



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년08월22일
(11) 등록번호 10-2570126
(24) 등록일자 2023년08월19일

(51) 국제특허분류(Int. Cl.)
H04N 21/8541 (2011.01) G06N 3/08 (2023.01)
G06T 7/20 (2017.01) H04N 21/845 (2011.01)
H04N 21/8549 (2011.01)
(52) CPC특허분류
H04N 21/8541 (2013.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2021-0097834
(22) 출원일자 2021년07월26일
심사청구일자 2021년07월26일
(65) 공개번호 10-2023-0016394
(43) 공개일자 2023년02월02일
(56) 선행기술조사문헌
KR101958263 B1*
KR1020090117771 A*
KR1020160097870 A*
KR1020210062256 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
김영갑
서울특별시 광진구 능동로 209, 대양AI센터 701호
팔라시 잉글
서울특별시 광진구 능동로 209, 대양AI센터 732호
(74) 대리인
특허법인이상

전체 청구항 수 : 총 14 항

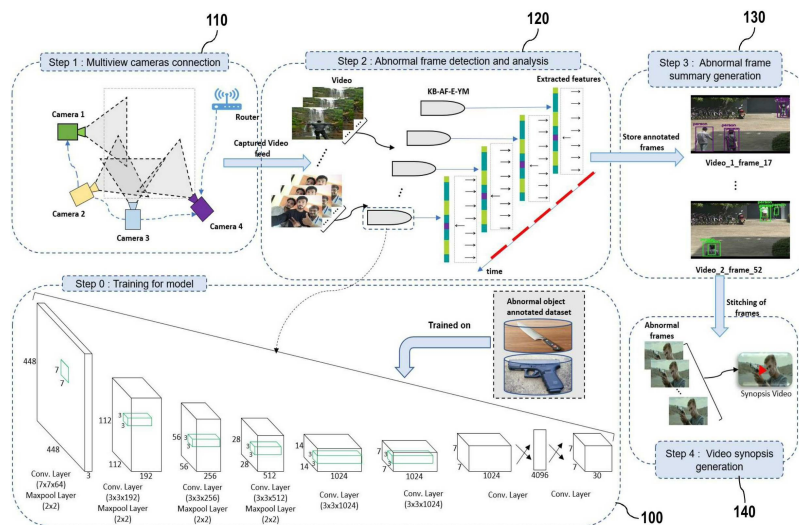
심사관 : 장우진

(54) 발명의 명칭 이상 객체 탐지 기반 영상 시놉시스 생성 방법 및 장치

(57) 요약

여러 대의 카메라로 구성된 다시점 영상감시시스템에 적용할 수 있고, 다양한 혼잡 상황에서도 위험 상황을 빠르게 탐지하여 실시간으로 영상 시놉시스를 생성할 수 있는 방법과 장치를 제공한다. 본 발명의 영상 시놉시스 생성 방법은 복수의 비디오를 받아들이고, 상기 복수의 비디오에서 소정의 객체가 포함된 비정상 프레임들을 추출하여 영상 시놉시스를 생성하는 것으로서, 복수의 비디오 각각에 대하여, 입력 영상 프레임에서 상기 비정상 프레임들을 추출하는 단계; 상기 비정상 프레임들을 결합하여 요약 영상을 생성하는 단계; 및 상기 복수의 비디오 각각에 대한 요약 영상을 결합하여 상기 영상 시놉시스를 생성하는 단계;를 포함한다. 상기 비정상 프레임들을 추출하는 단계는 상기 입력 영상 프레임에서 소정의 기준에 따라서 흐릿한 프레임을 구분하는 단계;를 포함한다. 이에 따라, 상기 흐릿한 프레임이 상기 요약 영상과 상기 영상 시놉시스에 포함되지 않게 된다.

대표도



(52) CPC특허분류

G06T 7/20 (2013.01)

H04N 21/845 (2013.01)

H04N 21/8549 (2013.01)

G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711126071
과제번호	2019-0-00231-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보보호핵심원천기술개발(R&D, 정보화)
연구과제명	공공 인프라 안전을 위한 인공지능 기반 영상보안 기술 및 시스템 개발
기 여 율	1/1
과제수행기관명	세종대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

복수의 카메라 각각을 통해서 획득된 복수의 비디오를 받아들이고, 상기 복수의 비디오에서 소정의 비정상 객체가 포함된 비정상 프레임들을 추출하여 상기 비정상 프레임들을 포함하는 영상 시뮬시스를 생성하는 방법으로서,

상기 복수의 비디오 각각에 대하여, 입력 영상 프레임에서 상기 비정상 프레임들을 추출하고, 추출된 비정상 프레임들을 포함하는 요약 영상을 생성하는 단계; 및

상기 복수의 비디오 각각에 대한 상기 요약 영상에 포함된 상기 비정상 프레임들을 포함하는 하나의 프레임 시퀀스 형태로 되어 있는 상기 영상 시뮬시스를 생성하는 단계;를 포함하며,

상기 비정상 프레임들을 추출하고 상기 요약 영상을 생성하는 단계가

상기 입력 영상 프레임에서 상기 소정의 객체를 추출하고, 추출된 객체가 실제로 상기 비정상 객체일 확률을 나타내는 비정상 점수를 산출하는 단계;

상기 추출된 객체에 대한 상기 비정상 점수를 토대로 상기 입력 영상 프레임에 대한 프레임 비정상 점수를 결정하는 단계; 및

상기 입력 영상 프레임에서 상기 프레임 비정상 점수를 토대로 흐릿한 프레임을 구분하는 단계;를 포함함으로써, 정상 프레임으로 판단되는 프레임과 상기 흐릿한 프레임으로 구분된 프레임이 상기 요약 영상과 상기 영상 시뮬시스에 포함되지 않게 하는 영상 시뮬시스 생성 방법.

청구항 2

청구항 1에 있어서, 상기 흐릿한 프레임을 구분하는 단계는

소정의 커널을 상기 입력 영상 프레임에 대하여 컨볼루션 연산함으로써 라플라시안 분산을 연산하는 단계;를 포함하여, 상기 라플라시안 분산을 소정의 임계치와 비교함으로써 상기 흐릿한 프레임을 구분하는 영상 시뮬시스 생성 방법.

청구항 3

청구항 1에 있어서, 상기 비정상 프레임들을 추출하는 단계에서, 소정의 인공지능망 모델에 의하여 상기 비정상 프레임을 추출하는 영상 시뮬시스 생성 방법.

청구항 4

청구항 3에 있어서, 상기 인공지능망 모델이 옴로(YOLO) 모델인 영상 시뮬시스 생성 방법.

청구항 5

청구항 3에 있어서, 상기 비정상 프레임들을 추출하는 단계는

상기 입력 영상 프레임의 색체계를 CIELUV 색공간으로 변환하는 단계;를 포함하며, 색공간 변환된 영상 프레임을 토대로 상기 인공지능망이 상기 비정상 프레임을 추출하는 영상 시뮬시스 생성 방법.

청구항 6

삭제

청구항 7

삭제

청구항 8

청구항 1에 있어서, 상기 요약 영상을 생성하는 단계는

상기 입력 영상 프레임에 대하여 교차 엔트로피를 결정하는 단계; 및

상기 교차 엔트로피에 따라 상기 요약 영상에 포함될 프레임들을 결정하는 단계;

를 더 포함하는 영상 시놉시스 생성 방법.

청구항 9

청구항 1에 있어서, 상기 소정의 객체는 총기와 칼 중 적어도 하나를 포함하는 영상 시놉시스 생성 방법.

청구항 10

복수의 카메라 각각을 통해서 획득된 복수의 비디오를 받아들이고, 상기 복수의 비디오에서 소정의 비정상 객체가 포함된 비정상 프레임들을 추출하여 상기 비정상 프레임들을 포함하는 영상 시놉시스를 생성하는 장치로서,

프로그램 명령들을 저장하는 메모리와; 상기 메모리에 통신가능하게 접속되고 상기 메모리에 저장된 상기 프로그램 명령들을 실행하는 프로세서;를 구비하며,

상기 프로그램 명령들은 상기 프로세서에 의해 실행될 때 상기 프로세서로 하여금:

상기 복수의 비디오 각각에 대하여, 입력 영상 프레임에서 상기 비정상 프레임들을 추출하고, 추출된 비정상 프레임들을 포함하는 요약 영상을 생성하는 동작; 및

상기 복수의 비디오 각각에 대한 상기 요약 영상에 포함된 상기 비정상 프레임들을 포함하는 하나의 프레임 시퀀스 형태로 되어 있는 상기 영상 시놉시스를 생성하는 동작;을 수행하게 하며,

상기 비정상 프레임들을 추출하고 상기 요약 영상을 생성하는 동작을 수행하는 명령들은

상기 입력 영상 프레임에서 상기 소정의 객체를 추출하고, 추출된 객체가 실제로 상기 비정상 객체일 확률을 나타내는 비정상 점수를 산출하는 동작;

상기 추출된 객체에 대한 상기 비정상 점수를 토대로 상기 입력 영상 프레임에 대한 프레임 비정상 점수를 결정하는 동작; 및

상기 입력 영상 프레임에서 상기 프레임 비정상 점수를 토대로 흐릿한 프레임을 구분하는 동작;을 수행하는 명령들을 포함함으로써, 정상 프레임으로 판단되는 프레임과 상기 흐릿한 프레임으로 구분된 프레임이 상기 요약 영상과 상기 영상 시놉시스에 포함되지 않게 하는 영상 시놉시스 생성 장치.

청구항 11

청구항 10에 있어서, 상기 흐릿한 프레임을 구분하는 동작을 수행하는 명령들은

소정의 커널을 상기 입력 영상 프레임에 대하여 컨볼루션 연산함으로써 라플라시안 분산을 연산하는 동작;을 수행하게 하는 명령들을 포함하여, 상기 라플라시안 분산을 소정의 임계치와 비교함으로써 상기 흐릿한 프레임을 구분하는 영상 시놉시스 생성 장치.

청구항 12

청구항 10에 있어서, 상기 비정상 프레임들을 추출하는 동작을 수행하는 명령들은, 소정의 인공지능망 모델에 의하여 상기 비정상 프레임을 추출하는 영상 시놉시스 생성 장치.

청구항 13

청구항 12에 있어서, 상기 인공지능망 모델이 옴로(YOLO) 모델인 영상 시놉시스 생성 장치.

청구항 14

청구항 12에 있어서, 상기 비정상 프레임들을 추출하는 명령들은

상기 입력 영상 프레임의 색체계를 CIELUV 색공간으로 변환하는 동작;을 수행하는 명령들을 포함하며, 색공간

변환된 영상 프레임들 토대로 상기 인공지능망이 상기 비정상 프레임을 추출하는 영상 시놉시스 생성 장치.

청구항 15

삭제

청구항 16

삭제

청구항 17

청구항 10에 있어서, 상기 요약 영상을 생성하는 명령들은

상기 입력 영상 프레임에 대하여 교차 엔트로피를 결정하는 동작; 및

상기 교차 엔트로피에 따라 상기 요약 영상에 포함될 프레임들을 결정하는 동작;

을 수행하는 명령들을 더 포함하는 영상 시놉시스 생성 장치.

청구항 18

청구항 10에 있어서, 상기 소정의 객체는 총기와 칼 중 적어도 하나를 포함하는 영상 시놉시스 생성 장치.

발명의 설명

기술 분야

[0001] 본 발명은 영상 감시 시스템에 관한 것으로서, 특히, 영상감시카메라에 의해 획득된 영상으로부터 시놉시스를 생성하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 영상감시카메라의 증가에 따라, 영상 데이터의 양이 과도하게 증가하였다. 이에 따라, 영상 데이터의 분석, 추출, 및 저장이 어려워지고, 더 많은 데이터 저장공간과 분석 인력이 필요하게 되었다. 영상 시놉시스는 긴 영상을 짧게 요약하는 것으로서, 영상 데이터의 양을 감소시키고 몇 시간 분량의 영상을 단 몇 분 만에 검토할 수 있게 해주기 때문에, 데이터 저장공간 문제를 해소하고 영상 분석에 소요되는 비용과 노동력을 크게 감소시킬 수 있게 해준다.

[0003] 그런데 기존의 영상 시놉시스 생성 방법은 한 대의 카메라에 의해 획득한 영상에 대한 시놉시스를 다루기 때문에, 여러 대의 카메라로 구성되어있는 영상감시시스템에서는 적합하지 않다. 또한 기존의 영상 시놉시스 생성 방법은 전체 영상 프레임을 처리하여 생성될 뿐만 아니라 모든 움직임 객체를 포함하도록 생성되기 때문에, 전 처리 부담이 커서 실시간 시나리오에 대한 성능이 낮고, 특히 무의미한 객체 데이터 처리 부담으로 인하여 혼잡 상황에서 성능이 크게 저하된다는 문제가 있다.

발명의 내용

해결하려는 과제

[0004] 본 발명은 이와 같은 문제점을 해결하기 위한 것으로서, 여러 대의 카메라로 구성된 다시점 영상감시시스템에 적용할 수 있고, 다양한 혼잡 상황에서도 위험 상황을 빠르게 탐지하여 실시간으로 영상 시놉시스를 생성할 수 있는 방법과 장치를 제공하는 것을 기술적 과제로 한다.

과제의 해결 수단

[0005] 상기 기술적 과제를 달성하기 위한 본 발명의 일 측면에 따르면, 복수의 비디오를 받아들이고, 상기 복수의 비디오에서 소정의 객체가 포함된 비정상 프레임들을 추출하여 영상 시놉시스를 생성하는 방법이 제공된다. 영상 시놉시스 생성 방법은 상기 복수의 비디오 각각에 대하여, 입력 영상 프레임에서 상기 비정상 프레임들을 추출하는 단계; 상기 비정상 프레임들을 결합하여 요약 영상을 생성하는 단계; 및 상기 복수의 비디오 각각에 대한

요약 영상을 결합하여 상기 영상 시놉시스를 생성하는 단계;를 포함한다. 상기 비정상 프레임들을 추출하는 단계는 상기 입력 영상 프레임에서 소정의 기준에 따라서 흐릿한 프레임을 구분하는 단계;를 포함한다. 이에 따라, 상기 흐릿한 프레임이 상기 요약 영상과 상기 영상 시놉시스에 포함되지 않게 된다.

- [0006] 상기 흐릿한 프레임을 구분하는 단계는 소정의 커널을 상기 입력 영상 프레임에 대하여 컨벌루션 연산함으로써 라플라시안 분산을 연산하는 단계;를 포함할 수 있다. 이 경우, 상기 라플라시안 분산을 소정의 임계치와 비교함으로써 상기 흐릿한 프레임이 구분될 수 있다.
- [0007] 상기 비정상 프레임들을 추출하는 단계에서, 상기 비정상 프레임은 소정의 인공신경망 모델에 의하여 추출될 수 있다.
- [0008] 상기 인공신경망 모델은 옴로(YOLO) 모델일 수 있다.
- [0009] 상기 비정상 프레임들을 추출하는 단계는 상기 입력 영상 프레임의 색체계를 CIELUV 색공간으로 변환하는 단계;를 포함할 수 있다. 이 경우, 상기 인공신경망은 색공간 변환된 영상 프레임을 토대로 상기 비정상 프레임을 추출할 수 있다.
- [0010] 상기 비정상 프레임들을 추출하는 단계는 상기 입력 영상 프레임에서 상기 소정의 객체를 추출하고, 상기 소정의 객체가 실제로 비정상 객체일 확률을 나타내는 비정상 점수를 산출하는 단계; 및 상기 객체에 대한 상기 비정상 점수를 토대로 상기 입력 영상 프레임에 대한 프레임 비정상 점수를 결정하는 단계;를 포함할 수 있다.
- [0011] 상기 요약 영상은 상기 입력 영상 프레임 중에서 상기 프레임 비정상 점수를 토대로 정상 프레임으로 판단되는 프레임과 상기 흐릿한 프레임으로 구분된 프레임을 배제하고 생성될 수 있다. 이 경우, 상기 요약 영상을 생성하는 단계는 상기 입력 영상 프레임에 대하여 교차 엔트로피를 결정하는 단계; 및 상기 교차 엔트로피에 따라 상기 요약 영상에 포함될 프레임들을 결정하는 단계;를 포함할 수 있다.
- [0012] 상기 소정의 객체는 총기와 칼 중 적어도 하나를 포함할 수 있다.
- [0013] 본 발명의 다른 측면에 따르면, 복수의 비디오를 받아들이고, 상기 복수의 비디오에서 소정의 객체가 포함된 비정상 프레임들을 추출하여 영상 시놉시스를 생성하는 영상 시놉시스 생성 장치가 제공된다. 영상 시놉시스 생성 장치는 프로그램 명령들을 저장하는 메모리와; 상기 메모리에 통신가능하게 접속되고 상기 메모리에 저장된 상기 프로그램 명령들을 실행하는 프로세서;를 구비한다. 상기 프로그램 명령들은 상기 프로세서에 의해 실행될 때 상기 프로세서로 하여금: 상기 복수의 비디오 각각에 대하여, 입력 영상 프레임에서 상기 비정상 프레임들을 추출하는 동작; 상기 비정상 프레임들을 결합하여 요약 영상을 생성하는 동작; 및 상기 복수의 비디오 각각에 대한 요약 영상을 결합하여 상기 영상 시놉시스를 생성하는 동작;을 수행하게 할 수 있다. 상기 비정상 프레임들을 추출하는 동작을 수행하는 명령들은 상기 입력 영상 프레임에서 소정의 기준에 따라서 흐릿한 프레임을 구분하는 동작;을 수행하는 명령들을 포함할 수 있다. 이에 따라 상기 흐릿한 프레임이 상기 요약 영상과 상기 영상 시놉시스에 포함되지 않게 된다.
- [0014] 상기 흐릿한 프레임을 구분하는 동작을 수행하는 명령들은 소정의 커널을 상기 입력 영상 프레임에 대하여 컨벌루션 연산함으로써 라플라시안 분산을 연산하는 동작;을 수행하게 하는 명령들을 포함하여, 상기 라플라시안 분산을 소정의 임계치와 비교함으로써 상기 흐릿한 프레임을 구분할 수 있다.
- [0015] 상기 비정상 프레임들을 추출하는 동작을 수행하는 명령들은, 소정의 인공신경망 모델에 의하여 상기 비정상 프레임을 추출할 수 있다. 상기 인공신경망 모델이 옴로(YOLO) 모델일 수 있다.
- [0016] 상기 비정상 프레임들을 추출하는 명령들은 상기 입력 영상 프레임의 색체계를 CIELUV 색공간으로 변환하는 동작;을 수행하는 명령들을 포함할 수 있다. 이 경우, 상기 인공신경망은 색공간 변환된 영상 프레임을 토대로 상기 비정상 프레임을 추출할 수 있다.
- [0017] 상기 비정상 프레임들을 추출하는 동작을 수행하는 명령들은 상기 입력 영상 프레임에서 상기 소정의 객체를 추출하고, 상기 소정의 객체가 실제로 비정상 객체일 확률을 나타내는 비정상 점수를 산출하는 동작; 및 상기 객체에 대한 상기 비정상 점수를 토대로 상기 입력 영상 프레임에 대한 프레임 비정상 점수를 결정하는 동작;을 수행하는 명령들을 포함할 수 있다.
- [0018] 상기 요약 영상은 상기 입력 영상 프레임 중에서 상기 프레임 비정상 점수를 토대로 정상 프레임으로 판단되는 프레임과 상기 흐릿한 프레임으로 구분된 프레임을 배제하고 생성될 수 있다. 이 경우, 상기 요약 영상을 생성하는 명령들은 상기 입력 영상 프레임에 대하여 교차 엔트로피를 결정하는 동작; 및 상기 교차 엔트로피에 따라

상기 요약 영상에 포함될 프레임들을 결정하는 동작;을 수행하는 명령들을 포함할 수 있다.

[0019] 상기 소정의 객체는 총기와 칼 중 적어도 하나를 포함할 수 있다.

발명의 효과

[0020] 본 발명의 일 실시예에 따르면, 다양한 혼잡 상황에서도 위험 상황을 빠르게 탐지하여 실시간으로 영상 시놉시스를 생성할 수 있게 된다. 특히, 본 발명에 의한 시놉시스 생성 방법과 장치는 여러 대의 카메라로 구성된 다시점 영상감시시스템에 적용할 수 있다.

도면의 간단한 설명

- [0021] 도 1은 본 발명의 일 실시예에 따른 영상 감시 시스템의 전체적인 구성을 보여주는 개략적인 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른 영상 시놉시스 생성 과정을 보여주는 흐름도이다.
- 도 3은 본 발명의 일 실시예에 따른 프레임 선택기 모듈의 블록도이다.
- 도 4는 CIELUV 색공간에 표시된 색채들의 예를 보여준다.
- 도 5는 다수의 카메라로부터 입력된 영상 프레임에 대하여 프레임 비정상 점수를 칼라로 시각화한 것을 보여준다.
- 도 6은 라플라시안 분산에 따른 프레임 선명도의 차이를 보여준다.
- 도 7은 비정상 요약 영상을 생성하는 프로세스에 대한 의사 코드(pseudo-code)의 일 예를 보여준다.
- 도 8은 저장된 다시점 비디오에 대한 비디오 시놉시스 생성 예를 보여준다.
- 도 9는 라이브 카메라 영상에 대한 비디오 시놉시스 생성 예를 보여준다.
- 도 10은 본 발명의 일 실시예에 따른 영상 시놉시스 생성 장치의 블록도이다.
- 도 11은 본 발명의 실험에서 객체 탐지 모델의 훈련에 사용된 테스트 영상에 대한 정보를 정리한 표이다.
- 도 12는 본 발명에 의한 객체 탐지 모델과 다른 최신 탐지 모델들의 총과 칼을 탐지하는 실험에서의 평균 정밀도를 정리한 표이다.
- 도 12는 본 발명에 의한 객체 탐지 모델과 다른 모델들의 탐지 정확도에 대한 정량적 평가 지표를 정리한 표이다.
- 도 13은 도 12에 정리된 정량적 평가 지표를 도시한 그래프이다.
- 도 14는 본 발명에 의한 영상 시놉시스 생성 방법과 다른 알고리즘들의 요약 영상 생성에 있어서의 정밀도에 대한 비교 결과를 도시한 그래프이다.
- 도 15는 본 발명에 의한 영상 시놉시스 생성 방법과 다른 알고리즘들의 영상 시놉시스 생성 결과를 대비하여 정리한 표이다.

발명을 실시하기 위한 구체적인 내용

[0022] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0023] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는 데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0024] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이

해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0025] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0026] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0027] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.

[0028] 도 1은 본 발명의 일 실시예에 따른 영상 감시 시스템의 전체적인 구성을 보여주는 개략적인 블록도이다. 도시된 영상 감시 시스템은 감시대상 지역에 분산설치되어 감시 영상을 획득하는 복수의 감시 카메라(10a~10n)와, 상기 복수의 감시 카메라(10a~10n)가 접속할 수 있는 감시제어 서버(20)를 구비한다. 일 실시예에 있어서, 감시 카메라들(10a~10n)은 유선망 및/또는 무선망을 기반으로 한 IP 네트워크를 통해서 감시제어 서버(20)에 접속된다. 감시 카메라들(10a~10n)은 유선망의 일 예로서 USB 케이블을 통해서 감시제어 서버(20)에 접속될 수도 있다.

[0029] 각 감시 카메라(10a~10n)는 감시제어 서버(20)의 제어신호에 응답하여 수평회전(panning) 및 수직회전(tilting)하면서 적절한 줌 배율로 주변 지역을 촬영할 수 있다. 감시제어 서버(20)는 각 감시 카메라(10a~10n)로부터의 영상에서 움직인 객체 또는 비정상 객체를 탐지하고, 필요에 따라 감시 카메라(10a~10n)의 팬, 틸트, 줌을 제어할 수 있다. 또한, 감시제어 서버(20)는 감시 카메라(10a~10n)로부터 획득한 영상에서 비정상 객체가 포함된 영상 프레임들을 추출하여 영상 시놉시스를 구성할 수 있다. 청구범위를 포함하여 본 명세서에서, "비정상 객체(Abnormal object)"이란 예컨대 총기, 칼, 또는 총기나 칼을 들고 있는 사람과 같이 타인에게 위협을 가할 수 있거나, 위험 상황을 야기할 수 있는 객체를 일컫는데, 비정상 객체의 종류가 이에 한정되는 것은 아니다. 예를 들어, 본 발명의 다른 실시예에서는, 비정상 객체가 ...

[0030] 감시제어 서버(20)는 영상 시놉시스를 디스플레이 장치를 통해서 운영자에게 표출할 수 있고, 저장 장치에 저장할 수도 있다. 감시제어 서버(20)는 후술하는 바와 같이 적어도 하나의 프로세서를 구비하며, 상기 적어도 하나의 프로세서에 의한 프로그램 코드의 실행을 통해 비정상 객체의 검출, 영상 시놉시스의 생성, 영상 디스플레이, 영상 저장 등의 기능을 수행한다.

[0031] 감시제어 서버(20)는 예컨대 클라우드 서버 내에서 구현될 수도 있다. 대부분의 카메라 네트워크에서 각 카메라에 의해 생성되는 비디오 피드는 클라우드 서버에 저장되기 때문에, 리소스와 대역폭이 낭비된다. 이 문제를 해소하기 위하여, 본 발명에서 상기 감시제어 서버(20)가 에지/클라우드 컴퓨팅 장치에서 직접 구현될 수 있으며, 그 결과, 각 감시 카메라(10a~10n)에 의해 획득된 원본 영상 전체가 감시제어 서버(20)에 저장되는 것이 아니라, 시놉시스 영상만이 저장될 수도 있다.

[0032] 도 2는 본 발명의 일 실시예에 따른 영상 시놉시스 생성 과정을 보여주는 흐름도이다. 본 실시예에 따른 영상 시놉시스 생성 과정은 이상 상황을 탐지하기 위한 객체 탐지 모델의 훈련 단계(제100단계), 복수의 카메라의 연결을 확인하는 단계(제110단계), 학습된 객체 탐지 모델을 사용하여 이상 상황을 탐지하고 추출하는 단계(제120단계), 비정상 프레임의 요약물 생성하는 단계(제130단계), 및 복수의 카메라들에 대한 이상 상황 비디오 시놉시스를 생성하는 단계(140)를 포함한다.

[0033] 위에서 언급한 바와 같이, 본 발명은 영상 프레임 내에 총이나 칼과 같은 비정상 객체가 있는지 탐지하여 해당 프레임의 이상 상황 여부를 판단한다. 본 발명의 예시적인 실시예에 있어서, 비정상 객체를 탐지하기 위한 객체 탐지 모델로는 YOLOv3가 사용되며, 파이썬(Python)으로 작성된 오픈소스 신경망 라이브러리인 케라스(Keras)를 토대로 구현된다. 즉, 본 실시예에 따른 객체 탐지 모델은 '케라스를 토대로 구현되며 비정상 프레임 객

체를 추출하는 YOLO 모델(KB-AF-E-YM: Keras based-abnormal frame-extraction-Yolo model)'이라고 칭할 수 있다. 그렇지만, 본 발명에서 객체 탐지모델이 YOLOv3 모델에 한정되는 것은 아니며, 예컨대 YOLOv2나 YOLOv5와 같은 다른 옴로 모델이 사용될 수도 있고, 옴로 모델 이외의 모델이 사용될 수도 있다.

- [0034] 도 2의 제100단계는 옴라인 전처리 단계로서, 이 단계에서는 객체 탐지 모델에 대한 훈련이 진행된다. 객체 탐지 모델의 훈련은 프레임이 입력되면 해당 프레임 내에 있는 비정상 객체를 탐지하고 실제 비정상 객체일 확률을 나타내는 비정상 점수(abnormal score)을 최대한 정확하게 출력하도록 이루어질 수 있다.
- [0035] 모델을 훈련하기 위해서는 방대한 데이터 세트를 대상으로 한 많은 계산이 필요하다. 일 실시예에 따르면, ImageNet 데이터 세트, Open-Image 데이터 세트, IMFD(Internet Movie Firearms Database) 및 Gettyimages에서 제공하는 이미지를 사용하여 보다 큰 커스텀 데이터 세트를 만들었다. 또한, 이러한 이미지에 스케일링(scaling), 뒤집기(flippping), 및 자르기(shearing)과 같은 데이터 변형 기술을 적용하여 데이터를 확대함으로써, 총 432427개의 훈련용 이미지 데이터를 구성하였다. 각 이미지에 주석을 추가한 후 Keras 기반 프레임워크를 사용하여 YOLOv3 가중치를 사전 훈련시킴으로써, 객체 탐지 모델을 재구성할 수 있는 업데이트된 가중치를 획득하였다.
- [0036] 제110단계에서는, 영상 시놉시스를 생성하는 서버 예컨대 감시제어 서버(20)는 복수의 카메라(10a~10n)의 연결 상태를 확인할 수 있다. 이때, 연결 상태의 확인은 복수의 카메라(10a~10n)에 대한 새로운 연결을 의미할 수도 있고, 연결되어 있는 카메라의 목록을 확인하는 작업을 의미할 수도 있으며, 기존에 연결된 카메라(10a~10n) 중에서 영상 시놉시스 생성에 사용할 원본 영상을 받아들일 일부 카메라를 선별하는 것을 의미할 수도 있다. 연결되는 카메라들은 유선 LAN 또는 무선 LAN과 같은 네트워크를 사용하여 감시제어 서버(20)에 연결될 수도 있고, USB 케이블을 통해 연결될 수도 있다. 각 카메라에는 영상 시놉시스 생성 시에 적용할 일련번호가 부여될 수 있다.
- [0037] 제120단계에서는, 제100단계에서 학습된 객체 탐지 모델을 사용하여 이상 상황을 탐지하고 추출한다. 도 3은 이상 상황의 탐지와 추출에 사용되는 프레임 선택기 모듈의 블록도이다. 프레임 선택기 모듈(122)은 프레임 추출부(124), 학습된 객체 탐지 모델(124: KB-AF-E-YM), 및 프레임 선별부(128)를 포함한다.
- [0038] 프레임 추출부(124)는 카메라들(10a~10n)로부터 입력되는 영상 프레임들의 색체계를 CIELUV 색공간으로 변환할 수 있다. 프레임 추출부(124)는 CIELUV 색공간의 절대 차(absolute difference) 값을 활용하여 이전 프레임과 일정 수준 이상으로 국부적 휘도 레벨이 다른 프레임들을 추출한다. 즉, 전처리 작업으로서, 프레임 추출부(124)는 수신되는 각 입력 영상 프레임에 대하여 픽셀 또는 소정의 이미지 블록 단위로 이전 프레임과의 CIELUV 색공간 상의 절대 차 값을 계산하고, 절대 차 값의 국부적인 또는 전체적인 합산치가 큰 프레임을 추출한다.
- [0039] 도 4는 CIELUV 색공간에 표시된 색체들의 예를 보여준다. CIELUV 색공간에서, 휘도(L)는 0~100의 범위에 있는 값을 가질 수 있고 u, v 좌표는 녹색/빨간색 및 파란색/노란색 축의 방향을 추정할 수 있게 해주며 -100에서 100 사이의 값을 갖는다. CIELUV 색공간은 인간의 시각 인지 특성을 감안하여 휘도를 더 세밀하게 표현할 수 있는 색체계로서, 상기 색체계 변환은 영상의 feature quality를 향상시켜 비정상 객체를 더 잘 탐지할 수 있게 해준다. 최소 색상 콘트라스트를 측정하는 것이 효율적인데, 이는 작은 물체를 검출할 때 많은 장점이 있다.
- [0040] 본 실시예에 따르면, 변화량이 일정 기준보다 높은 프레임만이 객체 탐지 모델(126)로 입력되어, 변화량이 큰 프레임에 대해서만 객체 탐지 모델(126)이 비정상 객체를 탐지하게 할 수 있다. 그렇지만, 변형된 실시예에서는 프레임 추출부(124)가 생략되어 모든 영상 프레임이 객체 탐지 모델(126)에 공급될 수 있다. 아울러, 프레임 추출부(124)의 기능 중에서 색공간 변환과 프레임 추출 중 어느 하나만이 생략될 수도 있다.
- [0041] 객체 탐지 모델(126)은 프레임 추출부(124)로부터 영상 프레임을 받아들이고, 영상 프레임에서 비정상 객체를 탐지한다. 비정상 객체의 탐지는 위에서 언급한 바와 같이 KB-AF-E-YM이라 명명되는 YOLO 모델에 의해 이루어질 수 있다. 상기 YOLO 모델은 다수의 컨벌루션 레이어들을 통해 특징 맵(feature map)을 추출하고 fully connected layer를 거쳐 바운딩 박스를 정하고 클래스 점수를 예측한다. 즉, YOLO 모델은 각 영상 프레임을 SxS 개의 그리드로 분할하고, 각 그리드에 비정상 객체가 포함되었는지 여부를 계산하기 위하여 객체 클래스 점수를 계산한다. 각 그리드에 대한 신뢰도를 높일 수 있도록 주변의 그리드를 합치고 바운딩 박스의 위치를 조정함으로써, 객체 인식의 정확도를 높이고 클래스 점수의 신뢰도를 높일 수 있다.
- [0042] 객체 탐지 모델(126)은 비정상 객체와 객체 클래스 점수를 출력할 수 있다. 일 실시예에 있어서, 비정상 객체의 정보를 바운딩 박스의 위치와 크기로 나타내어질 수 있다. 상기 객체 클래스 점수는 해당 객체가 실제 비정상 객체일 확률을 나타내는 비정상 점수(abnormal score)일 수 있다. 상기 비정상 점수는 0과 1 사이의 값으로

매겨져서 주식 형태로 영상 프레임 또는 바운딩 박스에 부가될 수 있다. 상기 비정상 점수는 1에 근접할수록 비정상일 확률이 높다는 것을 의미할 수 있다.

[0043] 객체에 대한 비정상 점수를 토대로 프레임의 비정상 점수가 계산될 수 있다. 예를 들어, 영상 프레임에서 탐지된 객체들의 비정상 점수 중 가장 높은 비정상 점수가 프레임의 비정상 점수로 정해질 수 있다. 프레임 비정상 점수(α)에 따라, 각 프레임은 정상 프레임과 비정상 프레임으로 구분될 수 있다. 즉, $\alpha=0$ 이라면 현재 처리중인 프레임에는 원하는 객체가 없으며, 해당 프레임은 정상 프레임(NF: Normal Frame)으로 표시된다. 만약 $\alpha=1$ 이라면 처리중인 프레임에 원하는 객체가 존재한다는 의미이며, 해당 프레임은 비정상 프레임(ANF: Abnormal Frame)으로 표시된다. 이와 같이, KB-AF-E-YM은 프레임에서 객체를 선택하고 프레임에 주석을 추가하게 된다. 주석이 부가된 프레임 정보는 매트릭스(MA) 시퀀스의 형태로 다음 단계로 넘겨진다. 이때, 상기 주석에는 비정상 객체에 대한 정보와, 프레임 비정상 점수 및/또는 프레임의 정상/비정상 여부에 대한 정보가 포함된다.

[0044] 도 5는 다수의 카메라로부터 입력된 영상 프레임에 대하여 프레임 비정상 점수를 칼라로 시각화한 것을 보여준다. 도 5의 (a)는 8개의 카메라로부터 입력되는 각 6개 프레임의 영상에 대한 프레임 시퀀스 열(sequential frame series)과 속성(attribute) 즉 프레임 비정상 점수를 보여준다. 도면에서, 보라색 부분은 비정상 프레임을 나타내고 다른 모든 색상은 정상 프레임임을 나타낸다. 아래로 향하는 화살표는 비정상 프레임의 표시를 의미하고, 위로 향하는 화살표는 정상 프레임의 표시를 의미한다.

[0045] 프레임 선별부(128)는 객체 탐지 모델(126)로부터 전달받은 프레임 중에서 흐릿한 프레임을 버리고 선명한 프레임만 선택함으로써, 우수한 화질의 프레임만으로 매끄러운 시놉시스가 생성될 수 있게 해준다. 일 실시예에 있어서, 프레임 선별부(128)는 영상 프레임에 라플라시안 커널을 적용하여 라플라시안 분산(laplacian variance)을 계산함으로써, 상기 라플라시안 분산을 토대로 흐릿한 프레임(blur frame)과 선명한 프레임을 구분할 수 있게 해준다. 수학적 식 1은 3×3 행렬로 되어 있는 라플라시안 커널의 일 예를 보여준다. 프레임 선별부(128)는 영상 프레임에 대하여 수평 및 수직 방향으로 일정한 이동간격(stride) 단위로 슬라이딩시켜 상대적으로 이동시키면서 컨벌루션 연산을 수행함으로써 라플라시안 분산을 계산할 수 있다.

수학적 식 1

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

[0047] 상기 라플라시안 분산이 일정한 임계값보다 작으면, 해당 프레임은 흐릿한 프레임으로 간주되어, 폐기되거나 주석에 표시될 수 있다. 반면에, 상기 라플라시안 분산이 일정한 임계값보다 크면, 해당 프레임은 선명한 프레임으로 간주될 수 있다. 도 6은 라플라시안 분산에 따른 프레임 선명도의 차이를 보여준다. 발명자들의 실험에 따르면, 라플라시안 분산이 (0, 159) 사이의 구간에 있으면 프레임 영상의 화질이 최악이었고, 라플라시안 분산이 (160, 349) 사이의 구간에 있으면 영상 화질이 평균 범위에 있다고 할 수 있었으며, 라플라시안 분산이 350 이상인 경우에는 프레임 영상의 화질이 우수하였다. 따라서, 중요한 정보인 비정상 프레임의 폐기를 방지하면서, 화질이 열악하지 않은 프레임 영상을 선별할 수 있게 임계값을 정하는 것이 중요하다. 응용분야에 따라 다르지만, 본 실시예에서 라플라시안 분산의 임계값은 160으로 설정되어 사용되었다.

[0048] 다시 도 2를 참조하면, 제130단계에서는, 프레임 선택기 모듈(122)로부터 출력되는 영상 프레임들 중에서 비정상 프레임을 모아서 비정상 요약 영상(abnormal summary video)을 생성한다. 제130단계는 다수의 감시 카메라(10a~10n)로부터 입력되는 다수의 입력 영상 각각에 대하여 별도로 수행되어, 모든 카메라 영상에 대한 요약을 생성하게 된다.

[0049] 일 실시예에 있어서, 비정상 요약 영상은 프레임 비정상 점수를 토대로 비정상 프레임을 선택하고, 화질이 기준에 미치지 못하는 프레임을 배제한 후, 남은 프레임들을 이어붙여서 생성될 수 있다. 예를 들어, 프레임 비정상 점수가 0.5 미만인 프레임은 배제하고 0.5 이상인 프레임만으로 비정상 요약 영상을 구성할 수 있다.

[0050] 도 7은 비정상 요약 영상을 생성하는 프로세스에 대한 의사 코드(pseudo-code)의 일 예를 보여준다. 객체 탐지 모델(126) 즉, KB-AF-E-YM이 프레임 내의 객체들의 비정상 점수와 객체들의 비정상 점수를 기반으로 한 해당 프레임의 비정상 점수 결정을 수행하며, 이를 어텐션 모델(attention model)이 백그라운드에서 동시에 수행해 실시간 시놉시스 생성이 가능해진다.

- [0051] 도 2의 제140단계에서는, 각 카메라 영상의 비정상 요약물 각 카메라에 대하여 사전에 할당된 시퀀스 넘버 순서대로 스티칭하여 연결한다. 즉, 도 8 및 도 9와 같이 각 카메라 영상에 대하여 요약된 비정상 프레임들을 순차적으로 연결하게 된다. 도 8은 저장된 다시점 비디오에 대한 비디오 시놉시스 생성 예를 보여주고, 도 9는 라이브 카메라 영상에 대한 비디오 시놉시스 생성 예를 보여준다.
- [0052] 제130단계에서 각 카메라 영상에 대하여 요약 영상을 생성하는 동안에, 또는 제140단계에서 여러 카메라 영상에 대한 요약 영상들을 스티칭하여 시놉시스를 생성하는 동안에, 영상에 지터, 고스팅, 또는 흔들림 문제가 생길 수 있다. 이를 감안하여, 스티칭은 스티칭 파이프라인에서 시작하여, 픽셀의 특징점을 계산한 다음, 이미지를 등록하고 블렌딩하는 과정을 거치는 것이 바람직하다.
- [0053] 도 10은 본 발명의 일 실시예에 따른 영상 시놉시스 생성 장치의 블록도이다. 본 발명의 일 실시예에 따른 영상 시놉시스 생성 장치는 적어도 하나의 프로세서(220), 메모리(240), 및 저장 장치(260)를 포함할 수 있다. 앞에서 언급한 바와 같이, 영상 시놉시스 생성 장치는 도 1에 도시된 영상 감시 시스템에서 감시제어 서버(20)에 의해 구현될 수 있다. 이러한 경우, 감시제어 서버(20)는 각 감시 카메라(10a~10n)가 획득한 감시 영상을 해당 카메라로부터 받아들이고, 수신된 감시 영상에서 비정상 객체가 포함된 비정상 프레임을 탐지하여 영상 시놉시스를 생성한다.
- [0054] 프로세서(220)는 메모리(240) 및/또는 저장 장치(260)에 저장된 프로그램 명령을 실행할 수 있다. 프로세서(220)는 적어도 하나의 중앙 처리 장치(central processing unit, CPU)나 그래픽 처리 장치(graphics processing unit, GPU)에 의해 구현될 수 있으며, 그밖에 본 발명에 따른 방법을 수행할 수 있는 여타의 프로세서일 수 있다.
- [0055] 메모리(240)는 예컨대 RAM(Random Access Memory)와 같은 휘발성 메모리와, ROM(Read Only Memory)과 같은 비휘발성 메모리를 포함할 수 있다. 메모리(240)는 저장 장치(260)에 저장된 프로그램 명령을 로드하여, 프로세서(220)에 제공함으로써 프로세서(220)가 이를 실행할 수 있도록 할 수 있다.
- [0056] 저장 장치(260)는 프로그램 명령과 데이터를 저장하기에 적합한 기록매체로서, 예컨대 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM(Compact Disk Read Only Memory), DVD(Digital Video Disk)와 같은 광 기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media), 플래시 메모리나 EPROM(Erasable Programmable ROM) 또는 이들을 기반으로 제작되는 SSD와 같은 반도체 메모리를 포함할 수 있다.
- [0057] 저장 장치(260)는 상기 프로그램 명령을 저장한다. 특히, 상기 프로그램 명령은 본 발명에 따른 영상 시놉시스 생성 방법을 구현하기 위한 영상 시놉시스 생성 프로그램을 포함할 수 있다. 상기 영상 시놉시스 생성 프로그램은 상기 복수의 비디오 각각에 대하여, 입력 영상 프레임에서 상기 비정상 프레임들을 추출하는 동작; 상기 비정상 프레임들을 결합하여 요약 영상을 생성하는 동작; 및 상기 복수의 비디오 각각에 대한 요약 영상을 결합하여 상기 영상 시놉시스를 생성하는 동작;을 수행하게 하는 영상 시놉시스 생성 방법을 구현하는데 필요한 프로그램 명령들을 포함한다. 특히, 상기 비정상 프레임들을 추출하는 동작을 수행하는 명령들은 상기 입력 영상 프레임에서 소정의 기준에 따라서 흐릿한 프레임을 구분하는 동작;을 수행하는 명령들을 포함함으로써 상기 흐릿한 프레임이 상기 요약 영상과 상기 영상 시놉시스에 포함되지 않게 하는 명령들을 포함할 수 있다. 이와 같은 프로그램 명령은 프로세서(220)의 제어에 의해 메모리(240)에 로드된 상태에서, 프로세서(220)에 의해 실행되어 본 발명에 의한 방법을 구현할 수 있다.
- [0058] 본 발명에 의한 영상 시놉시스 생성 방법의 성능을 평가하기 위해 실험을 실시하고, 실험 결과를 다른 최신 연구와 비교하였다. 모든 평가는 3.59GHz의 클럭 속도를 가지는 AMD Ryzen 5 3500X 6코어 프로세서, 16GB 램, Nvidia Gigabyte Geforce RTX 2060 그래픽 카드, Logitech C920 Pro H.D. 웹캠을 사용하여 평가하였다.
- [0059] 객체 탐지 모델(124) 즉, KB-AF-E-YM을 훈련하기 위하여 TensorFlow를 사용하였으며, Keras API를 사용하여 추가 변환을 하였다. 객체 탐지에 대한 KB-AF-E-YM 모델을 훈련하기 위하여 위해 ImageNet, IMFD, OpenImage dataset의 이미지들을 사용하였다.
- [0060] 영상 시놉시스 성능을 테스트하기 위해 YouTube, IMFD, 라이브 카메라에서 추출된 영상을 사용했다. 6개의 영상 중에서 영상 1, 5, 6번은 YouTube로부터 획득하였는데, 이 영상에는 각각 2, 1, 2개의 비정상 객체가 있었다. 영상 2번은 IMFD로부터 획득하였고, 이 영상에는 3개의 비정상 객체가 있었다. 영상 3, 4번은 라이브 카메라로부터 입력하였고, 이들 영상 각각에는 1개의 비정상 객체가 있었다. 각 영상에 대한 자세한 정보는 도 11의 표에 정리되어 있다.

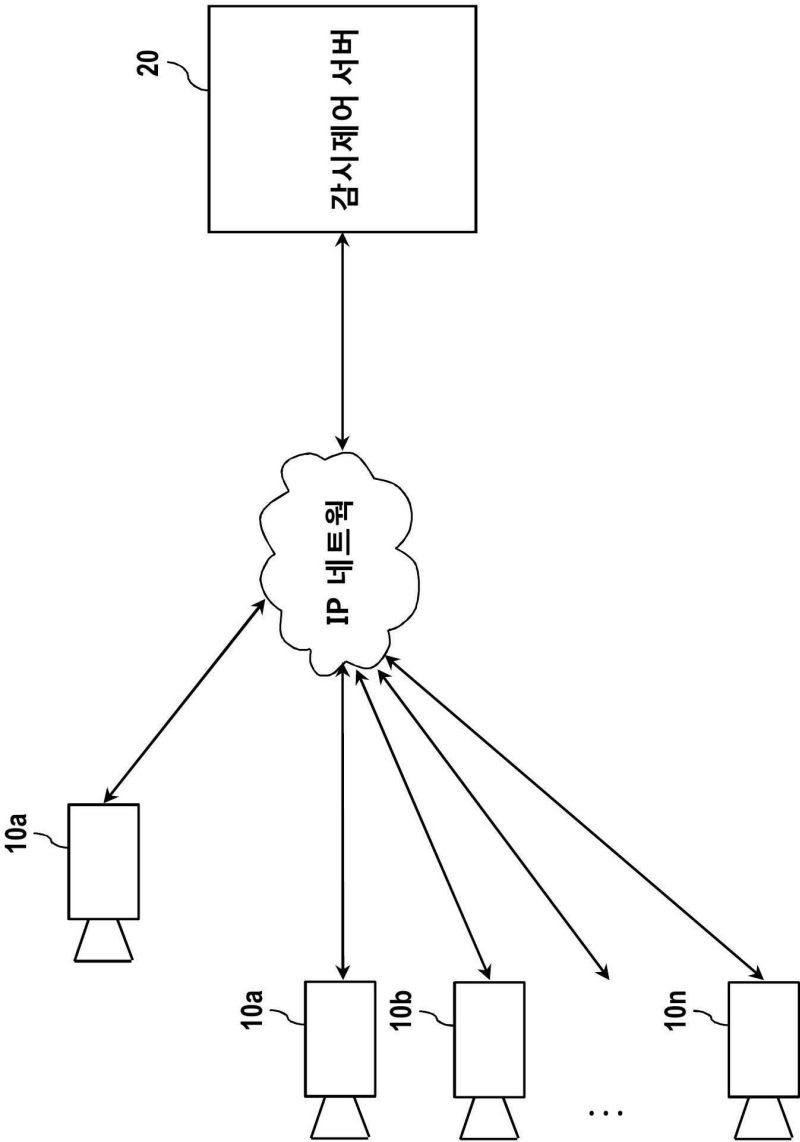
- [0061] 본 발명에 의한 객체 탐지 모델과 다른 최신 탐지 모델의 총과 칼을 탐지하는 실험에서의 평균 정밀도를 도 12에 정리하였다. 총과 칼의 두 가지 비정상 객체를 탐지하는 실험에서 본 발명에 의한 KB-AF-E-YM 모델은 평균 정밀도(mAP: mean Average Precision)가 96.54%에 달하였으며, 다른 최신 탐지 모델에 비하여 우수한 성능을 보였다.
- [0062] 또한, 본 발명에 의한 KB-AF-E-YM 모델을 3가지 기존 알고리즘 즉, 가우시안 혼합 모델(GMM: Gaussian Mixture Model)[R1], LOB-STERBGS[R2], 배경 차감법(Background subtraction method)[R3] 알고리즘과 Precision, Recall, F1, Similarity, PCC를 비교하였다. 비교 결과는 도 13의 표에 정리되어 있고, 도 14의 그래프에 도시되어 있다. 모든 척도에 있어서, 본 발명에 의한 KB-AF-E-YM 모델이 전체적으로 뛰어난 것을 확인할 수 있었다.
- [0063] 도 13 및 도 14에서, 정밀도(Precision)은 모델이 True라고 분류한 것 중에서 실제 True인 것의 비율을 말한다. 재현율(Recall)은 실제 True인 것 중에서 모델이 True라고 예측한 것의 비율을 말한다. F1은 정밀도(Precision)와 재현율(Recall)의 가중 평균이다. 유사도(Similarity)는 전체 객체와 비정상 객체라고 예측된 것의 합산치 중에서 실제 비정상 객체의 비율을 말한다(즉, $\text{TruePositive}/(\text{TruePositive}+\text{FalseNegative}+\text{FalsePositive})$).
- [0064] 한편, 요약 영상의 정밀도에 대해서도 다른 기법과 비교를 행하였다. 일반적으로, 영상 요약은 조명 조건의 변화, 혼잡 시나리오와 같은 다양한 환경 속성에 따라 결과가 달라질 수 있는데, 본 비교에서는 유용하지 않은 정보를 버리는 능력을 기준으로 정확도를 비교하였다. 그리고, 동일한 평가 척도를 평가에 사용하였다. 다른 기법과의 포괄적인 비교가 도 14에 도시되어 있다. 본 발명의 시놉시스 생성 방법이 다른 기법에 비하여 성능이 뒤지지는 않았지만, 비정상 프레임에 대한 요약만 생성하기 때문에 정밀도에 현저하게 큰 차이는 없었다.
- [0065] 마지막으로, 최종적으로 생성된 시놉시스에 대하여 본 발명의 방법과 다른 알고리즘을 비교하였다.
- [0066] 실제 실내 및 실외 위협 시나리오를 고려하여, 각각이 3500프레임 이상의 길이를 가지는 총 6개의 온라인 및 오프라인 비디오를 선택하였다. 본 발명의 방법을 Time Equidistant Algorithm(TEA)[T1], Constant Equidistant Algorithm(CEA)[T2], SeDiM[T3] 등의 알고리즘과 비교하였다. 도 15는 본 발명에 의한 영상 시놉시스 생성 방법과 다른 알고리즘들의 영상 시놉시스 생성 결과를 대비하여 정리한 표이다. 실험 결과, CEA가 TEA 알고리즘에 비해 성능이 더 우수함을 확인하였는데, 이 두가지 알고리즘 모두 SeDiM에 비해 결과 영상 시놉시스에서 왜곡이 발견되었다. 그리고 본 발명에 의한 방법이 T1, T2 및 T3 방법보다 성능이 비교적 우수하였다.
- [0067] 대비에 사용한 기존 알고리즘이 단일 카메라 뷰에 대한 시놉시스를 생성하는데 비하여, 본 발명의 방법은 멀티뷰 카메라에 대한 결합 영상 시놉시스를 생성하는 것임에도 불구하고, 본 발명의 방법은 최신 시놉시스 생성 기법에 비하여 Recall, F1, Similarity에서 뛰어난 성능을 가진 것을 확인할 수 있었고 영상 요약 시간 또한 다른 방법에 비해 짧아 더욱 빨리 영상을 분석할 수 있게 만들어 주었다.
- [0068] 위에서 언급한 바와 같이 본 발명의 실시예에 따른 장치와 방법은 컴퓨터로 읽을 수 있는 기록매체에 저장된 상태에서 컴퓨터가 읽어서 실행할 수 있는 프로그램 명령들 또는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산 방식으로 컴퓨터로 읽을 수 있는 프로그램 또는 코드가 저장되고 실행될 수 있다.
- [0069] 상기 컴퓨터가 읽을 수 있는 기록매체는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 프로그램 명령은 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.
- [0070] 본 발명의 일부 측면들은 장치의 문맥에서 설명되었으나, 그것은 상응하는 방법에 따른 설명 또한 나타낼 수 있고, 여기서 블록 또는 장치는 방법 단계 또는 방법 단계의 특징에 상응한다. 유사하게, 방법의 문맥에서 설명된 측면들은 또한 상응하는 블록 또는 아이템 또는 상응하는 장치의 특징으로 나타낼 수 있다. 방법 단계들의 몇몇 또는 전부는 예를 들어, 마이크로프로세서, 프로그램 가능한 컴퓨터 또는 전자 회로와 같은 하드웨어 장치에 의해서 또는 이를 이용하여 수행될 수 있다. 몇몇의 실시예에서, 가장 중요한 방법 단계들의 하나 이상은 이와 같은 장치에 의해 수행될 수 있다.
- [0071] 실시예들에서, 프로그램 가능한 로직 장치(예를 들어, 필드 프로그래머블 게이트 어레이)가 여기서 설명된 방법들의 기능의 일부 또는 전부를 수행하기 위해 사용될 수 있다. 실시예들에서, 필드 프로그래머블 게이트 어레이

는 여기서 설명된 방법들 중 하나를 수행하기 위한 마이크로프로세서와 함께 작동할 수 있다. 일반적으로, 방법들은 어떤 하드웨어 장치에 의해 수행되는 것이 바람직하다.

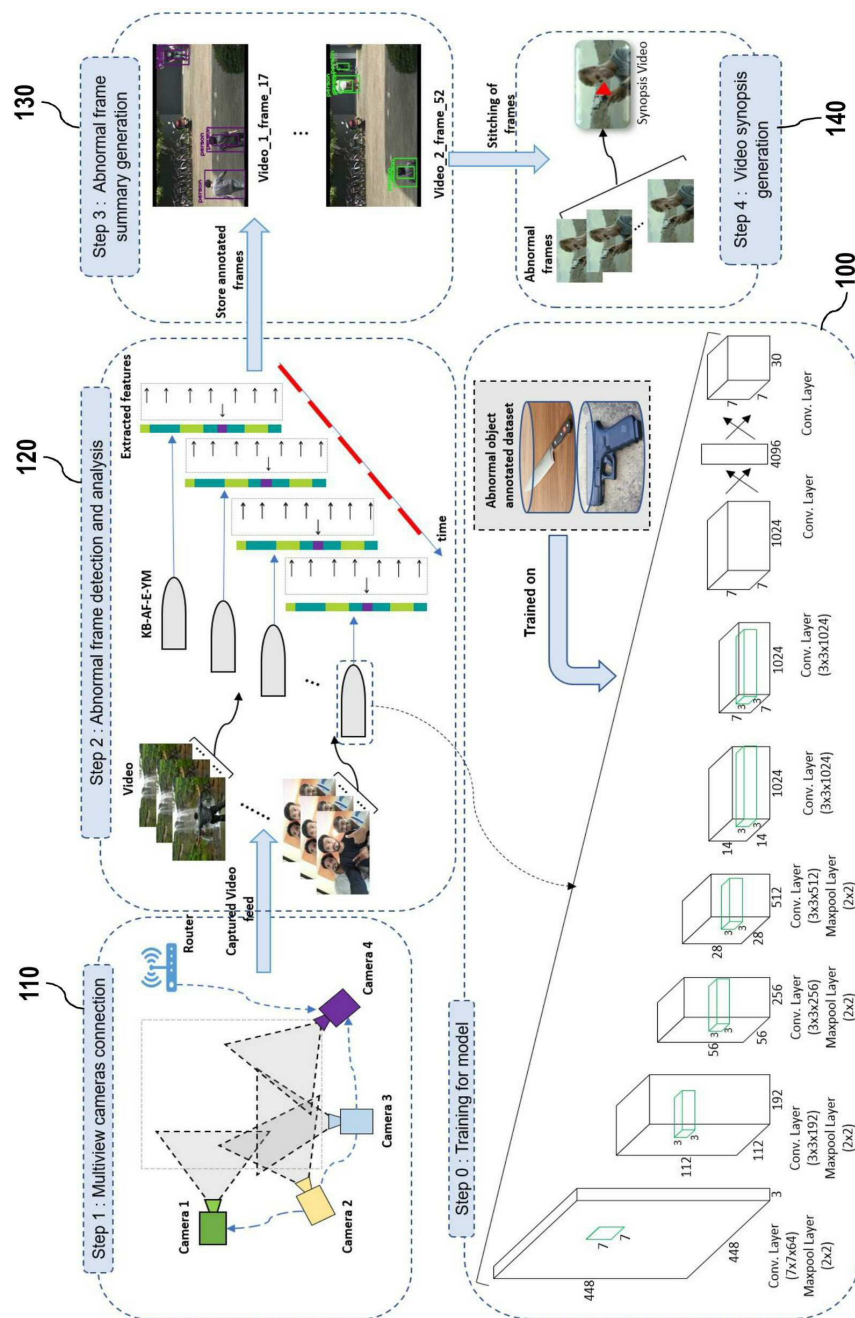
[0072] 위에서는 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

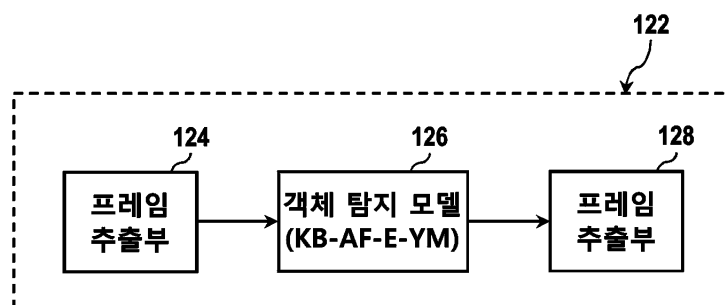
도면1



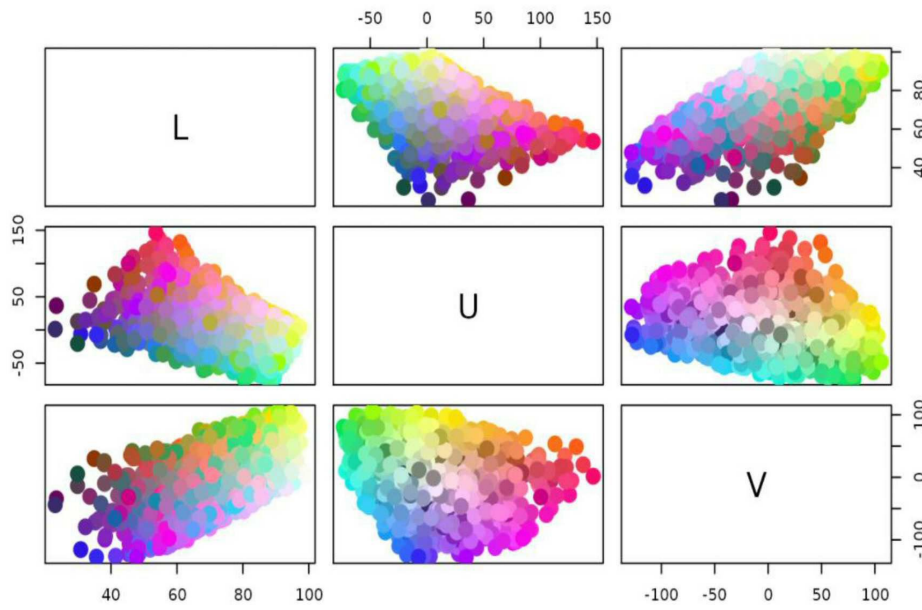
도면2



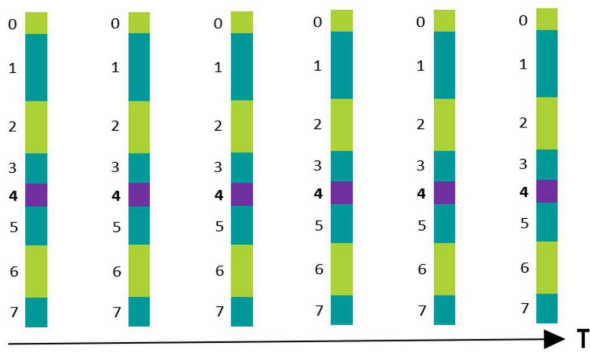
도면3



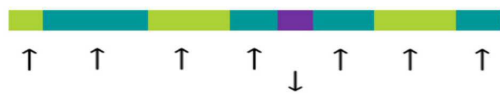
도면4



도면5

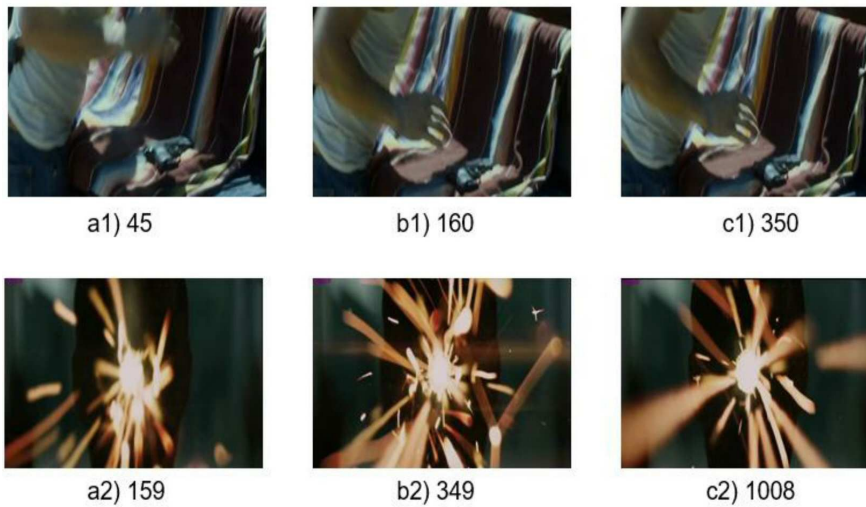


(a) 다수의 입력 영상에 대한 프레임 시퀀스



(b) 비정상 프레임 표시

도면6



도면7

Algorithm1

Input: The whole video that consists of many shots from serval cameras

Output: Video frame summary

Initialization:

1: Train Keras based yolo model as an auto feature extractor

2: Set auto feature extractor to extract only ANF frame

3: Set matrix **MA** to store ANF feature vectors shot

4: While (video frames) do

 /* Input video frame into KB-AF-E-YM model to generate a
 feature vector* /

5: If (input video == n):

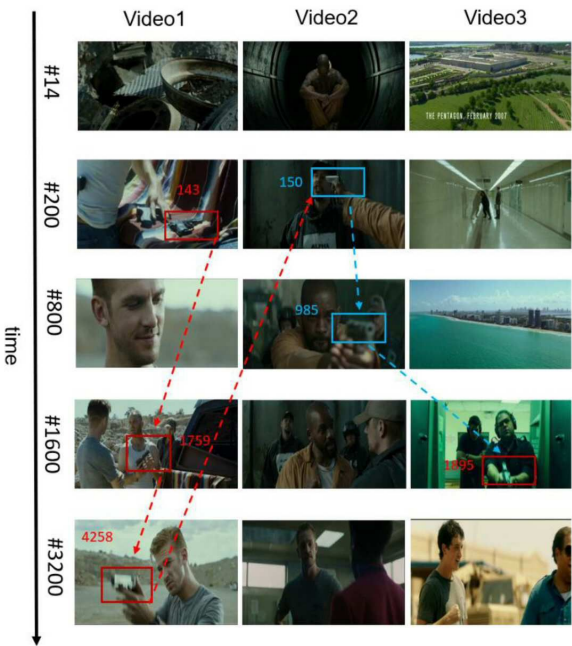
 /* Generate attention model to get key abnormal
 frames for the first input video sequence */

 /* Generate attention model to get key
 abnormal frames for the n+1 video sequence */

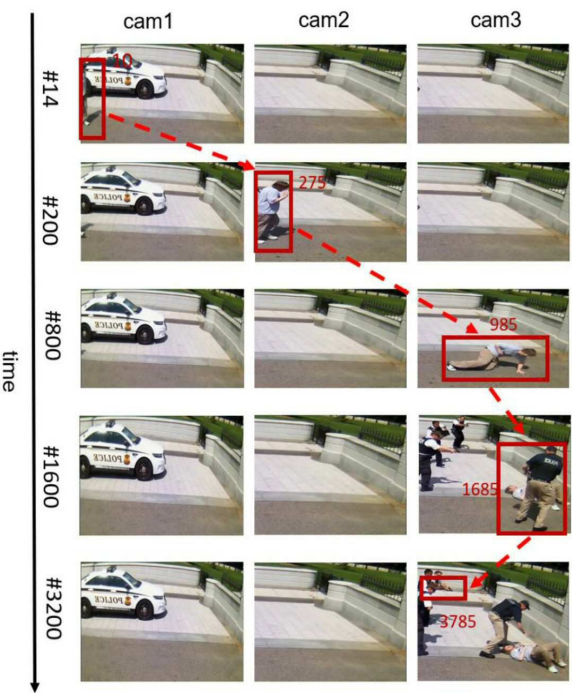
6: end if

7: end while

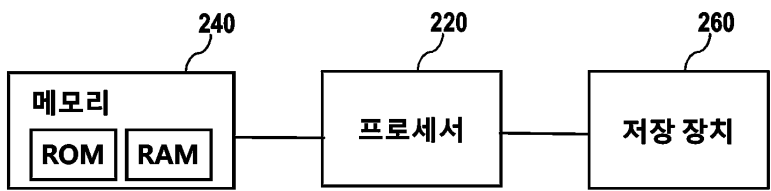
도면8



도면9



도면10



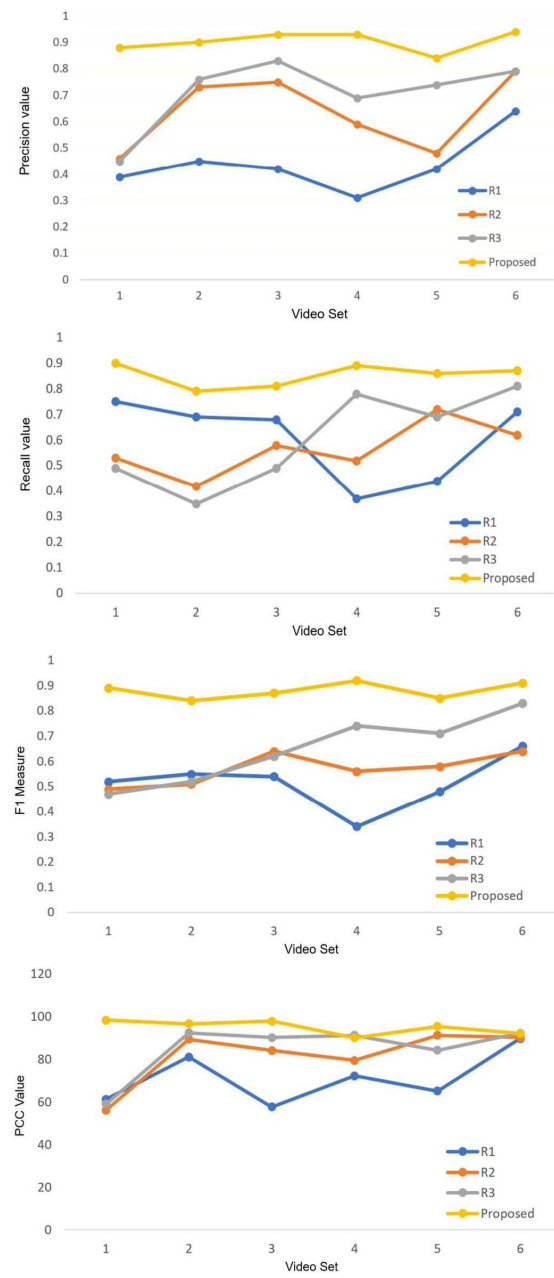
도면11

Video Number	Video Length (#Frames)	Frame Rate (fps)	Abnormal Objects	Number of Objects
1	4825	25	2	4
2	2904	24	3	22
3	5750	23	1	7
4	3168	24	1	256
5	9325	25	1	325
6	4485	23	2	24

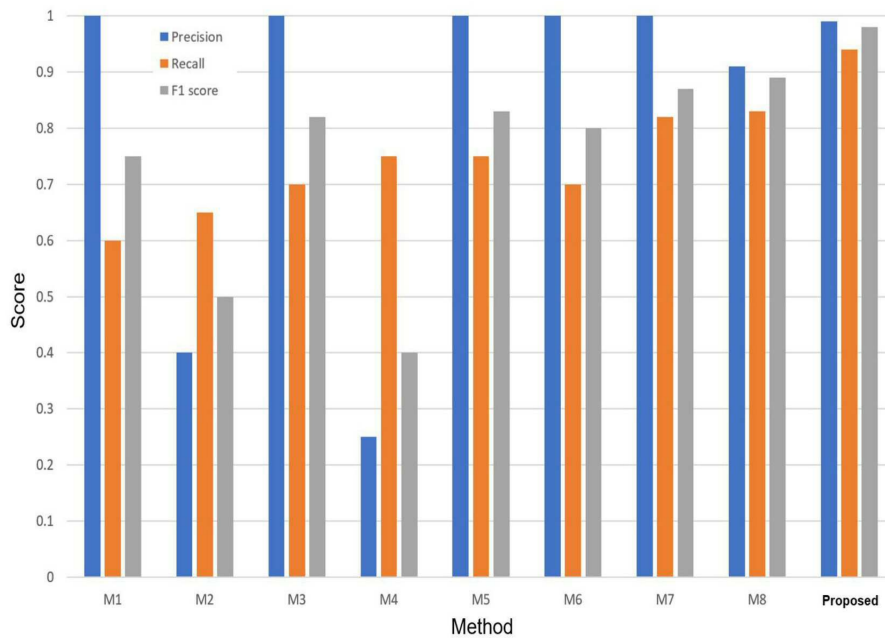
도면12

	Algorithm	Precision	Recall	F1	Similarity	PCC
Video#1	[R1]	0.39	0.75	0.52	0.54	61.42
	[R2]	0.46	0.53	0.49	0.45	56.32
	[R3]	0.45	0.49	0.47	0.45	59.25
	KB-AF-E-YM	0.88	0.90	0.89	0.80	98.32
Video#2	[R1]	0.45	0.69	0.55	0.39	81.15
	[R2]	0.73	0.42	0.51	0.43	89.35
	[R3]	0.76	0.35	0.52	0.45	92.45
	KB-AF-E-YM	0.90	0.79	0.84	0.92	96.65
Video#3	[R1]	0.42	0.68	0.54	0.41	57.85
	[R2]	0.75	0.58	0.64	0.53	84.14
	[R3]	0.83	0.49	0.62	0.43	90.23
	KB-AF-E-YM	0.93	0.81	0.87	0.84	97.85
Video#4	[R1]	0.31	0.37	0.34	0.24	72.38
	[R2]	0.59	0.52	0.56	0.41	79.48
	[R3]	0.69	0.78	0.74	0.58	91.24
	KB-AF-E-YM	0.95	0.89	0.92	0.90	90.14
Video#5	[R1]	0.42	0.44	0.48	0.39	65.21
	[R2]	0.48	0.72	0.58	0.42	91.51
	[R3]	0.74	0.69	0.71	0.63	84.31
	KB-AF-E-YM	0.84	0.86	0.85	0.76	95.45
Video#6	[R1]	0.64	0.71	0.66	0.54	89.75
	[R2]	0.71	0.62	0.64	0.52	90.21
	[R3]	0.79	0.81	0.83	0.67	92.57
	KB-AF-E-YM	0.94	0.87	0.91	0.87	92.12

도면13



도면14



도면15

	Techniques	Origin Video	Recall	F1	Similarity	Synopsis Video
Video#1	TEA [T1]	3.22 min	0.60	0.53	0.44	2.15 min
	CEA [T2]		0.62	0.59	0.45	1.21 min
	SeDiM [T3]		0.79	0.67	0.55	1.03 min
	Proposed		0.90	0.89	0.80	0.47 sec
Video#2	TEA [T1]	2.3 min	0.53	0.56	0.47	1.45 min
	CEA [T2]		0.62	0.58	0.60	1.05 min
	SeDiM [T3]		0.67	0.61	0.64	0.55 sec
	Proposed		0.79	0.84	0.92	0.25 sec
Video#3	TEA [T1]	4.18 min	0.54	0.47	0.35	3.03 min
	CEA [T2]		0.56	0.53	0.44	2.35 min
	SeDiM [T3]		0.64	0.73	0.63	1.56 min
	Proposed		0.81	0.87	0.84	0.58 sec
Video#4	TEA [T1]	2.20 min	0.47	0.54	0.36	1.15 min
	CEA [T2]		0.57	0.56	0.51	0.58 sec
	SeDiM [T3]		0.69	0.84	0.63	0.32 sec
	Proposed		0.89	0.92	0.90	0.27 sec
Video#5	TEA [T1]	6.23 min	0.53	0.42	0.38	4.57 min
	CEA [T2]		0.51	0.54	0.32	4.02 min
	SeDiM [T3]		0.68	0.69	0.65	3.55 min
	Proposed		0.86	0.85	0.76	3.41 sec
Video#6	TEA [T1]	3.26 min	0.49	0.61	0.54	2.35 min
	CEA [T2]		0.47	0.52	0.64	1.48 min
	SeDiM [T3]		0.74	0.85	0.73	1.13 min
	Proposed		0.87	0.91	0.87	0.15 sec