(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0067138
(43) 공개일자 2022년05월24일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2022.01) *G06V 10/46* (2022.01)
(52) CPC특허분류
G06V 20/46 (2022.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2020-0153515
(22) 출원일자 2020년11월17일
심사청구일자 2020년11월17일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
변혜란
서울특별시 서대문구 연세로 50, 연세대학교 제4공학관 810호 (신촌동)
이제욱
서울특별시 서대문구 연세로 50, 연세대학교 제4공학관 810호 (신촌동)
(뒷면에 계속)
(74) 대리인
특허법인(유한)아이시스

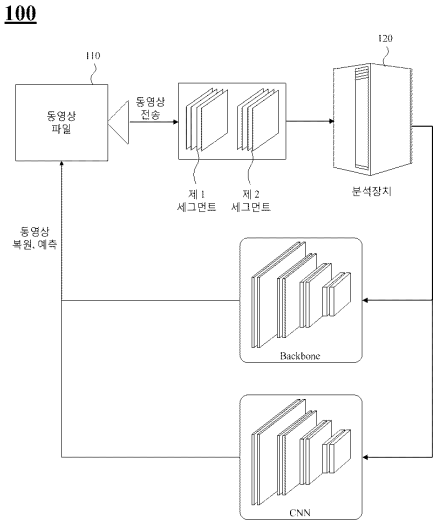
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **동영상 특징 추출 방법 및 장치**

(57) 요 약

개시된 기술은 동영상 특징 추출 방법 및 장치에 관한 것으로, 분석장치가 동영상을 수신하고 상기 동영상의 복수의 프레임들 중 일부의 프레임들을 포함하는 제 1 세그먼트를 특징 추출기에 입력하여 제 1 특징벡터를 출력하는 단계; 상기 분석장치가 상기 일부의 프레임들 후의 일부 프레임들을 포함하는 제 2 세그먼트를 상기 특징 추출기에 입력하여 제 2 특징벡터를 출력하는 단계; 상기 분석장치가 상기 제 1 특징벡터 및 상기 제 2 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터(Reference Label Data)를 생성하는 단계; 상기 분석장치가 상기 제 1 특징벡터를 합성곱 신경망에 입력하여 상기 제 1 세그먼트에 대한 예측 데이터를 출력하는 단계; 및 상기 분석장치가 상기 예측 데이터가 상기 레퍼런스 라벨 데이터와 유사해지도록 상기 특징 추출기를 학습하는 단계;를 포함한다. 따라서 라벨값을 생성하는 코스트를 방지하고 다음 세그먼트에 대한 예측 성능을 향상시키는 효과가 있다.

대 표 도 - 도1



(52) CPC특허분류
G06V 10/469 (2022.01)
(72) 발명자
이필현
서울특별시 서대문구 연세로 50, 연세대학교 제4공학관 810호 (신촌동)
홍기범
서울특별시 서대문구 연세로 50, 연세대학교 제4공학관 810호 (신촌동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711116905
과제번호	2020-0-01361-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원(한국연구재단부설)
연구사업명	정보통신방송연구개발사업
연구과제명	인공지능대학원지원사업
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

분석장치가 동영상을 수신하고 상기 동영상의 복수의 프레임들 중 일부의 프레임들을 포함하는 제 1 세그먼트를 특징 추출기에 입력하여 제 1 특징벡터를 출력하는 단계;

상기 분석장치가 상기 일부의 프레임들 후의 일부 프레임들을 포함하는 제 2 세그먼트를 상기 특징 추출기에 입력하여 제 2 특징벡터를 출력하는 단계;

상기 분석장치가 상기 제 1 특징벡터 및 상기 제 2 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터(Reference Label Data)를 생성하는 단계;

상기 분석장치가 상기 제 1 특징벡터를 합성곱 신경망에 입력하여 상기 제 1 세그먼트에 대한 예측 데이터를 출력하는 단계; 및

상기 분석장치가 상기 예측 데이터가 상기 레퍼런스 라벨 데이터와 유사해지도록 상기 특징 추출기를 학습하는 단계;를 포함하는 동영상 특징 추출 방법.

청구항 2

제 1 항에 있어서,

상기 제 1 세그먼트는 상기 분석장치에 입력되는 상기 동영상의 현재 세그먼트이고 상기 제 2 세그먼트는 상기 제 1 세그먼트에 연속하는 다음 세그먼트인 것을 특징으로 하는 동영상 특징 추출 방법.

청구항 3

제 1 항에 있어서,

상기 제 1 특징벡터 및 상기 제 2 특징벡터는 세그먼트의 너비(height), 높이(width) 및 시간(Time)*채널(channel)을 각각의 축으로 하는 3차원 데이터인 동영상 특징 추출 방법.

청구항 4

제 1 항에 있어서, 상기 레퍼런스 라벨 데이터를 생성하는 단계는,

상기 분석장치가 상기 제 1 특징벡터와 상기 제 2 특징벡터 사이의 변화량 차이를 계산하는 단계;

상기 분석장치가 상기 변화량 차이를 시간*채널 축으로 정규화하는 단계; 및

상기 분석장치가 상기 정규화 된 변화량 차이의 크기에 따라 점수를 매기는 단계;를 포함하는 동영상 특징 추출 방법.

청구항 5

제 1 항에 있어서,

상기 레퍼런스 라벨 데이터는 상기 제 1 특징벡터 및 상기 제 2 특징벡터 각각의 프레임 영역 단위의 변화량 차이의 정도를 나타내는 값을 포함하는 동영상 특징 추출 방법.

청구항 6

제 1 항에 있어서, 상기 특징 추출기를 학습하는 단계는,

상기 분석장치가 크로스 엔트로피(Cross Entropy) 손실함수를 이용하여 상기 특징 추출기의 매개변수 및 상기 합성곱 신경망의 매개변수를 조정하는 동영상 특징 추출 방법.

청구항 7

동영상을 입력받아 복수의 세그먼트들로 분할하는 입력장치;

상기 복수의 세그먼트들을 입력받아 복수의 특징벡터들을 출력하는 특징 추출기 및 상기 복수의 특징벡터들 중 하나를 입력받아 다음 세그먼트에 대한 예측 데이터를 출력하는 합성곱 신경망을 저장하는 저장장치; 및

상기 특징 추출기에 상기 복수의 세그먼트들 중 제 1 세그먼트를 입력하여 제 1 특징벡터를 출력하고 제 2 세그먼트를 입력하여 제 2 특징벡터를 출력하고, 상기 제 1 특징벡터 및 상기 제 2 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터를 생성하고, 상기 합성곱 신경망에 상기 제 1 특징벡터를 입력하여 출력된 예측 데이터와 상기 레퍼런스 라벨 데이터를 비교하는 계산장치;를 포함하는 동영상 특징 추출 장치.

청구항 8

제 7 항에 있어서,

상기 제 1 세그먼트는 상기 동영상의 현재 세그먼트이고 상기 제 2 세그먼트는 상기 제 1 세그먼트에 연속하는 다음 세그먼트인 것을 특징으로 하는 동영상 특징 추출 장치.

청구항 9

제 7 항에 있어서,

상기 제 1 특징벡터 및 상기 제 2 특징벡터는 세그먼트의 너비(height), 높이(width) 및 시간(Time)*채널(channel)을 각각의 축으로 하는 3차원 데이터인 동영상 특징 추출 장치.

청구항 10

제 7 항에 있어서,

상기 분석장치는 상기 제 1 특징벡터와 상기 제 2 특징벡터 사이의 변화량 차이를 계산하고, 상기 변화량 차이를 시간*채널 축으로 정규화하고, 상기 정규화 된 변화량 차이의 크기에 따라 점수를 매겨서 상기 레퍼런스 라벨 데이터를 생성하는 동영상 특징 추출 장치.

청구항 11

제 7 항에 있어서,

상기 레퍼런스 라벨 데이터는 상기 제 1 특징벡터 및 상기 제 2 특징벡터 각각의 프레임 영역 단위의 변화량 차이의 정도를 나타내는 값을 포함하는 동영상 특징 추출 장치.

청구항 12

제 7 항에 있어서,

상기 계산장치는 크로스 엔트로피(Cross Entropy) 손실함수를 이용하여 상기 특징 추출기의 매개변수 및 상기 합성곱 신경망의 매개변수를 조정하는 동영상 특징 추출 장치.

발명의 설명

기 술 분 야

[0001] 개시된 기술은 자기지도학습(Self Supervised Learning) 기반의 동영상 특징 추출 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 딥러닝 기반의 영상 분석은 합성곱 신경망(CNN)을 기반으로 하는 특징 추출기를 통해 이미지 또는 동영상을 인공지능이 이해할 수 있는 형태인 벡터나 행렬로 변환하는 것을 기반으로 하고 있다. 따라서, 분류, 추적, 생성, 변환 등의 다양한 목적을 위한 인공지능 모델에서 CNN 특징 추출기는 핵심 모듈로 배치되고 있다.

[0003] 이러한 CNN 특징 추출기의 표현력이 향상될 경우, 딥러닝 기반 컴퓨터 비전의 모든 분야에서 성능 향상을 야기할 수 있다. 종래에는 라벨값이 존재하는 매우 큰 용량의 데이터셋을 활용하여 CNN을 미리 학습시키고 이미지나 동영상을 분류하는데 이용하였다. 그러나 특징 추출기의 표현력을 충분한 수준으로 향상시키기 위해서는 라벨값

을 만들기 위한 시간적 금전적 비용이 매우 크게 발생하는 문제가 있었다. 특히. 동영상의 경우에는 매우 큰 비용이 발생하기 때문에 라벨값을 자체적으로 생성하여 특징 추출기의 표현력을 향상시키는 기술이 요구된다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 미국 공개특허 US2019-0228313호

발명의 내용

해결하려는 과제

[0005] 개시된 기술은 자기지도학습(Self Supervised Learning) 기반의 동영상 특징 추출 방법 및 장치를 제공하는데 있다.

과제의 해결 수단

[0006] 상기의 기술적 과제를 이루기 위하여 개시된 기술의 제 1 측면은 분석장치가 동영상을 수신하고 상기 동영상의 복수의 프레임들 중 일부의 프레임들을 포함하는 제 1 세그먼트를 특징 추출기에 입력하여 제 1 특징벡터를 출력하는 단계, 상기 분석장치가 상기 일부의 프레임들 후의 일부 프레임들을 포함하는 제 2 세그먼트를 상기 특징 추출기에 입력하여 제 2 특징벡터를 출력하는 단계, 상기 분석장치가 상기 제 1 특징벡터 및 상기 제 2 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터(Reference Label Data)를 생성하는 단계, 상기 분석장치가 상기 제 1 특징벡터를 합성곱 신경망에 입력하여 상기 제 1 세그먼트에 대한 예측 데이터를 출력하는 단계 및 상기 분석장치가 상기 예측 데이터가 상기 레퍼런스 라벨 데이터와 유사해지도록 상기 특징 추출기를 학습하는 단계를 포함하는 동영상 특징 추출 방법을 제공하는데 있다.

[0007] 상기의 기술적 과제를 이루기 위하여 개시된 기술의 제 2 측면은 동영상을 입력받아 복수의 세그먼트들로 분할하는 입력장치, 상기 복수의 세그먼트들을 입력받아 복수의 특징벡터들을 출력하는 특징 추출기 및 상기 복수의 특징벡터들 중 하나를 입력받아 다음 세그먼트에 대한 예측 데이터를 출력하는 합성곱 신경망을 저장하는 저장장치 및 상기 특징 추출기에 상기 복수의 세그먼트들 중 제 1 세그먼트를 입력하여 제 1 특징벡터를 출력하고 제 2 세그먼트를 입력하여 제 2 특징벡터를 출력하고, 상기 제 1 특징벡터 및 상기 제 2 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터를 생성하고, 상기 합성곱 신경망에 상기 제 1 특징벡터를 입력하여 출력된 예측 데이터와 상기 레퍼런스 라벨 데이터를 비교하는 계산장치를 포함하는 동영상 특징 추출 장치를 제공하는데 있다.

발명의 효과

[0008] 개시된 기술의 실시 예들은 다음의 장점들을 포함하는 효과를 가질 수 있다. 다만, 개시된 기술의 실시 예들이 이를 전부 포함하여야 한다는 의미는 아니므로, 개시된 기술의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.

[0009] 개시된 기술의 일 실시예에 따르면 동영상 특징 추출 방법 및 장치는 동영상의 라벨값을 자체적으로 생성하여 라벨링 수행에 따른 코스트를 방지하는 효과가 있다.

[0010] 또한, 특징 추출기의 동영상 표현에 대한 성능을 향상시키는 효과가 있다.

[0011] 또한, 다음 세그먼트에 대한 예측 정확도를 높여서 동영상을 복원하는 효과가 있다.

도면의 간단한 설명

[0012] 도 1은 개시된 기술의 일 실시예에 따른 동영상 특징 추출 과정을 나타낸 도면이다.

도 2는 개시된 기술의 일 실시예에 따른 동영상 특징 추출 방법을 나타낸 도면이다.

도 3은 개시된 기술의 일 실시예에 따른 분석장치를 나타낸 도면이다.

도 4는 개시된 기술의 일 실시예에 따라 레퍼런스 라벨 데이터와 예측 데이터를 비교하는 신경망의 구조를 나타

낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0013] 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.

[0014] 제 1, 제 2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제 1 구성요소는 제 2 구성요소로 명명될 수 있고, 유사하게 제 2 구성요소도 제 1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0015] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 한다. 그리고 "포함한다" 등의 용어는 설시된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.

[0016] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다.

[0017] 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다. 따라서, 본 명세서를 통해 설명되는 각 구성부들의 존재 여부는 기능적으로 해석되어야 할 것이다.

[0018] 도 1은 개시된 기술의 일 실시예에 따른 동영상 특징 추출 과정을 나타낸 도면이다. 도 1을 참조하면 분석장치(120)는 동영상 파일을 수신하여 이를 다수의 세그먼트로 분할하고, 각 세그먼트를 이용하여 CNN을 학습시킬 수 있다. 분석장치(120)는 동영상을 입력받는 입력장치, 백본과 CNN을 저장하는 저장장치 및 네트워크를 이용하여 벡터를 계산하는 계산장치를 포함한다.

[0019] 동영상은 단말기에서 재생 가능한 파일 형태일 수 있다. 동영상 파일은 저장공간의 용량을 고려하여 압축된 파일 형태일 수 있으며 동영상 플레이어(110) 내지는 저장장치에 저장될 수 있다. 동영상 파일은 동영상 플레이어(110)에서 MP4, AC3와 같은 동영상 코덱을 이용하여 압축을 해제하고 분석장치(120)로 전송될 수 있다. 여기에서 동영상은 별도의 라벨값이 입력되지 않은 상태의 로우데이터를 의미한다.

[0020] 한편, 동영상은 영상처리 장치에서 효율적으로 처리하기 위해 복수개의 프레임들 중 일부의 프레임을 하나의 단위로 묶어서 처리할 수 있다. 이러한 단위를 세그먼트라고 한다. 예컨대, 연속하는 8개의 프레임 또는 16개의 프레임을 하나의 세그먼트로 묶을 수 있다. 영상처리 장치는 입력장치를 통해 입력된 동영상을 복수개의 세그먼트로 분류할 수 있다.

[0021] 한편, 분석장치(120)는 두 개의 네트워크를 저장한다. 하나는 레퍼런스 라벨 데이터를 출력하는 백본(Backbone) 네트워크이고 다른 하나는 다음 세그먼트에 대한 예측 데이터를 출력하는 합성곱 신경망(CNN)이다. 백본 네트워크는 동영상 분류 작업에 이용하는 특징 추출기를 의미하며 동영상의 연속하는 두 개의 세그먼트를 입력값으로 하여 레퍼런스 라벨 데이터를 출력한다.

[0022] 종래의 동영상 특징 추출을 위한 학습 방법으로 동영상 분류를 통해 특징 추출기의 표현력을 동시에 학습하는 방법이 주류로 이용되었다. 이는 최대한 많은 동영상을 사용하기 위하여 가장 레이블을 확보하는데 드는 비용이 적은 분류작업을 통해 딥러닝 모델을 학습시키기 위함이다. 그러나 이미지와는 달리 동영상은 시간축이 데이터에 추가된 형태이므로 데이터의 다양성이 이미지와는 비교할 수 없을 만큼 커지기 때문에 더 많은 데이터를 활용하기 위해서는 라벨값 입력을 요구하지 않는 자기지도방식의 표현력 학습 방법을 이용하는 것이 보편화된 방법으로 이용되었다.

[0023] 그러나 자기주도학습 방법을 이용하더라도 시간 또는 공간적 특징만 이해하도록 학습을 수행하거나 난이도의 일

관성이 유지되지 않아 학습과정에서 충분한 표현력을 학습할 수 없는 문제가 추가적으로 발생되었다. 따라서, 이러한 문제점을 해결하기 위해서 시공간 특성을 동시에 학습하도록 분석장치(120)의 구조를 설계할 필요가 있었다.

[0024] 상술한 바와 같이 시공간 특성을 동시에 학습하기 위해서 우선 분석장치(120)의 백본 네트워크를 이용하여 레퍼런스 라벨 데이터를 생성한다. 일 실시예로, 분석장치가 백본 네트워크에 두 개의 세그먼트를 입력하여 각각에 대한 특징벡터를 출력할 수 있다. 여기에서 먼저 입력되는 세그먼트는 현재 세그먼트이고 다음에 입력되는 세그먼트는 현재 세그먼트에 연속하는 세그먼트를 의미한다. 이하부터는 현재 세그먼트를 제 1 세그먼트라고 하고, 다음 세그먼트를 제 2 세그먼트라고 한다. 백본 네트워크는 제 1 세그먼트를 입력받아 제 1 세그먼트의 특징을 나타내는 제 1 특징벡터를 출력하고, 제 2 세그먼트를 입력받아 제 2 특징벡터를 출력할 수 있다.

[0025] 한편, 분석장치(120)는 이와 같이 출력된 2개의 특징벡터의 차이값을 계산할 수 있다. 즉, 특징벡터 수준에서의 변화량을 계산할 수 있다. 제 1 특징벡터 및 제 2 특징벡터는 각각 너비(height), 높이(width) 및 채널(channel)을 포함한다. 두 특징벡터의 차이값을 계산하기 위해서 분석장치(120)는 두 특징벡터를 정규화(Normalization)할 수 있다. 예컨대, 너비(height), 높이(width) 및 시간(Time)*채널(channel)을 갖는 3차원 특징벡터를 시간과 채널을 곱한 값을 축으로 하여 2차원 특징벡터로 변환할 수 있다. 이러한 과정에 따라 동영상에서 지역별(Spatial) 표현을 살리는 것이 가능하다.

[0026] 이와 같이 변환된 특징벡터는 일종의 매트릭스와 같은 형태가 되며 두 개의 매트릭스의 각 요소들의 변화량을 계산할 수 있다. 특징벡터의 변화량은 제 1 특징벡터 및 제 2 특징벡터 각각의 프레임 영역 단위의 변화량 차이를 의미한다. 변화량은 큰 값일 수도 있고 작은 값일 수도 있다. 즉, 2개의 특징벡터의 차이값을 계산하여 변화량 차이의 정도를 나타내는 값을 레퍼런스 라벨 데이터로 생성한다. 변화량 정도의 차이는 스코어링을 수행하는 것으로 나타낼 수 있다. 분석장치는 변화량이 큰 순서대로 순위를 매길 수 있다. 이러한 과정에 따라 백본 네트워크는 영성분석 장치에서 사용할 레퍼런스 라벨 데이터를 생성할 수 있다.

[0027] 한편, 합성곱 신경망(CNN)은 세그먼트에 대한 특징벡터를 입력받아 다음 세그먼트를 예측한다. 일 실시예로 제 1 세그먼트에 대한 제 1 특징벡터를 입력받아 제 1 세그먼트의 다음을 예측하는 데이터를 출력할 수 있다. 여기에서 예측 데이터는 앞서 백본 네트워크가 생성한 바와 같이 2차원 매트릭스의 형태로 생성될 수 있다. 예컨대, 가로 축은 너비이고 세로 축은 높이인 매트릭스일 수 있다. 분석장치는 각 네트워크에서 출력된 데이터를 비교하여 합성곱 신경망을 트레이닝할 수 있다.

[0028] 한편, 네트워크 전반의 예측 정확도를 개선하기 위해서 분석장치는 합성곱 신경망에서 출력된 예측 데이터가 백본에서 출력된 레퍼런스 라벨 데이터와 유사해지도록 학습할 수 있다. 분석장치는 크로스 엔트로피(Cross Entropy) 손실함수를 이용하여 합성곱 신경망이 출력하는 예측 데이터가 레퍼런스 라벨 데이터와 유사해지도록 학습할 수 있다. 예컨대, 크로스 엔트로피 손실함수를 참조하여 백본 네트워크의 매개변수와 합성곱 신경망의 매개변수를 각각 조정할 수 있다. 이러한 과정을 반복하여 동영상에 대한 예측 정확도를 개선할 수 있다.

[0029] 도 2는 개시된 기술의 일 실시예에 따른 동영상 특징 추출 방법을 나타낸 도면이다. 도 2를 참조하면 동영상 특징 추출 방법은 제 1 세그먼트에 대한 제 1 특징벡터를 출력하는 단계(210), 제 2 세그먼트에 대한 제 2 특징벡터를 출력하는 단계(220), 두 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터를 생성하는 단계(230), 제 1 특징벡터를 토대로 제 1 세그먼트에 대한 예측 데이터를 생성하는 단계(240) 및 예측 데이터가 레퍼런스 라벨 데이터와 유사해지도록 학습하는 단계(250)를 포함한다.

[0030] 각 단계에 대한 설명에 앞서 동영상 특징 추출 방법은 컴퓨터에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다. 상기 프로그램은 일시적 또는 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.

[0031] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM(read-only memory), PROM (programmable read only memory), EPROM(Erasable PROM, EPROM) 또는 EEPROM(Electrically EPROM) 또는 플래시 메모리 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다

[0032] 일시적 판독 가능 매체는 스태틱 램(Static RAM, SRAM), 다이내믹 램(Dynamic RAM, DRAM), 싱크로너스 디램(Synchronous DRAM, SDRAM), 2배속 SDRAM(Double Data Rate SDRAM, DDR SDRAM), 증강형 SDRAM(Enhanced SDRAM

, ESDRAM), 동기화 DRAM(Synclink DRAM, SLDRAM) 및 직접 램버스 램(Direct Rambus RAM, DRRAM) 과 같은 다양한 RAM을 의미한다.

[0033] 210 단계에서 분석장치는 동영상을 수신하여 일부의 프레임끼리 묶음으로써 복수의 세그먼트를 생성한다. 그리고 복수의 세그먼트를 특징 추출기에 입력하여 복수의 세그먼트의 숫자만큼 특징벡터를 출력한다. 일 실시예로, 제 1 세그먼트를 특징 추출기에 입력하여 제 1 특징벡터를 출력할 수 있다. 여기에서 제 1 세그먼트는 동영상의 현재 세그먼트를 의미한다.

[0034] 220 단계에서 분석장치는 제 2 세그먼트를 상기 특징 추출기에 입력하여 제 2 특징벡터를 출력할 수 있다. 제 2 세그먼트는 제 1 세그먼트 이후의 세그먼트를 의미한다. 제 2 세그먼트는 제 1 세그먼트의 바로 다음에 연속되는 세그먼트일 수도 있고 이보다 더 뒤에 출력되는 세그먼트일 수도 있다. 두 개의 세그먼트는 분석장치의 특징 추출기인 백본 네트워크로 입력되며, 하나씩 순차적으로 입력될 수도 있고 두 세그먼트가 동시에 입력될 수도 있다. 특징 추출기는 입력된 세그먼트의 특징을 벡터값으로 출력할 수 있다.

[0035] 230 단계에서 분석장치는 제 1 특징벡터와 제 2 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터를 생성한다. 분석장치의 특징 추출기가 3차원 데이터인 특징벡터를 변환하여 2차원 데이터인 레퍼런스 라벨 데이터를 생성할 수 있다. 일 실시예로, 분석장치가 제 1 특징벡터와 제 2 특징벡터 사이의 변화량 차이를 계산하는 단계, 분석장치가 변화량 차이를 시간*채널 축으로 정규화하는 단계 및 정규화 된 변화량 차이의 크기에 따라 점수를 매기는 단계를 수행하여 레퍼런스 라벨 데이터를 생성할 수 있다.

[0036] 즉, 시간*채널을 축으로 두 특징벡터의 차이를 정규화하여 2차원 매트릭스 형태로 변환하고 변환된 매트릭스 내 프레임에 대응하는 값의 크기가 크면 높은 점수를 매기고 크기가 작으면 낮은 점수를 매기는 스코어링(Scoring) 내지는 넘버링(Numbering)을 수행함으로써 레퍼런스 라벨 데이터를 생성할 수 있다. 이와 같이 생성된 레퍼런스 라벨 데이터는 실제 관리자로부터 입력되는 라벨링 데이터와는 다소 다르게 분석장치가 자체적으로 생성한 의사 GT(Pseudo Ground Truth) 데이터를 의미한다. 분석장치는 제 1 특징벡터와 제 2 특징벡터 사이의 변화량이 큰 순서대로 스코어링을 수행하여 의사 GT(Pseudo Ground Truth) 데이터를 생성할 수 있다. 그리고 이 데이터를 레퍼런스 라벨 데이터로 이용할 수 있다.

[0037] 240 단계에서 분석장치는 제 1 특징벡터를 합성곱 신경망에 입력하여 제 1 세그먼트에 대한 예측 데이터를 출력한다. 백본 네트워크와는 별개로 구비되는 합성곱 신경망에 백본에서 출력된 제 1 특징벡터를 입력한다. 합성곱 신경망은 제 1 특징벡터를 입력값으로 하여 제1 세그먼트에 대한 예측 결과를 출력한다. 예측 데이터는 제 1 세그먼트가 이후 어떻게 변화할 것인지를 예측한 데이터를 의미한다. 즉, 영상의 뒷부분이 유실되거나 손상되더라도 예측 결과를 토대로 영상을 복원하는 것이 가능하다. 분석장치는 특정 순번의 세그먼트에 대한 특징벡터가 입력되면 그 다음을 예측한 데이터를 출력하는 과정을 영상 끝까지 반복할 수 있다.

[0038] 250 단계에서 분석장치는 예측 데이터가 레퍼런스 라벨 데이터와 유사해지도록 특징 추출기를 학습한다. 예컨대, 레퍼런스 라벨 데이터와 예측 데이터에 대한 확률분포의 차이를 크로스 엔트로피 함수를 이용하여 계산할 수 있다. 분석장치는 크로스 엔트로피(Cross Entropy) 손실함수를 이용하여 특징 추출기의 매개변수를 조정할 수 있으며 마찬가지로 합성곱 신경망이 출력하는 예측 데이터가 레퍼런스 라벨 데이터와 유사해지도록 합성곱 신경망의 매개변수도 조정할 수 있다.

[0039] 도 3은 개시된 기술의 일 실시예에 따른 분석장치를 나타낸 도면이다. 도 3을 참조하면 분석장치(300)는 입력장치(310), 저장장치(320) 및 계산장치(330)를 포함한다.

[0040] 입력장치(310)는 외부에서 전송되는 동영상을 수신하여 세그먼트 단위로 분할할 수 있다. 입력장치(310)는 분석장치(300)에 구비된 입력 인터페이스를 이용할 수 있다. 예컨대, 키보드, 마우스 등의 인터페이스로 구현될 수 있다. 입력장치는 동영상을 수신하면 이를 자동으로 사전에 설정된 단위의 프레임끼리 묶어서 세그먼트로 분류할 수 있다. 예컨대, 동영상 전체를 복수개의 프레임 단위로 분할한 뒤 사전에 설정된 개수만큼 각각 묶어서 복수의 세그먼트로 분류할 수 있다.

[0041] 저장장치(320)는 분석장치(300)에 구비된 HDD, SSD와 같은 메모리로 구현될 수 있다. 저장장치(320)는 영상을 예측하기 위해서 2개의 네트워크를 저장한다. 하나의 네트워크는 백본 네트워크로, 분석장치의 특징 추출기를 의미한다. 그리고 다른 하나는 영상에 대한 특징을 분석하는 합성곱 신경망을 의미한다. 백본 네트워크는 복수의 세그먼트들을 입력받아 복수의 특징벡터들을 출력하고, 합성곱 신경망은 복수의 특징벡터들 중 하나를 입력받아 다음 세그먼트에 대한 예측 데이터를 출력한다.

[0042] 계산장치(330)는 분석장치에 구비된 CPU 또는 AP와 같은 프로세서로 구현될 수 있다. 메모리에 저장된 특징 추

출기에 복수의 세그먼트들 중 제 1 세그먼트를 입력한다. 제 1 세그먼트는 입력장치에서 분류된 복수의 세그먼트들 중 가장 앞선 순번의 세그먼트일 수도 있고 시점 상 현재 시점의 세그먼트일 수 있다. 계산장치(330)는 제 1 세그먼트를 특징 추출기에 입력하여 제 1 특징벡터를 출력한다. 그리고 동일한 방식으로 제 2 세그먼트를 특징 추출기에 입력하여 제 2 특징벡터를 출력한다. 제 2 세그먼트는 제 1 세그먼트의 다음 순번의 세그먼트를 의미한다. 각 세그먼트는 입력장치에 따라 동일 숫자의 프레임으로 묶을 수 있다.

[0043] 한편, 계산장치(330)는 제 1 특징벡터 및 제 2 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터를 생성한다. 그리고, 합성곱 신경망에 제 1 특징벡터를 입력하여 출력된 제 1 예측 데이터와 레퍼런스 라벨 데이터를 비교한다. 두 데이터를 비교하는 것은 기본적으로 특징 추출기의 동영상 표현에 대한 성능을 높이기 위함이나, 특징 추출기 뿐만 아니라 합성곱 신경망의 예측 정확도를 개선하기 위해서도 이용될 수 있다.

[0044] 계산장치(330)는 특징 추출기를 통해 출력된 제 1, 제 2 특징벡터의 차이를 계산하고 이를 정규화 할 수 있다. 그리고 변화량의 차이에 따라 가장 변화량이 높은 순서부터 낮은 순서까지 점수를 매기는 것으로 레퍼런스 라벨 데이터를 생성할 수 있다. 즉, 제 1 특징벡터 및 제 2 특징벡터는 특징 추출기를 통해 출력되었을 때는 세그먼트의 너비(height), 높이(width) 및 채널(channel)을 포함하는 3차원 데이터지만, 시간*채널 축으로 정규화하는 과정에 따라 가로축이 너비이고 세로축이 높이인 매트릭스 데이터로 변환될 수 있다. 매트릭스 데이터의 내부에 변화량의 차이 정도를 나타내는 값이 저장될 수 있다. 이 데이터가 합성곱 신경망이 예측 정확도를 높이는데 이용하는 의사 GT 데이터로 이용된다. 당연하게도 초기 단계에서는 합성곱 신경망의 예측 성능이 낮은 수준이지만 학습이 반복됨에 따라 출력하는 예측 데이터가 의사 GT와 유사해질 수 있다. 따라서 분석장치의 동영상에 대한 충분한 표현력을 학습할 수 있다.

[0045] 도 4는 개시된 기술의 일 실시예에 따라 레퍼런스 라벨 데이터와 예측 데이터를 비교하는 신경망의 구조를 나타낸 도면이다. 도 4를 참조하면 특징 추출기인 백본 네트워크(401)와 합성곱 신경망(402)이 별도로 구비되어 있으며 백본 네트워크(401)에는 현재 세그먼트와 다음 세그먼트가 입력된다. 그리고 합성곱 신경망에는 현재 세그먼트의 특징벡터가 입력된다. 백본의 성능에 따라 세그먼트보다 더 많은 수의 프레임이 포함된 비디오 클립이 입력될 수도 있다.

[0046] 한편, 백본에 입력된 현재 세그먼트는 x축이 너비, y축이 높이, z축이 시간*채널인 특징벡터로 출력된다. 다음 세그먼트도 동일하게 3차원 형태의 특징벡터로 출력된다. 이후 두 특징벡터의 차이를 계산하고 이를 시간*채널 축으로 정규화한 뒤 스코어링을 수행함으로써 2차원 데이터로 축소할 수 있다. 축소된 데이터에는 현재 세그먼트의 특징과 다음 세그먼트의 특징을 비교한 결과가 저장된다. 즉, 두 세그면트의 특징 변화량이 저장될 수 있다. 특징의 변화량은 일부 특징값은 같은 변화량을 갖을 수 있고 나머지 특징값은 다른 변화량을 갖을 수 있다. 이는 영상에 따라 얼마든지 달라질 수 있다. 분석장치는 변화량의 차이에 대한 스코어링을 수행함으로써 합성곱 신경망이 이용하는 의사 GT 데이터(403)를 생성한다.

[0047] 한편, 합성곱 신경망(402)은 현재 세그먼트에 대한 특징벡터를 분석하여 예측 데이터(404)를 생성한다. 예측 데이터의 형태는 의사 GT 데이터(403)와 마찬가지로 x축이 너비이고 y축이 높이인 매트릭스 형태이다. 따라서 두 데이터를 비교하는 것이 가능하다. 분석장치는 예측 데이터(404)가 의사 GT 데이터(403)와 유사해지도록 크로스 엔트로피 손실함수를 적용할 수 있다. 분석장치는 크로스 엔트로피 손실함수를 통해 경사하강법을 이용하여 인공지능에 대한 학습을 반복 수행하게 된다. 분석장치는 특징 추출기와 합성곱 신경망의 매개변수를 각각 조정하는 것으로 네트워크 전반의 성능을 높일 수 있다.

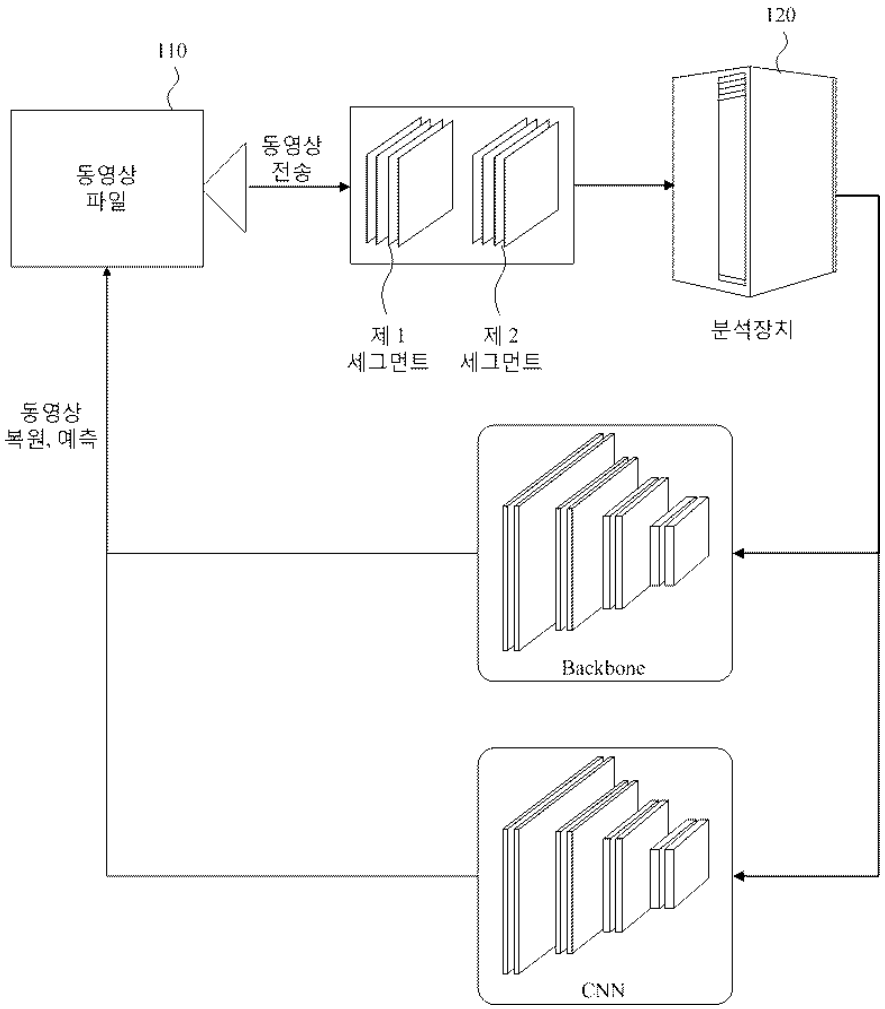
[0048] 이와 같이 분석장치는 먼저 특징벡터들을 추출한 다음 이를 기반으로 자가지도학습을 수행한다. 별도의 라벨링을 수행하지 않고 자체적으로 생성한 의사 GT 데이터를 레퍼런스 라벨 데이터로 이용하여 학습을 수행하기 때문에 동영상에 대한 라벨링에 소모되는 높은 비용을 방지할 수 있다.

[0049] 개시된 기술의 일 실시예에 따른 동영상 특징 추출 방법 및 장치는 이해를 돕기 위하여 도면에 도시된 실시 예를 참고로 설명되었으나, 이는 예시적인 것에 불과하며, 당해 분야에서 통상적 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 개시된 기술의 진정한 기술적 보호 범위는 첨부된 특허청구범위에 의해 정해져야 할 것이다.

도면

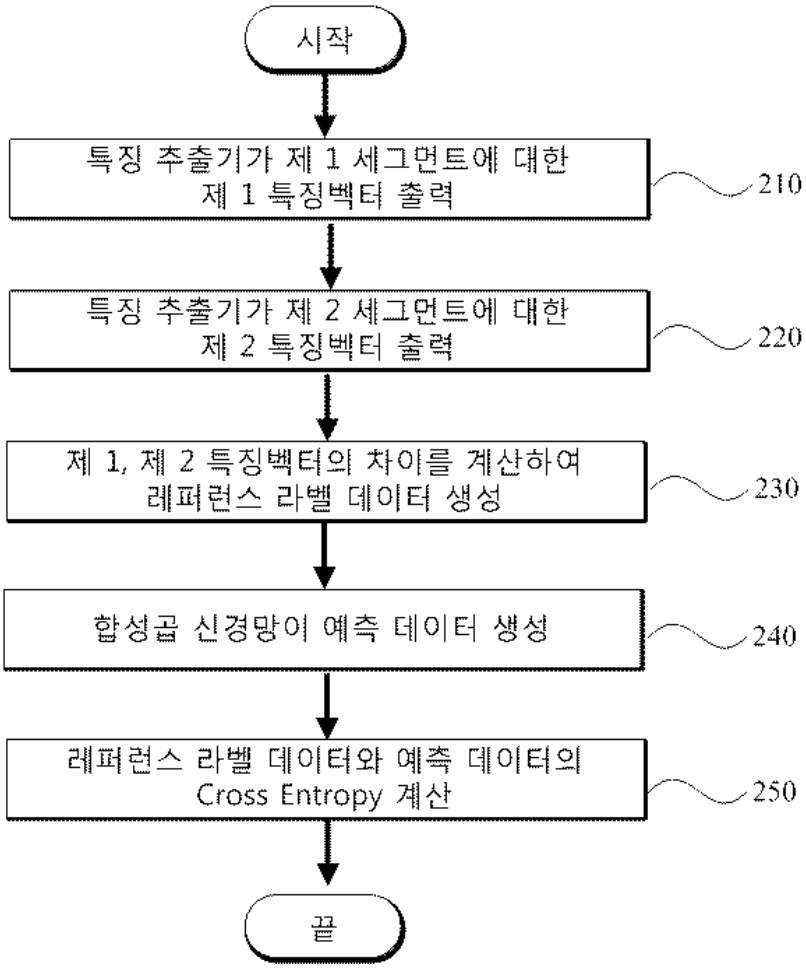
도면1

100

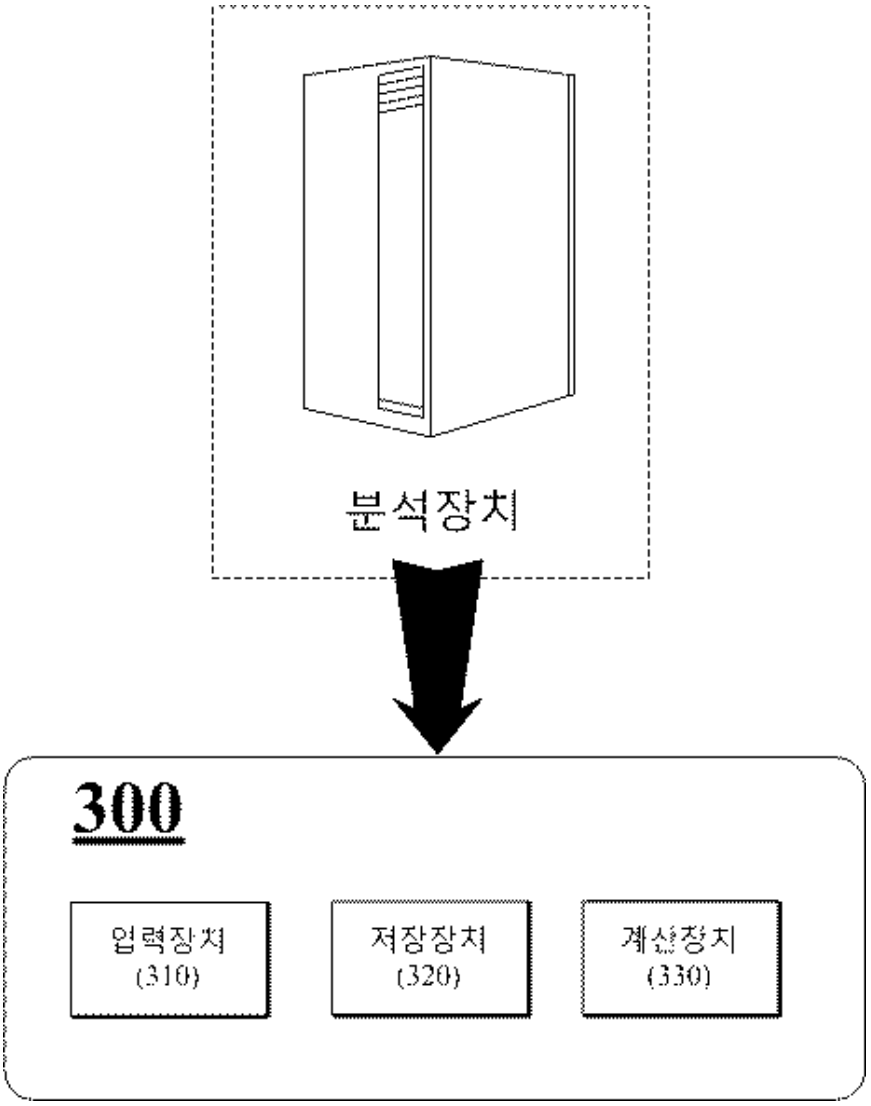

110
동영상
파일
동영상
전송
120
제 1
세그먼트
제 2
세그먼트
분석장치
동영상
복원, 예측
Backbone
CNN

도면2

200

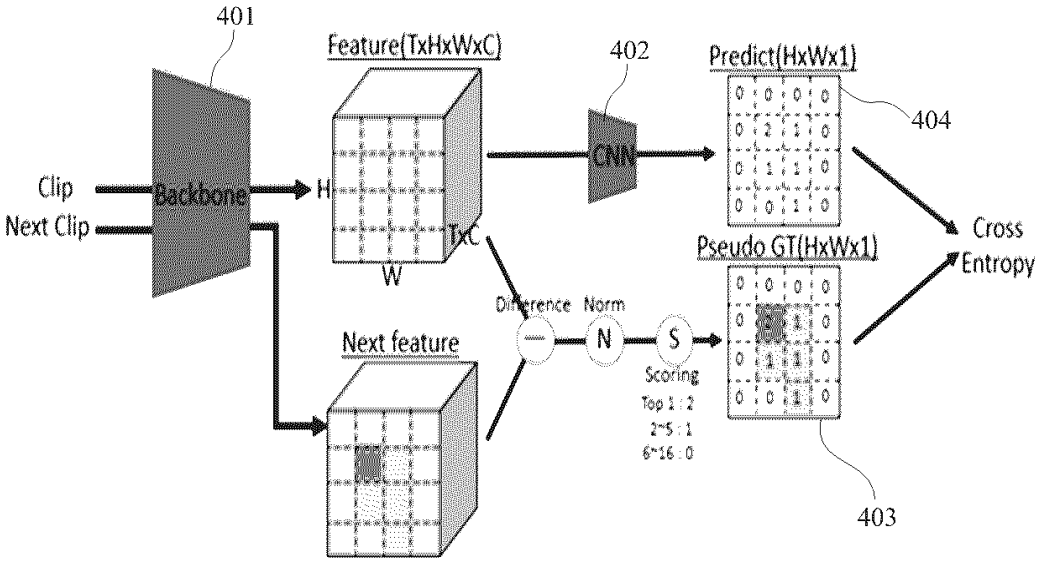

시작
특징 추출기가 제 1 세그먼트에 대한 제 1 특징벡터 출력
210
특징 추출기가 제 2 세그먼트에 대한 제 2 특징벡터 출력
220
제 1, 제 2 특징벡터의 차이를 계산하여 레퍼런스 라벨 데이터 생성
230
합성곱 신경망이 예측 데이터 생성
240
레퍼런스 라벨 데이터와 예측 데이터의 Cross Entropy 계산
250
끝

도면3



도면4

400


401
Feature(TxHxWxC)
402
Predict(HxWx1)
404
Clip
Next Clip
Backbone
H
W
TxC
CNN
Cross
Entropy
Pseudo GT(HxWx1)
Difference
Norm
Next feature
S
Scoring
Top 1 : 2
2~5 : 1
6~16 : 0
403