【서지사항】

【서류명】 특허출원서

【참조번호】 SDP20244386

【출원구분】 특허출원

【출원인】

【명칭】 연세대학교 산학협력단

【특허고객번호】 2-2005-009509-9

【대리인】

【명칭】 특허법인시공

【대리인번호】 9-2023-100041-2

【지정된변리사】 조예찬

【포괄위임등록번호】 2023-059479-9

【발명의 국문명칭】 F2LM 기반의 비디오 이상 탐지 방법 및 장치

【발명의 영문명칭】 VIDEO ANOMALY DETECTION METHOD AND APPARATUS BASED ON F2LM

【발명자】

【성명】 박상현

【성명의 영문표기】 SANGHYUN PARK

【주민등록번호】 670101-1XXXXXX

【우편번호】 08004

【주소】 서울특별시 양천구 오목로 300, 204동 3701호

【발명자】

【성명】 홍승균

【성명의 영문표기】 SEUNGKYUN HONG

【주민등록번호】 671120-1XXXXXX

【우편번호】 08083

【주소】 서울특별시 양천구 중앙로36길 15, 111동 1902호

【발명자】

【성명】 안성현

【성명의 영문표기】 SUNGHYUN AHN

【주민등록번호】 000104-3XXXXXX

【우편번호】 21368

【주소】 인천광역시 부평구 원적로269번길 15, 104동 1901호

【발명자】

【성명】 조영완

【성명의 영문표기】 YOUNGWAN JO

【주민등록번호】 990312-1XXXXXX

【우편번호】 03726

【주소】 서울특별시 서대문구 성산로17길 5, 402호

【출원언어】 국어

【심사청구】 청구

【이 발명을 지원한 국가연구개발사업】

【과제고유번호】 1711198526

【과제번호】 00229822

2024-04-25

【부처명】 과학기술정보통신부

【과제관리(전문)기관명】 한국연구재단

【연구사업명】 인공지능활용혁신신약발굴

【연구과제명】 난치성 질환 극복을 위한 인공지능 기반의 다중 약물 적응증 최적화 플랫폼 개발 및 혁신신약 발굴

【과제수행기관명】 연세대학교

【연구기간】 2024.01.01 ~ 2024.12.31

【취지】 위와 같이 특허청장에게 제출합니다.

대리인 특허법인시공 (서명 또는 인)

【수수료】

【출원료】	0	면	46,000	원
【가산출원료】	57	면	0	원
【우선권주장료】	0	건	0	원
【심사청구료】	18	항	1,084,000	원

【합계】 1,130,000원

【감면사유】 전담조직(50%감면)[1]

【감면후 수수료】 565,000 원

【발명의 설명】

【발명의 명칭】

F2LM 기반의 비디오 이상 탐지 방법 및 장치{VIDEO ANOMALY DETECTION METHOD AND APPARATUS BASED ON F2LM}

【기술분야】

【0001】본원 발명은 F2LM 기반의 비디오 이상 탐지 방법 및 장치에 관한 것으로, 구체적으로, 프레임 예측을 통해 비디오 이상 탐지를 수행하는 방법 및 장치에 관한 것이다.

【발명의 배경이 되는 기술】

【0002】최근 AI(artificial intelligence) 등의 기술이 발달함에 따라 비디오 이상 탐지를 위한 다양한 기술이 개발되고 있다. 여기서, 비디오 이상 탐지는 비정상적인 상황이나 행동을 감지하여 피해를 예방하는 기술을 나타낼 수 있다. 일반적으로, 비디오 이상 탐지를 위한 AI 모델은 비정상 데이터 확보의 어려움으로 인해 정상 데이터만을 이용하여 학습된다.

【0003】이와 같이 정상 데이터로 학습된 AI 모델은 정상 데이터에 대해서는 비디오의 프레임을 정확하게 예측하고, 비정상 데이터에 대해서는 비디오의 프레임을 정확하게 예측하지 못하는 것으로 가정한다. 즉, 학습된 AI 모델에 의해 예측된 프레임을 실제 프레임과 비교하여 유사도가 높은 경우 이상 행위가 발생하지 않은 것으로 판단하고, 유사도가 낮은 경우 이상 행위가 발생한 것으로 판단함으로써 비

디오 이상 탐지가 수행될 수 있다.

【0004】그러나, 이와 같이 정상 데이터만으로 학습을 수행하는 경우에도, AI 모델의 높은 일반화 성능으로 인해, AI 모델은 비정상 데이터에 대해서도 높은 정확도로 프레임을 예측한다. 따라서, AI 모델에 비정상 데이터가 입력되는 경우에도 예측된 프레임과 실제 프레임 사이의 유사도가 높게 산출되어 이상 행위가 정확하게 탐지되지 않는 문제가 있다.

【발명의 내용】

【해결하고자 하는 과제】

【0005】본원 발명은 상기와 같은 문제점을 해결하기 위한 F2LM 기반의 비디오 이상 탐지 방법, 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램, 컴퓨터 프로그램이 저장된 컴퓨터 판독 가능 매체 및 장치(시스템)를 제공한다.

【과제의 해결 수단】

【0006】본원 발명은 방법, 장치(시스템), 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램 또는 컴퓨터 프로그램이 저장된 컴퓨터 판독 가능 매체를 포함한 다양한 방식으로 구현될 수 있다.

【0007】본원 발명의 일 실시예에 따르면, 적어도 하나의 프로세서에 의해 수행되는 F2LM 기반의 비디오 이상 탐지 방법은, 비디오의 연속하는 프레임 시퀀스를 획득하는 단계, 획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하는 단계, 추출된 프레임 특징 정보, 라벨 특징 정

보 및 모션 특징 정보를 이용하여 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성하는 단계 및 생성된 제1 미래 프레임 중 품질이 임계치 미만인 비정상 영역을 결정하고, 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성하는 단계를 포함한다.

【0008】 본원 발명의 일 실시예에 따르면, 획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하는 단계는, 프레임 시퀀스에 대한 의미론적 분할을 수행하여 라벨 시퀀스를 생성하는 단계 및 생성된 라벨 시퀀스를 이용하여 라벨 특징 정보를 추출하는 단계를 포함한다.

【0009】 본원 발명의 일 실시예에 따르면, 획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하는 단계는, 프레임 시퀀스에 대한 광학 흐름 추정을 수행하여 모션 시퀀스를 생성하는 단계 및 생성된 모션 시퀀스를 이용하여 모션 특징 정보를 추출하는 단계를 포함한다.

【0010】 본원 발명의 일 실시예에 따르면, 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성하는 단계는, 라벨 특징 정보와 유사해지도록 프레임 특징 정보를 변환하여 변환된 라벨 특징 정보를 생성하는 단계, 모션 특징 정보와 유사해지도록 프레임 특징 정보를 변환하여 변환된 모션 특징 정보를 생성하는 단계 및 프레임 특징 정보, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보를 이용하여 제1 미래 프레임을 생성하는 단계를 포함한다.

【0011】본원 발명의 일 실시예에 따르면, 프레임 특징 정보, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보를 이용하여 제1 미래 프레임을 생성하는 단계는, 프레임 특징 정보, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보를 채널 방향으로 연결하여 채널 정보를 획득하는 단계 및 획득된 채널 정보로부터 추출된 특징맵을 이용하여 제1 미래 프레임을 생성하는 단계를 포함한다.

【0012】본원 발명의 일 실시예에 따르면, 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성하는 단계는, 제1 미래 프레임을 제1 세트의 패치로 분할하는 단계, 제1 세트의 패치 중 일부의 패치에 노이즈를 주입하여 제2 세트의 패치를 생성하는 단계, 생성된 제2 세트의 패치를 기초로 품질이 임계치 이상인 정상 영역 및 품질이 임계치 미만인 비정상 영역을 결정하는 단계 및 정상 영역의 패치를 제1 미래 프레임의 패치로 복원하고, 비정상 영역의 패치를 영벡터로 변환하여 제2 미래 프레임을 생성하는 단계를 포함한다.

【0013】본원 발명의 일 실시예에 따르면, 제1 미래 프레임과 정답 프레임 사이의 거리 및 제2 미래 프레임과 정답 프레임 사이의 거리를 기초로 이상 점수를 산출하는 단계 및 산출된 이상 점수가 임계치 이상인지 여부를 기초로 비디오 이상 탐지를 수행하는 단계를 더 포함한다.

【0014】본원 발명의 일 실시예에 따른 상술된 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램이 제공된다.

【0015】 본원 발명의 일 실시예에 따른 F2LM 기반의 비디오 이상 탐지 장치는, 비디오의 연속하는 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하고, 추출된 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 이용하여 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성하는 생성자 및 생성된 제1 미래 프레임 중 품질이 임계치 미만인 비정상 영역을 결정하고, 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성하는 파괴자를 포함한다.

【0016】 본원 발명의 일 실시예에 따른 생성자는, 프레임 특징 정보를 추출하기 위한 제1 인코더, 라벨 특징 정보를 추출하기 위한 제2 인코더 및 모션 특징 정보를 추출하기 위한 제3 인코더를 포함한다.

【0017】 본원 발명의 일 실시예에 따른 생성자는, 프레임 시퀀스에 대한 의미론적 분할을 수행하여 라벨 시퀀스를 생성하는 딥러닝 기반의 제1 모델을 포함한다.

【0018】 본원 발명의 일 실시예에 따른 생성자는, 프레임 시퀀스에 대한 광학 흐름 추정을 수행하여 모션 시퀀스를 생성하는 딥러닝 기반의 제2 모델을 포함한다.

【0019】 본원 발명의 일 실시예에 따른 생성자는, 라벨 특징 정보와 유사해지도록 프레임 특징 정보를 변환하여 변환된 라벨 특징 정보를 생성하는 제1 특징 변환 합성곱 블록 및 모션 특징 정보와 유사해지도록 모션 특징 정보를 변환하여

변환된 모션 특징 정보를 생성하는 제2 특징 변환 합성곱 블록을 포함한다.

【0020】 본원 발명의 일 실시예에 따른 생성자는, 프레임 특징 정보, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보를 채널 방향으로 연결하여 채널 정보를 획득하는 주의 모듈을 포함한다.

【0021】 본원 발명의 일 실시예에 따른 생성자는, 획득된 채널 정보로부터 추출된 특징맵을 이용하여 제1 미래 프레임을 생성하는 디코더를 포함한다.

【0022】 본원 발명의 일 실시예에 따른 생성자는, 프레임 특징 정보를 부정적 샘플로 하고, 라벨 특징 정보 및 모션 특징 정보를 긍정적 샘플로 하는 삼중항 손실 기반의 함수를 기초로 학습된다.

【0023】 본원 발명의 일 실시예에 따른 생성자는, 생성된 제1 미래 프레임과 프레임 시퀀스의 다음 프레임으로 결정된 정답 프레임이 유사해지도록 강도 손실 기반의 함수, 경사 손실 기반의 함수 및 적대적 손실을 기초로 학습된다.

【0024】 본원 발명의 일 실시예에 따른 파괴자는, 제1 미래 프레임 중 품질이 임계치 이상인 정상 영역은 복원하고, 품질이 임계치 미만인 비정상 영역은 영벡터로 변환하여 제2 미래 프레임을 생성한다.

【발명의 효과】

【0025】 본원 발명의 다양한 실시예에서 이상 행위를 포함하는 프레임 시퀀스가 입력되는 경우, 생성자는 품질이 낮은 미래 프레임을 생성할 수 있으며, 파괴자는 미래 프레임에서 품질이 낮은 비정상 영역을 더욱 파괴하여 정상 프레임과의

차이를 극대화할 수 있다. 이에 따라, 비디오 이상 탐지 장치는 모델의 높은 일반화 성능과 관계없이 높은 정확도로 비디오 이상 탐지를 수행할 수 있다.

【0026】본원 발명의 다양한 실시예에서 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 모두 활용하여 프레임 시퀀스의 다음 프레임을 예측하는 경우, 해당 프레임 시퀀스에 어떤 객체들이 존재하고, 각각의 객체가 어떤 동작을 수행하는지에 대한 정보를 참고할 수 있어 비디오 이상 탐지 장치는 더욱 정밀하게 미래 프레임을 예측할 수 있다.

【0027】본원 발명의 다양한 실시예에서 라벨 특징 정보 및 모션 특징 정보를 그대로 이용하지 않고, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보로 변환하여 사용하는 경우, 생성자는 이상 행위를 포함하는 비정상 데이터에 대해서는 제1 미래 프레임을 정밀하게 생성할 수 없어 정상 데이터와 비정상 데이터 간의 차이를 명확히 할 수 있다.

【도면의 간단한 설명】

【0028】본원 발명의 실시예들은, 이하 설명하는 첨부 도면들을 참조하여 설명될 것이며, 여기서 유사한 참조 번호는 유사한 요소들을 나타내지만, 이에 한정되지는 않는다.

도 1은 본원 발명의 일 실시예에 따른 비디오 이상 탐지 장치가 이상 탐지를 수행하는 예시를 나타내는 도면이다.

도 2는 본원 발명의 일 실시예에 따른 생성자가 프레임 특징 정보, 라벨 특

징 정보 및 모션 특징 정보를 생성하는 예시를 나타내는 도면이다.

도 3은 본원 발명의 일 실시예에 따른 생성자가 제1 미래 프레임을 생성하는 예시를 나타내는 도면이다.

도 4는 본원 발명의 일 실시예에 따른 주의 모듈의 구조를 나타내는 예시적인 도면이다.

도 5는 본원 발명의 일 실시예에 따른 파괴자가 제2 미래 프레임을 생성하는 예시를 나타내는 도면이다.

도 6은 본원 발명의 일 실시예에 따른 생성자의 학습 과정을 나타내는 예시적인 도면이다.

도 7은 본원 발명의 일 실시예에 따른 파괴자의 학습 과정을 나타내는 예시적인 도면이다.

도 8은 본원 발명의 일 실시예에 따른 특징 정보에 대한 확률적 임베딩의 예시를 나타내는 도면이다.

도 9는 본원 발명의 일 실시예에 따른 미래 프레임에 대한 확률적 임베딩의 예시를 나타내는 도면이다.

도 10은 본원 발명의 일 실시예에 따른 F2LM 기반의 비디오 이상 탐지 방법의 예시를 나타내는 흐름도이다.

도 11은 본원 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 나타내는 블록도이다.

【발명을 실시하기 위한 구체적인 내용】

【0029】이하, 본원 발명의 실시를 위한 구체적인 내용을 첨부된 도면을 참조하여 상세히 설명한다. 다만, 이하의 설명에서는 본원 발명의 요지를 불필요하게 흐릴 우려가 있는 경우, 널리 알려진 기능이나 구성에 관한 구체적 설명은 생략하기로 한다.

【0030】첨부된 도면에서, 동일하거나 대응하는 구성요소에는 동일한 참조부호가 부여되어 있다. 또한, 이하의 실시예들의 설명에 있어서, 동일하거나 대응되는 구성요소를 중복하여 기술하는 것이 생략될 수 있다. 그러나, 구성요소에 관한 기술이 생략되어도, 그러한 구성요소가 어떤 실시예에 포함되지 않는 것으로 의도되지는 않는다.

【0031】개시된 실시예의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명이 완전하도록 하고, 본 발명이 통상의 기술자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것일 뿐이다.

【0032】본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 개시된 실시예에 대해 구체적으로 설명하기로 한다. 본 명세서에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으

나, 이는 관련 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서, 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본원 발명의 전반에 걸친 내용을 토대로 정의되어야 한다.

【0033】본 명세서에서의 단수의 표현은 문맥상 명백하게 단수인 것으로 특정하지 않는 한, 복수의 표현을 포함한다. 또한, 복수의 표현은 문맥상 명백하게 복수인 것으로 특정하지 않는 한, 단수의 표현을 포함한다. 명세서 전체에서 어떤 부분이 어떤 구성요소를 포함한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다.

【0034】본 개시에서, "포함하다", "포함하는" 등의 용어는 특징들, 단계들, 동작들, 요소들 및/또는 구성 요소들이 존재하는 것을 나타낼 수 있으나, 이러한 용어가 하나 이상의 다른 기능들, 단계들, 동작들, 요소들, 구성 요소들 및/또는 이들의 조합이 추가되는 것을 배제하지는 않는다.

【0035】본 개시에서, 특정 구성 요소가 임의의 다른 구성 요소에 "결합", "조합", "연결" 되거나, "반응" 하는 것으로 언급된 경우, 특정 구성 요소는 다른 구성 요소에 직접 결합, 조합 및/또는 연결되거나, 반응할 수 있으나, 이에 한정되지 않는다. 예를 들어, 특정 구성 요소와 다른 구성 요소 사이에 하나 이상의 중간 구성 요소가 존재할 수 있다. 또한, 본 발명에서 "및/또는"은 열거된 하나 이상의

항목의 각각 또는 하나 이상의 항목의 적어도 일부의 조합을 포함할 수 있다.

【0036】본 개시에서, "제1", "제2" 등의 용어는 특정 구성 요소를 다른 구성 요소와 구별하기 위해 사용되는 것으로, 이러한 용어에 의해 상술된 구성 요소가 제한되진 않는다. 예를 들어, "제1" 구성 요소는 "제2" 구성 요소와 동일하거나 유사한 형태의 요소일 수 있다.

【0037】본 개시에서, "비디오 이상 탐지(video anomaly detection)"는 CCTV 등의 감시 카메라로부터 수집된 영상을 이용하여 싸움, 강도, 방화, 폭발 등의 이상 행위 및/또는 이상 상황을 탐지하는 것을 지칭할 수 있다. 예를 들어, 비디오 이상 탐지는 프레임을 재구축하거나 미래 프레임을 예측한 후 정답 프레임(또는 실제 프레임)과의 유사도를 비교하는 방식으로 수행될 수 있다.

【0038】본 개시에서, "이상(anomaly) 및/또는 이상 행위"는 사용자에 의해 사전 정의되는 비정상 행위로, 예를 들어, 싸우는 행위, 인도에서 자전거를 타는 행위 등의 사람에 관한 행위와 화재, 폭발 등의 재해 상황 등을 포함할 수 있다.

【0039】본 개시에서, "F2LM(feature to label and motion)"은 프레임 특징 정보를 변환된 라벨 특징 정보 및 변환된 모션 특징 정보로 변환하여 이용하는 방법을 지칭할 수 있다.

【0041】도 1은 본원 발명의 일 실시예에 따른 비디오 이상 탐지 장치(100)가 이상 탐지를 수행하는 예시를 나타내는 도면이다. 일 실시예에 따르면, 비디오

이상 탐지 장치(100)는 입력된 비디오(110)를 기초로 이상 및/또는 이상 행위를 탐지하기 위한 임의의 장치로서, 생성자(generator)(102) 및 파괴자(destroyer)(104)를 포함하도록 구성될 수 있다. 여기서, 생성자(102)는 비디오(110)에 포함된 프레임 시퀀스(frame sequence)의 다음 프레임을 예측하기 위한 모델일 수 있으며, 파괴자(104)는 생성자(102)에 의해 예측된 프레임에서 비정상 영역을 더욱 파괴하여 정상 영역과의 차이를 강조하기 위한 모델일 수 있다.

【0042】일 실시예에 따르면, 생성자(102)는 비디오(110)의 연속하는 프레임 시퀀스를 획득하고, 획득된 프레임 시퀀스를 기초로 프레임 특징 정보(frame feature information), 라벨 특징 정보(label feature information) 및 모션 특징 정보(motion feature information)를 추출할 수 있다. 또한, 생성자(102)는 추출된 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 이용하여 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성할 수 있다.

【0043】학습 과정에서 제1 미래 프레임은 정답 프레임과 유사하게 생성될 수 있으며, 추론 과정에서 제1 미래 프레임은 비디오(110)가 이상 행위를 포함하지 않는 경우 정답 프레임과 유사하게 생성되고, 비디오(110)가 이상 행위를 포함하는 경우 정답 프레임과 비유사하게 생성될 수 있다. 즉, 이상 행위를 포함하는 비디오(110)가 입력되는 경우, 생성자(102)는 정답 프레임과의 차이가 두드러지도록 품질이 낮은 제1 미래 프레임을 생성할 수 있다.

【0044】일 실시예에 따르면, 생성된 제1 미래 프레임은 파괴자(104)에게 제공될 수 있다. 예를 들어, 파괴자(104)는 생성된 제1 미래 프레임 중 품질이 임계

치 미만인 비정상 영역을 결정하고, 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성할 수 있다. 즉, 파괴자(104)는 생성자(102)에 의해 생성된 제1 미래 프레임 중 비정상 영역을 파괴하는 것에 의해 정답 프레임과의 차이가 극대화된 제2 미래 프레임을 생성할 수 있다.

【0045】 일 실시예에 따르면, 비디오 이상 탐지 장치(100)는 예측된 제1 미래 프레임 및/또는 제2 미래 프레임과 정답 프레임 간의 유사도가 높은 경우 이상 행위가 발생하지 않은 것으로 판단하고, 유사도가 낮은 경우 이상 행위가 발생한 것으로 판단할 수 있다. 이러한 유사도를 판단하기 위해, 비디오 이상 탐지 장치(100)는 제1 미래 프레임과 정답 프레임 사이의 거리 및 제2 미래 프레임과 정답 프레임 사이의 거리를 기초로 이상 점수(anomaly score)(120)를 산출할 수 있다.

【0046】 그리고 나서, 비디오 이상 탐지 장치(100)는 산출된 이상 점수(120)가 임계치 이상인지 여부를 기초로 비디오 이상 탐지를 수행할 수 있다. 예를 들어, 비디오 이상 탐지 장치(100)에 의해 예측된 미래 프레임과 정답 프레임 사이의 차이가 큰 경우, 이상 점수(120)가 높게 산출될 수 있으며, 이상 점수(120)가 임계치 이상으로 높게 산출되면 비디오(110)에서 이상 행위가 탐지된 것으로 판정될 수 있다.

【0047】 일 실시예에 따르면, 이상 점수(120)는 다음의 수학식 1과 같이 산출될 수 있다.

【0048】【수학식 1】

$$Anomaly\ Score = \frac{Scaled\ Score - MIN(Scaled\ Score)}{MAX(Scaled\ Score) - MIN(Scaled\ Score)}$$

【0049】즉, 이상 점수(120)는 변환 점수(scaled score)의 최소-최대 정규화에 의해 산출될 수 있다. 여기서, 변환 점수는 다음의 수학식 2와 같이 산출될 수 있다.

【0051】【수학식 2】

$$Scaled\ Score = \gamma_1 \cdot SL^{l}_{triplet} + \gamma_2 \cdot SL^{m}_{triplet} + \gamma_3 \cdot SL^{Generator}_{MSE} + \gamma_4 \cdot SL^{Destroyer}_{MSE}$$

【0052】 $$SL_{triplet} = \frac{L_{triplet} - \mu(L_{triplet})}{\sigma(L_{triplet})}$$

【0053】 $$SL_{MSE} = \frac{L_{MSE} - \mu(L_{MSE})}{\sigma(L_{MSE})}$$

【0054】여기서, SL은 손실 함수의 변환 값(scaled value)을 나타내고, γ_1 내지 γ_4는 각 항의 영향을 제어하기 위한 초매개변수(hyperparameter)를 나타낼 수 있다. 변환 점수의 첫번째 항은 라벨 특징 정보와 연관된 삼중항 손실(triplet loss)의 변환 값일 수 있으며, 두번째 항은 모션 특징 정보와 연관된 삼중항 손실

의 변환 값일 수 있다. 또한, 세번째 항은 생성자(102)에 의해 출력된 제1 미래 프레임과 정답 프레임 사이의 평균 제곱 오차(mean square error; MSE)의 변환 값을 나타내고, 네번째 항은 파괴자(104)에 의해 출력된 제2 미래 프레임과 정답 프레임 사이의 평균 제곱 오차의 변환 값을 나타낼 수 있다.

【0055】도 1에서는 비디오 이상 탐지 장치(100)가 생성자(102) 및 파괴자(104) 만을 포함하는 것으로 도시되었으나, 이에 한정되지 않는다. 예를 들어, 비디오 이상 탐지 장치(100)는 이상 점수 산출을 위한 별도의 제어부 등을 더 포함할 수 있다. 이와 같은 구성에 의해, 이상 행위를 포함하는 프레임 시퀀스가 입력되는 경우, 생성자(102)는 품질이 낮은 미래 프레임을 생성할 수 있으며, 파괴자(104)는 미래 프레임에서 품질이 낮은 비정상 영역을 더욱 파괴하여 정상 프레임과의 차이를 극대화할 수 있다. 이에 따라, 비디오 이상 탐지 장치(100)는 모델의 높은 일반화 성능과 관계없이 높은 정확도로 비디오 이상 탐지를 수행할 수 있다.

【0057】도 2는 본원 발명의 일 실시예에 따른 생성자가 프레임 특징 정보(232), 라벨 특징 정보(242) 및 모션 특징 정보(252)를 생성하는 예시를 나타내는 도면이다. 상술된 것과 같이, 생성자(도 1의 102)는 비디오의 연속하는 프레임 시퀀스(202)를 획득하고, 획득된 프레임 시퀀스(202)를 기초로 프레임 특징 정보(232), 라벨 특징 정보(242) 및 모션 특징 정보(252)를 추출할 수 있다. 도시된 것과 같이, 생성자는 제1 인코더(230), 제2 인코더(240), 제3 인코더(250) 등을 이용하여 프레임 특징 정보(232), 라벨 특징 정보(242) 및 모션 특징 정보(252)를 추출

할 수 있다.

【0058】 일 실시예에 따르면, 제1 인코더(230)는 연속하는 t개의 프레임 시퀀스($I_1, I_2, \cdots, I_t \in \mathbb{R}^{H \times W \times 3}$)(202)를 입력받아 프레임 특징 정보(232)를 추출할 수 있다. 예를 들어, 제1 인코더(230)는 비디오에 포함된 색상, 엣지(edge), 코너(corner) 등의 시각적인 특징을 기초로 프레임 특징 정보(232)를 추출할 수 있다.

【0059】 일 실시예에 따르면, 제2 인코더(240)는 프레임 시퀀스(202)를 변환하여 생성된 t개의 라벨 시퀀스($I_1^l, I_2^l, \cdots, I_t^l \in \mathbb{R}^{H \times W \times 21}$)(212)를 입력받아 라벨 특징 정보(242)를 추출할 수 있다. 예를 들어, 딥러닝 기반의 제1 모델(210)은 의미론적 분할(semantic segmentation)을 통해 프레임 시퀀스(202)에 포함된 객체들을 라벨링(labeling)하여 분할된 객체들에 대한 정보를 포함하는 라벨 시퀀스(212)를 생성할 수 있다. 여기서, 제1 모델(210)은 DeepLab 기반의 모델(예: DeepLab v3)을 포함할 수 있으나, 이에 한정되지 않는다. 이와 같이, 제2 인코더(240)는 프레임 시퀀스(202)를 그대로 이용하지 않고, 전처리된 라벨 시퀀스(212)를 입력으로 사용하여 각각의 객체들의 특징을 더욱 명확히 나타내는 라벨 특징 정보(242)를 추출할 수 있다.

【0060】 일 실시예에 따르면, 제3 인코더(250)는 프레임 시퀀스(202)를 변환하여 생성된 t-1개의 모션 시퀀스($I_1^m, I_2^m, \cdots, I_{t-1}^m \in \mathbb{R}^{H \times W \times 2}$)(222)를 입력받아 모션 특징 정보(252)를 추출할 수 있다. 예를 들어, 딥러닝 기반의 제2 모델(220)은 광학 흐름 추정(optical flow estimation)을 통해 프레임 시퀀스(202)에 포함된 객

체들의 시간에 따른 동작에 대한 정보를 포함하는 모션 시퀀스(222)를 생성할 수 있다. 여기서, 제2 모델(220)은 FlowNet 기반의 모델(예: FlowNet v2)을 포함할 수 있으나, 이에 한정되지 않는다. 이와 같이, 제3 인코더(250)는 프레임 시퀀스(202)를 그대로 이용하지 않고, 전처리된 모션 시퀀스(222)를 입력으로 사용하여 객체들의 동작을 더욱 명확히 나타내는 모션 특징 정보(252)를 추출할 수 있다.

【0061】이와 같은 구성에 의해, 프레임 특징 정보(232), 라벨 특징 정보(242) 및 모션 특징 정보(252)를 모두 활용하여 프레임 시퀀스(202)의 다음 프레임을 예측하는 경우, 해당 프레임 시퀀스(202)에 어떤 객체들이 존재하고, 각각의 객체가 어떤 동작을 수행하는지에 대한 정보를 참고할 수 있어 비디오 이상 탐지 장치(도 1의 100)는 더욱 정밀하게 미래 프레임을 예측할 수 있다.

【0063】도 3은 본원 발명의 일 실시예에 따른 생성자가 제1 미래 프레임(342)을 생성하는 예시를 나타내는 도면이다. 상술된 것과 같이, 생성자(도 1의 102)는 추출된 프레임 특징 정보(232), 라벨 특징 정보(242) 및 모션 특징 정보(252)를 이용하여 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임(342)을 생성할 수 있다. 도시된 것과 같이, 생성자는 제1 특징 변환 합성곱 블록(feature transform convolutional block)(310), 제2 특징 변환 합성곱 블록(320), 주의 모듈(attention module)(330), 디코더(decoder)(340) 등을 이용하여 제1 미래 프레임(342)을 생성할 수 있다.

【0064】일 실시예에 따르면, 제1 특징 변환 합성곱 블록(310)은 라벨 특징

정보(242)와 유사해지도록 프레임 특징 정보(232)를 변환하여 변환된 라벨 특징 정보(312)를 생성할 수 있다. 또한, 제2 특징 변환 합성곱 블록(320)은 모션 특징 정보(252)와 유사해지도록 프레임 특징 정보(232)를 변환하여 변환된 모션 특징 정보(322)를 생성할 수 있다. 예를 들어, 제1 특징 변환 합성곱 블록(310) 및 제2 특징 변환 합성곱 블록(320)은 다음의 수학식 3의 삼중항 손실(triplet loss) 기반의 함수를 기초로 학습될 수 있다.

【0065】【수학식 3】

$$L(a,p,n) = max\{d(a_i,p_i) - d(a_i,n_i) + margin, 0\}$$

【0066】여기서, d는 유클리드 거리(Euclidean distance)를 나타내고, a_i는 기준점(anchor point)으로서 변환된 라벨 특징 정보(312) 및 변환된 모션 특징 정보(322)를 나타낼 수 있다. 또한, p_i는 긍정적 샘플(positive sample)로서 라벨 특징 정보(242) 및 모션 특징 정보(252)를 나타내고, n_i는 부정적 샘플(negative sample)로서 프레임 특징 정보(232)를 나타낼 수 있다. 즉, 제1 특징 변환 합성곱 블록(310) 및 제2 특징 변환 합성곱 블록(320)은 프레임 특징 정보(232)와 거리가 멀어지고, 라벨 특징 정보(242) 및 모션 특징 정보(252)와 거리가 가까워지도록 프레임 특징 정보(232)를 변환할 수 있다.

【0067】일 실시예에 따르면, 프레임 특징 정보(232), 변환된 라벨 특징 정보(312) 및 변환된 모션 특징 정보(322)는 채널 방향 연결(channel-wise concatenation)을 통해 결합되어 채널 정보를 형성할 수 있다. 이와 같이 형성된 채널 정보는 주의 모듈(330)에 제공될 수 있다. 주의 모듈(330)은 획득된 채널 정보로부터 중요한 채널에 집중하기 위한 모듈을 지칭할 수 있으며, 주의 모듈(330)에 의해 중요 특징이 포함된 특징맵(feature map)이 생성될 수 있다. 이 때, 디코더(340)는 주의 모듈(330)로부터 전달된 특징맵을 기초로 제1 미래 프레임 $\hat{I}_{t+1} \in \mathbb{R}^{H \times W \times 3}$ (342)을 생성할 수 있다.

【0068】이와 같은 구성에 의해, 라벨 특징 정보(242) 및 모션 특징 정보(252)를 그대로 이용하지 않고, 변환된 라벨 특징 정보(312) 및 변환된 모션 특징 정보(322)로 변환하여 사용하는 경우, 생성자는 이상 행위를 포함하는 비정상 데이터에 대해서는 제1 미래 프레임을 정밀하게 생성할 수 없어 정상 데이터와 비정상 데이터 간의 차이를 명확히 할 수 있다.

【0070】도 4는 본원 발명의 일 실시예에 따른 주의 모듈(330)의 구조를 나타내는 예시적인 도면이다. 상술된 것과 같이, 주의 모듈(330)은 프레임 특징 정보, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보를 채널 방향으로 연결한 채널 정보를 획득하여 특징맵을 생성할 수 있다.

【0071】 도시된 예에서, 프레임 특징 정보, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보 $X \in \mathbb{R}^{H \times W \times 3C}$는 3 x 3의 합성곱 레이어(convolutional layer)를 통과하여 채널 방향으로 연결된 채널 정보 $U \in \mathbb{R}^{H \times W \times C}$로 변환될 수 있다. 그리고 나서, 채널 정보에 포함된 각 채널에 대해 전역 풀링(global pooling)이 수행되어 벡터 $Z \in \mathbb{R}^{1 \times 1 \times C}$가 생성될 수 있다.

【0072】 이 경우, 스칼라 인자(scalar factor) $S \in \mathbb{R}^{1 \times 1 \times C}$는 다음의 수학식 4와 같이 산출될 수 있다.

【0073】 【수학식 4】

$$S = \sigma(\delta(Z \cdot W_1) \cdot W_2)$$

【0074】 여기서, σ는 시그모이드 활성 함수를 나타내고, δ는 ReLU 활성 함수를 나타낼 수 있다. 또한, $W_1 \in \mathbb{R}^{C \times \frac{C}{r}}$ 및 $W_2 \in \mathbb{R}^{\frac{C}{r} \times C}$는 2개의 연속하는 완전 연결 레이어(fully connected layer)의 가중치 행렬을 나타내고, r은 감속비(reduction ratio)를 나타낼 수 있다.

【0075】 그리고 나서, 각각의 채널에 대해 S와 U의 값을 곱하여 재보정된 특징맵을 포함하는 텐서(tensor) $\widetilde{U} \in \mathbb{R}^{H \times W \times C}$가 생성될 수 있다. 이와 같은 구성에 의해, 주의 모듈(330)은 특징 중 중요한 채널만을 추출함으로써 미래 프레임의 생성

능력을 향상시키고, 잘못된 예측을 효과적으로 감소시킬 수 있다.

【0077】도 5는 본원 발명의 일 실시예에 따른 파괴자(104)가 제2 미래 프레임(522)을 생성하는 예시를 나타내는 도면이다. 상술된 것과 같이, 파괴자(104)는 제1 미래 프레임(342) 중 품질이 임계치 미만인 비정상 영역을 결정하고, 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임(522)을 생성할 수 있다. 도시된 것과 같이, 파괴자(104)는 인코더(510) 및 디코더(520)를 포함할 수 있다. 여기서, 인코더(510)는 U-net 구조의 생성자의 인코더들과 유사한 구조를 가질 수 있으며, 디코더(520)는 U-net 구조의 생성자의 디코더와 동일한 구조를 가질 수 있다.

【0078】일 실시예에 따르면, 파괴자(104)는 제1 미래 프레임(342)에서 정상 영역의 패치를 제1 미래 프레임(342)의 패치로 복원하고, 비정상 영역의 패치를 영 벡터(zero vector)로 변환하여 제2 미래 프레임(522)을 생성할 수 있다. 여기서, 정상 영역의 패치는 이미지의 품질이 임계치 이상인 영역을 나타내고, 비정상 영역의 패치는 이미지의 품질이 임계치 미만인 영역을 나타낼 수 있다. 즉, 생성자는 프레임에서 이상 행위에 대응하는 영역의 품질은 낮게 예측하고, 정상에 대응하는 영역의 품질은 높게 예측하여 정상 영역과 비정상 영역을 구분할 수 있다. 또한, 디코더(520)는 이상 행위에 대응하는 비정상 영역의 품질이 더 낮아지도록 파괴하여 정상 영역과 비정상 영역의 대비를 강조할 수 있다.

【0080】도 6은 본원 발명의 일 실시예에 따른 생성자(102)의 학습 과정을 나타내는 예시적인 도면이다. 일 실시예에 따르면, 생성자(102)는 4개의 손실 함수를 기초로 학습될 수 있다. 예를 들어, 생성자(102)는 삼중항 손실(triplet loss), 강도 손실(intensity loss), 경사 손실(gradient loss) 및 적대적 손실(adversarial loss)을 기초로 학습될 수 있다.

【0081】상술된 것과 같이, 생성자(102)는 라벨 특징 정보와 유사해지도록 프레임 특징 정보를 변환하여 변환된 라벨 특징 정보를 생성하고, 모션 특징 정보와 유사해지도록 프레임 특징 정보를 변환하여 변환된 모션 특징 정보를 생성할 수 있다. 이 때, 생성자(102)는 삼중항 손실을 기초로 기준점과 긍정적 샘플 사이의 거리를 최소화하고, 기준점과 부정적 샘플 사이의 거리를 최대화하도록 학습될 수 있다. 예를 들어, 삼중항 손실은 다음의 수학식 5와 같이 정의될 수 있다.

【0082】【수학식 5】

$$L^{l}_{triplet}(Z_f, Z_l, Z'_l) = max\{d(Z'_l, Z_l) - d(Z'_l, Z_f) + \alpha, 0\}$$

【0083】 $$L^{m}_{triplet}(Z_f, Z_m, Z'_m) = max\{d(Z'_m, Z_m) - d(Z'_m, Z_f) + \alpha, 0\}$$

【0084】

$$L_{triplet}(Z_f, Z_l, Z_m, Z'_l, Z'_m) = L^{l}_{triplet}(Z_f, Z_l, Z'_l) + L^{m}_{triplet}(Z_f, Z_m, Z'_m)$$

【0085】 여기서, Z_f 는 부정적 샘플인 프레임 특징 정보일 수 있고, Z_l 및 Z_m 는 긍정적 샘플인 라벨 특징 정보와 모션 특징 정보일 수 있다. 또한, Z_l' 및 Z_m' 는 기준점인 변환된 라벨 특징 정보와 변환된 모션 특징 정보일 수 있다. 또한, d는 유클리드 거리를 나타내고, α 는 기준점과 부정적 샘플 사이의 거리를 증가시키기 위한 초매개변수를 나타낼 수 있다.

【0086】 삼중항 손실을 이용하는 경우, 생성자(102)(또는 생성자(102)에 포함된 특징 변환 합성곱 블록)는 정상 프레임 시퀀스에 대해서는 기준점과 긍정적 샘플 간의 거리를 최소화하도록 학습될 수 있다. 그러나, 학습 후 비정상 프레임 시퀀스가 입력되는 경우, 기준점과 긍정적 샘플 간의 거리는 증가할 수 있으며, 이에 따라 생성자(102)는 저품질의 미래 프레임을 예측할 수 있다.

【0087】 일 실시예에 따르면, 생성자(102)는 예측된 미래 프레임(예: 제1 미래 프레임)과 정답 프레임이 유사해지도록 강도 손실 및 경사 손실을 기초로 학습될 수 있다. 여기서, 강도 손실은 RGB 공간에 있는 모든 픽셀(pixel)의 유사성을 보장할 수 있으며, 경사 손실은 예측된 미래 프레임과 정답 프레임에 포함된 각각의 영역의 주변 픽셀들의 차이를 줄일 수 있다. 예를 들어, 강도 손실 및 경사 손실은 다음의 수학식 6 및 7과 같이 정의될 수 있다.

【0088】【수학식 6】

$$L_{int}(\hat{I},I) = \|\hat{I} - I\|_2^2$$

【0089】【수학식 7】

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

【0090】여기서, $L_{int}(\hat{I},I)$는 강도 손실을 나타내고, $L_{gd}(\hat{I},I)$는 경사 손실을 나타낼 수 있으며, I는 정답 프레임을 나타내고, $\hat{I}$는 예측된 미래 프레임을 나타낼 수 있다. 또한, i 및 j는 프레임 프레임의 공간 인덱스(spatial indices)를 나타낼 수 있다.

【0091】일 실시예에 따르면, 생성자(102)는 생성적 적대 신경망(generative adversarial network; GAN)을 기초로 학습될 수 있다. 여기서, 생성적 적대 신경망은 생성자(102)와 구별자(예: pix2pix의 PatchGAN 모델의 구별자 등)(610)의 경쟁을 통해 이미지 또는 프레임의 생성 및 분류를 학습하는 기계 학습 방식 또는 그러한 기계 학습 모델을 지칭할 수 있다. 생성적 적대 신경망에서, 생성자(102)는 정답 프레임과 더욱 가까운 프레임을 생성하도록 학습되고, 구별자(610)는 생성자(102)에 의해 생성된 프레임을 정답 프레임과 구분하도록 학습될 수 있다. 예를 들

어, 생성자(102)는 다음의 수학식 8의 적대적 손실을 기초로 구별자(610)와의 적대적 학습을 통해 미래 프레임을 정밀하게 예측하도록 학습될 수 있다.

【0092】【수학식 8】

$$L_{adv}^{D}(\hat{I},I)=\sum_{i,j}\frac{1}{2}L_{MSE}(D(I)_{i,j},1)+\sum_{i,j}\frac{1}{2}L_{MSE}\left(D(\hat{I})_{i,j},0\right)$$

【0093】

$$L_{adv}^{G}(\hat{I})=\sum_{i,j}\frac{1}{2}L_{MSE}\left(D(\hat{I})_{i,j},1\right)$$

【0094】여기서, $L_{adv}^{D}(\hat{I},I)$ 및 $L_{adv}^{G}(\hat{I})$는 적대적 손실을 나타내고, MSE는 평균 제곱 오차를 나타낼 수 있다. 수학식 8을 통해 구별자(610)는 정답 프레임에 대해 1을 출력하고, 예측된 미래 프레임에 대해 0을 출력하도록 학습될 수 있으며, 생성자(102)는 구별자(610)가 예측된 미래 프레임에 대해 1을 출력하도록 학습될 수 있다.

【0095】상술된 것과 같이, 생성자(102)의 학습 시 4가지의 손실 함수가 종합적으로 사용될 수 있으며, 생성자(102)의 최종 손실 함수는 다음의 수학식 9와 같이 정의될 수 있다.

【0096】【수학식 9】

$$L_{generator}=\delta_{int}L_{int}(\hat{I},I)+\delta_{gd}L_{gd}(\hat{I},I)+\delta_{adv}L_{adv}^{G}(\hat{I})+\delta_{tri}L_{triplet}(Z_f,Z_l,Z_m,Z_l',Z_m')$$

【0097】 여기서, δ_{int}, δ_{gd}, δ_{adv} 및 δ_{tri}는 각각의 손실의 영향을 제어하기 위한 초매개변수를 나타낼 수 있다.

【0099】 도 7은 본원 발명의 일 실시예에 따른 파괴자(104)의 학습 과정을 나타내는 예시적인 도면이다. 일 실시예에 따르면, 파괴자(104)는 프레임을 구성하는 패치 사이의 거리를 기준으로 정의된 파괴자 손실(destroyer loss)을 기초로 학습될 수 있다.

【0100】 학습 과정에서, 비정상 프레임을 만들기 위해, 프레임을 패치 단위로 분할하고, 분할된 패치 중 일부에 노이즈(noise)를 주입하여 과정이 수행될 수 있다. 즉, 생성자에 의해 예측된 제1 미래 프레임은 제1 세트의 패치(702)로 분할될 수 있으며, 분할되어 생성된 제1 세트의 패치(702) 중 일부에 노이즈를 주입하는 노이징(710)이 수행되어 제2 세트의 패치(704)가 생성될 수 있다.

【0101】 파괴자(104)는 제2 세트의 패치(704)를 입력받아 제3 세트의 패치(706)를 생성할 수 있다. 그리고 나서, 파괴자(104)는 제3 세트의 패치(706)를 기초로 품질이 임계치 이상인 정상 영역 및 품질이 임계치 미만인 비정상 영역을 결정할 수 있다. 이 때, 파괴자(104)는 파괴자 손실을 기초로 정상 영역의 패치를 제1 미래 프레임의 패치로 복원하고, 비정상 영역의 패치를 영 벡터(708)로 변환하여 제2 미래 프레임을 생성할 수 있다. 이 경우, 파괴자 손실은 다음의 수학식 10과 같이 정의될 수 있다.

【0102】【수학식 10】

$$L_{destroyer}(diff,R,Z,A)=\sum_{p=1}^{P}(diff_p\cdot\|Z_p-R_p\|_2^2+(1-diff_p)\cdot\|A_p-R_p\|_2^2)$$

【0103】여기서, A_p는 제1 세트의 패치(702)를 나타내고, R_p는 제3 세트의 패치(706)를 나타내며, Z_p는 영 벡터(708)를 나타낼 수 있다. 또한, $diff_p$는 다음의 수학식 11과 같이 정의될 수 있다.

【0104】【수학식 11】

$$diff_p=MIN(\lambda\left(1-SSIM\left(N_p,GT_p\right)\right),1)$$

【0105】여기서, GT_p는 정답 프레임이 분할되어 생성된 패치를 나타내고, N_p는 노이즈가 주입된 제2 세트의 패치(704)를 나타낼 수 있다. 또한, λ는 초매개변수를 나타낼 수 있다. $diff_p$는 0에서 1 사이의 값을 가질 수 있으며, 정상 영역과 비정상 영역을 구별하는 척도로 사용될 수 있다.

【0106】상술된 수학식에서 노이즈가 주입된 패치와 정상 패치 사이의 품질 차이를 확인하기 위해, 구조적 유사성 지수 측정(structural similarity index measure; SSIM)이 사용될 수 있다. 이에 따라, 각 패치에 대해 $diff_p$의 값이 1에

가까울수록 비정상 영역을 나타낼 수 있고, $diff_p$의 값이 0에 가까울수록 정상 영역을 나타낼 수 있다.

【0107】이와 같은 학습 과정을 통해, 파괴자(104)는 학습 과정에서 비정상 영역을 영 벡터로 변환하여 파괴할 수 있으며, 비정상 영역에 대해 정상 프레임과의 차이를 더욱 강조하여 비디오 이상 탐지의 성능을 효과적으로 향상시킬 수 있다.

【0109】도 8은 본원 발명의 일 실시예에 따른 특징 정보에 대한 확률적 임베딩의 예시를 나타내는 도면이다. 일 실시예에 따르면, 제1 이미지(810)는 정상 데이터의 t-분포 확률적 임베딩(t-distributed stochastic neighbor embedding; t-SNE)을 나타내고, 제2 이미지(820)는 비정상 데이터의 t-분포 확률적 임베딩을 나타낼 수 있다. 여기서, t-분포 확률적 임베딩은 고차원 공간의 데이터를 2차원이나 3차원의 저차원으로 표현하는 차원 축소 기법 및 시각화 기법을 지칭할 수 있다.

【0110】도시된 제1 이미지(810) 및 제2 이미지(920)에서 파란색의 f 포인트는 프레임 특징 정보를 나타내고, 주황색의 l 포인트는 라벨 특징 정보를 나타내고, 초록색의 m 포인트는 모션 특징 정보를 나타낼 수 있다. 또한, f-to-l 포인트는 변환된 라벨 특징 정보를 나타내고, f-to-m은 변환된 모션 특징 정보를 나타낼 수 있다.

【0111】 제1 이미지(810)를 참조하면, 보라색의 변환된 모션 특징 정보(f-to-m)가 파란색의 프레임 특징 정보(f)와 멀고, 초록색의 모션 특징 정보(m)와 가까운 분포를 갖는 것을 확인할 수 있다. 반면에, 제2 이미지(820)를 참조하면, 변환된 모션 특징 정보(f-to-m)와 모션 특징 정보(m) 사이의 거리가 모션 특징 정보(f-to-m)와 프레임 특징 정보(f) 사이의 거리와 유사한 것을 확인할 수 있다.

【0112】 즉, 비디오 탐지 장치의 생성자는 정상 데이터가 입력되는 경우, 프레임 특징 정보의 변환을 잘 수행할 수 있으나, 비정상 데이터가 입력되는 경우, 프레임 특징 정보의 변환을 잘 수행하지 못할 수 있다. 따라서, 이상 행위와 연관된 비정상 데이터가 입력되는 경우, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보는 프레임의 노이즈로 작용하여 생성자에 의해 생성되는 미래 프레임의 품질이 낮아질 수 있다.

【0114】 도 9는 본원 발명의 일 실시예에 따른 미래 프레임에 대한 확률적 임베딩의 예시를 나타내는 도면이다. 일 실시예에 따르면, 제1 이미지(910)는 생성자의 출력값의 t-분포 확률적 임베딩을 나타내고, 제2 이미지(920)는 파괴자의 출력값의 t-분포 확률적 임베딩을 나타낼 수 있다. 여기서, 파란색 포인트는 정상 데이터(normal data)를 나타내고, 주황색 포인트는 비정상 데이터(anomaly data)를 나타낼 수 있다.

【0115】 제1 이미지(910) 및 제2 이미지(920)를 참조하면, 생성자 및 파괴자 모두 정상 데이터와 비정상 데이터를 잘 구분하였으나, 파괴자에 의해 출력된 값에서 정상 데이터와 비정상 데이터 사이의 거리가 더 먼 것을 확인할 수 있다. 즉, 생성자에 의해 생성되는 제1 미래 프레임을 그대로 이용하지 않고, 파괴자에 의해 비정상 영역을 파괴하는 동작을 통해 제2 미래 프레임을 생성하는 경우, 정상 데이터와 비정상 데이터가 더욱 명확히 구분될 수 있다.

【0117】 도 10은 본원 발명의 일 실시예에 따른 F2LM 기반의 비디오 이상 탐지 방법(1000)의 예시를 나타내는 흐름도이다. F2LM 기반의 비디오 이상 탐지 방법(1000)은 프로세서(예를 들어, 컴퓨팅 장치의 적어도 하나의 프로세서)에 의해 수행될 수 있다. F2LM 기반의 비디오 이상 탐지 방법(1000)은 프로세서가 비디오의 연속하는 프레임 시퀀스를 획득함으로써 개시될 수 있다(S1010).

【0118】 프로세서는 획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출할 수 있다(S1020). 예를 들어, 프로세서는 프레임 시퀀스에 대한 의미론적 분할을 수행하여 라벨 시퀀스를 생성하고, 생성된 라벨 시퀀스를 이용하여 라벨 특징 정보를 추출할 수 있다. 또한, 프로세서는 프레임 시퀀스에 대한 광학 흐름 추정을 수행하여 모션 시퀀스를 생성하고, 생성된 모션 시퀀스를 이용하여 모션 특징 정보를 추출할 수 있다.

【0119】프로세서는 추출된 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 이용하여 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성할 수 있다(S1030). 예를 들어, 프로세서는 라벨 특징 정보와 유사해지도록 프레임 특징 정보를 변환하여 변환된 라벨 특징 정보를 생성하고, 모션 특징 정보와 유사해지도록 프레임 특징 정보를 변환하여 변환된 모션 특징 정보를 생성하고, 프레임 특징 정보, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보를 이용하여 제1 미래 프레임을 생성할 수 있다. 이 경우, 프로세서는 프레임 특징 정보, 변환된 라벨 특징 정보 및 변환된 모션 특징 정보를 채널 방향으로 연결하여 채널 정보를 획득하고, 획득된 채널 정보로부터 추출된 특징맵을 이용하여 제1 미래 프레임을 생성할 수 있다.

【0120】프로세서는 생성된 제1 미래 프레임 중 품질이 임계치 미만인 비정상 영역을 결정하고, 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성할 수 있다(S1040). 예를 들어, 프로세서는 제1 미래 프레임을 제1 세트의 패치로 분할하고, 제1 세트의 패치 중 일부의 패치에 노이즈를 주입하여 제2 세트의 패치를 생성할 수 있다. 그리고 나서, 프로세서는 생성된 제2 세트의 패치를 기초로 품질이 임계치 이상인 정상 영역 및 품질이 임계치 미만인 비정상 영역을 결정하고, 정상 영역의 패치를 제1 미래 프레임의 패치로 복원하고, 비정상 영역의 패치를 영 벡터로 변환하여 제2 미래 프레임을 생성할 수 있다.

【0121】프로세서는 제1 미래 프레임과 정답 프레임 사이의 거리 및 제2 미래 프레임과 정답 프레임 사이의 거리를 기초로 이상 점수를 산출할 수 있다. 그리

고 나서, 프로세서는 산출된 이상 점수가 임계치 이상인지 여부를 기초로 비디오 이상 탐지를 수행할 수 있다.

【0123】도 11은 본원 발명의 일 실시예에 따른 컴퓨팅 장치(1100)의 내부 구성을 나타내는 블록도이다. 일 실시예에 따르면, 컴퓨팅 장치(1100)는 상술된 비디오 이상 탐지 장치(도 1의 100)를 포함할 수 있다. 예를 들어, 컴퓨팅 장치(1100)는 메모리(1110), 프로세서(1120), 통신 모듈(1130) 및 입출력 인터페이스(1140)를 포함할 수 있으며, 도 11에 도시된 바와 같이, 컴퓨팅 장치(1100)는 통신 모듈(1130)을 이용하여 네트워크를 통해 정보 및/또는 데이터를 통신할 수 있도록 구성될 수 있다.

【0124】메모리(1110)는 비-일시적인 임의의 컴퓨터 판독 가능한 기록매체를 포함할 수 있다. 일 실시예에 따르면, 메모리(1110)는 RAM(random access memory), ROM(read only memory), 디스크 드라이브, SSD(solid state drive), 플래시 메모리(flash memory) 등과 같은 비소멸성 대용량 저장 장치(permanent mass storage device)를 포함할 수 있다. 다른 예로서, ROM, SSD, 플래시 메모리, 디스크 드라이브 등과 같은 비소멸성 대용량 저장 장치는 메모리와는 구분되는 별도의 영구 저장 장치로서 컴퓨팅 장치(1100)에 포함될 수 있다. 또한, 메모리(1110)에는 운영체제와 적어도 하나의 프로그램 코드가 저장될 수 있다.

【0125】 이러한 소프트웨어 구성요소들은 메모리(1110)와는 별도의 컴퓨터에서 판독 가능한 기록매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록매체는 이러한 컴퓨팅 장치(1100)에 직접 연결가능한 기록 매체를 포함할 수 있는데, 예를 들어, 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록매체를 포함할 수 있다. 다른 예로서, 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록매체가 아닌 통신 모듈(1130)을 통해 메모리(1110)에 로딩될 수도 있다. 예를 들어, 적어도 하나의 프로그램은 개발자들 또는 애플리케이션의 설치 파일을 배포하는 파일 배포 시스템이 통신 모듈(1130)을 통해 제공하는 파일들에 의해 설치되는 컴퓨터 프로그램에 기반하여 메모리(1110)에 로딩될 수 있다.

【0126】 프로세서(1120)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(1110) 또는 통신 모듈(1130)에 의해 다른 사용자 단말(미도시) 또는 다른 외부 시스템으로 제공될 수 있다.

【0127】 통신 모듈(1130)은 네트워크를 통해 사용자 단말(미도시)과 컴퓨팅 장치(1100)가 서로 통신하기 위한 구성 또는 기능을 제공할 수 있으며, 컴퓨팅 장치(1100)가 외부 시스템(일례로 별도의 클라우드 시스템 등)과 통신하기 위한 구성 또는 기능을 제공할 수 있다. 일례로, 컴퓨팅 장치(1100)의 프로세서(1120)의 제어에 따라 제공되는 제어 신호, 명령, 데이터 등이 통신 모듈(1130)과 네트워크를 거쳐 사용자 단말 및/또는 외부 시스템의 통신 모듈을 통해 사용자 단말 및/또는 외

부 시스템으로 전송될 수 있다.

【0128】 또한, 컴퓨팅 장치(1100)의 입출력 인터페이스(1140)는 컴퓨팅 장치(1100)와 연결되거나 컴퓨팅 장치(1100)가 포함할 수 있는 입력 또는 출력을 위한 장치(미도시)와의 인터페이스를 위한 수단일 수 있다. 도 11에서는 입출력 인터페이스(1140)가 프로세서(1120)와 별도로 구성된 요소로서 도시되었으나, 이에 한정되지 않으며, 입출력 인터페이스(1140)가 프로세서(1120)에 포함되도록 구성될 수 있다. 컴퓨팅 장치(1100)는 도 11의 구성요소들보다 더 많은 구성요소들을 포함할 수 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다.

【0129】 컴퓨팅 장치(1100)의 프로세서(1120)는 복수의 사용자 단말 및/또는 복수의 외부 시스템으로부터 수신된 정보 및/또는 데이터를 관리, 처리 및/또는 저장하도록 구성될 수 있다.

【0131】 상술된 방법 및/또는 다양한 실시예들은, 디지털 전자 회로, 컴퓨터 하드웨어, 펌웨어, 소프트웨어 및/또는 이들의 조합으로 실현될 수 있다. 본원 발명의 다양한 실시예들은 데이터 처리 장치, 예를 들어, 프로그래밍 가능한 하나 이상의 프로세서 및/또는 하나 이상의 컴퓨팅 장치에 의해 실행되거나, 컴퓨터 판독 가능한 기록 매체 및/또는 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 상술된 컴퓨터 프로그램은 컴파일된 언어 또는 해석된 언어를 포함하여 임의의 형태의 프로그래밍 언어로 작성될 수 있으며, 독립 실행형

프로그램, 모듈, 서브 루틴 등의 임의의 형태로 배포될 수 있다. 컴퓨터 프로그램은 하나의 컴퓨팅 장치, 동일한 네트워크를 통해 연결된 복수의 컴퓨팅 장치 및/또는 복수의 상이한 네트워크를 통해 연결되도록 분산된 복수의 컴퓨팅 장치를 통해 배포될 수 있다.

【0132】상술된 방법 및/또는 다양한 실시예들은, 입력 데이터를 기초로 동작하거나 출력 데이터를 생성함으로써, 임의의 기능, 함수 등을 처리, 저장 및/또는 관리하는 하나 이상의 컴퓨터 프로그램을 실행하도록 구성된 하나 이상의 프로세서에 의해 수행될 수 있다. 예를 들어, 본원 발명의 방법 및/또는 다양한 실시예는 FPGA(Field Programmable Gate Array) 또는 ASIC(Application Specific Integrated Circuit)과 같은 특수 목적 논리 회로에 의해 수행될 수 있으며, 본원 발명의 방법 및/또는 실시예들을 수행하기 위한 장치 및/또는 시스템은 FPGA 또는 ASIC와 같은 특수 목적 논리 회로로서 구현될 수 있다.

【0133】컴퓨터 프로그램을 실행하는 하나 이상의 프로세서는, 범용 목적 또는 특수 목적의 마이크로 프로세서 및/또는 임의의 종류의 디지털 컴퓨팅 장치의 하나 이상의 프로세서를 포함할 수 있다. 프로세서는 읽기 전용 메모리, 랜덤 액세스 메모리의 각각으로부터 명령 및/또는 데이터를 수신하거나, 읽기 전용 메모리와 랜덤 액세스 메모리로부터 명령 및/또는 데이터를 수신할 수 있다. 본 발명에서, 방법 및/또는 실시예들을 수행하는 컴퓨팅 장치의 구성 요소들은 명령어들을 실행하기 위한 하나 이상의 프로세서, 명령어들 및/또는 데이터를 저장하기 위한 하나 이상의 메모리 디바이스를 포함할 수 있다.

【0134】일 실시예에 따르면, 컴퓨팅 장치는 데이터를 저장하기 위한 하나 이상의 대용량 저장 장치와 데이터를 주고받을 수 있다. 예를 들어, 컴퓨팅 장치는 자기 디스크(magnetic disc) 또는 광 디스크(optical disc)로부터 데이터를 수신하거나/수신하고, 자기 디스크 또는 광 디스크로 데이터를 전송할 수 있다. 컴퓨터 프로그램과 연관된 명령어들 및/또는 데이터를 저장하기에 적합한 컴퓨터 판독 가능한 저장 매체는, EPROM(Erasable Programmable Read-Only Memory), EEPROM(Electrically Erasable PROM), 플래시 메모리 장치 등의 반도체 메모리 장치를 포함하는 임의의 형태의 비 휘발성 메모리를 포함할 수 있으나, 이에 한정되지 않는다. 예를 들어, 컴퓨터 판독 가능한 저장 매체는 내부 하드 디스크 또는 이동식 디스크와 같은 자기 디스크, 광 자기 디스크, CD-ROM 및 DVD-ROM 디스크를 포함할 수 있다.

【0135】사용자와의 상호 작용을 제공하기 위해, 컴퓨팅 장치는 정보를 사용자에게 제공하거나 디스플레이하기 위한 디스플레이 장치(예를 들어, CRT (Cathode Ray Tube), LCD(Liquid Crystal Display) 등) 및 사용자가 컴퓨팅 장치 상에 입력 및/또는 명령 등을 제공할 수 있는 포인팅 장치(예를 들어, 키보드, 마우스, 트랙볼 등)를 포함할 수 있으나, 이에 한정되지 않는다. 즉, 컴퓨팅 장치는 사용자와의 상호 작용을 제공하기 위한 임의의 다른 종류의 장치들을 더 포함할 수 있다. 예를 들어, 컴퓨팅 장치는 사용자와의 상호 작용을 위해, 시각적 피드백, 청각 피드백 및/또는 촉각 피드백 등을 포함하는 임의의 형태의 감각 피드백을 사용자에게 제공할 수 있다. 이에 대해, 사용자는 시각, 음성, 동작 등의 다양한 제스처를 통해 컴

퓨팅 장치로 입력을 제공할 수 있다.

【0136】본 발명에서, 다양한 실시예들은 백엔드 구성 요소(예: 데이터 서버), 미들웨어 구성 요소(예: 애플리케이션 서버) 및/또는 프론트 엔드 구성 요소를 포함하는 컴퓨팅 시스템에서 구현될 수 있다. 이 경우, 구성 요소들은 통신 네트워크와 같은 디지털 데이터 통신의 임의의 형태 또는 매체에 의해 상호 연결될 수 있다. 예를 들어, 통신 네트워크는 LAN(Local Area Network), WAN(Wide Area Network) 등을 포함할 수 있다.

【0137】본 명세서에서 기술된 예시적인 실시예들에 기반한 컴퓨팅 장치는, 사용자 디바이스, 사용자 인터페이스(UI) 디바이스, 사용자 단말 또는 클라이언트 디바이스를 포함하여 사용자와 상호 작용하도록 구성된 하드웨어 및/또는 소프트웨어를 사용하여 구현될 수 있다. 예를 들어, 컴퓨팅 장치는 랩톱(laptop) 컴퓨터와 같은 휴대용 컴퓨팅 장치를 포함할 수 있다. 추가적으로 또는 대안적으로, 컴퓨팅 장치는, PDA(Personal Digital Assistants), 태블릿 PC, 게임 콘솔(game console), 웨어러블 디바이스(wearable device), IoT(internet of things) 디바이스, VR(virtual reality) 디바이스, AR(augmented reality) 디바이스 등을 포함할 수 있으나, 이에 한정되지 않는다. 컴퓨팅 장치는 사용자와 상호 작용하도록 구성된 다른 유형의 장치를 더 포함할 수 있다. 또한, 컴퓨팅 장치는 이동 통신 네트워크 등의 네트워크를 통한 무선 통신에 적합한 휴대용 통신 디바이스(예를 들어, 이동 전화, 스마트 전화, 무선 셀룰러 전화 등) 등을 포함할 수 있다. 컴퓨팅 장치는, 무선 주파수(RF; Radio Frequency), 마이크로파 주파수(MWF; Microwave Frequency)

및/또는 적외선 주파수(IRF; Infrared Ray Frequency)와 같은 무선 통신 기술들 및/또는 프로토콜들을 사용하여 네트워크 서버와 무선으로 통신하도록 구성될 수 있다.

【0138】 본 발명에서 특정 구조적 및 기능적 세부 사항을 포함하는 다양한 실시예들은 예시적인 것이다. 따라서, 본원 발명의 실시예들은 상술된 것으로 한정되지 않으며, 여러 가지 다른 형태로 구현될 수 있다. 또한, 본 발명에서 사용된 용어는 일부 실시예를 설명하기 위한 것이며 실시예를 제한하는 것으로 해석되지 않는다. 예를 들어, 단수형 단어 및 상기는 문맥상 달리 명확하게 나타내지 않는 한 복수형도 포함하는 것으로 해석될 수 있다.

【0139】 본 발명에서, 달리 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함하여 본 명세서에서 사용되는 모든 용어는 이러한 개념이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 갖는다. 또한, 사전에 정의된 용어와 같이 일반적으로 사용되는 용어들은 관련 기술의 맥락에서의 의미와 일치하는 의미를 갖는 것으로 해석되어야 한다.

【0140】 본 명세서에서는 본 발명이 일부 실시예들과 관련하여 설명되었지만, 본원 발명의 발명이 속하는 기술분야의 통상의 기술자가 이해할 수 있는 본원 발명의 범위를 벗어나지 않는 범위에서 다양한 변형 및 변경이 이루어질 수 있다. 또한, 그러한 변형 및 변경은 본 명세서에 첨부된 특허청구의 범위 내에 속하는 것으로 생각되어야 한다.

【부호의 설명】

【0141】100: 비디오 이상 탐지 장치

102: 생성자

104: 파괴자

110: 비디오

120: 이상 점수

【청구범위】

【청구항 1】

적어도 하나의 프로세서에 의해 수행되는 비디오 이상 탐지 방법으로서,

비디오의 연속하는 프레임 시퀀스(frame sequence)를 획득하는 단계;

상기 획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하는 단계;

상기 추출된 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 이용하여 상기 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성하는 단계; 및

상기 생성된 제1 미래 프레임 중 품질이 임계치 미만인 비정상 영역을 결정하고, 상기 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 2】

제1항에 있어서,

상기 획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하는 단계는,

상기 프레임 시퀀스에 대한 의미론적 분할(semantic segmentation)을 수행하

여 라벨 시퀀스(label sequence)를 생성하는 단계; 및

상기 생성된 라벨 시퀀스를 이용하여 상기 라벨 특징 정보를 추출하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 3】

제1항에 있어서,

상기 획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하는 단계는,

상기 프레임 시퀀스에 대한 광학 흐름 추정(optical flow estimation)을 수행하여 모션 시퀀스(motion sequence)를 생성하는 단계; 및

상기 생성된 모션 시퀀스를 이용하여 상기 모션 특징 정보를 추출하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 4】

제1항에 있어서,

상기 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성하는 단계는,

상기 라벨 특징 정보와 유사해지도록 상기 프레임 특징 정보를 변환하여 변환된 라벨 특징 정보를 생성하는 단계;

상기 모션 특징 정보와 유사해지도록 상기 프레임 특징 정보를 변환하여 변환된 모션 특징 정보를 생성하는 단계; 및

상기 프레임 특징 정보, 상기 변환된 라벨 특징 정보 및 상기 변환된 모션 특징 정보를 이용하여 상기 제1 미래 프레임을 생성하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 5】

제4항에 있어서,

상기 프레임 특징 정보, 상기 변환된 라벨 특징 정보 및 상기 변환된 모션 특징 정보를 이용하여 상기 제1 미래 프레임을 생성하는 단계는,

상기 프레임 특징 정보, 상기 변환된 라벨 특징 정보 및 상기 변환된 모션 특징 정보를 채널(channel) 방향으로 연결하여 채널 정보를 획득하는 단계; 및

상기 획득된 채널 정보로부터 추출된 특징맵(feature map)을 이용하여 상기 제1 미래 프레임을 생성하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 6】

제1항에 있어서,

상기 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성하는 단계는,

상기 제1 미래 프레임을 제1 세트의 패치(patch)로 분할하는 단계;

상기 제1 세트의 패치 중 일부의 패치에 노이즈(noise)를 주입하여 제2 세트의 패치를 생성하는 단계;

상기 생성된 제2 세트의 패치를 기초로 품질이 임계치 이상인 정상 영역 및 품질이 임계치 미만인 비정상 영역을 결정하는 단계; 및

상기 정상 영역의 패치를 상기 제1 미래 프레임의 패치로 복원하고, 상기 비정상 영역의 패치를 영 벡터(zero vector)로 변환하여 상기 제2 미래 프레임을 생성하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 7】

제1항에 있어서,

상기 제1 미래 프레임과 정답 프레임 사이의 거리 및 상기 제2 미래 프레임과 정답 프레임 사이의 거리를 기초로 이상 점수(anomaly score)를 산출하는 단계; 및

상기 산출된 이상 점수가 임계치 이상인지 여부를 기초로 비디오 이상 탐지를 수행하는 단계를 더 포함하는

비디오 이상 탐지 방법.

【청구항 8】

제1항 내지 제7항 중 어느 한 항에 따른 비디오 이상 탐지 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램.

【청구항 9】

비디오의 연속하는 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하고, 추출된 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 이용하여 상기 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성하는 생성자(generator); 및

상기 생성된 제1 미래 프레임 중 품질이 임계치 미만인 비정상 영역을 결정하고, 상기 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성하는 파괴자(destroyer)를 포함하는

비디오 이상 탐지 장치.

【청구항 10】

제9항에 있어서,

상기 생성자는,

상기 프레임 특징 정보를 추출하기 위한 제1 인코더(encoder), 상기 라벨 특징 정보를 추출하기 위한 제2 인코더 및 상기 모션 특징 정보를 추출하기 위한 제3 인코더를 포함하는

비디오 이상 탐지 장치.

【청구항 11】

제10항에 있어서,

상기 생성자는,

상기 프레임 시퀀스에 대한 의미론적 분할을 수행하여 라벨 시퀀스를 생성하는 딥러닝 기반의 제1 모델을 포함하는

비디오 이상 탐지 장치.

【청구항 12】

제10항에 있어서,

상기 생성자는,

상기 프레임 시퀀스에 대한 광학 흐름 추정을 수행하여 모션 시퀀스를 생성

하는 딥러닝 기반의 제2 모델을 포함하는

비디오 이상 탐지 장치.

【청구항 13】

제9항에 있어서,

상기 생성자는,

상기 라벨 특징 정보와 유사해지도록 상기 프레임 특징 정보를 변환하여 변환된 라벨 특징 정보를 생성하는 제1 특징 변환 합성곱 블록(feature transform convolutional block); 및

상기 모션 특징 정보와 유사해지도록 상기 모션 특징 정보를 변환하여 변환된 모션 특징 정보를 생성하는 제2 특징 변환 합성곱 블록을 포함하는

비디오 이상 탐지 장치.

【청구항 14】

제13항에 있어서,

상기 생성자는,

상기 프레임 특징 정보, 상기 변환된 라벨 특징 정보 및 상기 변환된 모션 특징 정보를 채널 방향으로 연결하여 채널 정보를 획득하는 주의 모듈(attention module)을 포함하는

비디오 이상 탐지 장치.

【청구항 15】

제14항에 있어서,

상기 생성자는,

상기 획득된 채널 정보로부터 추출된 특징맵을 이용하여 상기 제1 미래 프레임을 생성하는 디코더(decoder)를 포함하는

비디오 이상 탐지 장치.

【청구항 16】

제13항에 있어서,

상기 생성자는,

상기 프레임 특징 정보를 부정적 샘플(negative sample)로 하고, 상기 라벨 특징 정보 및 상기 모션 특징 정보를 긍정적 샘플(positive sample)로 하는 삼중항 손실(triplet loss) 기반의 함수를 기초로 학습되는

비디오 이상 탐지 장치.

【청구항 17】

제9항에 있어서,

상기 생성자는,

상기 생성된 제1 미래 프레임과 상기 프레임 시퀀스의 다음 프레임으로 결정된 정답 프레임이 유사해지도록 강도 손실(intensity loss) 기반의 함수, 경사 손실(gradient loss) 기반의 함수 및 적대적 손실(adversarial loss)을 기초로 학습되는

비디오 이상 탐지 장치.

【청구항 18】

제9항에 있어서,

상기 파괴자는,

상기 제1 미래 프레임 중 품질이 임계치 이상인 정상 영역은 복원하고, 품질이 임계치 미만인 비정상 영역은 영 벡터로 변환하여 상기 제2 미래 프레임을 생성하는

비디오 이상 탐지 장치.

【요약서】

【요약】

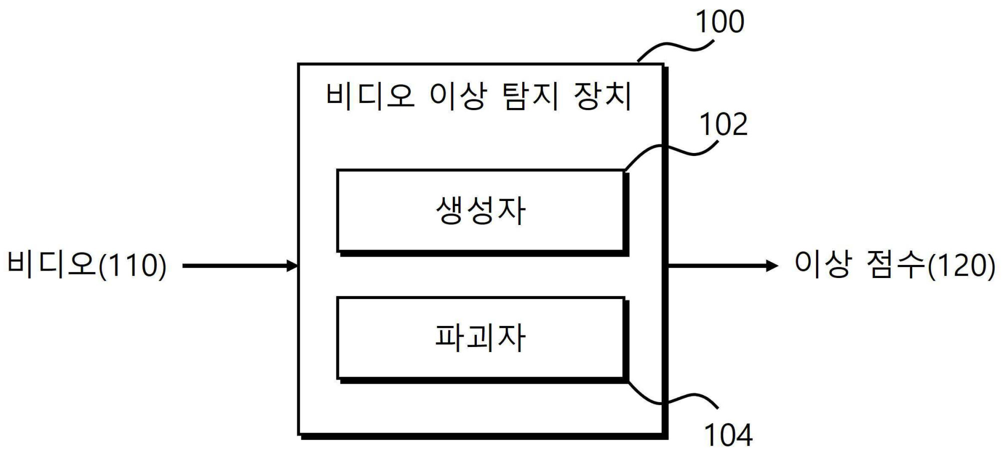
본원 발명의 F2LM 기반의 비디오 이상 탐지 방법은, 비디오의 연속하는 프레임 시퀀스를 획득하는 단계, 획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출하는 단계, 추출된 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 이용하여 프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성하는 단계 및 생성된 제1 미래 프레임 중 품질이 임계치 미만인 비정상 영역을 결정하고, 결정된 비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성하는 단계를 포함한다.

【대표도】

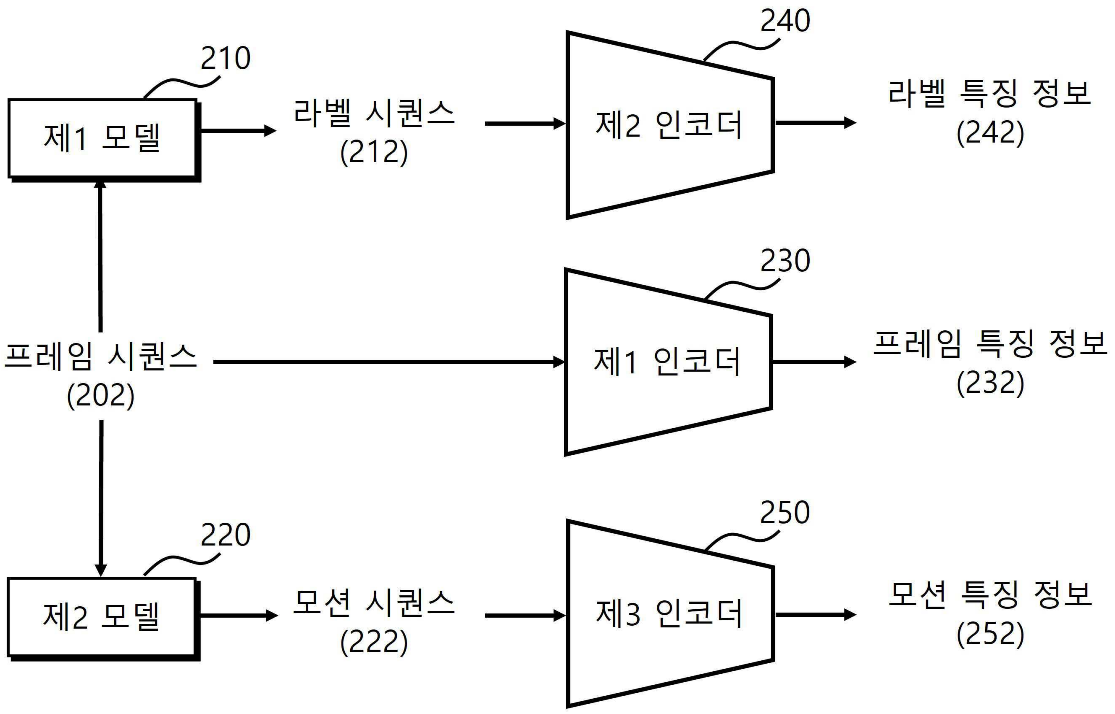
도 1

【도면】

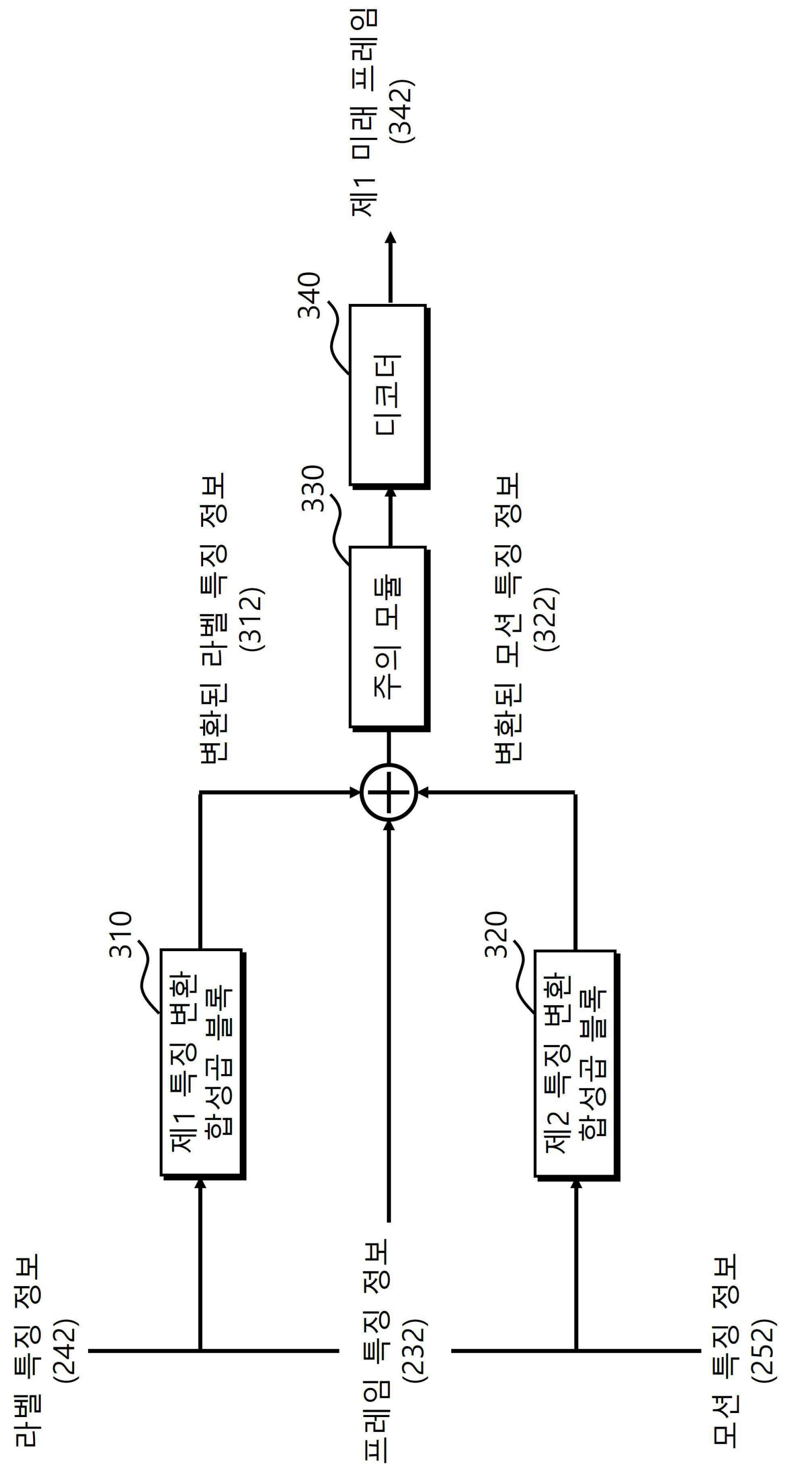
【도 1】



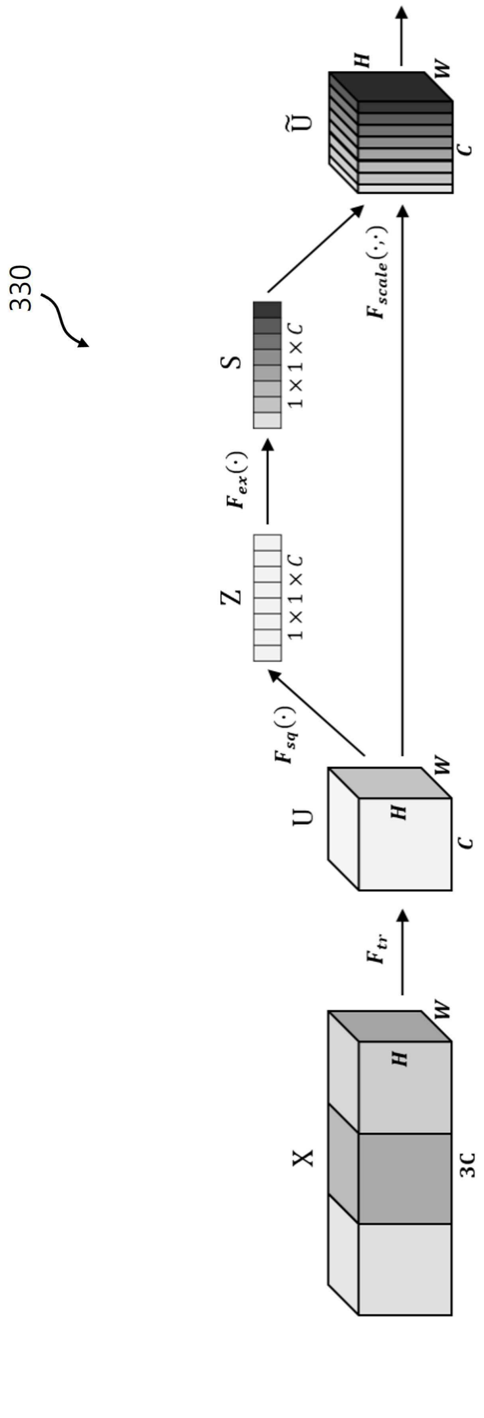
【도 2】


240
210
제1 모델
라벨 시퀀스
(212)
제2 인코더
라벨 특징 정보
(242)
230
프레임 시퀀스
(202)
제1 인코더
프레임 특징 정보
(232)
250
220
제2 모델
모션 시퀀스
(222)
제3 인코더
모션 특징 정보
(252)

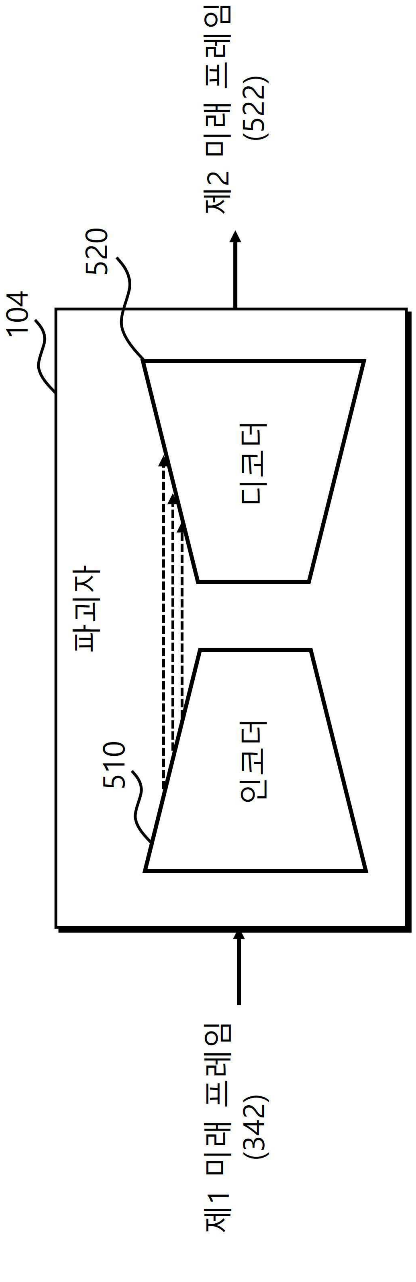
【도 3】


라벨 특징 정보
(242)
프레임 특징 정보
(232)
모션 특징 정보
(252)
310
제1 특징 변환
합성곱 블록
320
제2 특징 변환
합성곱 블록
변환된 라벨 특징 정보
(312)
변환된 모션 특징 정보
(322)
330
주의 모듈
340
디코더
제1 미래 프레임
(342)

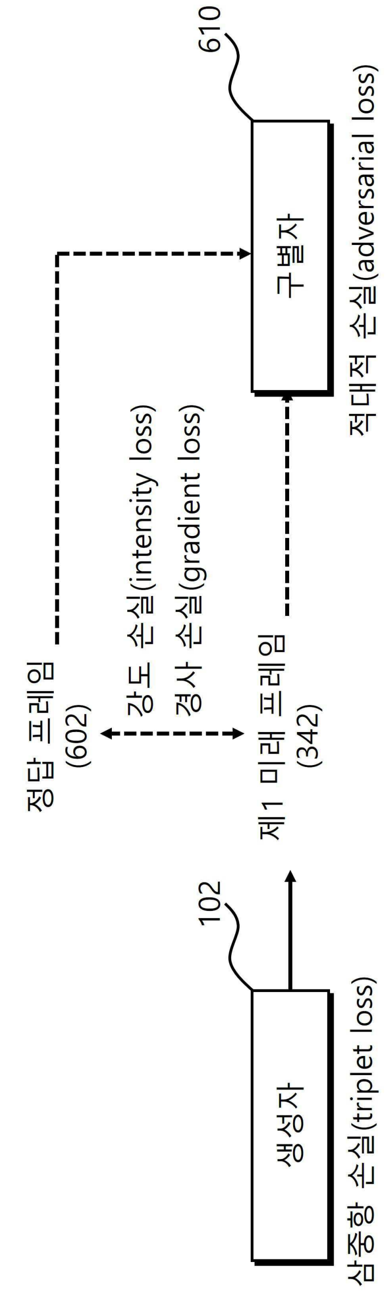
【도 4】



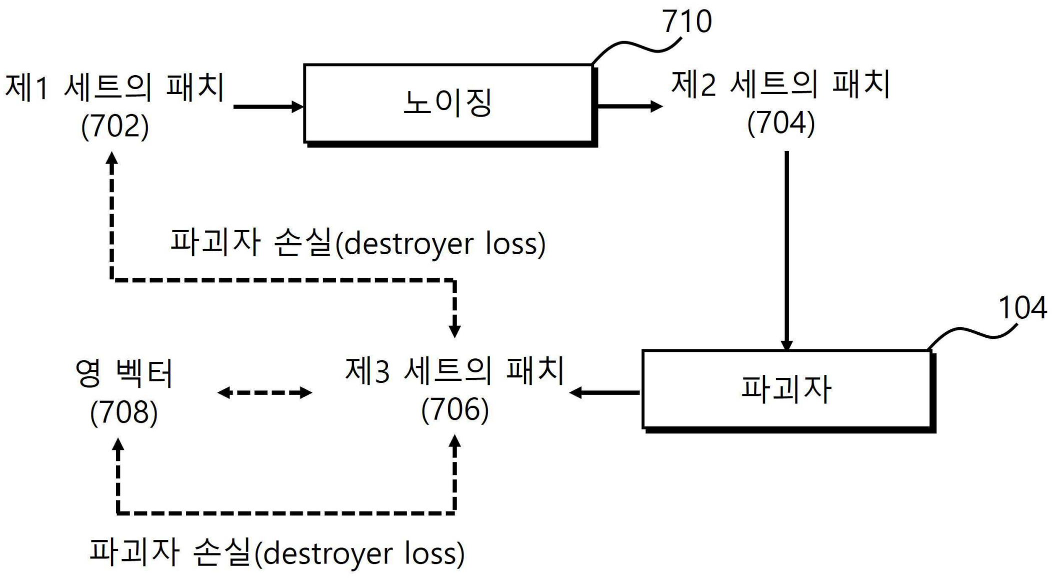
【도 5】



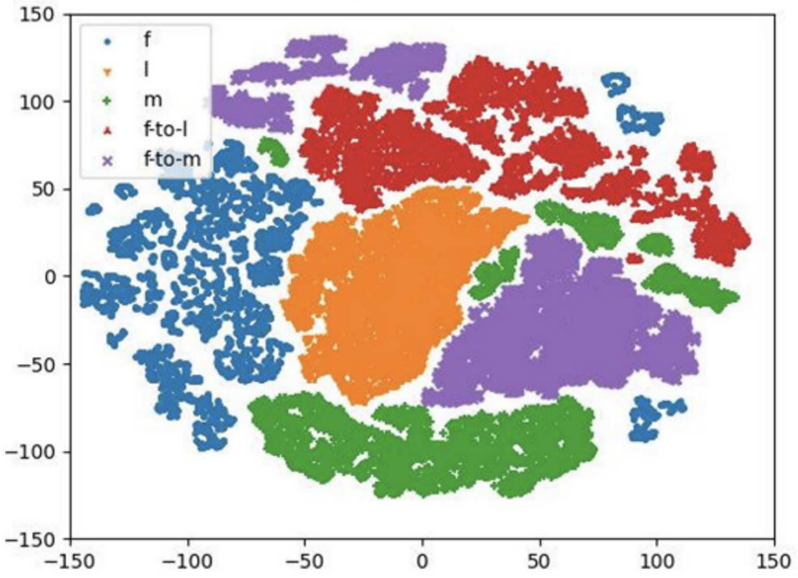
【도 6】


610
구별자
적대적 손실(adversarial loss)
정답 프레임
(602)
강도 손실(intensity loss)
경사 손실(gradient loss)
제1 미래 프레임
(342)
102
생성자
삼중항 손실(triplet loss)

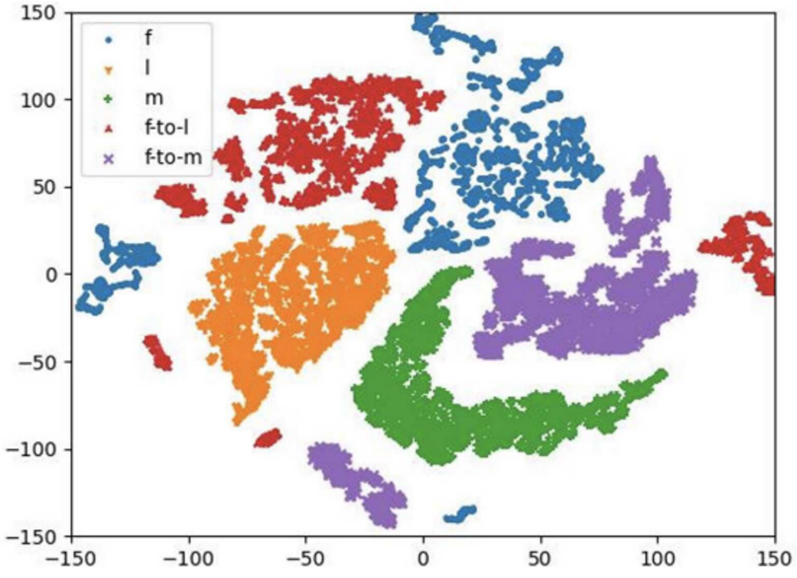
【도 7】


710
제1 세트의 패치
(702)
노이징
제2 세트의 패치
(704)
파괴자 손실(destroyer loss)
104
영 벡터
(708)
제3 세트의 패치
(706)
파괴자
파괴자 손실(destroyer loss)

【도 8】

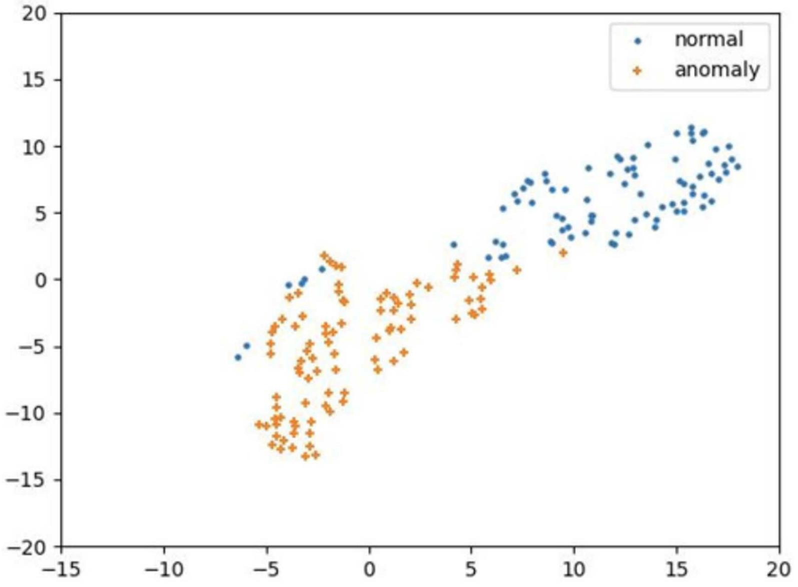

f
l
m
f-to-l
f-to-m
150
100
50
0
−50
−100
−150
−150
−100
−50
0
50
100
150

810


f
l
m
f-to-l
f-to-m
150
100
50
0
−50
−100
−150
−150
−100
−50
0
50
100
150

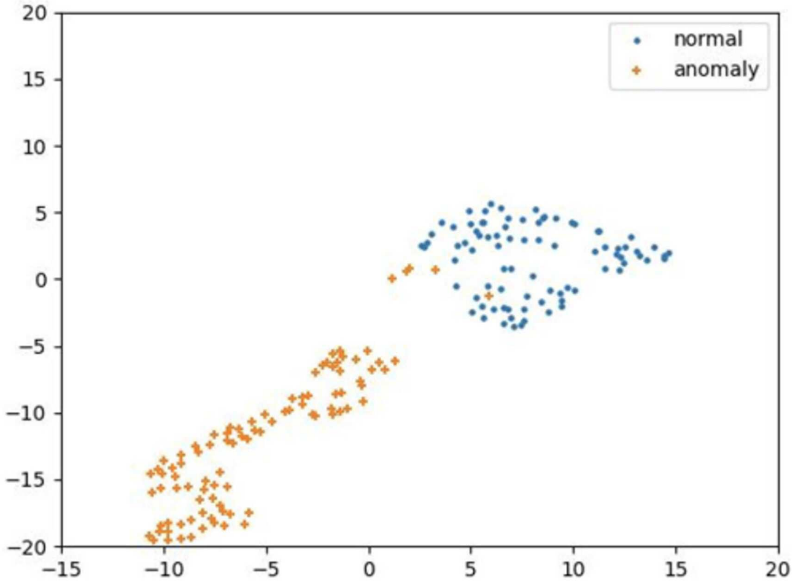
820

【도 9】



normal
anomaly
20
15
10
5
0
−5
−10
−15
−20
−15
−10
−5
0
5
10
15
20

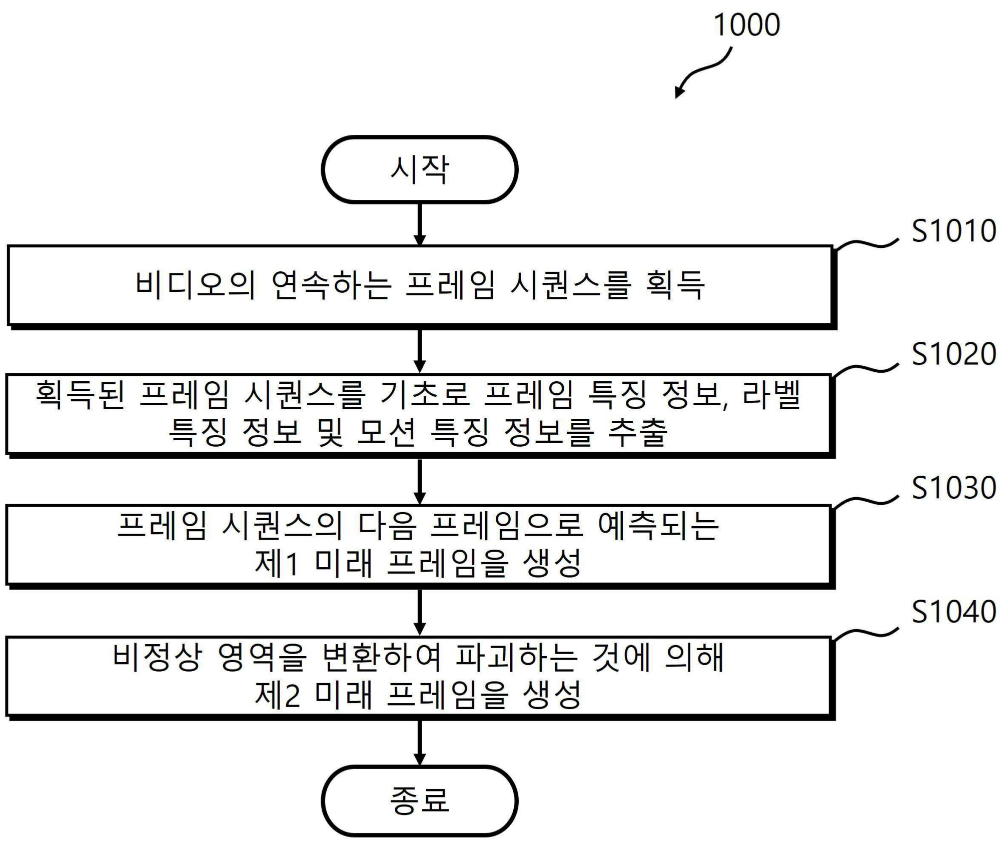
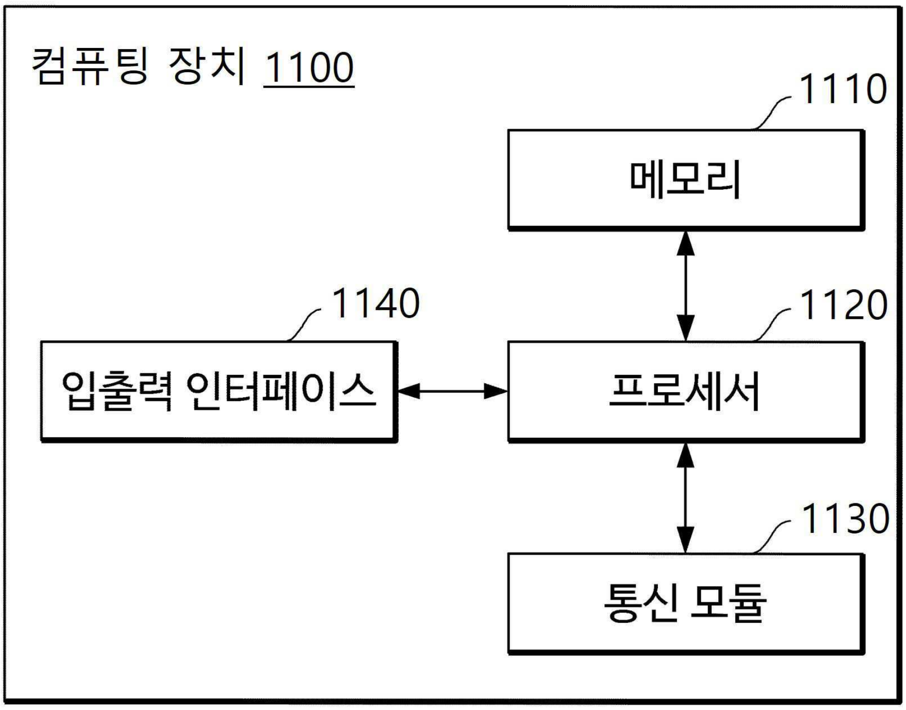
910



normal
anomaly
20
15
10
5
0
−5
−10
−15
−20
−15
−10
−5
0
5
10
15
20

920

【도 10】


1000
시작
S1010
비디오의 연속하는 프레임 시퀀스를 획득
S1020
획득된 프레임 시퀀스를 기초로 프레임 특징 정보, 라벨 특징 정보 및 모션 특징 정보를 추출
S1030
프레임 시퀀스의 다음 프레임으로 예측되는 제1 미래 프레임을 생성
S1040
비정상 영역을 변환하여 파괴하는 것에 의해 제2 미래 프레임을 생성
종료

【도 11】