



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0054702
(43) 공개일자 2019년05월22일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 3/04 (2006.01)
G06T 7/20 (2017.01)

(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/0454 (2013.01)

(21) 출원번호 10-2017-0151575
(22) 출원일자 2017년11월14일
심사청구일자 2017년11월14일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자
이성환
서울특별시 강남구 언주로30길 13, B동 2506호 (도곡동, 대림아크로빌)

조남규
서울특별시 관악구 호암로 399, 309동 304호 (신림동, 삼성산주공아파트)

윤다혜
인천광역시 부평구 마장로 287, 7동 201호 (산곡동, 산곡동한양아파트)

(74) 대리인
특허법인엠에이피에스

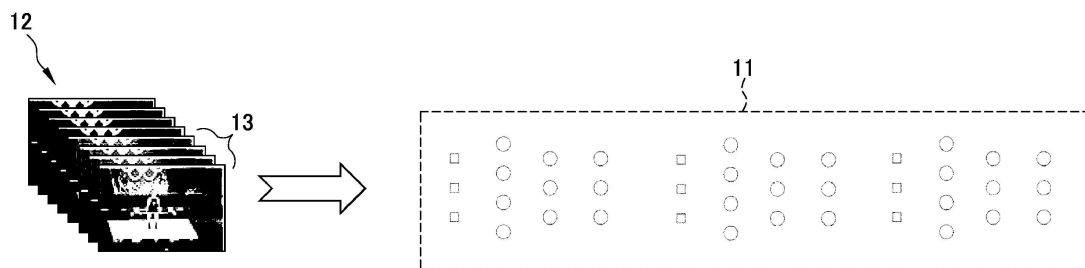
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 영상에서 객체의 행동을 인식하는 방법 및 그 장치

(57) 요약

본 발명은 영상으로부터 영상 내 객체의 행동을 인식하는 방법 및 그 장치를 제공한다. 보다 구체적으로, 본 방법은 영상 세그먼트를 획득하는 단계, 제1 뉴럴 네트워크를 통해 영상 세그먼트 내에 객체의 행동이 존재하는지 여부를 판별하는 단계, 영상 세그먼트 내에 객체의 행동이 존재하면, 제 2 뉴럴 네트워크를 통해 영상 세그먼트에 연속하는 하나 이상의 미래영상 프레임을 생성하여 통합 영상 세그먼트를 구성하는 단계, 및 제3 뉴럴 네트워크를 통해 상기 통합 영상 세그먼트 내의 객체의 행동 유형 및 행동 시점을 검출하는 단계를 포함한다. 이때, 제1 뉴럴 네트워크는 독립적으로 학습되며, 제2 뉴럴 네트워크와 제3 뉴럴 네트워크는 의존적으로 학습된다.

대표도



(52) CPC특허분류

G06T 7/20 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711054902

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 미래성장동력

연구과제명 Deep Learning 기반의 보행자 안전 지원 스마트카 비전 기술의 개발

기 여 율 1/1

주관기관 고려대학교 산학협력단

연구기간 2017.01.01 ~ 2017.12.31

명세서

청구범위

청구항 1

행동 인식 장치가 영상 내의 객체의 행동을 인식하는 방법에 있어서,

영상 세그먼트(segment)를 획득하는 단계;

제1 뉴럴 네트워크(neural network)를 통해 상기 영상 세그먼트 내에 객체의 행동이 존재하는지 여부를 판별하는 단계;

상기 영상 세그먼트 내에 객체의 행동이 존재하면, 제2 뉴럴 네트워크를 통해 상기 영상 세그먼트에 연속하는 하나 이상의 미래영상 프레임을 생성하여 통합 영상 세그먼트를 구성하는 단계; 및

제3 뉴럴 네트워크를 통해 상기 통합 영상 세그먼트 내의 상기 객체의 행동 유형 및 행동 시점을 검출하는 단계를 포함하되,

상기 제1 뉴럴 네트워크는 독립적으로 학습되며, 상기 제2 뉴럴 네트워크와 상기 제3 뉴럴 네트워크는 의존적으로 학습되는 것인 행동 인식 방법.

청구항 2

제 1 항에 있어서,

상기 제2 뉴럴 네트워크는 상기 제3 뉴럴 네트워크의 손실함수의 값을 기초로 학습되며, 상기 제3 뉴럴 네트워크는 상기 제2 뉴럴 네트워크의 결과값을 기초로 학습되는 것인 행동 인식 방법.

청구항 3

제 1 항에 있어서,

상기 객체의 행동이 존재하는지 여부를 판별하는 단계는

상기 제1 뉴럴 네트워크에 포함된 복수의 레이어 중 마지막 레이어에서 추출된 속성 정보를 조합하여 특징맵을 획득하는 단계; 및

상기 제1 뉴럴 네트워크에 포함된 이진분류기에 상기 특징맵을 입력 데이터로 적용하여 상기 영상 세그먼트 내에 상기 객체의 행동이 존재하는지 여부를 판별하는 단계;를 포함하는 것인 행동 인식 방법.

청구항 4

제 3 항에 있어서,

상기 속성 정보는 상기 영상 세그먼트에 포함된 폴리곤(polygon), 에지(edge), 깊이(depth), 선명도(sharpness), 채도, 명도, 깊이 및 이들의 시공간적 변화값 중 적어도 하나를 포함하는 것인 행동 인식 방법.

청구항 5

제 1 항에 있어서,

상기 통합 영상 세그먼트를 구성하는 단계는

상기 제2 뉴럴 네트워크에 포함된 복수의 제1 레이어 중에서 마지막 레이어에서 추출된 속성 정보를 조합하여 특징맵을 획득하는 단계;

상기 제2 뉴럴 네트워크에 포함된 복수의 제2 레이어를 통해, 상기 특징맵을 기초로 상기 영상 세그먼트의 속성 정보가 변형되거나, 기 저장된 다른 이미지들의 특징 정보가 합성된 하나 이상의 미래영상 프레임을 생성하는 단계; 및

상기 영상 세그먼트에 상기 하나 이상의 미래영상 프레임을 더하여 상기 통합 영상 세그먼트를 구성하는 단계를

포함하는 것인 행동 인식 방법.

청구항 6

제 5 항에 있어서,

상기 통합 영상 세그먼트를 구성하는 단계는

상기 제2 뉴럴 네트워크에 포함된 복수의 제3 레이어를 통해, 상기 생성된 하나 이상의 미래영상 프레임이 상기 영상 세그먼트에 실질적으로 연속할 확률을 판별하는 단계; 및

상기 판별된 확률값을 상기 복수의 제1 레이어 및 상기 제2 레이어로 피드백하여, 상기 하나 이상의 미래영상 프레임을 재생성하는 단계를 포함하는 것인 행동 인식 방법.

청구항 7

제 1 항에 있어서,

상기 행동 유형 및 행동 시점을 판별하는 단계는

상기 제3 뉴럴 네트워크에 포함된 복수의 레이어 중 마지막 레이어에서 추출된 속성 정보를 조합하여 특징맵을 획득하는 단계; 및

상기 특징맵을 제1 분류기에 입력 데이터로 적용하여 상기 통합 영상 세그먼트 내의 객체의 행동 유형을 식별하고, 상기 제1 분류기의 결과값을 제2 분류기의 입력 데이터로 적용하여 상기 통합 영상 세그먼트 내의 객체의 행동이 발현되는 시점이 상기 행동 유형의 끝 또는 시작 시점에 대응되는지 여부를 검출하는 단계를 포함하는 것인 행동 인식 방법.

청구항 8

영상에 포함된 객체의 행동을 인식하는 프로그램이 저장된 메모리; 및

상기 프로그램을 실행하는 프로세서를 포함하고,

상기 프로세서는, 상기 프로그램이 실행됨에 따라,

영상 세그먼트를 획득하고, 제1 뉴럴 네트워크를 통해 상기 영상 세그먼트 내에 객체의 행동이 존재하는지 여부를 판별하며,

상기 영상 세그먼트 내에 객체의 행동이 존재하면, 제2 뉴럴 네트워크를 통해 상기 영상 세그먼트에 연속하는 하나 이상의 미래영상 프레임을 생성하여 통합 영상 세그먼트를 구성하고, 제3 뉴럴 네트워크를 통해 상기 통합 영상 세그먼트 내의 상기 객체의 행동 유형 및 행동 시점을 검출하되,

상기 제1 뉴럴 네트워크는 독립적으로 학습되며, 상기 제2 뉴럴 네트워크와 상기 제3 뉴럴 네트워크는 의존적으로 학습되는 것인 행동 인식 장치.

청구항 9

제 8 항에 있어서,

상기 제2 뉴럴 네트워크는 상기 제3 뉴럴 네트워크의 손실함수의 값을 기초로 학습되며, 상기 제3 뉴럴 네트워크는 상기 제2 뉴럴 네트워크의 결과값을 기초로 학습되는 것인 행동 인식 장치.

청구항 10

제 8 항에 있어서,

상기 프로세서는

상기 제1 뉴럴 네트워크에 포함된 복수의 레이어 중 마지막 레이어에서 추출된 속성 정보를 조합하여 특징맵을 획득하고, 상기 제1 뉴럴 네트워크에 포함된 이진분류기에 상기 특징맵을 입력 데이터로 적용하여 상기 영상 세그먼트 내에 상기 객체의 행동이 존재하는지 여부를 판별하는 것인 행동 인식 장치.

청구항 11

제 8 항에 있어서,

상기 프로세서는

상기 제2 뉴럴 네트워크에 포함된 복수의 제1 레이어 중에서 마지막 레이어에서 추출된 속성 정보를 조합하여 특징맵을 획득하고, 상기 제2 뉴럴 네트워크에 포함된 복수의 제2 레이어를 통해 상기 특징맵을 기초로 상기 영상 세그먼트의 속성 정보가 변형되거나 기 저장된 다른 이미지들의 특징 정보가 합성된 하나 이상의 미래영상 프레임을 생성하며,

상기 영상 세그먼트에 상기 하나 이상의 미래영상 프레임을 더하여 상기 통합 영상 세그먼트를 구성하는 것인 행동 인식 장치.

청구항 12

제 8 항에 있어서,

상기 프로세서는

상기 제3 뉴럴 네트워크에 포함된 복수의 레이어 중 마지막 레이어에서 추출된 속성 정보를 조합하여 특징맵을 획득하고, 상기 특징맵을 제1 분류기에 입력 데이터로 적용하여 상기 통합 영상 세그먼트 내의 객체의 행동 유형을 식별하고, 상기 제1 분류기의 결과값을 제2 분류기의 입력 데이터로 적용하여 상기 통합 영상 세그먼트 내의 객체의 행동이 발현되는 시점이 상기 행동 유형의 끝 또는 시작 시점에 대응되는지 여부를 검출하는 것인 행동 인식 장치.

청구항 13

제 1 항 내지 제 7 항 중 어느 한 항에 기재된 방법을 컴퓨터 상에서 수행하기 위한 프로그램을 기록한 컴퓨터 판독 가능한 기록 매체.

발명의 설명

기술 분야

[0001] 본 발명은 영상으로부터 영상에 포함된 객체의 행동을 인식하는 방법 및 그 장치에 관한 것이다.

배경 기술

[0002] 컴퓨터 비전 분야에서의 행동인식 기술은 이미지 센서로부터 색상(RGB) 또는 깊이(depth) 영상을 입력 받아 얻어진 영상 정보에 의존하여 영상 내의 객체의 행동을 분류한다. 이미지 센서는 다양한 장소에 용이하게 설치 운용될 수 있기 때문에, 최근 이미지 센서를 활용하여 위험 상황 감시, 특정 이벤트 탐지 등이 이루어지고 있다. 그러나 행동인식 기술을 여러 응용 분야에 활용하기 위해서는 이 기술을 사용하고자 하는 상황에 따라 영상의 어떤 피쳐(feature)를 사용할지 결정하는 것이 중요하다. 영상에서 사용될 수 있는 피쳐의 범주는 크게 두 가지로, 연구자가 직접 설계한 핸드크래프트(hand-crafted) 피쳐와 딥러닝을 통해 추출되는 딥(deep) 피쳐로 나누어질 수 있다. 영상에서 많이 이용 되는 핸드크래프트 피쳐는 사람의 궤적정보를 표현하는 Dense Trajectories 기술자(descriptor)와 외형정보를 나타내는 Histogram of Oriented Gradients (HOG) 기술자를 포함하며, 딥 피쳐에는 시간 정보를 학습하는 Recurrent Neural Network(RNN)와 외형정보를 학습하는 Convolutional Neural Network(CNN)가 주로 사용된다. 최근에는 딥러닝 기술의 비약적 발전으로 인해, 딥러닝 기반의 기술자가 대표적으로 사용되고 있다.

[0003] 이와 같이, 행동인식 기술은 딥러닝 기술의 발전으로 인한 동반 발전을 보이며 높은 성능을 보이고 있다. 그러나 종래의 행동인식 기술은 특정 행동이 발생된 영상을 학습데이터로 사용하므로, 다양한 행동이 연이어 일어나거나 행동이 일어나지 않는 프레임들을 포함하는 실생활 영상으로부터 행동인식을 수행하는 데에는 여전히 한계가 있다. 즉, 이러한 연속적인 행동을 인식하기 위해서는 인식기술을 적용하기 전에 연속적인 행동을 하나의 행동만이 포함되도록 클립단위로 자르는 전처리 과정이 요구된다. 따라서 이러한 전처리 과정 없이 연속적인 행동을 인식하기 위해서는 영상에서 행동이 발생하는 부분을 검출하는 동시에 어떤 행동이 발생되었는지 인식하는 기술이 요구된다.

선행기술문헌

[0004] 미국등록특허 제 9,648,035 호 (발명의 명칭: User behavioral risk assessment)

발명의 내용

해결하려는 과제

[0005] 본 발명은 전술한 종래 기술의 문제점을 해결하기 위한 것으로, 실시간 영상에서 행동인식을 위해서 행동 유형과 함께 행동이 발생하는 구간의 시작점과 끝점을 학습함으로써, 보다 정확하게 행동을 검출하는 방법 및 시스템을 제시하고자 한다.

과제의 해결 수단

[0006] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제1 측면에 따른 객체 인식 장치가 영상 내의 객체의 행동을 인식하는 방법은 영상 세그먼트를 획득하는 단계; 제1 뉴럴 네트워크(neural network)를 통해 영상 세그먼트 내에 객체의 행동이 존재하는지 여부를 판별하는 단계; 영상 세그먼트 내에 객체의 행동이 존재하면, 제2 뉴럴 네트워크를 통해 영상 세그먼트에 연속하는 하나 이상의 미래영상 프레임을 생성하여 통합 영상 세그먼트를 구성하는 단계; 및 제3 뉴럴 네트워크를 통해 통합 영상 세그먼트 내의 객체의 행동 유형 및 행동 시점을 검출하는 단계를 포함한다. 이때, 제1 뉴럴 네트워크는 독립적으로 학습되며, 제2 뉴럴 네트워크와 제3 뉴럴 네트워크는 의존적으로 학습된다.

[0007] 또한, 본 발명의 제2 측면에 따른 행동 인식 장치는, 영상에 포함된 객체의 행동을 인식하는 프로그램이 저장된 메모리; 및 상기 프로그램을 실행하는 프로세서를 포함한다. 이때, 프로세서는, 상기 프로그램이 실행됨에 따라, 영상 세그먼트를 획득하고, 제1 뉴럴 네트워크를 통해 영상 세그먼트 내에 객체의 행동이 존재하는지 여부를 판별하며, 영상 세그먼트 내에 객체의 행동이 존재하면, 제2 뉴럴 네트워크를 통해 영상 세그먼트에 연속하는 하나 이상의 미래영상 프레임을 생성하여 통합 영상 세그먼트를 구성하고, 제3 뉴럴 네트워크를 통해 통합 영상 세그먼트 내의 객체의 행동 유형 및 행동 시점을 검출한다. 이때, 제1 뉴럴 네트워크는 독립적으로 학습되며, 제2 뉴럴 네트워크와 제3 뉴럴 네트워크는 의존적으로 학습된다.

[0008] 또한, 본 발명의 제3 측면은, 상기 제1 측면의 방법을 컴퓨터 상에서 수행하기 위한 프로그램을 기록한 컴퓨터 판독 가능한 기록 매체를 제공한다.

발명의 효과

[0009] 전술한 과제 해결 수단에 따르면, 본 발명의 일 실시예는 미래영상 프레임들을 생성함으로써 실시간으로 영상이 입력되는 환경에서 위험 상황 감지 또는 특정 상황 인지를 수행할 수 있다. 또한, 본 발명의 일 실시예는 영상 내에 행동이 일어나는 구간과 그렇지 않은 구간을 분류하여 행동이 일어나는 구간을 검출하고, 나아가 영상 내의 행동에 대한 행동 시점(즉, 행동의 끝 또는 시작 시점 등)을 검출함으로써, 시간에 대한 사전적 학습 없이 시각적 정보에만 의존하여 행동을 반복적으로 검출함으로써, 지속적 관찰이 요구되는 로봇 제어 등에 효율적으로 적용될 수 있다.

도면의 간단한 설명

[0010] 도 1은 본 발명의 일 실시예에 따른 행동 인식 장치를 도시한다.

도 2는 본 발명의 일 실시예에 따른 뉴럴 네트워크를 도시한 일례이다.

도 3은 본 발명의 일 실시예에 따른 행동 인식 장치의 구성을 도시한다.

도 4는 본 발명의 일 실시예에 따른 탐지 네트워크의 일례를 도시한다.

도 5는 본 발명의 일 실시예에 따라 도 3의 프로세서가 영상 내의 객체의 행동을 인식하는 방법을 도시한 순서도이다.

도 6은 본 발명의 일 실시예에 따른 프로세서를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0011] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0012] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0013] 이하, 도면을 참조하여, 본 발명의 일 실시예에 대하여 구체적으로 설명하도록 한다.
- [0014] 도 1은 본 발명의 일 실시예에 따른 행동 인식 장치(10)를 도시한다.
- [0015] 도 1에 도시된 바와 같이, 행동 인식 장치(10)는 뉴럴 네트워크(neural network)(11)를 이용하여 실시간 영상(12)으로부터 객체의 행동을 인식한다. 뉴럴 네트워크(11)는, 통계학적 기계 학습의 결과를 이용하여, 실시간 영상으로부터 다양한 속성 정보들을 추출하고, 추출된 속성 정보들을 기초로 실시간 영상 내 객체의 행동을 식별하는 알고리즘 집합일 수 있다. 또한, 뉴럴 네트워크(11)는 전술한 알고리즘 집합을 실행하기 위한 소프트웨어 또는 엔진(engine) 등으로 구현될 수 있다. 소프트웨어 또는 엔진 등으로 구현된 뉴럴 네트워크는 행동 인식 장치(10) 내의 프로세서에 의해 실행될 수 있다.
- [0016] 한편, 행동 인식 장치(10)는 카메라, CCTV(closed circuit television), 블랙박스(black-box) 등과 같이 영상 장치일 수 있으나, 이에 한정되지 않으며, 영상 장치를 포함하거나 영상 장치와 통신하여 영상을 제공받을 수 있는 컴퓨팅 기기일 수 있으며, 비한정적인 예로서, 스마트폰, 태블릿 PC, PC, 스마트 TV, 휴대폰, PDA(personal digital assistant), 랩톱, 미디어 플레이어, 마이크로 서버, IoT 허브, IoT 서버, 네비게이션, 키오스크, 가전기기 등일 수 있다.
- [0017] 도 2는 본 발명의 일 실시예에 따른 뉴럴 네트워크(11)를 도시한 일례이다.
- [0018] 도 2에 도시된 바와 같이, 본 발명의 일 실시예에 따른 뉴럴 네트워크(11)는 프로포절 네트워크(proposal network)(210), 미래영상 생성자 네트워크(future frame generation network)(220) 및 탐지 네트워크(detection network)(230)를 포함한다.
- [0019] 먼저, 프로포절 네트워크(210)는 영상 세그먼트(13)를 입력받아, 영상 세그먼트 내에 포함된 다양한 속성들을 추상화함으로써, 상기 영상 세그먼트(13)가 객체의 행동을 포함하는지 여부(즉, 행동 존부)를 판별할 수 있다. 여기서, 영상 세그먼트(13)는 연속되는 복수의 영상 프레임들의 집합으로서 실시간으로 획득되는 것일 수 있다. 그리고 영상 세그먼트 내 속성들을 추상화한다는 것은, 영상 세그먼트 내 객체 외형 정보, 객체 움직임 정보 등과 같은 속성 정보들을 검출하고, 검출된 속성 정보들 중에서 영상 세그먼트를 대표할 수 있는 핵심 속성을 판단하는 것일 수 있다.
- [0020] 이때, 프로포절 네트워크(210)는 복수의 추상화 레이어(211) 및 이진 분류기(binary classifier)(212)를 포함할 수 있으며, 행동 인식 장치(10)는 복수의 추상화 레이어(211)를 기초로 영상 세그먼트(13)로부터 속성 정보를 추출할 수 있다. 여기서, 속성 정보는, 비한정적인 예로서, 폴리곤(polygon), 에지(edge), 깊이(depth), 선명도(sharpness), 채도, 명도, 깊이 및 이들의 시공간적 변화값 등을 포함할 수 있다. 행동 인식 장치(10)는 마지막 추상화 레이어에서 추출된 속성 정보를 기초로 특징맵을 획득한다. 여기에서, 특징맵은 추출된 속성 정보의 조합으로서, 영상 세그먼트의 속성을 대표하는 적어도 하나의 속성 벡터를 포함할 수 있다.
- [0021] 특징맵은 이진분류기(212)의 입력 데이터로 적용될 수 있다. 이진분류기(212)는 특징맵을 기초로 영상 세그먼트에 객체의 행동이 존재하는지 여부를 판별할 수 있다. 행동 인식 장치(10)는 이진분류기(212)에 특징맵을 입력한 결과로서 객체 행동 존부(즉, 행동 존재 또는 비존재)를 획득할 수 있다.
- [0022] 미래 영상 생성자 네트워크(220)는 상기한 프로포절 네트워크(210)의 결과값이 영상 세그먼트(13) 내에 객체의 행동이 존재하는 것임을 나타내는 경우에 동작한다. 미래 영상 생성자 네트워크(220)는 영상 세그먼트에 연속하는 하나 이상의 영상 프레임을 생성한다.
- [0023] 구체적으로, 미래 영상 생성자 네트워크(220)는 영상 세그먼트(13)로부터 다양한 속성 정보를 추출하는 복수의 추상화 레이어(221)와 추출된 속성 정보를 기초로 영상 세그먼트에 연속하는 하나 이상의 미래 영상 프레임을 생성하는 복수의 생성자 레이어(222)를 포함한다. 이때, 영상 세그먼트의 속성 정보는, 비한정적인 예로서, 폴

리전, 예지, 깊이, 선명도, 채도, 명도, 깊이 및 이들의 시공간적 변화값 등을 포함할 수 있다.

- [0024] 행동 인식 장치(10)는 복수의 추상화 레이어(221) 중 마지막 추상화 레이어에서 추출된 속성 정보를 조합한 특징맵을 획득하고, 이를 첫번째 생성자 레이어의 입력 데이터로 적용할 수 있다. 복수의 생성자 레이어(222)는 기 학습된 정보를 기초로 속성 정보를 변형하거나, 기 저장된 다른 이미지들의 특징 정보를 합성하여 미래 영상 프레임을 생성한다. 이때, 기 학습된 정보는 객체의 행동에 기반한 각 속성 정보의 시공간적 변화값으로서, 각 생성자 레이어의 파라미터로 표현된다. 또한, 다른 이미지들의 특징 정보는, 비한정적인 예로서, 폴리곤, 예지, 깊이, 선명도, 채도, 명도, 깊이 및 이들의 시공간적 변화값 등을 포함할 수 있다.
- [0025] 또한, 구현예에 따라, 미래영상 생성자 네트워크(220)는 생성된 미래영상 프레임이 영상 세그먼트에 실질적으로 연속할 확률을 판별하고, 이를 추상화 레이어(221) 및 생성자 레이어(222)로 피드백하는 복수의 판별자 레이어(도시되지 않음)를 더 포함할 수 있다. 이러한, 미래 영상 생성자 네트워크는 예시적으로, 라플라시안 생산적 대 네트워크(Laplacian Generative Adversarial Network)일 수 있으나, 이에 한정되는 것은 아니다.
- [0026] 이어서, 행동 인식 장치(10)는 미래 영상 생성자 네트워크(220)를 통해 획득된 하나 이상의 미래 영상 프레임을 영상 세그먼트(13)에 더하여 확장 영상 세그먼트를 구성한다.
- [0027] 다음으로, 탐지 네트워크(230)는 확장 영상 세그먼트를 입력받아, 확장 영상 세그먼트에 포함된 다양한 속성들을 추상화함으로써, 확장 영상 세그먼트에 포함된 행동 유형 및 행동 시점을 검출할 수 있다. 여기서, 행동 유형은 상기 탐지 네트워크(230)의 학습 데이터에 의해 결정될 수 있으며, 비한정적인 예로서, 사람의 운동, 공격, 사고, 부상 등을 포함할 수 있으며, 기계의 고장, 사고 등을 포함할 수 있다. 또한, 행동 시점은 확장 영상 세그먼트에 포함된 객체의 행동이 해당 행동의 시작과 끝 시점 내의 어느 시점을 나타내는지를 포함한다.
- [0028] 탐지 네트워크(230)는 복수의 추상화 레이어(231)와 행동 유형 및 행동 시점을 판별하는 하나 이상의 분류기(232)를 포함한다. 행동 인식 장치(10)는 복수의 추상화 레이어(231)를 통해 행동 인식 장치(10)는 복수의 추상화 레이어(231)를 기초로 확장 영상 세그먼트로부터 속성 정보를 추출하고, 마지막 추상화 레이어로부터 추출된 속성 정보로부터 획득된 특징맵을 분류기(232)의 입력 데이터로 적용할 수 있다. 예시적으로, 탐지 네트워크(230)는 상기 특징맵을 제1 분류기(도시되지 않음)에 입력 데이터로 적용하여 통합 영상 세그먼트 내의 객체의 행동 유형을 식별하고, 제1 분류기의 결과값 및/또는 특징맵을 제2 분류기(도시되지 않음)의 입력 데이터로 적용하여 통합 영상 세그먼트 내의 객체의 행동이 발현되는 시점이 행동 유형의 끝 또는 시작 시점에 대응되는지 여부를 검출할 수 있다. 이때, 제1 분류기는 다중 클래스 분류기일 수 있으며, 제2 분류기는 이진분류기일 수 있으나, 이에 한정되는 것은 아니며, 상기 제1 및 제2 분류기는 복수의 이진분류기로 구현될 수도 있다.
- [0029] 도 3은 본 발명의 일 실시예에 따른 행동 인식 장치(10)의 구성을 도시한다. 행동 인식 장치(10)는 메모리(310) 및 프로세서(320)를 포함한다.
- [0030] 메모리(310)에는 프로세서(320)의 처리 및 제어를 위한 프로그램들(하나 이상의 인스트럭션들)을 저장할 수 있다. 메모리(310)에 저장된 프로그램들은 기능에 따라 복수 개의 모듈들로 구분될 수 있다. 일 실시예에 따라 메모리(310)는 영상으로부터 영상에 포함된 객체의 행동을 인식하는 프로그램을 저장할 수 있다. 상기 프로그램은 뉴럴 네트워크 모듈을 포함할 수 있다.
- [0031] 뉴럴 네트워크 모듈은 프로포절 네트워크, 미래 영상 생성자 네트워크 및 탐지 네트워크에 포함된 복수의 레이어들과 분류기들을 포함할 수 있다. 이때, 상기한 네트워크들에 포함된 각 추상화 레이어는 입력 영상로부터 이미지의 속성 정보를 추출하여 특징맵을 생성하는 하나 이상의 인스트럭션을 포함하는 3D 컨볼루션 레이어(3D convolutional layer), 및/또는 추출된 속성 정보로부터 대표값을 결정하는 하나 이상의 인스트럭션을 포함하는 풀링 레이어(pooling layer)를 포함할 수 있다. 또한, 미래 영상 생성자 네트워크에 포함된 각 생성자 레이어는 특징맵을 기초로 추출된 속성 정보를 변형하거나, 기 저장된 다른 이미지들의 특징 정보를 합성하는 하나 이상의 인스트럭션을 포함하는 디컨볼루션 레이어(deconvolutional layer)를 포함할 수 있다.
- [0032] 도 4는 본 발명의 일 실시예에 따른 탐지 네트워크의 일례를 도시한다.
- [0033] 프로세서(320)는 하나 이상의 코어(core, 도시되지 않음) 및 그래픽 처리부(도시되지 않음) 및/또는 다른 구성요소와 신호를 송수신하는 연결 통로(예를 들어, 버스(bus) 등)를 포함할 수 있다.
- [0034] 일 실시예에 따라 프로세서(320)는 뉴럴 네트워크 모듈 내의 각 네트워크에 포함된 하나 이상의 인스트럭션들을 병렬적으로 처리할 수 있다.

- [0035] 이하, 도 5를 참조하여, 프로세서(320)가 영상 내의 객체의 행동을 인식하는 방법을 설명한다.
- [0036] 먼저, 프로세서(320)는 영상으로부터 영상 세그먼트를 획득한다(S510). 이때, 영상은 행동 인식 장치(10)에 구비된 이미지 센서로부터 실시간으로 획득된 것일 수 있으며, 외부 영상 장치로부터 수신된 것일 수도 있다. 그리고 영상 세그먼트는 연속하는 영상 프레임들의 집합이다.
- [0037] 이후, 프로세서(320)는 프로포절 네트워크를 통해 영상 세그먼트 내에 객체의 행동이 존재하는지 여부를 판별한다(S520).
- [0038] 전술한 바와 같이 프로포절 네트워크는 복수의 추상화 레이어와 이진분류기를 포함한다. 프로세서(320)는 복수의 추상화 네트워크를 통해 영상 세그먼트의 속성 정보를 추출한다. 예를 들어, 프로세서(320)는 복수의 레이어 중 제 1 레이어를 이용하여 영상 프레임으로부터 직선 정보를 추출할 수 있다. 또한, 디바이스는 추출된 직선 정보를 제 1 레이어와 연결된 제 2 레이어에 입력 데이터로 적용하여, 제 2 레이어로부터 직선의 변화값을 추출할 수 있다. 전술한 방식과 같이 디바이스는 복수의 레이어 각각에 영상 세그먼트를 입력하거나 이전 레이어로부터 추출된 속성 정보를 입력 데이터로 적용함으로써, 다양한 속성 정보를 추출할 수 있다. 이어서 프로세서(320)는 마지막 레이어에서 추출된 속성 정보를 조합하여 특징맵을 추출한다.
- [0039] 프로세서(320)는 추출된 특징맵을 이진분류기의 입력 데이터로 적용하여 영상 세그먼트에 객체의 행동이 존재하는지 여부를 판별할 수 있다.
- [0040] 만약, 상기 영상 세그먼트 내에 객체의 행동이 존재하지 않는 것으로 판별되면, 프로세서(320)는 다음 영상 세그먼트에 대해 S510 및 S520 단계를 반복 수행한다. 이를 통해, 프로세서(320)는 불필요한 연산 부하를 최소화할 수 있다. 한편, 다음 영상 세그먼트는 현재 영상 세그먼트의 일부 영상 프레임(예컨대, 현재 영상 세그먼트의 마지막 영상 프레임 등)을 중복하여 포함할 수 있다.
- [0041] 그러나, 영상 세그먼트 내에 객체의 행동이 존재하면, 프로세서(320)는 미래영상 생성자 네트워크를 통해 영상 세그먼트에 연속하는 하나 이상의 미래영상 프레임을 생성하여 통합 영상 세그먼트를 구성한다 (S530).
- [0042] 미래영상 생성자 네트워크는 복수의 추상화 레이어와 복수의 생성자 레이어를 포함한다. 프로세서(320)는, 프로포절 네트워크에서와 마찬가지로, 미래영상 생성자 네트워크에 포함된 복수의 추상화 레이어를 통해 특징맵을 획득하며, 복수의 생성자 레이어를 통해 특징맵을 기초로 영상 세그먼트의 속성 정보가 변형되거나, 기 저장된 다른 이미지들의 특징 정보가 합성된 하나 이상의 미래 영상 프레임을 생성한다. 프로세서(320)는 영상 세그먼트에 하나 이상의 미래영상 프레임을 더하여 통합 영상 세그먼트를 구성한다.
- [0043] 추가로, 미래영상 생성자 네트워크는 생성된 하나 이상의 미래영상 프레임이 상기 영상 세그먼트에 실질적으로 연속할 확률을 판별하고, 이를 상기 추상화 레이어 및 생성자 레이어로 피드백하는 복수의 판별자 레이어를 더 포함할 수 있다. 프로세서(320)는 복수의 판별자 레이어를 통해 산출된 확률값을 기초로 미래영상 프레임을 반복적으로 생성하여 상기 판별자 레이어의 확률값을 향상시켜, 보다 정확한 미래영상 프레임을 생성할 수 있다.
- [0044] 이후, 프로세서(320)는 탐지 네트워크를 통해 통합 영상 세그먼트 내의 객체의 행동 유형 및 행동 시점을 검출한다(S540).
- [0045] 전술한 바와 같이 탐지 네트워크는 복수의 추상화 레이어와 하나 이상의 분류기를 포함한다. 프로세서(320)는 탐지 네트워크에 포함된 복수의 추상화 레이어를 통해 통합 영상 세그먼트의 속성 정보를 추출하여, 마지막 추상화 레이어에서 추출된 속성 정보를 조합하여 특징맵을 획득한다. 그리고 프로세서(320)는 특징맵을 하나 이상의 분류기의 입력 데이터로 적용한다. 예시적으로, 프로세서(320)는 특징맵을 제1 분류기에 입력 데이터로 적용하여 통합 영상 세그먼트 내의 객체의 행동 유형을 식별하고, 상기 제1 분류기의 결과값 및/또는 상기 특징맵을 상기 제2 분류기의 입력 데이터로 적용하여 상기 통합 영상 세그먼트 내의 객체의 행동이 발현되는 시점이 상기 행동 유형의 끝 또는 시작 시점에 대응되는지 여부를 검출할 수 있다.
- [0046] 이후, 프로세서(320)는 다시 S510 내지 S540을 반복 수행하여, 실시간으로 수신되는 영상으로부터 객체의 행동 유형 및 행동 시점을 탐지할 수 있다. 또한, 프로세서(320)는 탐지된 행동 유형이 기 설정된 이상 행동에 대응되는 경우, 이를 행동 인식 장치(10)에 구비된 알림 장치를 통해 행동 유형 및 행동 시점에 대한 정보를 알림하거나, 외부 장치로 상기한 정보를 전송할 수 있다.
- [0047] 한편, 단계 S520 내지 S540의 각 뉴럴 네트워크는 기 수행된 학습을 통해 상기한 속성 정보 추출 동작, 분류 동작, 미래영상 프레임 생성 등을 수 있다. 예를 들어, 프로세서(320)는 S510 단계 이전에 수행된 영상 세그먼트를 행동 또는 비행동으로 분류한 결과, 영상 세그먼트에 연속하는 미래영상 프레임을 생성한 결과, 영상 세그

먼트에 미래영상 프레임을 더한 결과로부터 행동 유형 및 행동 시점 판별 결과 등의 정확도를 높일 수 있는 방향으로 각 네트워크에 포함된 각 레이어 및 분류기를 학습시킬 수 있다. 이에 대해서는, 도 6을 참조하여 보다 구체적으로 후술한다.

[0048] 도 6은 본 발명의 일 실시예에 따른 프로세서(320)를 설명하기 위한 도면이다. 도 6을 참조하면, 프로세서(320)는 데이터 학습부(610) 및 데이터 인식부(620)를 포함한다.

[0049] 데이터 학습부(610)는 프로포절 네트워크를 독립 학습을 수행하며, 미래영상 생성자 네트워크와 탐지 네트워크는 상호 의존적 학습을 수행한다. 이때, 학습은 각 레이어 및/또는 분류기 클래스의 파라미터(예컨대, 가중치 등)를 조정하거나, 연산에서 적어도 하나의 레이어를 생략 및/또는 추가하는 것일 수 있다.

[0050] 구체적으로, 데이터 학습부(610)는 프로포절 네트워크의 입력 데이터(즉, 영상 세그먼트)로부터 객체의 행동이 존재하는지 여부를 검출하기 위한 제1 기준을 독립적으로 학습시킨다. 예를 들어, 데이터 학습부(610)는 학습 영상 세그먼트와 해당 학습 영상 세그먼트 내에 객체의 행동 존부에 대한 정보(즉, "행동" 또는 "비행동")를 프로포절 네트워크에 입력하여 학습시킬 수 있다.

[0051] 또한, 데이터 학습부(610)는 학습 영상 세그먼트와 상기 학습 영상 세그먼트에 연속하는 프레임을 미래영상 생성자 네트워크에 입력하여 학습시킨다. 또한, 데이터 학습부(610)는 미래영상 생성자 네트워크의 결과값(즉, 통합 영상 세그먼트)과 통합 영상 세그먼트 내의 객체의 행동 유형 및 행동 시점을 탐지 네트워크에 입력하여 학습시킨다. 이때, 데이터 학습부(610)는 탐지 네트워크의 손실함수(loss function)의 값을 미래영상 생성자 네트워크에 더 입력하여, 미래영상 생성자 네트워크가 상기 손실함수의 값을 기초로 학습되도록 한다. 여기서, 손실함수는, 예시적으로, 음의 로그-우도(negative log-likelihood) 함수가 이용될 수 있으나, 이에 한정되는 것은 아니다. 이와 같이, 데이터 학습부(610)는 미래영상 생성자 네트워크와 탐지 네트워크에 대해 상호 의존적 학습을 수행함으로써 보다 정확하게 객체의 행동 유형 및 행동 시점을 탐지할 수 있도록 할 수 있다.

[0052] 데이터 인식부(620)는 데이터 학습부(610)를 통해 학습된 기준에 기초하여, 영상 세그먼트로부터 객체의 행동 유형 및 행동 시점을 판별할 수 있다. 이에 대해서는, 도 1 내지 도 5를 참조하여 전술하였으므로, 자세한 설명은 생략한다.

[0053] 한편, 데이터 학습부(610) 및 데이터 인식부(620) 중 적어도 하나는, 적어도 하나의 하드웨어 칩 형태로 제작되어 행동 인식 장치(10)에 탑재될 수 있다. 또는, 데이터 학습부(610) 및 데이터 인식부(620) 중 적어도 하나는 소프트웨어 모듈로 구현될 수 있다. 데이터 학습부(610) 및 데이터 인식부(620) 중 적어도 하나가 소프트웨어 모듈(또는, 인스트럭션(instruction) 포함하는 프로그램 모듈)로 구현되는 경우, 소프트웨어 모듈은 컴퓨터로 읽을 수 있는 판독 가능한 비일시적 판독 가능 기록매체(non-transitory computer readable media)에 저장될 수 있다. 또한, 이 경우, 적어도 하나의 소프트웨어 모듈은 OS(Operating System)에 의해 제공되거나, 소정의 어플리케이션에 의해 제공될 수 있다.

[0054] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체를 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.

[0055] 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수 있다.

[0056] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성요소들도 결합된 형태로 실시될 수 있다.

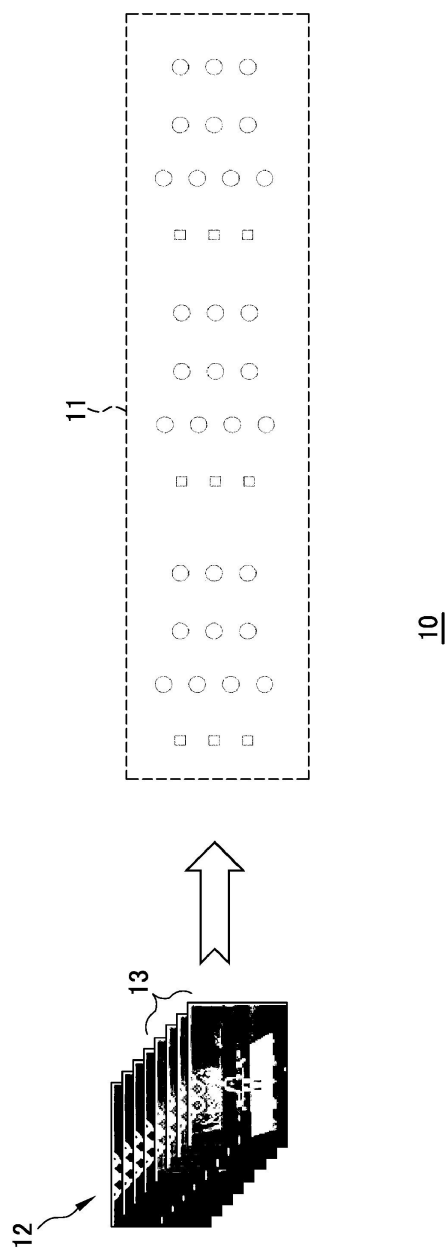
부호의 설명

[0057] 10: 행동 인식 장치

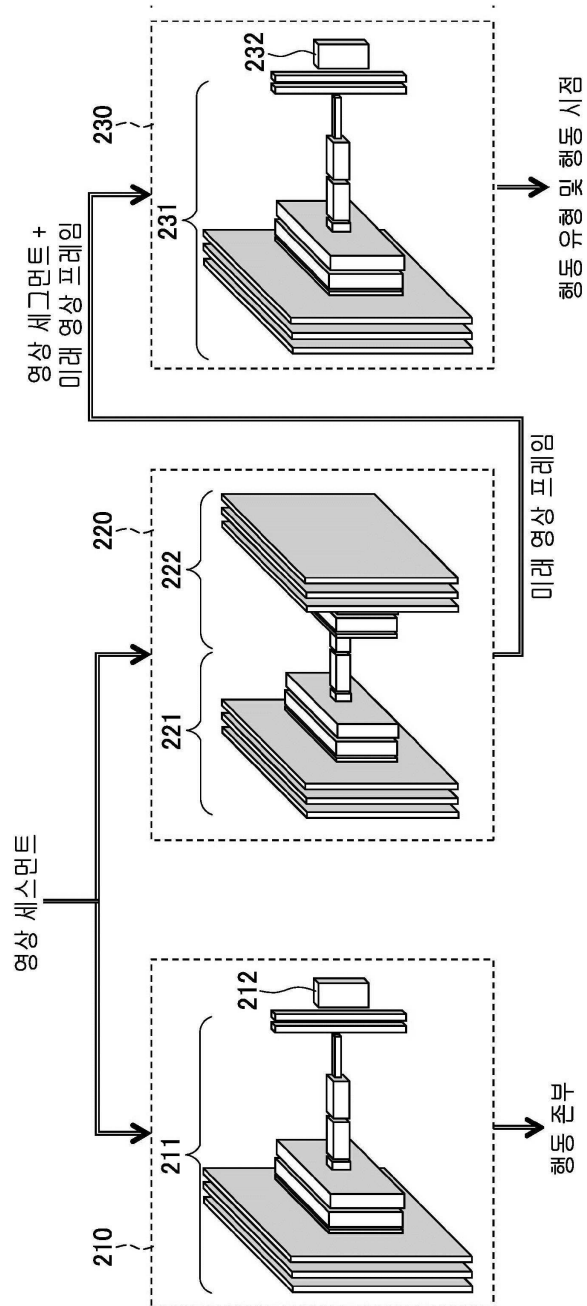
- 11: 뉴럴 네트워크
 12: 실시간 영상 13: 영상 세그먼트
 210: 프로포절 네트워크(proposal network)
 220: 미래영상 생성자 네트워크
 230: 탐지 네트워크(detection network)
 310: 메모리 320: 프로세서
 610: 데이터 학습부 620: 데이터 인식부

도면

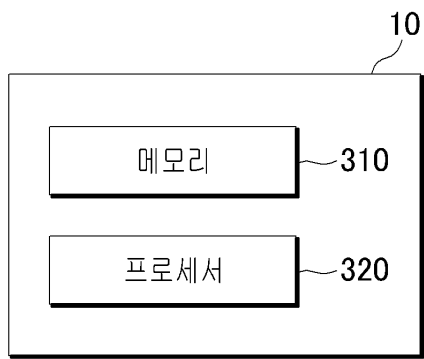
도면1



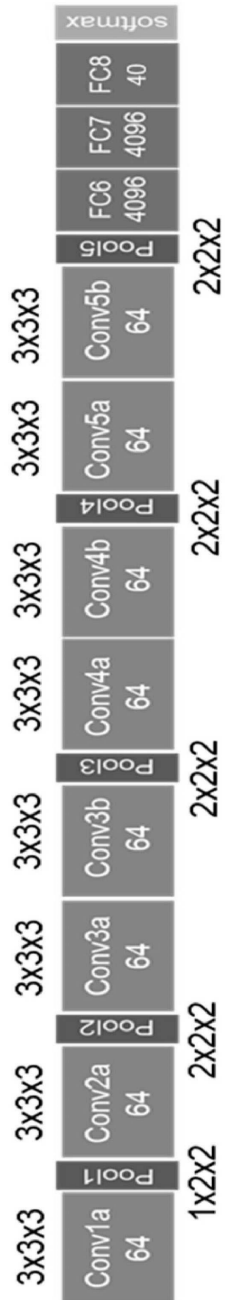
도면2



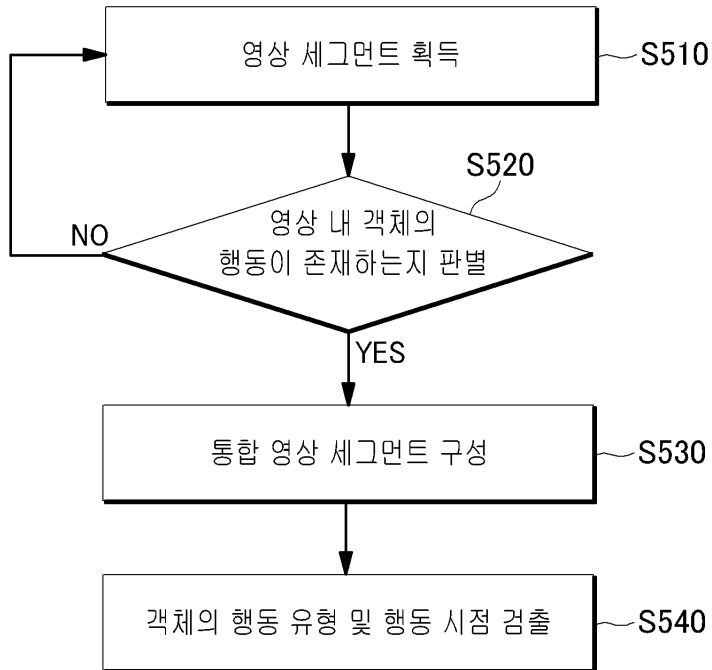
도면3



도면4



도면5



도면6

