(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0073539
(43) 공개일자 2022년06월03일

(51) 국제특허분류(Int. Cl.)
G06N 20/20 (2019.01) *G06N 3/08* (2006.01)
G06T 7/20 (2017.01)
(52) CPC특허분류
G06N 20/20 (2021.08)
G06N 3/08 (2013.01)
(21) 출원번호 10-2020-0161705
(22) 출원일자 2020년11월26일
심사청구일자 없음

(71) 출원인
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
조민수
경상북도 포항시 남구 청암로 77
유기현
서울특별시 관악구 청룡5길 11, 1106호
김상현
경기도 안양시 동안구 관평로138번길 63, 712동 701호
(74) 대리인
특허법인이상

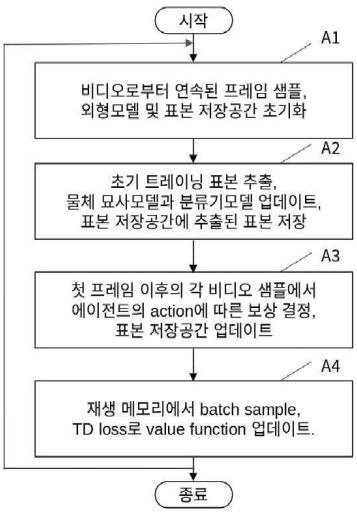
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 **온라인 학습 정책을 위한 강화학습 방법 및 장치**

(57) 요 약

본 발명은 강화학습 방법을 개시한다.

대 표 도 - 도1



(52) CPC특허분류
G06T 7/20 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711116284
과제번호 2017M3C4A7069369
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 차세대정보·컴퓨팅기술개발
연구과제명 비전 모델 기반 공간 상황 인지 원천기술 연구
기 여 율 1/1
과제수행기관명 한양대학교
연구기간 2020.04.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

강화 학습 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 강화학습 방법 및 장치에 관한 것으로, 더욱 상세하게는 영상 추적 모델의 효과적인 온라인 학습 정책을 위한 강화학습 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 영상 추적 문제는 첫 번째 프레임에서 주어진 표적의 정답 경계 박스를 기반으로 이어지는 프레임들에서 표적의 위치를 추정하는 문제이다. 딥 러닝 알고리즘의 발전과 많은 데이터셋의 이용으로 딥 러닝을 이용한 영상추적은 빠르면서도 정확한 영상 추적을 가능하게 하였다.

[0003] 딥 러닝을 이용한 영상 추적은 크게 세가지 부류로 나뉘어진다. 즉, 영상 추적 방법은 샴 네트워크 (Siamese network)로부터 얻어진 첫 번째 프레임의 정답 경계 박스의 정보와 현재 프레임 이미지의 정보의 상호 상관(cross correlation)으로 유사도를 측정하는 샴 네트워크 기반의 방법, 현재 프레임의 이미지에서 표적의 위치를 히트 맵 (heat map) 출력으로 찾는 상관 필터 (correlation filter) 기반의 방법, 이미지 패치 입력을 받아 표적과 백그라운드 클러터 (background clutter) 구분하는 분류기를 이용하는 디텍션 기반 (detection)의 방법으로 분류될 수 있다. 세 가지 방법 모두 온라인 학습을 채용함으로써 표적의 외모나 형태 변화에 강인하게 되어 장기적인 영상 추적에 있어서 큰 정확도 향상을 얻을 수 있다.

[0004] 하지만 온라인 학습은 학습 표본들을 모으고 파라미터를 갱신하기 위한 시간 소모와, 잘못된 학습 표본에 의한 오류 전파의 문제점을 가지고 있다. 이러한 문제점을 시사하고 시간 소모와 정확도를 절충하는 학습을 위해서는 잘 고안된 학습 정책을 세우는 것이 온라인 학습의 효과를 극대화하는 방법이다.

[0005] 대부분의 온라인 학습을이용하는 온라인 영상 추적 모델은잘 못된 표본에 의한 오류 전파를막기 위해 오랫동안 사용되어온 휴리스틱한 학습 방법을 차용하고 있다. 휴리스틱 학습 정책은 학습 표본 추출, 정기적 학습, 비정기적 학습의 세 가지 정책을 이용하여 외모모델을 갱신한다. 현재 프레임에서 추정된 표적의 확실도가 높을 경우 이 표적을 주변으로 학습표본들을 추출하고, 그렇지않을 경우 비정기적 학습으로 외모모델을 학습시킨다. 또한 그와 무관하게 정기적으로 매 특정 프레임마다 모델을 업데이트시킨다. 이 휴리스틱 학습 정책은 중복 프로세스를 발생시킨다.

[0006] 상관 필터의 온라인 학습을 외부의 에이전트를 통해 관리하는시도가 있는데, 외부의 에이전트는 영상 추적 결과를 입력으로 받아 무거운 Q네트워크를 거치기 때문에 해당하는 물체에 대한 정보가 부족하고 효율적이지 못하다.

발명의 내용

해결하려는 과제

[0007] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은 강화학습을 이용한 온라인 학습 정책으로 온라인 학습의 시간 소모의 문제와 잘못된 학습 표본에 의한 오류의 문제를 동시에 해결하는 데 있다.

[0008] 또한, 상기와 같은 문제점을 해결하기 위한 본 발명의 다른 목적은 강화 학습 에이전트가 강화학습에 이용되지 않은 영상 추적 모델에 일반화되어 적용하는 방법을 제공하는 데 있다.

과제의 해결 수단

[0009] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 강화학습 방법은 영상 추적 모델의 효과적인 온라인 학습 정책을 위한 강화학습 방법을 제공할 수 있다.

발명의 효과

[0010] 본 발명의 일 실시예에 따르면 딥러닝과 강화학습을 결합하여 영상 내 물체를 정확하게 추정함으로써 영상 추적 모델을 위한 빠르고 효과적인 온라인 학습 정책을 제공할 수 있다.

도면의 간단한 설명

[0011] 도 1은 본 발명의 일 실시예에 따른 강화 학습 방법의 동작 순서도이다.

도 2는 본 발명의 일 실시예에 따른 강화 학습 방법의 다른 동작 순서도이다.

도 3은 강화학습 에이전트의 학습 표본을 결정하는 방법의 예시도이다.

도 4는 표준화를 통해 정보 공간 안정화하는 방법의 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0012] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0013] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는 데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0014] 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.

[0015] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 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.

[0016] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0018] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0019] 도 1은 본 발명의 일 실시예에 따른 강화 학습 방법의 동작 순서도이다.

[0020] 도 2는 본 발명의 일 실시예에 따른 강화 학습 방법의 다른 동작 순서도이다.

[0021] 도 3은 강화학습 에이전트의 학습 표본을 결정하는 방법의 예시도이다.

[0022] 도 4는 표준화를 통해 정보 공간 안정화하는 방법의 예시도이다.

[0024] 본 발명은 온라인 영상 추적 모델 중 tracking-by-detection 체제의 딥러닝을 이용한 분류기 학습을 이용할 수 있다. 이 때, 해당하는 분류기는 이미지 패치 입력에 대해서 해당하는 패치가 표적인지 아니면 백그라운드 클러터인지 구분할 수 있다. 또한, 본 발명은 여러 딥러닝 알고리즘 중에서도 표적의 클래스(class)나 도메인(domain)에 구애받지 않는 특징을 학습하기 위해, 오프라인 학습으로 여러 도메인 지식을 사용하는 MDNet기반의 영상 추적을 활용할 수 있다. 본 발명은 온라인 추적 모델 중에서MDNet에 대해 적용할 수 있다.

[0026] MDNet기반의 분류기 f 의 구조는 정보 추출 (feature extract)을 하는 세 개의 합성곱 신경망, 물체 묘사(object representation)를 배우는 두 개의완전 연결 신 경 그리고 표적과 백그라운드 클러터를 구분하는 분류기 (classifier)의 세 가지의 구성으로 이루어져 있다.

[0027] 본 발명에서는 세 가지 구성을 각각 f^{FE}, f^{OR}, f^{CL} 로 명명하고 각각의 파라미터 역시 θ^{FE}, θ^{OR}, θ^{CL} 로 정의할 수 있다.

[0028] RGB 이미지 패치 입력 x에 대한 분류기의 출력 $\hat{y} \in R^2$ 은 수학식1과 같이 계산될 수 있다.

수학식 1

[0029] $$\hat{y} = f^{CL}(f^{OR}(f^{FE}(x;\theta^{FE});\theta^{OR});\theta^{CL})$$

[0030] MDNet기반의 영상 추적 모델은 프레임마다 이전 프레임에서 추정된 표적 근처에서 뽑힌 N개의 표적 후보들 $x^1, \ldots, x^N$ 중 가장 표적에 가까운 최적의 표적 x^* 을 찾는 것이 목표이다. 각 표적 후보 x^i 들은 f 에 의해서 positive 점수와 $f^+(x^i)$ 와 negative 점수 $f^-(x^i)$ 를 받게 되고, 최적의 표적은 각 표적 후보들중 positive 점수가 가장 높은 후보를 고르는 것으로 수학식 2와 같이 정해질 수 있다.

수학식 2

[0031] $$x^* = \arg\max_{x^i} f^+(x^i)$$

[0032] MDNet 영상 추적 모델의 학습 과정은 오프라인으로 이루어지는 사전 학습, 첫 번째 프레임에서 표적의 정답 경계 박스로 학습하는 초기 학습, 그리고 이어지는 프레임들에서 추정된 표적으로 학습하는 온라인 학습의 세 가지 과정으로 이루어질 수 있다.

[0033] 학습을 위해서 사용되는 표본은 정답 경계 박스(ground-truth bounding box) 혹은 추정 경계 박스와의 Intersection of Union(IoU) 기준으로 positive 표본 과negative 표본 S^- 으로 나뉠 수 있다.

[0034] 이 때, 학습에 사용되는 손실 함수는 분류기를 학습시킬 때 사용하는 binary cross-entropy가 수학식3과 같이 사용될 수 있다.

수학식 3

[0035] $$L_{ds}(S^+,S^-) = -E_{x^i \in S^+,S^-}[y^i \log(f^+(x^i)) + (1-y^i)\log(1-f^-(x^i))]$$

[0036] 여기서, $x^i \in S^+$ 일 때 $y^i = 1$ 이고 $x^i \in S^-$ 일 때 $y^i = 0$ 이다.

[0037] 사전 학습에서는 추적해야 할 비디오가 주어지기 전 도메인에 구애받지 않는 일반적인 물체의 묘사에 대해 학습하며, 큰 비디오 데이터셋에서 온라인 영상 추적에 사용될 사전 지식을 전체 모델 파라미터 Θ를 갱신함으로써 학습할 수 있다.

[0038] 초기 학습에서는 추적해야 할 비디오에서 첫 번째 프레임의 정답경계박스가 주어지면, 영상추적모델이 해당하는 표적에 적응하기 위해 θ^{α}를 무작위로 초기화시킬 수 있다. 그 다음 정답 경계 박스 주변으로 S^+와 S^-를 샘플링(sampling)한 후, 초기 학습을 통해 주어진 표적에 대해서 θ^{OR}과 θ^{α}을 적응시킬 수 있다.

[0039] 온라인 학습을 수행하는 영상 추적 모델은 파라미터 갱신을 위한 학습표본을 저장공간 M에 저장할 수 있다. M에는 최근 $N_M = 100$ 프레임들에서 추출된 학습 표본들의 feature를 가지고 있으며, 각 학습 표본은 IoU 기준에 의해서 그려지고 프레임t에서 뽑힌 표본들을 S^+, S^-로 정의할 수 있다.

[0040] 온라인 학습은 정기적 적응과 비정기적 적응의 두 가지 정책으로 θ^{OR}과 θ^{α}을 갱신할 수 있다.

[0041] 정기적 적응은 장기적 외모 변화에 강인함을 위해 학습이 되며 매 TR = 10프레임마다 M 에 있는 모든 프레임의 표본들로 학습할 수 있다. 정기적 적응은 어떤 표본들이 쌓이건 계속 일어나기 때문에 따라서 TR 프레임 동안 비슷한 정보를 가진 샘플들이 계속해서 들어온다면 불필요한 과정이 된다.

[0042] 비정기적 적응은 영상 추적이 잘되고 있지 않을 때 표적과의 적응성을 위해 최근 20 프레임들의 표본으로 학습할 수 있다. 하지만 표적의 위치를 놓쳐 계속해서 영상 추적이 되지 않는다면, 모델이 이미 주어진 학습 표본에 적응이 되었음에도 불구하고 새로운 학습표본 없이 비정기적 적응이 계속 일어나게 되어 큰 계산 손해로 이어진질 수 있다.

[0043] 강화학습 문제는 일반적으로 Markov Decision Process (MDP)로 표현이 된다. MDP는 상태 s와 행동 a, 그리고 상태와 행동에 따른 보상 r(s,a)으로 이루어져 있다. 먼저 본 출원에서 제안하는 방법의 상태와 행동으로 온라인 학습 정책을 어떻게 구성하는지 설명하고, 그에 따른 보상과 학습방법에 대해서 설명한다.

[0044] 제안하는 방법의 목표는 강화학습을 사용하여 온라인 학습 정책을 최적화하는 것으로, 제안된 모델을 사용할 때의 영상 추적 상황을 묘사하고 있다.

[0045] 불필요한 온라인 영상 추적 모델의 적응과 학습 표본 축적을 줄이기 위해, 본 출원에서는 학습을 할지 말지, 학습 표본을 모을지 말지 정하는 강화학습 에이전트를 만들었다. 본 발명은 학습 표본 축적의 여부를 결정하는 행동 공간∈ {collect, pass}과 학습의 여부를 결정하는 행동 공간 a_2 ∈ {adapt, pass}를 정의하였고, 제안된 에이전트는 매 프레임 두 가지 행동 공간에서 하나씩 독립 적으로 행동을 취하게 된다.

[0046] 프레임 t에서 취하는 행동 은 두 가지 행동을 이어 붙인 $a_t = \{a_1^t, a_2^t\}$와 같이 나타내어진다.

[0047] 정의된 행동 공간은 두 가지 행동이 모두 pass일 때 불필요한 무거운 계산을 피할 수 있다는 점에서 휴리스틱한 학습 정책과 다르다. 잘 학습된 에이전트는 현재 추적되어진 표적이 이미 축적된 학습 표본의 물체 묘사와 다를 때 collect를 해야 하고, 축적된 학습 표본들을 표적인지 아닌지 구분할 수 없을 때 adapt해야 한다. 그런 점에서 에이전트는 현재 외모 모델의 물체 묘사에 대한 정보가 필요하며, 분류기의 입력으로 사용되는 f^{OR}을 사용

하였다.

[0048] f^{OR} 은 영상 추적에서 이미 계산되어있는 정보이기 때문에 이미지로부터 또 다른 정보 추출없이 효율적인 에이전트를 만들어낼 수 있지만, 다른 비디오 입력이 들어오거나, 외모 모델을 갱신할 경우에 특징 공간(feature space)의 크기와 특성이 변화한다는 단점이 있다.

[0049] 본 발명은 저장된 학습 표본의 통계를 이용하여 현재 추정된 표적을 표준화된 벡터 공간(standardized vector space)으로 보냄으로써 이 문제를 해결하였다.

[0050] 도 2을 참조하면, 저장공간 M 에 있는 positive 클래스와 negative 클래스의 통계로 표준 정규 공간으로 보내어 현재 추정된 표적의 물체 묘사를 해석할 수 있다. 이는 온라인 영상 추적의 경우, 외모 모델의 파라미터가 계속해서 바뀜에 따라 정보 공간의 의미와 크기가 달라지는 점을 해결하기 위함이다.

[0051] 본 발명은 positive 표본들과 negative 표본들의 물체 묘사는 각각 정규분포를 따르고 있다고 가정하고, positive 표본들의 평균을 수학식4를 통해 계산할 수 있다.

수학식 4

[0052] $$\mu_t^+ = E_{F_z^{FE} \in M_t^+}[f^{OR}(F_x^{FE};\theta_t^{OR})]$$

[0053] 또한, 본 발명은 positive 표본들의 분산을 수학식 5를 통해 계산할 수 있다.

수학식 5

[0054] $$\sigma_t^+ = \sqrt{Var_{F_z^{FE} \in M_t^+}(f^{OR}(F^{FE_z};\theta_t^{OR}))}$$

[0055] 한편, negative 표본들의 평균도 상기 수학식 4와 유사한 방법으로 계산될 수 있다. 따라서, negative 표본들의 분산도 상기 수학식 5와 유사한 방식으로 계산될 수 있다.

[0056] positive 표본들의 평균과 표준편차를 이용하여, 현재 추정된 표적에 대한표준화된 positive 벡터를 구한다. 표준화된 positive 벡터도 상기 수학식 6과 유사한 방법으로 계산될 수 있다.

수학식 6

[0057] $$z_t^+ = (F_{x^*}^{OR} - \mu^{+_t})/\sigma^{+_t}$$

[0058] 구해진 표준화된 벡터는 온라인으로 변하는 불안정한 공간을 안정화된 표준 정규 분포 공간으로 보낼 뿐만 아니라, 현재 추정 표적이 두 분포상에서 어느 위치에 있는지 알 수 있다.

[0059] 에이전트는 두 개의 은닉 층을 가진 다층 퍼셉트론(multi-layer perceptron) 딥 Q 네트워 크(deep Q network)으로 이루어져 있고 그 파라미터는 a로 정의하였다. Q 네트워크는 프레임 t에서 입력 상태 st가 주어지고, 가장 큰 Q 값을 가지는 행동들 a_t^* 을 취한다.

[0060] 한편, 입력 상태 st는 수학식7과 같이 두 개의 표준화된 벡터와 행동 히스토리를 덧붙인 벡터가 될 수 있다.

수학식 7

[0061] $$s_t = \{z^+_t, z^-_t, h_t\}$$

[0062] 행동 히스토리 h_t 는 수학식8과 같이 최근 N프레임에서의 에이전트 행동을 모아놓은 정보들을 의미할 수 있다.

수학식 8

[0063] $$h_t = \{a_{t-N_h}, \ldots, a_{t-1}\}$$

[0064] 본 발명에서 제안하는 강화학습 에이전트는 딥 큐러닝(Deep Q-learning)을 사용하여 학습될 수 있다. 딥 큐러닝의 목적은 미래의 받을 보상들의 합을 최대화하는 행동들을 결정하는 정책을 배우는 것이다. 미래의 받을 보상들의 합은 하이퍼 파라미터인 할인율 $\gamma \in (0, 1]$ 와 함께 수학식9와 같이 정의될 수 있다.

수학식 9

[0065] $$G_t = \sum_{k=0}^{\infty} \gamma^k r_{k+t}(s_{k+t}, a_{k+t})$$

[0066] 한편, 이 보상은 종종 가치 함수인 큐 함수로 추론하는 값이 되고, 이 함수는 현재의 보상과 다음 상태의 큐 함수의 합으로 나누어질 수 있다. 이는 수학식 10과 같은 형태로 나타낼 수 있다.

수학식 10

[0067] $$Q(s_t, a_t; \psi) = r_t(s_t, a_t) + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}, \psi)$$

[0068] 수학식 10 형태로 나타내어지는 큐함수를 학습시키기 위해서는 주로 시간차 학습(temporal difference learning)이 이용되며 그 손실 함수는 수학식 11과 같다.

수학식 11

[0069] $$L_t = l_\sigma(q(s_t, a_t; \psi), r_t(s_t, a_t) + \gamma \max_{a'} \bar{q}(s_{t+1}, a'; \bar{\psi}))$$

수학식 12

[0070] $$l_\sigma(x) = \begin{cases} 0.5x^2 & |x| < 1 \text{ 일 때} \\ |x| - 0.5 & \text{이외의 경우} \end{cases}$$

[0071] 한편, 수학식 12형태로 나타내어지는 l_σ 는 Huber loss로 Smooth-L1 loss일 수 있다. 또한, 본 발명은 최근 딥

큐러닝의 알고리즘 발전을 활용하여 에러 기반의 샘플링 기법인 priority experience replay (PER), 과대평가(overestimate)를 막기 위한 더블 큐러닝(double q learning) 가치와 행동에 따른 이점으로 나눈 듀얼 네트워크(Dual network)를 사용하여 에이전트를 학습 시킬 수 있다.

수학식 13

$$r_t(s_t,a_t)=\begin{cases} r'_t-\eta_c*\rho^{N_c} & a_t^1\text{가 collect 일 때} \\ r'_t-\eta_a*\rho^{N_a} & a_t^1\text{가 adapt 일 때} \\ r'_t & \text{이외의 경우} \end{cases}$$

[0072]

[0073] 한편, 수학식 13은 본 발명에서 강화학습 에이전트를 학습하는데 사용 하는 보상함수를 의미할 수 있다. 여기서, 영상 추적이 잘되고 있는지에 대한 판단으로 추정된 표적과 정답 경계 박스와의 IoU기준으로 기본적인 보상 r_0^t를 활용할 수 있다. 또한, r_0^t는 IoU가 0.6이상 일 때 1을, 아닐 때는 0을 할당해주어 에이전트가 해당하는 표적을 잘따라가는 온라인 학습정책을 배우게 할 수 있다.

[0074] 또한, 무거운 계산량이 드는 collect와 adapt에 대해서는 각각 η_c와 η_a는 하이퍼파라미터로 영상 추적의 정확도와 속도를 절충할 수 있도록 해주는 인자일 수 있다. 한편, 연속된 프레임에서의 collect와 adapt를 줄이기 위해서 가중치 ρ를 곱하여 연속된 collect와 adapt를 더 큰 불이익을 줄 수 있다.

[0075] 또한, N_c와 N_a는 행동 히스토리 안에 존재하는 collect와 adapt의 개수를 의미할 수 있다. 강화 학습 에이언트를 학습 시키기 위한 학습 에피소드는 기존 tracking-by-detection의 온라인 학습 과정을 기초로해서 학습을 진행할 수 있다.

[0076] 본 발명의 강화학습의 학습 과정은 에이전트의 초기화(A1), 첫 프레임 표본의 특징 저장 (A2), 첫 프레임 이후의 표본의 저장 및 보상 결정 (A3), value function 업데이트 (A4)로 구성될 수 있다,

[0077] 첫 프레임 이후의 표본의 저장 및 보상 결정(A3)는 각 프레임의 학습을 위한 기본 보상 결정 및 행동 결정 (B1), 행동의 collect 여부 확인 (B2), collect 시 feature 추출과 패널티 추가 (B3), 행동의 adapt 여부 확인 (B4), adapt 시 모델 업데이트와 패널티 추가 (B5), 재생메모리에 기록 (B6)로 구성할 수 있다.

[0078] 비디오 데이터셋 T의 한 비디오로부터 연속된 프레임들을 샘플링하고, 영상 추적 모델 는 기존의 사전 학습된 파라미터로 초기화할 수 있다. 첫번째 프레임에서 정답 경계 박스가 주어지고, 영상 추적 모델은 초기 학습으로 분류기를 학습한다. 이어지는 프레임에서는 에이전트의 행동에 의해서 외모 모델을 갱신하거나 학습 표본을 모은다. 활용(exploitation)과 탐험(exploration)을 위해서 ϵ-greedy 방법을 채용할 수 있다. 행동에 의한 상태변환 정보를 재생 메모리에 저장하고, 비디오가 끝나면 재생 메모리에서 배치(batch)를 생플링해 에이전트를 학습할 수 있다.

[0080] 본 발명의 일 실시예에 따르면, 강화학습 에이전트가 영상 추적 과정에서 어떤 프레임에서 데이터를 모으고, 언제 학습을 해야 할지 결정함으로써 효과적인 학습정책을 제공한다. 또한 본 발명의 온라인 영상 추적 모델은 MDNet일 수 있다.

[0081] 본 발명에 있어서, 강화학습 에이전트는 (학습 표본을 추적하거나 학습여부를 결정하는) 두개의 은닉층을 가진 다층 퍼셉트론(multi-layer perceptron) 딥 Q 네트워크(deep Q network)일 수 있다.

[0082] 본 발명에 있어서 강화학습 에이전트는 표적의 외모 변화를 감지하고 현재 외모 모델의 적응성을 감지하여 학습 표본을 모으고 파라미터를 갱신시킬 수 있다.

[0083] 본 발명에 있어서, 온라인추적모델의 은닉층들의 파라미터 갱신과정에서 생기는 출력의 크기와 정보공간(feature space)이 불안정한 문제에서, 출력을 학습 표본의 통계를 이용해 표준 정규 분포로 변환함으로써 현재

추정 표적에 대한 확률적인 해석을 이용할 수 있다.

[0084] 본 발명은 우리의 온라인 학습 정책은 불필요한 과정을 줄여 속도를 늘릴뿐만 아니라, 학습 표본들을 잘 선택함으로써 정확도도 높일 수 있다.

[0085] 본 발명은 표준화를 이용하여 변화하는 정보공간을 안정화시키고, 강화학습에 이용되지 않은 영상 추적 모델로의 일반화를 가능하게 한다.

[0087] 본 발명의 실시예에 따른 방법의 동작은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 프로그램 또는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽혀질 수 있는 정보가 저장되는 모든 종류의 기록장치를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산 방식으로 컴퓨터로 읽을 수 있는 프로그램 또는 코드가 저장되고 실행될 수 있다.

[0088] 또한, 컴퓨터가 읽을 수 있는 기록매체는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 프로그램 명령은 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.

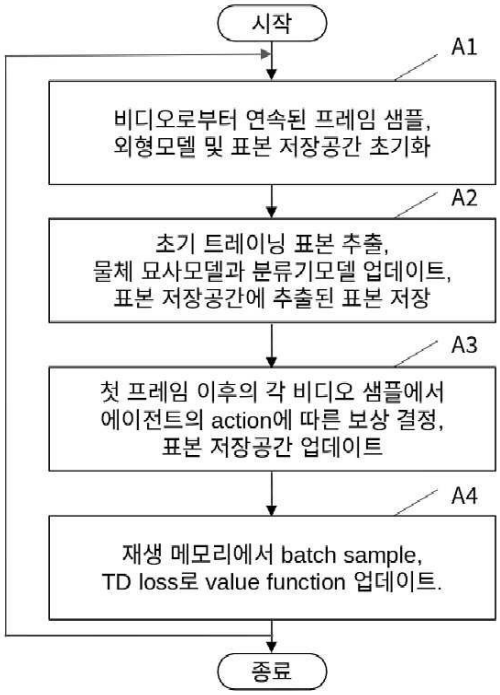
[0089] 본 발명의 일부 측면들은 장치의 문맥에서 설명되었으나, 그것은 상응하는 방법에 따른 설명 또한 나타낼 수 있고, 여기서 블록 또는 장치는 방법 단계 또는 방법 단계의 특징에 상응한다. 유사하게, 방법의 문맥에서 설명된 측면들은 또한 상응하는 블록 또는 아이템 또는 상응하는 장치의 특징으로 나타낼 수 있다. 방법 단계들의 몇몇 또는 전부는 예를 들어, 마이크로프로세서, 프로그램 가능한 컴퓨터 또는 전자 회로와 같은 하드웨어 장치에 의해(또는 이용하여) 수행될 수 있다. 몇몇의 실시예에서, 가장 중요한 방법 단계들의 하나 이상은 이와 같은 장치에 의해 수행될 수 있다.

[0090] 실시예들에서, 프로그램 가능한 로직 장치(예를 들어, 필드 프로그래머블 게이트 어레이)가 여기서 설명된 방법들의 기능의 일부 또는 전부를 수행하기 위해 사용될 수 있다. 실시예들에서, 필드 프로그머블 게이트 어레이는 여기서 설명된 방법들 중 하나를 수행하기 위한 마이크로프로세서와 함께 작동할 수 있다. 일반적으로, 방법들은 어떤 하드웨어 장치에 의해 수행되는 것이 바람직하다.

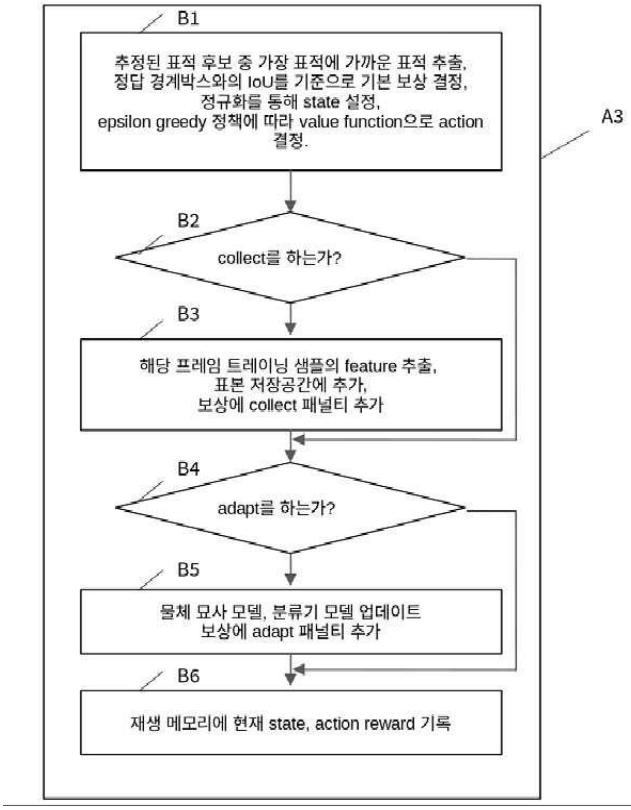
[0091] 이상 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

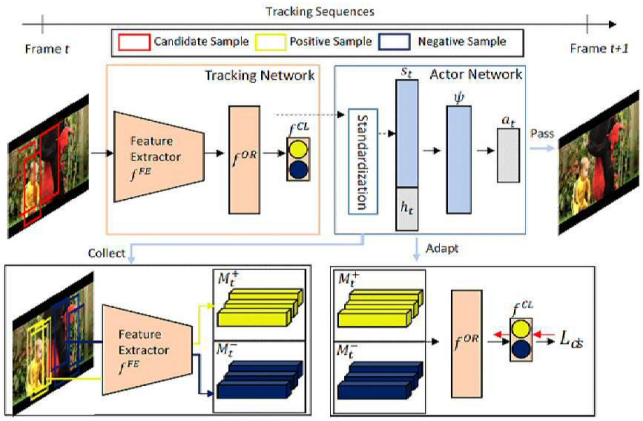
도면1


시작
A1
비디오로부터 연속된 프레임 샘플,
외형모델 및 표본 저장공간 초기화
A2
초기 트레이닝 표본 추출,
물체 묘사모델과 분류기모델 업데이트,
표본 저장공간에 추출된 표본 저장
A3
첫 프레임 이후의 각 비디오 샘플에서
에이전트의 action에 따른 보상 결정,
표본 저장공간 업데이트
A4
재생 메모리에서 batch sample,
TD loss로 value function 업데이트.
종료

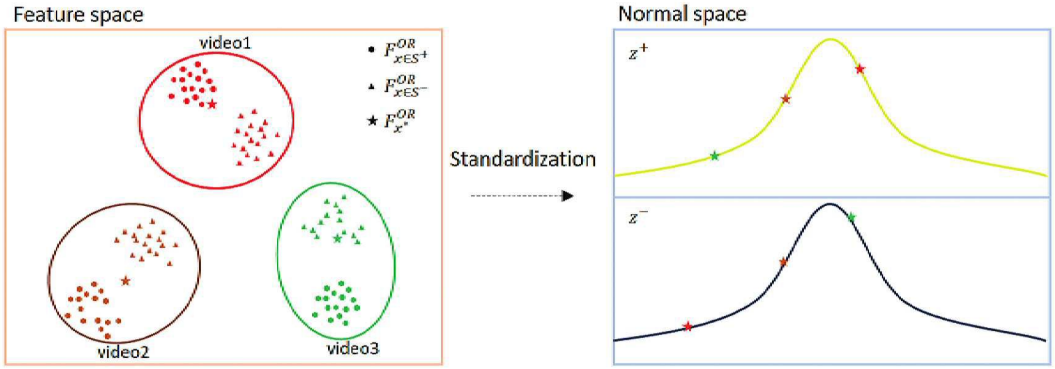
도면2


B1
추정된 표적 후보 중 가장 표적에 가까운 표적 추출,
정답 경계박스와의 IoU를 기준으로 기본 보상 결정,
정규화를 통해 state 설정,
epsilon greedy 정책에 따라 value function으로 action
결정.
A3
B2
collect를 하는가?
B3
해당 프레임 트레이닝 샘플의 feature 추출,
표본 저장공간에 추가,
보상에 collect 패널티 추가
B4
adapt를 하는가?
B5
물체 묘사 모델, 분류기 모델 업데이트
보상에 adapt 패널티 추가
B6
재생 메모리에 현재 state, action reward 기록

도면3


Tracking Sequences
Candidate Sample
Positive Sample
Negative Sample
Frame t
Frame t+1
Tracking Network
Feature Extractor
Actor Network
Standardization
Pass
Collect
Adapt
Feature Extractor

도면4


Feature space
video1
video2
video3
Standardization
Normal space