



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0067159  
(43) 공개일자 2022년05월24일

(51) 국제특허분류(Int. Cl.)  
G06K 9/00 (2022.01) G06N 3/04 (2006.01)  
G06V 10/40 (2022.01)  
(52) CPC특허분류  
G06V 20/52 (2022.01)  
G06N 3/0454 (2013.01)  
(21) 출원번호 10-2020-0153560  
(22) 출원일자 2020년11월17일  
심사청구일자 2020년11월17일

(71) 출원인  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)  
(72) 발명자  
이상윤  
서울특별시 서초구 청계산로7길 43, 504동 1302호(신원동, 서초포레스타5단지)  
조명아  
경기도 성남시 분당구 판교로 421, 204동 403호(야탑동, 탑마을대우아파트)  
(74) 대리인  
민영준

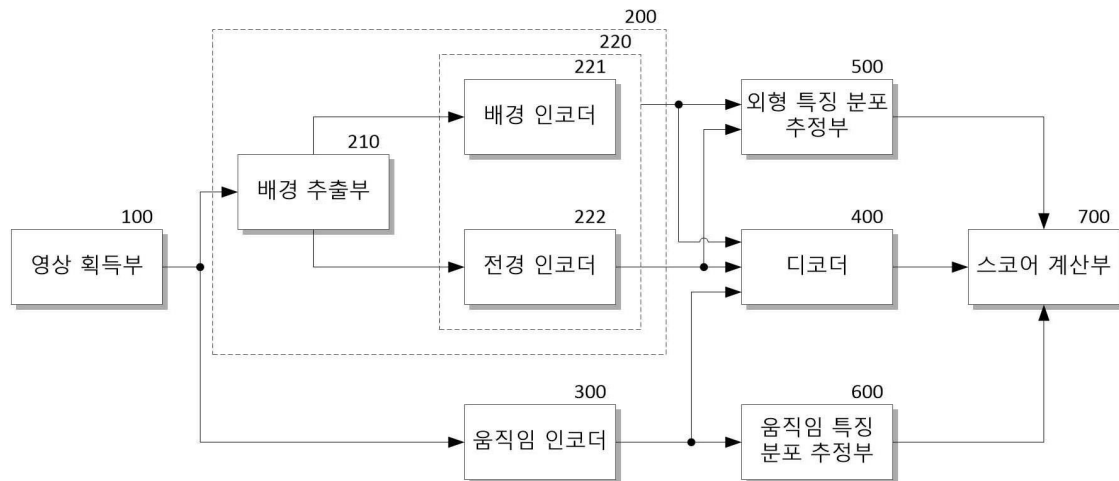
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 이상 상황 감지 장치 및 방법

(57) 요약

본 발명은 미리 학습된 다수의 인공 신경망을 포함하여, 인가된 입력 영상의 다수의 프레임으로부터 움직임 특징을 추출하고, 각 프레임의 전경 및 배경 각각에 대한 전경 특징과 배경 특징 구분하여 추출하는 인코더, 미리 학습된 인공 신경망으로 구현되어 상기 움직임 특징과 상기 전경 특징 및 상기 배경 특징을 기반으로 상기 입력 영상을 재건축하여 재건축 영상을 획득하는 디코더 및 상기 입력 영상과 상기 재건축 영상 사이의 오차를 기반으로 재건축 스코어를 계산하고, 계산된 재건축 스코어를 이용하여 이상 상황을 검출하는 스코어 계산부를 포함하여, 이상 상황을 정확하게 검출할 수 있으며, 비지도 학습으로 학습되므로, 학습 데이터 획득이 용이할 뿐만 아니라 다양한 이상 상황을 검출할 수 있는 이상 상황 감지 장치 및 방법을 제공할 수 있다.

대표도



(52) CPC특허분류

*G06N 3/08* (2013.01)  
*G06T 7/194* (2017.01)  
*G06V 10/40* (2022.01)

**조수환**

서울특별시 양천구 목동동로 350, 510동 304호(목동, 목동신시가자아파트5단지)

(72) 발명자

**정태영**

서울특별시 서대문구 연희로10길 24-10, 101호(연희동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711119679
과제번호	2018M3E3A1057289
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	복합인지기술개발사업
연구과제명	이종 CCTV 영상에서의 딥러닝 기반 실종자 초동 신원확인 및 추적 시스템 (2
단계)(1/3)	
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2020.04.23 ~ 2020.12.31

---

## 명세서

### 청구범위

#### 청구항 1

미리 학습된 다수의 인공 신경망을 포함하여, 인가된 입력 영상의 다수의 프레임으로부터 움직임 특징을 추출하고, 각 프레임의 전경 및 배경 각각에 대한 전경 특징과 배경 특징 구분하여 추출하는 인코더;

미리 학습된 인공 신경망으로 구현되어 상기 움직임 특징과 상기 전경 특징 및 상기 배경 특징을 기반으로 상기 입력 영상을 재건축하여 재건축 영상을 획득하는 디코더; 및

상기 입력 영상과 상기 재건축 영상 사이의 오차를 기반으로 재건축 스코어를 계산하고, 계산된 재건축 스코어를 이용하여 이상 상황을 검출하는 스코어 계산부를 포함하는 이상 상황 감지 장치.

#### 청구항 2

제1항에 있어서, 상기 인코더는

미리 학습된 인공 신경망으로 구현되어 학습된 방식에 따라 인가된 상기 입력 영상을 인코딩하여 다수의 프레임 사이의 변화에 따른 움직임 특징을 추출하는 움직임 인코더; 및

미리 학습된 다수의 인공 신경망을 포함하여, 상기 입력 영상의 각 프레임에서 전경과 배경을 구분하고, 구분된 전경과 배경 각각을 학습된 방식으로 인코딩하여 전경 특징과 배경 특징을 추출하는 영상 인코더를 포함하는 이상 상황 감지 장치.

#### 청구항 3

제2항에 있어서, 상기 영상 인코더는

상기 입력 영상을 인가받고, 기지정된 방식에 따라 각 프레임에서 전경과 배경을 구분하여 추출하는 배경 추출부;

미리 학습된 인공 신경망으로 구현되어, 추출된 전경을 학습된 방식으로 인코딩하여 전경 특징을 획득하는 전경 인코더; 및

미리 학습된 인공 신경망으로 구현되어, 추출된 배경을 학습된 방식으로 인코딩하여 배경 특징을 획득하는 배경 인코더를 포함하는 이상 상황 감지 장치.

#### 청구항 4

제3항에 있어서, 상기 배경 추출부는

상기 입력 영상의 다수의 프레임에서 기지정된 간격 단위로 프레임을 선택하고 선택된 프레임의 전경과 배경을 구분하여 추출하는 이상 상황 감지 장치.

#### 청구항 5

제3항에 있어서, 상기 이상 상황 감지 장치는

학습 시에 학습 영상으로 정상 상황에 대한 입력 영상(I)을 인가받아 재건축 영상( $\hat{I}$ )을 획득하고, 각각 수학적

$$L_2 = \left\| I - \hat{I} \right\|_2$$

으로 계산되는  $L_2$  손실( $L_2$ )과

수학식

$$L_{grd} = \sum_{i,j} \left\| |I_{i,j} - I_{i-1,j}| - |\hat{I}_{i,j} - \hat{I}_{i-1,j}| \right\|_1 + \sum_{i,j} \left\| |I_{i,j} - I_{i,j-1}| - |\hat{I}_{i,j} - \hat{I}_{i,j-1}| \right\|_1$$

(여기서  $i, j$ 는 각각 입력 영상( $I$ )과 재건축 영상( $\hat{I}$ )의 픽셀 좌표를 나타낸다.)

으로 계산되는 그래디언트 손실( $L_{grd}$ )의 합으로 나타나는 재건축 손실( $L_{recon}$ )이 역전파됨으로써 상기 인코더와 디코더가 학습되는 이상 상황 감지 장치.

#### 청구항 6

제3항에 있어서, 상기 이상 상황 감지 장치는

미리 학습된 흐름 기반 생성 모델로 구현되어, 상기 전경 특징과 배경 특징이 결합된 외형 특징에 대한 특징 분포를 네거티브 로그 가능도(negative log likelihood: 이하 NLL) 형식의 가능도로 추정하는 외형 특징 분포 추정부; 및

미리 학습된 흐름 기반 생성 모델로 구현되어, 상기 움직임 특징의 특징 분포를 NLL 형식의 가능도로 추정하는 움직임 특징 분포 추정부를 더 포함하는 이상 상황 감지 장치.

#### 청구항 7

제6항에 있어서, 상기 이상 상황 감지 장치는

학습 시에 학습 영상으로 정상 상황에 대한 입력 영상( $I$ )을 인가받아 상기 외형 특징 분포 추정부와 상기 움직임 특징 분포 추정부 각각에서 추정된 가능도로부터 수학식

$$L(x_t^a, x_t^m) = \frac{\lambda_1 NLL(x_t^a) + \lambda_2 NLL(x_t^m)}{\lambda_1 + \lambda_2}$$

$$\lambda_1 = \max(x^a) - \min(x^a), \lambda_2 = \max(x^m) - \min(x^m)$$

(여기서  $x_t^a$  와  $x_t^m$  는 각각 입력 영상( $I$ )의  $t$ 번째 프레임( $I_t$ )의 외형 특징과 움직임 특징을 나타내고,  $NLL()$ 은 네거티브 로그 가능도(negative log likelihood) 함수이며,  $\lambda_1$  과  $\lambda_2$  는 각각 외형 특징 분포 추정부(500)의 손실인 네거티브 로그 가능도( $NLL(x_t^a)$ )와 움직임 특징 분포 추정부(600)의 손실인 네거티브 로그 가능도( $NLL(x_t^m)$ )를 정규화하기 위한 정규화 파라미터이다.)

으로 계산된 특징 분포 손실( $L(x_t^a, x_t^m)$ )이 역전파됨으로써 상기 외형 특징 분포 추정부와 상기 특징 분포 추정부가 학습되는 이상 상황 감지 장치.

#### 청구항 8

제7항에 있어서, 상기 스코어 계산부는

입력 영상에 대한 이상 상황 검출 시에 입력 영상( $I$ )의  $t$ 번째 프레임( $I_t$ )과 재건축 영상( $\hat{I}$ )의  $t$ 번째 프레임( $\hat{I}_t$ )으로부터 상기 재건축 스코어를

수학식

$$R(I_t, \hat{I}_t) = -\log_{10} (1 - |I_t - \hat{I}_t|)$$

에 따라 계산하여 획득하는 이상 상황 감지 장치.

#### 청구항 9

제8항에 있어서, 상기 스코어 계산부는

상기 특징 분포 손실( $L(x_t^a, x_t^m)$ )을 NLL 스코어로 획득하며, 획득된 상기 재건축 스코어와 상기 NLL 스코어를 기지정된 방식으로 정규화하고 가산하여 총 스코어를 획득하며, 획득된 총 스코어가 기지정된 문턱값 이상인 프레임은 이상 상황 프레임으로 검출하는 이상 상황 감지 장치.

#### 청구항 10

미리 학습된 다수의 인공 신경망을 이용하여 인가된 입력 영상의 다수의 프레임으로부터 움직임 특징을 추출하고, 각 프레임의 전경 및 배경 각각에 대한 전경 특징과 배경 특징 구분하여 추출하는 단계;

미리 학습된 인공 신경망으로 이용하여 상기 움직임 특징과 상기 전경 특징 및 상기 배경 특징을 기반으로 상기 입력 영상을 재건축하여 재건축 영상을 획득하는 단계; 및

상기 입력 영상과 상기 재건축 영상 사이의 오차를 기반으로 재건축 스코어를 계산하고, 계산된 재건축 스코어를 이용하여 이상 상황을 검출하는 단계를 포함하는 이상 상황 감지 방법.

#### 청구항 11

제10항에 있어서, 상기 구분하여 추출하는 단계는

미리 학습된 인공 신경망을 이용하여, 학습된 방식에 따라 인가된 상기 입력 영상을 인코딩하여 다수의 프레임 사이의 변화에 따른 움직임 특징을 추출하는 단계;

상기 입력 영상의 각 프레임에서 전경과 배경을 구분하여 추출하는 단계; 및

미리 학습된 다수의 인공 신경망을 이용하여 구분된 전경과 배경 각각을 학습된 방식으로 인코딩하여 전경 특징과 배경 특징을 추출하는 단계를 포함하는 이상 상황 감지 방법.

#### 청구항 12

제11항에 있어서, 상기 배경 특징을 추출하는 단계는

미리 학습된 인공 신경망을 이용하여 추출된 전경을 학습된 방식으로 인코딩하여 전경 특징을 획득하는 단계; 및

미리 학습된 인공 신경망을 이용하여 추출된 배경을 학습된 방식으로 인코딩하여 배경 특징을 획득하는 단계를 포함하는 이상 상황 감지 방법.

#### 청구항 13

제12항에 있어서, 상기 전경과 배경을 구분하여 추출하는 단계는

상기 입력 영상의 다수의 프레임에서 기지정된 간격 단위로 프레임을 선택하고 선택된 프레임의 전경과 배경을 구분하여 추출하는 이상 상황 감지 방법.

#### 청구항 14

제12항에 있어서, 상기 이상 상황 감지 방법은

학습 단계를 더 포함하고,

상기 학습 단계는 학습 영상으로 정상 상황에 대한 입력 영상(I)을 인가받아 재건축 영상( $\hat{I}$ )을 획득하고, 각각

수학식

$$L_2 = \left\| \mathbf{I} - \hat{\mathbf{I}} \right\|_2$$

으로 계산되는  $L_2$  손실( $L_2$ )과

수학식

$$L_{grd} = \sum_{i,j} \left\| |I_{i,j} - I_{i-1,j}| - |\hat{I}_{i,j} - \hat{I}_{i-1,j}| \right\|_1 + \sum_{i,j} \left\| |I_{i,j} - I_{i,j-1}| - |\hat{I}_{i,j} - \hat{I}_{i,j-1}| \right\|_1$$

(여기서  $i, j$ 는 각각 입력 영상( $I$ )과 재건축 영상( $\hat{I}$ )의 픽셀 좌표를 나타낸다.)

으로 계산되는 그래디언트 손실( $L_{grd}$ )의 합으로 나타나는 재건축 손실( $L_{recon}$ )이 역전파하는 단계를 포함하는 이상 상황 감지 방법.

#### 청구항 15

제14항에 있어서, 상기 이상 상황 감지 방법은

미리 학습된 흐름 기반 생성 모델을 이용하여 상기 전경 특징과 배경 특징이 결합된 외형 특징에 대한 특징 분포를 네거티브 로그 가능도(negative log likelihood: 이하 NLL) 형식의 가능도로 추정하는 단계; 및

미리 학습된 흐름 기반 생성 모델로 이용하여 상기 움직임 특징의 특징 분포를 NLL 형식의 가능도로 추정하는 단계를 더 포함하는 이상 상황 감지 방법.

#### 청구항 16

제15항에 있어서, 상기 학습 단계는

NLL 형식으로 추정된 외형 특징 분포와 움직임 특징 분포로부터 수학식

$$L(\mathbf{x}_t^a, \mathbf{x}_t^m) = \frac{\lambda_1 NLL(\mathbf{x}_t^a) + \lambda_2 NLL(\mathbf{x}_t^m)}{\lambda_1 + \lambda_2}$$

$$\lambda_1 = \max(\mathbf{x}^a) - \min(\mathbf{x}^a), \lambda_2 = \max(\mathbf{x}^m) - \min(\mathbf{x}^m)$$

(여기서  $\mathbf{x}_t^a$  와  $\mathbf{x}_t^m$  는 각각 입력 영상( $I$ )의  $t$ 번째 프레임( $I_t$ )의 외형 특징과 움직임 특징을 나타내고,  $NLL()$ 은 네거티브 로그 가능도(negative log likelihood) 함수이며,  $\lambda_1$  과  $\lambda_2$  는 각각 외형 특징 분포 추정부(500)의 손실인 네거티브 로그 가능도( $NLL(\mathbf{x}_t^a)$ )와 움직임 특징 분포 추정부(600)의 손실인 네거티브 로그 가능도( $NLL(\mathbf{x}_t^m)$ )를 정규화하기 위한 정규화 파라미터이다.)

으로 계산된 특징 분포 손실( $L(\mathbf{x}_t^a, \mathbf{x}_t^m)$ )을 역전파하는 단계를 더 포함하는 이상 상황 감지 방법.

#### 청구항 17

제16항에 있어서, 상기 이상 상황을 검출하는 단계는

입력 영상에 대한 이상 상황 검출 시에 입력 영상( $I$ )의  $t$ 번째 프레임( $I_t$ )과 재건축 영상( $\hat{I}$ )의  $t$ 번째 프레임( $\hat{I}_t$ )으로부터 상기 재건축 스코어를

수학식

$$R(I_t, \hat{I}_t) = -\log_{10} (1 - |I_t - \hat{I}_t|)$$

에 따라 계산하여 획득하는 이상 상황 감지 방법.

## 청구항 18

제17항에 있어서, 상기 이상 상황을 검출하는 단계는

상기 특정 분포 손실( $L(x_t^a, x_t^m)$ )을 NLL 스코어로 획득하는 단계;

획득된 상기 재건축 스코어와 상기 NLL 스코어를 기지정된 방식으로 정규화하고 가산하여 총 스코어를 획득하는 단계; 및

획득된 총 스코어가 기지정된 문턱값 이상인 프레임을 이상 상황 프레임으로 검출하는 단계를 더 포함하는 이상 상황 감지 방법.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 이상 상황 감지 장치 및 방법에 관한 것으로, 영상에 포함된 객체의 외형과 움직임에 기반한 이상 상황 감지 장치 및 방법에 관한 것이다.

### 배경 기술

[0002] 최근 감시 카메라(CCTV)는 은행, 도로, 건물 등 일상 속 대부분의 장소에 존재하며, 공공 장소의 상황을 기록하고 범죄나 사고 현장에서 중요한 역할을 한다. 그러나 인간이 매 순간을 모니터링하거나 수많은 감시 비디오에서 사건, 사고를 찾아내는 것은 비효율적이다. 이러한 현실을 반영하여 정상 상황(normal) 감시비디오 중에서 차 사고, 폭발과 같은 이상 상황(anomalous event)을 감지하는 연구가 딥러닝 분야에서 활발히 진행 중이다.

[0003] 도 1은 이상 상황 감지 장치의 개념을 설명하기 위한 도면이다.

[0004] 도 1에 도시된 바와 같이, 이상 상황 감지 장치는 다수의 프레임으로 구성된 입력 영상에서 각 프레임에 폭발 사고, 폭력 사고, 차량 추돌 사고 및 강도 사고와 같은 다양한 이상 상황이 포함되어 있는지 여부를 판별하여 이상 상황이 포함된 프레임을 검출해야 한다.

[0005] 그러나 이상 상황은 다양한 형태로 나타날 수 있다. 도 1에서는 일 예로 4개의 이상 상황만을 도시하였으나, 이상 상황은 무한하게 많은 다양한 형태로 나타날 수 있다. 따라서 모든 이상 상황을 정의 내리고 분류하는 것은 매우 어려우며, 현실적으로 모든 종류의 존재하는 이상 상황 데이터를 수집하는 것 또한 불가능하다.

[0006] 따라서 이상 상황 감지 장치에 딥러닝 기반 지도 학습(supervised-learning) 방법을 적용하기에 어려움이 존재한다.

[0007] 이상 상황 검출을 위해 딥러닝 기반 지도 학습을 수행하기 위해서는 정상 상황과 이상 상황이 구분되어 레이블된 대량의 학습 데이터를 이용하여 인공 신경망으로 구현되는 이상 상황 감지 장치가 정상 상황과 이상 상황을 정확하게 분류할 수 있도록 학습시키야 한다. 그러나 이상 상황은 정상 상황에 비해 매우 드물게 발생하므로, 정상 상황에 대한 학습 데이터와 이상 상황에 대한 학습 데이터를 획득할 수 있는 데이터량의 차이가 매우 커서, 데이터 불균형(class-imbalance) 문제가 발생한다. 이는 이상 상황을 학습시키기 어렵게 할 뿐만 아니라 오버피팅(overfitting) 문제가 쉽게 발생된다는 문제가 있다.

[0008] 또한 이상 상황에 대해 지도 학습을 수행할 지라도 상기한 바와 같이, 이상 상황은 무한히 다양한 형태로 발생할 수 있으므로, 다양한 이상 상황 각각을 개별적으로 학습시킬 수 없다는 한계가 있다.

[0009] 이에 최근에는 방대한 양의 학습 데이터에 대한 레이블링을 하지 않고 학습하는 비지도 학습 기반 이상 상황 검출 기술 연구가 활발하게 연구되고 있다. 비지도 학습 기반 이상 상황 검출 기법 중 대표적인 예는 오토인코더를 이용하는 재건축(reconstruction) 기법(또는 프레임 예측(frame prediction) 기법이라고도 함)이 있다. 오

토큰코더를 이용하는 재건축 기법에서는 이용하여 입력된 영상의 프레임에서 특징을 추출한 후, 추출된 특징으로부터 다시 프레임을 재건축하는 오토인코더를 정상 상황에 대한 학습 데이터를 이용하여 학습시킨다. 그리고 정상 상황을 기반으로 학습된 오토인코더가 재건축한 프레임과 원본 프레임 사이의 차이를 비교하여 이상 상황을 검출하도록 한다. 정상 상황이 학습된 오토 프레임의 경우, 이상 상황을 재건축하기 어려워 오차가 크게 발생한다는 개념에 기초한 기법이다.

- [0010] 도 2는 비지도 학습에 기반하는 오토인코더를 이용하는 이상 상황 검출 장치의 이상 상황 검출 결과의 일 예를 나타낸다.
- [0011] 도 2에서 (a)는 정상 상황 프레임과 이상 상황이 검출된 프레임에 따른 오차 스코어의 변화를 나타내고, (b)는 이상 상황 프레임과 재건축된 프레임 및 이상 상황 프레임과 재건축된 프레임 사이의 오차가 반영된 오차맵의 일 예를 나타낸다.
- [0012] 도 2에서 이상 상황 검출 장치는 보행자 도로에 보행자가 걸어다니는 정상 상황 데이터로 학습된 오토인코더가 보행자 도로에 차량이 등장하는 이상 상황 프레임이 입력됨에 따라 차량을 재건축하지 못하여 큰 오차가 발생하는 일 예를 나타낸다. 이와 같이, 비지도 학습 방식으로 학습된 오토인코더를 이용하는 재건축 기법은 재건축 가능한 정상 상황 이외의 모든 상황을 이상 상황으로 검출할 수 있어, 효과적인 이상 상황 검출 장치로 이용되고 있다.
- [0013] 다만 이상 상황은 영상에서 이상 객체 출현과 같은 객체의 외형으로 구분되는 외형적 이상 상황뿐만 아니라, 객체의 행동으로 구분되는 움직임 이상 상황도 발생하게 된다.
- [0014] 도 3은 기존의 이상 상황 검출 장치가 이상 상황으로 검출하지 못하는 경우의 일 예를 나타낸다.
- [0015] 도 3에서 (a)는 보행자 도로에 자전거 탑승자가 나타난 경우를 도시하였고, (b)는 이상 행동을 하는 어린이가 나타난 경우를 도시하였다. 그리고 (a) 및 (b)에서 우측 도면은 이상 상황 검출 장치가 검출한 오차맵을 나타낸다.
- [0016] 도 3에서 (a) 및 (b)의 우측 도면을 살펴보면, 기존에 객체의 외형을 기준으로 이상 상황을 검출하는 이상 상황 검출 장치는 객체의 움직임에 따른 이상 상황이 발생하였음에도 오차가 크게 나타나지 않아 이상 상황을 검출하지 못함을 알 수 있다.
- [0017] 즉 기존에 오토인코더를 이용하는 이상 상황 검출 장치는 객체의 외형에 대한 정보를 기준으로 학습되어 이상 상황을 검출함에 따라, 객체의 이상 행동에 따른 이상 상황을 검출하기 어렵다는 문제가 있다.

## 선행기술문헌

### 특허문헌

- [0018] (특허문헌 0001) 한국 공개 특허 제10-2020-0056183호 (2020.05.22 공개)

## 발명의 내용

### 해결하려는 과제

- [0019] 본 발명의 목적은 비지도 학습 기법을 이용하여 다양한 이상 상황을 정확하게 검출할 수 있는 이상 상황 감지 장치 및 방법을 제공하는데 있다.
- [0020] 본 발명의 다른 목적은 객체의 외형과 객체의 움직임을 함께 반영하여 이상 상황을 정확하게 검출할 수 있는 이상 상황 감지 장치 및 방법을 제공하는데 있다.

### 과제의 해결 수단

- [0021] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 이상 상황 감지 장치는 미리 학습된 다수의 인공 신경망을 포함하여, 인가된 입력 영상의 다수의 프레임으로부터 움직임 특징을 추출하고, 각 프레임의 전경 및 배경 각각에 대한 전경 특징과 배경 특징 구분하여 추출하는 인코더; 미리 학습된 인공 신경망으로 구현되어 상기 움직임 특징과 상기 전경 특징 및 상기 배경 특징을 기반으로 상기 입력 영상을 재건축하여 재건축 영상을 획득하



는 디코더; 및 상기 입력 영상과 상기 재건축 영상 사이의 오차를 기반으로 재건축 스코어를 계산하고, 계산된 재건축 스코어를 이용하여 이상 상황을 검출하는 스코어 계산부를 포함한다.

[0022] 상기 인코더는 미리 학습된 인공 신경망으로 구현되어 학습된 방식에 따라 인가된 상기 입력 영상을 인코딩하여 다수의 프레임 사이의 변화에 따른 움직임 특징을 추출하는 움직임 인코더; 및 미리 학습된 다수의 인공 신경망을 포함하여, 상기 입력 영상의 각 프레임에서 전경과 배경을 구분하고, 구분된 전경과 배경 각각을 학습된 방식으로 인코딩하여 전경 특징과 배경 특징을 추출하는 영상 인코더를 포함할 수 있다.

[0023] 상기 영상 인코더는 상기 입력 영상을 인가받고, 기지정된 방식에 따라 각 프레임에서 전경과 배경을 구분하여 추출하는 배경 추출부; 미리 학습된 인공 신경망으로 구현되어, 추출된 전경을 학습된 방식으로 인코딩하여 전경 특징을 획득하는 전경 인코더; 및 미리 학습된 인공 신경망으로 구현되어, 추출된 배경을 학습된 방식으로 인코딩하여 배경 특징을 획득하는 배경 인코더를 포함할 수 있다.

[0024] 상기 배경 추출부는 상기 입력 영상의 다수의 프레임에서 기지정된 간격 단위로 프레임을 선택하고 선택된 프레임의 전경과 배경을 구분하여 추출할 수 있다.

[0025] 상기 이상 상황 감지 장치는 미리 학습된 흐름 기반 생성 모델로 구현되어, 상기 전경 특징과 배경 특징이 결합된 외형 특징에 대한 특징 분포를 네거티브 로그 가능도(negative log likelihood: 이하 NLL) 형식의 가능도로 추정하는 외형 특징 분포 추정부; 및 미리 학습된 흐름 기반 생성 모델로 구현되어, 상기 움직임 특징의 특징 분포를 NLL 형식의 가능도로 추정하는 움직임 특징 분포 추정부를 더 포함할 수 있다.

[0026] 상기 스코어 계산부는 상기 특징 분포 손실을 NLL 스코어로 획득하며, 획득된 상기 재건축 스코어와 상기 NLL 스코어를 기지정된 방식으로 정규화하고 가산하여 총 스코어를 획득하며, 획득된 총 스코어가 기지정된 문턱값 이상인 프레임을 이상 상황 프레임으로 검출할 수 있다.

[0027] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 이상 상황 감지 방법은 미리 학습된 다수의 인공 신경망을 이용하여 인가된 입력 영상의 다수의 프레임으로부터 움직임 특징을 추출하고, 각 프레임의 전경 및 배경 각각에 대한 전경 특징과 배경 특징 구분하여 추출하는 단계; 미리 학습된 인공 신경망으로 이용하여 상기 움직임 특징과 상기 전경 특징 및 상기 배경 특징을 기반으로 상기 입력 영상을 재건축하여 재건축 영상을 획득하는 단계; 및 상기 입력 영상과 상기 재건축 영상 사이의 오차를 기반으로 재건축 스코어를 계산하고, 계산된 재건축 스코어를 이용하여 이상 상황을 검출하는 단계를 포함 포함한다.

## 발명의 효과

[0028] 따라서, 본 발명의 실시예에 따른 이상 상황 감지 장치 및 방법은 입력된 영상의 각 프레임에서 전경과 배경을 구분하여 개별적으로 특징을 추출할 뿐만 아니라, 영상의 움직임 특징을 별도로 추출한 후, 추출된 전경, 배경 및 움직임 특징을 기반으로 프레임을 재건축하고, 외형과 움직임 특징을 이용하여 데이터 분포를 추정하고, 추정된 데이터 분포를 기반으로 각 특징들의 가능도를 계산하여 이상 상황을 검출하므로, 이상 상황을 정확하게 검출할 수 있다. 또한 정상 상황에 대한 학습 데이터를 이용하여 비지도 학습으로 학습되므로, 학습 데이터 획득이 용이할 뿐만 아니라 다양한 이상 상황을 검출할 수 있다.

## 도면의 간단한 설명

[0029] 도 1은 이상 상황 감지 장치의 개념을 설명하기 위한 도면이다.

도 2는 비지도 학습에 기반하는 오토인코더를 이용하는 이상 상황 검출 장치의 이상 상황 검출 결과의 일 예를 나타낸다.

도 3은 기존의 이상 상황 검출 장치가 이상 상황으로 검출하지 못하는 경우의 일 예를 나타낸다.

도 4는 본 발명의 일 실시예에 따른 이상 상황 검출 장치의 개략적 구성을 나타낸다.

도 5는 도 4의 이상 상황 검출 장치의 동작을 설명하기 위한 도면이다.

도 6은 본 실시예의 이상 상황 검출 장치의 이상 상황 검출 성능을 기존의 이상 상황 검출 장치와 비교한 결과를 나타낸다.

도 7은 도 4의 스코어 계산부가 재건축 스코어와 가능도 스코어에 따라 총 스코어를 계산하여 이상 프레임을 검출하는 일 예를 나타낸다.

도 8은 본 발명의 일 실시예에 따른 이상 상황 검출 방법을 나타낸다.

### 발명을 실시하기 위한 구체적인 내용

- [0030] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.
- [0031] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.
- [0032] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0033] 도 4는 본 발명의 일 실시예에 따른 이상 상황 검출 장치의 개략적 구성을 나타내고, 도 5는 도 4의 이상 상황 검출 장치의 동작을 설명하기 위한 도면이다.
- [0034] 도 4를 참조하면, 본 실시예에 따른 이상 상황 검출 장치는 영상 획득부(100), 영상 인코더(200), 움직임 인코더(300), 디코더(400), 외형 특징 분포 추정부(500), 움직임 특징 분포 추정부(600) 및 스코어 계산부(700)를 포함할 수 있다.
- [0035] 영상 획득부(100)는 이상 상황을 검출할 입력 영상을 획득한다. 여기서 입력 영상은 시간의 흐름에 따라 연속하는 다수의 프레임으로 구성되는 영상으로서, 감시 카메라(CCTV) 등에서 획득될 수 있다. 그리고 영상 획득부(100)는 직접 영상을 획득하는 카메라 등으로 구현될 수 있으나, 경우에 따라서는 카메라 등을 통해 미리 획득된 영상이 저장되는 저장 장치로 구현되거나, 외부 기기로부터 영상을 인가받는 통신 수단으로 구현될 수도 있다.
- [0036] 영상 인코더(200)는 영상 획득부(100)가 획득한 입력 영상을 인가받아 미리 학습된 방식에 따라 인코딩하여 입력 영상의 외형적 특징을 추출한다. 이때 영상 인코더(200)는 입력 영상을 그대로 인코딩하지 않고, 전경 및 배경으로 구분하고, 구분된 전경 및 배경 각각을 별도로 인코딩하여 전경 특징과 외형 특징을 추출하여 전경과 배경에 대한 외형적 특징을 더욱 정확하게 추출할 수 있도록 한다.
- [0037] 또한 영상 인코더(200)는 입력 영상의 모든 프레임을 인가받아 외형적 특징을 추출하는 것이 아니라, 입력 영상의 다수의 프레임 중 기지정된 간격 단위로 프레임을 추출하고, 추출된 프레임들로부터 전경 및 배경의 외형적 특징을 추출할 수 있다. 이는 영상에 포함된 다수의 프레임에서 외형 정보는 움직임 정보에 비해 상대적으로 변화가 작기 때문이다.
- [0038] 영상 인코더(200)는 배경 추출부(210)와 외형 인코더(220)를 포함할 수 있다. 배경 추출부(210)는 영상 획득부(100)가 획득한 입력 영상을 인가받아 기지정된 방식으로 전경과 배경을 구분하여 추출한다. 여기서 배경 추출부(210)가 전경과 배경을 구분하여 추출하는 것은 이상 상황은 전경과 배경 사이의 관계에 따라 서로 다르게 분류될 수 있기 때문이다. 일 예로 사람이 운동장을 달리는 상황은 정상 상황이지만, 차도를 무단으로 횡단하는 경우는 이상 상황이 된다. 또한 차량이 도로를 주행하는 것은 정상 상황인 반면, 보행자 도로를 주행하는 경우에는 이상 상황으로 볼 수 있다. 즉 영상의 이상 상황은 전경과 배경 사이의 관계에 따라 상이하게 판별될 수 있다. 이에 본 실시예에서는 정확한 이상 상황 검출이 가능하도록 우선 배경 추출부(210)가 전경과 배경을 구분하여 추출한다.
- [0039] 여기서는 일 예로 배경 추출부(210)가 가우시안 혼합(Mixture-of-Gaussians: MoG) 모델을 기반으로 영상에서 전경과 배경을 구분하여 추출하는 것으로 가정하지만, 이에 한정되지 않는다. 영상에서 전경과 배경을 추출하는 기법은 인공 신경망을 이용하는 기법을 포함하여 매우 다양한 기법이 공지되어 있으므로, 배경 추출부(210)는 공지된 기법을 이용하여 전경과 배경을 추출할 수 있다.
- [0040] 또한 상기한 바와 같이 배경 추출부(210)는 입력 영상의 모든 프레임을 인가받는 것이 아니라, 기지정된 간격 단위(여기서는 일 예로 10 프레임 당 1프레임)로 프레임을 선택하고, 선택된 프레임을 인가받아 전경과 배경을 구분할 수 있다.

- [0041] 외형 인코더(220)는 배경 추출부(210)에서 구분된 전경과 배경을 각각 인가받아 미리 학습된 방식으로 인코딩하여 전경 특징과 배경 특징을 획득한다. 외형 인코더(220)는 배경 인코더(221)와 전경 인코더(222)를 포함할 수 있다. 배경 인코더(221)는 배경 추출부(210)에서 추출된 배경을 인가받아 미리 학습된 방식으로 인코딩하여 배경 특징을 추출하고, 전경 인코더(222)는 배경 추출부(210)에서 추출된 전경을 인가받아 미리 학습된 방식으로 인코딩하여 전경 특징을 추출한다.
- [0042] 본 실시예에서는 외형 인코더(220)가 배경 인코더(221)와 전경 인코더(222)를 포함하여 배경과 전경을 구분하여 개별적으로 인코딩하여 배경 특징과 전경 특징을 추출함으로써, 입력 영상에 포함된 배경과 전경 각각에 대한 특징을 더욱 정확하게 추출할 수 있도록 한다.
- [0043] 한편, 움직임 인코더(300)는 영상 획득부(100)가 획득한 입력 영상을 인가받아 인코딩하여 다수의 프레임 사이의 변화, 즉 움직임 특징을 추출한다. 본 실시예에서 이상 상황 검출 장치가 영상 인코더(200) 이외에 움직임 인코더(300)를 더 구비하여 움직임 특징을 추출하는 것은 객체 또는 배경의 외형적 이상 상황 뿐만 아니라 입력 영상의 다수의 프레임 사이의 변화로 나타나는 움직임에 기반한 이상 상황을 정확하게 검출할 수 있도록 하기 위해서이다.
- [0044] 여기서 움직임 인코더(300)는 입력 영상에서 각 프레임 사이에 변화량이 많은 움직임 특징을 추출해야 하므로, 영상 인코더(200)와 달리 입력 영상의 모든 프레임을 인가받아 인코딩하여 움직임 특징을 추출할 수 있다.
- [0045] 이에 도 5에 도시된 바와 같이, 영상 인코더(200)의 배경 인코더(221)와 전경 인코더(222)는 다수의 커널을 포함하여 구성될 수 있으며, 입력 영상에서 프레임이 선택적으로 추출되어 인가되므로, 시간축(T)으로 크기는 작은 반면, 채널축(C) 크기가 상대적으로 큰 형태의 커널을 포함하는 인공 신경망으로 구현될 수 있다. 반면 움직임 인코더(300)는 입력 영상의 모든 프레임이 인가되므로, 채널축(C)보다 시간축(T)에서의 크기가 상대적으로 작은 형태의 커널을 포함하는 인공 신경망으로 구현되어 시간의 변화에 따른 움직임 특징을 더 정확하게 추출하도록 구성될 수 있다.
- [0046] 도 5에서 FG와 BG는 각각 배경 추출부(210)에서 추출된 전경과 배경을 나타내고,  $E_{app}$  와  $E_{mot}$  는 외형 인코더(220)의 배경 인코더(221)와 전경 인코더(222)를 나타내며, D는 디코더(400)를 나타내며,  $F_{app}$  와  $F_{mot}$  는 각각 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)를 나타낸다.
- [0047] 영상 인코더(200)와 움직임 인코더(300)는 입력 영상을 학습된 방식으로 인코딩하여 특징을 추출하는 인코더로 볼 수 있다.
- [0048] 디코더(400)는 영상 인코더(200)에서 추출된 배경 특징과 전경 특징 및 움직임 인코더(300)에서 추출된 움직임 특징을 모두 인가받고, 인가된 배경 특징과 전경 특징 및 움직임 특징을 기반으로 미리 학습된 방식에 따라 디코딩하여, 입력 영상의 다수의 프레임을 재건축하여 재건축 영상을 획득할 수 있다.
- [0049] 여기서 디코더(400)는 배경과 전경의 외형뿐만 아니라 움직임을 반영하여 입력 영상의 다수의 프레임을 재건축할 수 있도록, 인가된 배경 특징과 전경 특징 및 움직임 특징을 결합(concatenation)한 후 학습된 방식에 따라 디코딩하여 재건축 영상을 획득할 수 있다.
- [0050] 표 1은 외형 인코더의 배경 인코더(221)와 전경 인코더(222) 및 움직임 인코더(300)의 커널 구성의 일 예를 나타낸다.

표 1

Layer	Encoder		Output size
	Appearance	Motion	
Conv1	(1, 3 <sup>2</sup> ) stride: 1, 2 <sup>2</sup>	(5, 3 <sup>2</sup> ) stride: 1, 2 <sup>2</sup>	app: 96 × 4 × 128 <sup>2</sup> motion: 12 × 16 × 128 <sup>2</sup>
Conv2	(1, 3 <sup>2</sup> ) stride: 1, 2 <sup>2</sup>	(3, 3 <sup>2</sup> ) stride: 1, 2 <sup>2</sup>	app: 128 × 4 × 64 <sup>2</sup> motion: 16 × 16 × 64 <sup>2</sup>
Conv3	(3, 3 <sup>2</sup> ) stride: 1, 2 <sup>2</sup>	(3, 3 <sup>2</sup> ) stride: 1, 2 <sup>2</sup>	app: 256 × 4 × 32 <sup>2</sup> motion: 32 × 16 × 32 <sup>2</sup>
Conv4	(3, 3 <sup>2</sup> ) stride: 1, 1 <sup>2</sup>	(3, 3 <sup>2</sup> ) stride: 1, 1 <sup>2</sup>	app: 256 × 4 × 32 <sup>2</sup> motion: 32 × 16 × 32 <sup>2</sup>
Layer	Decoder		Output size
DeConv1	(3, 3 <sup>2</sup> ) stride: 1, 1 <sup>2</sup>		256 × 4 × 32 <sup>2</sup>
DeConv2	(3, 3 <sup>2</sup> ) stride: 2, 2 <sup>2</sup>		128 × 8 × 64 <sup>2</sup>
DeConv3	(3, 3 <sup>2</sup> ) stride: 2, 2 <sup>2</sup>		96 × 16 × 128 <sup>2</sup>
DeConv4	(3, 3 <sup>2</sup> ) stride: 1, 2 <sup>2</sup>		3 × 16 × 256 <sup>2</sup>

[0051]

[0052]

표 1에 나타난 바와 같이, 배경 인코더(221)와 전경 인코더(222)는 움직임 인코더(300)는 다수의 컨볼루션 커널을 포함하여 구성될 수 있으며, 배경 인코더(221)와 전경 인코더(222)는 동일한 커널을 갖는 반면, 움직임 인코더(300)는 다른 크기의 컨볼루션 커널을 포함하도록 구성될 수 있다. 그리고 디코더(400)는 다수의 디콘볼루션 커널을 포함하여 구성될 수 있다.

[0053]

본 실시예에서는 입력 영상을 그대로 인코딩하여 특징을 추출하고, 추출된 특징으로부터 영상을 재건축하는 기존의 이상 상황 검출 장치와 달리, 입력 영상에서 전경과 배경 및 움직임을 구분하여 특징을 추출하고, 구분되어 추출된 전경 특징, 배경 특징 및 움직임 특징 각각을 개별적으로 반영하여 영상을 재건축함으로써, 전경과 배경 각각의 외형적 변화와 움직임 변화가 정확하게 반영되는 재건축 영상을 획득할 수 있다.

[0054]

여기서 영상 인코더(200)와 움직임 인코더(300)를 포함하는 인코더와 디코더(400)는 오토인코더를 구성하며, 오토인코더는 정상 상황에 대한 영상을 입력 영상으로 인가받아 학습될 수 있다. 즉 정상 상황에 대한 영상이 학습 영상으로 이용될 수 있다. 이때 오토인코더는 학습 영상이 입력 영상으로 인가된 오토인코더는 입력 영상(I)과 재건축 영상( $\hat{I}$ ) 사이의 차이를 재건축 손실( $L_{recon}$ )로 계산하고, 계산된 재건축 손실( $L_{recon}$ )을 역전과 함으로써, 인코더와 디코더(400)를 학습시킬 수 있다.

[0055]

재건축 손실( $L_{recon}$ )은  $L_2$  놈 함수를 이용한  $L_2$  손실( $L_2$ )과 입력 영상과 재건축 영상 각각에서 각 픽셀과 인접 픽셀 사이의 변화 차를 누적한 그래디언트 손실( $L_{grd}$ )의 합으로 수학식 1 내지 3에 따라 계산될 수 있다.

수학식 1

$$L_2 = \|I - \hat{I}\|_2$$

[0056]

## 수학식 2

$$L_{grd} = \sum_{i,j} \left\| |I_{i,j} - I_{i-1,j}| - |\hat{I}_{i,j} - \hat{I}_{i-1,j}| \right\|_1 + \sum_{i,j} \left\| |I_{i,j} - I_{i,j-1}| - |\hat{I}_{i,j} - \hat{I}_{i,j-1}| \right\|_1$$

[0057]

[0058] 여기서 i, j는 각각 입력 영상(I)과 재건축 영상( $\hat{I}$ )의 픽셀 좌표를 나타낸다.

## 수학식 3

$$L_{recon} = L_2 + L_{grd}$$

[0059]

[0060] 상기와 같이 학습된 오토인코더는 입력 영상이 정상 상황에 대한 영상인 경우, 입력 영상과 유사한 영상을 재건축할 수 있는 반면, 배경이나 전경에서 외형적으로 이상 상황이 발생하거나, 움직임에서 이상 상황이 발생하게 되면, 영상을 정상적으로 재건축할 수 없게 되어 입력 영상과 상이한 재건축 영상을 출력하게 된다.

[0061] 따라서 스코어 계산부(700)가 입력 영상과 재건축 영상 사이의 오차를 분석하여 정확하게 입력 영상에서 이상 상황 발생 여부를 판별할 수 있다. 또한 디코더(400)가 입력 영상의 다수의 프레임에 대응하는 다수 프레임을 포함하는 영상을 재건축하므로, 입력 영상의 다수의 프레임과 재건축 영상의 대응하는 프레임을 비교하여 오차를 분석하여 입력 영상에서 이상 상황이 발생한 프레임을 별도로 추출할 수도 있다.

[0062] 도 6은 본 실시예의 이상 상황 검출 장치의 이상 상황 검출 성능을 기존의 이상 상황 검출 장치와 비교한 결과를 나타낸다.

[0063] 도 6에서 (a)는 입력 이미지를 나타내고, (b)는 기존의 오토 인코더를 이용하여 획득된 재건축 오차를 나타내며, (c)는 본 실시예에 따른 오토 인코더를 이용하여 획득된 재건축 오차를 나타낸다.

[0064] 도 6의 (b)와 (c)를 비교하면, 기존의 이상 상황 검출 장치에서는 단지 입력 영상 전체의 특징을 추출하여 재건축함에 따라 외형적 특징에만 치중되어 움직임에 따른 재건축 오차가 잘 나타나지 않는 반면, 본 실시예에 따른 이상 상황 검출 장치는 입력 영상의 전경과 배경을 분리하여 특징을 추출할 뿐만 아니라, 움직임 특징도 추출하여 전경 특징과 배경 특징 및 움직임 특징이 모두 반영된 재건축 영상을 획득하므로, 다양한 이상 상황, 특히 움직임에 따른 이상 상황 발생을 정확하게 검출할 수 있다.

[0065] 다만 본 발명에서는 이상 상황을 더욱 정확하게 검출하기 위해, 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)를 더 구비한다.

[0066] 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600) 각각은 일 예로 알려진 인공 신경망의 한 종류인 흐름 기반 생성 모델(flow-based generative model)로 구현될 수 있다. 흐름 기반 생성 모델은 데이터의 분포를 추정하는 인공 신경망으로 외형 특징 분포 추정부(500)는 배경 인코더(221)와 전경 인코더(222)에서 획득된 배경 특징과 전경 특징을 인가받아 결합하고, 결합된 외형 특징에 대한 데이터 분포를 추정한다. 그리고 움직임 특징 분포 추정부(600)는 움직임 인코더(300)에서 획득된 움직임 특징을 인가받아 움직임 특징에 대한 데이터 분포를 추정한다.

[0067] 즉 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)는 입력 영상의 외형적 특징의 분포와 움직임 특징의 분포를 추정하여, 외형 특징과 움직임 특징의 분포가 정상 상황에 대한 분포인지 여부를 판정할 수 있도록 한다.

[0068] 이때, 흐름 기반 생성 모델로 구현되는 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)는 음의 로그 형태를 갖는 네거티브 로그 가능도(negative log likelihood: 이하 NLL)를 손실로 하여 학습될 수 있다. 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600) 또한 정상 상황의 입력 영상이 학습 영상으로 인가되면, 학습 영상에 대한 외형 특징과 움직임 특징의 분포를 출력한다. 그리고 출력된 외형 특징 분포와 움직임



특징 분포 각각의 NLL을 계산하여 역전파함으로써, 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)를 학습시킬 수 있다.

[0069] 다만, 본 실시예에서는 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)가 동일한 입력 이미지에 대한 특징으로부터 데이터 분포를 추정하므로, 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)의 손실을 각각 추정하여 개별적으로 학습이 수행되기 보다는 함께 학습이 수행되어야 한다. 이에 본 실시예에서는 흐름 기반 생성 모델로 구현되는 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)의 가능도를 정규화하여 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)에 대한 통합된 손실을 특징 분포 손실( $L(x_t^a, x_t^m)$ )로 수학식 4와 같이 계산하여 역전파함으로써 학습시킬 수 있다.

#### 수학식 4

$$L(x_t^a, x_t^m) = \frac{\lambda_1 NLL(x_t^a) + \lambda_2 NLL(x_t^m)}{\lambda_1 + \lambda_2}$$

$$\lambda_1 = \max(x^a) - \min(x^a), \lambda_2 = \max(x^m) - \min(x^m)$$

[0070]

[0071] 여기서  $x_t^a$  와  $x_t^m$  는 각각 입력 영상(I)의 t번째 프레임( $I_t$ )의 외형 특징과 움직임 특징을 나타내고, NLL()은 네거티브 로그 가능도(negative log likelihood) 함수이며,  $\lambda_1$  과  $\lambda_2$  는 각각 외형 특징 분포 추정부(500)의 손실인 네거티브 로그 가능도( $NLL(x_t^a)$ )와 움직임 특징 분포 추정부(600)의 손실인 네거티브 로그 가능도( $NLL(x_t^m)$ )를 정규화하기 위한 정규화 파라미터이다.

[0072] 여기서 외형 특징 분포 추정부(500)와 움직임 특징 분포 추정부(600)를 통합하여 특징 분포 추정부라 불릴 수 있다.

[0073] 스코어 계산부(700)는 입력 영상(I)의 각 프레임( $I_t$ )과 디코더(400)에서 획득된 재건축 영상( $\hat{I}$ )의 각 프레임( $\hat{I}_t$ ) 사이의 오차에 따른 재건축 스코어(Recon.score)와 특징 분포 추정부에서 추정된 특징 분포 손실( $L(x_t^a, x_t^m)$ )을 기반으로 총 스코어(Total score)( $S_t$ )를 계산한다.

[0074] 스코어 계산부(700)는 특징 분포 손실( $L(x_t^a, x_t^m)$ )을 NLL 스코어(NLL score)로 획득하고, NLL 스코어에 대응하여 재건축 스코어(Recon.score)를 수학식 5와 같이 계산한다.

#### 수학식 5

$$R(I_t, \hat{I}_t) = -\log_{10} (1 - |I_t - \hat{I}_t|)$$

[0075]

[0076] 그리고 스코어 계산부(700)는 NLL 스코어(NLL score)와 재건축 스코어(Recon.score)로부터 총 스코어( $S_t$ )를 수학식 6에 따라 계산하여 획득한다.

#### 수학식 6

$$S_t = \text{norm}(R(I_t, \hat{I}_t)) + \text{norm}(L(x_t^a, x_t^m))$$

[0077]

[0078] 여기서 norm은 정규화 함수로서, 재건축 스코어(Recon.score)에 대한 정규화는

$$\text{norm}(R(I_t, \hat{I}_t)) = \frac{R(I_t, \hat{I}_t) - \min(R(I, \hat{I}))}{\max(R(I, \hat{I})) - \min(R(I, \hat{I}))}$$
로 계산될 수 있으며, NLL 스코어(NLL score)에 대한 정규화 또한 동일한 방식으로 계산될 수 있다.

- [0079] 한편, 스코어 계산부(700)는 각 프레임에 대한 총 스코어( $S_t$ )가 계산되면, 계산된 총 스코어( $S_t$ )가 기지정된 문턱값 이상인 프레임을 이상 상황이 나타난 프레임으로 추출한다.
- [0080] 도 7은 도 4의 스코어 계산부가 재건축 스코어와 가능도 스코어에 따라 총 스코어를 계산하여 이상 프레임을 검출하는 일 예를 나타낸다.
- [0081] 도 7에서 녹색 점선은 정규화된 재건축 스코어(Recon.score)이고, 파란색 점선은 가능도를 나타내는 정규화된 NLL 스코어(NLL score)이다. 그리고 빨간색 실선은 총 스코어( $S_t$ )를 나타낸다.
- [0082] 도 7에 도시된 바와 같이, 본 실시예의 이상 상황 검출 장치는 입력 영상을 재건축함에 있어 전경과 배경 및 움직임을 구분하여 특징을 추출하고, 추출된 전경 특징과 배경 특징 및 움직임 특징을 모두 반영하여 재건축함으로써, 기존의 이상 상황 검출 장치보다 더욱 정확하게 이상 상황을 검출할 수 있다. 또한 추가적으로 전경 특징과 배경 특징 및 움직임 특징의 데이터 분포에 따른 가능도를 함께 고려하여 이상 상황을 검출할 수 있도록 함으로써, 오토인코더를 이용한 재건축 오차만을 고려하는 이상 상황 검출 방식보다 더욱 정확하게 이상 상황을 검출할 수 있다.
- [0083] 도 8은 본 발명의 일 실시예에 따른 이상 상황 검출 방법을 나타낸다.
- [0084] 도 4 내지 도 7을 참조하여, 도 8의 이상 상황 검출 방법을 설명하면, 우선 다수의 프레임을 포함하는 입력 영상을 획득한다(S11).
- [0085] 그리고 획득된 입력 영상의 다수의 프레임에서 기지정된 방식에 따라 전경과 배경을 구분하여 추출한다(S21). 이때 입력 영상의 다수의 프레임 중 기지정된 간격으로 프레임을 선택하고 선택된 프레임에 대해 전경과 배경을 구분하여 추출할 수 있다. 전경과 배경이 구분되면, 미리 학습된 인공 신경망으로 구현된 인코더를 이용하여 전경과 배경을 각각 인코딩하여 전경 특징과 배경 특징을 추출한다(S22).
- [0086] 한편 미리 학습된 인공 신경망으로 구현된 인코더를 이용하여 입력 영상으로부터 다수의 프레임 사이의 변화를 인코딩하여 움직임 특징을 추출한다(S23).
- [0087] 그리고 전경 특징과 배경 특징 및 움직임 특징이 추출되면, 추출된 전경 특징과 배경 특징 및 움직임 특징을 결합한다(S31). 전경 특징과 배경 특징 및 움직임 특징이 결합된 결합 특징이 획득되면, 결합 특징을 미리 학습된 인공 신경망으로 구현된 디코더를 이용하여 디코딩하여 입력 영상에 대응하는 재건축 영상을 획득한다(S32). 재건축 영상이 획득되면, 입력 영상과 재건축 영상 사이의 차이를 기반으로 재건축 스코어를 수학적 식 5에 따라 계산한다.
- [0088] 이와 함께 미리 학습된 흐름 기반 생성 모델을 기반으로 전경 특징과 배경 특징이 결합된 외형 특징에 대한 특징 분포를 NLL 형식으로 추정한다(S41). 또한 미리 학습된 흐름 기반 생성 모델을 기반으로 움직임 특징에 대한 특징 분포를 NLL 형식으로 추정한다(S42).
- [0089] 외형 특징과 움직임 특징 각각에 대한 특징 분포가 추정되면, 추정된 외형 특징 분포와 움직임 특징 분포를 정규화하여 통합하여 NLL 스코어를 수학적 식 4에 따라 계산한다(S43).
- [0090] 이후, 계산된 재건축 스코어와 NLL 스코어를 수학적 식 6과 같이 통합하여 총 스코어( $S_t$ )를 계산한다(S51). 입력 영상의 각 프레임에 대한 총 스코어( $S_t$ )가 계산되면, 계산된 각 프레임별 총 스코어( $S_t$ ) 중 기지정된 문턱값 이상인 총 스코어에 대응하는 프레임을 탐색하여 이상 상황으로 검출한다(S52).
- [0091] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.
- [0092] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의

지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

[0093] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

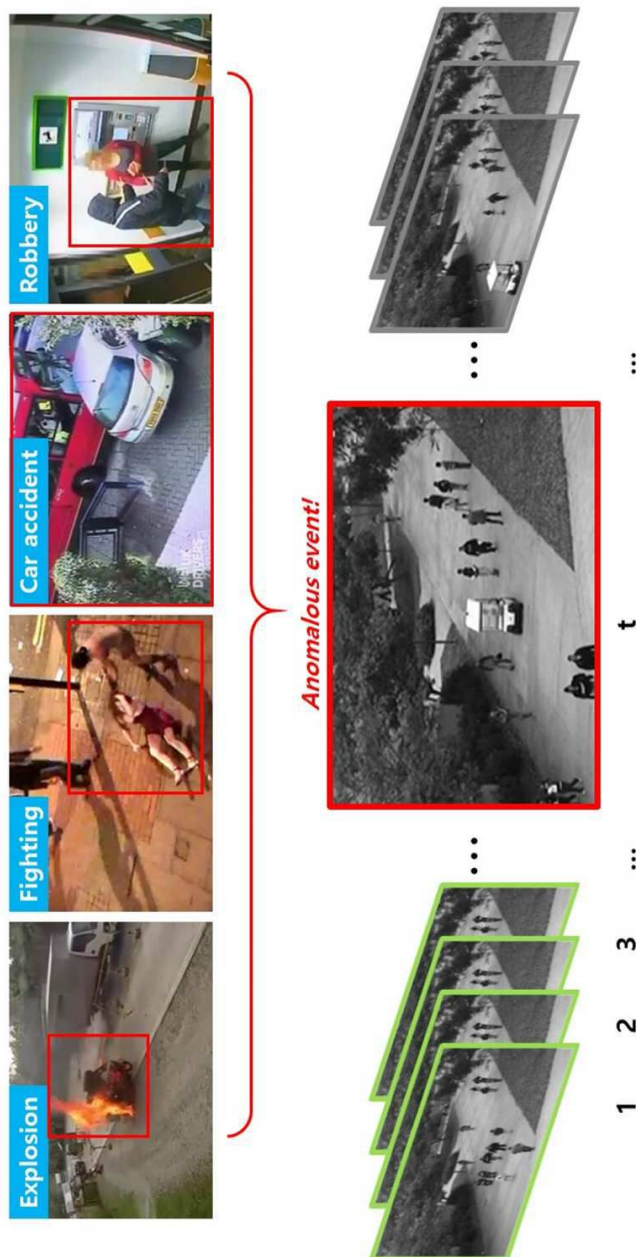
### 부호의 설명

[0094] 100: 영상 획득부 200: 영상 인코더  
210: 배경 추출부 220: 외형 인코더  
221: 배경 인코더 222: 전경 인코더  
300: 움직임 인코더 400: 디코더  
500: 외형 특징 분포 추정부 600: 움직임 특징 분포 추정부  
700: 스코어 계산부

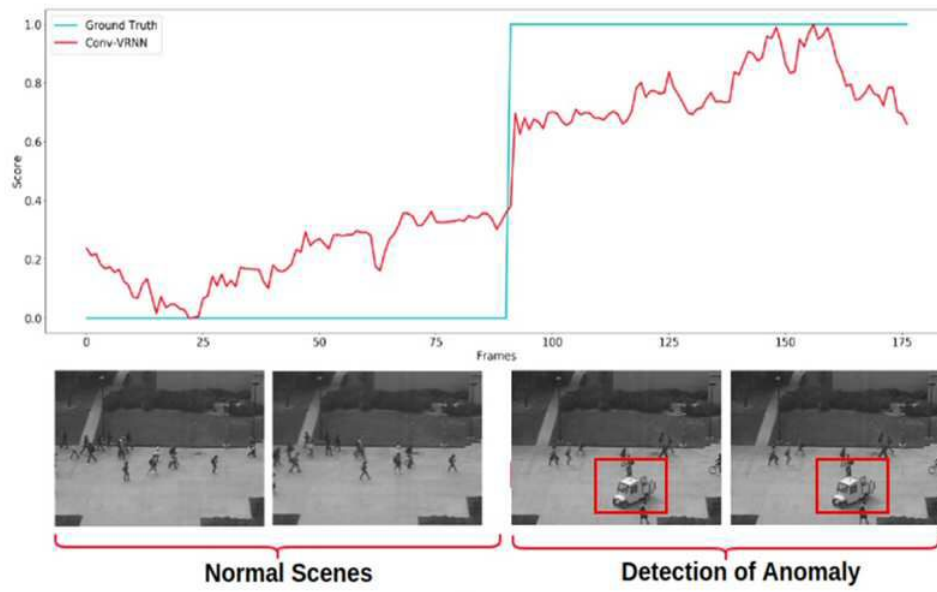


도면

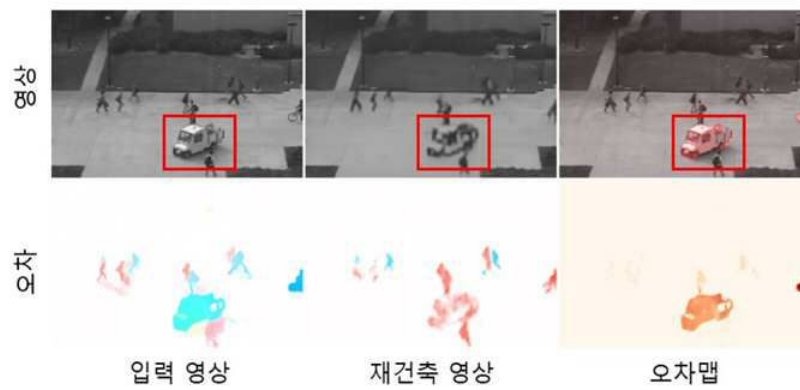
도면1



도면2



(a)



(b)

도면3

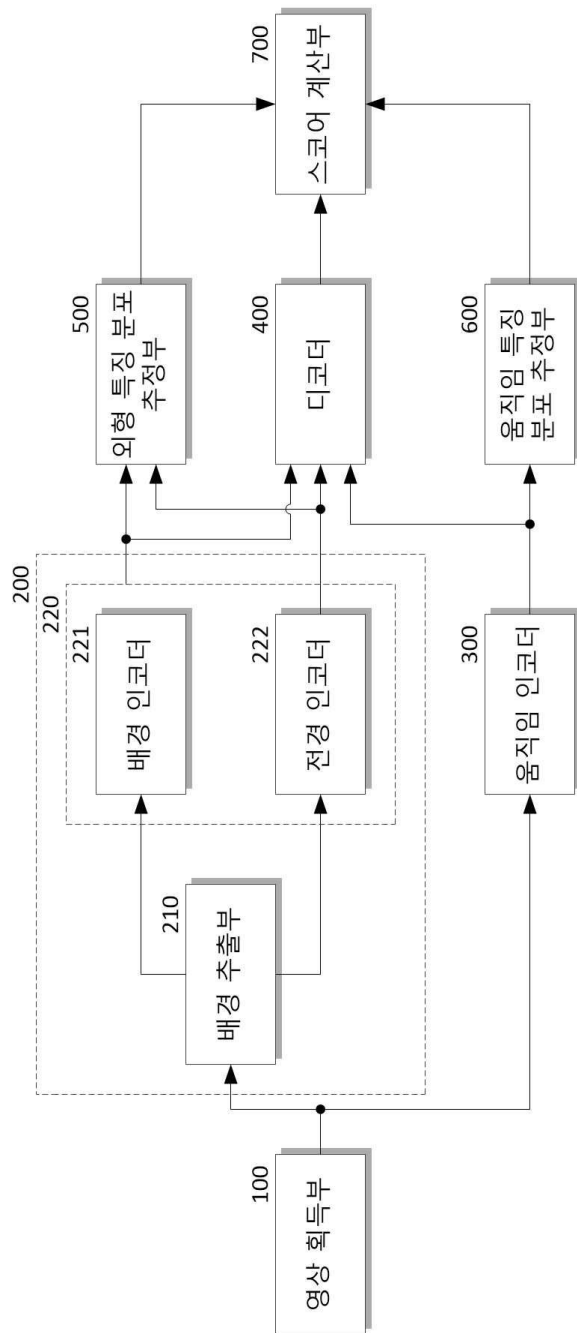


(a)

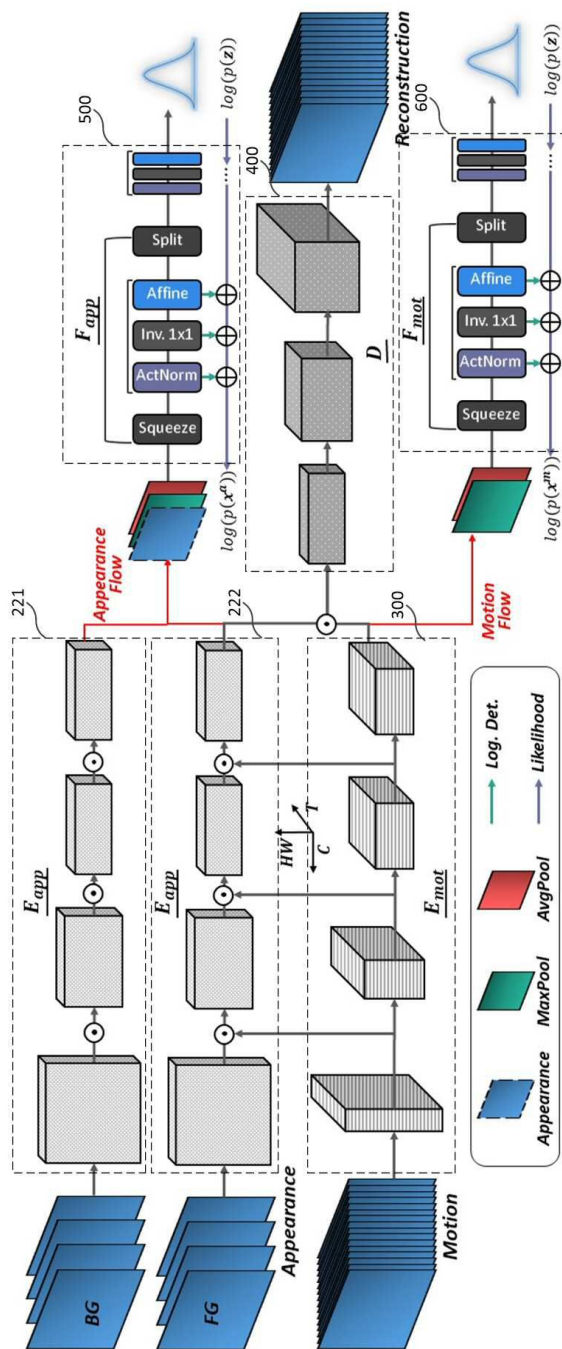


(b)

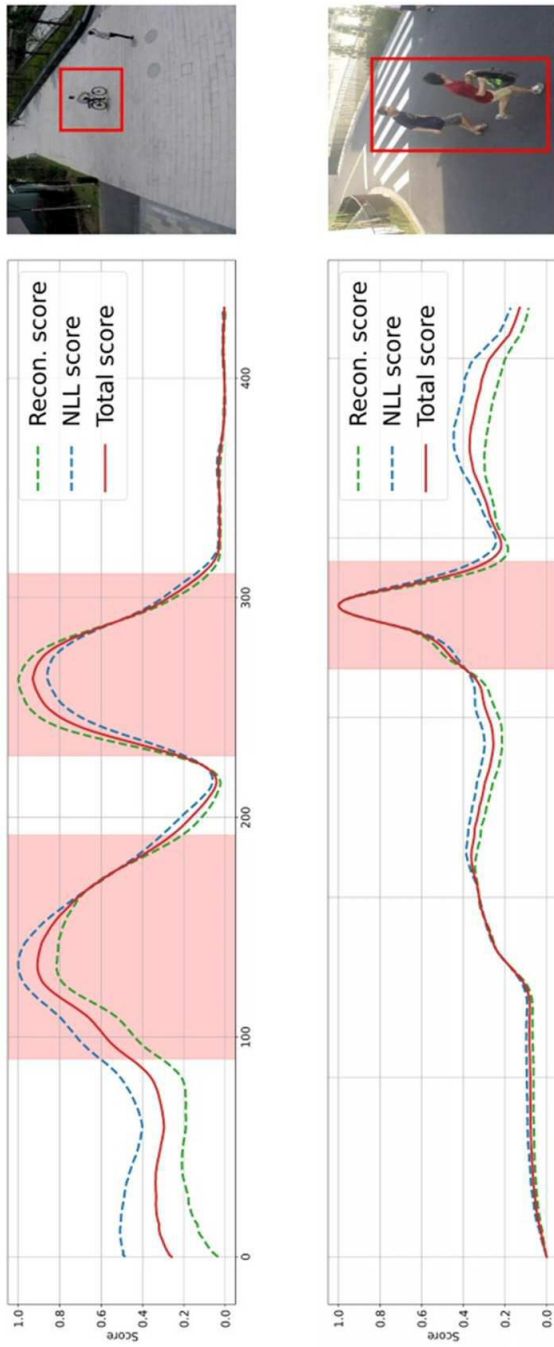
도면4



도면5



도면6



도면7



도면8

