

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2021-0132998
(43) 공개일자 2021년11월05일

- (51) 국제특허분류(Int. Cl.)
G06T 7/292 (2017.01) *G06N 3/08* (2006.01)
G06T 7/20 (2017.01) *H04N 19/139* (2014.01)
H04N 19/513 (2014.01)
- (52) CPC특허분류
G06T 7/292 (2017.01)
G06N 3/08 (2013.01)
- (21) 출원번호 10-2020-0051607
 (22) 출원일자 2020년04월28일
 심사청구일자 없음

- (71) 출원인
 한국전력공사
 전라남도 나주시 전력로 55(빛가람동)
 경희대학교 산학협력단
 경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
- (72) 발명자
 서덕영
 경기도 성남시 분당구 수내로 174, 201동 202호(수내동, 푸른마을벽산아파트)
 윤지혁
 대전광역시 유성구 북유성대로488번길 160, 316동 109호(수남동)
 차은영
 인천광역시 계양구 효서로 385, 205동 308호(작전동)
- (74) 대리인
 특허법인 정안

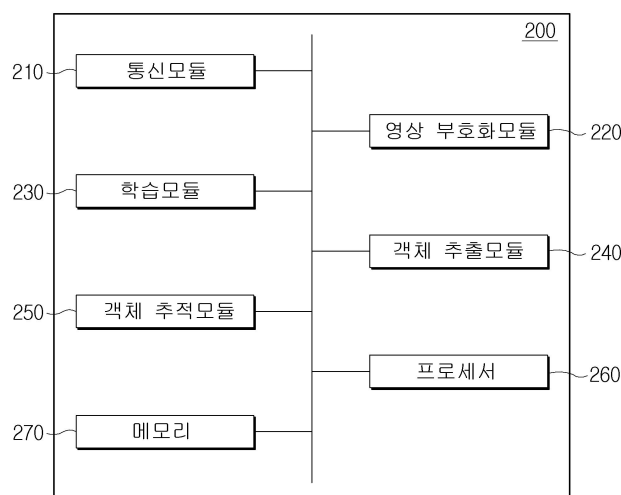
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **인공 신경망을 이용한 영상 내 객체 추적 장치 및 방법**

(57) 요약

본 발명은 영상 내 객체 추적 장치로서, 외부 장치로부터 연속하는 영상을 수신하는 통신모듈; 상기 영상을 부호화하여 움직임 정보를 획득하는 영상 부호화모듈; 상기 움직임 정보를 이용하여 인공 신경망을 학습시키는 학습모듈; 상기 인공 신경망을 기초로 상기 영상 내 적어도 하나의 객체를 추출하는 객체 추출모듈; 및 상기 움직임 정보를 기초로 상기 객체의 이동을 추적하는 객체 추적모듈을 포함한다.

대표도 - 도2



(52) CPC특허분류

G06T 7/20 (2013.01)

H04N 19/139 (2015.01)

H04N 19/513 (2015.01)

G06T 2207/20084 (2013.01)

명세서

청구범위

청구항 1

영상 내 객체 추적 장치로서,
외부 장치로부터 연속하는 영상을 수신하는 통신모듈;
상기 영상을 부호화하여 움직임 정보를 획득하는 영상 부호화모듈;
상기 움직임 정보를 이용하여 인공 신경망을 학습시키는 학습모듈;
상기 인공 신경망을 기초로 상기 영상 내 적어도 하나의 객체를 추출하는 객체 추출모듈; 및
상기 움직임 정보를 기초로 상기 객체의 이동을 추적하는 객체 추적모듈을 포함하는, 객체 추적 장치.

청구항 2

제 1 항에 있어서,
상기 영상 부호화모듈은,
상기 영상 간 예측을 포함하는지 여부를 판단하고, 상기 영상 간 예측을 포함하는 경우에 상기 영상을 부호화하는, 객체 추적 장치.

청구항 3

제 1 항에 있어서,
상기 영상 부호화모듈은,
상기 영상의 현재 블록과 유사도가 가장 높은 예측 블록을 선택하고, 상기 예측 블록과 상기 현재 블록 사이의 이동 벡터를 산출하여 상기 움직임 정보를 획득하는, 객체 추적 장치.

청구항 4

제 1 항에 있어서,
상기 객체 추출모듈은 상기 객체의 위치를 기반으로 적어도 하나의 관심 영역을 설정하는, 객체 추적 장치.

청구항 5

제 1 항에 있어서,
상기 객체 추적모듈은 상기 관심 영역을 기초로 상기 객체의 이동을 추적하는, 객체 추적 장치.

청구항 6

영상 내 객체 추적 방법으로서,
외부 장치로부터 연속하는 영상을 수신하는 동작;

상기 수신된 영상을 부호화하여 움직임 정보를 획득하는 동작;
 상기 획득한 움직임 정보를 이용하여 인공 신경망을 학습시키는 동작;
 상기 학습된 인공 신경망을 기초로 상기 영상 내 적어도 하나의 객체를 추출하는 동작; 및
 상기 추출된 객체의 이동을 추적하는 동작을 포함하는, 객체 추적 방법.

청구항 7

제 6 항에 있어서,
 상기 영상을 부호화하는 동작은,
 상기 영상 간 예측을 포함하는지 여부를 판단하는 동작; 및
 상기 예측을 포함하는 경우에 상기 영상을 부호화하는 동작을 포함하는, 객체 추적 방법.

청구항 8

제 6 항에 있어서,
 상기 움직임 정보를 획득하는 동작은,
 상기 영상의 현재 블록과 유사도가 가장 높은 예측 블록을 인접하는 영상에서 선택하는 동작; 및
 상기 선택된 예측 블록과 상기 현재 블록 사이의 이동 벡터를 산출하여 상기 움직임 정보를 획득하는 동작을 포함하는, 객체 추적 방법.

청구항 9

제 6 항에 있어서,
 상기 객체의 이동을 추적하는 동작은,
 상기 움직임 정보를 기초로 상기 객체의 위치를 기반으로 관심 영역을 설정하는 동작; 및
 상기 설정된 관심 영역을 기초로 상기 객체의 이동을 추적하는 동작을 포함하는, 객체 추적 방법,

청구항 10

컴퓨터 판독 가능 저장 매체에 있어서,
 컴퓨터 상에서 실행될 때, 제 6 항 내지 제 9 항 중 어느 하나의 항에 따른 객체 추적 방법을 수행하는 컴퓨터 프로그램을 포함하는 컴퓨터 판독가능 저장 매체.

발명의 설명

기술 분야

[0001] 다양한 실시 예는 인공 신경망을 이용한 영상 내 객체 추적 장치 및 방법에 관한 것이다.

배경 기술

[0003] 일반적으로 비주얼 트래킹(visual tracking)이라고 불리는 객체 트래킹은 일련의 이미지 또는 영상에서 타겟

(target)이 되는 객체를 감지, 추출, 식별 및 위치를 찾는 영상 처리 프로세스(process)이다.

- [0004] 이러한 영상 처리 프로세스는 의료 진단, 교통 흐름 모니터링, 등 다양한 분야에서 사용되고 있으며, 따라서 영상 처리 프로세서를 수행하기 위한 다양한 방법들이 지속적으로 연구되고 있다.
- [0005] 그러나, 종래의 영상 처리 프로세스 방법에 의하면, 타겟이 되는 객체의 정보(semantic information)를 캡처(capture)할 수 없고, 객체의 모양 변화에 견고하지 않으며, 차별적 일률(discriminative power)이 제한적이어서 낮은 레벨의 수작업에 의존하므로 영상의 화질 저하를 발생시킨다는 문제점이 존재하였다.
- [0006] 따라서, 영상 처리 프로세서에 있어 특화된 알고리즘을 이용하여 실시간으로 입력된 영상으로부터 객체 트래킹을 고속화할 수 있는 인공 신경망을 이용한 객체 추적 장치 및 방법에 대한 요구가 당업계에서 점차 증가하고 있는 상황이다.

발명의 내용

해결하려는 과제

- [0008] 따라서 본 발명은 상기한 문제점들을 해결하고자 안출된 것으로서, 본 발명의 목적은 인공 신경망을 이용하여 실시간으로 입력되는 영상으로부터 객체 트래킹을 고속화할 수 있는 객체 추적 장치 및 방법을 제공하는 것이다.
- [0009] 본 문서에서 이루고자 하는 기술적 과제는 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0011] 본 발명의 다양한 실시 예들에 따르면, 영상 내 객체 추적 장치로서, 외부 장치로부터 연속하는 영상을 수신하는 통신모듈; 상기 영상을 부호화하여 움직임 정보를 획득하는 영상 부호화모듈; 상기 움직임 정보를 이용하여 인공 신경망을 학습시키는 학습 모듈; 상기 인공 신경망을 기초로 상기 영상 내 적어도 하나의 객체를 추출하는 객체 추출모듈; 및 상기 움직임 정보를 기초로 상기 객체의 이동을 추적하는 객체 추적모듈을 포함한다.
- [0012] 또한, 상기 영상 부호화모듈은, 상기 영상 간 예측을 포함하는지 여부를 판단하고, 상기 영상 간 예측을 포함하는 경우에 상기 영상을 부호화한다.
- [0013] 또한, 상기 영상 부호화모듈은, 상기 영상의 현재 블록과 유사도가 가장 높은 예측 블록을 선택하고, 상기 예측 블록과 상기 현재 블록 사이의 이동 벡터를 산출하여 상기 움직임 정보를 획득한다.
- [0014] 또한, 상기 객체 추출모듈은 상기 객체의 위치를 기반으로 적어도 하나의 관심 영역을 설정한다.
- [0015] 또한, 상기 객체 추적모듈은 상기 관심 영역을 기초로 상기 객체의 이동을 추적한다.
- [0016] 추가로, 상기의 기술적 과제를 해결하기 위한 본 발명의 또 다른 실시 예에 따른 영상 내 객체 추적 방법으로서, 외부 장치로부터 연속하는 영상을 수신하는 동작; 상기 수신된 영상을 부호화하여 움직임 정보를 획득하는 동작; 상기 획득한 움직임 정보를 이용하여 인공 신경망을 학습시키는 동작; 상기 학습된 인공 신경망을 기초로 상기 영상 내 적어도 하나의 객체를 추출하는 동작; 및 상기 추출된 객체의 이동을 추적하는 동작을 포함한다.
- [0017] 또한, 상기 영상을 부호화하는 동작은, 상기 영상 간 예측을 포함하는지 여부를 판단하는 동작; 및 상기 예측을 포함하는 경우에 상기 영상을 부호화하는 동작을 포함한다.
- [0018] 또한, 상기 움직임 정보를 획득하는 동작은, 상기 영상의 현재 블록과 유사도가 가장 높은 예측 블록을 인접하는 영상에서 선택하는 동작; 및 상기 선택된 예측 블록과 상기 현재 블록 사이의 이동 벡터를 산출하여 상기 움직임 정보를 획득하는 동작을 포함한다.
- [0019] 또한, 상기 객체의 이동을 추적하는 동작은, 상기 움직임 정보를 기초로 상기 객체의 위치를 기반으로 관심 영

역을 설정하는 동작; 및 상기 설정된 관심 영역을 기초로 상기 객체의 이동을 추적하는 동작을 포함한다.

[0020] 추가로, 상기의 기술적 과제를 해결하기 위한 본 발명의 다른 실시 예에 따른 컴퓨터 판독 가능 저장 매체는, 컴퓨터상에서 실행될 때, 상기의 객체 추적 방법을 수행하기 위한 프로그램이 기록될 수 있다.

발명의 효과

[0022] 본 발명은 인공 신경망을 이용하여 실시간으로 입력되는 영상으로부터 객체 트래킹을 고속화할 수 있다.

[0023] 또한, 본 발명은 인공 신경망을 이용하여 객체 추출 및 객체 추적에 대한 오차를 최소화함으로써, 객체 트래킹 프로세스에 대한 정확도 및 신뢰도를 향상시킬 수 있다.

[0024] 또한, 본 발명은 객체 트래킹 프로세스를 최적화, 자동화함으로써, 객체 트래킹을 위한 비용과 시간을 최소화할 수 있다.

[0025] 본 개시에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0027] 도 1은 본 발명의 일 실시 예에 따른 인공 신경망 구조의 일 예를 도시한다.

도 2는 본 발명의 일 실시 예에 따른 인공 신경망을 이용한 영상 내 객체 추적 장치(200)의 기능적 블록도를 도시한다.

도 3은 본 발명의 일 실시 예에 따른 인공 신경망을 이용한 영상 내 객체 추적 방법을 설명하기 위한 순서도(S300)를 도시한다.

도 4는 본 발명의 일 실시 예에 따라 영상 부호화를 위해 복수의 예측 유닛(221)으로 분할된 영상의 일 예를 도시한다.

도 5는 본 발명의 일 실시 예에 따라 추출된 움직임 벡터(222)의 일 예를 도시한다.

도 6a는 본 발명의 일 실시 예에 따른 객체 추출모듈(240)을 통해 추출된 영상 내 객체와 관심 영역의 일 예를 도시한다.

도 6b는 본 발명의 일 실시 예에 따른 객체 추적모듈(250)을 통해 영상 내 객체의 이동을 추적하는 일 예를 도시한다.

도 6c는 본 발명의 일 실시 예에 따른 객체 추출모듈(240)을 통해 추출된 영상 내 관심 영역 기반 확장 영역의 일 예를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0028] 이하 다양한 실시 예들이 첨부된 도면을 참고하여 상세히 설명된다.

[0029] 도면 부호에 관계없이 동일하거나 유사한 구성요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략할 수 있다.

[0030] 이하의 설명에서 사용되는 구성요소에 대한 접미사 '모듈' 또는 '부'는 명세서 작성의 용이함만이 고려되어 부여되거나 혼용되는 것으로서, 그 자체로 서로 구별되는 의미 또는 역할을 갖는 것은 아니다. 또한, '모듈' 또는 '부'는 소프트웨어 또는 FPGA(field programmable gate array) 또는 ASIC(application specific integrated circuit)과 같은 하드웨어 구성요소를 의미하나, 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. '부' 또는 '모듈'은 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '부' 또는 '모듈'은 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로코드, 회로, 데이터, 데이터

터베이스, 데이터 구조들, 테이블들, 어레이들, 및 변수들을 포함할 수 있다. 하나의 구성요소, '부' 또는 '모듈'들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '부' 또는 '모듈'들로 결합되거나 추가적인 구성요소들과 '부' 또는 '모듈'들로 더 분리될 수 있다.

- [0031] 본 발명의 몇몇 실시예들과 관련하여 설명되는 방법 또는 알고리즘의 단계는 프로세서에 의해 실행되는 하드웨어, 소프트웨어 모듈, 또는 그 2 개의 결합으로 직접 구현될 수 있다. 소프트웨어 모듈은 RAM 메모리, 플래시 메모리, ROM 메모리, EPROM 메모리, EEPROM 메모리, 레지스터, 하드 디스크, 착탈형 디스크, CD-ROM, 또는 당업계에 알려진 임의의 다른 형태의 기록 매체에 상주할 수도 있다. 예시적인 기록 매체는 프로세서에 커플링되며, 그 프로세서는 기록 매체로부터 정보를 판독할 수 있고 저장 매체에 정보를 기입할 수 있다. 다른 방법으로, 기록 매체는 프로세서와 일체형일 수도 있다. 프로세서 및 기록 매체는 주문형 집적회로(ASIC) 내에 상주할 수도 있다. ASIC은 사용자 단말기 내에 상주할 수도 있다.
- [0032] 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.
- [0033] 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.
- [0034] 우선 본 명세서에서 사용되는 용어들에 대하여 간략히 설명한다.
- [0035] 인공지능은 인간의 지능을 갖춘 컴퓨터 시스템 또는 장치이며, 인간의 지능을 기계 등에 인공적으로 구현한 것을 의미할 수 있다. 인공 지능은 또한 지능을 만들 수 있는 방법론이나 실현 가능성 등을 연구하는 과학 분야를 지칭하기도 한다.
- [0036] 인공 신경망은 인공지능을 구현하는 모델 또는 알고리즘으로써, 기계학습에서 생물학의 신경망을 모사하여 모델링한 통계학적 학습 알고리즘으로, 시냅스의 결합으로 네트워크를 형성한 인공 뉴런이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 또는 학습 알고리즘이라 할 수 있다.
- [0037] 인공 신경망(artificial neural network)은 입력 층, 출력 층 그리고 하나 이상의 은닉 층을 포함할 수 있다. 인공 신경망의 각 층은 신경망의 뉴런에 대응하는 복수의 노드를 포함하고, 인공 신경망의 한 층의 노드와 다른 층의 노드 간은 시냅스로 연결될 수 있다. 일 실시 예로 각 층의 모든 노드와 다음 층의 모든 노드가 시냅스로 연결된 인공 신경망을 완전 연결된 인공 신경망이라 칭할 수 있다.
- [0038] 인공 신경망에서 각 노드는 시냅스를 통해 입력되는 입력 신호들을 받고 각 입력 신호들에 대한 가중치 및 편향에 대한 활성 함수에 기초하여 출력 값을 생성할 수 있다.
- [0039] 심층신경망(deep neural network)은 입력층과 출력층 사이에 복수의 은닉층을 포함하는 인공 신경망을 통칭할 수 있다. 심층신경망은 복잡한 비선형 관계들을 모델링할 수 있으며, 그 목적에 따라 다양한 구조를 가질 수 있다. 예를 들면, 심층 신경망 구조로, 합성곱신경망(convolutional neural network), 순환신경망(recurrent neural network, RNN), LSTM(long short term memory), 등이 있을 수 있다.
- [0040] 합성곱 신경망은 이미지, 동영상, 문자열과 같은 구조적 공간 데이터의 특징을 식별하여 학습함으로서 이미지, 동영상을 분류하고 식별하는데 효과적일 수 있다. 순환신경망은 내부에 순환 구조가 들어 있어 과거 시간의 학습이 가중치와 곱해져 현재 학습에 반영될 수 있는 구조이며, 현재의 출력 결과는 과거 시간에서의 출력 결과에 영향을 받으며, 은닉 층은 일종의 메모리 기능을 수행한다. 따라서, 순차적인 데이터를 학습하여 분류 또는 예측을 수행하는 데 효과적일 수 있다.
- [0041] 콘볼루션 LSTM은 순환신경망의 일종으로써 순환신경망의 오래된 과거 데이터가 영향을 미치지 못하고 사라지는 문제점을 해소하는 신경망이며, 순환신경망과 마찬가지로, 순차적인 데이터를 학습하여 분류 또는 예측을 수행하는 데 효과적일 수 있다.
- [0042] 도 1은 일 실시 예에 따른 콘볼루션 LSTM 구조를 도시한 도면이다.
- [0043] 도 1을 참조하면, 콘볼루션 LSTM은 복수 개의 동일한 구조의 셀(cell)로 구성될 수 있다. 콘볼루션 LSTM의 각 셀은 지금 시간의 셀 상태(예: C_t) 및 출력(예: h_t)을 이전 시간의 셀 상태(예: C_{t-1}) 및 출력(예: h_{t-1})과 지

금 시간의 입력(예: x_t)에 기초하여 결정할 수 있다. 이에 따라, 콘볼루션 LSTM의 각 셀은 4개의 신경망 층(neural network layer)을 포함할 수 있다.

[0044] 첫 번째 신경망 층은 망각 게이트 층(forget gate layer)라고도 불리며 과거 및 현재의 정보를 반영할지를 결정하는 게이트이다. 일 실시 예에 따라 게이트는 시그모이드 함수(sigmoid function)를 이용하여 결정을 수행할 수 있다. 시그모이드 함수는 S자형 곡선을 갖는 수학 함수로 일 실시 예로 수학식 1의 함수로 정의될 수 있다.

수학식 1

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

[0045]

[0046] 수학식 1을 참조하면, 시그모이드 함수는 입력에 따라 0과 1 사이의 값을 출력할 수 있다.

[0047] 다음 수학식 2는 망각 게이트 층의 출력을 획득하기 위한 함수를 정의한다.

수학식 2

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

[0048]

[0049] 수학식 2를 참조하면, 망각 게이트 층의 출력(f_t)은 이전 셀 출력(예: h_{t-1})과 현재 입력(예: x_t)에 기초하여 이전 셀 상태(예: C_{t-1})를 망각할 것인지 아니면 보존할 것인지를 결정할 수 있다. 여기서, 망각 게이트 층의 가중치(W_f)는 학습에 의하여 변경될 수 있는 값이며, 망각 게이트 층의 바이어스(b_f)는 미리 설정되는 값이다. 그리고 σ 는 시그모이드 함수를 나타낸다.

[0050] 두 번째 신경망 층은 입력 게이트 층(input gate layer)라고도 불리며 새로운 정보 중 어떤 것을 셀 상태(예: C_t)에 저장할 것인지를 결정하며, 다음 수학식 3에 의해 정의될 수 있다.

수학식 3

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

[0051]

[0052] 수학식 3을 참조하면, 입력 게이트 층의 출력(i_t)은 이전 셀 출력(예: h_{t-1})과 현재 입력(예: x_t)에 기초하여 정보 중 어떤 것을 셀 상태에 저장할 것인지를 결정할 수 있다. 여기서, 입력 게이트 층의 가중치(W_i)는 학습에 의하여 변경될 수 있는 값이며, 입력 게이트 층의 바이어스(b_i)는 미리 설정되는 값이다. 그리고 σ 는 시그모이드 함수를 나타낸다.

[0053] 세 번째 신경망 층은 셀 상태에 저장하기 위한 새로운 후보 값($\tilde{C}_t$)을 결정한다. 새로운 후보 값($\tilde{C}_t$)은 다음 수학식 4에 따라 결정될 수 있다.

수학식 4

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

[0054]

[0055] 수학식 4를 참조하면, 새로운 후보 값($\tilde{C}_t$)은 이전 셀 출력(예: h_{t-1})과 현재 입력(예: x_t)에 기초하는 쌍곡탄젠트(hyperbolic tangent; $\tanh$) 함수의 출력 값일 수 있다. 쌍곡탄젠트 함수는 -1에서 1사이의 값을 가질 수 있다. 여기서, 가중치(W_C)는 학습에 의하여 변경될 수 있는 값이며, 바이어스(b_C)는 미리 설정되는 값이다. 그리

고 $\tanh$ 는 쌍곡탄젠트 함수를 나타낸다.

[0056] 상술한 세 개의 신경망 층의 결과에 기초하여 새로운 셀 상태(예: C_t)가 결정될 수 있다. 수학식 5는 새로운 셀 상태(예: C_t)를 결정하는 함수를 정의한다.

수학식 5

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

[0057]

[0058] 수학식 5를 참조하면, 과거의 셀 상태(C_{t-1})를 망각할 것인지 아니면 보존할 것인지를 나타내는 망각 게이트 층의 출력(f_t)을 곱하여 과거 셀 상태(C_{t-1})의 보존 여부를 결정하고, 갱신하고자 하는 후보 값($\tilde{C}_t$)과 갱신 정도를 나타내는 입력 게이트 층의 출력(i_t)을 곱하여 최종 갱신할 값을 결정함으로써 새로운 셀 상태(C_t)를 결정할 수 있다.

[0059] 마지막 네 번째 신경망 층은 출력 게이트 층(output gate layer)라고도 불리며 무엇을 출력으로 내 보낼 것인지를 결정하며, 다음 수학식 6에 의해 정의될 수 있다.

수학식 6

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$

[0060]

[0061] 수학식 6을 참조하면, 출력 게이트 층의 출력(o_t)은 이전 셀 출력(예: h_{t-1})과 현재 입력(예: x_t)에 기초하여 갱신된 셀 상태(C_t) 중에서 어떤 것을 출력할 것인지를 결정할 수 있다. 여기서, 출력 게이트 층의 가중치(W_o)는 학습에 의하여 변경될 수 있는 값이며, 출력 게이트 층의 바이어스(b_o)는 미리 설정되는 값이다. 그리고 σ 는 시그모이드 함수를 나타낸다.

[0062] 최종적으로 콘볼루션 LSTM의 셀 출력(h_t)은 다음 수학식 7에 따라 결정될 수 있다.

수학식 7

$$h_t = o_t * \tanh(C_t)$$

[0063]

[0064] 여기서, $\tanh$ 는 쌍곡 탄젠트 함수를 나타낸다.

[0065] 상술한 수학식 2 내지 7에 따라 셀 상태 및 출력을 결정하는 콘볼루션 LSTM은 반복 학습을 통하여 가중치(W_f , W_i , W_c , W_o)를 변경하여 최적의 결과가 출력될 수 있도록 할 수 있다.

[0066] 콘볼루션 LSTM의 출력에서 최적의 결과가 출력되도록 가중치를 변경하는 것을 학습, 특히 기계 학습이라 칭할 수 있다. 기계 학습은 실측 값을 이용하여 입력 데이터를 생성하고, 입력 데이터의 입력 후 콘볼루션 LSTM의 출력이 실측에 따른 결과 값이 나올 수 있도록 콘볼루션 LSTM을 학습시키는 것일 수 있다. 이때, 기계 학습은 일 실시 예에 따라, 경사 하강법(gradient descent) 및 역전파(back propagation)를 이용하여 콘볼루션 LSTM의 출력과 실측 출력 값 사이의 오차를 최소화하도록 가중치(W_f , W_i , W_c , W_o) 값들을 정하는 것일 수 있다.

[0067] 상술한 콘볼루션 LSTM과 같은 인공 신경망을 사용하여 본 발명에서는 연속하는 영상 내 객체를 감지한 이후에 감지된 객체의 추적을 고속화하는 객체 추적 장치 및 방법을 제공하고자 한다.

[0068] 참고로, 이하의 본 명세서의 객체 추적 장치(200)는 본 발명의 일 실시 예에 따른 객체 추적 방법을 구성하는 각 동작들을 실행하도록 구성될 수 있으며, 예를 들어, 도 3에 예시적으로 도시되는 바와 같이, 객체 추적 방법(S300)의 각 동작을 실행하도록 구성될 수 있다.

[0069] 한편, 도 3에서 각 동작들은 순차적으로 수행될 수 있으나, 반드시 순차적으로 수행되는 것은 아니다. 예를 들

어, 각 동작들의 순서가 변경될 수도 있으며, 적어도 두 동작들이 병렬적으로 수행될 수도 있다.

- [0070] 도 2 및 도 3을 참조하면, 본 발명의 일 실시 예에 따른 객체 추적 장치(200)는 통신모듈(210), 영상 부호화모듈(220), 학습모듈(230), 객체 추출모듈(240), 객체 추적모듈(250), 프로세서(260) 및 메모리(270)를 포함할 수 있다.
- [0071] 통신모듈(210)은 외부 장치(서버 또는 다른 전자 장치)로부터 연속하는 영상을 수신할 수 있다(S310). 예를 들어, 통신모듈(210)은 외부 장치(서버 또는 다른 전자 장치)와의 직접 연결 또는 네트워크를 통한 연결을 위해 제공되는 것으로서, 유선 및/또는 무선 통신모듈(210)로 구현될 수 있다. 연속하는 영상은 움직이는 순간적인 동작을 나타내는 정지영상이 순차적으로 출력되는 연속된 영상일 수 있다.
- [0072] 영상 부호화모듈(220)은 동작 S310을 통해 수신된 영상을 부호화할 수 있다(S320). 예를 들어, 영상 부호화모듈(220)은 영상을 효율적으로 분할하여 부호화 및 복호화를 수행할 수 있으며, 이러한 영상 부호화 및 복호화의 기반 단위로서 코딩 유닛(coding unit, CU)이 이용될 수 있다. 즉, 코딩 유닛은 예측, 변환, 양자화, 역변환, 역양자화 및 변환 계수의 부호화 및 복호화에 있어 기반 단위일 수 있다.
- [0073] 일 실시 예에서, 영상 부호화모듈(220)은 영상을 64X64 픽셀의 최대 코딩 유닛(largest coding unit, LCU)의 단위로 순차적으로 분할할 수 있다. 또한, 영상 부호화모듈(220)은 각각의 최대 코딩 유닛을 64X64 픽셀 내지 8X8 픽셀을 가진 쿼드 트리(quad tree) 형태의 코딩 유닛으로 분할하여 부호화 및 복호화를 수행할 수 있다.
- [0074] 일 실시 예에서, 영상 부호화 모듈(220)은 각각의 코딩 유닛이 인터 예측을 수행하기 위한 예측 유닛(prediction unit, PU)으로 분할되었을 경우에 움직임 추정(motion estimation) 과정을 통해 영상의 현재 프레임과 인접하는 다른 프레임에서 부호화 및 복호화를 위한 예측 유닛과 유사도가 가장 높은 예측 블록을 선택할 수 있다. 이를 통해, 영상 부호화모듈(220)은 예측 블록과 현재 블록 간의 움직임 정보를 산출할 수 있다. 움직임 정보는 움직임 파라미터(motion parameter), 움직임 벡터(motion vector), 등을 포함할 수 있다. 움직임 벡터는 인터 예측 또는 움직임 보상에서 사용되는 2차원 벡터를 의미한다. 따라서, 영상 부호화모듈(220)은 영상에서의 움직임 정보를 획득할 수 있는 동시에 분할된 예측 유닛의 조합을 결정한 후 변환, 양자화, 엔트로피 부호화 과정을 거쳐 영상을 압축할 수 있다.
- [0075] 일 실시 예에서, 영상 부호화모듈(220)은 연속하는 영상 간 예측을 포함하는지 여부를 판단하여, 영상 간 예측을 포함하는 경우에 부호화 및 복호화를 수행할 수 있다. 예를 들어, 영상 간 예측을 포함하지 않는 경우에 영상 부호화모듈(220)은 영상의 부호화 및 복호화를 수행할 수 있으며(S320), 유사도 또는 움직임 벡터가 존재하지 않으므로 객체 추출모듈(240)은 일반적인 객체 탐지 과정을 수행할 수 있다. 즉, 영상 부호화모듈(220)은 영상을 예측하는 모든 정보를 그대로 저장하거나 전송하는 고속 객체 추적 프레임워크를 구성할 수 있다. 이와 달리, 영상 간 예측을 포함하는 경우에 영상 부호화모듈(220)은 예측 과정을 거쳐 영상의 부호화 및 복호화를 수행할 수 있다.
- [0076] 학습모듈(230)은 동작 S320을 통해 부호화된 영상을 이용하여 인공 신경망을 학습시키고(S330), 학습된 인공 신경망을 이용하여 영상 내 객체의 이동을 예측할 수 있다. 즉, 학습모듈(230)은 학습된 인공 신경망을 이용하여 다음 영상 내의 객체 위치를 추정하여 제공할 수 있다. 이때, 학습모듈(230)은 학습된 인공 신경망 모델에 대한 성능평가를 추가적으로 수행할 수도 있다.
- [0077] 객체 추출모듈(240)은 동작 S330을 통해 학습된 인공 신경망을 이용하여 영상 내 적어도 하나의 객체를 추출할 수 있다(S340). 예를 들어, 객체 추출모듈(240)은 움직임 정보를 이용하여 영상 내에서 움직이는 객체(object)에 대한 대략적인 위치를 파악할 수 있다.
- [0078] 일 실시 예에서, 객체 추출모듈(240)은 영상 내에서 움직임 정보(예를 들어, 움직임 벡터)가 존재하지 않거나, 노이즈로 인해 추적된 객체를 찾지 못하거나, 또는 추적된 객체 이외의 다른 객체가 추출되는 경우에 객체를 다시 추출하여 분류할 수도 있다.
- [0079] 일 실시 예에서, 객체 추출모듈(240)은 영상에서 유의미한 움직임 즉, 유효 움직임이 추출되면 객체를 추출하고, 추출된 영상 내 객체의 위치를 기반으로 관심 영역(region of interest, RoI)을 설정할 수 있다. 나아가, 객체 추출모듈(240)은 관심 영역을 중심으로 하여 인접하는 복수의 영상 블록을 확장 영역으로 설정할 수도 있으며, 확장 영역은 영상 내 객체의 위치로부터 시작되는 예측 유닛들의 합이 될 수 있다. 이에 대해서는, 이하에서 보다 상술하기로 한다.
- [0080] 일 실시 예에서, 객체 추출모듈(240)은 대상으로 설정된 객체의 경계선을 추출하는 방법을 통해 객체 추출에 대

한 정확도를 향상시킬 수 있다. 예를 들어, 객체 추출모듈(240)은 서로 상이한 색을 가진 두 영역 간의 차이값을 계산하여 객체의 경계선을 추출할 수 있다.

[0081] 객체 추적모듈(250)은 동작 S340을 통해 추출된 객체의 이동을 추적할 수 있다(S350). 예를 들어, 객체 추적모듈(250)은 추적 대상으로 설정된 객체의 관심 영역에 대해 사각형 좌표 등을 산출하여 객체의 이동을 추적할 수 있다. 사각형 좌표의 일 예로서, 관심 영역은 객체를 최적으로 포함하도록 가상으로 형성된 사각형에 대한 좌상단 좌표(x,y), 가로축 길이(dx), 세로축 길이(dy)를 통해 형성될 수 있다. 즉, 관심 영역을 표시하기 위한 사각형 좌표는 (x, y, dx, dy)의 형태로 이루어질 수 있다.

[0082] 프로세서(260)는 영상 내 객체를 추출하고, 추출된 객체의 이동을 추적하기 위해 객체 추적 장치(200)에 포함된 구성요소들 중 적어도 둘 이상을 서로 조합하여 동작시킬 수 있다. 참고로, 프로세서(260)는 컨트롤러(controller), 프로세서(processor), 마이크로-프로세서(micro-processor), 마이크로-컨트롤러(micro-controller), 등으로 구현되어 본 발명에 따른 객체 추적 장치(200)의 동작, 기능, 등을 총괄적으로 제어할 수 있다.

[0083] 일 실시 예에서, 프로세서(260)는 영상 처리에 소요되는 시간을 고려하여 객체 추적 장치(200)를 동작시킬 수 있다. 예를 들어, 프로세서(260)는 영상 처리 프레임율이 10FPS(frame per second, FPS)로 적용되는 경우에 다음 수학식 8에 따라 프로세싱 타임을 결정할 수 있다.

수학식 8

$$T_p^{tot} = T_d + T_f + (T_{ROI}^I + T_{ROI}^O) + (T_1 + T_c^I) + T_r \leq 6ms$$

[0084] 수학식 8을 참조하면, T_{ROI}^I 는 움직임 벡터 기반 추적 관심 영역 탐색 시간, T_{ROI}^O 는 탐색되지 않은 부분에 대한 관심 영역 탐색 시간, T_1 은 이전 객체와 추적된 객체를 비교하는데 소요되는 시간, T_c^I 은 추적된 객체가 아닌 경우에 새로운 객체를 분류하는데 소요되는 시간을 각각 의미할 수 있다. 즉, 프로세서(260)는 기존의 객체 추적 프로세싱과 비교하여 고속의 객체 추적 기능을 지원할 수 있다.

[0086] 메모리(270)는 객체 추적 장치(200)를 통해 생성된 다양한 데이터가 저장될 수 있다. 참고로, 메모리(270)는 통상의 기술자에게 알려진 바와 같이, HDD(Hard Disk Drive), ROM(Read Only Memory), RAM(Random Access Memory), EEPROM(Electrically Erasable and Programmable Read Only Memory), 플래시 메모리(flash memory), CF(Compact Flash) 카드, SD(Secure Digital) 카드, SM(Smart Media) 카드, MMC(Multimedia) 카드 또는 메모리 스틱(Memory Stick) 등 정보의 입출력이 가능한 다양한 형태의 저장 장치로 구현될 수 있다.

[0087] 또한, 본 발명의 추가 실시 예에 따르면, 메모리(270)에는 또는 메모리(270)와는 별개로 데이터 백업(data backup)을 위한 추가 메모리가 더 구비될 수 있으며, 프로세서(260)는 메모리(270)에 저장된 다양한 데이터를 백업하여 상기 추가 메모리에 저장함으로써, 데이터 유실 또는 분실에 대한 능동적인 대처를 가능하게 한다.

[0088] 참고로, 도 2에 도시되는 영상 내 객체 추적 장치(200)의 각 엘리먼트(210, 220, 230, 240, 250, 260, 270)는 객체 추적 장치(200)의 동작 및 기능을 설명하기 위한 예시적인 엘리먼트에 해당할 뿐 이에 한정되지 않으며, 추가의 엘리먼트(예를 들어, 객체의 관심 영역이 표시되도록 영상을 출력하는 디스플레이, 사용자 입력부, 등)가 더 구현될 수 있음은 명백할 것이다. 이하에서는 객체 추적 장치(200)에 의해 객체를 추출 및 추적하기 위한 예시도를 각각 설명하기로 한다.

[0090] 도 4는 본 발명의 일 실시 예에 따라 영상 부호화를 위해 복수의 예측 유닛(221)으로 분할된 영상의 일 예를 도시한다.

[0091] 도 4를 참조하면, 영상 부호화모듈(220)은 코딩 유닛(coding unit, CU) 또는 인터 예측을 위한 복수의 예측 유닛(221)으로 영상을 분할하여 움직임 정보를 획득할 수 있다. 예를 들어, 영상 부호화모듈(220)은 움직임 추정(motion estimation) 과정을 통해 영상의 현재 프레임과 인접하는 다른 프레임에서 부호화 및 복호화를 위한 예측 유닛(221)과 유사도가 가장 높은 예측 블록을 선택할 수 있다. 이를 통해, 영상 부호화모듈(220)은 예측 블록과 현재 블록 간의 움직임 정보를 산출할 수 있다. 이때, 움직임 정보는 움직임 파라미터(motion parameter),

움직임 벡터(motion vector), 등을 포함할 수 있다.

- [0093] 도 5는 본 발명의 일 실시 예에 따라 추출된 움직임 벡터(222)의 일 예를 도시한다.
- [0094] 도 5를 참조하면, 영상 부호화모듈(220)은 예측 블록과 현재 블록 간의 복수의 움직임 벡터(222)를 산출할 수 있다. 움직임 벡터(222)는 인터 예측 또는 움직임 보상에서 사용되는 2차원 벡터를 의미할 수 있다. 예컨대, 움직임 벡터(222)는 시간상으로 인접하는 영상 내에 존재하는 예측 유닛과 가장 유사한 블록을 선택하고, 선택된 블록의 크기와 블록의 이동 벡터를 추정된 정보를 포함할 수 있다.
- [0096] 도 6a는 본 발명의 일 실시 예에 따른 객체 추출모듈(240)을 통해 추출된 영상 내 객체와 관심 영역의 일 예를 도시하고, 도 6b는 본 발명의 일 실시 예에 따른 객체 추적모듈(250)을 통해 영상 내 객체의 이동을 추적하는 일 예를 도시한다.
- [0097] 도 6a 및 도 6b를 참조하면, 객체 추적모듈(250)은 영상 부호화모듈(220)을 통해 산출된 움직임 벡터(222)를 기초로 영상 내 객체(object, 241)의 이동을 추적할 수 있다. 이를 구현하기 위해, 객체 추출모듈(240)은 영상에서 유의미한 움직임 즉, 유효 움직임이 추출되면 객체(241)를 추출하고, 추출된 영상 내 객체(241)의 위치를 기반으로 관심 영역(region of interest, RoI, 242)을 설정할 수 있다. 이때, 영상 내 관심 영역(242)은 움직임 벡터의 누적된 값이 기설정된 임계값을 초과하는 영상 블록에 대해 사각 바운딩 박스로 표시되어 디스플레이상에 출력될 수 있다. 즉, 객체 추적모듈(250)은 추출된 움직임 벡터(222)와 설정된 관심 영역(242)을 기초로 객체(241)의 이동을 보다 신속하게 추적할 수 있다.
- [0099] 도 6c는 본 발명의 일 실시 예에 따른 객체 추출모듈(240)을 통해 추출된 영상 내 관심 영역 기반 확장 영역의 일 예를 도시한다.
- [0100] 도 6c를 참조하면, 객체 추출모듈(240)은 관심 영역(242)을 중심으로 하여 인접하는 복수의 영상 블록을 확장 영역(243)으로 설정할 수 있다. 이때, 확장 영역(243)은 영상 내 객체의 위치로부터 시작되는 예측 유닛들의 합이 될 수 있다. 이를 위해, 객체 추출모듈(240)은 움직임 벡터(241)의 이동 추세와 유사한 영역 즉, 관심 영역(242)을 우선적으로 선정하여 객체(241)의 유사도를 비교할 수 있으며, 나아가 움직임 벡터(241)의 추세와 유사한 순서로 영상 블록을 확장하여 확장 영역(243)을 설정할 수 있다. 즉, 객체 추적 장치(200)는 관심 영역(242) 및 확장 영역(243)을 활용하여 객체(241)를 추적하는 시간을 감소시킬 수 있다.
- [0102] 상술한 바와 같이, 본 발명은 인공 신경망을 이용하여 실시간으로 입력되는 영상으로부터 객체 트래킹을 고속화할 수 있다.
- [0103] 또한, 본 발명은 인공 신경망을 이용하여 객체 추출 및 객체 추적에 대한 오차를 최소화함으로써, 객체 트래킹 프로세스에 대한 정확도 및 신뢰도를 향상시킬 수 있다.
- [0104] 또한, 본 발명은 객체 트래킹 프로세스를 최적화, 자동화함으로써, 객체 트래킹을 위한 비용과 시간을 최소화할 수 있다.
- [0106] 한편, 본 명세서에 기재된 다양한 실시예들은 하드웨어, 미들웨어, 마이크로코드, 소프트웨어 및/또는 이들의 조합에 의해 구현될 수 있다. 예를 들어, 다양한 실시예들은 하나 이상의 주문형 반도체(ASIC)들, 디지털 신호 프로세서(DSP)들, 디지털 신호 프로세싱 디바이스(DSPD)들, 프로그램어블 논리 디바이스(PLD)들, 필드 프로그램 어블 게이트 어레이(FPGA)들, 프로세서들, 컨트롤러들, 마이크로컨트롤러들, 마이크로프로세서들, 여기서 제시되는 기능들을 수행하도록 설계되는 다른 전자 유닛들 또는 이들의 조합 내에서 구현될 수 있다.
- [0107] 또한, 예를 들어, 다양한 실시예들은 명령들을 포함하는 컴퓨터-판독가능한 매체에 수록되거나 인코딩될 수 있다. 컴퓨터-판독가능한 매체에 수록 또는 인코딩된 명령들은 프로그램 가능한 프로세서 또는 다른 프로세서로 하여금 예컨대, 명령들이 실행될 때 방법을 수행하게끔 할 수 있다. 저장 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수도 있다. 예를 들어, 이러한 컴퓨터-판독가능한 매체는 RAM, ROM, EEPROM, CD-ROM

또는 기타 광학 디스크 저장 매체, 자기 디스크 저장 매체 또는 기타 자기 저장 디바이스, 또는 원하는 프로그램 코드를 컴퓨터에 의해 액세스가능한 명령들 또는 데이터 구조들의 형태로 저장하는데 이용될 수 있는 임의의 다른 매체를 포함할 수 있다.

[0108] 이러한 하드웨어, 소프트웨어, 펌웨어 등은 본 명세서에 기술된 다양한 동작들 및 기능들을 지원하도록 동일한 디바이스 내에서 또는 개별 디바이스들 내에서 구현될 수 있다. 추가적으로, 본 발명에서 "~부"로 기재된 구성요소들, 유닛들, 모듈들, 컴포넌트들 등은 함께 또는 개별적이지만 상호 운용가능한 로직 디바이스들로서 개별적으로 구현될 수 있다. 모듈들, 유닛들 등에 대한 서로 다른 특징들의 묘사는 서로 다른 기능적 실시예들을 강조하기 위해 의도된 것이며, 이들이 개별 하드웨어 또는 소프트웨어 컴포넌트들에 의해 실현되어야만 함을 필수적으로 의미하지 않는다. 오히려, 하나 이상의 모듈들 또는 유닛들과 관련된 기능은 개별 하드웨어 또는 소프트웨어 컴포넌트들에 의해 수행되거나 또는 공통의 또는 개별의 하드웨어 또는 소프트웨어 컴포넌트들 내에 통합될 수 있다.

[0109] 특정한 순서로 동작들이 도면에 도시되어 있지만, 이러한 동작들이 원하는 결과를 달성하기 위해 도시된 특정한 순서, 또는 순차적인 순서로 수행되거나, 또는 모든 도시된 동작이 수행되어야 할 필요가 있는 것으로 이해되지 말아야 한다. 임의의 환경에서는, 멀티태스킹 및 병렬 프로세싱이 유리할 수 있다. 더욱이, 상술한 실시예에서 다양한 구성요소들의 구분은 모든 실시예에서 이러한 구분을 필요로 하는 것으로 이해되어서는 안되며, 기술된 구성요소들이 일반적으로 단일 소프트웨어 제품으로 함께 통합되거나 다수의 소프트웨어 제품으로 패키징될 수 있다는 것이 이해되어야 한다.

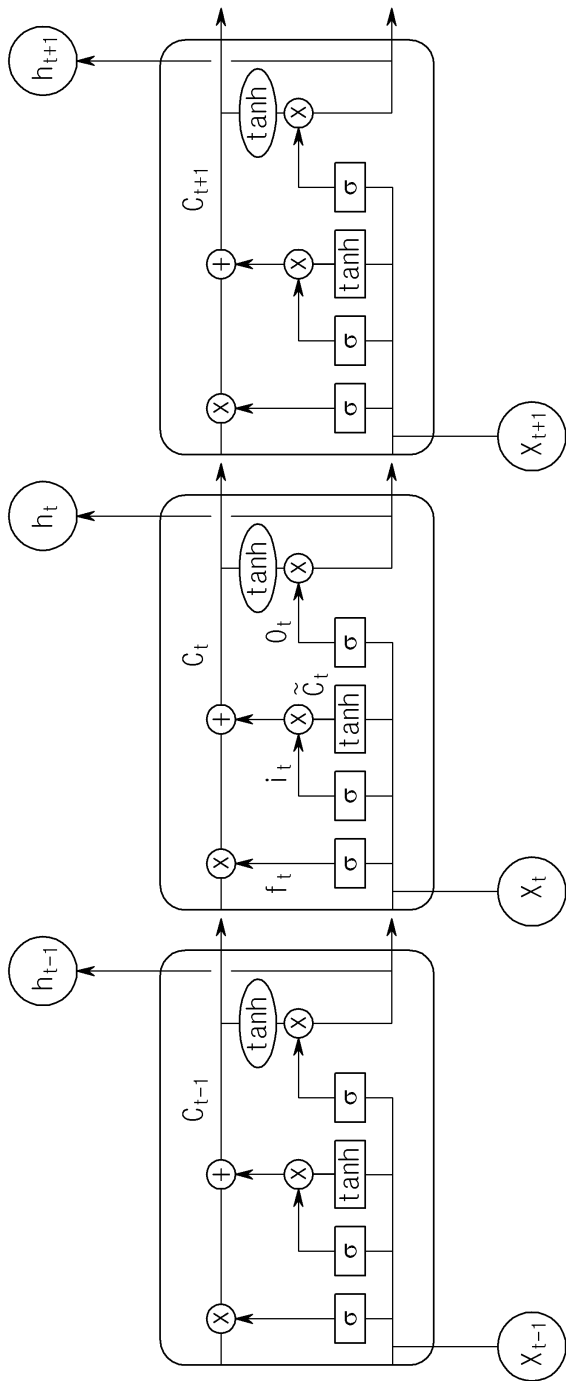
[0110] 이상에서와 같이 도면과 명세서에서 최적 실시예가 개시되었다. 여기서 특정한 용어들이 사용되었으나, 이는 단지 본 발명을 설명하기 위한 목적에서 사용된 것이지 의미한정이나 특허청구범위에 기재된 본 발명의 범위를 제한하기 위하여 사용된 것은 아니다. 그러므로, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0112] 200: 객체 추적 장치 210: 통신모듈
220: 영상 부호화모듈 230: 학습모듈
240: 객체 추출모듈 250: 객체 추적모듈
260: 프로세서 270: 메모리

도면

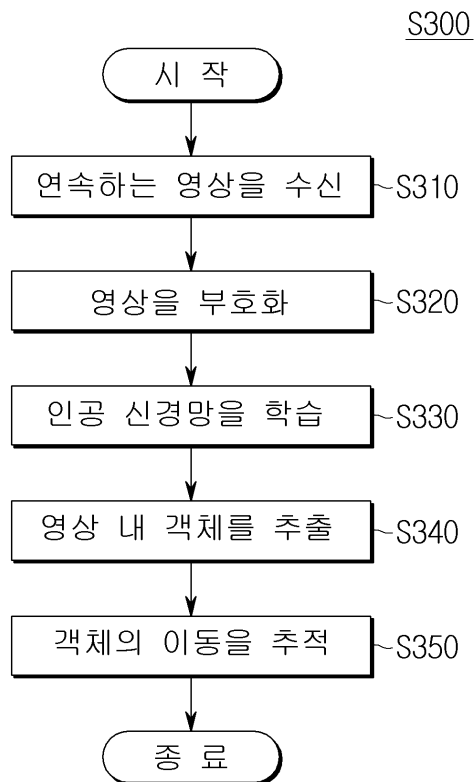
도면1



도면2



도면3

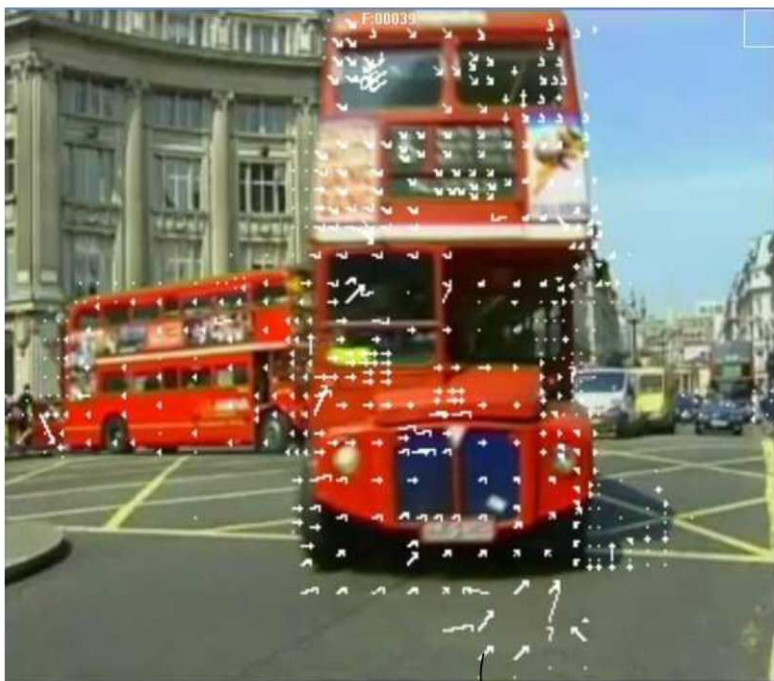


도면4



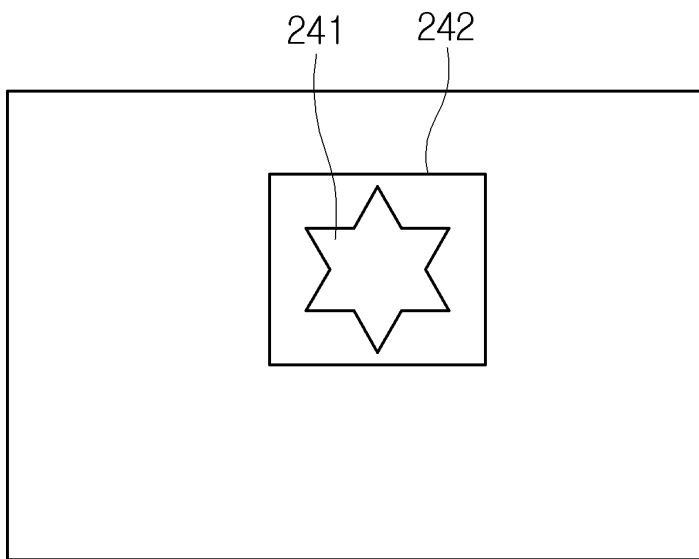
221

도면5

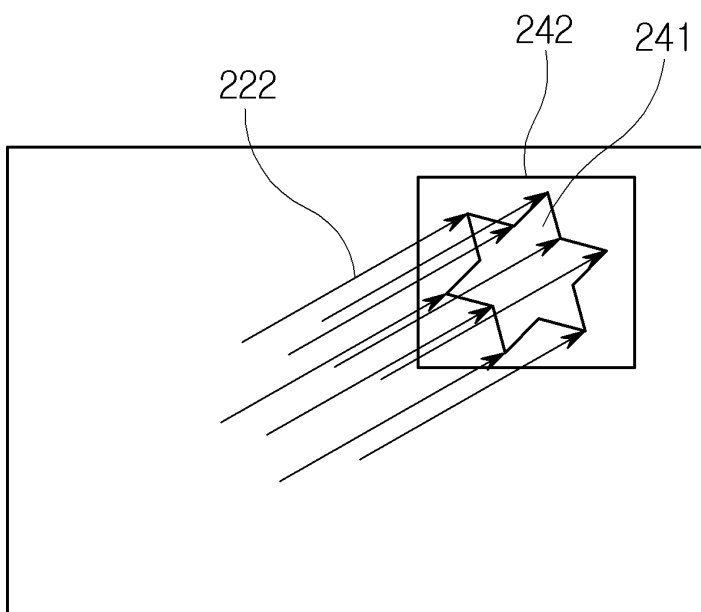


222

도면6a



도면6b



도면6c

