한 도면이다.

[0045] 도 1 내지 9를 참조하면, 본 발명의 실시예에 따른 동영상 기반 행동인식 장치(10)는 합성 채널 프레임 제공부(100), 뉴럴 네트워크부(200) 및 행동인식 결과 제공부(300)를 포함할 수 있다. 행동인식 결과 제공부(300)는 풀링부, 가중치 생성부(320) 및 행동인식 결과부(330)를 포함할 수 있다. 일 실시예에 있어서, 풀링부(310)는 중간 프레임(MF)을 에버리지 풀링하여 풀링 데이터(PD)를 생성하고, 풀링 데이터(PD)는 풀리 커넥티드 레이어를 포함하는 가중치 생성부(320)에 제공될 수 있다. 가중치 생성부(320)는 풀링 데이터(PD)에 기초하여 가중치 값(WT)을 생성할 수 있다. 행동인식 결과 제공부(300)는 가중치 값(WT)에 기초하여 행동인식 결과(ARR)를 제공할 수 있다. 행동인식 결과 제공부(300)로부터 출력되는 행동인식 결과(ARR)는 동영상 내의 이벤트 발생을 보다 효과적으로 검출하기 위하여 사용될 수 있다.

[0046] 일 실시예에 있어서, 행동인식 결과 제공부(300)는 제1 풀링부(311) 및 제2 풀링부(313)를 포함할 수 있다. 제1 풀링부(311)는 중간 프레임(MF)을 각각 에버리지 풀링하여 중간 풀링 데이터(MPD)를 생성할 수 있다. 제2 풀링부(313)는 중간 풀링 데이터(MPD)를 에버리지 풀링하여 풀링데이터(PD)를 제공할 수 있다. 또한, 일 실시예에 있어서, 행동인식 결과 제공부(300)는 중간 풀링 데이터(MPD) 및 가중치 값(WT)에 기초하여 행동인식 결과(ARR)를 제공할 수 있다.

[0047] 본 발명의 실시예에 따른 동영상 기반 행동인식 장치(10)는 합성 채널 프레임 제공부(100), 뉴럴 네트워크부(200) 및 행동인식 결과 제공부(300)를 포함할 수 있다. 합성 채널 프레임 제공부(100)는 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보(HIN)를 생성하고, 채널 프레임들과 하이라이트 정보(HIN)를 합성하여 합성 채널 프레임(HCF)을 제공할 수 있다. 뉴럴 네트워크부(200)는 합성 채널 프레임(HCF) 및 멀티프레임 컨볼루션 뉴럴네트워크(220)에 기초하여 중간 프레임(MF)을 제공할 수 있다. 행동인식 결과 제공부(300)는 중간 프레임(MF)에 따라 생성되는 가중치 값(WT) 및 중간 프레임(MF)에 기초하여 행동인식 결과(ARR)를 제공할 수 있다.

[0048] 본 발명에 따른 동영상 기반 행동인식 장치(10)에서는, 채널별로 제공되는 채널 프레임들로부터 합성되는 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 행동인식 결과(ARR)를 제공함으로써 동영상 내의 이벤트 발생을 보다 효과적으로 검출할 수 있다.

[0049] 또한, 본 발명에 따른 동영상 기반 행동인식 장치의 동작방법에서는, 3채널 이미지를 1채널 이미지로 변환하여 멀티프레임 컨볼루션 뉴럴네트워크(2D CNN)에 입력하기 때문에 기존의 2D CNN보다 실시간 인식을 위한 연산량을 약 1/3 배 줄일 수 있어 행동인식의 성능을 높일 수 있다.

[0050] 도 10은 본 발명의 실시예들에 따른 동영상 기반 행동인식 시스템의 동작방법을 나타내는 순서도이고, 도 11은 본 발명의 실시예들에 따른 동영상 기반 행동인식 시스템을 나타내는 도면이다.

[0051] 도 1 내지 11을 참조하면, 본 발명의 실시예에 따른 동영상 기반 행동인식 시스템의 동작방법에서는, 합성 채널

프레임 제공부(100)가 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보(HIN)를 생성하고, 채널 프레임(CHF)들과 하이라이트 정보(HIN)를 합성하여 합성 채널 프레임을 제공할 수 있다(S100). 뉴럴 네트워크부(200)가 합성 채널 프레임(CHF) 및 멀티프레임 컨볼루션 뉴럴네트워크(220)에 기초하여 중간 프레임(MF)을 제공할 수 있다(S200). 행동인식 결과 제공부(300)가 중간 프레임(MF)에 따라 생성되는 가중치 값(WT) 및 중간 프레임(MF)에 기초하여 행동인식 결과(ARR)를 제공할 수 있다(S300). 디스플레이부(400)가 행동인식 결과(ARR)에 따라 생성되는 알림 정보를 디스플레이할 수 있다(S400).

[0052] 일 실시예에 있어서, 뉴럴 네트워크부(200)는 그레이 합성 프레임 생성부(210) 및 멀티프레임 컨볼루션 뉴럴네트워크(220)를 포함할 수 있다. 그레이 합성 프레임 생성부(210)는 복수의 채널들의 각각에 상응하는 합성 채널 프레임을 합성하여 그레이 합성 프레임(GHF)을 생성할 수 있다. 멀티프레임 컨볼루션 뉴럴네트워크(220)는 그레이 합성 프레임(GHF)에 기초하여 중간 프레임(MF)을 생성할 수 있다.

[0053] 본 발명의 실시예에 따른 동영상 기반 행동인식 시스템은 합성 채널 프레임 제공부(100), 뉴럴 네트워크부(200), 행동인식 결과 제공부(300) 및 디스플레이부(400)를 포함할 수 있다. 합성 채널 프레임 제공부(100)는 복수의 채널들 중 각 채널에 상응하는 채널 프레임들을 비교하여 하이라이트 정보(HIN)를 생성하고, 채널 프레임들과 하이라이트 정보(HIN)를 합성하여 합성 채널 프레임을 제공할 수 있다. 뉴럴 네트워크부(200)는 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 중간 프레임(MF)을 제공할 수 있다. 행동인식 결과 제공부(300)는 중간 프레임(MF)에 따라 생성되는 가중치 값(WT) 및 중간 프레임(MF)에 기초하여 행동인식 결과(ARR)를 제공할 수 있다. 디스플레이부(400)는 행동인식 결과(ARR)에 따라 생성되는 알림 정보를 디스플레이할 수 있다. 일 실시예에 있어서, 복수의 채널들은 제1 채널(C1), 제2 채널(C2) 및 제3 채널(C3)을 포함할 수 있다. 제1 채널(C1)은 레드(Red) 채널이고, 제2 채널(C2)은 그린(Green)이고, 제3 채널(C3)은 블루(Blue) 채널일 수 있다.

[0054] 본 발명에 따른 동영상 기반 행동인식 시스템에서는, 채널별로 제공되는 채널 프레임들로부터 합성되는 합성 채널 프레임 및 멀티프레임 컨볼루션 뉴럴네트워크에 기초하여 행동인식 결과(ARR)를 제공함으로써 동영상 내의 이벤트 발생을 보다 효과적으로 검출할 수 있다.

[0055] 또한, 본 발명에 따른 동영상 기반 행동인식 시스템의 동작방법에서는, 3채널 이미지를 1채널 이미지로 변환하여 멀티프레임 컨볼루션 뉴럴네트워크(2D CNN)에 입력하기 때문에 기존의 2D CNN보다 실시간 인식을 위한 연산량을 약 1/3 배 줄일 수 있어 행동인식의 성능을 높일 수 있다.

[0056] 도 1 내지 11을 참조하면, 본 발명에 따른 동영상 기반 행동인식 장치(10)의 동작방법에서는, 합성 채널 프레임(HCF)은 각 채널에 상응하는 채널 프레임(CHF)들일 수도 있고, 채널 프레임(CHF)들을 하이라이트 정보(HIN)와 합성하여 생성된 프레임일 수도 있다. 뉴럴 네트워크부(200)가 합성 채널 프레임(HCF)에 기초하여 중간 프레임(MF)을 제공할 수 있다. 행동인식 결과 제공부(300)가 중간 프레임(MF)에 기초하여 행동인식 결과(ARR)를 제공할 수 있다. 합성 채널 프레임 제공부는 프레임 제공부일 수도 있고, 입력 프레임 제공부일 수도 있다.

[0057] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

부호의 설명

[0058] 10: 동영상 기반 행동인식 장치 20: 동영상 기반 행동인식 시스템

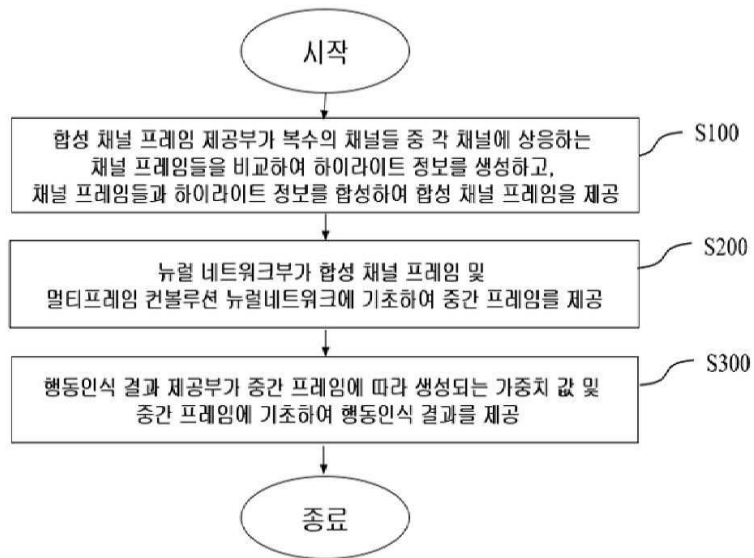
100: 합성 채널 프레임 제공부 200: 뉴럴 네트워크부

300: 행동인식 결과 제공부 210: 그레이 합성 프레임 생성부

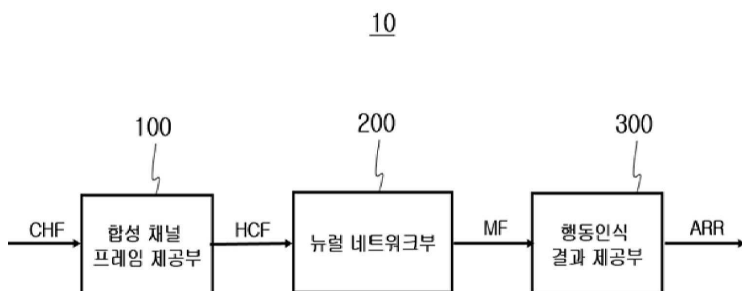
220: 멀티프레임 컨볼루션 뉴럴네트워크

도면

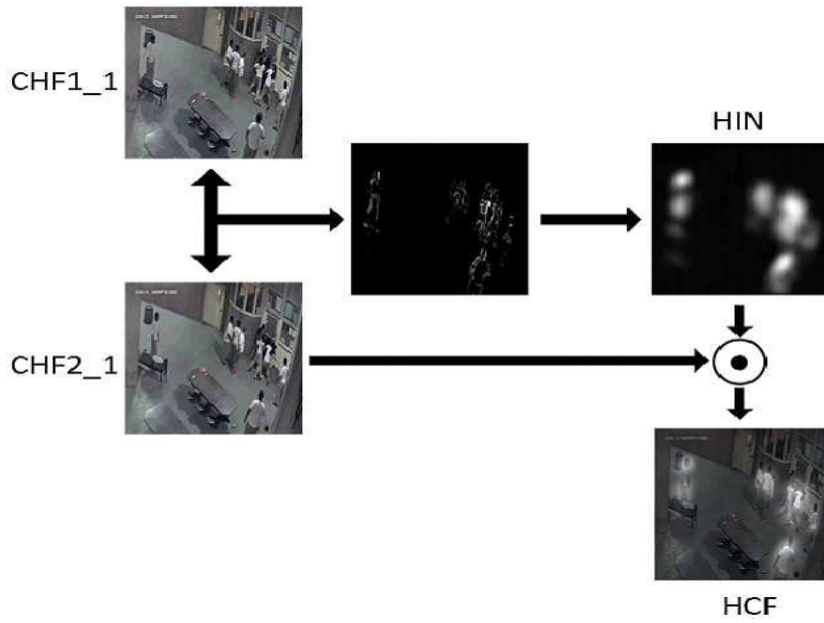
도면1



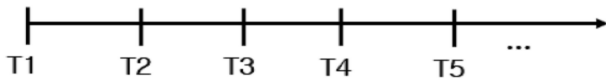
도면2



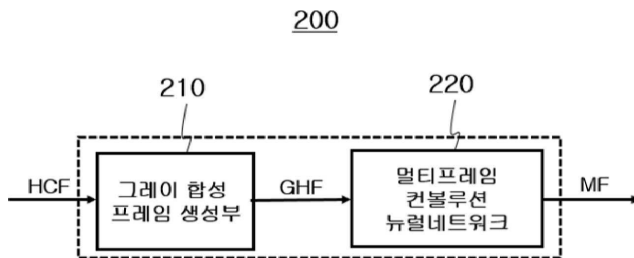
도면3



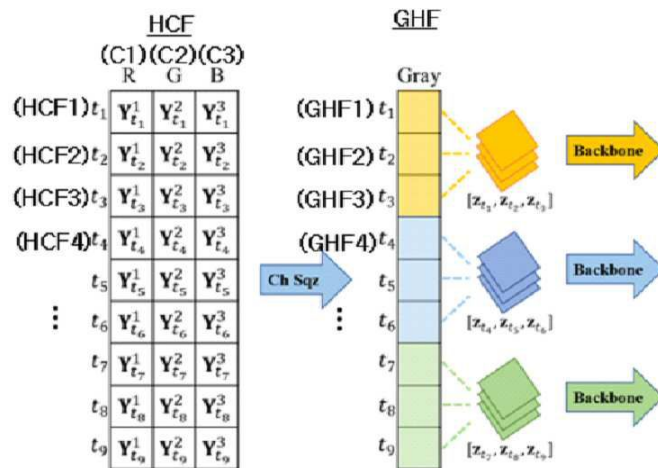
도면4



도면5

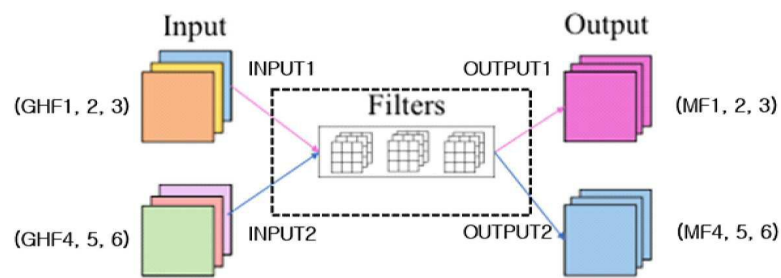


도면6

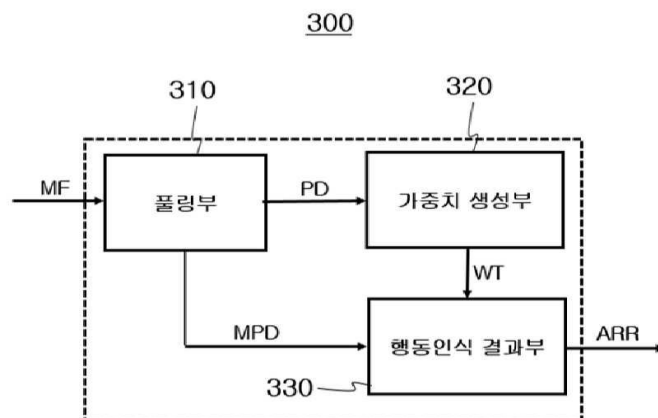


도면7

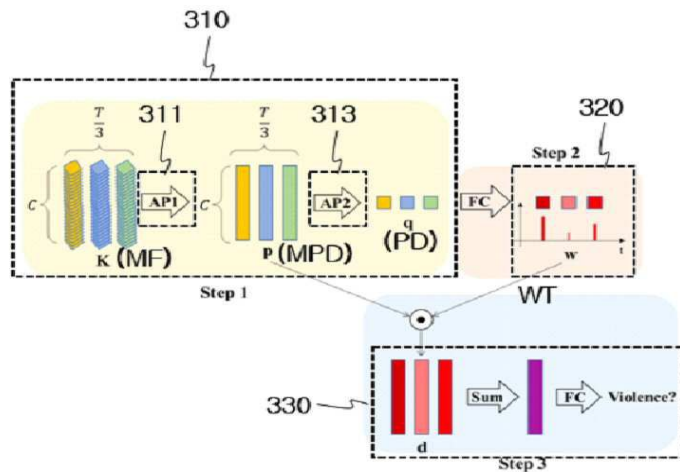
멀티프레임 컨볼루션 뉴럴네트워크 (220)



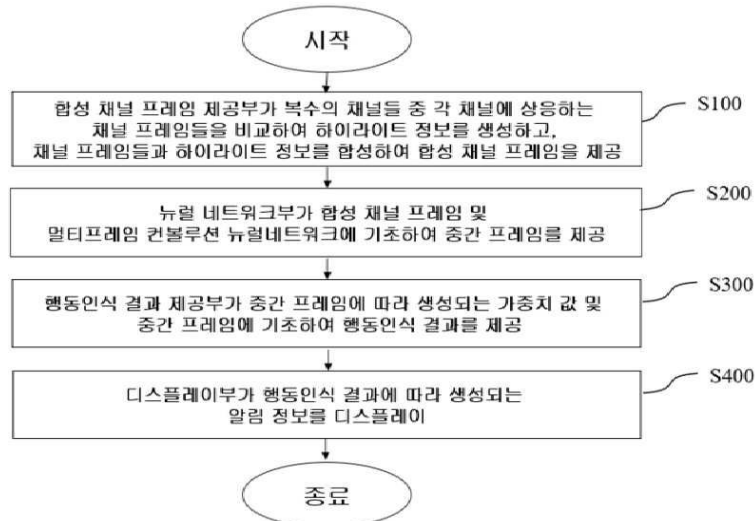
도면8



도면9



도면10



도면11

