



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년11월25일
(11) 등록번호 10-2182660
(24) 등록일자 2020년11월18일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2006.01) G06K 9/62 (2006.01)
G08B 13/196 (2006.01) H04N 7/18 (2006.01)
(52) CPC특허분류
G06K 9/00664 (2013.01)
G06K 9/00335 (2013.01)
(21) 출원번호 10-2019-0155473
(22) 출원일자 2019년11월28일
심사청구일자 2019년11월28일
(56) 선행기술조사문헌
KR101942808 B1
(뒷면에 계속)

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
백성욱
서울특별시 광진구 아차산로 262, B동 1304호 (자양동, 더샵스타시티)
노승민
서울특별시 은평구 진관2로 111-7, 214동 805호(진관동, 은평뉴타운우물골)
(뒷면에 계속)
(74) 대리인
특허법인현

전체 청구항 수 : 총 15 항

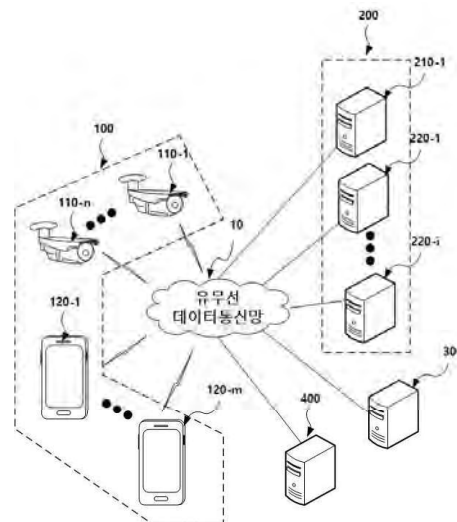
심사관 : 황승희

(54) 발명의 명칭 시공간 특징을 이용한 폭력 감지 시스템, 방법 및 상기 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록매체

(57) 요약

본 발명은 시공간 특징을 이용한 폭력 감지 시스템, 방법 및 상기 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록매체에 관한 것으로, 더욱 상세하게는 획득되는 영상으로부터 불필요한 프레임을 줄이기 위해 CNN모델을 적용하여 사람을 검출하고, 사람이 검출된 연속된 프레임열인 프레임 시퀀스로부터 3D-CNN을 적용하여 시공간 특징을 추출하여 영상 내의 폭력적인 활동을 예측 및 판단하고, 폭력적인 활동의 예측 및 판단 시 관련 보안 부서로 경고를 제공하는 시공간 특징을 이용한 폭력 감지 시스템, 방법 및 상기 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록매체에 관한 것이다.

대표도 - 도1



- (52) CPC특허분류
G06K 9/6267 (2013.01)
G08B 13/196 (2013.01)
H04N 7/181 (2013.01)
- (72) 발명자
이미영
 서울특별시 강남구 도산대로101길 34, 603호 (청담동, 구산아파트)
- 파튜 유 민 올라**
 서울특별시 광진구 능동로 209, 율곡관 601B(군자동)
- 올라 아민**
 서울특별시 광진구 능동로 209, 율곡관 601B(군자동)
- 이자즈 울 하크**
 서울특별시 광진구 능동로 209, 율곡관 601B(군자동)
- (56) 선행기술조사문헌
 KR1020140037354 A
 KR1020150088613 A
 KR1020160057503 A
 파튜 유 민 올라 외 4인 공저, '스마트 감시 애플리케이션을 위해 Deep CNN을 이용한 폭력인식,' 한 국차세대컴퓨팅학회 논문지 14:5, 53-59 (2018.10.)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711094394
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능 융합선도 프로젝트 사업
연구과제명	스마트시티 산업 생산성 혁신을 위한 AI융합 기술 개발
기 여 율	1/1
과제수행기관명	세종대학교 산학협력단
연구기간	2019.04.01 ~ 2021.12.31
공지예외적용	: 있음

명세서

청구범위

청구항 1

감시영역을 촬영한 영상을 획득하여 출력하는 영상획득부;

상기 영상획득부로부터 출력되는 영상으로부터 사람을 검출하고, 사람이 검출된 프레임을 캡처하여 출력하는 사람 검출부;

상기 사람 검출부로부터 출력되는 프레임들을 일정 수의 프레임으로 그룹화한 프레임 시퀀스를 생성하여 출력하는 프레임 시퀀스 생성부;

상기 프레임 시퀀스 생성부로부터 출력되는 프레임 시퀀스로부터 시공간특성을 검출하고, 그에 따른 시공간특성 정보를 분류하여 상기 프레임 시퀀스에 폭력이 존재하는지의 폭력 존재 여부 정보를 출력하는 시공간특성 추출부; 및

상기 폭력 존재 여부 정보에 의해 상기 프레임 시퀀스가 폭력이 존재하는 프레임 시퀀스인지 폭력이 없는 프레임 시퀀스인지를 판단하는 폭력 감지 판단부를 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 2

제1항에 있어서,

상기 영상획득부는,

상기 감시영역을 감시하는 폐쇄회로텔레비전(CCTV)을 통해 촬영된 영상을 획득하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 3

제1항에 있어서,

상기 영상획득부는,

상기 감시영역을 촬영한 임의의 모바일 단말기 및 상기 감시영역을 촬영한 영상에 대한 임의의 서비스를 제공하는 서비스 어플리케이션 서버 중 어느 하나로부터 상기 영상을 획득하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 4

제1항에 있어서,

상기 사람 검출부는,

획득되는 영상의 프레임의 깊이 및 점별 컨벌루션이 개별적으로 계산되는 경우 28개의 레이어를 포함하고, 마지막 완전 연결 레이어(Fully Connected Layer: FC Layer)를 제외한 레이어 뒤에 비선형 배치 표준 및 ReLU가 적용된 모바일 넷(Mobilenet)-SSD CNN 모델을 적용하여 각 프레임별로 사람을 검출하여 출력하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 5

제4항에 있어서,

상기 프레임 시퀀스 생성부는,

사람이 검출된 16개의 프레임을 포함하는 프레임 시퀀스를 생성하여 출력하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 6

제1항 또는 제5항에 있어서,

상기 시공간특성 추출부는,

상기 프레임 시퀀스 생성부로부터 입력되는 프레임 시퀀스의 사람이 검출된 프레임들 간 공간 상관 정보 및 시간 상관 정보를 포함하는 시공간특성 정보를 생성하고, 생성된 시공간특성정보에 근거하여 프레임 시퀀스의 영상에 폭력이 존재하는지 존재하지 않는지를 분류하여 출력하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 7

제6항에 있어서,

상기 시공간특성 추출부는,

8개의 회선(Conv), 5개의 풀링(Pooling) 및 폭력이 존재하는 영상인지 폭력이 없는 영상인지를 분류하는 소프트웨어 출력 레이어가 있는 2개의 완전 연결 레이어(FC Layer)를 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 8

제1항에 있어서,

상기 영상획득부는,

상기 감시영역에 대한 위치정보를 더 획득하여 출력하고,

상기 폭력 감지 판단부를 통해 입력된 상기 프레임 시퀀스가 폭력이 있는 프레임인 것으로 판단되면, 상기 프레임 시퀀스의 영상이 획득된 위치정보를 포함하는 폭력 발생 통지 정보를 유관기관으로 통지하는 경보부를 더 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 9

제8항에 있어서,

상기 위치정보는,

상기 감시영역을 촬영하는 폐쇄회로텔레비전이 설치된 위치정보를 포함하는 폐쇄회로텔레비전 관리서버로부터 획득되는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

청구항 10

영상획득부가 감시영역을 촬영한 영상을 획득하여 출력하는 영상획득 과정;

사람 검출부가 상기 영상획득부로부터 출력되는 영상으로부터 사람을 검출하고, 사람이 검출된 프레임을 캡처하여 출력하는 사람 검출 프레임 추출 과정;

프레임 시퀀스 생성부가 상기 사람 검출부로부터 출력되는 프레임을 일정 수의 프레임으로 그룹화한 프레임 시퀀스를 생성하여 출력하는 프레임 시퀀스 생성과정;

시공간특성 추출부가 상기 프레임 시퀀스 생성부로부터 출력되는 프레임 시퀀스로부터 시공간특성을 검출하고, 그에 따른 시공간특성정보를 분류하여 상기 프레임 시퀀스에 폭력이 존재하는지의 폭력 존재 여부 정보를 출력하는 시공간특성 추출 과정; 및

폭력 감지 판단부가 상기 폭력 존재 여부 정보에 의해 상기 프레임 시퀀스가 폭력이 존재하는 프레임 시퀀스인지 폭력이 없는 프레임 시퀀스인지를 판단하는 폭력 감지 판단부를 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 방법.

청구항 11

제10항에 있어서,

상기 시공간특성 추출 과정은,

상기 프레임 시퀀스 생성부로부터 입력되는 프레임 시퀀스의 사람이 검출된 프레임들 간 공간 상관 정보를 획득하는 공간 상관 정보 획득 단계;

상기 프레임 시퀀스의 프레임들 간 시간 상관 정보를 획득하는 시간 상관 정보 획득 단계;

상기 공간 상관 정보 및 시간 상관 정보를 포함하는 시공간특성 정보를 생성하는 시공간특성 정보 생성 단계; 및

생성된 상기 시공간특성정보에 근거하여 프레임 시퀀스의 영상에 폭력이 존재하는지 존재하지 않는지를 분류하여 출력하는 분류 단계를 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 방법.

청구항 12

제10항에 있어서,

상기 영상획득 과정은,

상기 감시영역에 대한 영상을 획득하는 영상 획득 단계; 및

상기 감시영역의 위치정보를 획득하는 위치정보 획득 단계를 포함하되,

경보부가 상기 폭력 감지 판단부를 통해 입력된 상기 프레임 시퀀스가 폭력이 있는 프레임인 것으로 판단되면, 상기 영상획득 과정의 위치정보 획득 단계에서 획득된 프레임 시퀀스의 영상이 획득된 위치정보를 포함하는 폭력 발생 통지 정보를 유관기관으로 통지하는 경보 과정을 더 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 방법

청구항 13

영상획득부가 감시영역을 촬영한 영상을 획득하여 출력하는 영상획득 과정;

사람 검출부가 상기 영상획득부로부터 출력되는 영상으로부터 사람을 검출하고, 사람이 검출된 프레임을 캡처하여 출력하는 사람 검출 프레임 추출 과정;

프레임 시퀀스 생성부가 상기 사람 검출부로부터 출력되는 프레임을 일정 수의 프레임으로 그룹화한 프레임 시퀀스를 생성하여 출력하는 프레임 시퀀스 생성과정;

시공간특성 추출부가 상기 프레임 시퀀스 생성부로부터 출력되는 프레임 시퀀스로부터 시공간특성을 검출하고, 그에 따른 시공간특성정보를 분류하여 상기 프레임 시퀀스에 폭력이 존재하는지의 폭력 존재 여부 정보를 출력하는 시공간특성 추출 과정; 및

폭력 감지 판단부가 상기 폭력 존재 여부 정보에 의해 상기 프레임 시퀀스가 폭력이 존재하는 프레임 시퀀스인

지 폭력이 없는 프레임 시퀀스인지를 판단하는 폭력 감지 판단부를 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록 매체.

청구항 14

제13항에 있어서,

상기 시공간특성 추출 과정은,

상기 프레임 시퀀스 생성부로부터 입력되는 프레임 시퀀스의 사람이 검출된 프레임들 간 공간 상관 정보를 획득하는 공간 상관 정보 획득 단계;

상기 프레임 시퀀스의 프레임들 간 시간 상관 정보를 획득하는 시간 상관 정보 획득 단계;

상기 공간 상관 정보 및 시간 상관 정보를 포함하는 시공간특성 정보를 생성하는 시공간특성 정보 생성 단계; 및

생성된 상기 시공간특성정보에 근거하여 프레임 시퀀스의 영상에 폭력이 존재하는지 존재하지 않는지를 분류하여 출력하는 분류 단계를 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록 매체.

청구항 15

제13항에 있어서,

상기 영상획득 과정은,

상기 감시영역에 대한 영상을 획득하는 영상 획득 단계; 및

상기 감시영역의 위치정보를 획득하는 위치정보 획득 단계를 포함하되,

경보부가 상기 폭력 감지 판단부를 통해 입력된 상기 프레임 시퀀스가 폭력이 있는 프레임인 것으로 판단되면, 상기 영상획득 과정의 위치정보 획득 단계에서 획득된 프레임 시퀀스의 영상이 획득된 위치정보를 포함하는 폭력 발생 통지 정보를 유관기관으로 통지하는 경보 과정을 더 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록 매체.

발명의 설명

기술 분야

[0001] 본 발명은 시공간 특징을 이용한 폭력 감지 시스템, 방법 및 상기 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록매체에 관한 것으로, 더욱 상세하게는 획득되는 영상으로부터 불필요한 프레임을 줄이기 위해 CNN 모델을 적용하여 사람을 검출하고, 사람이 검출된 연속된 프레임열인 프레임 시퀀스로부터 3D-CNN을 적용하여 시공간 특징을 추출하여 영상 내의 폭력적인 활동을 예측 및 판단하고, 폭력적인 활동의 예측 및 판단 시 관련 보안 부서로 경고를 제공하는 시공간 특징을 이용한 폭력 감지 시스템, 방법 및 상기 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록매체에 관한 것이다.

배경 기술

[0002] 일반적으로 폭력이란 신체적인 손상을 가져오고, 정신적 및 심리적인 압박을 가하는 물리적인 강제력을 말한다. 폭력은 법에서 정하고 있는 바와 같이 다른 사람에게 상해를 입히거나, 협박하거나 하는 등의 행위, 사람을 감금하는 행위, 주거에 침입하는 행위, 기물을 파손하는 행위 등과 같이 다양한 형태로 발생한다.

[0003] 사회가 집단이기주의화, 개인주의화, 핵가족화되고 고용불안 등의 심리적 불안감 및 학업 중압감이 증가함에 따라 폭력도 증가하고 있으며 이러한 폭력은 학교, 거리, 공원, 의료센터 등과 같은 다양한 장소에서 빈번하게 발생하고 있다.

[0004] 이러한 폭력을 방지하고 폭력 발생 시의 후 조치를 위해 학교, 유치원, 거리, 공원, 의료센터, 보호시설 등의

감시영역에는 수많은 폐쇄회로텔레비전(Closed Circuit Television: CCTV)이 설치되고 있다.

- [0005] 다수의 CCTV를 통해 촬영된 영상들은 CCTV를 설치한 주최 측의 서버에 저장되고, 주최 측의 관리자에 의해 실시간 모니터링되고 있으며, 경찰 등의 유관기관의 증거자료서 제공되고 있다.
- [0006] 그러나 관리자가 CCTV를 통해 촬영된 다수의 영상을 실시간 모니터링한다 할지라도, 관리자가 다수의 영상을 한꺼번에 보아야 하므로 정밀하게 모니터링하기 힘들며, 24시간 지속해서 모니터링하는 것은 매우 어려운 문제점이 있었다.
- [0007] 또한 관리자가 영상을 정밀하게 모니터링하여 폭력에 대응하는 영상을 확인했더라도, 이 영상이 실제 폭력인지를 판단하는 데 어려움이 있으며, 폭력을 판단하여 조치하는 데 있어 많은 절차를 걸쳐야 하므로 많은 시간이 소요되는 문제점이 있었다.
- [0008] 이러한 문제점을 해결하기 위해 획득된 영상으로부터 폭력을 감지하는 Datta,et.al, Ngynen.et.al, Mahadevan.et.al, Hassner, Huang,et.al, Guo.et.al 등의 수작업 기반 접근 방식 및 Chen, Lloyed.et.al, Fu.et.al, Sudhakaran.et.al 등의 딥러닝 기반 접근법들이 개발되었다.
- [0009] 위에서 언급한 방식들은 카메라 보기, 복잡한 군중 패턴 및 강도 변화를 포함하여 폭력 감지의 많은 문제를 해결하려고 시도했으나, 폭력탐지를 위해 인체에 관점, 중요한 상호 교합 및 규모로 인해 발생하는 변화가 발생했을 때 추출을 통해 차별적이고 효과적인 특징을 포착하지 못하는 문제점이 있었다.
- [0010] 특히, 종래의 방식들은 저수준의 기능을 사용하므로 복잡한 패턴을 인식하기에는 비효율적이었으며, 실시간 감시에서는 구현하기 어려운 문제점이 있었다.
- [0011] 또한, 종래 방식들은 많은 수의 중요하지 않은 프레임을 처리하는 문제로 인해 더 많은 메모리를 차지하고 시간이 오래 걸리는 문제점이 있었다.
- [0012] 또한, 종래 방식들은 폭력탐지 벤치마크 데이터 세트의 데이터 부족과 낮은 정확도로 인해 효과적인 패턴을 학습할 수 없는 문제점이 있었다.

선행기술문헌

특허문헌

- [0013] (특허문헌 0001) 대한민국 등록특허공보 제10-1541272호(2015.08.03. 공고)

발명의 내용

해결하려는 과제

- [0014] 따라서 본 발명의 목적은 획득되는 영상으로부터 불필요한 프레임을 줄이기 위해 CNN 모델을 적용하여 사람을 검출하고, 사람이 검출된 연속된 프레임열인 프레임 시퀀스로부터 3D-CNN을 적용하여 시공간 특징을 추출하여 영상 내의 폭력적인 활동을 예측 및 판단하고, 폭력적인 활동의 예측 및 판단 시 관련 보안 부서로 경고를 제공하는 시공간 특징을 이용한 폭력 감지 시스템, 방법 및 상기 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록매체를 제공함에 있다.

과제의 해결 수단

- [0015] 상기와 같은 목적을 달성하기 위한 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템은: 감시영역을 촬영한 영상을 획득하여 출력하는 영상획득부; 상기 영상획득부로부터 출력되는 영상으로부터 사람을 검출하고, 사람이 검출된 프레임을 캡처하여 출력하는 사람 검출부; 상기 사람 검출부로부터 출력되는 프레임들을 일정 수의 프레임으로 그룹화한 프레임 시퀀스를 생성하여 출력하는 프레임 시퀀스 생성부; 상기 프레임 시퀀스 생성부로부터 출력되는 프레임 시퀀스로부터 시공간특성을 검출하고, 그에 따른 시공간특성정보를 분류하여 상기 프레임 시퀀스에 폭력이 존재하는지의 폭력 존재 여부 정보를 출력하는 시공간특성 추출부; 및 상기 폭력 존재 여부 정보에 의해 상기 프레임 시퀀스가 폭력이 존재하는 프레임 시퀀스인지 폭력이 없는 프레임 시퀀스인지를 판단하는 폭력 감지 판단부를 포함하는 것을 특징으로 한다.

- [0016] 상기 영상획득부는, 상기 감시영역을 감시하는 폐쇄회로텔레비전(CCTV)을 통해 촬영된 영상을 획득하는 것을 특징으로 한다.
- [0017] 상기 영상획득부는, 상기 감시영역을 촬영한 임의의 모바일 단말기 및 상기 감시영역을 촬영한 영상에 대한 임의의 서비스를 제공하는 서비스 어플리케이션 서버 중 어느 하나로부터 상기 영상을 획득하는 것을 특징으로 한다.
- [0018] 상기 사람 검출부는, 획득되는 영상의 프레임의 깊이 및 점별 컨벌루션이 개별적으로 계산되는 경우 28개의 레이어를 포함하고, 마지막 완전 연결 레이어(Fully Connected Layer: FC Layer)를 제외한 레이어 뒤에 비선형 배치 표준 및 ReLU가 적용된 모바일 넷(Mobilenet)??SSD CNN 모델을 적용하여 각 프레임별로 사람을 검출하여 출력하는 것을 특징으로 한다.
- [0019] 상기 프레임 시퀀스 생성부는, 사람이 검출된 16개의 프레임을 포함하는 프레임 시퀀스를 생성하여 출력하는 것을 특징으로 한다.
- [0020] 상기 시공간특성 추출부는, 상기 프레임 시퀀스 생성부로부터 입력되는 프레임 시퀀스의 사람이 검출된 프레임들 간 공간 상관 정보 및 시간 상관 정보를 포함하는 시공간특성 정보를 생성하고, 생성된 시공간특성정보에 근거하여 프레임 시퀀스의 영상에 폭력이 존재하는지 존재하지 않는지를 분류하여 출력하는 것을 특징으로 한다.
- [0021] 상기 시공간특성 추출부는, 8개의 회전(Conv), 5개의 풀링(Pooling) 및 폭력이 존재하는 영상인지 폭력이 없는 영상인지를 분류하는 소프트맥스 출력 레이어가 있는 2개의 완전 연결 레이어(FC Layer)를 포함하는 것을 특징으로 한다.
- [0022] 상기 영상획득부는, 상기 감시영역에 대한 위치정보를 더 획득하여 출력하고, 상기 폭력 감지 판단부를 통해 입력된 상기 프레임 시퀀스가 폭력이 있는 프레임인 것으로 판단되면, 상기 프레임 시퀀스의 영상이 획득된 위치정보를 포함하는 폭력 발생 통지 정보를 유관기관으로 통지하는 경보부를 더 포함하는 것을 특징으로 한다.
- [0023] 상기 위치정보는, 상기 감시영역을 촬영하는 폐쇄회로텔레비전이 설치된 위치정보를 포함하는 폐쇄회로텔레비전 관리서버로부터 획득되는 것을 특징으로 한다.
- [0024] 상기와 같은 목적을 달성하기 위한 본 발명에 따른 시공간 특징을 이용한 폭력 감지 방법은: 영상획득부가 감시영역을 촬영한 영상을 획득하여 출력하는 영상획득 과정; 사람 검출부가 상기 영상획득부로부터 출력되는 영상으로부터 사람을 검출하고, 사람이 검출된 프레임을 캡처하여 출력하는 사람 검출 프레임 추출 과정; 프레임 시퀀스 생성부가 상기 사람 검출부로부터 출력되는 프레임을 일정 수의 프레임으로 그룹화한 프레임 시퀀스를 생성하여 출력하는 프레임 시퀀스 생성과정; 시공간특성 추출부가 상기 프레임 시퀀스 생성부로부터 출력되는 프레임 시퀀스로부터 시공간특성을 검출하고, 그에 따른 시공간특성정보를 분류하여 상기 프레임 시퀀스에 폭력이 존재하는지의 폭력 존재 여부 정보를 출력하는 시공간특성 추출 과정; 및 폭력 감지 판단부가 상기 폭력 존재 여부 정보에 의해 상기 프레임 시퀀스가 폭력이 존재하는 프레임 시퀀스인지 폭력이 없는 프레임 시퀀스인지를 판단하는 폭력 감지 판단부를 포함하는 것을 특징으로 한다.
- [0025] 상기 시공간특성 추출 과정은, 상기 프레임 시퀀스 생성부로부터 입력되는 프레임 시퀀스의 사람이 검출된 프레임들 간 공간 상관 정보를 획득하는 공간 상관 정보 획득 단계; 상기 프레임 시퀀스의 프레임들 간 시간 상관 정보를 획득하는 시간 상관 정보 획득 단계; 상기 공간 상관 정보 및 시간 상관 정보를 포함하는 시공간특성 정보를 생성하는 시공간특성 정보 생성 단계; 및 생성된 상기 시공간특성정보에 근거하여 프레임 시퀀스의 영상에 폭력이 존재하는지 존재하지 않는지를 분류하여 출력하는 분류 단계를 포함하는 것을 특징으로 한다.
- [0026] 상기 영상획득 과정은, 상기 감시영역에 대한 영상을 획득하는 영상 획득 단계; 및 상기 감시영역의 위치정보를 획득하는 위치정보 획득 단계를 포함하되, 경보부가 상기 폭력 감지 판단부를 통해 입력된 상기 프레임 시퀀스가 폭력이 있는 프레임인 것으로 판단되면, 상기 영상획득 과정의 위치정보 획득 단계에서 획득된 프레임 시퀀스의 영상이 획득된 위치정보를 포함하는 폭력 발생 통지 정보를 유관기관으로 통지하는 경보 과정을 더 포함하는 것을 특징으로 한다.
- [0027] 상기와 같은 목적을 달성하기 위한 본 발명에 따른 시공간 특징을 이용한 폭력 감지 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터 판독 가능한 기록매체는: 영상획득부가 감시영역을 촬영한 영상을 획득하여 출력하는 영상획득 과정; 사람 검출부가 상기 영상획득부로부터 출력되는 영상으로부터 사람을 검출하고, 사람이 검출된 프레임을 캡처하여 출력하는 사람 검출 프레임 추출 과정; 프레임 시퀀스 생성부가 상기 사람 검출부로부터 출력되는 프레임을 일정 수의 프레임으로 그룹화한 프레임 시퀀스를 생성하여 출력하는 프레임 시퀀스 생성과정;

시공간특성 추출부가 상기 프레임 시퀀스 생성부로부터 출력되는 프레임 시퀀스로부터 시공간특성을 검출하고, 그에 따른 시공간특성정보를 분류하여 상기 프레임 시퀀스에 폭력이 존재하는지의 폭력 존재 여부 정보를 출력하는 시공간특성 추출 과정; 및 폭력 감지 판단부가 상기 폭력 존재 여부 정보에 의해 상기 프레임 시퀀스가 폭력이 존재하는 프레임 시퀀스인지 폭력이 없는 프레임 시퀀스인지를 판단하는 폭력 감지 판단부를 포함하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 방법을 기록하는 것을 특징으로 한다.

[0028]

발명의 효과

[0029]

본 발명은 첫 번째 단계에서 효율적인 CNN 모델을 사용하여 원하지 않는 프레임을 제거하여 사람을 감지하여 전체 처리 시간을 줄일 수 있는 효과가 있다.

[0030]

또한, 본 발명은 사람이 있는 프레임 시퀀스가 3개의 벤치마크 데이터 세트에 대해 훈련된 3D CNN 모델에 적용되어 시공간적 특징을 추출하므로 비교적 적은 데이터 세트에도 불구하고 효과적으로 패턴을 학습할 수 있으며, 정확도를 향상시킬 수 있는 효과가 있다.

[0031]

또한, 본 발명은 최적화 모듈을 적용하여 모델을 최적화하므로 최종 플랫폼에서 속도를 높이고 성능을 향상시킬 수 있는 효과가 있다.

도면의 간단한 설명

[0032]

도 1은 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템을 포함하는 통신시스템의 구성을 나타낸 도면이다.

도 2는 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템의 구성을 나타낸 도면이다.

도 3은 본 발명에 따른 시공간 특징을 이용한 폭력 감지 과정을 나타낸 도면이다.

도 4는 본 발명의 일실시예에 따른 시공간 특징을 이용한 폭력 감지 시스템에서 사람이 검출된 영상 프레임을 나타낸 도면이다.

도 5는 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템의 시공간 특징 추출을 위한 3D CNN의 구조를 나타낸 도면이다.

도 6은 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템의 컨볼루션 깊이에 따른 영상 프레임을 나타낸 도면이다.

도 7은 본 발명에 따른 시공간 특징을 이용한 폭력 감지 방법을 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0033]

이하 첨부된 도면을 참조하여 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템의 구성 및 동작을 설명하고, 시스템에서의 폭력 감지 방법을 상세히 설명한다.

[0034]

도 1은 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템을 포함하는 통신시스템의 구성을 나타낸 도면이다.

[0035]

도 1을 참조하면, 본 발명의 시공간 특징을 이용한 폭력 감지 시스템(300)은 유무선데이터통신망(10)에 연결되어 유무선데이터통신망(10)에 연결된 영상제공장치(100), 영상제공부(200) 등과 데이터 통신을 수행하여 영상을 획득할 수 있을 것이다.

[0036]

영상제공장치(100)는 특정 위치의 감시영역을 촬영하는 다수의 폐쇄회로텔레비전(Closed Circuit Television: CCTV)(110) 및 감시영역을 촬영하는 스마트폰, 스마트패드 등과 같은 모바일 단말기(120)일 수 있을 것이다. 상기 감시영역은 CCTV(110) 및 모바일 단말기(120)가 촬영하는 영역을 의미한다.

[0037]

영상제공장치(100)는 감시영역을 촬영한 영상을 유무선데이터통신망(10)을 통해 본 발명의 폭력 감지 시스템(300)으로 전송할 수도 있고, 영상제공부(200)로 전송할 수도 있을 것이다.

[0038]

영상제공부(200)는 CCTV(110)들에 대한 CCTV 식별정보 및 상기 CCTV(110)가 설치된 위치정보를 저장하여 관리하고, 상기 CCTV(110)들로부터 수신되는 영상을 CCTV 식별정보 및 위치정보에 맵핑하여 저장하고, 실시간으로 본

발명의 폭력 감지 시스템(300)으로 제공하는 CCTV 관리 서버(210-1)일 수도 있고, 모바일 단말기(120)들에 단말기 식별정보를 저장하여 관리하고, 설치된 영상 촬영 어플리케이션을 통해 영상을 수신받아 영상을 단말기 식별 정보에 맵핑하여 저장하며, 상기 영상을 본 발명의 폭력 감지 시스템(300)으로 전송하는 적어도 하나 이상의 어플리케이션 서버(220)를 포함할 수 있을 것이다.

- [0039] 상기 모바일 단말기(120)는 촬영된 영상 및 상기 감시영역의 위치정보(모바일 단말기의 위치정보)를 폭력 감지 시스템(300) 및 해당 어플리케이션 서버(220) 중 어느 하나 이상으로 전송할 수도 있을 것이다.
- [0040] 유관기관 서버(400)는 경찰서, 병원 등과 같은 유관기관에 설치되는 서버로 폭력 감지 시스템(300)으로부터 폭력 발생 통지 정보의 수신 시 별도의 경보기(미도시)를 통해 경보를 발생하고 관리자 단말기(미도시)에 폭력 발생 통지 정보를 표시할 수 있을 것이다. 상기 폭력 발생 통지 정보에는 폭력이 발생한 감시영역의 위치정보가 포함되는 것이 바람직할 것이다.
- [0041] 폭력 감지 시스템(300)은 상술한 영상제공장치(100) 및 영상제공부(200) 중 어느 하나로부터 영상 또는 위치정보를 포함하는 영상을 획득하고, 획득된 영상을 분석하여 폭력이 발생하는지를 모니터링하고, 폭력 발생을 감지 시 상기 유관기관 서버(400)로 폭력 발생 통지 정보를 전송하여 상기 발생한 폭력에 빠르게 대처할 수 있도록 한다.
- [0042] 상기 폭력 감지 시스템(300)의 상세 구성 및 동작은 다음의 도 2를 참조하여 상세히 설명한다.
- [0043] 도 2는 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템의 구성을 나타낸 도면이고, 도 3은 본 발명에 따른 시공간 특징을 이용한 폭력 감지 과정을 나타낸 도면이며, 도 4는 본 발명의 일실시예에 따른 시공간 특징을 이용한 폭력 감지 시스템에서 사람이 검출된 영상 프레임을 나타낸 도면이고, 도 5는 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템의 시공간 특징 추출을 위한 3D CNN의 구조를 나타낸 도면이며, 도 6은 본 발명에 따른 시공간 특징을 이용한 폭력 감지 시스템의 컨볼루션 깊이에 따른 영상 프레임을 나타낸 도면이다. 이하 도 2 내지 도 6을 참조하여 설명한다.
- [0044] 본 발명에 따른 폭력 감지 시스템(300)은 영상획득부(310), 사람 검출부(320), 프레임 시퀀스 생성부(330), 시공간특성 추출부(340) 및 폭력 감지 판단부(350)를 포함하고, 실시예에 따라 경보부(360) 및 학습부(380)를 더 포함할 수 있을 것이다.
- [0045] 상기 영상획득부(310)는 도 3의 510의 511에서 보이는 바와 같이 상기 유무선데이터통신망(10)을 통해 영상제공장치(100) 및 영상제공부(200) 중 어느 하나 이상으로부터 영상 또는, 위치정보 및 영상을 획득하여 출력한다. 상기 획득되는 영상은 스트림 형태로 획득되어 출력될 수도 있고, 영상 데이터 형태로 획득되어 출력될 수 있으나, 스트림 형태로 획득되어 실시간 감시될 수 있도록 하는 것이 바람직할 것이다.
- [0046] 사람 검출부(320)는 상기 영상획득부(310)로부터 입력되는 영상에서 프레임 단위로 사람이 존재하는지를 모니터링하고, 사람이 감지되면 해당 프레임을 출력한다.
- [0047] 상기 사람 검출부(320)는 도 3의 512에서 보이는 바와 같이 미리 훈련된 모바일넷(MobileNet)-SSD(single Shot MultiBox Detector) 컨볼루션 신경망(Convolution Neural Network: CNN) 모델을 적용하여 획득된 영상의 프레임 내 물체인 사람을 검출한다.
- [0048] 상기 MobileNet은 규칙적인 컨볼루션 대신 객체를 감지하기 위해 깊이 분리 가능한 컨볼루션을 가지고 있다.
- [0049] 본 발명에 따른 MobileNet-SSD CNN 모델은 획득되는 영상의 프레임의 깊이 및 점별 컨볼루션이 개별적으로 계산되는 경우 28개의 레이어를 포함하고, 마지막 완전 연결 레이어(Fully Connected Layer: FC Layer)를 제외한 레이어 뒤에 비선형 배치 표준 및 ReLU가 적용된다. 상기 ReLU는 기계 학습에서 많이 사용되는 액티베이션 함수 중 하나이다.
- [0050] 본 발명에 따른 상기 MobileNet-SSD CNN 모델의 첫 번째 컨볼루션 층은 3*3*3*32의 필터 형상을 갖는 2개의 스트라이드를 포함하고, 입력 크기는 224*224*3이며, 그다음 깊이 방향 컨볼루션에는 하나의 스트라이드가 있고, 필터 모양은 3*3*32이며, 입력 크기는 112*112*32이다.
- [0051] 상기 MobileNet은 주로 분류에 적용되는 반면, SSD는 멀티 박스 검출기의 정확한 위치를 찾는 데 사용되고, 그 조합은 물체 감지를 수행한다. 이를 위해 SSD는 네트워크 끝에 추가되어 피드 포워드 컨볼루션을 수행하고 고정 크기 그룹의 경계 상자 그룹을 생성하여 기능 맵을 추출하고 적용하여 해당 상자의 객체 인스턴스 존재 및 감지를 보장한다. 컨볼루션 필터 및 경계 상자는 각 클래스에 대한 확률을 가진 예측 클래스로 구성되고, 확률이 가

장 높은 클래스는 객체를 나타낸다.

- [0052] 본 발명에 따른 상기 MobileNet-SSD CNN 모델, 즉 사람 검출부(320)는 도 4와 같이 아이스하키 영상에서 사람들을 검출하고 사람이 검출된 프레임을 캡처하여 출력할 것이다. 이처럼 영상 내에서 불필요한 프레임을 제거하여 사람만 존재하는 프레임만을 출력하므로 효율적으로 메모리를 사용할 수 있으며, 효율적인 분석을 수행할 수 있다.
- [0053] 프레임 시퀀스 생성부(330)는 상기 사람 검출부(320)로부터 출력되는 연속되는 프레임을 일정 개수 이상 수집 후 상기 수집된 개수의 프레임 시퀀스를 생성하여 출력한다. 상기 프레임 시퀀스에는 16개의 프레임으로 구성되는 것이 바람직하나, 이에 한정되는 것은 아니다.
- [0054] 시공간특성 추출부(340)는 상기 프레임 시퀀스 생성부(330)로부터 출력되는 프레임 시퀀스를 입력받고, 상기 프레임 시퀀스로부터 시공간특성을 검출하고, 그에 따른 시공간특성정보를 분류하여 상기 프레임 시퀀스에 폭력이 존재하는지의 폭력 존재 여부 정보를 출력한다.
- [0055] 상기 시공간특성 추출부(340)는 3D CNN 모델을 적용하여 시공간특성을 추출하고, 추출된 시공간특성에 대한 시공간특성정보를 출력한다.
- [0056] 본 발명의 시공간특성 추출부(340)는 도 3의 520 및 도 5에서 보이는 바와 같이 2D CNN을 통해서 공간정보를 추출하고, 3D 컨볼루션(Convolution: Conv) 및 풀링(Pooling) 작업으로 인해 시간정보를 더 잘 추출할 수 있다. 본 발명의 3D 컨볼루션은 프레임 시퀀스의 프레임 조립을 통해 설계된 큐브에서 3D 마스크를 회전시켜 작도하며, 컨볼루션 계층으로부터 획득된 특징 맵은 모션정보를 캡처하여 이전 계층의 다수의 부착된 프레임에 링크된다. 따라서 바이어스 t_{pq} 를 갖는 p번 계층의 q 번째 특징 맵에서의 위치, x, y, z의 값은 수학식 1에 의해 나타낼 수 있을 것이다.

수학식 1

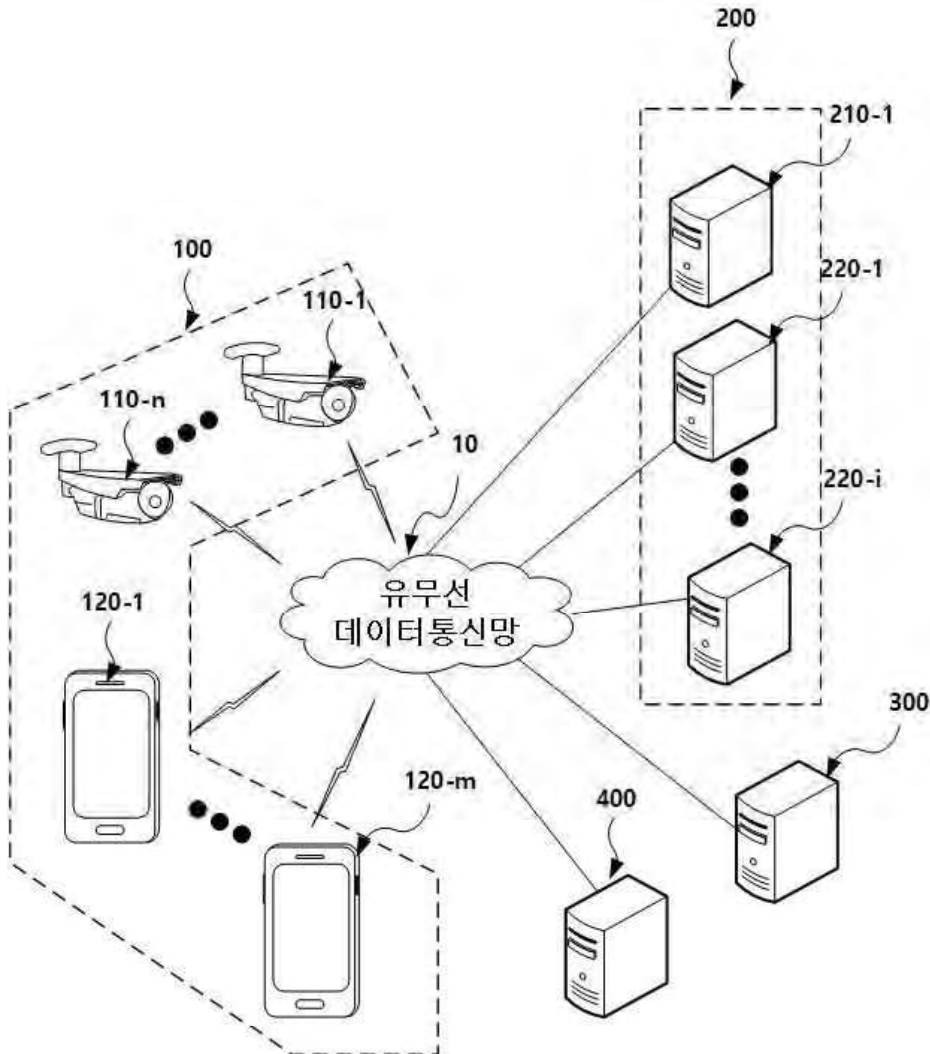
[0057]
$$N_{pq}^{xyz} = \tanh(t_{pq} + \sum_k \sum_{a=0}^{A_p-1} \sum_{b=0}^{B_p-1} \sum_{c=0}^{C_p-1} w_{pqk}^{abc} N_{(p-1)k}^{(x+a)(y+b)(z+c)})$$

- [0058] 여기서, C_p 는 시간 차원이 있는 3D 마스크 크기이고, w_{pqk}^{abc} 는 이전 레이어에서 k 번째 특징 맵에 연결된 3D 마스크의 (a, b, c) 값이다. 커널의 가중치가 전체 큐브에 복제되므로 프레임 큐브에서 3D 컨볼루션 마스크로 한 가지 유형의 특징만 추출된다.
- [0059] 상기 시공간특성 추출부(340)는 도 6에서와 같이 입력 프레임 시퀀스를 입력받고 컨볼루션 프로세스가 3번째 및 5번째 컨볼루션 계층으로 진행할수록 더 깊은 특징을 획득할 수 있을 것이다.
- [0060] 이러한 시공간특성 추출부(340)는 8개의 회선(컨볼루션 레이어: Convolution Layer=Conv), 5개의 풀링(Pooling) 및 소프트맥스(Softmax) 출력 레이어가 있는 2개의 완전 연결된 레이어(FC Layer=fc)를 포함한다.
- [0061] 각 컨볼루션 레이어는 하나의 스트라이드를 갖는 3*3*3 커널이 있으며, 모든 풀링 레이어는 커널 크기가 1*2*2인 두 번째 스트라이드가 있는 첫 번째 풀링 레이어를 제외하고 2*2*2 커널 크기를 갖는 맥스 풀링(Max Pooling)이다. 상기 맥스 풀링은 시간 기반 정보를 보존한다. 각 컨볼루션의 필터 수는 각각 첫 번째, 두 번째 및 세 번째 계층에 대해 64, 128, 256이다. 각 컨볼루션 레이어의 커널은 크기가 D로 정의된 시간 깊이를 갖는다. 컨볼루션을 적용하는 데 사용되는 커널 크기와 패딩은 각각 3과 1로 유지된다. 두 개의 완전 연결 레이어(fc6 및 fc7)는 4096개의 뉴런이 포함되어 있으며, N개의 출력을 포함하는 소프트맥스레이어는 데이터 세트에 따라 N값이 다르다. 본 발명에서는 두 가지 클래스, 즉 폭력적 장면과 비폭력적 장면만 있기 때문에 상기 소프트맥스 레이어의 N은 2일 것이다.
- [0062] 본 발명의 시공간특성 추출부(340)는 16개의 프레임의 짧은 프레임 시퀀스를 획득하지만, 훈련 시 원래 입력 시퀀스에서 3*16*112*112 크기의 임의 자르기를 사용하고, 효과적인 학습을 위해 그 후, 도 3 및 도 5에서와 같이 프레임 시퀀스 다음에 3D 컨볼루션 및 풀링 작업을 수행한다.
- [0063] 폭력 감지 판단부(350)는 시공간특성 추출부(340)로부터 프레임 시퀀스에 대한 소프트맥스의 결과값(0 or 1)이

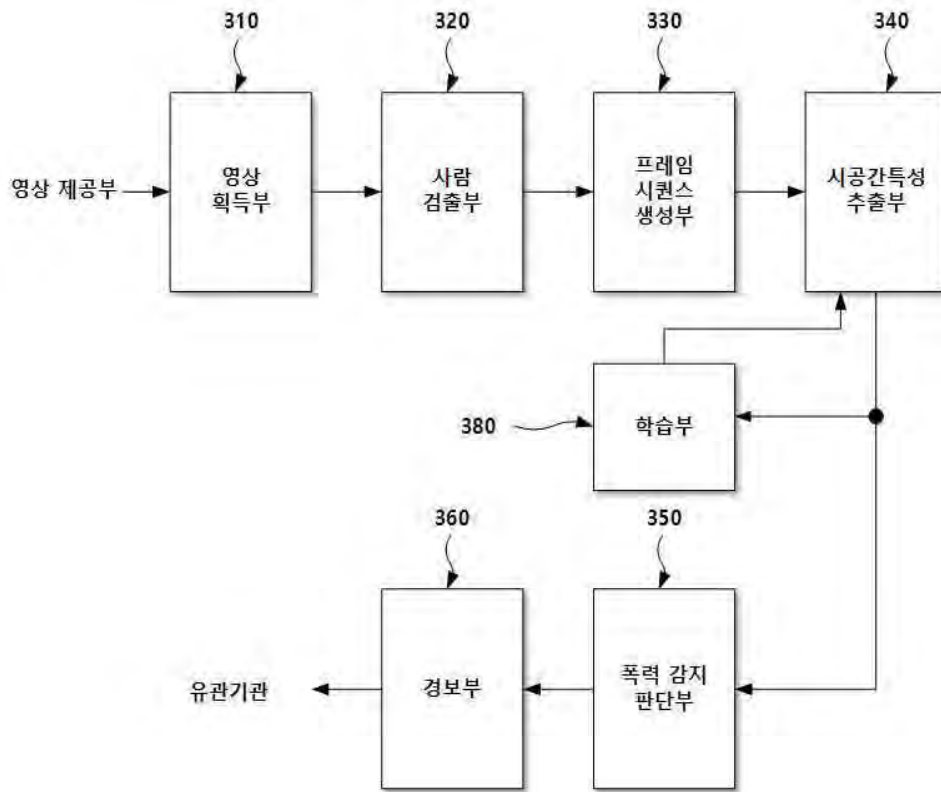
- 200: 영상제공부
- 300: 폭력 감지 시스템
- 310: 영상획득부
- 320: 사람 검출부
- 330: 프레임 시퀀스 생성부
- 340: 시공간특징 추출부
- 350: 폭력 감지 판단부
- 360: 경보부
- 380: 학습부
- 400: 유관기관 서버

도면

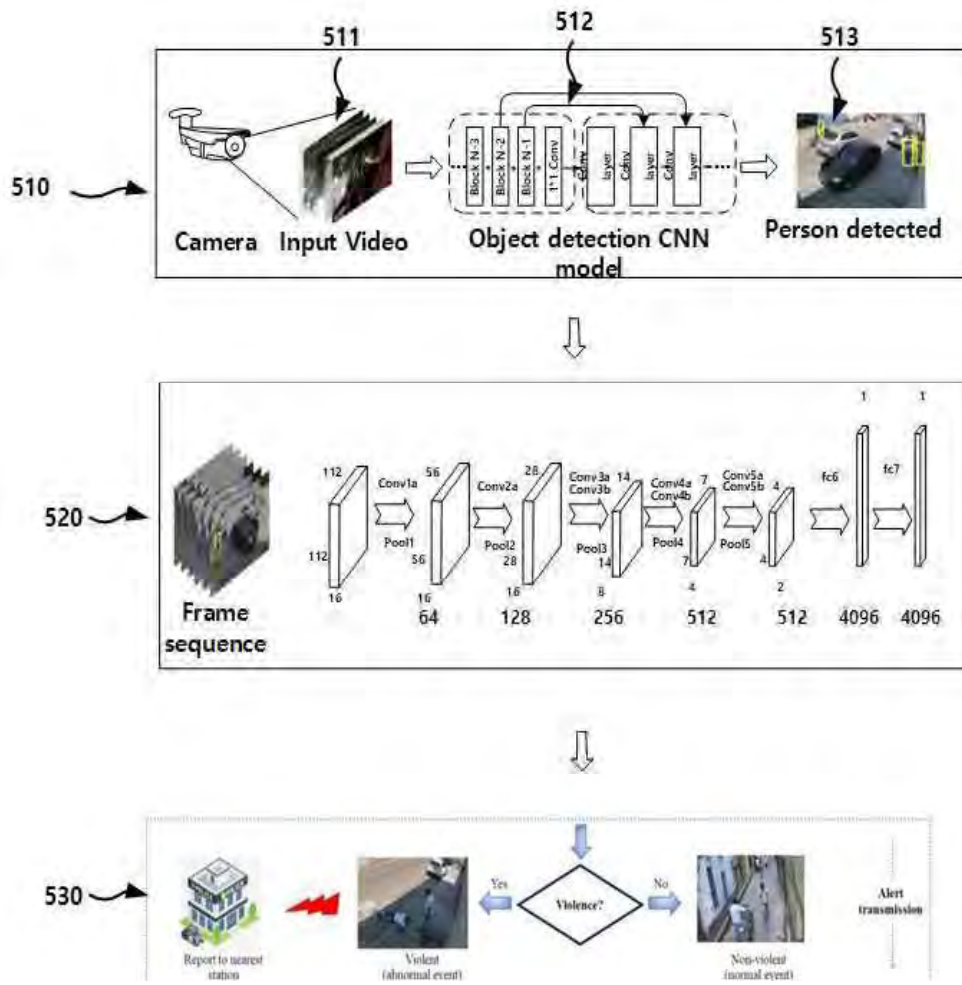
도면1



도면2



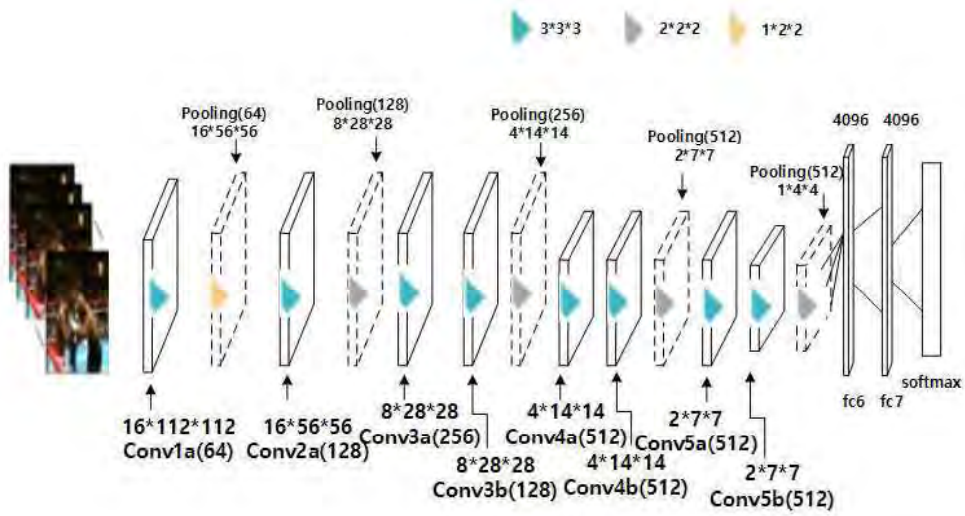
도면3



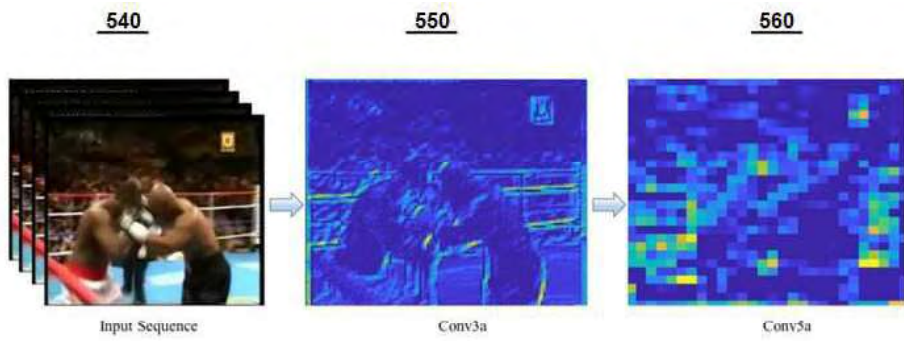
도면4



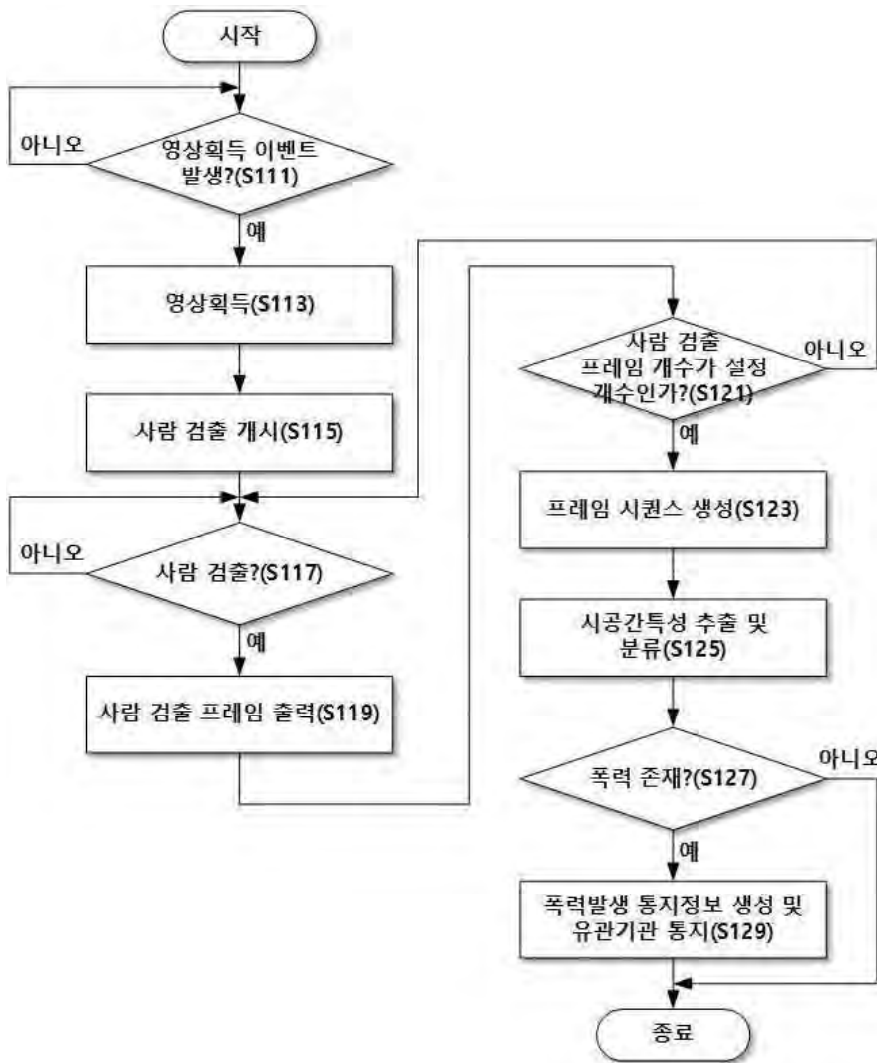
도면5



도면6



도면7



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 4

【변경전】

제1항에 있어서,

상기 사람 검출부는,

획득되는 영상의 프레임의 깊이 및 점별 컨벌루션이 개별적으로 계산되는 경우 28개의 레이어를 포함하고, 마지막 완전 연결 레이어(Fully Connected Layer: FC Layer)를 제외한 레이어 뒤에 비선형 배치 표준 및 ReLU가 적용된 모바일 넷(Mobilenet)??SSD CNN 모델을 적용하여 각 프레임별로 사람을 검출하여 출력하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.

【변경후】

제1항에 있어서,

상기 사람 검출부는,

획득되는 영상의 프레임의 깊이 및 점별 컨벌루션이 개별적으로 계산되는 경우 28개의 레이어를 포함하고, 마지막 완전 연결 레이어(Fully Connected Layer: FC Layer)를 제외한 레이어 뒤에 비선형 배치 표준 및 ReLU가 적

용된 모바일 넷(Mobilenet)-SSD CNN 모델을 적용하여 각 프레임별로 사람을 검출하여 출력하는 것을 특징으로 하는 시공간 특징을 이용한 폭력 감지 시스템.