【서지사항】

【서류명】 특허출원서

【참조번호】 SDP20244393

【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】 다중모달 확산 기반의 비디오 이상 탐지 방법 및 장치

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

【발명자】

【성명】 박상현

【성명의 영문표기】 SANGHYUN PARK

【주민등록번호】 670101-1XXXXXX

【우편번호】 08004

【주소】 서울특별시 양천구 오목로 300, 204동 3701호

【발명자】

【성명】 이기정

【성명의 영문표기】 KIJUNG LEE

【주민등록번호】 981013-1XXXXXX

【우편번호】 03725

【주소】 서울특별시 서대문구 연희로10길 24-4, 지동 201호

【발명자】

【성명】 조영완

【성명의 영문표기】 YOUNGWAN JO

【주민등록번호】 990312-1XXXXXX

【우편번호】 03726

【주소】 서울특별시 서대문구 성산로17길 5, 402호

【발명자】

【성명】 안성현

【성명의 영문표기】 SUNGHYUN AHN

【주민등록번호】 000104-3XXXXXX

【우편번호】 21368

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

【출원언어】 국어

【심사청구】 청구

【이 발명을 지원한 국가연구개발사업】

【과제고유번호】 1711198526

【과제번호】 00229822

【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】	50	면	0 원
【우선권주장료】	0	건	0 원
【심사청구료】	14	항	880,000 원

【합계】 926,000원

【감면사유】 전담조직(50%감면)[1]

【감면후 수수료】 463,000 원

【발명의 설명】

【발명의 명칭】

다중모달 확산 기반의 비디오 이상 탐지 방법 및 장치{VIDEO ANOMALY DETECTION METHOD AND APPARATUS BASED ON MULTIMODAL DIFFUSION}

【기술분야】

【0001】 본원 발명은 다중모달 확산 기반의 비디오 이상 탐지 방법 및 장치에 관한 것으로, 복수의 특징을 이용하여 비디오의 이상을 탐지하는 방법 및 장치에 관한 것이다.

【발명의 배경이 되는 기술】

【0002】 최근 AI(artificial intelligence) 등의 기술이 발달함에 따라 CCTV 등의 감시 카메라로부터 수집된 영상을 통해 안전 사고의 발생 등의 이상 행위를 인식하기 위한 다양한 기술이 개발되고 있다. 예를 들어, 정상일 때의 영상과 이상 행위가 발생했을 때의 영상을 구분하도록 AI 모델의 학습 및 개발이 수행되고 있다. 그러나, 이상 행위의 발생 빈도가 낮아 이러한 AI 모델을 학습시키기 위한 영상 데이터를 충분히 확보하는데 어려움이 있다. 또한, 현재의 대부분의 모델은 프레임 이미지 등의 단편적인 정보만을 활용할 수 있어 정확도가 떨어지는 문제가 있다.

【발명의 내용】

【해결하고자 하는 과제】

【0003】본원 발명은 상기와 같은 문제점을 해결하기 위한 다중모달 확산 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】

【0004】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.

【0005】본원 발명의 일 실시예에 따르면, 적어도 하나의 프로세서에 의해 수행되는 다중모달 확산 기반의 비디오 이상 탐지 방법은, 복수의 프레임을 포함하는 비디오 데이터를 획득하는 단계, 복수의 프레임에 포함된 객체를 탐지하는 단계, 탐지된 객체에 대한 시각적 특징 벡터, 텍스트 특징 벡터 및 모션 특징 벡터를 포함하는 다중모달 특징 벡터를 추출하는 단계, 시각적 특징 벡터에 노이즈를 주입하여 노이즈 벡터를 생성하는 단계, 확산 모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터 및 모션 특징 벡터를 조건으로 활용하여 노이즈가 제거된 복원 벡터를 생성하는 단계 및 시각적 특징 벡터와 복원 벡터를 비교하여 비디오 데이터에 대한 이상 탐지를 수행하는 단계를 포함한다.

【0006】본원 발명의 일 실시예에 따르면, 다중모달 특징 벡터를 추출하는 단계는, 탐지된 객체와 연관된 정보를 학습된 I3D 기반의 모델에 제공하여, 객체에 대한 시각적 특징 벡터를 추출하는 단계를 포함한다.

【0007】본원 발명의 일 실시예에 따르면, 다중모달 특징 벡터를 추출하는 단계는, 탐지된 객체와 연관된 정보를 BERT 기반의 모델에 제공하여 객체에 대한 설명을 나타내는 캡션을 생성하는 단계 및 생성된 캡션을 학습된 SimCSE 기반의 모델에 제공하여 객체에 대한 설명에 대응하는 텍스트 특징 벡터를 추출하는 단계를 포함한다.

【0008】본원 발명의 일 실시예에 따르면, 다중모달 특징 벡터를 추출하는 단계는, 탐지된 객체와 연관된 정보를 학습된 HRNet 기반의 모델에 제공하여 객체에 대응하는 뼈대 정보를 추출하는 단계 및 추출된 뼈대 정보를 이용하여 객체의 동작을 나타내는 모션 특징 벡터를 추출하는 단계를 포함한다.

【0009】본원 발명의 일 실시예에 따르면, 추출된 뼈대 정보를 이용하여 객체의 동작을 나타내는 모션 특징 벡터를 추출하는 단계는, 추출된 뼈대 정보를 학습된 PoseConv3D 기반의 모델에 제공하여 모션 특징 벡터를 추출하는 단계를 포함한다.

【0010】본원 발명의 일 실시예에 따르면, 시각적 특징 벡터에 노이즈를 주입하여 노이즈 벡터를 생성하는 단계는, 시간 단계의 범위에 따라 결정된 양의 가우시안 노이즈를 시각적 특징 벡터에 주입하여 노이즈 벡터를 생성하는 단계를 포함한다.

【0011】본원 발명의 일 실시예에 따르면, 확산 모델은 제1 확산 모델 및 제2 확산 모델을 포함한다. 노이즈가 제거된 복원 벡터를 생성하는 단계는, 제1 확산

모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터를 조건 벡터로 활용하여 노이즈 벡터에 포함된 적어도 일부의 노이즈를 제거하는 제1 복원 단계 및 제2 확산 모델에 노이즈 벡터를 입력하고, 모션 특징 벡터를 조건 벡터로 활용하여 노이즈 벡터에 포함된 적어도 일부의 노이즈를 제거하는 제2 복원 단계를 포함한다.

【0012】 본원 발명의 일 실시예에 따르면, 노이즈가 제거된 복원 벡터를 생성하는 단계는, 제1 복원 단계 및 제2 복원 단계를 반복 수행하여 노이즈가 제거된 복원 벡터를 생성하는 단계를 더 포함한다.

【0013】 본원 발명의 일 실시예에 따르면, 비디오 데이터에 대한 이상 탐지를 수행하는 단계는, 시각적 특징 벡터와 복원 벡터 사이의 거리에 따른 이상 점수를 산출하는 단계 및 산출된 이상 점수가 임계값 이상인지 여부를 기초로 비디오 데이터에 대한 이상 탐지를 수행하는 단계를 포함한다.

【0014】 본원 발명의 일 실시예에 따르면, 이상 점수를 산출하는 단계는, 시각적 특징 벡터와 복원 벡터 사이의 평균 제곱 오차를 이용하여 거리에 따른 이상 점수를 산출하는 단계를 포함한다.

【0015】 본원 발명의 일 실시예에 따르면, 확산 모델은, 복수의 디노이징 주의 블록을 포함하는 인코더, 병목 및 디코더를 포함한다.

【0016】 본원 발명의 일 실시예에 따르면, 디노이징 주의 블록은, 스킵 연결로 연결된 복수의 선형 레이어를 포함하는 잔차 블록 및 자기 주의 레이어, 교차 주의 레이어 및 피드 포워드 네트워크를 포함하는 트랜스포머 블록을 포함한다.

【0017】본원 발명의 일 실시예에 따른 상술된 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램이 제공된다.

【0018】본원 발명의 일 실시예에 따른 컴퓨팅 장치는, 통신 모듈, 메모리 및 메모리와 연결되고, 메모리에 포함된 컴퓨터 판독 가능한 적어도 하나의 프로그램을 실행하도록 구성된 적어도 하나의 프로세서를 포함한다. 적어도 하나의 프로그램은, 복수의 프레임을 포함하는 비디오 데이터를 획득하고, 복수의 프레임에 포함된 객체를 탐지하고, 탐지된 객체에 대한 시각적 특징 벡터, 텍스트 특징 벡터 및 모션 특징 벡터를 포함하는 다중모달 특징 벡터를 추출하고, 시각적 특징 벡터에 노이즈를 주입하여 노이즈 벡터를 생성하고, 확산 모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터 및 모션 특징 벡터를 조건 벡터로 활용하여 노이즈가 제거된 복원 벡터를 생성하고, 시각적 특징 벡터와 복원 벡터를 비교하여 비디오 데이터에 대한 이상 탐지를 수행하기 위한 명령어들을 포함한다.

【발명의 효과】

【0019】본원 발명의 다양한 실시예에서 컴퓨팅 장치는 다중모달 특징 벡터를 상호 보완적으로 활용하여 비디오 이상 탐지의 성능을 향상시킬 수 있다.

【0020】본원 발명의 다양한 실시예에서 트랜스포머 블록 및 잔차 블록의 연산 시 텍스트 특징 벡터 및/또는 모션 특징 벡터가 조건으로 참조되는 것에 의해, 컴퓨팅 장치는 객체의 시각적 특징과 함께 객체를 설명하는 텍스트 및/또는 객체의 동작을 모두 참조하여 효과적으로 노이즈 제거 및 벡터 복원을 수행할 수 있다.

【0021】본원 발명의 다양한 실시예에서 하나의 확산 모델을 이용하는 것이 아닌 조건이 상이한 제1 확산 모델 및 제2 확산 모델을 모두 이용함으로써 복원 성능이 향상될 수 있으며, 이에 따라 더 높은 정확도로 비디오 이상 탐지가 수행될 수 있다.

【도면의 간단한 설명】

【0022】본원 발명의 실시예들은, 이하 설명하는 첨부 도면들을 참조하여 설명될 것이며, 여기서 유사한 참조 번호는 유사한 요소들을 나타내지만, 이에 한정되지는 않는다.

도 1는 본원 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 나타내는 기능적인 블록도이다.

도 2는 본원 발명의 일 실시예에 따른 다중모달 특징 벡터가 추출되는 과정을 나타내는 블록도이다.

도 3은 본원 발명의 일 실시예에 따른 확산 모델의 구조를 나타내는 예시적인 도면이다.

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

도 5는 본원 발명의 일 실시예에 따른 제1 확산 모델 및 제2 확산 모델에 의해 복원 과정이 수행되는 예시를 나타내는 도면이다.

도 6은 본원 발명의 제2 실시예에 따른 제1 확산 모델 및 제2 확산 모델에

의해 복원 과정이 수행되는 예시를 나타내는 도면이다.

도 7은 본원 발명의 제3 실시예에 따른 제1 확산 모델 및 제2 확산 모델에 의해 복원 과정이 수행되는 예시를 나타내는 도면이다.

도 8은 본원 발명의 제4 실시예에 따른 제1 확산 모델 및 제2 확산 모델에 의해 복원 과정이 수행되는 예시를 나타내는 도면이다.

도 9은 본원 발명의 일 실시예에 따른 다중모달 확산 기반의 비디오 이상 탐지 방법의 예시를 나타내는 흐름도이다.

도 10은 본원 발명의 일 실시예에 따른 컴퓨팅 장치의 하드웨어 구성을 나타내는 블록도이다.

【발명을 실시하기 위한 구체적인 내용】

【0023】 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.

【0024】 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.

【0025】 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.

【0026】 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 선택하였으나, 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.

【0027】 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.

【0028】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.

【0029】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.

【0030】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.

【0031】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.

【0032】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.

【0033】본 개시에서, "다중모달(multimodal)"은 시각 데이터, 텍스트 데이터 등의 다양한 유형의 데이터를 함께 처리하는 것을 지칭할 수 있다.

【0034】본 개시에서, "확산 모델(diffusion model)"은 데이터에 노이즈를 조금씩 더해가거나, 노이즈로부터 조금씩 복원해가는 과정을 통해 데이터를 생성하는 생성형 모델을 지칭할 수 있다. 예를 들어, 확산 모델은 텍스트 특징 벡터를 조건(condition)으로 활용하는 제1 확산 모델과 모션 특징 벡터를 조건으로 활용하는 제2 확산 모델을 포함할 수 있다. 여기서, 제1 확산 모델과 제2 확산 모델은 학습 시 별도로 구분되어 학습되나, 추론 시 함께 사용될 수 있다.

【0035】본 개시에서, "시각적 특징 벡터(visual feature vector)"는 객체의 색상, 형상 등의 외형적인 정보를 나타내는 벡터이고, "텍스트 특징 벡터(text feature vector)"는 객체를 설명하는 텍스트를 나타내는 벡터이고, "모션 특징 벡터(motion feature vector)"는 객체의 동작을 나타내는 벡터를 지칭일 수 있다. 또한, 본 개시에서, "노이즈 벡터(noise vector)"는 시각적 특징 벡터에 적어도 일부의 노이즈가 주입된 상태의 벡터를 지칭하는 것으로, 확산 프로세스에 의해 생성되는 벡터와 확산 모델을 충분히 거치지 않아 아직 노이즈가 남아있는 벡터를 모두 포함할 수 있다. 또한, 본 개시에서, "복원 벡터(restoration vector)"는 시각적 특징 벡터에 주입된 노이즈가 모두 제거된 형태의 벡터를 지칭할 수 있다.

【0037】도 1는 본원 발명의 일 실시예에 따른 컴퓨팅 장치(100)의 내부 구성을 나타내는 기능적인 블록도이다. 일 실시예에 따르면, 컴퓨팅 장치(100)는 비디오 이상 탐지를 수행하기 위한 임의의 장치로서, 객체 탐지부(110), 다중모달 특징 추출부(120), 노이즈 주입부(130), 벡터 복원부(140), 이상 탐지부(150) 등을 포함할 수 있다. 예를 들어, 컴퓨팅 장치(100)는 CCTV 등의 감시 카메라 등으로부터 복수의 프레임을 포함하는 비디오 데이터를 획득하는 경우, 해당 비디오 데이터에서 이상 행위가 발생하는지 여부를 탐지할 수 있다.

【0038】일 실시예에 따르면, 컴퓨팅 장치(100)는 비디오 데이터에 포함된 각각의 객체가 이상 행위를 수행하는지 여부를 검출하기 위해, 먼저 해당 비디오 데이터를 구성하는 복수의 프레임에 포함된 객체를 탐지할 수 있다. 예를 들어, 객체 탐지부(110)는 임의의 객체 추적 알고리즘(예: object detector, multi object tracker 등) 및/또는 기계학습 모델을 통해 복수의 프레임에 포함된 객체를 탐지할 수 있다. 이 경우, 연속하는 프레임 상에서 다음의 수학식 1과 같은 객체 트랙렛(object tracklet)이 추출될 수 있다.

【0039】**【수학식 1】**

$$\{O_n | O_n \in \mathbb{R}^{l\times 3\times H\times W}\}_{n=1}^{N}$$

【0040】여기서, O_n은 객체 트랙렛을 나타내고, N은 객체의 수를 나타내며, L, H 및 W는 각각 객체 트랙렛의 길이, 높이 및 너비를 나타낼 수 있다. 여기서,

객체 트랙렛은 복수의 프레임 상에서 검출된 동일한 객체의 시간에 따른 이동을 나타내는 배열을 포함할 수 있다. 즉, 객체 탐지부(110)는 각각의 프레임에서 추출된 객체를 연관시켜 해당 객체의 시간에 따른 동작을 탐지할 수 있다.

【0041】일 실시예에 따르면, 객체 탐지부(110)는 추출된 프레임 레벨(frame-level)의 객체 트랙렛을 세그먼트 레벨(segment-level)의 객체 트랙렛으로 변환할 수 있다. 여기서, 세그먼트는 16개의 연속하는 프레임으로 구성될 수 있으나, 이에 한정되지 않는다. 세그먼트 레벨의 객체 트랙렛으로 변환하는 경우, 객체 트랙렛은 $\mathbb{R}^{S\times 16\times 3\times H\times W}$ ($S = l/16$)의 형상을 가질 수 있다. 이와 같이 세그먼트 레벨로 변형된 객체 트랙렛은 탐지된 객체와 연관된 정보로서, 다중모달 특징 추출을 위한 정보로 사용될 수 있다.

【0042】일 실시예에 따르면, 다중모달 특징 추출부(120)는 탐지된 객체에 대한 시각적 특징 벡터, 텍스트 특징 벡터 및 모션 특징 벡터를 포함하는 다중모달 특징 벡터를 추출할 수 있다. 예를 들어, 다중모달 특징 추출부(120)는 탐지된 객체와 연관된 정보를 학습된 I3D(inflated 3D ConvNet) 기반의 모델에 제공하여, 객체에 대한 시각적 특징 벡터를 추출할 수 있다. 여기서, I3D 기반의 모델은 객체의 색상, 형태 등의 시각적인 정보를 추출하기 위한 모델을 지칭할 수 있다.

【0043】추가적으로, 다중모달 특징 추출부(120)는 탐지된 객체와 연관된 정보를 BERT(bidirectional encoder representations from transformers) 기반의 모델(예: SwinBERT 모델)에 제공하여 객체에 대한 설명을 나타내는 캡션(caption)을

생성할 수 있다. 여기서, 캡션은 탐지된 객체를 해설하기 위한 텍스트로서, 예를 들어, '자전거를 타는 사람'의 객체 트랙렛이 입력으로 제공된 경우, "a man is riding a bicycle with a bicycle on a street."과 같은 캡션이 추출될 수 있다. 이 경우, 다중모달 특징 추출부(120)는 생성된 캡션을 학습된 SimCSE(simple contrastive learning of sentence embeddings) 기반의 모델에 제공하여 객체에 대한 설명에 대응하는 텍스트 특징 벡터를 추출할 수 있다.

【0044】추가적으로, 다중모달 특징 추출부(120)는 탐지된 객체와 연관된 정보를 학습된 HRNet(high resolution network) 기반의 모델에 제공하여 객체에 대응하는 뼈대 정보를 추출할 수 있다. 여기서, 뼈대 정보는 객체의 주요 특징점(예: 인체의 관절 등)을 추출하고, 추출된 특징점을 연결하여 생성되는 골격 정보일 수 있다. 이 경우, 다중모달 특징 추출부(120)는 추출된 뼈대 정보를 학습된 PoseConv3D 기반의 모델에 제공하여 객체의 동작을 나타내는 모션 특징 벡터를 추출할 수 있다.

【0045】일 실시예에 따르면, 컴퓨팅 장치(100)는 확산 모델(diffusion model)의 입력으로 사용하기 위해 시각적 특징 벡터 상에 노이즈(noise)를 주입할 수 있다. 예를 들어, 노이즈 주입부(130)는 시간 단계(time step)의 범위에 따라 결정된 양의 가우시안 노이즈(Gaussian noise)를 시각적 특징 벡터에 주입하여 노이즈 벡터를 생성할 수 있다. 노이즈 주입부(130)는 시간 단계가 $t \in [1, T]$의 범위를 가질 때, 다음의 수학식 2를 기초로 시각적 특징 벡터에 노이즈를 주입할 수 있

다.

【0046】 【수학식 2】

$$f_{vis}^{t} = \sqrt{\alpha_t} f_{vis}^{0} + \sqrt{1-\alpha_t}\epsilon, \epsilon \sim N(0, I), \alpha_t = \prod_{0}^{t}(1-\beta_t)$$

【0047】 여기서, f_{vis}^{0}는 시각적 특징 벡터이고, f_{vis}^{t}는 t의 시간 단계 만큼의 노이즈가 주입된 시각적 특징 벡터, 즉 노이즈 벡터일 수 있다. 또한, β_t는 주입할 노이즈의 양을 결정하는데 사용되는 스케쥴(schedule)일 수 있다. 즉, β_t가 증가할수록, α_t가 더욱 감소하기 때문에 더 많은 노이즈가 주입될 수 있다.

【0048】 일 실시예에 따르면, 벡터 복원부(140)는 시각적 특징 벡터 상에 노이즈가 주입되어 생성된 노이즈 벡터를 확산 모델에 입력하여 원본 벡터를 복원할 수 있다. 예를 들어, 벡터 복원부(140)는 확산 모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터 및 모션 특징 벡터를 조건으로 활용하여 노이즈가 제거된 복원 벡터를 생성할 수 있다. 여기서, 조건은 확산 모델의 동작 시 참조되는 정보를 지칭할 수 있으며, 확산 모델은 조건으로 입력된 정보를 참조하여 데이터를 생성할 수 있다.

【0049】 일 실시예에 따르면, 벡터 복원부(140)는 제1 확산 모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터를 조건으로 활용하여 노이즈 벡터에 포함된 적어도 일부의 노이즈를 제거하는 제1 복원 단계 및 제2 확산 모델에 노이즈 벡터를

입력하고, 모션 특징 벡터를 조건으로 활용하여 노이즈 벡터에 포함된 적어도 일부의 노이즈를 제거하는 제2 복원 단계를 반복 수행하여 노이즈가 제거된 복원 벡터를 생성할 수 있다.

【0050】일 실시예에 따르면, 이상 탐지부(150)는 시각적 특징 벡터와 복원 벡터를 비교하여 비디오 데이터에 대한 이상 탐지를 수행할 수 있다. 예를 들어, 이상 탐지부(150)는 시각적 특징 벡터와 복원 벡터 사이의 거리에 따른 이상 점수를 산출하고, 산출된 이상 점수가 임계값 이상인지 여부를 기초로 비디오 데이터에 대한 이상 탐지를 수행할 수 있다. 여기서, 시각적 특징 벡터와 복원 벡터 사이의 거리에 따른 이상 점수는 다음의 수학식 3과 같이 평균 제곱 오차(mean squared error; MSE)를 이용하여 산출될 수 있다.

【0051】**【수학식 3】**

$$Loss = \left\| f_{vis}^{0} - \hat{f}_{vis}^{0} \right\|_{2}^{2}$$

【0052】여기서, Loss는 평균 제곱 오차 손실을 나타낼 수 있다. 또한, f_{vis}^{0}는 초기의 시각적 특징 벡터를 나타내고, $\hat{f}_{vis}^{0}$는 확산 모델에 의해 노이즈가 제거되어 복원된 복원 벡터를 나타낼 수 있다.

【0053】도 1에서는 컴퓨팅 장치(100)에 포함된 각각의 기능적인 구성이 구분되어 상술되었으나, 이는 발명의 이해를 돕기 위한 것일 뿐이며, 하나의 연산 장

치에서 둘 이상의 기능을 수행할 수도 있다. 이와 같은 구성에 의해, 컴퓨팅 장치(100)는 다중모달 특징 벡터를 상호 보완적으로 활용하여 비디오 이상 탐지의 성능을 향상시킬 수 있다.

【0055】 도 2는 본원 발명의 일 실시예에 따른 다중모달 특징 벡터가 추출되는 과정을 나타내는 블록도이다. 상술된 것과 같이, 컴퓨팅 장치(도 1의 100)는 비디오 데이터에서 객체를 탐지하고, 탐지된 객체에 대한 시각적 특징 벡터(f_{vis})(212), 모션 특징 벡터(f_{mot})(232) 및 텍스트 특징 벡터(f_{text})(252)를 포함하는 다중모달 특징 벡터를 추출할 수 있다.

【0056】 비디오 데이터에서 객체가 탐지되는 경우, 탐지된 객체와 연관된 정보(202)를 기초로 다중모달 특징 벡터가 추출될 수 있다. 여기서, 객체와 연관된 정보(202)는 객체 트랙렛을 나타낼 수 있다. 일 실시예에 따르면, 객체와 연관된 정보(202)를 시각적 추출기(210)에 입력하는 경우, 시각적 특징 벡터(f_{vis})(212)가 추출될 수 있다. 여기서, 시각적 추출기(210)는 객체의 색상, 형상 등의 외형 정보를 인식 및/또는 분류하는 모델로서, I3D 기반의 모델, C3D(convolutional 3D network) 기반의 모델 등을 포함할 수 있다. 예를 들어, 시각적 특징 벡터(f_{vis})(212)는 다음의 수학식 4와 같이 추출될 수 있다.

【0057】【수학식 4】

$$f_{vis} = \Phi_{vis}(O_n)$$

【0058】일 실시예에 따르면, 객체와 연관된 정보(202)를 뼈대 추출기(220)에 입력하는 경우, 객체에 대응하는 뼈대 정보(222)가 추출될 수 있다. 여기서, 뼈대 추출기(220)는 객체의 골격 정보를 추출하기 위한 HRNet 기반의 모델을 포함할 수 있다. 또한, 뼈대 정보(222)를 모션 추출기(230)에 입력하는 경우, 모션 특징 벡터(f_{mot})(232)가 추출될 수 있다. 여기서, 모션 추출기(230)는 뼈대 정보(222)를 기반으로 객체의 자세 및/또는 동작을 추정하기 위한 PoseConv3D 기반의 모델 등을 포함할 수 있다. 예를 들어, 모션 특징 벡터(f_{mot})(232)는 다음의 수학식 5와 같이 추출될 수 있다.

【0059】【수학식 5】

$$f_{mot} = \Phi_{mot}(\Phi_{skl}(O_n))$$

【0060】일 실시예에 따르면, 객체와 연관된 정보(202)를 캡션 추출기(240)에 입력하는 경우, 객체의 동작을 해설하는 캡션(242)이 추출될 수 있다. 여기서, 캡션 추출기(240)는 SwinBERT 모델 등과 같은 비디오 캡셔닝(video captioning) 모델을 포함할 수 있다. 또한, 캡션(242)을 텍스트 추출기(250)에 입력하는 경우, 텍

스트 특징 벡터(f_{text})(252)가 추출될 수 있다. 여기서, 텍스트 추출기(250)는 캡션(242)을 기반으로 문장 기반의 텍스트 특징을 추출하기 위한 SimCSE 등의 모델을 포함할 수 있다. 예를 들어, 텍스트 특징 벡터(f_{text})(252)는 다음의 수학식 6과 같이 추출될 수 있다.

【0061】 **【수학식 6】**

$$f_{text} = \Phi_{text}(\Phi_{cap}(O_n))$$

【0062】 도 3은 본원 발명의 일 실시예에 따른 확산 모델(300)의 구조를 나타내는 예시적인 도면이다. 일 실시예에 따르면, 확산 모델(300)은 복수의 디노이징 주의 블록(denoising attention block; DAB)을 포함하는 인코더(encoder)(310), 병목(bottleneck)(미도시) 및 디코더(decoder)(320)를 포함할 수 있다. 예를 들어, 확산 모델(300)은 도시된 구조를 통해 노이즈 벡터(312)에 주입된 노이즈를 제거하여 복원 벡터(322)를 생성할 수 있다.

【0063】 일 실시예에 따르면, 확산 모델(300)은 노이즈 벡터(312)를 입력으로 하고, 텍스트 특징 벡터 또는 모션 특징 벡터를 조건(314)으로 참조하여 시각적 특징 벡터를 복원한 복원 벡터(322)를 생성할 수 있다. 즉, 확산 모델(300)은 노이즈 벡터(312)를 복원할 때, 객체에 대응하는 텍스트 또는 모션을 참조하여 정답과 가깝도록 복원 벡터(322)를 생성할 수 있다.

【0064】일 실시예에 따르면, 노이즈 벡터(312)가 확산 모델(300)을 통과하는 경우, 시간 단계에 따라 정해진 양의 노이즈가 제거될 수 있다. 예를 들어, 노이즈 벡터(312)가 확산 모델(300)을 한번 통과하는 경우, 하나의 시간 단계에 대응하는 양의 노이즈가 제거될 수 있다. 다른 예에서, 노이즈 벡터(312)가 확산 모델(300)을 한번 통과하는 경우, 1/2의 시간 단계에 대응하는 양의 노이즈가 제거될 수도 있다. 즉, 노이즈 벡터(312)가 노이즈 주입 과정에서 부여된 시간 단계의 구간만큼 반복하여 확산 모델(300)을 통과하는 경우, 노이즈가 모두 제거된 복원 벡터(322)가 생성될 수 있다.

【0066】도 4는 본원 발명의 일 실시예에 따른 디노이징 주의 블록의 구조를 나타내는 예시적인 도면이다. 일 실시예에 따르면, 디노이징 주의 블록은 트랜스포머 블록(transformer block)(410)과 잔차 블록(residual block)(420)의 스택(stack)으로 구성될 수 있다. 또한, 트랜스포머 블록(410)은 자기 주의 레이어(self-attention layer)(416), 교차 주의 레이어(cross-attention layer)(414) 및 피드 포워드 네트워크(feed forward network)(412)를 포함할 수 있으며, 잔차 블록(420)은 스킵 연결(skip connection)로 연결된 복수의 선형 레이어(linear layer)(422_1, 422_2)를 포함할 수 있다.

【0067】일 실시예에 따르면, 확산 모델은 시간 단계에 따라 동작하기 때문에, 시간 임베딩 벡터(404)가 텍스트 특징 벡터 또는 모션 특징 벡터인 조건(402)에 결합될 수 있다. 예를 들어, 벡터 결합은 하기의 수학식 7과 같이 수행될 수 있

다.

【0068】 **【수학식 7】**

$$f'_{cond} = W_1 f_{cond} + W_2 f_{ftime}$$

【0069】 여기서, f'_{cond}는 결합 벡터를 나타내고, f_{cond}는 텍스트 특징 벡터 또는 모션 특징 벡터인 조건(402)을 나타내며, f_{ftime}는 시간 임베딩 벡터(404)를 나타낼 수 있다. 또한, $W_1 \in \mathbb{R}^{D_{cond} \times D_{vis}}$ 및 $W_2 \in \mathbb{R}^{D_{time} \times D_{vis}}$는 각각 조건(402)의 차원 D_{cond}과 시간 임베딩 벡터(404)의 차원 D_{time}을 입력된 시각적 특징 벡터(406)의 차원 D_{vis}과 일치시키기 위한 투영 행렬을 나타낼 수 있다. 즉, 이와 같이 생성된 결합 벡터가 트랜스포머 블록(410) 및 잔차 블록(420)에 제공되어 확산 모델이 조건(402)을 더욱 효과적으로 참조하도록 도울 수 있다.

【0070】 일 실시예에 따르면, 트랜스포머 블록(410)은 시각적 특징과 조건 특징들 간의 상관관계를 인식하기 위한 블록일 수 있다. 상관관계를 인식하기 위해, 세그먼트 레벨의 다중모달 특징 벡터는 클립 레벨(clip-level)의 다중모달 특징 벡터로 변환될 수 있다. 여기서, 클립은 8개의 연속하는 세그먼트의 집합으로 구성될 수 있으나, 이에 한정되지 않는다. 이 경우, 클립 레벨의 다중모달 특징 벡터는 $\mathbb{R}^{C \times 8 \times d}$ ($C = S/16$)의 형상을 가질 수 있다. 예를 들어, 트랜스포머 블록

(410)은 다음의 수학식 8과 같이 연산을 수행할 수 있다.

【0071】 **【수학식 8】**

$$TA(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d}})V$$

【0072】 여기서, $Q = W_q f_{vis} \in \mathbb{R}^{C\times 8\times d}$, $K = W_k f'_{cond} \in \mathbb{R}^{C\times 8\times d}$, $V = W_v f'_{cond} \in \mathbb{R}^{C\times 8\times d}$일 수 있다. 또한, $W_q, W_k, W_v \mathbb{R}^{D_{vis}\times d}$는 각각 투영 행렬을 나타내고, d는 쿼리(query), 키(key) 및 밸류(value)의 차원을 나타낼 수 있다. 이와 같이 트랜스포머 블록(410) 및 잔차 블록(420)의 연산 시 텍스트 특징 벡터 및/또는 모션 특징 벡터가 조건(402)으로 참조되는 것에 의해, 컴퓨팅 장치(도 1의 100)는 객체의 시각적 특징과 함께 객체를 설명하는 텍스트 및/또는 객체의 동작을 모두 참조하여 효과적으로 노이즈 제거 및 벡터 복원을 수행할 수 있다.

【0074】 도 5는 본원 발명의 일 실시예에 따른 제1 확산 모델(520) 및 제2 확산 모델(530)에 의해 복원 과정이 수행되는 예시를 나타내는 도면이다. 상술된 것과 같이, 컴퓨팅 장치(도 1의 100)는 확산 모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터(512) 및 모션 특징 벡터(514)를 조건으로 활용하여 노이즈가 제거된 복원 벡터를 생성할 수 있다. 여기서, 확산 모델은 제1 확산 모델(520) 및 제2 확산 모델(530)을 포함할 수 있다.

【0075】일 실시예에 따르면, 시각적 특징 벡터(f_{vis}^{0})(502)에 노이즈를 주입하는 확산 프로세스(510)에 의해 제1 노이즈 벡터(f_{vis}^{τ})(504)가 생성될 수 있다. 예를 들어, 시간 단계가 τ로 정해진 경우, 시간 단계 τ에 대응하는 양의 가우시안 노이즈가 시각적 특징 벡터(f_{vis}^{0})(502)에 주입되어 제1 노이즈 벡터(f_{vis}^{τ})(504)가 생성될 수 있다.

【0076】일 실시예에 따르면, 이와 같이 생성된 제1 노이즈 벡터(f_{vis}^{τ})(504)는 텍스트 특징 벡터(512)를 조건으로 학습된 제1 확산 모델(520)에 입력될 수 있다. 이 경우, 제1 확산 모델(520)은 텍스트 특징 벡터(512)를 참고로 노이즈를 1회 제거하여 제2 노이즈 벡터($\hat{f}_{vis}^{\tau-1}$)(506)를 생성할 수 있다. 그리고 나서, 생성된 제2 노이즈 벡터($\hat{f}_{vis}^{\tau-1}$)(506)는 모션 특징 벡터(514)를 조건으로 학습된 제2 확산 모델(530)에 입력될 수 있다. 이 경우, 제2 확산 모델(530)은 모션 특징 벡터(514)를 참고로 노이즈를 2회 제거하여 제3 노이즈 벡터($\hat{f}_{vis}^{\tau-2}$)(508)를 생성할 수 있다.

【0077】이 때, 생성된 제3 노이즈 벡터($\hat{f}_{vis}^{\tau-2}$)(508)는 다시 제1 확산 모델(520)에 제공되어 제1 확산 모델(520) 및 제2 확산 모델(530) 사이에 순환이 형성될 수 있다. 상술된 과정에 의해, 노이즈가 시간 단계 τ만큼 반복적으로 제거되는 결과 복원 벡터가 생성될 수 있다. 이와 같은 구성에 의해, 하나의 확산 모델을 이

용하는 것이 아닌 조건이 상이한 제1 확산 모델(520) 및 제2 확산 모델(530)을 모두 이용함으로써 복원 성능이 향상될 수 있으며, 이에 따라 더 높은 정확도로 비디오 이상 탐지가 수행될 수 있다.

【0079】도 6은 본원 발명의 제2 실시예에 따른 제1 확산 모델(520) 및 제2 확산 모델(530)에 의해 복원 과정이 수행되는 예시를 나타내는 도면이다. 상술된 복원 과정과 달리 제1 확산 모델(520) 및 제2 확산 모델(530)을 사용하는 순서 및/또는 방식은 상이하게 결정될 수 있다. 도 6의 예에서, 시간 단계 τ에 대응하는 양의 가우시안 노이즈가 시각적 특징 벡터에 주입되어 제1 노이즈 벡터(602)가 생성될 수 있다. 이 경우, 제1 노이즈 벡터(602)는 시간 단계 τ만큼 반복적으로 제1 확산 모델(520)에 입력될 수 있으며, 이에 따라 제1 확산 모델(520)은 노이즈가 모두 제거된 제1 복원 벡터(604)를 생성할 수 있다.

【0080】그리고 나서, 제1 복원 벡터(604)에 대한 확산 프로세스(510)가 다시 수행될 수 있다. 예를 들어, 시간 단계 τ에 대응하는 양의 가우시안 노이즈가 제1 복원 벡터(604)에 주입되어 제2 노이즈 벡터(606)가 생성될 수 있다. 이 경우, 제2 노이즈 벡터(606)는 시간 단계 τ만큼 반복적으로 제2 확산 모델(530)에 입력될 수 있으며, 이에 따라 제2 확산 모델(530)은 노이즈가 모두 제거된 제2 복원 벡터(608)를 생성할 수 있다.

【0082】도 7은 본원 발명의 제3 실시예에 따른 제1 확산 모델(520) 및 제2 확산 모델(530)에 의해 복원 과정이 수행되는 예시를 나타내는 도면이다. 상술된 복원 과정과 달리 제1 확산 모델(520) 및 제2 확산 모델(530)을 사용하는 순서 및/또는 방식은 상이하게 결정될 수 있다. 도 7의 예에서, 시간 단계 T에 대응하는 양의 가우시안 노이즈가 시각적 특징 벡터에 주입되어 제1 노이즈 벡터(702)가 생성될 수 있다. 이 경우, 제1 노이즈 벡터(602)는 제1 확산 모델(520)에 제공될 수 있으며, 이에 따라, 제1 확산 모델(520)은 노이즈가 1회 제거된 제2 노이즈 벡터(704)를 생성할 수 있다.

【0083】일 실시예에 따르면, 제2 노이즈 벡터(704)에 대해 다시 하나의 시간 단계 만큼의 노이즈를 주입하는 확산 프로세스(510)가 수행되어 제3 노이즈 벡터(706)가 생성될 수 있다. 이 경우, 제3 노이즈 벡터(706)는 제2 확산 모델(530)에 제공될 수 있으며, 이에 따라, 제2 확산 모델(530)은 다시 노이즈가 1회 제거된 제4 노이즈 벡터(708)를 생성할 수 있다. 즉, 제1 확산 모델(520) 및 제2 확산 모델(530)을 한번 순환할 때마다 노이즈가 1회 제거될 수 있으며, 시간 단계 T만큼 해당 순환이 반복되는 경우 복원 벡터가 생성될 수 있다.

【0085】도 8은 본원 발명의 제4 실시예에 따른 제1 확산 모델(520) 및 제2 확산 모델(530)에 의해 복원 과정이 수행되는 예시를 나타내는 도면이다. 상술된 복원 과정과 달리 제1 확산 모델(520) 및 제2 확산 모델(530)을 사용하는 순서 및/

또는 방식을 상이하게 결정될 수 있다. 도 8의 예에서, 시간 단계 τ에 대응하는 양의 가우시안 노이즈가 시각적 특징 벡터에 주입되어 제1 노이즈 벡터(802)가 생성될 수 있다. 이 경우, 제1 노이즈 벡터(802)는 시간 단계 τ의 절반만큼 반복적으로 제1 확산 모델(520)에 입력될 수 있으며, 이에 따라 제1 확산 모델(520)은 노이즈가 절반만큼 제거된 제2 노이즈 벡터(804)를 생성할 수 있다.

【0086】 그리고 나서, 제2 노이즈 벡터(804)는 시간 단계 τ의 절반만큼 반복적으로 제2 확산 모델(530)에 입력될 수 있으며, 이에 따라 제2 확산 모델(530)은 노이즈가 모두 제거된 복원 벡터(806)를 생성할 수 있다. 도 5 내지 도 8에서 상술된 것과 같이, 제1 확산 모델(520) 및 제2 확산 모델(530)을 이용하는 방식은 다양하게 결정될 수 있으며, 이상 및/또는 이상 행위의 탐지 조건에 따라 최적의 성능을 갖도록 복원 과정이 수행될 수 있다.

【0088】 도 9은 본원 발명의 일 실시예에 따른 다중모달 확산 기반의 비디오 이상 탐지 방법(900)의 예시를 나타내는 흐름도이다. 다중모달 확산 기반의 비디오 이상 탐지 방법(900)은 프로세서(예를 들어, 컴퓨팅 장치의 적어도 하나의 프로세서)에 의해 수행될 수 있다. 다중모달 확산 기반의 비디오 이상 탐지 방법(900)은 프로세서가 복수의 프레임을 포함하는 비디오 데이터를 획득함으로써 개시될 수 있다(S910). 예를 들어, 비디오 데이터는 CCTV 등의 감시 카메라로부터 획득될 수 있으나, 이에 한정되지 않는다.

【0089】일 실시예에 따르면, 프로세서는 복수의 프레임에 포함된 객체를 탐지할 수 있다(S920). 예를 들어, 프로세서는 임의의 객체 탐지 알고리즘 및/또는 기계학습 모델을 이용하여 복수의 프레임에 포함된 객체를 탐지할 수 있다. 그리고 나서, 프로세서는 탐지된 객체에 대한 시각적 특징 벡터, 텍스트 특징 벡터 및 모션 특징 벡터를 포함하는 다중모달 특징 벡터를 추출할 수 있다(S930).

【0090】일 실시예에 따르면, 프로세서는 탐지된 객체와 연관된 정보를 학습된 I3D 기반의 모델에 제공하여, 객체에 대한 시각적 특징 벡터를 추출할 수 있다. 추가적으로 또는 대안적으로, 프로세서는 탐지된 객체와 연관된 정보를 BERT 기반의 모델에 제공하여 객체에 대한 설명을 나타내는 캡션을 생성하고, 생성된 캡션을 학습된 SimCSE 기반의 모델에 제공하여 객체에 대한 설명에 대응하는 텍스트 특징 벡터를 추출할 수 있다. 추가적으로 또는 대안적으로, 프로세서는 탐지된 객체와 연관된 정보를 학습된 HRNet 기반의 모델에 제공하여 객체에 대응하는 뼈대 정보를 추출하고, 추출된 뼈대 정보를 이용하여 객체의 동작을 나타내는 모션 특징 벡터를 추출할 수 있다.

【0091】일 실시예에 따르면, 프로세서는 시각적 특징 벡터에 노이즈를 주입하여 노이즈 벡터를 생성할 수 있다(S940). 예를 들어, 프로세서는 시간 단계의 범위에 따라 결정된 양의 가우시안 노이즈를 시각적 특징 벡터에 주입하여 노이즈 벡터를 생성할 수 있다. 여기서, 노이즈 벡터는 시각적 특징 벡터에 적어도 일부의 노이즈가 주입된 상태의 벡터를 지칭하는 것으로, 확산 프로세스에 의해 초기에 생성되는 벡터와 확산 모델을 충분히 거치지 않아 아직 노이즈가 남아있는 벡터를 모

두 포함할 수 있다.

【0092】일 실시예에 따르면, 프로세서는 확산 모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터 및 모션 특징 벡터를 조건으로 활용하여 노이즈가 제거된 복원 벡터를 생성할 수 있다(S950). 여기서, 확산 모델은 제1 확산 모델 및 제2 확산 모델을 포함할 수 있다. 프로세서는 제1 확산 모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터를 조건으로 활용하여 노이즈 벡터에 포함된 적어도 일부의 노이즈를 제거하는 제1 복원 단계 및 제2 확산 모델에 노이즈 벡터를 입력하고, 모션 특징 벡터를 조건으로 활용하여 노이즈 벡터에 포함된 적어도 일부의 노이즈를 제거하는 제2 복원 단계를 반복 수행하여 노이즈가 제거된 복원 벡터를 생성할 수 있다.

【0093】일 실시예에 따르면, 프로세서는 시각적 특징 벡터와 복원 벡터를 비교하여 비디오 데이터에 대한 이상 탐지를 수행할 수 있다(S960). 이 경우, 프로세서는 시각적 특징 벡터와 복원 벡터 사이의 거리에 따른 이상 점수를 산출하고, 산출된 이상 점수가 임계값 이상인지 여부를 기초로 비디오 데이터에 대한 이상 탐지를 수행할 수 있다. 예를 들어, 이상 점수는 시각적 특징 벡터와 복원 벡터 사이의 평균 제곱 오차를 이용하여 산출될 수 있다.

【0095】도 10은 본원 발명의 일 실시예에 따른 컴퓨팅 장치(100)의 하드웨어 구성을 나타내는 블록도이다. 컴퓨팅 장치(100)는 메모리(1010), 프로세서(1020), 통신 모듈(1030) 및 입출력 인터페이스(1040)를 포함할 수 있으며, 도 10

에 도시된 바와 같이, 컴퓨팅 장치(100)는 통신 모듈(1030)을 이용하여 네트워크를 통해 정보 및/또는 데이터를 통신할 수 있도록 구성될 수 있다.

【0096】 메모리(1010)는 비-일시적인 임의의 컴퓨터 판독 가능한 기록매체를 포함할 수 있다. 일 실시예에 따르면, 메모리(1010)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(100)에 포함될 수 있다. 또한, 메모리(1010)에는 운영체제와 적어도 하나의 프로그램 코드가 저장될 수 있다.

【0097】 이러한 소프트웨어 구성요소들은 메모리(1010)와는 별도의 컴퓨터에서 판독 가능한 기록매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록매체는 이러한 컴퓨팅 장치(100)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(1030)을 통해 메모리(1010)에 로딩될 수도 있다. 예를 들어, 적어도 하나의 프로그램은 개발자들 또는 애플리케이션의 설치 파일을 배포하는 파일 배포 시스템이 통신 모듈(1030)을 통해 제공하는 파일들에 의해 설치되는 컴퓨터 프로그램에 기반하여 메모리(1010)에 로딩될 수 있다.

【0098】프로세서(1020)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(1010) 또는 통신 모듈(1030)에 의해 다른 사용자 단말(미도시) 또는 다른 외부 시스템으로 제공될 수 있다.

【0099】통신 모듈(1030)은 네트워크를 통해 사용자 단말(미도시)과 컴퓨팅 장치(100)가 서로 통신하기 위한 구성 또는 기능을 제공할 수 있으며, 컴퓨팅 장치(100)가 외부 시스템(일례로 별도의 클라우드 시스템 등)과 통신하기 위한 구성 또는 기능을 제공할 수 있다. 일례로, 컴퓨팅 장치(100)의 프로세서(1020)의 제어에 따라 제공되는 제어 신호, 명령, 데이터 등이 통신 모듈(1030)과 네트워크를 거쳐 사용자 단말 및/또는 외부 시스템의 통신 모듈을 통해 사용자 단말 및/또는 외부 시스템으로 전송될 수 있다.

【0100】또한, 컴퓨팅 장치(100)의 입출력 인터페이스(1040)는 컴퓨팅 장치(100)와 연결되거나 컴퓨팅 장치(100)가 포함할 수 있는 입력 또는 출력을 위한 장치(미도시)와의 인터페이스를 위한 수단일 수 있다. 도 10에서는 입출력 인터페이스(1040)가 프로세서(1020)와 별도로 구성된 요소로서 도시되었으나, 이에 한정되지 않으며, 입출력 인터페이스(1040)가 프로세서(1020)에 포함되도록 구성될 수 있다. 컴퓨팅 장치(100)는 도 10의 구성요소들보다 더 많은 구성요소들을 포함할 수 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다.

【0101】 컴퓨팅 장치(100)의 프로세서(1020)는 복수의 사용자 단말 및/또는 복수의 외부 시스템으로부터 수신된 정보 및/또는 데이터를 관리, 처리 및/또는 저

장하도록 구성될 수 있다.

【0103】 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.

【0104】 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.

【0105】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.

【0106】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.

【0107】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 컴퓨팅 장치로 입력을 제공할 수 있다.

【0108】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.

【0109】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.

【0110】 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.

【0111】 본 발명에서, 달리 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함하여 본 명세서에서 사용되는 모든 용어는 이러한 개념이 속하는 기술 분야

에서 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.

【0112】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】

【0113】100: 컴퓨팅 장치

110: 객체 탐지부

120: 다중모달 특징 추출부

130: 노이즈 주입부

140: 벡터 복원부

150: 이상 탐지부

【청구범위】

【청구항 1】

적어도 하나의 프로세서에 의해 수행되는 비디오 이상 탐지 방법으로서,

복수의 프레임(frame)을 포함하는 비디오 데이터를 획득하는 단계;

상기 복수의 프레임에 포함된 객체를 탐지하는 단계;

상기 탐지된 객체에 대한 시각적 특징 벡터(visual feature vector), 텍스트 특징 벡터(text feature vector) 및 모션 특징 벡터(motion feature vector)를 포함하는 다중모달 특징 벡터를 추출하는 단계;

상기 시각적 특징 벡터에 노이즈(noise)를 주입하여 노이즈 벡터를 생성하는 단계;

확산 모델(diffusion model)에 상기 노이즈 벡터를 입력하고, 상기 텍스트 특징 벡터 및 상기 모션 특징 벡터를 조건으로 활용하여 노이즈가 제거된 복원 벡터를 생성하는 단계; 및

상기 시각적 특징 벡터와 상기 복원 벡터를 비교하여 상기 비디오 데이터에 대한 이상 탐지를 수행하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 2】

제1항에 있어서,

상기 다중모달 특징 벡터를 추출하는 단계는,

상기 탐지된 객체와 연관된 정보를 학습된 I3D(inflated 3D ConvNet) 기반의 모델에 제공하여, 상기 객체에 대한 시각적 특징 벡터를 추출하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 3】

제1항에 있어서,

상기 다중모달 특징 벡터를 추출하는 단계는,

상기 탐지된 객체와 연관된 정보를 BERT(bidirectional encoder representations from transformers) 기반의 모델에 제공하여 상기 객체에 대한 설명을 나타내는 캡션(caption)을 생성하는 단계; 및

상기 생성된 캡션을 학습된 SimCSE(simple contrastive learning of sentence embeddings) 기반의 모델에 제공하여 상기 객체에 대한 설명에 대응하는 상기 텍스트 특징 벡터를 추출하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 4】

제1항에 있어서,

상기 다중모달 특징 벡터를 추출하는 단계는,

상기 탐지된 객체와 연관된 정보를 학습된 HRNet(high resolution network) 기반의 모델에 제공하여 상기 객체에 대응하는 뼈대 정보를 추출하는 단계; 및

상기 추출된 뼈대 정보를 이용하여 상기 객체의 동작을 나타내는 모션 특징 벡터를 추출하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 5】

제4항에 있어서,

상기 추출된 뼈대 정보를 이용하여 상기 객체의 동작을 나타내는 모션 특징 벡터를 추출하는 단계는,

상기 추출된 뼈대 정보를 학습된 PoseConv3D 기반의 모델에 제공하여 상기 모션 특징 벡터를 추출하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 6】

제1항에 있어서,

상기 시각적 특징 벡터에 노이즈를 주입하여 노이즈 벡터를 생성하는 단계는,

시간 단계(time step)의 범위에 따라 결정된 양의 가우시안 노이즈(Gaussian

noise)를 상기 시각적 특징 벡터에 주입하여 상기 노이즈 벡터를 생성하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 7】

제1항에 있어서,

상기 확산 모델은 제1 확산 모델 및 제2 확산 모델을 포함하고,

상기 노이즈가 제거된 복원 벡터를 생성하는 단계는,

상기 제1 확산 모델에 노이즈 벡터를 입력하고, 상기 텍스트 특징 벡터를 조건으로 활용하여 상기 노이즈 벡터에 포함된 적어도 일부의 노이즈를 제거하는 제1 복원 단계; 및

상기 제2 확산 모델에 노이즈 벡터를 입력하고, 상기 모션 특징 벡터를 조건으로 활용하여 상기 노이즈 벡터에 포함된 적어도 일부의 노이즈를 제거하는 제2 복원 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 8】

제7항에 있어서,

상기 노이즈가 제거된 복원 벡터를 생성하는 단계는,

상기 제1 복원 단계 및 상기 제2 복원 단계를 반복 수행하여 상기 노이즈가 제거된 복원 벡터를 생성하는 단계를 더 포함하는

비디오 이상 탐지 방법.

【청구항 9】

제1항에 있어서,

상기 비디오 데이터에 대한 이상 탐지를 수행하는 단계는,

상기 시각적 특징 벡터와 상기 복원 벡터 사이의 거리에 따른 이상 점수(anomaly score)를 산출하는 단계; 및

상기 산출된 이상 점수가 임계값 이상인지 여부를 기초로 상기 비디오 데이터에 대한 이상 탐지를 수행하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 10】

제9항에 있어서,

상기 이상 점수를 산출하는 단계는,

상기 시각적 특징 벡터와 상기 복원 벡터 사이의 평균 제곱 오차(mean squared error; MSE)를 이용하여 상기 거리에 따른 이상 점수를 산출하는 단계를 포함하는

비디오 이상 탐지 방법.

【청구항 11】

제1항에 있어서,

상기 확산 모델은,

복수의 디노이징 주의 블록(denoising attention block; DAB)을 포함하는 인코더(encoder), 병목(bottleneck) 및 디코더(decoder)를 포함하는

비디오 이상 탐지 방법.

【청구항 12】

제11항에 있어서,

상기 디노이징 주의 블록은,

스킵 연결(skip connection)로 연결된 복수의 선형 레이어(linear layer)를 포함하는 잔차 블록(residual block); 및

자기 주의 레이어(self-attention layer), 교차 주의 레이어(cross-attention layer) 및 피드 포워드 네트워크(feed forward network; FFN)를 포함하는 트랜스포머 블록(transformer block)을 포함하는

비디오 이상 탐지 방법.

【청구항 13】

제1항 내지 제12항 중 어느 한 항에 따른 비디오 이상 탐지 방법을 컴퓨터에서 실행하기 위해 컴퓨터 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램.

【청구항 14】

컴퓨팅 장치로서,

통신 모듈;

메모리; 및

상기 메모리와 연결되고, 상기 메모리에 포함된 컴퓨터 판독 가능한 적어도 하나의 프로그램을 실행하도록 구성된 적어도 하나의 프로세서를 포함하고,

상기 적어도 하나의 프로그램은,

복수의 프레임을 포함하는 비디오 데이터를 획득하고,

상기 복수의 프레임에 포함된 객체를 탐지하고,

상기 탐지된 객체에 대한 시각적 특징 벡터, 텍스트 특징 벡터 및 모션 특징 벡터를 포함하는 다중모달 특징 벡터를 추출하고,

상기 시각적 특징 벡터에 노이즈를 주입하여 노이즈 벡터를 생성하고,

확산 모델에 상기 노이즈 벡터를 입력하고, 상기 텍스트 특징 벡터 및 상기 모션 특징 벡터를 조건 벡터로 활용하여 노이즈가 제거된 복원 벡터를 생성하고,

상기 시각적 특징 벡터와 상기 복원 벡터를 비교하여 상기 비디오 데이터에

대한 이상 탐지를 수행하기 위한 명령어들을 포함하는

컴퓨팅 장치.

【요약서】

【요약】

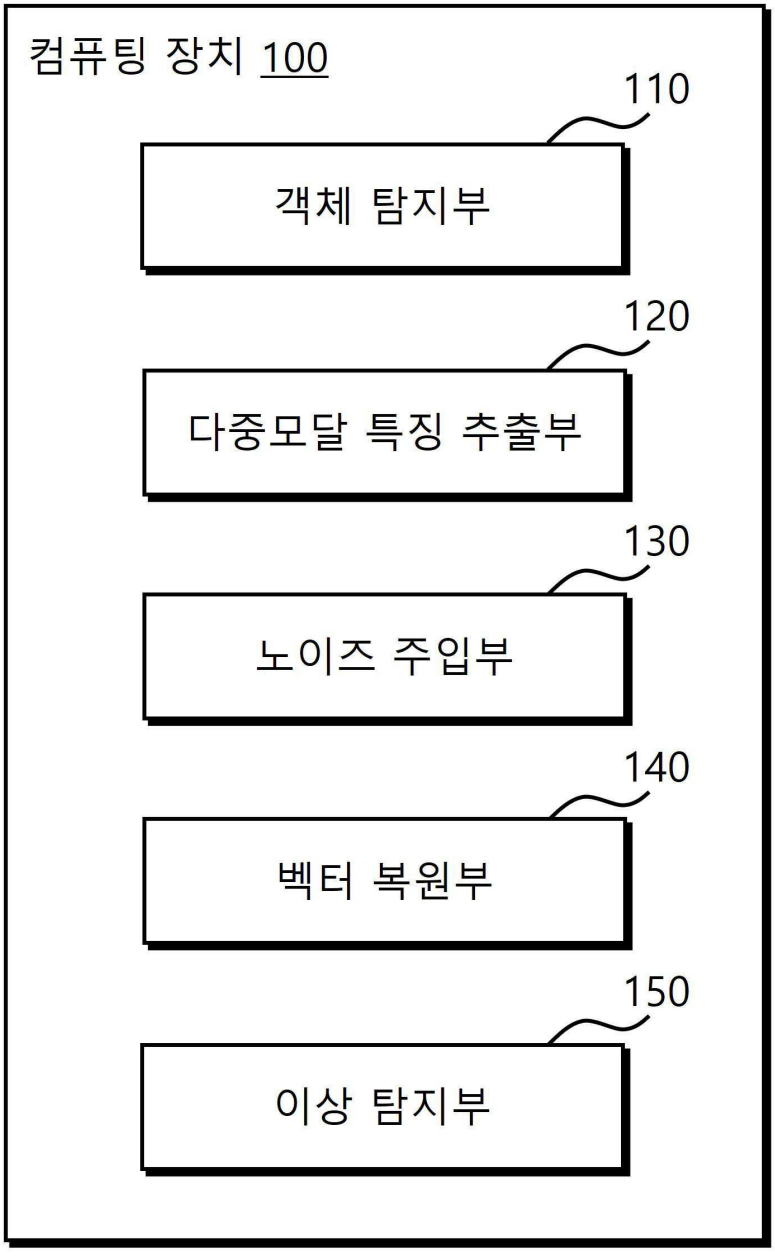
본원 발명의 다중모달 확산 기반의 비디오 이상 탐지 방법은 복수의 프레임을 포함하는 비디오 데이터를 획득하는 단계, 복수의 프레임에 포함된 객체를 탐지하는 단계, 탐지된 객체에 대한 시각적 특징 벡터, 텍스트 특징 벡터 및 모션 특징 벡터를 포함하는 다중모달 특징 벡터를 추출하는 단계, 시각적 특징 벡터에 노이즈를 주입하여 노이즈 벡터를 생성하는 단계, 확산 모델에 노이즈 벡터를 입력하고, 텍스트 특징 벡터 및 모션 특징 벡터를 조건 벡터로 활용하여 노이즈가 제거된 복원 벡터를 생성하는 단계 및 시각적 특징 벡터와 복원 벡터를 비교하여 비디오 데이터에 대한 이상 탐지를 수행하는 단계를 포함한다.

【대표도】

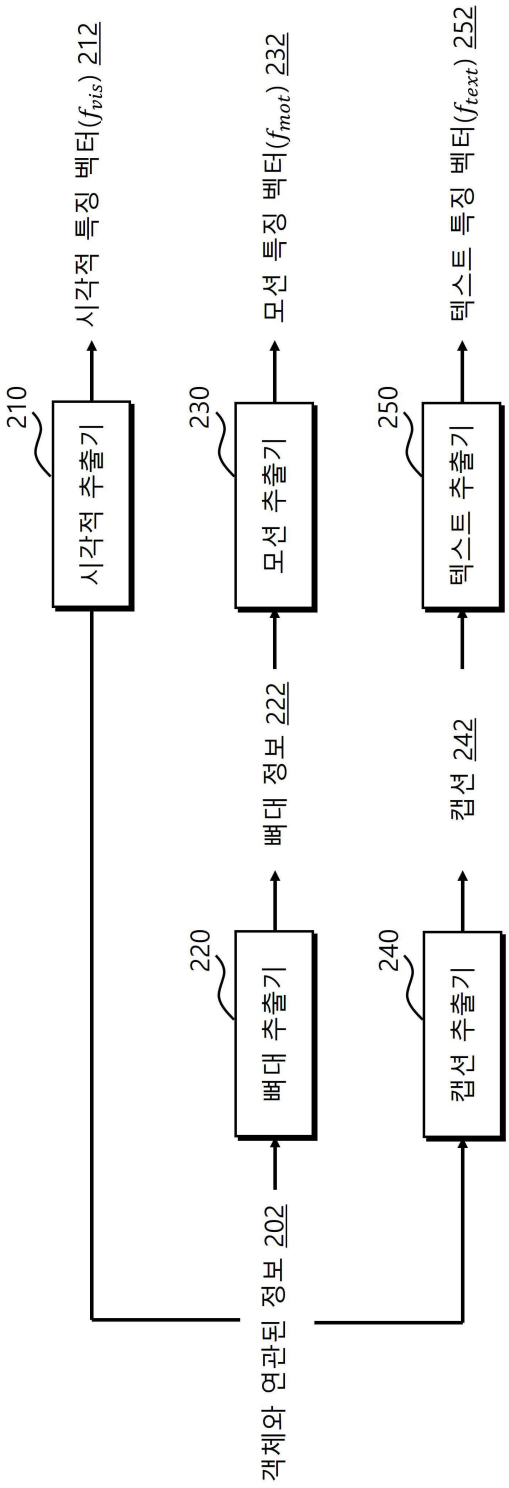
도 1

【도면】

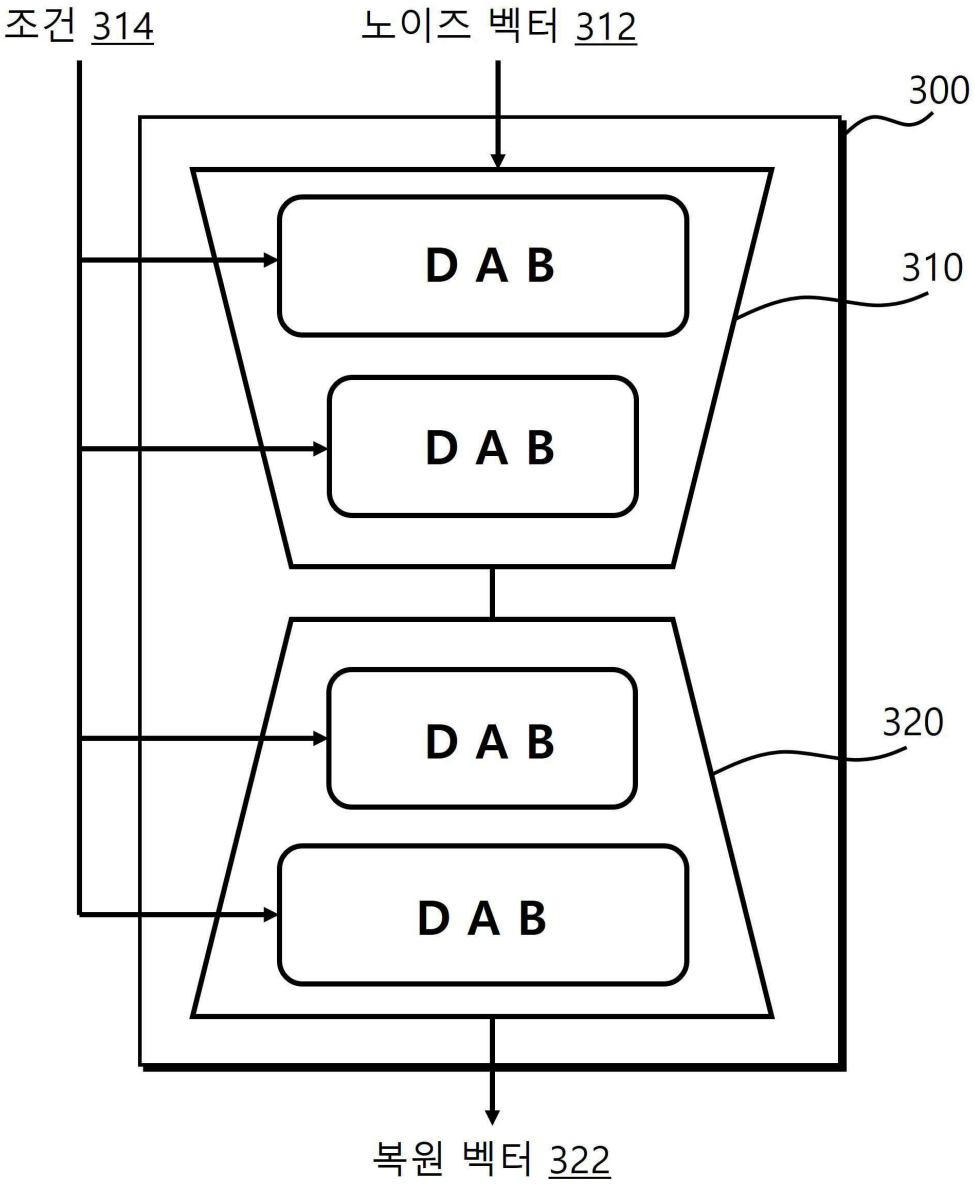
【도 1】


컴퓨팅 장치 100
110
객체 탐지부
120
다중모달 특징 추출부
130
노이즈 주입부
140
벡터 복원부
150
이상 탐지부

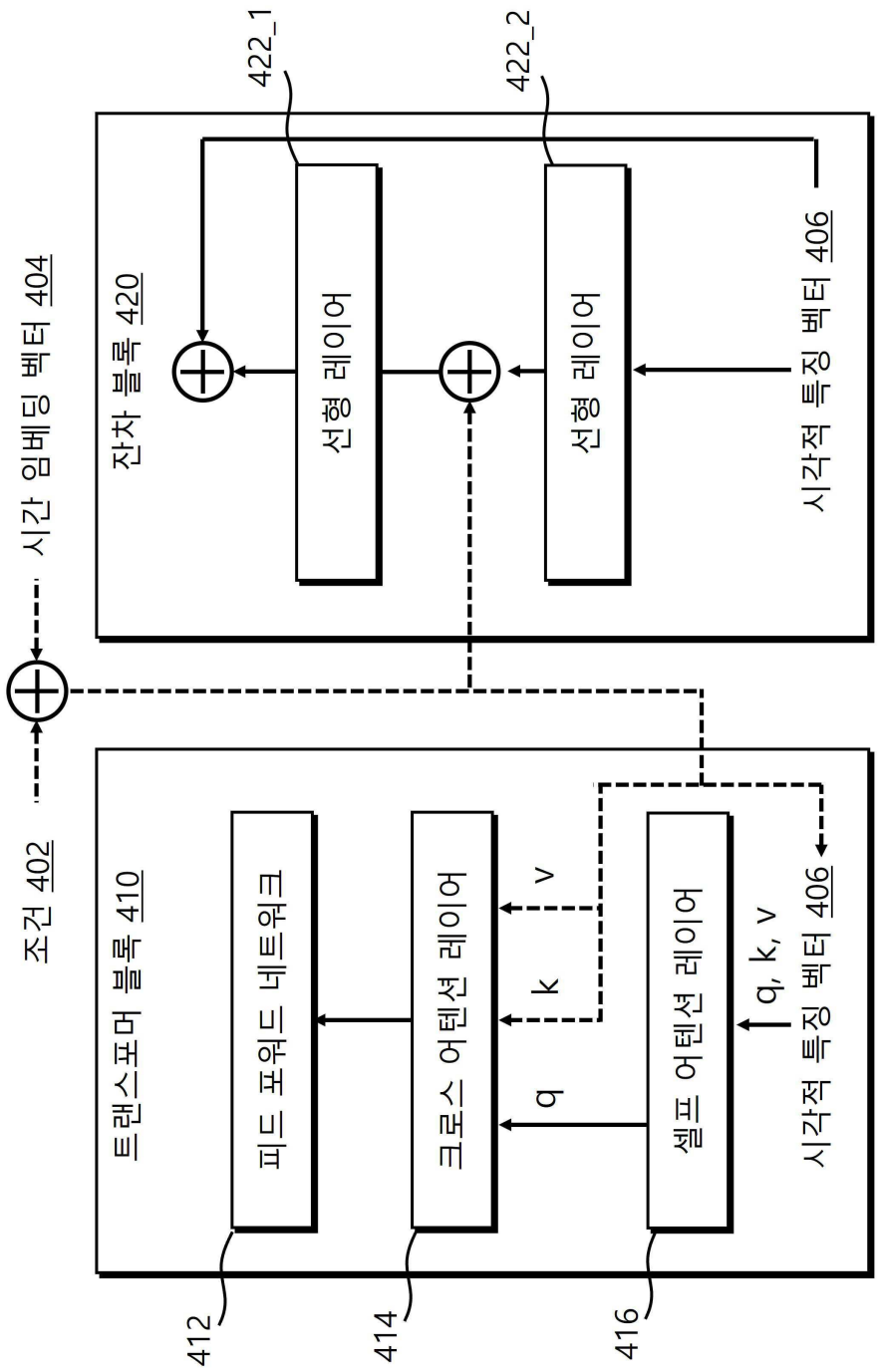
【도 2】


객체와 연관된 정보 202
시각적 추출기 210
시각적 특징 벡터(f_{vis}) 212
뼈대 추출기 220
뼈대 정보 222
모션 추출기 230
모션 특징 벡터(f_{mot}) 232
캡션 추출기 240
캡션 242
텍스트 추출기 250
텍스트 특징 벡터(f_{text}) 252

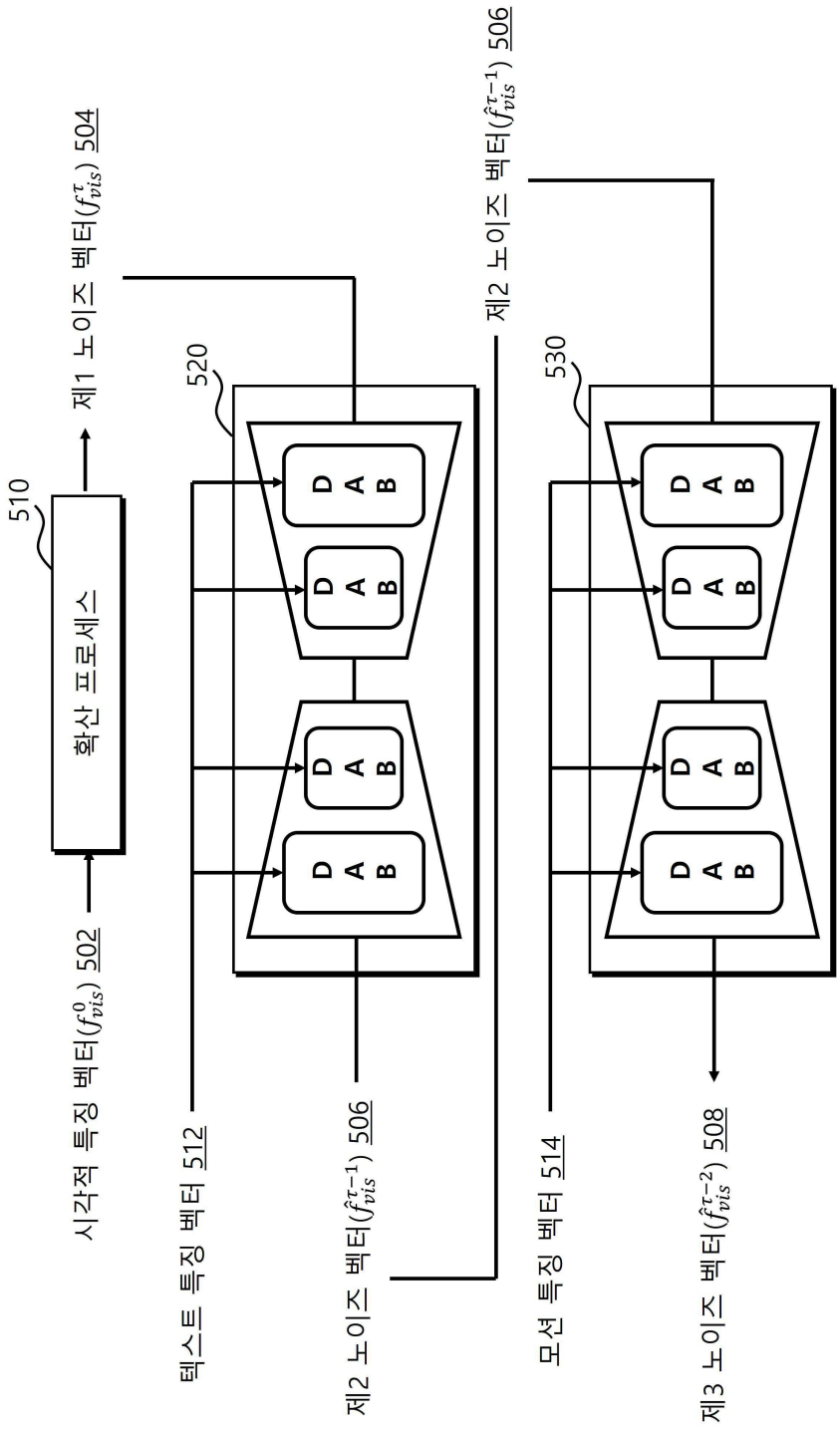
【도 3】



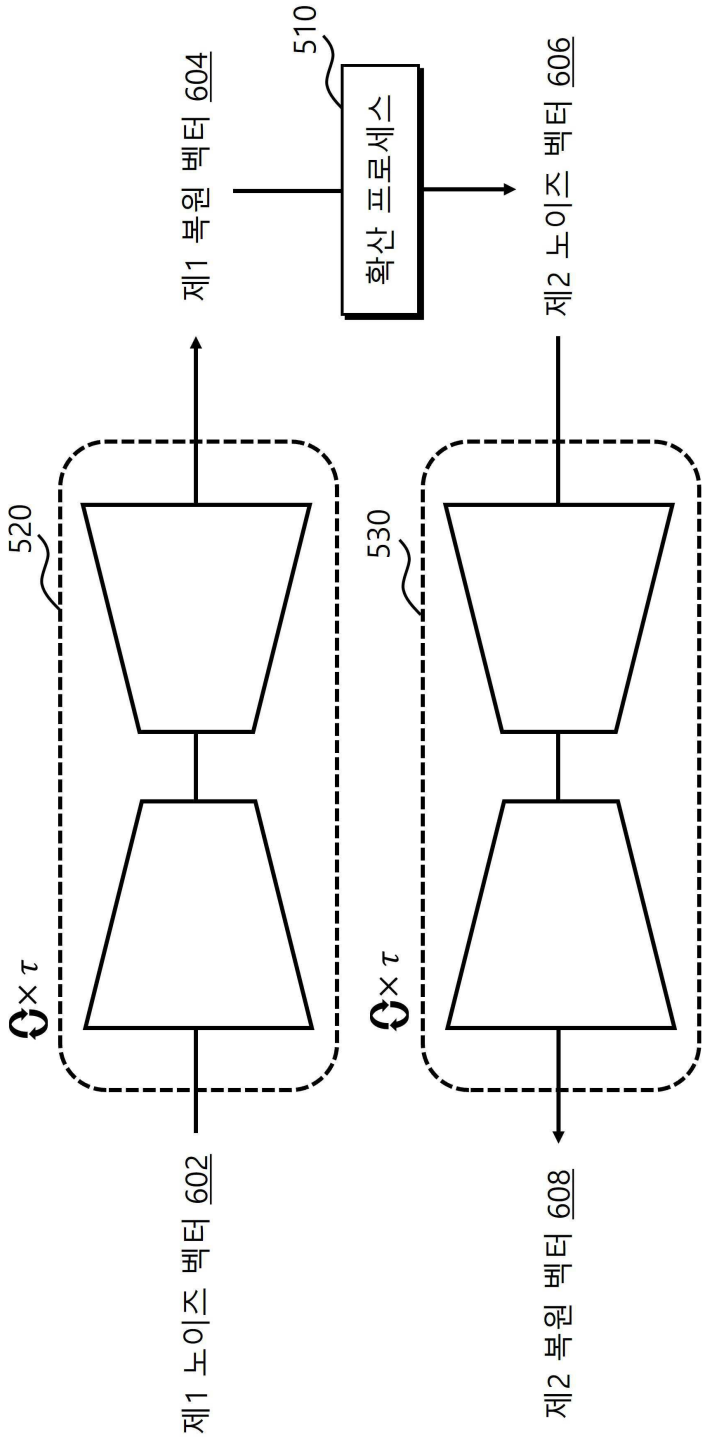
【도 4】


조건 402
시간 임베딩 벡터 404
트랜스포머 블록 410
피드 포워드 네트워크
412
크로스 어텐션 레이어
414
셀프 어텐션 레이어
416
q
k
v
q, k, v
시각적 특징 벡터 406
잔차 블록 420
선형 레이어
422_1
선형 레이어
422_2
시각적 특징 벡터 406

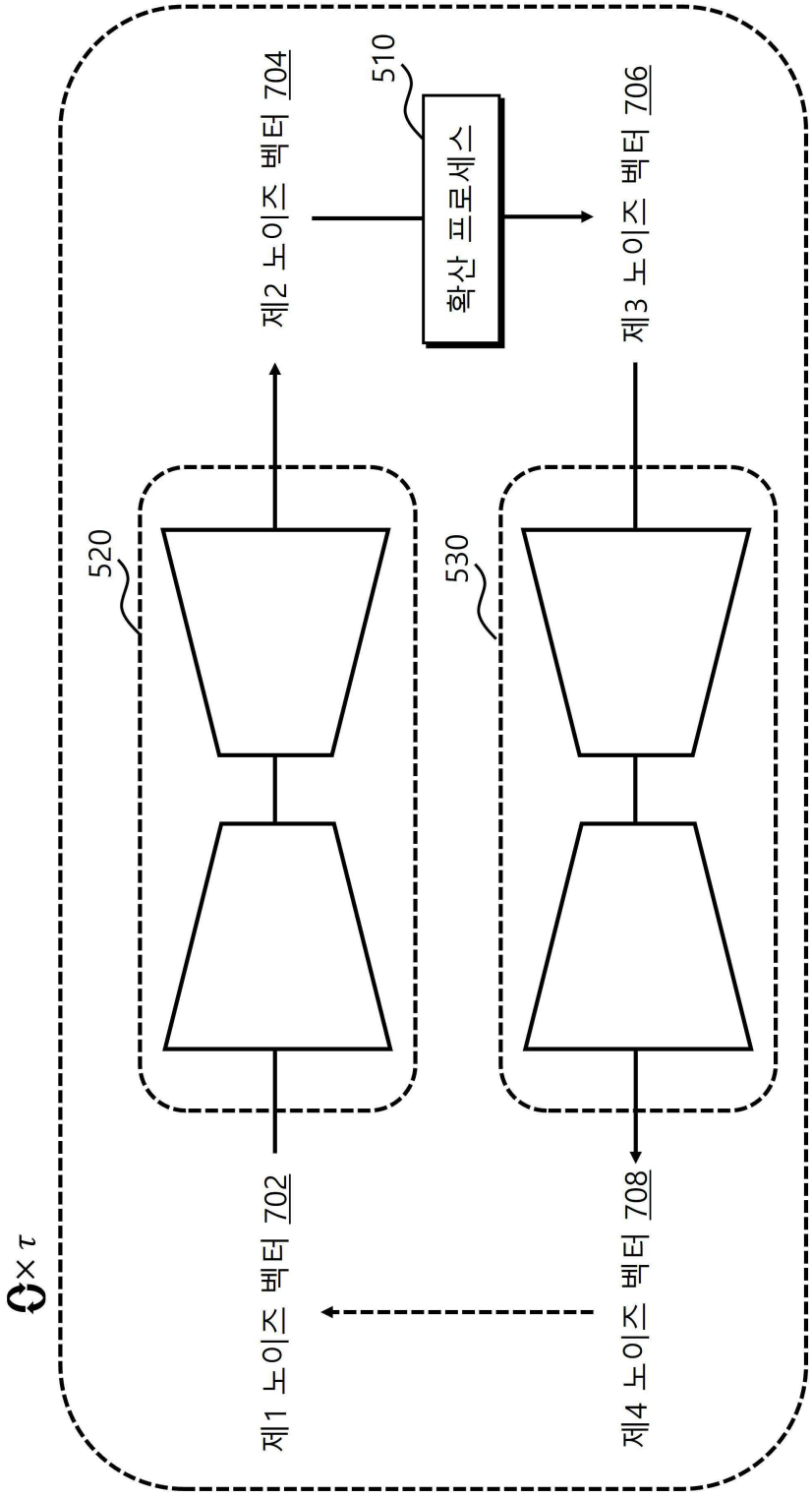
【도 5】


시각적 특징 벡터(f_{vis}^{0}) 502
확산 프로세스
510
제1 노이즈 벡터(f_{vis}^{τ}) 504
520
D
A
B
텍스트 특징 벡터 512
제2 노이즈 벡터($\hat{f}_{vis}^{\tau-1}$) 506
제2 노이즈 벡터($\hat{f}_{vis}^{\tau-1}$) 506
530
모션 특징 벡터 514
제3 노이즈 벡터($\hat{f}_{vis}^{\tau-2}$) 508

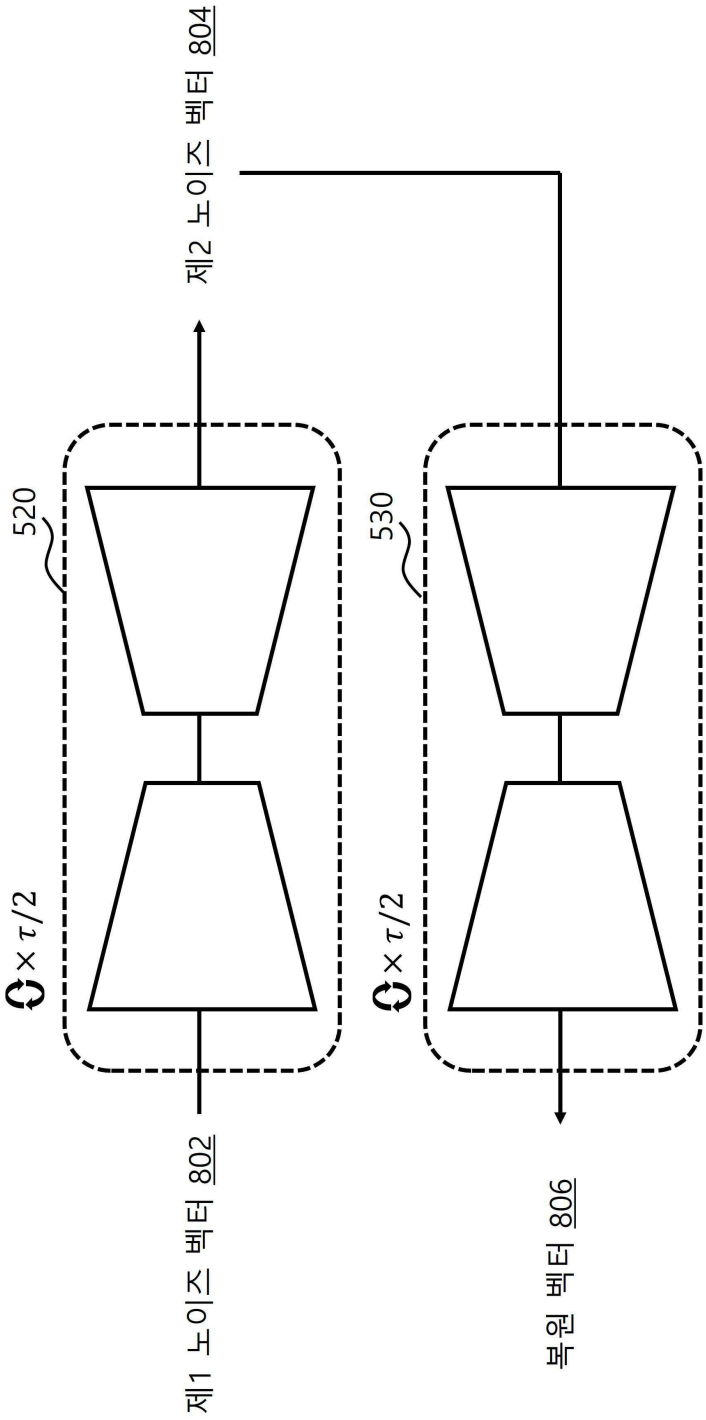
【도 6】


제1 노이즈 벡터 602
520
× τ
제1 복원 벡터 604
510
확산 프로세스
제2 노이즈 벡터 606
530
× τ
제2 복원 벡터 608

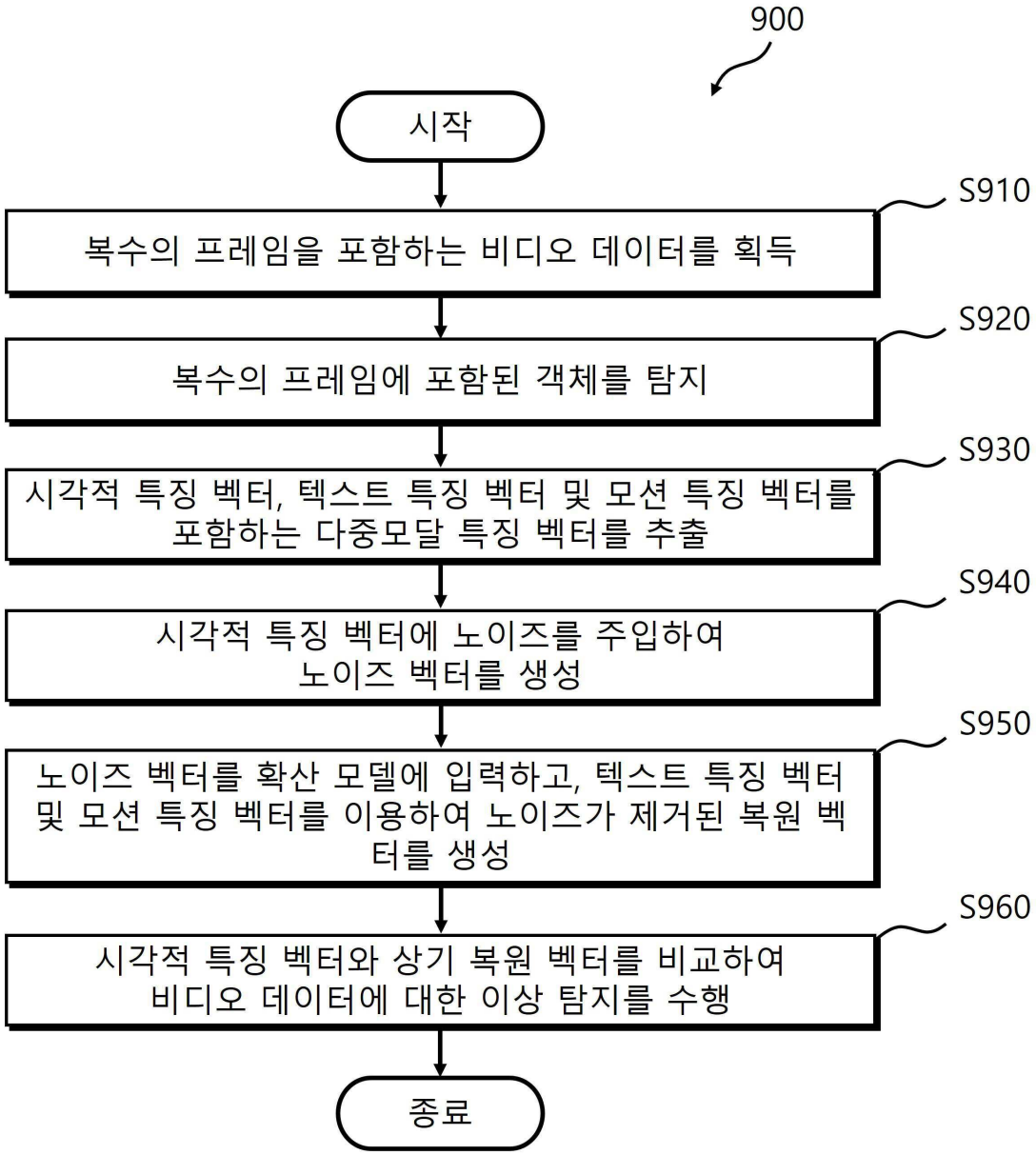
【도 7】


510
확산 프로세스
제2 노이즈 벡터 704
제3 노이즈 벡터 706
520
530
제1 노이즈 벡터 702
제4 노이즈 벡터 708
$\mathcal{G} \times \tau$

【도 8】


제2 노이즈 벡터 804
520
530
$\times\tau/2$
$\times\tau/2$
제1 노이즈 벡터 802
복원 벡터 806

【도 9】


900
시작
S910
복수의 프레임을 포함하는 비디오 데이터를 획득
S920
복수의 프레임에 포함된 객체를 탐지
S930
시각적 특징 벡터, 텍스트 특징 벡터 및 모션 특징 벡터를 포함하는 다중모달 특징 벡터를 추출
S940
시각적 특징 벡터에 노이즈를 주입하여 노이즈 벡터를 생성
S950
노이즈 벡터를 확산 모델에 입력하고, 텍스트 특징 벡터 및 모션 특징 벡터를 이용하여 노이즈가 제거된 복원 벡터를 생성
S960
시각적 특징 벡터와 상기 복원 벡터를 비교하여 비디오 데이터에 대한 이상 탐지를 수행
종료

【도 10】