



등록특허 10-2156818



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2020년09월16일  
(11) 등록번호 10-2156818  
(24) 등록일자 2020년09월10일

- (51) 국제특허분류(Int. Cl.)  
*G06K 9/00* (2006.01) *G06T 7/20* (2017.01)
- (52) CPC특허분류  
*G06K 9/00771* (2013.01)  
*G06T 7/20* (2013.01)
- (21) 출원번호 10-2017-0148159  
(22) 출원일자 2017년11월08일  
심사청구일자 2020년05월04일
- (65) 공개번호 10-2018-0054453  
(43) 공개일자 2018년05월24일
- (30) 우선권주장  
16198678.1 2016년11월14일  
유럽특허청(EPO)(EP)
- (56) 선행기술조사문헌  
US09158974 B1  
WO2013093173 A1

- (73) 특허권자  
**엑시스 에이비**  
스웨덴왕국 룬트 에스-223 69, 엠달라베겐 14
- (72) 발명자  
**다니엘손 니클라스**  
스웨덴 룬트 223 69 엠달라베겐 14 엑시스 커뮤니  
케이션스 에이비 씨/오  
**몰린 시몬**  
스웨덴 룬트 223 69 엠달라베겐 14 엑시스 커뮤니  
케이션스 에이비 씨/오
- (74) 대리인  
**박장원**

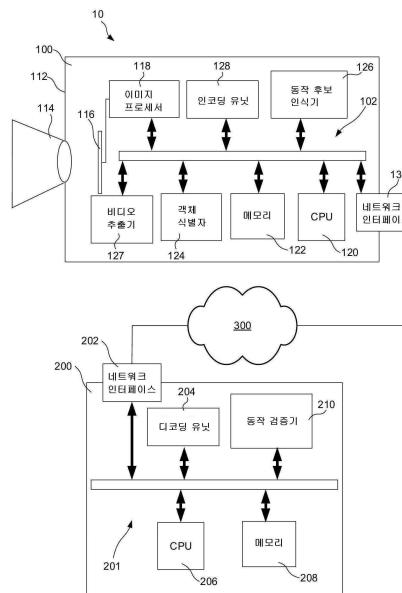
전체 청구항 수 : 총 12 항

심사관 : 하은주

(54) 발명의 명칭 **비디오 시퀀스에서 동작 인식****(57) 요 약**

본 발명은 비디오 시퀀스에서 동작 인식을 위한 시스템에 관한 것이다. 상기 시스템은 상기 비디오 시퀀스를 캡처하도록 구성된 카메라(100) 및 동작 인식을 수행하도록 구성된 서버(200)를 포함한다. 상기 카메라는, 상기 비디오 시퀀스의 객체 이미지 프레임에서 관심 객체를 식별하도록 구성된 객체 식별자(124)와; 상기 객체 이미지

(뒷면에 계속)

**대 표 도** - 도1

프레임에 제1 동작 인식 알고리즘을 적용하여 동작 후보의 존재를 검출하도록 구성된 동작 후보 인식기(126)와; 상기 비디오 시퀀스로부터 복수의 이미지 프레임들에 관한 비디오 데이터를 추출함으로써 동작 비디오 시퀀스의 동작 이미지 프레임들을 생성하도록 구성된 비디오 추출기(127) - 상기 비디오 데이터가 추출되는 상기 복수의 이미지 프레임들 중 하나 이상은 상기 관심 객체를 포함하며 - 와; 그리고 상기 동작 비디오 시퀀스를 상기 서버로 전송하도록 구성된 네트워크 인터페이스(130)를 포함한다. 상기 서버는, 제2 동작 인식 알고리즘을 상기 동작 비디오 시퀀스에 적용하여, 상기 동작 후보가 동작임을 확인 또는 거부하는 동작 검증기(210)를 포함한다. 또한, 비디오 시퀀스에서 동작 인식을 위한 방법이 제시된다.

(52) CPC특허분류

***H04N 5/144*** (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

카메라에 의해 캡처된 비디오 시퀀스에서의 동작 인식을 위한 방법으로서,

상기 방법은, 상기 카메라의 회로에 의해,

상기 비디오 시퀀스의 이미지 프레임에서 관심 객체(object of interest)를 식별하는 단계와;

동작 후보를 검출하기 위해 상기 이미지 프레임에 제1 동작 인식 알고리즘을 적용하는 단계 - 상기 이미지 프레임은 상기 관심 객체를 포함하는 단일 이미지이고, 상기 제1 동작 인식 알고리즘은 상기 이미지 프레임 내의 상기 동작 후보를 검출하기 위해 단일 이미지 프레임의 콘텍스트 및/또는 공간 인식 정보를 사용하며 - 와;

상기 비디오 시퀀스로부터 복수의 이미지 프레임들에 관한 비디오 데이터를 추출함으로써 동작 비디오 시퀀스의 이미지 프레임들을 생성하는 단계 - 상기 비디오 데이터가 추출되는 상기 복수의 이미지 프레임들 중 하나 이상은 상기 관심 객체를 포함하며 - 와; 그리고

동작 인식을 수행하는 서버로 상기 동작 비디오 시퀀스를 전송하는 단계를 포함하고,

상기 방법은, 상기 서버의 회로에 의해, 제2 동작 인식 알고리즘을 상기 동작 비디오 시퀀스에 적용하여, 상기 동작 후보가 미리 정의된 유형의 동작임을 확인 또는 거부하는 단계를 포함하며,

상기 제2 동작 인식 알고리즘은 상기 동작 비디오 시퀀스의 복수의 이미지 프레임들의 시간 정보를 이용하는 것을 특징으로 하는

동작 인식을 위한 방법.

#### 청구항 2

삭제

#### 청구항 3

제1항에 있어서,

상기 동작 비디오 시퀀스의 이미지 프레임들을 생성하는 단계는, 상기 관심 객체를 포함하는 상기 이미지 프레임들이 상기 관심 객체의 적어도 일부를 포함하도록 상기 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑(cropping)하는 단계를 포함하는 것을 특징으로 하는

동작 인식을 위한 방법.

#### 청구항 4

제3항에 있어서,

상기 관심 객체를 포함하는 상기 동작 비디오 시퀀스의 이미지 프레임들은 상기 관심 객체를 적어도 부분적으로 둘러싸는 배경의 일 부분을 포함하는 것을 특징으로 하는

동작 인식을 위한 방법.

#### 청구항 5

제1항에 있어서,

상기 동작 비디오 시퀀스를 전송하는 단계는 상기 동작 비디오 시퀀스 내의 좌표들을 상기 관심 객체로 전송하는 단계를 포함하는 것을 특징으로 하는

동작 인식을 위한 방법.

## 청구항 6

제1항에 있어서,

상기 방법은, 상기 카메라의 회로에 의해,

상기 비디오 시퀀스에서 관심 객체를 검출하는 단계를 더 포함하고,

상기 동작 비디오 시퀀스의 이미지 프레임들을 생성하는 단계는, 상기 관심 객체의 검출 이전의 시점에 관련된 상기 비디오 시퀀스의 제1 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하는 단계를 포함하는 것을 특징으로 하는

동작 인식을 위한 방법.

## 청구항 7

제1항에 있어서,

상기 방법은, 상기 카메라의 회로에 의해,

상기 비디오 시퀀스에서 관심 객체를 검출하는 단계를 더 포함하고,

상기 동작 비디오 시퀀스의 이미지 프레임들을 생성하는 단계는 상기 관심 객체의 검출 이후의 시점에 관련된 상기 비디오 시퀀스의 제2 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하는 단계를 포함하는 것을 특징으로 하는

동작 인식을 위한 방법.

## 청구항 8

제1항에 있어서,

상기 카메라 및 상기 서버는, 서로 거리를 두고 위치되고 그리고 디지털 네트워크를 통해 서로 통신하는 별개의 물리적 개체들인 것을 특징으로 하는

동작 인식을 위한 방법.

## 청구항 9

비디오 시퀀스에서 동작 인식을 위한 시스템으로서,

상기 비디오 시퀀스를 캡처하는 카메라 및 동작 인식을 수행하는 서버를 포함하고,

상기 카메라는:

상기 비디오 시퀀스의 이미지 프레임에서 관심 객체를 식별하는 객체 식별자와;

동작 후보를 검출하기 위해 상기 이미지 프레임에 제1 동작 인식 알고리즘을 적용하는 동작 후보 인식기 - 상기 이미지 프레임은 상기 관심 객체를 포함하는 단일 이미지이고, 상기 제1 동작 인식 알고리즘은 상기 이미지 프레임 내의 상기 동작 후보를 검출하기 위해 단일 이미지 프레임의 콘텍스트 및/또는 공간 인식 정보를 사용하며 - 와;

상기 비디오 시퀀스로부터 복수의 이미지 프레임들에 관한 비디오 데이터를 추출함으로써 동작 비디오 시퀀스의 이미지 프레임들을 생성하는 비디오 추출기 - 상기 비디오 데이터가 추출되는 상기 복수의 이미지 프레임들 중 하나 이상은 상기 관심 객체를 포함하며 - 와; 그리고

상기 동작 비디오 시퀀스를 상기 서버로 전송하는 네트워크 인터페이스를 포함하고,

상기 서버는:

제2 동작 인식 알고리즘을 상기 동작 비디오 시퀀스에 적용하여, 상기 동작 후보가 미리 정의된 유형의 동작임을 확인 또는 거부하는 동작 검증기를 포함하며,

상기 제2 동작 인식 알고리즘은 상기 동작 비디오 시퀀스의 복수의 이미지 프레임들의 시간 정보를 이용하는 것

을 특징으로 하는  
동작 인식을 위한 시스템.

### 청구항 10

제9항에 있어서,

상기 비디오 추출기는, 또한, 상기 관심 객체를 포함하는 상기 비디오 시퀀스의 이미지 프레임들이 상기 관심 객체의 적어도 일부를 포함하도록 상기 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑하는 것을 특징으로 하는

동작 인식을 위한 시스템.

### 청구항 11

제9항에 있어서,

상기 비디오 추출기는, 또한, 상기 관심 객체를 포함하는 상기 비디오 시퀀스의 이미지 프레임들이 상기 관심 객체를 적어도 부분적으로 둘러싸는 배경의 일 부분을 포함하도록 상기 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑하는 것을 특징으로 하는

동작 인식을 위한 시스템.

### 청구항 12

제9항에 있어서,

상기 객체 식별자는, 또한, 상기 비디오 시퀀스에서 관심 객체를 검출하고,

상기 비디오 추출기는, 또한, 상기 관심 객체의 검출 이전의 시점에 관련된 상기 비디오 시퀀스의 제1 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하는 것을 특징으로 하는

동작 인식을 위한 시스템.

### 청구항 13

제9항에 있어서,

상기 객체 식별자는, 또한, 상기 비디오 시퀀스에서 관심 객체를 검출하고,

상기 비디오 추출기는, 또한, 상기 관심 객체의 검출 이후의 시점에 관련된 상기 비디오 시퀀스의 제2 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하는 것을 특징으로 하는

동작 인식을 위한 시스템.

### 청구항 14

삭제

## 발명의 설명

### 기술 분야

[0001] 본 발명은 비디오 시퀀스에서 동작 인식에 관한 것이다.

### 배경 기술

[0002] 비디오 시퀀스의 동작 인식은 비디오 시퀀스에서 미리 정의된 특정 동작들을 검출하는 작업이다. 검출될 동작의 예들은, 예를 들어, 싸우는 것, 달리기, 먹기, 스포츠 게임하는 것일 수 있다. 동작 인식은 일반적으로 비디오 시퀀스를 캡처하는 디지털 비디오 카메라에서 수행하기에는 부적합할 수 있는 과도한 프로세싱 능력을 요구한다. 대신, 동작 인식은 통상적으로 필요한 프로세싱 능력을 가진 서버에 의해 수행된다. 하지만, 디지털 네트워크들을 통해 많은 양의 비디오를 전송하는 것은 다수의 대역폭을 요구한다.

[0003] 따라서, 비디오 시퀀스에서 동작 인식을 향상시킬 필요가 있다.

### 발명의 내용

#### 해결하려는 과제

[0004] 상기를 고려하여, 본 발명의 목적은 비디오 시퀀스에서 개선된 동작 인식을 제공하는 것이다.

#### 과제의 해결 수단

[0005] 제1 양상에 따라, 카메라에 의해 캡처된 비디오 시퀀스에서의 동작 인식을 위한 방법이 제공된다. 상기 방법은, 상기 카메라의 회로에 의해, 상기 비디오 시퀀스의 객체 이미지 프레임에서 관심 객체를 식별하는 단계와; 상기 객체 이미지 프레임에 제1 동작 인식 알고리즘을 적용하여 동작 후보의 존재를 검출하는 단계와; 상기 비디오 시퀀스로부터 복수의 이미지 프레임들에 관한 비디오 데이터를 추출함으로써 동작 비디오 시퀀스의 동작 이미지 프레임들을 생성하는 단계 - 상기 비디오 데이터가 추출되는 상기 복수의 이미지 프레임들 중 하나 이상은 상기 관심 객체를 포함하며 - 와; 그리고 동작 인식을 수행하는 서버로 상기 동작 비디오 시퀀스를 전송하는 단계를 포함하고, 상기 방법은, 상기 서버의 회로에 의해, 제2 동작 인식 알고리즘을 상기 동작 비디오 시퀀스에 적용하여, 상기 동작 후보가 미리 정의된 유형의 동작임을 확인 또는 거부하는 단계를 포함한다.

[0006] 동작 인식에 대한 본 방법은, 카메라와 서버 간의 통신에서 지나치게 많은 대역폭을 차지하는 문제없이 동작 인식을 위해 필요한 프로세싱 전력이 2개의 상이한 디바이스들에 분산되도록하기 때문에 바람직하다. 또한, 동작 인식에서의 본 방법은, 서버에 전체 비디오 시퀀스를 끊임없이 스트리밍할 필요없이 동시에 카메라의 프로세서를 너무 많이 로딩하지 않으면서 프로세서 집중 동작 인식을 수행하는 것을 가능하게 한다. 따라서, 본 방법은 카메라상에서 국부적으로 제1 유형의 동작 인식 알고리즘을 요구하는 프로세싱을 적게 사용하고, 서버상에서 제2 유형의 동작 인식 알고리즘을 요구하는 프로세싱을 보다 많이 사용하게 한다. 연속적인 비디오 스트림이 아닌 동작 비디오 시퀀스만을 추출하여 전송함으로써 대역폭이 절약될 수 있다. 동작 인식에서의 본 방법에 따르면, 후보 동작은 카메라에서 검출될 수 있다. 후보 동작은 동작 비디오 시퀀스를 서버로 전달하는 것을 야기하며, 여기서 더 진보된 동작 인식 분석이 수행된다. 따라서, 카메라상의 동작 인식의 목적은 동작 후보들의 검출을 야기하는 것이다. 동작 후보의 검출은 추가 분석을 위해 서버로 동작 비디오 시퀀스를 전송하도록 초래하고, 동작 후보를 실제 동작으로 확인 또는 거부한다.

[0007] 상기 제1 동작 인식 알고리즘은 주로 객체 이미지 프레임에서 콘텍스트 및/또는 공간 정보를 사용하는 콘텍스트 및/또는 공간 동작 인식 알고리즘에 기초할 수 있다. 콘텍스트 및/또는 공간 동작 인식 알고리즘은 일반적으로 프로세싱 전력을 그다지 많이 요구하지 않는다. 따라서, 이러한 종류의 알고리즘들을 사용하는 동작 인식은 카메라에서 로컬적으로 보다 쉽게 수행할 수 있다.

[0008] 상기 제2 동작 인식 알고리즘은 주로 상기 동작 비디오 시퀀스의 복수의 이미지 프레임들의 시간 정보를 이용하는 시간 동작 인식 알고리즘에 기초할 수 있다. 시간 동작 인식 알고리즘들은 통상적으로 동작을 인식할 때 더 정확하다. 따라서, 더 정확한 동작 인식이 수행될 수 있다.

[0009] 상기 동작 이미지 프레임들을 생성하는 단계는, 상기 관심 객체를 포함하는 상기 동작 이미지 프레임들이 상기 관심 객체의 적어도 일부를 포함하도록 상기 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑(cropping)하는 단계를 포함할 수 있다. 이는, 카메라와 서버간의 대역폭을 절약할 수 있다.

[0010] 상기 동작 이미지 프레임들을 생성하는 단계는, 상기 관심 객체를 포함하는 상기 동작 이미지 프레임들은 상기 관심 객체를 적어도 부분적으로 둘러싸는 배경의 일 부분을 포함하도록 상기 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑하는 단계를 포함할 수 있다. 상기 관심 객체를 적어도 부분적으로 둘러싸는 배경을 포함시킴으로써, 상기 동작 비디오 시퀀스의 시간적 분석뿐만 아니라 문맥적 분석 및/또는 공간적 분석이 수행될 수 있다.

[0011] 상기 동작 비디오 시퀀스를 전송하는 단계는 상기 동작 비디오 시퀀스 내의 좌표들을 상기 관심 객체로 전송하는 단계를 포함할 수 있다. 상기 좌표들은, 관심 객체를 포함하는 그리고/또는 각각의 이미지 프레임에서 관심 객체가 위치되는 이미지 프레임 또는 이미지 프레임들을 참조할 수 있다.

[0012] 상기 방법은, 상기 카메라의 회로에 의해, 상기 비디오 시퀀스에서 관심 객체를 검출하는 단계를 더 포함할 수 있다. 상기 동작 이미지 프레임들을 생성하는 단계는, 상기 관심 객체의 검출 이전의 시점에 관련된 상기 비디오 시퀀스의 제1 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하는 단계를 포함할 수 있다. 상기 동작 이미지 프레임들을 생성하는 단계는, 상기 관심 객체의 검출 이후의 시점에 관련된 상기 비디오 시퀀스에서 관심 객체를 검출하는 단계를 더 포함할 수 있다.

스의 제2 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하는 단계를 포함할 수 있다. 이를 통해, 서버에서 동작을 인식할 수 있는 올바른 임시 창을 서버로 보낼 수 있는 가능성을 높일 수 있다.

[0013] 상기 카메라 및 서버는, 서로 거리를 두고 위치되는 별도의 물리적 개체일 수 있다. 상기 카메라 및 상기 서버는, 디지털 네트워크를 통해 서로 통신하도록 구성될 수 있다.

[0014] 제2 양상에 따라, 비디오 시퀀스에서 동작 인식을 위한 시스템이 제공된다. 상기 시스템은, 상기 비디오 시퀀스를 캡처하는 카메라 및 동작 인식을 수행하는 서버를 포함한다. 상기 카메라는, 상기 비디오 시퀀스의 객체 이미지 프레임에서 관심 객체를 식별하는 객체 식별자와; 상기 객체 이미지 프레임에 제1 동작 인식 알고리즘을 적용하여 동작 후보의 존재를 검출하는 동작 후보 인식기와; 상기 비디오 시퀀스로부터 복수의 이미지 프레임들에 관한 비디오 데이터를 추출함으로써 동작 비디오 시퀀스의 동작 이미지 프레임들을 생성하는 비디오 추출기 - 상기 비디오 데이터가 추출되는 상기 복수의 이미지 프레임들 중 하나 이상은 상기 관심 객체를 포함하며 - 와; 그리고 상기 동작 비디오 시퀀스를 상기 서버로 전송하는 네트워크 인터페이스를 포함한다. 상기 서버는, 제2 동작 인식 알고리즘을 상기 동작 비디오 시퀀스에 적용하여, 상기 동작 후보가 미리 정의된 유형의 동작임을 확인 또는 거부하는 동작 검증기를 포함한다.

[0015] 상기 비디오 추출기는, 또한, 상기 관심 객체를 포함하는 상기 동작 이미지 프레임들이 상기 관심 객체의 적어도 일부를 포함하도록 상기 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑하도록 구성될 수 있다.

[0016] 상기 비디오 추출기는, 또한, 상기 관심 객체를 포함하는 상기 동작 이미지 프레임들은 상기 관심 객체를 적어도 부분적으로 둘러싸는 배경의 일 부분을 포함하도록 상기 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑하도록 구성될 수 있다.

[0017] 상기 객체 식별자는, 또한, 상기 비디오 시퀀스에서 관심 객체를 검출하도록 구성될 수 있다. 상기 비디오 추출기는, 또한, 상기 관심 객체의 검출 이전의 시점에 관련된 상기 비디오 시퀀스의 제1 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하도록 구성될 수 있다. 상기 비디오 추출기는, 또한, 상기 관심 객체의 검출 이후의 시점에 관련된 상기 비디오 시퀀스의 제2 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하도록 구성될 수 있다.

[0018] 적용 가능한 경우 상기에서 언급한 방법의 피처들은 이 두 번째 양상에도 적용된다. 과도한 반복을 피하기 위해 위의 내용을 참조할 수 있다.

[0019] 본 발명의 또 다른 적용 범위는 이하의 상세한 설명으로부터 명백해질 것이다. 하지만, 본 발명의 범위 내의 다양한 변경들 및 수정들이 이 상세한 설명으로부터 통상의 기술자에게 명백해질 것이기 때문에, 본 발명의 바람직한 실시예를 나타내는 상세한 설명 및 특정예들은 단지 예시로서 주어진 것임을 이해해야 한다.

[0020] 따라서, 상기 서술된 디바이스 및 방법이 변할 수 있기 때문에, 본 발명은, 상기 서술된 디바이스의 특정 컴포넌트 부분들 또는 서술된 방법들의 단계들로 제한되지 않음이 이해되어야 한다. 또한, 여기에서 사용된 용어는 특정 실시예들을 설명하기 위한 것일 뿐이며, 제한하도록 의도되지는 않음을 알아야 한다. 명세서 및 첨부된 청구 범위에서 사용되는 바와 같이, 용어 "하나의" 및 "상기"는, 문맥에서 명확하게 다르게 나타내지 않으면, 하나 이상의 요소들이 존재함을 의도한다고 이해해야 한다. 따라서, 예를 들어, "하나의 유닛" 또는 "상기 유닛"에 대한 언급은 다수의 디바이스들을 포함할 수 있다. 또한, 단어 "포함하는" 및 유사한 의미의 단어들은 다른 요소들 또는 단계들을 배제하지 않는다.

## 도면의 간단한 설명

[0021] 본 발명의 상기 양상 및 다른 양상은 본 발명의 실시예들을 도시하는 첨부 된 도면들을 참조하여 보다 상세히 설명될 것이다. 도면들은 본 발명을 특정 실시 예로 제한하는 것으로 고려되어서는 안된다. 대신에, 도면들은 발명을 설명하고 이해하기 위해 사용된다.

도면들에 도시된 바와 같이, 층들 및 영역들의 크기들은 설명의 목적으로 과장되어 있으며, 따라서 본 발명의 실시예들의 일반적인 구조들을 예시하기 위해 제공된다. 동일한 도면 부호들은 동일한 요소들을 지칭한다.

도 1은 비디오 시퀀스에서 동작 인식을 위한 시스템을 도시한다.

도 2는 비디오 시퀀스에서 동작 인식 방법의 블록도이다.

## 발명을 실시하기 위한 구체적인 내용

- [0022] 본 발명은, 본 발명의 현재 바람직한 실시예들이 도시되는 첨부된 도면들을 참조하여 이하에서 상세하게 설명될 것이다. 그러나, 본 발명은 많은 상이한 형태들로 구체화될 수 있으며, 여기에 설명된 실시예들에 한정되는 것으로 해석되어서는 안된다. 오히려, 이들 실시예들은 완성도를 위해 제공되며, 본 발명의 범위를 당업자에게 충분하게 전달하기 위해 제공된다.
- [0023] 도 1은 비디오 시퀀스에서 동작 인식을 위한 시스템을 도시한다. 이 시스템은 디지털 네트워크 카메라(100) 및 서버(200)를 포함한다. 디지털 네트워크 카메라(100)는 디지털 네트워크(300)를 통해 서버(200)에 접속된다. 디지털 네트워크 카메라(100)와 서버(200)는 서로 떨어져 있는 개별적인 물리적 개체들로서, 디지털 네트워크(300)를 통해 서로 통신하도록 구성된다.
- [0024] 디지털 네트워크(300)는, 디바이스들, 예를 들어, 디지털 네트워크 카메라(100) 및 서버(200)가 접속되어 디지털 데이터를 교환하도록 하는 네트워크에 관한 것이다. 디지털 네트워크(300)에 접속된 디바이스들 간의 접속들은 케이블들 또는 무선을 사용하여 설립된다. 디지털 네트워크들의 비-제한적 예들은 인터넷, 인트라넷, 로컬 영역 네트워크 및 셀룰러 네트워크이다. 디지털 네트워크의 일부들은 사설 디지털 네트워크일 수 있다. 디지털 네트워크의 일부들은 공용 디지털 네트워크일 수 있다. 사설 디지털 네트워크는 네트워크 액세스 제한 디바이스(도시되지 않음)에 의해 공용 디지털 네트워크에 접속될 수 있다. 네트워크 액세스 제한 디바이스는 사설 디지털 네트워크를 보호하기 위해 설치된 방화벽 일 수 있다. 네트워크 액세스 제한 디바이스는 NAT(Network Address Translation)를 수행하는 디바이스일 수 있다.
- [0025] 디지털 네트워크 카메라(100)는 장면을 묘사하는 비디오 시퀀스를 캡처하도록 구성된다. 디지털 네트워크 카메라(100)는 하우징(112), 렌즈(114) 및 회로(102)를 포함한다. 디지털 네트워크 카메라(100)는 비디오 시퀀스를 캡처 및 프로세싱(그리고 가능하게는 저장)하도록 구성된다. 회로(102)는 이미지 센서(116), 이미지 프로세싱 유닛(118), 객체 식별자(124), 동작 후보 인식기(126), 비디오 추출기(127) 및 네트워크 인터페이스(130)를 포함한다. 회로(102)는 중앙 처리 유닛(CPU)(120), 디지털 데이터 저장 매체(메모리)(122) 및 인코딩 유닛(128) 중 하나 이상을 더 포함할 수 있다. 이미지 프로세싱 유닛(118), 객체 식별자(124), 동작 후보 인식기(126), 비디오 추출기(127) 및/또는 인코딩 유닛(128) 중 임의의 하나는 전용 하드웨어 회로 및/또는 소프트웨어 모듈로서 구현될 수 있다. 소프트웨어로 구현되는 경우, 소프트웨어는 CPU(120)상에서 실행될 수 있다. CPU(120)는 디지털 데이터 프로세싱을 수행하기 위한 임의의 적합한 CPU일 수 있다. 또한, 임의의 전용 하드웨어 회로는 부분적으로 전용 프로세서 또는 CPU(120) 상에서 실행되는 소프트웨어 부분들을 포함할 수 있음을 알아야 한다.
- [0026] 메모리(122)는 임의의 종류의 휴대용 또는 비-휴대용 메모리 일 수 있다. 또한, 메모리(122)는 복수의 메모리 유닛들을 포함할 수 있다. 복수의 메모리 유닛들 중 적어도 하나는, 예를 들어, 비디오 시퀀스의 콘텐트를 프로세싱하는 동안 데이터를 버퍼링하기 위한 버퍼 메모리로서 사용될 수 있다.
- [0027] 디지털 네트워크 카메라(100)는 네트워크 인터페이스(130)를 통해 디지털 네트워크(300)에 접속되도록 구성된다. 디지털 네트워크에 대한 접속은 유선 또는 무선일 수 있다. 따라서, 네트워크 인터페이스(130)는 이더넷 포트와 같은 10/100/1000 Mbps 데이터 트래픽에 적합한 네트워크 포트일 수 있고, 예를 들어, RJ45 커넥터와 같은 모듈러 커넥터를 수신하도록 구성된 모듈 포트일 수 있다. 일반적으로, 이러한 RJ45 커넥터 포트는 (예를 들어, cat 5, cat 5e 또는 cat 6의) 트위스트 페어 케이블과 같은 네트워크 케이블을 수신하도록 구성된다. 대안으로, 네트워크 포트의 I/O 수단은 모바일 인터넷 통신 표준(예를 들어, 1G, 2G, 2.5G, 2.75G, 3G, 3.5G, 3.75G, 3.9G, 4G, 5G)을 사용하는 또는 WiFi를 사용하는 무선 I/O 수단일 수 있다.
- [0028] 카메라 컴포넌트들, 즉, 렌즈(114) 및 이미지 센서(116)는 미가공 이미지들(raw images)을 캡처하도록 구성될 수 있고, 각각의 미가공 이미지는 상이한 과정들의 광으로 기술될 수 있고 그리고 상이한 객체들 및 객체들의 부분들로부터 기원할 수 있다. 그 후, 이러한 미가공 이미지들은 아날로그 포맷으로부터 디지털 포맷으로 변환되어 이미지 프로세싱 유닛(118)으로 전송된다. 이러한 실시예에 따르면, 디지털 네트워크 카메라(100)는 사진 이미지들을 캡처하도록 구성된 카메라이다. 대안으로 또는 조합하여, 디지털 네트워크 카메라(100)의 이미지 센서(116)는 열 이미지(thermal image)들을 캡처하도록 구성될 수 있다. 또한, 대안적으로 또는 조합하여, 디지털 네트워크 카메라(100)의 이미지 센서(116)는 레이더 이미지들을 캡처하도록 구성될 수 있다. 따라서, 디지털 네트워크 카메라(100)에 의해 캡처된 비디오 시퀀스는 사진 이미지들의 표현, 열 이미지들의 표현, 레이더 이미지들의 표현 또는 이들의 조합일 수 있다.
- [0029] 객체 식별자(124)는 카메라(100)에 의해 캡처된 비디오 시퀀스에서 관심 객체들을 검출하도록 구성된다. 관심 객체는, 예를 들어, 인간, 얼굴, 차량, 컨베이어 밴드 상의 제품, 동물, 지형 컴포넌트들, 무기일 수 있다. 객체 식별자(124)는 또한 검출된 관심 객체들을 분류하도록 구성될 수 있다. 관심 객체는, 예를 들어, 특정 유형

의 객체에 속하는 것으로 분류될 수 있다. 특정 유형의 객체들의 예들은 인간들, 얼굴들, 차량들, 특정 유형의 제품들이 있다. 객체 식별자(124)는, 또한, 관심 객체가 처음 검출된 비디오 시퀀스에서의 시점을 식별하도록 구성될 수 있다. 이와 관련하여, 메모리(122)는 미리 결정된 수의 이미지 프레임들을 저장하도록 구성된 이미지 프레임 버퍼로서 더 사용될 수 있다. 따라서, 먼저 검출된 관심 객체에서의 비디오 시퀀스에서 시점 이전의 이미지 프레임들을 나타내는 이미지 프레임들은, 이미지 프레임 버퍼로서 동작하는 메모리(122)에 저장될 수 있다.

[0030] 객체 식별자(124)는 또한 비디오 시퀀스의 하나 이상의 이미지 프레임들에서 관심 객체를 식별하도록 구성되며, 관심 객체가 식별된 이미지 프레임은 여기서 객체 이미지 프레임이라 언급될 것이다.

[0031] 비디오 시퀀스에서 동작 인식은 비디오 시퀀스에서 하나 이상의 미리 정의 된 유형들의 동작들을 검색하는 작업이다. 미리 정의된 유형들의 동작들의 예로는 싸우기, 달리기, 먹기, 특정 게임하기 등이 있다. 미리 정의된 유형들의 동작들의 다른 예는, 음주 운전 탐지, 점프 탐지, 분노 탐지, 미소 감지, 수신호 감지, 낙하 감지, 정체 감지, 보행 감지, 위협 행동 감지, 의심스러운 행동 감지(예를 들어, 비정상적인 행동의 감지)들이다.

[0032] 동작 인식은 정지 이미지 문맥 및/또는 공간 분석 또는 시간 분석(또는 이들의 조합)에 의해 수행될 수 있다. 문맥 및/또는 공간 동작 인식 알고리즘들은 정지 이미지들, 예를 들어, 비디오 시퀀스의 단일 이미지 프레임 상에서 수행된다. 시간 동작 인식 알고리즘들은 비디오 시퀀스의 복수의 이미지 프레임들에 대해 수행된다.

[0033] 문맥적 동작 인식 알고리즘의 예는, Georgia Gkioxari, Ross Girshick 및 Jitendra Malik에 의해 "R\*CNN과의 문맥 액션 인식"에서 공개되었다; arXiv : 1505.01197. 동작 인식 알고리즘들의 추가 예들로서, 문맥 및/또는 공간 동작 인식 알고리즘들 및 시간 동작 인식 알고리즘들은, 예를 들어, CN102855462, CN103106394, 및 Karen Simonyan 및 Andrew Zisserman에 의한 "비디오들에서 동작 인식을 위한 2-스트림 컨볼루션 네트워크들"에서 기술된다; arXiv : 1406.2199.

[0034] 따라서, 동작 인식에는 두 가지 주요 접근법들; 정지 이미지 문맥 및/또는 공간 분석 및 시간 분석을 가진다. 가장 유망한 접근법들은 시간 분석을 기본 동작 인식 알고리즘으로 사용하지만, 정지 이미지 접근법들은 일부 경우들에서 상당히 잘 작동한다. 하지만, 시간적 동작 인식은 어렵고 그리고, 예를 들어, 순환 신경망들을 사용하는 프로세싱 집약적인 알고리즘들을 포함한다. 이는, 그와 같은 시간 동작 인식 알고리즘들이 디지털 네트워크 카메라(100)와 같은 내장형 디바이스들 상에서 수행 되기에는 부적합하게 만든다.

[0035] 본 발명은 문맥 및/또는 공간 동작 인식 알고리즘들 사이의 시너지들을 사용하는 것에 관한 것으로, 종종 집중적으로 합리적인 처리를 하고, 프로세싱 집약적인 시간 동작 인식 알고리즘들을 사용한다. 제1 동작 인식 알고리즘은 동작 후보들을 찾기 위해 디지털 네트워크 카메라(100)상에서 실행되고, 그리고 제2 동작 인식 알고리즘은 동작 후보가 실제 동작인 것을 확인 또는 거부하기 위해 서버(200)상에서 실행된다. 제2 동작 인식 알고리즘은 제1 동작 인식 알고리즘보다 많은 프로세싱을 요구한다. 제1 동작 인식 알고리즘은 주로 문맥 및/또는 공간 동작 인식에 기초한다. 비-제한적인 예로서, 제1 동작 인식 알고리즘은 단지 문맥 및/또는 공간 동작 인식 알고리즘에 기초할 수 있다. 제2 동작 인식 알고리즘은 주로 시간 동작 인식에 기초한다. 하지만, 제2 동작 인식 알고리즘은 문맥 및/또는 공간 동작 인식의 요소들을 포함할 수 있다.

[0036] 동작 후보 인식기(126)는 객체 식별자(124)에 의해 식별된 객체 이미지 프레임들 중 적어도 하나에 제1 동작 인식 알고리즘을 적용하도록 구성된다. 제1 동작 인식 알고리즘을 적용함으로써, 동작 후보의 존재가 검출된다. 동작 후보 인식기(126)에 의해 수행되는 제1 동작 인식 알고리즘 분석은 동작 유형을 검출할 필요가 없다. 단지 일반 동작 후보들만 검색하면 된다. 하지만, 제1 동작 인식 알고리즘은 상이한 유형들의 동작들을 필터링하도록 구성될 수 있다. 따라서, 제1 동작 인식 알고리즘을 적용함으로써, 미리 결정된 유형의 동작의 동작 후보의 존재가 검출될 수 있다.

[0037] 제1 동작 인식 알고리즘을 적용함으로써, 동작 후보 인식기(126)는 동작을 나타낼 가능성 있는 비정상적인 포즈들을 검출하도록 구성될 수 있다. 그 결과, 동작을 위한 동작 후보자를 찾는다. 또한, 동작 후보 인식기(126)는 미리 정의된 유형의 동작에 대한 동작 후보를 찾기 위해 검출된 포즈들을 필터링하도록 구성될 수 있다. 제1 동작 인식 알고리즘은 상대적으로 가벼운 알고리즘이다. 거짓 긍정들이 어느 정도 허용된다. 동작 후보자의 검출은 제2 동작 인식 알고리즘을 사용하는 보다 진보된 동작 인식 분석을 동기로 삼는다. 따라서, 동작 후보 인식기(126)는, 본 명세서에서 언급되는 바와 같이, 가능성 있는 동작 제안들 또는 동작 후보들을 야기하도록 구성된다.

[0038] 동작 후보를 검출하는 것은, 동작 비디오 시퀀스에 제2 동작 인식 알고리즘을 적용함으로써 동작 후보가 실제

동작인지 여부를 결정하기 위해, 동작 후보를 포함하는 비디오 시퀀스의 일부인 동작 비디오 시퀀스를 서버(200)에 송신하는 것을 야기한다.

[0039] 비디오 추출기(127)는 동작 비디오 시퀀스의 동작 이미지 프레임들을 생성하도록 구성된다. 동작 비디오 시퀀스는 동작 후보를 포함하는 비디오 시퀀스의 일부이다. 동작 비디오 시퀀스는 비디오 시퀀스와 동일한 프레임 레이트를 가질 수 있다. 동작 비디오 시퀀스는 비디오 시퀀스의 프레임 레이트보다 작은 프레임 레이트를 가질 수 있다. 즉, 동작 비디오 시퀀스는 비디오 시퀀스보다 작은 초당 프레임들 수(fps)를 갖는 프레임 레이트를 갖는다. 예를 들어, 비디오 시퀀스의 프레임 레이트는 60fps일 수 있고, 동작 비디오 시퀀스의 프레임 레이트는 30fps일 수 있다.

[0040] 여기에서, 동작 이미지 프레임들이라고 불리는 동작 비디오 시퀀스의 이미지 프레임들은 비디오 시퀀스로부터의 복수의 이미지 프레임들에 관한 비디오 데이터를 추출함으로써 생성된다. 비디오 추출기(127)는, 비디오 데이터가 추출되는 하나 이상의 복수의 이미지 프레임들이 관심 객체를 포함하도록 구성된다. 따라서, 복수의 동작 이미지 프레임들 중 적어도 하나 이상은 객체 이미지 프레임(들)이다.

[0041] 또한, 비디오 추출기(127)는, 관심이 있는 객체가 검출된 비디오 시퀀스에서 식별된 시점 이전의 시점과 관련된 비디오 시퀀스의 제1 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하도록 구성될 수 있다. 그러므로, 비디오 시퀀스의 이미지 프레임들은 동작 비디오 시퀀스에서 나중에 사용하기 위해 메모리(122)에서 캐싱될 수 있다. 이는, 제1 동작 인식 알고리즘의 실행을 야기하는 관심 객체를 포함하는 객체 이미지 프레임 및 동작 비디오 시퀀스에 포함될 객체 이미지 프레임에 선행하는 이미지 프레임(들) 모두를 포함할 수 있게 한다. 이것은, 동작 비디오 시퀀스의 시간 윈도우가 관심 객체에 의해 수행되는 동작에 관한 모든 관련 정보를 포함할 가능성을 향상시킨다. 비-제한적인 예로서, 제1 미리결정된 수의 이미지 프레임들은, 관심 객체의 유형 또는 동작 후보의 동작의 유형 중 하나 이상에 따라 설정될 수 있다. 더욱이, 추가의 비-제한적인 예로서, 객체 및 가능하게는 그 주변의 모션 분석은 어떤 이전의 이미지 프레임들에서 동작 후보가 시작되었는지를 나타내기 위해 사용될 수 있다. 이를 통해, 비교적 큰 프리버퍼(prebuffer)로부터, 실제로 제2 동작 인식에 관련될 가능성이 있는 이미지 프레임들만을 추출할 수 있게 된다. 따라서, 제1 미리결정된 수의 이미지 프레임들이 동적으로 설정될 수 있다.

[0042] 또한, 비디오 추출기(127)는, 관심이 있는 객체가 검출된 비디오 시퀀스에서 식별된 시점 이후의 시점과 관련된 비디오 시퀀스의 제2 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하도록 구성될 수 있다. 이는, 제1 동작 인식 알고리즘의 실행을 야기하는 관심 객체를 포함하는 객체 이미지 프레임 및 동작 비디오 시퀀스에 포함될 객체 이미지 프레임에 후속하는 이미지 프레임(들) 모두를 포함하게 한다. 이것은, 동작 비디오 시퀀스의 시간 윈도우가 관심 객체에 의해 수행되는 동작에 관한 모든 관련 정보를 포함할 가능성을 향상시킨다. 제2 미리결정된 수의 이미지 프레임들은 관심 객체의 유형 또는 동작 후보의 동작 유형 중 하나 이상에 따라 설정될 수 있다. 더욱이, 추가의 비-제한적인 예로서, 객체 및 가능하게는 그 주변의 모션 분석은 어떤 이미지 프레임에서 동작 후보가 종료되었는지를 나타내기 위해 사용될 수 있다. 이를 통해, 실제로 제2 동작 인식에 관련될 가능성이 있는 이미지 프레임들만을 추출할 수 있게 된다. 따라서, 제2 미리결정된 수의 이미지 프레임들이 동적으로 설정될 수 있다.

[0043] 비디오 추출기(127)는, 또한, 관심 객체에 대해 동작 비디오 시퀀스 내의 좌표들에 관한 정보를 추출하도록 구성될 수 있다. 좌표들은, 관심 객체를 포함하는 그리고/또는 각각의 이미지 프레임에서 관심 객체가 위치되는 이미지 프레임 또는 이미지 프레임들을 참조할 수 있다. 이러한 좌표들은 동작 비디오 시퀀스와 함께 서버로 전송될 수 있다.

[0044] 비디오 추출기(127)는 또한 동작 이미지 프레임들을 생성할 때 복수의 이미지 프레임들의 비디오 데이터를 자르도록 구성될 수 있다. 관심 객체를 포함하는 동작 이미지 프레임들이 관심 객체의 적어도 일부를 포함하도록 복수의 이미지 프레임들의 비디오 데이터가 잘릴 수 있다. 더욱이, 관심 객체를 포함하는 동작 이미지 프레임들이 관심 객체를 적어도 부분적으로 둘러싸는 배경의 부분을 포함하도록 복수의 이미지 프레임들의 비디오 데이터가 잘릴 수 있다. 관심 객체를 적어도 부분적으로 둘러싸는 배경을 통합함으로써, 동작 비디오 시퀀스의 시간 분석뿐만 아니라 문맥 분석 및/또는 공간 분석이 수행될 수 있다.

[0045] 인코딩 유닛(128)은 비디오 인코딩을 이용하여 비디오 시퀀스의 디지털 비디오 데이터를 인코딩하도록 구성된다. 비디오 인코딩의 비-제한적인 예들은 ISO/MPEG 또는 ITU-H.26X 계열의 비디오 인코딩 표준들이다. 인코딩 유닛(128)은 이후에 인코딩된 디지털 비디오 데이터로 지정되는, 디지털 비디오 데이터의 이미지를 인코딩하도록 구성된다. 인코딩된 디지털 비디오 데이터는 네트워크 인터페이스(130)를 거쳐 디지털 네트워크(300)

를 통해 직접 전송될 수 있다. 대안으로, 인코딩된 디지털 비디오 데이터는, 디지털 네트워크(300)를 통해 네트워크 인터페이스(130)를 거쳐 추후 전송을 위해 메모리(122)에 저장될 수 있다. 인코딩 유닛(128)은, 동작 비디오 시퀀스가 서버(200)로 전송되기 전에 동작 비디오 시퀀스를 인코딩하도록 구성될 수 있다.

[0046] 네트워크 인터페이스(130)는 동작 비디오 시퀀스를 서버(200)로 전송하도록 구성된다.

[0047] 서버(200)는 동작 검증기(210)를 포함하는 회로(201)를 포함한다. 회로(201)는 네트워크 인터페이스(202), 디코딩 유닛(204), 중앙 처리 유닛(CPU)(206) 및 디지털 데이터 저장 매체(메모리)(208) 중 하나 이상을 더 포함할 수 있다. 디코딩 유닛(204) 및/또는 동작 검증기(210) 중 임의의 하나는 전용 하드웨어 회로 및/또는 소프트웨어 모듈로서 구현될 수 있다. 소프트웨어로 구현되는 경우, 소프트웨어는 CPU(206) 상에서 실행될 수 있다. CPU(206)는 디지털 데이터 프로세싱을 수행하기 위한 임의의 적합한 CPU일 수 있다. 임의의 전용 하드웨어 회로는 부분적으로 전용 프로세서 또는 CPU(206) 상에서 실행되는 소프트웨어 부분들을 포함할 수 있음에 유의해야 한다.

[0048] 서버(200)는 네트워크 인터페이스(202)를 통해 디지털 네트워크(300)에 접속되도록 구성된다. 디지털 네트워크에 대한 접속은 유선 또는 무선일 수 있다. 따라서, 네트워크 인터페이스(202)는 이더넷 포트와 같은 10/100/1000 Mbps 데이터 트래픽에 적합한 네트워크 포트일 수 있고, 예를 들어, RJ45 커넥터와 같은 모듈러 커넥터를 수신하도록 구성된 모듈 포트일 수 있다. 일반적으로, 이러한 RJ45 커넥터 포트는 (예를 들어, cat 5, cat 5e 또는 cat 6의) 트위스트 페어 케이블과 같은 네트워크 케이블을 수신하도록 구성된다. 대안으로, 네트워크 포트의 I/O 수단은 모바일 인터넷 통신 표준(예를 들어, 1G, 2G, 2.5G, 2.75G, 3G, 3.5G, 3.75G, 3.9G, 4G, 5G)을 사용하는 또는 WiFi를 사용하는 무선 I/O 수단일 수 있다.

[0049] 동작 비디오 시퀀스가 인코딩되는 경우, 디코딩 유닛(204)은 인코딩된 비디오 시퀀스를 디코딩하도록 구성된다. 따라서, 디코딩 유닛(204)은 비디오 디코딩을 사용하여 비디오 시퀀스의 디지털 비디오 데이터를 디코딩하도록 구성된다.

[0050] 메모리(122)는 임의의 종류의 휘발성 또는 비-휘발성 메모리 일 수 있다. 또한, 메모리(122)는 복수의 메모리 유닛들을 포함할 수 있다. 복수의 메모리 유닛들 중 적어도 하나는, 예를 들어, 동작 비디오 시퀀스를 프로세싱하는 동안 데이터를 버퍼링하기 위한 버퍼 메모리로서 사용될 수 있다. 메모리(122)는 동작 비디오 시퀀스의 전체 또는 일부를 더 저장할 수 있다.

[0051] 동작 검증기(210)는 제2 동작 인식 알고리즘을 동작 비디오 시퀀스에 적용하도록 구성된다. 따라서, 동작 후보가 실제 행동인 것으로 확인되거나 거절될 수 있다. 특히, 동작 후보가 미리 정의된 행동 유형의 행동임을 확인하거나 거절할 수 있다. 동작 검증기(210)에 의해 수행되는 동작 인식은 반드시 실시간으로 수행될 필요는 없다. 이것은 행동이 끊임없이 지속되는 것이 아니라 짧은 시간의 사건이기 때문에, 중요한 것은 특정 유형의 행동에 대한 알람이 발생되는지를 결정하는 것이다.

[0052] 도 2를 참조하면, 카메라(100)에 의해 캡처된 비디오 시퀀스에서의 동작 인식 방법이 도시된다. 상기 방법은, 카메라(100)의 회로(102)에 의해,

[0053] 비디오 시퀀스의 객체 이미지 프레임에서 관심 객체를 식별하는 단계(S501)와;

[0054] 제1 동작 인식 알고리즘을 객체 이미지 프레임에 적용(S502)하여 동작 후보의 존재를 검출하는 단계와;

[0055] 비디오 시퀀스로부터 복수의 이미지 프레임들에 관한 비디오 데이터를 추출함으로써 동작 비디오 시퀀스의 동작 이미지 프레임들을 생성하는 단계(S504) - 상기 비디오 데이터가 추출되는 상기 복수의 이미지 프레임들 중 하나 이상은 상기 관심 객체를 포함하며 - 와; 그리고

[0056] 상기 동작 비디오 시퀀스를 서버(200)로 전송하는 단계(S506)를 포함한다.

[0057] 상기 방법은, 서버(200)의 회로(201)에 의해, 제2 동작 인식 알고리즘을 동작 비디오 시퀀스에 적용(S508)하여, 동작 후보가 미리 정의된 유형의 동작임을 확인 또는 거부하는 단계를 더 포함한다.

[0058] 동작 이미지 프레임들을 생성하는 단계(S504)는, 관심 객체를 포함하는 동작 이미지 프레임들이 관심 객체의 적어도 일부를 포함하도록 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑(cropping)하는 단계를 포함할 수 있다.

[0059] 동작 이미지 프레임들을 생성하는 단계(S504)는, 관심 객체를 포함하는 동작 이미지 프레임들이 관심 객체를 적어도 부분적으로 둘러싸는 배경 부분을 포함하도록 비디오 시퀀스의 복수의 이미지 프레임들을 크로핑하는 단계

를 포함할 수 있다.

[0060] 동작 비디오 시퀀스를 전송하는 단계(S506)는 동작 비디오 시퀀스 내의 좌표들을 관심 객체로 전송하는 단계를 포함할 수 있다. 좌표들은, 어떤 이미지 프레임 또는 이미지 프레임들이 관심 객체를 포함하는지 그리고/또는 각각의 이미지 프레임에서 관심 객체가 어디에 위치하는지를 나타낼 수 있다.

[0061] 상기 방법은, 카메라(100)의 회로(102)에 의해, 비디오 시퀀스에서 관심 객체를 검출하는 단계(S500)를 더 포함할 수 있다. 동작 이미지 프레임들을 생성하는 단계(S504)는, 관심 객체의 검출 이전의 시점과 관련된 비디오 시퀀스의 제1 미리결정된 수의 이미지 프레임들에 관한 비디오 데이터를 추출하는 단계를 포함할 수 있다. 동작 이미지 프레임들을 생성하는 단계(S504)는, 관심 객체의 검출 이후의 시점에 관련된 비디오 시퀀스의 제2 미리 결정된 수의 이미지 프레임들에 관련된 비디오 데이터를 추출하는 단계를 포함할 수 있다.

[0062] 통상의 기술자는, 본 발명이 결코 전술한 바람직한 실시예들에 제한되지 않는다는 것을 알 수 있다. 반대로, 첨부된 청구항의 범위 내에서 많은 수정들 및 변형들이 가능하다.

[0063] 예를 들어, 동작 후보가 실제 동작인지를 검증한 후에, 서버(200)는 알람 트리거를 전송하도록 구성될 수 있다. 알람 트리거는, 추가적인 프로세싱을 위해 디지털 네트워크 카메라(100)로 전송될 수 있다. 예를 들어, 디지털 네트워크 카메라(100)는 알람 트리거를 수신한 후 카메라 설정을 변경하도록 구성될 수 있다. 변경될 수 있는 카메라 설정들의 비-제한적인 예들은, 프레임 레이트, 해상도, 광 감도, HDR로 전환, 표준 메시지로 카메라에 연결된 스피커 트리거, 관심 객체의 PTZ 추적 시작, 탐지된 객체의 개선된 외양의 모델의 트리거를 생성하여 여러 카메라들에서 추적하거나, 레이더 추적을 시작하거나, 열 모드로 전환하거나, 추가 동작 인식을 위한 임계값을 변경하거나 관련 유형의 후속 작업을 확인하는 것이다(예를 들어, 사람이 넘어지면 "다시 일어나십시오 (rise up again)"라는 행동을 검색하고 일정한 시간 내에 일어나지 않으면 알람을 울림).

[0064] 대안으로 또는 조합하여, 알람 트리거는 비디오 관리 센터(VMS)로 전송될 수 있다. 알람 트리거는, 미리 정의된 유형의 동작이 발생했다는 통지들을 전송하기 위해 VMS에서 사용될 수 있다.

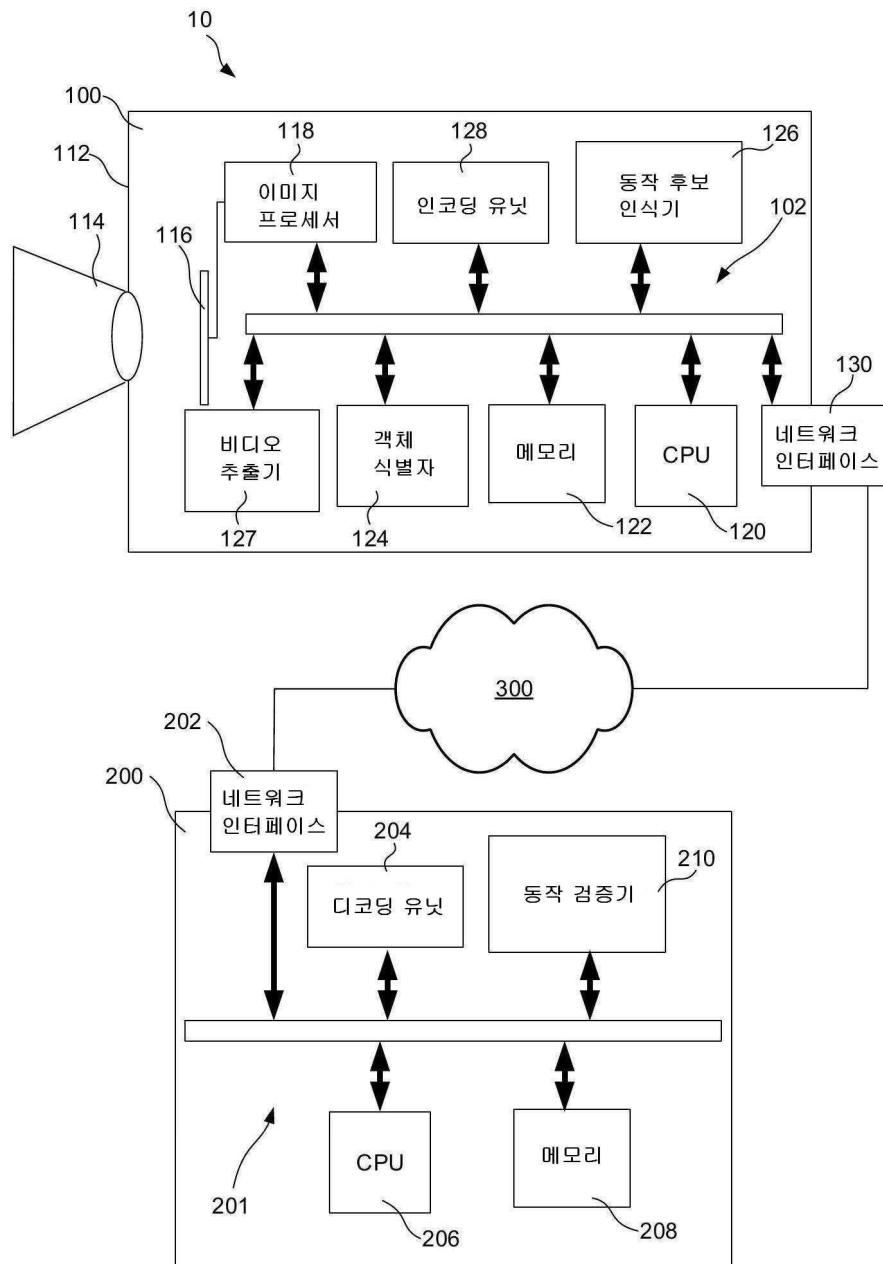
[0065] 또한, 서버(200)는 다양한 유형들의 디바이스들로 구현될 수 있다. 서버(200)로서 구현되는 디바이스들의 비-제한적인 예들은, 전용 컴퓨터, 다른 카메라 디바이스, 비디오 관리 시스템, 클라우드 서버, 카메라 근처의 분석 박스, 액세스 제어 유닛, 컴퓨팅 성능을 갖는 IoT 장치이다. 또한, 서버 기능은 다른 디바이스들에 분산될 수 있다. 특히, 동작 검증기(210)의 경우에, 적어도 부분적으로는 복수의 프로세서들 상에서 실행되는 소프트웨어 코드 부분으로서 구현된다.

[0066] 더욱이, 관심 객체는 비디오 시퀀스의 복수의 객체 이미지 프레임들에서 식별될 수 있다. 제1 동작 인식 알고리즘은 복수의 객체 이미지 프레임들 각각에 개별적으로 적용될 수 있다. 이후, 제1 동작 인식 알고리즘의 개별적인 애플리케이션의 결과는 동작 후보를 찾는 데 사용될 수 있다. 예를 들어, 복수의 이미지들의 개별 분석은 사람의 다리들이 항상 상이한 각도들에 있음을 개시할 수 있다. 이는 사람이 걷거나 달리는 것을 나타낼 수 있다. 이후, 동작 비디오 시퀀스의 동작 이미지 프레임들은 비디오 시퀀스로부터의 복수의 이미지 프레임들에 관한 비디오 데이터를 추출함으로써 생성되며, 비디오 데이터가 추출되는 하나 이상의 복수의 이미지 프레임들은 관심 객체를 포함한다.

[0067] 추가적으로, 개시된 실시예들에 대한 변형들은, 도면들, 개시된 내용 및 첨부된 청구 범위의 연구로부터 청구된 발명을 실시하는 당업자에 의해 이해되고 영향을 받을 수 있다.

## 도면

## 도면1



## 도면2

