

명세서

발명의 명칭: 영상 이상을 감지하는 방법 및 장치

기술분야

- [1] 본 발명은 영상에 포함된 이상 징후를 감지할 수 있는 영상 이상을 감지하는 방법 및 장치에 관한 것이다.

[2]

배경기술

- [3] 최근에는 사람이 직접 관찰하지 않고도 영상 속에 포함된 비정상적인 데이터를 판단하는 영상 이상 감지 작업이 다양한 분야에 적용되어 사용되고 있다. 예컨대, 제조 분야에서는 생산 공정에서 비정상적인 행동과 불규칙한 작업을 감지할 수 있고, 의료 분야에서는 환자의 상태를 모니터링하고 사고가 발생하면 자동으로 의료진에게 알람을 송출하도록 할 수 있다. 또한, 공공 안전 분야에서는 영상 속에서 불법 행위 또는 사고 행위와 같은 행위를 미리 인식하여 인명 및 재산을 보호하고자 하였다.
- [4] 이러한 영상 이상 감지 작업이 원활하게 수행되기 위해서는 정상적인 데이터와 비정상적인 데이터 각각이 균형적으로 확보될 필요가 있다. 그러나, 일상에서는 비정상적인 데이터에 대응되는 이벤트의 발생이 적음에 따라 비정상적인 데이터가 부족하여 정상적인 데이터와 비정상적인 데이터 간의 불균형으로 인한 원활한 영상 이상 감지 작업이 수행되기 어려운 문제가 있다. 또한, 정상적인 데이터와 비정상적인 데이터를 구분하는 기준이 명확하지 않아 원활한 영상 이상 감지 작업이 수행되기 어려운 문제도 있다.
- [5] 종래에는 이러한 문제를 해결하기 위해 영상 관찰자의 감독 관여 정도에 따라 영상에 포함된 시각적 특징에 대한 정보를 기반으로 영상 이상 감지 작업이 수행되도록 하였다. 그러나, 시각적 특징에 대한 정보만을 기반으로 영상에 포함된 의미론적 의미를 명확히 판단하기가 어렵고, 그로 인한 영상 이상 감지 작업의 정확성이 저하되는 문제가 빈번히 발생하고 있다.
- [6]
- [7] 상기의 배경기술로서 설명된 사항들은 본 발명의 배경에 대한 이해 증진을 위한 것일 뿐, 이 기술분야에서 통상의 지식을 가진 자에게 이미 알려진 종래기술에 해당함을 인정하는 것으로 받아들여져서는 안 될 것이다.
- [8]

발명의 상세한 설명

기술적 과제

- [9] 본 발명은 이러한 문제점을 해결하기 위하여 제안된 것으로, 영상에 포함된 시각적 특징 뿐만 아니라 텍스트 특징을 함께 이용하여 영상에 포함된 이상 징후를 감지하는 영상 이상을 감지하는 방법 및 장치를 제공하고자 함이다.

- [10]
- [11] 본 발명에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.
- [12]
- 과제 해결 수단**
- [13] 상기의 목적을 달성하기 위한 본 발명에 따른 영상 이상을 감지하는 방법은, 적어도 하나의 영상 정보를 수집하여 상기 적어도 하나의 영상 정보를 복수의 스니펫으로 분할하는 단계; 상기 분할된 복수의 스니펫 각각에 대한 시각적 특징 정보 및 텍스트 특징 정보를 추출하는 단계; 상기 추출된 시각적 특징 정보 및 텍스트 특징 정보를 합성하여 합성 정보를 생성하는 단계; 및 상기 생성된 합성 정보를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지하는 단계;를 포함할 수 있다.
- [14] 예를 들어, 상기 추출하는 단계는 상기 분할된 복수의 스니펫 각각에 대해 기 설정된 크롭횟수에 따라 프레임이 크롭되는 전처리를 수행하는 단계; 및 상기 전처리가 수행된 복수의 스니펫 각각에 대한 상기 시각적 특징 정보를 추출하는 단계;를 포함할 수 있다.
- [15] 예를 들어, 상기 추출하는 단계는 상기 분할된 복수의 스니펫 각각에 포함된 캡션을 생성하는 단계; 및 상기 생성된 캡션을 기반으로 문장 임베딩을 수행하여 상기 텍스트 특징 정보를 추출하는 단계;를 포함할 수 있다.
- [16] 예를 들어, 상기 추출하는 단계는 상기 복수의 스니펫 각각에 대한 상기 시각적 특징 정보 및 상기 텍스트 특징 정보에 다중 스케일에 따른 보정을 수행하는 단계; 및 상기 보정이 수행된 시각적 특징 정보 및 텍스트 특징 정보를 추출하는 단계;를 포함할 수 있다.
- [17] 예를 들어, 상기 생성하는 단계는 상기 추출된 시각적 특징 정보 및 텍스트 특징 정보를 복수의 합성 방식 중 적어도 하나의 방식을 이용하여 상기 합성 정보를 생성하는 단계;를 포함할 수 있다.
- [18] 예를 들어, 상기 감지하는 단계는 상기 생성된 합성 정보를 기반으로 상기 적어도 하나의 영상 정보의 총 손실값에 대응되는 제1 손실값 및 상기 적어도 하나의 영상 정보의 이상 징후를 감지하기 위한 학습값에 대응되는 제2 손실값을 각각 판단하는 단계; 상기 판단된 제1 손실값 및 제2 손실값을 기반으로 손실함수를 생성하는 단계; 및 상기 생성된 손실함수를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지하는 단계;를 포함할 수 있다.
- [19] 예를 들어, 상기 제1 손실값 및 제2 손실값을 각각 판단하는 단계는 상기 생성된 합성 정보를 기반으로 상기 복수의 스니펫 각각에 대응되는 특징 크기를 판단하는 단계; 상기 판단된 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는

특징 크기들의 평균 특징 크기를 판단하는 단계; 및 상기 판단된 평균 특징 크기를 기반으로 상기 제1 손실값을 판단하는 단계;를 포함할 수 있다.

[20] 예를 들어, 상기 제1 손실값 및 제2 손실값을 각각 판단하는 단계는 상기 생성된 합성 정보를 기반으로 상기 복수의 스니펫 각각에 대응되는 특징 크기를 판단하는 단계; 상기 판단된 특징 크기를 기반으로 상기 복수의 스니펫 각각의 프레임 수준 예측값을 판단하는 단계; 상기 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기에 대응되는 프레임 수준 예측값을 기반으로 상기 적어도 하나의 영상 정보의 이상 점수를 판단하는 단계; 및 상기 판단된 이상 점수를 기반으로 상기 제2 손실값을 판단하는 단계;를 포함할 수 있다.

[21] 예를 들어, 상기 감지하는 단계 이후, 상기 이상 징후가 감지되면 상기 적어도 하나의 영상 정보의 이상 점수 및 상기 적어도 하나의 영상 정보에 포함된 캡션을 출력하는 단계;를 더 포함할 수 있다.

[22] 예를 들어, 상기 출력하는 단계는 상기 적어도 하나의 영상 정보의 이상 점수가 기 설정된 기준 점수를 초과하면 경고 알람을 출력하는 단계;를 포함할 수 있다.

[23]

[24] 또한, 상기의 목적을 달성하기 위한 본 발명에 따른 영상 이상을 감지하는 장치는, 적어도 하나의 영상 정보를 수집하여 상기 적어도 하나의 영상 정보를 복수의 스니펫으로 분할하는 영상 수집부; 상기 영상 수집부에서 분할된 복수의 스니펫 각각에 대한 시각적 특징 정보 및 텍스트 특징 정보를 추출하는 정보 추출부; 및 상기 추출된 시각적 특징 정보 및 텍스트 특징 정보를 합성하여 합성 정보를 생성하고, 상기 생성된 합성 정보를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지하는 감지 제어부;를 포함할 수 있다.

[25] 예를 들어, 상기 정보 추출부는 상기 분할된 복수의 스니펫 각각에 대해 기 설정된 크롭횟수에 따라 프레임이 크롭되는 전처리를 수행하고, 상기 전처리가 수행된 복수의 스니펫 각각에 대한 상기 시각적 특징 정보를 추출할 수 있다.

[26] 예를 들어, 상기 정보 추출부는 상기 분할된 복수의 스니펫 각각에 포함된 캡션을 생성하고, 상기 생성된 캡션을 기반으로 문장 임베딩을 수행하여 상기 텍스트 특징 정보를 추출할 수 있다.

[27] 예를 들어, 상기 정보 추출부는 상기 복수의 스니펫 각각에 대한 상기 시각적 특징 정보 및 상기 텍스트 특징 정보에 다중 스케일에 따른 보정을 수행하고, 상기 보정이 수행된 시각적 특징 정보 및 텍스트 특징 정보를 추출할 수 있다.

[28] 예를 들어, 상기 감지 제어부는 상기 추출된 시각적 특징 정보 및 텍스트 특징 정보를 복수의 합성 방식 중 적어도 하나의 방식을 이용하여 상기 합성 정보를 생성할 수 있다.

[29] 예를 들어, 상기 감지 제어부는 상기 생성된 합성 정보를 기반으로 상기 적어도 하나의 영상 정보의 총 손실값에 대응되는 제1 손실값 및 상기 적어도 하나의 영상 정보의 이상 징후를 감지하기 위한 학습값에 대응되는 제2 손실값을 각각 판단하고, 상기 판단된 제1 손실값 및 제2 손실값을 기반으로 손실함수를 생성하

며, 상기 생성된 손실함수를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지할 수 있다.

[30] 예를 들어, 상기 감지 제어부는 상기 생성된 합성 정보를 기반으로 상기 복수의 스니펫 각각에 대응되는 특징 크기를 판단하고, 상기 판단된 특징 크기 중 크기 순으로 기 설정된 기준개수에 해당하는 특징 크기들의 평균 특징 크기를 판단하며, 상기 판단된 평균 특징 크기를 기반으로 상기 제1 손실값을 판단할 수 있다.

[31] 예를 들어, 상기 감지 제어부는 상기 생성된 합성 정보를 기반으로 상기 복수의 스니펫 각각에 대응되는 특징 크기를 판단하고, 상기 판단된 특징 크기를 기반으로 상기 복수의 스니펫 각각의 프레임 수준 예측값을 판단하며, 상기 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기에 대응되는 프레임 수준 예측값을 기반으로 상기 적어도 하나의 영상 정보의 이상 점수를 판단하고, 상기 판단된 이상 점수를 기반으로 상기 제2 손실값을 판단할 수 있다.

[32] 예를 들어, 상기 감지 제어부는 상기 이상 징후가 감지되면 상기 적어도 하나의 영상 정보의 이상 점수 및 상기 적어도 하나의 영상 정보에 포함된 캡션이 출력되도록 할 수 있다.

[33] 예를 들어, 상기 감지 제어부는 상기 적어도 하나의 영상 정보의 이상 점수가 기 설정된 기준 점수를 초과하면 경고 알람이 출력되도록 할 수 있다.

[34]

발명의 효과

[35] 상술한 바에 따르면, 본 발명의 영상 이상을 감지하는 방법 및 장치는 영상에서 추출된 시각적 특징 및 텍스트 특징을 기반으로 영상에 포함된 이상 징후를 감지함으로써 영상 이상을 감지함에 있어 정확도를 향상시킬 수 있다.

[36] 그리고, 시각적 특징 및 텍스트 특징 각각에 대해 다중 스케일에 따른 보정을 수행함으로써 영상 이상을 감지하는 방법 및 장치의 견고성을 향상시킬 수 있다.

[37] 또한, 영상에 대한 이상 점수와 영상에 포함된 캡션이 출력되도록 함으로써 본 발명에 따른 영상 이상을 감지하는 방법 및 장치를 사용하는 사용자가 영상에 포함된 이상 징후에 대해 효율적으로 대응할 수 있다.

[38] 아울러, 시각적 특징 뿐만 아니라 텍스트 특징을 함께 이용하여 영상에 포함된 이상 징후의 의미론적 의미를 포착할 수 있어 다양한 분야에 적용시킬 수 있다.

[39]

[40] 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

[41]

도면의 간단한 설명

[42] 도 1 내지 4는 본 발명의 일 실시예에 따른 영상 이상을 감지하는 장치의 구성을 설명하기 위한 블록도이다.

[43] 도 5는 본 발명의 일 실시예에 따른 영상 이상을 감지하는 방법을 설명하기 위한 순서도이다.

[44]

발명의 실시를 위한 형태

[45] 본 명세서에 개시된 실시 예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 명세서에 개시된 실시 예의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 첨부된 도면은 본 명세서에 개시된 실시 예를 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 명세서에 개시된 기술적 사상이 제한되지 않으며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[46] 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.

[47] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[48] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다.

[49] 본 명세서에서, "포함한다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[50] 이하, 첨부된 도면을 참조하여 본 명세서에 개시된 실시 예를 상세히 설명하되, 도면 부호에 관계없이 동일하거나 유사한 구성요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.

[51] 도 1 내지 4를 참조하여 본 발명의 실시예에 따른 영상 이상을 감지하는 장치에 대해 설명하도록 한다.

[52]

[53] 도 1 내지 4는 본 발명의 일 실시예에 따른 영상 이상을 감지하는 장치의 구성을 설명하기 위한 블록도이다.

[54] 도 1을 참조하면, 본 발명의 실시예에 따른 영상 이상을 감지하는 장치는 영상 수집부(100), 정보 추출부(200) 및 감지 제어부(300)를 포함할 수 있다. 도 1은 본

발명의 실시예와 관련된 구성 요소를 위주로 나타낸 것으로, 실제 영상 이상을 감지하는 장치를 구현함에 있어서는 이보다 더 적거나 많은 구성 요소를 포함할 수 있음은 물론이다.

- [55] 한편, 본 발명의 실시예에 따른 영상 이상을 감지하는 장치의 구성 요소 각각 (100, 200, 300)은 다른 구성 요소와 통신하는 통신 장치, 운영체제나 로직 명령어와 입출력 정보 등을 저장하는 메모리 및 담당 기능 제어에 필요한 판단, 연산, 결정 등을 수행하는 하나 이상의 프로세서를 포함할 수 있다.
- [56] 이하에서는 도 2 내지 4를 참조하여 각 구성 요소에 대해 설명하도록 한다.
- [57] 먼저, 도 2를 참조하여 본 발명의 실시예에 따른 영상 수집부(100)에 대해 설명한다.
- [58] 도 2를 참조하면, 영상 수집부(100)는 적어도 하나의 영상 정보를 수집할 수 있고, 적어도 하나의 영상 정보를 복수의 스니펫으로 분할할 수 있다. 예컨대, 영상 수집부(100)는 실시간으로 영상을 촬영하도록 구비된 외부 촬영장치에서 촬영된 적어도 하나의 영상 정보를 입력 정보로써 수집할 수 있다. 그리고, 영상 수집부(100)는 적어도 하나의 영상 정보를 복수의 스니펫으로 분할하되, 적어도 하나의 영상 정보를 사전 설정된 개수(예컨대, T개)만큼 복수의 스니펫(SN₁, SN₂, ..., SN_T)으로 분할할 수 있다.
- [59] 또한, 영상 수집부(100)는 적어도 하나의 영상 정보를 복수의 스니펫으로 분할하고, 적어도 하나의 영상 정보 간에 서로 대응되는 스니펫을 그룹핑하여 관리할 수도 있다. 다만, 이는 예시적인 것으로 반드시 이에 한정되지 않음은 물론이다. 예컨대, 영상 수집부(100)는 외부 촬영장치에서 촬영된 영상 정보만을 수집하도록 구성될 수 있으며, 그에 따라 영상 정보를 복수의 스니펫으로 분할하는 기능은 영상 수집부(100)와 별도의 장치로 분리되어 수행될 수도 있다.
- [60] 영상 수집부(100)는 분할된 복수의 스니펫에 대한 정보를 정보 추출부(200)에 전달할 수 있다.
- [61] 다음으로는 도 3을 참조하여 본 발명의 실시예에 따른 정보 추출부(200)에 대해 설명한다.
- [62]
- [63] 도 3을 참조하면, 정보 추출부(200)는 분할된 복수의 스니펫에 대한 정보를 제공받고, 이를 기반으로 복수의 스니펫 각각에 대한 시각적 특징 정보 및 텍스트 특징 정보를 추출할 수 있다. 이를 위해, 정보 추출부(200)는 시각적 특징 추출부(210) 및 텍스트 특징 추출부(220)를 포함할 수 있다. 그리고, 시각적 특징 추출부(210)는 전처리부(211) 및 다중 스케일 보정부(212)를 포함할 수 있으며, 텍스트 특징 추출부(220)는 캡션 생성부(221), 문장 임베딩부(222) 및 다중 스케일 보정부(223)를 포함할 수 있다. 도 3은 본 발명의 실시예와 관련된 구성 요소를 위주로 나타낸 것으로, 실제 정보 추출부(200)를 구현함에 있어서는 이보다 더 적거나 많은 구성 요소를 포함할 수 있음은 물론이다.

- [64] 이하, 각 구성 요소에 대해 설명한다.
- [65] 시각적 특징 추출부(210)는 영상 수집부(100)에서 제공된 복수의 스니펫에 대한 정보를 기반으로 복수의 스니펫 각각에 포함된 시각적 특징 정보를 추출할 수 있다. 예컨대, 시각적 특징 추출부(210)는 사전 학습된 신경망인 ResNet-50을 통해 복수의 스니펫 각각에 포함된 시각적 특징 정보를 추출할 수 있으며, 시각적 특징 정보는 I3D(Inflated 3D Convnet) 방식에 따른 정보일 수 있다. 다만, 이는 예시적인 것으로 반드시 이에 한정되지 않음은 물론이다.
- [66] 그리고, 시각적 특징 추출부(210)는 시각적 특징 정보 추출시 정보의 정합성을 높이기 위해 전처리부(211)를 통해 전처리 과정을 수행할 수 있다. 예를 들어, 전처리부(211)는 분할된 복수의 스니펫 각각에 대해 기 설정된 크롭횟수에 따라 프레임이 크롭(crop)되는 전처리를 수행할 수 있다. 예컨대, 기 설정된 크롭횟수가 5회인 경우, 전처리부(211)는 복수의 스니펫 각각의 프레임이 네 모서리 부분과 중앙 부분으로 크롭(crop)되도록 할 수 있다.
- [67] 시각적 특징 추출부(210)는 복수의 스니펫이 전처리부(211)를 거쳐 전처리가 수행되면, 전처리가 수행된 복수의 스니펫 각각에 대한 시각적 특징 정보를 추출할 수 있다. 복수의 스니펫 각각에 대해 프레임이 크롭되는 전처리를 수행함으로써 기존 복수의 스니펫에 대한 정보보다 처리할 수 있는 정보의 양이 증가하게 되고, 이를 통해 시각적 특징 정보를 추출함으로써 정보의 정합성을 높일 수 있다.
- [68] 텍스트 특징 추출부(220)는 영상 수집부(100)에서 제공된 복수의 스니펫에 대한 정보를 기반으로 복수의 스니펫 각각에 포함된 텍스트 특징 정보를 추출할 수 있다. 예컨대, 텍스트 특징 추출부(220)는 캡션 생성부(221)를 통해 복수의 스니펫 각각에 포함된 캡션(caption)을 생성할 수 있다. 이때, 캡션 생성부(221)는 복수의 스니펫 각각에 대하여 슬라이딩 윈도우 전략을 사용하여 연속된 프레임에 대한 캡션을 생성할 수도 있다. 또한, 캡션은 복수의 스니펫 각각에 포함된 이벤트를 설명하는 것으로, 예를 들어, 복수의 스니펫 각각에 포함된 사람의 동작 및 사물의 배치 상태 등을 설명하는 것일 수 있다. 다만, 이는 예시적인 것으로 반드시 이에 한정되지 않음은 물론이다.
- [69] 캡션 생성부(221)는 생성된 캡션에 대한 정보를 문장 임베딩부(222)에 전달할 수 있으며, 문장 임베딩부(222)는 생성된 캡션을 기반으로 문장 임베딩을 수행할 수 있다. 예컨대, 문장 임베딩부(222)는 드롭아웃 노이즈(Drop-out Noise) 방식을 이용하거나 자연어 처리 방식(NLP: Natural Language Processing) 중 하나인 자연어 추론 데이터셋에 기록된 쌍을 기반으로 문장 임베딩을 수행할 수 있다. 다만, 이는 예시적인 것으로, 반드시 이에 한정되지 않음은 물론이다. 텍스트 특징 추출부(220)는 캡션 생성부(221) 및 문장 임베딩부(222)를 통해 복수의 스니펫 각각에 대하여 캡션을 생성하고, 문장 임베딩을 수행함으로써 텍스트 특징 정보를 추출할 수 있다.

- [70] 한편, 시각적 특징 추출부(210) 및 텍스트 특징 추출부(220)는 시각적 특징 정보 및 텍스트 특징 정보를 외부로 전달하기 이전, 복수의 스니펫 각각에 대해 1차적으로 추출된 시각적 특징 정보 및 텍스트 특징 정보를 다중 스케일에 따른 보정을 수행할 수 있다. 이를 위해, 시각적 특징 추출부(210) 및 텍스트 특징 추출부(220)에는 각각 다중 스케일 보정부(212, 223)가 구비될 수 있다.
- [71] 예컨대, 다중 스케일 보정부(212, 223)는 시간적 네트워크(Temporal Network)를 통해 1차적으로 추출된 시각적 특징 정보 및 텍스트 특징 정보를 다중 스케일에 따른 보정을 수행할 수 있다. 시간적 네트워크는 다층 피라미드 확장 컨벌루션(Multiple Layer Pyramid Dilated Convolution) 영역과 비로컬 영역(NLB : Non-Local Block)이 포함될 수 있다. 시간 범위에 대한 피라미드 확장 컨벌루션은 여러 스케일로 복수의 스니펫에 대한 표현을 학습하는데 사용될 수 있으며, 비로컬 영역은 복수의 스니펫 간의 전반적인 시간적 종속성을 학습하는데 사용될 수 있다. 다중 스케일 보정부(212, 223)에서 시각적 특징 정보 및 텍스트 특징 정보 각각에 시간적 네트워크를 통한 다중 스케일에 따른 보정을 수행함으로써 이후 시각적 특징 추출부(210) 및 텍스트 특징 추출부(220)에서 보정되어 추출된 시각적 특징 정보 및 텍스트 특징 정보의 합성시 시간적 종속성이 잘 포착되도록 할 수 있다.
- [72] 상술한 바와 같이, 정보 추출부(200)는 시각적 특징 추출부(210)를 통해 시각적 특징 정보를 추출하고, 텍스트 특징 추출부(220)를 통해 텍스트 특징 정보를 추출할 수 있으며, 추출된 시각적 특징 정보와 텍스트 특징 정보를 감지 제어부(300)에 제공할 수 있다.
- [73] 다음으로는 도 4를 참조하여 본 발명의 실시예에 따른 감지 제어부(300)에 대해 설명한다.
- [74]
- [75] 도 4를 참조하면, 본 발명의 실시예에 따른 감지 제어부(300)는 정보 추출부(200)에서 추출된 시각적 특징 정보 및 텍스트 특징 정보를 합성하여 합성 정보를 생성하고, 생성된 합성 정보를 기반으로 적어도 하나의 영상 정보의 이상 징후를 감지할 수 있다. 이를 위해, 감지 제어부(300)는 합성부(310), 판단부(320) 및 이상 감지부(330)를 포함할 수 있다. 도 4는 본 발명의 실시예와 관련된 구성 요소를 위주로 나타낸 것으로, 실제 감지 제어부(300)를 구현함에 있어서는 이보다 더 적거나 많은 구성 요소를 포함할 수 있음은 물론이다.
- [76] 이하, 각 구성 요소에 대해 설명한다.
- [77] 합성부(310)는 정보 추출부(200)에서 추출되어 제공된 시각적 특징 정보 및 텍스트 특징 정보를 합성하여 합성 정보를 생성하되, 복수의 합성 방식 중 적어도 하나의 합성 방식을 이용하여 합성 정보를 생성할 수 있다. 복수의 합성 방식에는 연결(Concatenation), 더하기(Addition) 및 곱하기(Product) 방식이 존재할 수 있으나, 이는 예시적인 것으로 반드시 이에 한정되지 않음은 물론이다.
- [78] 그리고, 더하기 및 곱하기 방식을 통해 합성 정보가 생성될 경우, 합성부(310)는 시각적 특징 정보에 대한 차원을 텍스트 특징 정보에 대한 차원과 동일하게 줄이

기 위해 완전 연결된 레이어를 추가할 수 있으며, 이후 시각적 특징 정보와 텍스트 특징 정보에 대한 합성 정보를 생성할 수 있다. 다만, 이는 예시적인 것으로 반드시 이에 한정되지 않음은 물론이다.

[79] 판단부(320)는 합성부(310)에서 생성된 합성 정보를 기반으로 영상 손실값을 판단할 수 있으며, 영상 손실값을 판단하기 위한 제1 손실값 및 제2 손실값을 판단할 수 있다. 예컨대, 제1 손실값은 영상 수집부(100)에서 수집된 적어도 하나의 영상 정보의 총 손실값에 대응되는 값일 수 있으며, 제2 손실값은 적어도 하나의 영상 정보의 이상 징후를 감지하기 위한 학습값에 대응되는 값일 수 있다.

[80] 구체적으로, 판단부(320)는 합성부(310)에서 생성된 합성 정보를 기반으로 복수의 스니펫 각각에 대응되는 특징 크기를 판단할 수 있다. 이때, 판단부(320)는 l_2 norm 방식을 이용한 특징 크기를 판단할 수 있으나, 이는 예시적인 것으로 반드시 이에 한정되지 않음은 물론이다.

[81] 그리고, 판단부(320)는 판단된 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기들의 평균 특징 크기를 판단할 수 있다. 예컨대, 기 설정된 기준개수가 k 개인 경우, 판단부(320)는 복수의 스니펫 각각에 대응되는 특징 크기 중 크기순으로 상위 k 개에 해당하는 특징 크기들을 추출하고, 이를 기반으로 평균 특징 크기를 판단할 수 있다. 평균 특징 크기($f_{FM}(v;k)$)는 아래 수학식 1을 통해 판단될 수 있다. 특히, 평균 특징 크기는 적어도 하나의 영상 정보에 대응되는 특징 크기로 판단될 수도 있다.

[82] [수식1]

$$f_{FM}(v;k)=\frac{1}{k} \sum_{X_i \in topK(v;k)} \| X_i \|_2$$

[83] 판단부(320)는 위 수학식 1을 통해 판단된 평균 특징 크기를 기반으로 아래의 수학식 2를 이용하여 적어도 하나의 영상 정보의 총 손실값에 대응되는 제1 손실값(

$$L_{fm}$$

)을 판단할 수 있다.

[84] [수식2]

$$L_{fm}=\begin{cases} \sum_{j=1}^{|v|} (c-f_{FM}(v_j;k)), \text{if } y_j=1 \\ \sum_{j=1}^{|v|} f_{FM}(v_j;k), \text{if } y_j=0 \end{cases}$$

[85] 제1 손실값(

$$L_{fm}$$

)은 적어도 하나의 영상 정보에서 정상적인 데이터와 비정상적인 데이터를 모두 포함하는 총 손실값을 의미할 수 있다.

[86] 또한, 판단부(320)는 생성된 합성 정보를 기반으로 복수의 스니펫 각각에 대응되는 특징 크기를 판단하고, 이를 기반으로 복수의 스니펫 각각의 프레임 수준 예측값을 판단할 수 있다. 예를 들어, 판단부(320)는 판단된 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기들을 기반으로 별도 구비된 이진 스니펫 분류기(Binary Snippet Classifier)를 학습시킬 수 있다. 그리고, 판단부(320)는 학습된 이진 스니펫 분류기를 기반으로 복수의 스니펫 각각에 포함된 개별 프레임으로 전파되는 스니펫 수준 예측값을 판단할 수 있으며, 판단된 스니펫 수준 예측값을 기반으로 복수의 스니펫 각각의 프레임 수준 예측값을 판단할 수 있다.

[87] 그리고, 판단부(320)는 복수의 스니펫 각각에 대응되는 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기에 대응되는 프레임 수준 예측값을 기반으로 적어도 하나의 영상 정보의 이상 점수를 판단할 수 있다. 판단부(320)는 복수의 스니펫 각각에 대응되는 프레임 수준 예측값 중 특징 크기 순으로 상위 기준개수에 해당하는 값을 기반으로 적어도 하나의 영상 정보 전체에 대한 이상 점수를 판단할 것을 가정한다. 이에 판단부(320)는 아래 수학적식 3을 통해 적어도 하나의 영상 정보에 대한 이상 점수(

$$f_s(v;k)$$

)를 판단할 수 있다.

[88] [수식3]

$$f_s(v;k)=\frac{1}{k}\sum_{X_i\in topK(v;k)}f_{pred}(X_i;\delta)$$

[89] 이때,

$$f_{pred}(X_i;\delta)$$

는 복수의 스니펫 각각에 대응되는 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기에 대응되는 프레임 수준 예측값을 의미할 수 있다. 본 발명은 정상적인 데이터를 포함하고 있는 영상 정보의 이상 점수와 비정상적인 데이터를 포함하고 있는 영상 정보의 이상 점수의 차이가 커지도록 하여 정상적인 데이터와 비정상적인 데이터의 구분을 용이하게 할 것을 부수적인 목적으로 한다. 이를 달성하기 위해, 복수의 스니펫 각각에 대응되는 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기를 기반으로 이진 스니펫 분류기를 학습시킬 수 있다. 그리고, 학습된 이진 스니펫 분류기에서 판단된 프레임 수준 예측값을 기반으로 영상 정보의 이상 점수를 판단함으로써 정상적인 데이터를 포함하고 있는 영상 정보의 이상 점수와 비정상적인 데이터를 포함하고 있는 영상 정보의 이상 점수의 차이가 커지게 할 수 있다.

[90] 이후, 판단부(320)는 위 수학적식 3을 통해 판단된 적어도 하나의 영상 정보에 대한 이상 점수(

$$f_s(v;k)$$

)를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지하기 위한 학습 값에 대응되는 제2 손실값을 판단할 수 있다. 이때, 판단부(320)는 이진 교차 엔트로피(Binary Cross Entropy)를 기반으로 형성된 아래 수학적 식 4를 통해 제2 손실값(L_{bce})을 판단할 수 있다.

[91] [수식4]

$$L_{bce} = -\frac{1}{|v|} \sum_{j=1}^{|v|} (y_j \log(f_s(v_j; k)) + (1 - y_j) \log(1 - f_s(v_j; k)))$$

[92] 그리고, 판단부(320)는 상술한 바와 같이 판단된 제1 손실값(L_{fm}) 및 제2 손실값(L_{bce})을 기반으로 손실함수를 생성할 수 있다. 이때, 판단부(320)는 아래 수학적 식 5를 이용하여 손실함수(L)를 생성할 수 있다.

[93] [수식5]

$$L = \alpha L_{fm} + L_{bce}$$

[94] 위 수학적 식 5에서 α 는 손실 구성 요소의 가중치를 조정하기 위한 매개변수로써, 본 발명의 실시예에서는 제1 손실값의 가중치를 조정할 수 있다. 즉, 손실함수는 위 수학적 식 5와 같이 매개변수 α , 제1 손실값(L_{fm}) 및 제2 손실값(L_{bce})을 기반으로 한 일차함수 형태로 생성될 수 있다. 다만, 이는 예시적인 것으로 반드시 이에 한정되지 않음은 물론이다.

[95] 판단부(320)는 생성된 손실함수를 이상 감지부(330)에 제공할 수 있으며, 이상 감지부(330)는 생성된 손실함수를 기반으로 적어도 하나의 영상 정보의 이상 징후를 감지할 수 있다. 그리고, 이상 감지부(330)는 적어도 하나의 영상 정보에서

이상 징후가 감지되면, 외부 사용자가 이상 징후에 대해 인식할 수 있도록 감지된 이상 징후에 대한 정보가 출력되도록 할 수 있다.

- [96] 예를 들어, 이상 감지부(330)는 적어도 하나의 영상 정보에서 이상 징후가 감지되면, 적어도 하나의 영상 정보의 이상 점수 및 적어도 하나의 영상 정보에 포함된 캡션이 출력되도록 할 수 있다. 이때, 적어도 하나의 영상 정보의 이상 점수는 적어도 하나의 영상 정보를 분할한 복수의 스니펫에 대한 이상 점수일 수 있고, 적어도 하나의 영상 정보에 포함된 캡션은 적어도 하나의 영상 정보를 분할한 복수의 스니펫에 대한 캡션일 수 있다.
- [97] 또한, 이상 감지부(330)는 상술한 적어도 하나의 영상 정보의 이상 점수 및 적어도 하나의 영상 정보에 포함된 캡션을 출력하고자 할 때 경고 알람도 함께 출력되도록 할 수 있다. 예를 들어, 이상 감지부(330)는 이상 징후 감지시 판단된 적어도 하나의 영상 정보의 이상 점수를 기반으로 기 설정된 기준 점수와 비교할 수 있다. 이상 감지부(330)는 앞선 비교 결과, 적어도 하나의 영상 정보의 이상 점수가 기 설정된 기준 점수를 초과하면 경고 알람을 생성할 수 있다. 이상 감지부(330)가 경고 알람을 생성하고, 생성된 경고 알람이 출력되도록 함으로써 외부 사용자가 인식하고 이상 징후에 대응하도록 할 수 있다. 다만, 이는 예시적인 것으로 반드시 이에 한정되지 않음은 물론이다.
- [98]
- [99] 이하에서는 도 1 내지 4를 참조하여 상술한 영상 이상을 감지하는 장치의 구성을 바탕으로 본 발명의 실시예에 따른 영상 이상을 감지하는 방법을 도 5를 참조하여 설명한다. 단, 도 5의 각 단계에 대한 구체적인 설명은 도 1 내지 4를 참조하여 상술하였으므로, 이하에서는 생략하기로 한다.
- [100] 도 5는 본 발명의 일 실시예에 따른 영상 이상을 감지하는 방법을 설명하기 위한 순서도이다.
- [101] 도 5를 참조하면, 영상 수집부(100)는 외부의 촬영장치로부터 적어도 하나의 영상 정보를 수집할 수 있고(S510), 수집된 적어도 하나의 영상 정보를 복수의 스니펫으로 분할할 수 있다(S520).
- [102] 그리고, 정보 추출부(200)는 분할된 복수의 스니펫을 기반으로 복수의 스니펫 각각에 대한 시각적 특징 정보 및 텍스트 특징 정보를 추출할 수 있다(S530).
- [103] 그리고, 감지 제어부(300)는 정보 추출부(200)에서 추출된 시각적 특징 정보 및 텍스트 특징 정보를 합성하여 합성 정보를 생성할 수 있고(S540), 생성된 합성 정보를 기반으로 적어도 하나의 영상 정보의 총 손실값에 대응되는 제1 손실값 및 적어도 하나의 영상 정보의 이상 징후를 감지하기 위한 학습값에 대응되는 제2 손실값을 각각 판단할 수 있다(S550).
- [104] 감지 제어부(300)는 판단된 제1 손실값 및 제2 손실값을 기반으로 손실함수를 생성할 수 있으며(S560), 생성된 손실함수를 기반으로 적어도 하나의 영상 정보에 대한 이상 감지를 수행할 수 있다(S570).

- [105] 이후, 이상 감지 수행시 적어도 하나의 영상 정보에서 이상 징후가 감지될 경우 (S580의 Yes), 감지 제어부(300)는 적어도 하나의 영상 정보의 이상 점수 및 적어도 하나의 영상 정보에 포함된 캡션이 출력되도록 할 수 있다(S590). 또한, 감지 제어부(300)는 적어도 하나의 영상 정보의 이상 점수가 기 설정된 기준 점수를 초과할 경우, 이상 점수 및 캡션 출력시 경고 알람이 함께 출력되도록 할 수도 있다.
- [106]
- [107] 상술한 바에 따르면, 본 발명의 영상 이상을 감지하는 방법 및 장치는 영상에서 추출된 시각적 특징 및 텍스트 특징을 기반으로 영상에 포함된 이상 징후를 감지함으로써 영상 이상을 감지함에 있어 정확도를 향상시킬 수 있다.
- [108] 그리고, 시각적 특징 및 텍스트 특징 각각에 대해 다중 스케일에 따른 보정을 수행함으로써 영상 이상을 감지하는 방법 및 장치의 견고성을 향상시킬 수 있다.
- [109] 또한, 영상에 대한 이상 점수와 영상에 포함된 캡션이 출력되도록 함으로써 본 발명에 따른 영상 이상을 감지하는 방법 및 장치를 사용하는 사용자가 영상에 포함된 이상 징후에 대해 효율적으로 대응할 수 있다.
- [110] 아울러, 시각적 특징 뿐만 아니라 텍스트 특징을 함께 이용하여 영상에 포함된 이상 징후의 의미론적 의미를 포착할 수 있어 다양한 분야에 적용시킬 수 있다.
- [111]
- [112] 본 발명의 특정한 실시예에 관련하여 도시하고 설명하였지만, 이하의 청구범위에 의해 제공되는 본 발명의 기술적 사상을 벗어나지 않는 한도 내에서, 본 발명이 다양하게 개량 및 변화될 수 있다는 것은 당 업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.
- [113] 전술한 본 발명은, 프로그램이 기록된 매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 매체는, 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 매체의 예로는, HDD(Hard Disk Drive), SSD(Solid State Disk), SDD(Silicon Disk Drive), ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있다. 따라서, 상기의 상세한 설명은 모든 면에서 제한적으로 해석되어서는 아니되고 예시적인 것으로 고려되어야 한다. 본 발명의 범위는 첨부된 청구항의 합리적 해석에 의해 결정되어야 하고, 본 발명의 등가적 범위 내에서의 모든 변경은 본 발명의 범위에 포함된다.
- [114]
- [115] [부호의 설명]
- [116] 100 : 영상 수집부
- [117] 200 : 정보 추출부
- [118] 300 : 감지 제어부

청구범위

- [청구항 1] 적어도 하나의 영상 정보를 수집하여 상기 적어도 하나의 영상 정보를 복수의 스니펫으로 분할하는 단계;
상기 분할된 복수의 스니펫 각각에 대한 시각적 특징 정보 및 텍스트 특징 정보를 추출하는 단계;
상기 추출된 시각적 특징 정보 및 텍스트 특징 정보를 합성하여 합성 정보를 생성하는 단계; 및
상기 생성된 합성 정보를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지하는 단계;를 포함하는 영상 이상을 감지하는 방법.
- [청구항 2] 청구항 1에 있어서,
상기 추출하는 단계는
상기 분할된 복수의 스니펫 각각에 대해 기 설정된 크롭횟수에 따라 프레임이 크롭되는 전처리를 수행하는 단계; 및
상기 전처리가 수행된 복수의 스니펫 각각에 대한 상기 시각적 특징 정보를 추출하는 단계;를 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.
- [청구항 3] 청구항 1에 있어서,
상기 추출하는 단계는
상기 분할된 복수의 스니펫 각각에 포함된 캡션을 생성하는 단계; 및
상기 생성된 캡션을 기반으로 문장 임베딩을 수행하여 상기 텍스트 특징 정보를 추출하는 단계;를 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.
- [청구항 4] 청구항 1에 있어서,
상기 추출하는 단계는
상기 복수의 스니펫 각각에 대한 상기 시각적 특징 정보 및 상기 텍스트 특징 정보에 다중 스케일에 따른 보정을 수행하는 단계; 및
상기 보정이 수행된 시각적 특징 정보 및 텍스트 특징 정보를 추출하는 단계;를 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.
- [청구항 5] 청구항 1에 있어서,
상기 생성하는 단계는
상기 추출된 시각적 특징 정보 및 텍스트 특징 정보를 복수의 합성 방식 중 적어도 하나의 방식을 이용하여 상기 합성 정보를 생성하는 단계;를 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.
- [청구항 6] 청구항 1에 있어서,
상기 감지하는 단계는
상기 생성된 합성 정보를 기반으로 상기 적어도 하나의 영상 정보의 총 손실값에 대응되는 제1 손실값 및 상기 적어도 하나의 영상 정보의 이상 징

후를 감지하기 위한 학습값에 대응되는 제2 손실값을 각각 판단하는 단계;
 상기 판단된 제1 손실값 및 제2 손실값을 기반으로 손실함수를 생성하는 단계; 및
 상기 생성된 손실함수를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지하는 단계;를 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.

[청구항 7]

청구항 6에 있어서,
 상기 제1 손실값 및 제2 손실값을 각각 판단하는 단계는
 상기 생성된 합성 정보를 기반으로 상기 복수의 스니펫 각각에 대응되는 특징 크기를 판단하는 단계;
 상기 판단된 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기들의 평균 특징 크기를 판단하는 단계; 및
 상기 판단된 평균 특징 크기를 기반으로 상기 제1 손실값을 판단하는 단계;를 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.

[청구항 8]

청구항 6에 있어서,
 상기 제1 손실값 및 제2 손실값을 각각 판단하는 단계는
 상기 생성된 합성 정보를 기반으로 상기 복수의 스니펫 각각에 대응되는 특징 크기를 판단하는 단계;
 상기 판단된 특징 크기를 기반으로 상기 복수의 스니펫 각각의 프레임 수준 예측값을 판단하는 단계;
 상기 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기에 대응되는 프레임 수준 예측값을 기반으로 상기 적어도 하나의 영상 정보의 이상 점수를 판단하는 단계; 및
 상기 판단된 이상 점수를 기반으로 상기 제2 손실값을 판단하는 단계;를 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.

[청구항 9]

청구항 1에 있어서,
 상기 감지하는 단계 이후,
 상기 이상 징후가 감지되면 상기 적어도 하나의 영상 정보의 이상 점수 및 상기 적어도 하나의 영상 정보에 포함된 캡션을 출력하는 단계;를 더 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.

[청구항 10]

청구항 9에 있어서,
 상기 출력하는 단계는
 상기 적어도 하나의 영상 정보의 이상 점수가 기 설정된 기준 점수를 초과하면 경고 알람을 출력하는 단계;를 포함하는 것을 특징으로 하는 영상 이상을 감지하는 방법.

[청구항 11]

적어도 하나의 영상 정보를 수집하여 상기 적어도 하나의 영상 정보를 복수의 스니펫으로 분할하는 영상 수집부;

상기 영상 수집부에서 분할된 복수의 스니펫 각각에 대한 시각적 특징 정보 및 텍스트 특징 정보를 추출하는 정보 추출부; 및
 상기 추출된 시각적 특징 정보 및 텍스트 특징 정보를 합성하여 합성 정보를 생성하고, 상기 생성된 합성 정보를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지하는 감지 제어부;를 포함하는 영상 이상을 감지하는 장치.

- [청구항 12] 청구항 11에 있어서,
 상기 정보 추출부는
 상기 분할된 복수의 스니펫 각각에 대해 기 설정된 크롭횟수에 따라 프레임이 크롭되는 전처리를 수행하고, 상기 전처리가 수행된 복수의 스니펫 각각에 대한 상기 시각적 특징 정보를 추출하는 것을 특징으로 하는 영상 이상을 감지하는 장치.
- [청구항 13] 청구항 11에 있어서,
 상기 정보 추출부는
 상기 분할된 복수의 스니펫 각각에 포함된 캡션을 생성하고, 상기 생성된 캡션을 기반으로 문장 임베딩을 수행하여 상기 텍스트 특징 정보를 추출하는 것을 특징으로 하는 영상 이상을 감지하는 장치.
- [청구항 14] 청구항 11에 있어서,
 상기 정보 추출부는
 상기 복수의 스니펫 각각에 대한 상기 시각적 특징 정보 및 상기 텍스트 특징 정보에 다중 스케일에 따른 보정을 수행하고, 상기 보정이 수행된 시각적 특징 정보 및 텍스트 특징 정보를 추출하는 것을 특징으로 하는 영상 이상을 감지하는 장치.
- [청구항 15] 청구항 11에 있어서,
 상기 감지 제어부는
 상기 추출된 시각적 특징 정보 및 텍스트 특징 정보를 복수의 합성 방식 중 적어도 하나의 방식을 이용하여 상기 합성 정보를 생성하는 것을 특징으로 하는 영상 이상을 감지하는 장치.
- [청구항 16] 청구항 11에 있어서,
 상기 감지 제어부는
 상기 생성된 합성 정보를 기반으로 상기 적어도 하나의 영상 정보의 총 손실값에 대응되는 제1 손실값 및 상기 적어도 하나의 영상 정보의 이상 징후를 감지하기 위한 학습값에 대응되는 제2 손실값을 각각 판단하고, 상기 판단된 제1 손실값 및 제2 손실값을 기반으로 손실함수를 생성하며, 상기 생성된 손실함수를 기반으로 상기 적어도 하나의 영상 정보의 이상 징후를 감지하는 것을 특징으로 하는 영상 이상을 감지하는 장치.
- [청구항 17] 청구항 16에 있어서,
 상기 감지 제어부는

상기 생성된 합성 정보를 기반으로 상기 복수의 스니펫 각각에 대응되는 특징 크기를 판단하고, 상기 판단된 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기들의 평균 특징 크기를 판단하며, 상기 판단된 평균 특징 크기를 기반으로 상기 제1 손실값을 판단하는 것을 특징으로 하는 영상 이상을 감지하는 장치.

- [청구항 18] 청구항 16에 있어서,
상기 감지 제어부는
상기 생성된 합성 정보를 기반으로 상기 복수의 스니펫 각각에 대응되는 특징 크기를 판단하고, 상기 판단된 특징 크기를 기반으로 상기 복수의 스니펫 각각의 프레임 수준 예측값을 판단하며, 상기 특징 크기 중 크기순으로 기 설정된 기준개수에 해당하는 특징 크기에 대응되는 프레임 수준 예측값을 기반으로 상기 적어도 하나의 영상 정보의 이상 점수를 판단하고, 상기 판단된 이상 점수를 기반으로 상기 제2 손실값을 판단하는 것을 특징으로 하는 영상 이상을 감지하는 장치.
- [청구항 19] 청구항 11에 있어서,
상기 감지 제어부는
상기 이상 징후가 감지되면 상기 적어도 하나의 영상 정보의 이상 점수 및 상기 적어도 하나의 영상 정보에 포함된 캡션이 출력되도록 하는 것을 특징으로 하는 영상 이상을 감지하는 장치.
- [청구항 20] 청구항 19에 있어서,
상기 감지 제어부는
상기 적어도 하나의 영상 정보의 이상 점수가 기 설정된 기준 점수를 초과하면 경고 알람이 출력되도록 하는 것을 특징으로 하는 영상 이상을 감지하는 장치.