



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년01월05일
(11) 등록번호 10-2198480
(24) 등록일자 2020년12월29일

(51) 국제특허분류(Int. Cl.)
G06F 16/738 (2019.01) G06K 9/00 (2006.01)
H04N 21/845 (2011.01) H04N 21/8549 (2011.01)
(52) CPC특허분류
G06F 16/739 (2019.01)
G06K 9/00751 (2013.01)
(21) 출원번호 10-2020-0024805
(22) 출원일자 2020년02월28일
심사청구일자 2020년02월28일
(56) 선행기술조사문헌
KR1020190100320 A
KR1020180126362 A

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
손광훈
서울특별시 서대문구 연세로 50, 연세대학교 제3공학관 C129호(신촌동)
박정인
서울특별시 서대문구 연세로 50, 연세대학교 제3공학관 C129호(신촌동)
(74) 대리인
민영준

전체 청구항 수 : 총 18 항

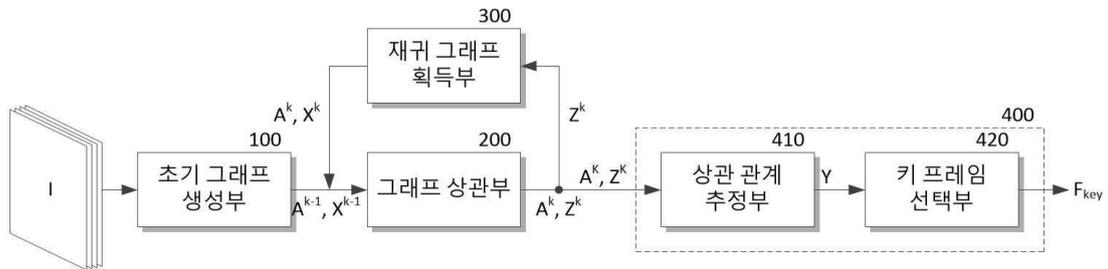
심사관 : 윤태섭

(54) 발명의 명칭 재귀 그래프 모델링을 통한 비디오 요약 생성 장치 및 방법

(57) 요약

본 발명은 비디오 영상의 다수 프레임을 관계 그래프의 노드로 간주하고, 재귀적 추론 구조를 갖는 그래프 컨볼루션 신경망을 이용하여 다수의 프레임 사이의 시멘틱 유사성을 추론함으로써 다수 프레임 사이의 전역적이고 장기적인 상호 관계를 고려한 정확한 키 프레임을 추출할 수 있어, 상대적으로 적은 파라미터 개수와 높은 정확성으로 다양한 비디오 플랫폼에서 비디오 요약을 생성할 수 있는 비디오 요약 생성 장치 및 방법을 제공할 수 있다.

대표도



(52) CPC특허분류

G06N 3/08 (2013.01)

H04N 21/845 (2013.01)

H04N 21/8549 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	2019070646
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	원천기술개발사업
연구과제명	(2세부)딥러닝 기반 의미론적 상황 이해 원천기술 연구 (2단계)(1/2)
기여율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2019.06.01 ~ 2020.03.31

명세서

청구범위

청구항 1

다수 프레임으로 구성되는 입력 비디오를 인가받고, 미리 학습된 패턴 추정 방식에 따라 다수의 프레임에서 각각 추출된 다수의 특징맵을 노드 벡터로 간주하여 초기 특징 그래프를 생성하며, 상기 노드 벡터 사이의 유사도를 계산하여 초기 인접 행렬을 생성하는 초기 그래프 생성부;

패턴 추정 방식이 미리 학습된 다수의 그래프 컨볼루션 모듈을 이용하여 상기 초기 특징 그래프와 상기 초기 인접 행렬 또는 반복적으로 재귀되어 인가되는 특징 그래프와 인접 행렬 사이의 패턴으로부터 상기 초기 특징 그래프 또는 상기 특징 그래프의 노드 사이의 패턴을 보정하여 보정 그래프를 추출하는 그래프 상관부;

상기 보정 그래프를 인가받아 미리 지정된 서로 다른 가중치 함수로 가중하고, 가중된 2개의 보정 그래프 사이의 유사도를 나타내는 보정 유사도와 이전 인접 행렬을 기반으로 다음 보정 그래프를 추출하기 위한 인접 행렬과 특징 그래프를 획득하는 재귀 그래프 획득부; 및

패턴 추정 방식이 미리 학습된 별도의 그래프 컨볼루션 네트워크를 이용하여, 기지정된 횟수로 반복 재귀하여 획득된 최종 보정 그래프와 최종 인접 행렬로부터 상기 최종 보정 그래프의 각 노드 사이의 시멘틱 유사도에 따라 다수의 프레임 각각이 키 프레임이 될 확률을 추정하여 다수의 키 프레임을 선택하는 키 프레임 추출부를 포함하는 비디오 요약 생성 장치.

청구항 2

제1 항에 있어서, 상기 초기 그래프 생성부는

미리 학습된 패턴 추정 방식에 따라 상기 입력 비디오의 다수의 프레임 각각에 대한 특징을 추출하여 상기 다수의 특징맵을 생성하는 특징맵 획득부;

상기 다수의 특징맵 각각을 다수의 노드와 다수의 노드를 서로 연결하는 다수의 에지로 구성되는 그래프의 노드로 간주하여 상기 초기 특징 그래프를 획득하는 초기 그래프 획득부; 및

상기 초기 특징 그래프의 다수의 노드 사이의 유사도를 계산하여 초기 인접 행렬을 획득하는 초기 인접 행렬 획득부를 포함하는 비디오 요약 생성 장치.

청구항 3

제2 항에 있어서, 상기 초기 그래프 획득부는

상기 다수의 특징맵 각각을 1차원의 노드 벡터로 변환하여 상기 초기 특징 그래프를 획득하는 비디오 요약 생성 장치.

청구항 4

제3 항에 있어서, 상기 초기 인접 행렬 획득부는

상기 초기 특징 그래프(X^0)의 각 노드 벡터(x_1, \dots, x_T) 사이의 유사도를 수학식

$$a_{ij} = \frac{x_i^T x_j}{\|x_i\|_2 \cdot \|x_j\|_2}$$

(여기서 a_{ij} 는 초기 인접 행렬(A^0)의 원소를 나타내고, x_i 와 x_j 는 각각 초기 특징 그래프(X^0)의 다수의 노드 벡터(x_1, \dots, x_T) 중 i 번째 노드 벡터와 j 번째 노드 벡터를 나타내고, T 는 전치 행렬을 나타내며, $\| \cdot \|_2$ 는 L2 놈(L2

norm) 함수를 나타낸다.)

에 따라 계산하여 상기 초기 인접 행렬(A^0)을 획득하는 비디오 요약 생성 장치.

청구항 5

제2 항에 있어서, 상기 초기 그래프 생성부는

상기 입력 비디오의 다수의 프레임에서 기지정된 시간 구간 단위로 프레임을 추출하여 상기 특징맵 획득부로 전달하는 프레임 선별부를 더 포함하는 비디오 요약 생성 장치.

청구항 6

제2 항에 있어서, 상기 그래프 상관부는

상기 다수의 그래프 컨볼루션 모듈이 직렬로 순차 연결되고, 상기 다수의 그래프 상관 모듈 중 초기단의 그래프 상관 모듈은 초기 특징 그래프 또는 재귀된 특징 그래프와 초기 인접 행렬 또는 재귀된 인접 행렬을 인가받아 패턴을 추정하여 제1 중간 특징 그래프를 추출하고, 나머지 그래프 상관 모듈은 각각 이전단에서 추출된 중간 특징 그래프와 초기단의 그래프 상관 모듈에 인가된 인접 행렬을 인가받아 패턴을 추정하여 중간 특징 그래프 또는 상기 보정 그래프를 추출하는 비디오 요약 생성 장치.

청구항 7

제6 항에 있어서, 상기 재귀 그래프 획득부는

이전 획득된 상기 보정 그래프를 각각 인가받고, 상기 보정 그래프가 서로 다른 선형 임베딩 공간에 투사되도록 서로 다른 기지정된 가중치 함수를 상기 보정 그래프에 가중하여 2개의 투영 보정 그래프를 획득하는 투사부;

상기 2개의 투영 보정 그래프 사이의 유사도를 계산하여 상기 보정 유사도를 획득하는 인접 보정값 획득부; 및

이전 인접 행렬과 획득된 보정 유사도를 가산하여 상기 인접 행렬을 획득하는 인접 행렬 획득부를 포함하는 비디오 요약 생성 장치.

청구항 8

제7 항에 있어서, 상기 인접 보정값 획득부는

상기 보정 유사도(dA^k)를 수학식

$$dA^k = \frac{(W_\theta Z^k)^T (W_\phi Z^k)}{\|W_\theta Z^k\|_2 \cdot \|W_\phi Z^k\|_2}$$

(여기서 $W_\theta Z^k$, $W_\phi Z^k$ 는 보정 그래프(Z^k)에 가중치 함수(W_θ , W_ϕ)가 가중된 투영 보정 그래프를 나타내고, T는 전치 행렬을 나타내며, $\| \cdot \|_2$ 는 L2 놈(L2 norm) 함수를 나타낸다.)

에 따라 계산하는 비디오 요약 생성 장치.

청구항 9

제7 항에 있어서, 상기 키 프레임 추출부는

패턴 추정 방식이 미리 학습된 별도의 그래프 컨볼루션 네트워크에 상기 최종 보정 그래프와 상기 최종 인접 행렬을 인가하여, 상기 최종 보정 그래프의 각 노드 사이의 시멘틱 유사도 패턴에 따라 상기 입력 비디오의 다수의 프레임 각각이 키 프레임일 확률을 나타내는 키 프레임 확률맵을 추출하는 상관 관계 추정부; 및

상기 키 프레임 확률맵으로부터 상기 입력 비디오의 다수의 프레임 각각이 키 프레임될 확률을 분석하여 다수의 키 프레임을 선택하는 키 프레임 선택부를 포함하는 비디오 요약 생성 장치.

청구항 10

다수 프레임으로 구성되는 입력 비디오를 인가받고, 미리 학습된 패턴 추정 방식에 따라 다수의 프레임에서 각각 추출된 다수의 특징맵을 노드 벡터로 간주하여 초기 특징 그래프를 생성하며, 상기 노드 벡터 사이의 유사도를 계산하여 초기 인접 행렬을 생성하는 단계;

패턴 추정 방식이 미리 학습된 다수의 그래프 컨볼루션 모듈을 이용하여 상기 초기 특징 그래프와 상기 초기 인접 행렬 또는 반복적으로 재귀되어 인가되는 특징 그래프와 인접 행렬 사이의 패턴으로부터 상기 초기 특징 그래프 또는 상기 특징 그래프의 노드 사이의 패턴을 보정하여 보정 그래프를 추출하는 단계;

상기 보정 그래프를 인가받아 미리 지정된 서로 다른 가중치 함수로 가중하고, 가중된 2개의 보정 그래프 사이의 유사도를 나타내는 보정 유사도와 이전 인접 행렬을 기반으로 다음 보정 그래프를 추출하기 위한 인접 행렬과 특징 그래프를 획득하는 단계; 및

패턴 추정 방식이 미리 학습된 별도의 그래프 컨볼루션 네트워크를 이용하여, 기지정된 횡수로 반복 재귀하여 획득된 최종 보정 그래프와 최종 인접 행렬로부터 상기 최종 보정 그래프의 각 노드 사이의 시멘틱 유사도에 따라 다수의 프레임 각각이 키 프레임이 될 확률을 추정하여 다수의 키 프레임을 선택하는 단계를 포함하는 비디오 요약 생성 방법.

청구항 11

제10 항에 있어서, 상기 초기 인접 행렬을 생성하는 단계는

미리 학습된 패턴 추정 방식에 따라 상기 입력 비디오의 다수의 프레임 각각에 대한 특징을 추출하여 상기 다수의 특징맵을 생성하는 단계;

상기 다수의 특징맵 각각을 다수의 노드와 다수의 노드를 서로 연결하는 다수의 에지로 구성되는 그래프의 노드로 간주하여 상기 초기 특징 그래프를 획득하는 단계; 및

상기 초기 특징 그래프의 다수의 노드 사이의 유사도를 계산하여 초기 인접 행렬을 획득하는 단계를 포함하는 비디오 요약 생성 방법.

청구항 12

제11 항에 있어서, 상기 초기 특징 그래프를 획득하는 단계는

상기 다수의 특징맵 각각을 1차원의 노드 벡터로 변환하여 상기 초기 특징 그래프를 획득하는 비디오 요약 생성 방법.

청구항 13

제11 항에 있어서, 상기 초기 인접 행렬을 획득하는 단계는

상기 초기 특징 그래프(X^0)의 각 노드 벡터(x_1, \dots, x_T) 사이의 유사도를 수학식

$$a_{ij} = \frac{x_i^T x_j}{\|x_i\|_2 \cdot \|x_j\|_2}$$

(여기서 a_{ij} 는 초기 인접 행렬(A^0)의 원소를 나타내고, x_i 와 x_j 는 각각 초기 특징 그래프(X^0)의 다수의 노드 벡터(x_1, \dots, x_T) 중 i 번째 노드 벡터와 j 번째 노드 벡터를 나타내고, T 는 전치 행렬을 나타내며, $\| \cdot \|_2$ 는 L2 놈(L2 norm) 함수를 나타낸다.)

에 따라 계산하여 상기 초기 인접 행렬(A^0)을 획득하는 비디오 요약 생성 방법.

청구항 14

제11 항에 있어서, 상기 초기 인접 행렬을 생성하는 단계는

상기 다수의 특징맵을 생성하는 단계 이전, 상기 입력 비디오의 다수의 프레임에서 기지정된 시간 구간 단위로 프레임을 추출하는 단계를 더 포함하는 비디오 요약 생성 방법.

청구항 15

제11 항에 있어서, 상기 보정 그래프를 추출하는 단계는

상기 다수의 그래프 컨볼루션 모듈이 직렬로 순차 연결되고, 상기 다수의 그래프 상관 모듈 중 초기단의 그래프 상관 모듈은 초기 특징 그래프 또는 재귀된 특징 그래프와 초기 인접 행렬 또는 재귀된 인접 행렬을 인가받아 패턴을 추정하여 제1 중간 특징 그래프를 추출하고, 나머지 그래프 상관 모듈은 각각 이전단에서 추출된 중간 특징 그래프와 초기단의 그래프 상관 모듈에 인가된 인접 행렬을 인가받아 패턴을 추정하여 중간 특징 그래프 또는 상기 보정 그래프를 추출하는 비디오 요약 생성 방법.

청구항 16

제15 항에 있어서, 상기 인접 행렬과 특징 그래프를 획득하는 단계는

이전 획득된 상기 보정 그래프를 각각 인가받고, 상기 보정 그래프가 서로 다른 선형 임베딩 공간에 투사되도록 서로 다른 기지정된 가중치 함수를 상기 보정 그래프에 가중하여 2개의 투영 보정 그래프를 획득하는 단계;

상기 2개의 투영 보정 그래프 사이의 유사도를 계산하여 상기 보정 유사도를 획득하는 단계; 및

이전 인접 행렬과 획득된 보정 유사도를 가산하여 상기 인접 행렬을 획득하는 단계를 포함하는 비디오 요약 생성 방법.

청구항 17

제16 항에 있어서, 상기 보정 유사도를 획득하는 단계는

상기 보정 유사도(dA^k)를 수학식

$$dA^k = \frac{(W_\theta Z^k)^T (W_\phi Z^k)}{\|W_\theta Z^k\|_2 \cdot \|W_\phi Z^k\|_2}$$

(여기서 $W_\theta Z^k$, $W_\phi Z^k$ 는 보정 그래프(Z^k)에 가중치 함수(W_θ , W_ϕ)가 가중된 투영 보정 그래프를 나타내고, T 는 전치 행렬을 나타내며, $\| \cdot \|_2$ 는 L2 놈(L2 norm) 함수를 나타낸다.)

에 따라 계산하는 비디오 요약 생성 방법.

청구항 18

제16 항에 있어서, 상기 다수의 키 프레임을 선택하는 단계는

패턴 추정 방식이 미리 학습된 별도의 그래프 컨볼루션 네트워크에 상기 최종 보정 그래프와 상기 최종 인접 행렬을 인가하여, 상기 최종 보정 그래프의 각 노드 사이의 시멘틱 유사도 패턴에 따라 상기 입력 비디오의 다수의 프레임 각각이 키 프레임일 확률을 나타내는 키 프레임 확률맵을 추출하는 단계; 및

상기 키 프레임 확률맵으로부터 상기 입력 비디오의 다수의 프레임 각각이 키 프레임될 확률을 분석하여 다수의 키 프레임을 선택하는 단계를 포함하는 비디오 요약 생성 방법.

발명의 설명

기술 분야

본 발명은 비디오 요약 생성 장치 및 방법에 관한 것으로, 재귀 그래프 모델링을 통한 비디오 요약 생성 장치

[0001]

및 방법에 관한 것이다.

배경 기술

- [0002] 최근 온라인 스트리밍 서비스와 같은 비디오 플랫폼의 변화로 인해, 사용자가 원하는 비디오 데이터에 액세스하기 어려워졌다. 또한 비디오의 길이가 점차로 증가하고 있기 때문에 사용자는 비디오에서 유용한 정보를 획득하는 것이 더욱 어려워졌다. 이에 사용자가 비디오 데이터를 효율적으로 탐색할 수 있도록 하기 위한 연구가 수행되고 있다. 이러한 연구의 일환으로 비디오에서 유용한 장면을 선출하여 비디오의 내용을 간결하게 묘사할 수 있도록 하는 비디오 요약은 사용자 편의성을 크게 향상시킬 수 있는 도구로서 주목받고 있다.
- [0003] 다만 비디오 요약은 비디오의 전체 프레임 사이의 상호 연관성을 고려하여 유용한 프레임을 선출해야 한다는 어려움을 갖는다. 이에 인공 신경망 (artificial neural network)을 이용하여 키 프레임(key frame)을 추출하는 비디오 요약 방식에 대한 다양한 연구가 수행되고 있다.
- [0004] 기존에 인공 신경망을 이용하는 비디오 요약 알고리즘들은 주로 컨볼루션 신경망(Convolutional neural networks: CNN) 또는 순환 신경망(Recurrent neural networks: RNN)을 기반으로 구현되었다. 그러나 CNN이나 RNN과 같은 인공 신경망은 작은 수용 영역(receptive field)와 표현자를 전역적으로 표현하지 못하는 국지성(locality)의 문제를 가지고 있어 비디오의 장기 의존성(long-term dependency)을 반영하기에 적절하지 않다는 한계가 있다.

선행기술문헌

특허문헌

- [0005] (특허문헌 0001) 한국 공개 특허 제10-2019-0099027호 (2019.08.23 공개)

발명의 내용

해결하려는 과제

- [0006] 본 발명의 목적은 비디오의 다수의 프레임을 관계 그래프의 노드로 간주하여 그래프 컨볼루션 신경망에 적용함으로써 비디오 요약을 위한 키 프레임을 정확하게 추출할 수 있는 비디오 요약 생성 장치 및 방법을 제공하는데 있다.
- [0007] 본 발명의 다른 목적은 다수의 프레임 사이의 시멘틱 유사성을 재귀적으로 추론하여 다수 프레임 사이의 전역적이고 장기적인 상호 관계를 고려한 키 프레임을 추출할 수 있는 비디오 요약 생성 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

- [0008] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 비디오 요약 생성 장치는 다수 프레임으로 구성되는 입력 비디오를 인가받고, 미리 학습된 패턴 추정 방식에 따라 다수의 프레임에서 각각 추출된 다수의 특징맵을 노드 벡터로 간주하여 초기 특징 그래프를 생성하며, 상기 노드 벡터 사이의 유사도를 계산하여 초기 인접 행렬을 생성하는 초기 그래프 생성부; 패턴 추정 방식이 미리 학습된 다수의 그래프 컨볼루션 모듈을 이용하여 상기 초기 특징 그래프와 상기 초기 인접 행렬 또는 반복적으로 재귀되어 인가되는 특징 그래프와 인접 행렬 사이의 패턴으로부터 상기 초기 특징 그래프 또는 상기 특징 그래프의 노드 사이의 패턴을 보정하여 보정 그래프를 추출하는 그래프 상관부; 상기 보정 그래프를 인가받아 미리 지정된 서로 다른 가중치 함수로 가중하고, 가중된 2개의 보정 그래프 사이의 유사도를 나타내는 보정 유사도와 이전 인접 행렬을 기반으로 다음 보정 그래프를 추출하기 위한 인접 행렬과 특징 그래프를 획득하는 재귀 그래프 획득부; 및 패턴 추정 방식이 미리 학습된 별도의 그래프 컨볼루션 네트워크를 이용하여, 기지정된 횟수로 반복 재귀하여 획득된 최종 보정 그래프와 최종 인접 행렬로부터 상기 최종 보정 그래프의 각 노드 사이의 시멘틱 유사도에 따라 다수의 프레임 각각이 키 프레임이 될 확률을 추정하여 다수의 키 프레임을 선택하는 키 프레임 추출부를 포함한다.
- [0009] 상기 초기 그래프 생성부는 미리 학습된 패턴 추정 방식에 따라 상기 입력 비디오의 다수의 프레임 각각에 대한 특징을 추출하여 상기 다수의 특징맵을 생성하는 특징맵 획득부; 상기 다수의 특징맵 각각을 다수의 노드와

다수의 노드를 서로 연결하는 다수의 에지로 구성되는 그래프의 노드로 간주하여 상기 초기 특징 그래프를 획득하는 초기 그래프 획득부; 및 상기 초기 특징 그래프의 다수의 노드 사이의 유사도를 계산하여 초기 인접 행렬을 획득하는 초기 인접 행렬 획득부를 포함할 수 있다.

[0010] 상기 초기 그래프 획득부는 상기 다수의 특징맵 각각을 1차원의 노드 벡터로 변환하여 상기 초기 특징 그래프를 획득할 수 있다.

[0011] 상기 초기 인접 행렬 획득부는 상기 초기 특징 그래프(X^0)의 각 노드 벡터(x_1, \dots, x_T) 사이의 유사도를 수학식

$$a_{ij} = \frac{x_i^T x_j}{\|x_i\|_2 \cdot \|x_j\|_2}$$

[0012]

[0013] (여기서 a_{ij} 는 초기 인접 행렬(A^0)의 원소를 나타내고, x_i 와 x_j 는 각각 초기 특징 그래프(X^0)의 다수의 노드 벡터(x_1, \dots, x_T) 중 i 번째 노드 벡터와 j 번째 노드 벡터를 나타내고, T 는 전체 행렬을 나타내며, $\| \cdot \|_2$ 는 L2 놈(L2 norm) 함수를 나타낸다.)에 따라 계산하여 상기 초기 인접 행렬(A^0)을 획득할 수 있다.

[0014] 상기 초기 그래프 생성부는 상기 입력 비디오의 다수의 프레임에서 기지정된 시간 구간 단위로 프레임을 추출하여 상기 특징맵 획득부로 전달하는 프레임 선별부를 더 포함할 수 있다.

[0015] 상기 그래프 상관부는 상기 다수의 그래프 컨볼루션 모듈이 직렬로 순차 연결되고, 상기 다수의 그래프 상관 모듈 중 초기단의 그래프 상관 모듈은 초기 특징 그래프 또는 재귀된 특징 그래프와 초기 인접 행렬 또는 재귀된 인접 행렬을 인가받아 패턴을 추정하여 제1 중간 특징 그래프를 추출하고, 나머지 그래프 상관 모듈은 각각 이전단에서 추출된 중간 특징 그래프와 초기단의 그래프 상관 모듈에 인가된 인접 행렬을 인가받아 패턴을 추정하여 중간 특징 그래프 또는 상기 보정 그래프를 추출할 수 있다.

[0016] 상기 재귀 그래프 획득부는 이전 획득된 상기 보정 그래프를 각각 인가받고, 상기 보정 그래프가 서로 다른 선형 임베딩 공간에 투사되도록 서로 다른 기지정된 가중치 함수를 상기 보정 그래프에 가중하여 2개의 투영 보정 그래프를 획득하는 투사부; 상기 2개의 투영 보정 그래프 사이의 유사도를 계산하여 상기 보정 유사도를 획득하는 인접 보정값 획득부; 및 이전 인접 행렬과 획득된 보정 유사도를 가산하여 상기 인접 행렬을 획득하는 인접 행렬 획득부를 포함할 수 있다.

[0017] 상기 인접 보정값 획득부는 상기 보정 유사도(dA^k)를 수학식

$$dA^k = \frac{(W_\theta Z^k)^T (W_\phi Z^k)}{\|W_\theta Z^k\|_2 \cdot \|W_\phi Z^k\|_2}$$

[0018]

[0019] (여기서 $W_\theta Z^k$, $W_\phi Z^k$ 는 보정 그래프(Z^k)에 가중치 함수(W_θ , W_ϕ)가 가중된 투영 보정 그래프를 나타내고, T 는 전체 행렬을 나타내며, $\| \cdot \|_2$ 는 L2 놈(L2 norm) 함수를 나타낸다.)에 따라 계산할 수 있다.

[0020] 상기 키 프레임 추출부는 패턴 추정 방식이 미리 학습된 별도의 그래프 컨볼루션 네트워크에 상기 최종 보정 그래프와 상기 최종 인접 행렬을 인가하여, 상기 최종 보정 그래프의 각 노드 사이의 시멘틱 유사도 패턴에 따라 상기 입력 비디오의 다수의 프레임 각각이 키 프레임일 확률을 나타내는 키 프레임 확률맵을 추출하는 상관 관계 추정부; 및 상기 키 프레임 확률맵으로부터 상기 입력 비디오의 다수의 프레임 각각이 키 프레임될 확률을 분석하여 다수의 키 프레임을 선택하는 키 프레임 선택부를 포함할 수 있다.

[0021] 상기 다른 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 비디오 요약 생성 방법은 다수 프레임으로 구성되는 입력 비디오를 인가받고, 미리 학습된 패턴 추정 방식에 따라 다수의 프레임에서 각각 추출된 다수의 특징맵을 노드 벡터로 간주하여 초기 특징 그래프를 생성하며, 상기 노드 벡터 사이의 유사도를 계산하여 초기 인접 행렬을 생성하는 단계; 패턴 추정 방식이 미리 학습된 다수의 그래프 컨볼루션 모듈을 이용하여 상기 초기 특징

그래프와 상기 초기 인접 행렬 또는 반복적으로 재귀되어 인가되는 특징 그래프와 인접 행렬 사이의 패턴으로부터 상기 초기 특징 그래프 또는 상기 특징 그래프의 노드 사이의 패턴을 보정하여 보정 그래프를 추출하는 단계; 상기 보정 그래프를 인가받아 미리 지정된 서로 다른 가중치 함수로 가중하고, 가중된 2개의 보정 그래프 사이의 유사도를 나타내는 보정 유사도와 이전 인접 행렬을 기반으로 다음 보정 그래프를 추출하기 위한 인접 행렬과 특징 그래프를 획득하는 단계; 및 패턴 추정 방식이 미리 학습된 별도의 그래프 컨볼루션 네트워크를 이용하여, 기지정된 횟수로 반복 재귀하여 획득된 최종 보정 그래프와 최종 인접 행렬로부터 상기 최종 보정 그래프의 각 노드 사이의 시멘틱 유사도에 따라 다수의 프레임 각각이 키 프레임이 될 확률을 추정하여 다수의 키 프레임을 선택하는 단계를 포함한다.

발명의 효과

[0022] 따라서, 본 발명의 실시예에 따른 비디오 요약 생성 장치 및 방법은 비디오 영상의 다수 프레임을 관계 그래프의 노드로 간주하고, 재귀적 추론 구조를 갖는 그래프 컨볼루션 신경망을 이용하여 다수의 프레임 사이의 시멘틱 유사성을 추론함으로써 다수 프레임 사이의 전역적이고 장기적인 상호 관계를 고려한 정확한 키 프레임을 추출할 수 있다. 그러므로 상대적으로 적은 파라미터 개수와 높은 정확성으로 다양한 비디오 플랫폼에서 비디오 요약을 사용자에게 용이하게 제공할 수 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 일 실시예에 따른 비디오 요약 생성 장치의 개략적 구조를 나타낸다.
 도 2는 도 1의 비디오 요약 생성 장치의 초기 그래프 생성부의 상세 구성을 나타낸다.
 도 3은 도 1의 비디오 요약 생성 장치의 그래프 상관부와 재귀 그래프 획득부의 상세 구성을 나타낸다.
 도 4는 재귀 반복 횟수에 따라 추출된 키 프레임의 일예를 나타낸다.
 도 5는 본 발명의 일 실시예에 따른 비디오 요약 생성 방법을 나타낸다.
 도 6은 본 실시예에 따른 비디오 요약 생성 방법과 기존의 비디오 요약 생성 방법에 따라 생성된 비디오 요약을 비교한 결과를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0024] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0025] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.

[0026] 명세서 전체에서, 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. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0027] 도 1은 본 발명의 일 실시예에 따른 비디오 요약 생성 장치의 개략적 구조를 나타낸다.

[0028] 도 1을 참조하면, 본 실시예에 따른 비디오 요약 생성 장치는 초기 그래프 생성부(100), 그래프 상관부(200), 재귀 그래프 획득부(300) 및 키 프레임 추출부(400)를 포함할 수 있다.

[0029] 초기 그래프 생성부(100)는 다수의 프레임으로 구성된 입력 비디오(I)를 인가받아 각 프레임에 대한 특징을 추출하여 다수의 특징맵($X = \{x_1, \dots, x_T\}$)을 생성하고, 생성된 다수의 특징맵(x_1, \dots, x_T) 각각을 노드 벡터로 간주하여, 초기 특징 그래프(X^0)를 생성하며, 초기 특징 그래프(X^0)의 각 노드 벡터 사이의 유사도(affinity)를 계산하여 초기 인접 행렬(A^0)을 생성한다. 여기서 초기 인접 행렬(A^0)은 $T \times T$ 크기로 획득될 수 있다.

[0030] 그래프 상관부(200)는 초기 그래프 생성부(100)에서 생성된 초기 특징 그래프(X^0)와 초기 인접 행렬(A^0) 또는 이

전 획득된 특징 그래프(X^{k-1})와 재귀 그래프 획득부(300)에서 획득된 인접 행렬(A^{k-1})을 인가받고, 학습을 통해 미리 획득된 가중치 행렬(W)을 이용한 패턴 추정 방식에 따라 초기 특징 그래프(X^0)와 초기 인접 행렬(A^0) 또는 특징 그래프(X^{k-1})와 인접 행렬(A^{k-1}) 사이의 패턴으로부터 초기 특징 그래프(X^0) 또는 특징 그래프(X^{k-1})의 노드 사이의 패턴을 보정하여 보정 그래프(Z^k)를 추출한다.

[0031] 재귀 그래프 획득부(300)는 그래프 상관부(200)에서 추출된 보정 그래프(Z^k)를 인가받아 미리 지정된 가중치 함수(W_θ, W_ϕ)로 가중하여 서로 다른 선형 임베딩 공간에 투사하고, 서로 다른 선형 임베딩 공간에 투사된 2개의 보정 그래프($W_\theta Z^k, W_\phi Z^k$) 사이의 보정 유사도(dA^k)와 이전 인접 행렬(A^{k-1})을 이용하여 인접 행렬(A^k)을 획득한다. 그리고 이미 획득된 보정 그래프(Z^k)를 특징 그래프($X^k = Z^k$)로서 인접 행렬(A^k)과 함께 그래프 상관부(200)로 전달하여, 그래프 상관부(200)가 다음 보정 그래프(Z^{k+1})를 추출할 수 있도록 한다.

[0032] 키 프레임 추출부(400)는 그래프 상관부(200)와 재귀 그래프 획득부(300)에 의해 지정된 횟수로 반복 재귀하여 획득한 최종 보정 그래프(Z^k)와 최종 인접 행렬(A^k)을 인가받고, 미리 학습된 패턴 추정 방식에 따라 인가된 최종 보정 그래프(Z^k)의 각 노드 사이의 시멘틱 유사도 패턴을 추정하여 입력 비디오(I)의 다수의 프레임 각각이 키 프레임(F_{key})일 가능성을 나타내는 키 프레임 확률맵(Y)을 추출하고, 추출된 키 프레임 확률맵(Y)에 따라 키 프레임 키 프레임(F_{key})을 선택한다.

[0033] 즉 본 실시예에 따른 비디오 요약 생성 장치는 입력 비디오(I)의 다수의 프레임 각각에서 추출된 특징맵을 노드로 간주하여 다수의 노드로 구성된 초기 특징 그래프(X^0)와 노드 사이의 유사도를 나타내는 초기 인접 행렬(A^0)을 생성하고, 생성된 초기 특징 그래프(X^0)와 초기 인접 행렬(A^0)로부터 각 노드의 특징을 인접 행렬의 유사도에 기반하여 보정한 보정 특징 그래프(Z^k)와 인접 행렬(A^k)을 지정된 횟수로 재귀적으로 반복 추출하여 최종 획득된 최종 보정 특징 그래프(Z^k)와 최종 인접 행렬(A^k)을 획득한 이후, 최종 보정 특징 그래프(Z^k)와 최종 인접 행렬(A^k)로부터 입력 비디오(I)의 다수의 프레임 각각이 키 프레임(F_{key})일 확률을 시멘틱 유사도 패턴에 따라 추정한다.

[0034] 따라서 입력 비디오(I)의 다수의 프레임 사이의 상관 관계가 보정된 최종 보정 특징 그래프(Z^k)와 최종 인접 행렬(A^k)을 기반으로 시멘틱 유사도가 높은 키 프레임을 정확하게 선택할 수 있도록 한다.

[0035] 이하에서는 도 1의 비디오 요약 생성 장치의 각 구성을 상세하게 설명한다.

[0036] 도 2는 도 1의 비디오 요약 생성 장치의 초기 그래프 생성부의 상세 구성을 나타낸다.

[0037] 도 2를 참조하면, 초기 그래프 생성부(100)는 프레임 선별부(110), 특징맵 획득부(120), 초기 그래프 획득부(130) 및 초기 인접 행렬 획득부(140)를 포함할 수 있다.

[0038] 프레임 선별부(110)는 다수의 프레임을 포함하는 입력 비디오(I)를 인가받아 지정된 시간 구간 단위(예를 들면 1초)로 프레임을 선별하여 특징맵 획득부(120)로 전달한다. 일반적으로 비디오 영상은 초당 30 프레임 내지 60 프레임으로 구성되며, 대부분 짧은 시간 구간 내에서는 다수의 프레임이 서로 매우 유사하다. 따라서 입력 비디오(I)의 모든 프레임 사이의 시멘틱 유사성을 분석하여 비디오 요약을 생성하기 위한 키 프레임을 선별하는 것은 비효율적이다.

[0039] 이에 프레임 선별부(110)는 비디오 요약 생성 장치의 효율성을 향상시키기 위해 인가된 입력 비디오(I)에서 지정된 시간 구간 단위로 프레임을 선별하여 특징맵 획득부(120)로 전달한다.

[0040] 다만 이는 일반적으로 비디오 요약은 비디오 영상에서 프레임 단위로 정밀하게 키 프레임을 선별할 것을 요구하지 않는다는 점을 고려하여 효율성을 높이기 위한 것으로, 키 프레임을 매우 정확하게 추출해야 하는 경우에는 프레임을 선별하지 않을 수 있다. 즉 프레임 선별부(110)는 생략될 수도 있다.

[0041] 특징맵 획득부(120)는 패턴 추정 방식이 미리 학습된 인공 신경망으로 구현되어 입력 비디오(I) 또는 프레임 선별부(110)에서 프레임이 선별된 입력 비디오를 인가받고, 미리 학습된 패턴 추정 방식에 따라 각 프레임의 특징

을 추출하여 다수의 특징맵($X = \{x_1, \dots, x_T\}$)을 획득한다.

[0042] 특징맵 획득부(120)는 다양한 인공 신경망으로 구현될 수 있으며, 각 프레임 이미지로부터 특징맵(x_1, \dots, x_T)을 추출하는 다양한 인공 신경망이 이미 공개되어 있으므로, 이러한 공개된 인공 신경망 중 하나를 이용하여 구현될 수도 있다. 특징맵 획득부(120)는 일 예로 미리 학습된 컨볼루션 신경망(Convolutional Neural Network: 이하 CNN) 등으로 구현될 수 있다.

[0043] 초기 그래프 획득부(130)는 다수의 특징맵(X) 각각을 노드 벡터로 간주하여 초기 특징 그래프(X^0)를 획득한다. 초기 그래프 획득부(130)는 일 예로 특징맵 획득부(120)에서 지정된 2차원 또는 3차원의 크기로 획득된 다수의 특징맵(X)을 1차원의 노드 벡터로 변환하여, 초기 특징 그래프(X^0)를 획득할 수 있다.

[0044] 초기 그래프 획득부(130)는 다수의 특징맵(X) 각각을 다수의 노드와 다수의 노드를 서로 연결하는 다수의 에지로 구성되는 그래프의 노드로 간주함으로써, 이후 그래프 컨볼루션 네트워크(Graph Convolutional Network: GCN)로 구성되는 그래프 상관부(200)에 적합한 초기 특징 그래프(X^0)로 획득한다. 이때 초기 그래프 획득부(130)는 다수의 특징맵(X) 각각을 지정된 형식을 갖는 노드 벡터로 변환하여 초기 특징 그래프(X^0)를 획득하도록 구성될 수도 있다. 일 예로 초기 그래프 획득부(130)는 다수의 특징맵(X) 각각이 $T \times D$ 크기의 2차원 벡터로 획득된 경우, 이를 $T \times D$ 길이의 1차원 벡터 형태로 변환할 수 있다.

[0045] 초기 인접 행렬 획득부(140)는 초기 그래프 획득부(130)에서 획득된 초기 특징 그래프(X^0)의 각 노드 벡터(x_1, \dots, x_T), 즉 다수의 특징맵(X) 사이의 유사도를 수학적 식 1에 따라 계산하여 초기 인접 행렬(A^0)을 획득한다.

수학적 식 1

$$a_{ij} = \frac{x_i^T x_j}{\|x_i\|_2 \cdot \|x_j\|_2}$$

[0046] 여기서 a_{ij} 는 초기 인접 행렬(A^0)의 원소를 나타내고, x_i 와 x_j 는 각각 초기 특징 그래프(X^0)의 다수의 노드 벡터(x_1, \dots, x_T) 중 i 번째 노드 벡터와 j 번째 노드 벡터를 나타내고, T 는 전치 행렬을 나타내며, $\| \cdot \|_2$ 는 L2 놈(L2 norm) 함수를 나타낸다.

[0048] 즉 초기 인접 행렬(A^0)은 초기 특징 그래프(X^0)의 노드 벡터(x_1, \dots, x_T) 사이를 유사도에 따라 연결하는 에지의 가중치 행렬로 볼 수 있다.

[0049] 초기 그래프 생성부(100)는 초기 인접 행렬 획득부(130)와 초기 인접 행렬 획득부(140)에서 생성된 초기 특징 그래프(X^0)와 초기 인접 행렬(A^0)을 그래프 상관부(200)로 전달한다.

[0050] 도 3은 도 1의 비디오 요약 생성 장치의 그래프 상관부와 재귀 그래프 획득부의 상세 구성을 나타낸다.

[0051] 도 3을 참조하면, 그래프 상관부(200)는 직렬로 순차 연결된 다수의 그래프 상관 모듈(210 ~ 230)을 포함하여 구성된다. 다수의 그래프 상관 모듈(210 ~ 230) 각각은 패턴 추정 방식이 미리 학습된 그래프 컨볼루션 네트워크(GCN)로 구현되어, 초기 그래프 생성부(100)로부터 인가되는 초기 특징 그래프(X^0)와 초기 인접 행렬(A^0)에서 추정되는 패턴을 보정하여 보정 그래프(Z^1)를 추출한다. 그리고 그래프 상관부(200)는 추출된 보정 그래프(Z^1)를 재귀 그래프 획득부(300)로 전달하고, 재귀 그래프 획득부(300)로부터 전달되는 인접 행렬(A^1)과 특징 그래프(X^1)를 인가받아 다음 보정 그래프(Z^2)를 추출하는 과정을 반복한다.

[0052] 즉 그래프 상관부(200)는 이미 획득된 특징 그래프(X^{k-1})와 인접 행렬(A^{k-1})을 인가받아 수학적 식 2로 표현될 수 있

는 그래프 컨볼루션 네트워크(GCN)의 패턴 추정 방식에 따라 보정 그래프(Z^k)를 추출한다.

[0053] 이때, 다수의 그래프 상관 모듈(210 ~ 230) 중 초기단의 그래프 상관 모듈(210)은 특징 그래프(X^{k-1})와 인접 행렬(A^{k-1})을 인가받아 패턴을 추정하여 제1 중간 특징 그래프(X')를 추출하고, 이후 나머지 그래프 상관 모듈(220, 230)은 각각 이전단에서 추출된 중간 특징 그래프(X', X'')와 인접 행렬(A^{k-1})을 인가받아 패턴을 추정하여 제1 중간 특징 그래프(X'') 또는 보정 그래프(Z^k)를 추출한다.

수학식 2

[0054]
$$Z^k = \sigma(A^{k-1} X^{k-1} W)$$

[0055] 여기서 W 는 학습된 그래프 컨볼루션 네트워크(GCN)의 가중치를 나타내고, σ 는 ReLU(Rectified Linear Unit)와 같은 비선형 활성화 함수를 나타낸다.

[0056] 본 실시예에서는 일 예로 그래프 상관부(200)가 직렬로 순차 연결된 3개의 그래프 상관 모듈(210 ~ 230)을 포함하는 것으로 도시하였으나, 그래프 상관부(200)에 포함되는 그래프 상관 모듈(210 ~ 230)의 개수는 보정 그래프(Z^k)를 추출하는 성능을 실험적으로 분석하여 다양하게 조절될 수 있다. 즉 그래프 상관 모듈(210 ~ 230)의 개수에 의해 보정 그래프(Z^k)를 추출하는 성능이 조절될 수 있으며, 여기서는 실험적 결과로서 3개의 그래프 상관 모듈(210 ~ 230)을 포함하는 경우를 도시하였다.

[0057] 한편, 재귀 그래프 획득부(300)는 제1 및 제2 투사부(310, 320), 인접 보정값 획득부(330) 및 인접 행렬 획득부(340)를 포함할 수 있다.

[0058] 제1 및 제2 투사부(310, 320)는 각각 그래프 상관부(200)에서 획득된 보정 그래프(Z^k)를 인가받고, 보정 그래프(Z^k)가 서로 다른 선형 임베딩 공간에 투사되도록 서로 다르게 기 지정된 가중치 함수(W_θ, W_ϕ)를 인가된 보정 그래프(Z^k)에 가중하여 투영 보정 그래프($W_\theta Z^k, W_\phi Z^k$)를 획득한다.

[0059] 인접 보정값 획득부(330)는 제1 및 제2 투사부(310, 320)에서 획득된 투영 보정 그래프($W_\theta Z^k, W_\phi Z^k$)를 인가받아 보정 유사도(dA^k)를 수학식 3에 따라 획득한다.

수학식 3

[0060]
$$dA^k = \frac{(W_\theta Z^k)^T (W_\phi Z^k)}{\|W_\theta Z^k\|_2 \cdot \|W_\phi Z^k\|_2}$$

[0061] 여기서도 T 는 전치 행렬을 나타내며, $\| \cdot \|_2$ 는 L2 놈(L2 norm) 함수를 나타낸다.

[0062] 인접 행렬 획득부(340)는 이전 인접 행렬(A^{k-1})과 인접 보정값 획득부(330)에서 획득된 보정 유사도(dA^k)를 가산하여 그래프 상관부(200)가 다음 보정 그래프(Z^{k+1})를 추정하도록 하기 위한 인접 행렬(A^k)을 획득한다.

[0063] 즉 인접 행렬 획득부(340)는 초기 인접 행렬(A^0)로부터 이후 획득되는 보정 유사도(dA^k)를 누적 가산하여 다음 보정 그래프(Z^{k+1}) 추정을 위한 인접 행렬(A^k)을 획득한다. 따라서 만일 재귀 그래프 획득부(300)가 기 지정된 K 회만큼 반복적으로 재귀하여 인접 행렬(A^k)을 획득하는 경우, 최종적으로 획득되는 최종 인접 행렬(A^K)은 수학식

4와 같이 계산될 수 있다.

수학식 4

$$A^K = A^0 + \sum_{k=1}^K dA^k$$

[0064]

[0065] 이때 인접 행렬 획득부(340)는 이미 획득된 보정 그래프(Z^k)를 다음 특징 그래프(X^k)로서 인접 행렬(A^k)과 함께 그래프 상관부(200)로 전달 수 있다.

[0066] 다시 도 1을 참조하면, 키 프레임 추출부(400)는 상관 관계 추정부(410) 및 키 프레임 선택부(420)를 포함할 수 있다.

[0067] 상관 관계 추정부(410)는 패턴 추정 방식이 미리 학습된 그래프 컨볼루션 네트워크(GCN)으로 구현되어 인가되는 최종 보정 그래프(Z^k)와 최종 인접 행렬(A^k)의 패턴으로부터 최종 보정 그래프(Z^k)의 다수의 노드 벡터 사이의 상관 관계를 추정하여, 각 노드 벡터가 나타내는 입력 비디오(I)의 다수의 프레임 각각이 키 프레임(F_{key})일 가능성을 나타내는 확률로 구성되는 키 프레임 확률맵(Y)을 시멘틱 유사도에 따라 추출한다.

[0068] 상관 관계 추정부(410)는 수학식 5로 표현되는 그래프 컨볼루션 네트워크(GCN)의 패턴 추정 방식에 따라 키 프레임 확률맵(Y)을 추출할 수 있다.

수학식 5

$$Y = \sigma(A^K Z^K W_C)$$

[0069]

[0070] 여기서 W_C 는 학습된 그래프 컨볼루션 네트워크(GCN)의 가중치를 나타내고, σ 는 활성화 함수를 나타낸다.

[0071] 키 프레임 확률맵(Y)은 프레임의 개수(T)와 각 노드 벡터에 대응하는 프레임이 키 프레임일 확률과 키 프레임이 아닐 확률을 각각 표현하도록 $T \times 2$ 의 크기로 획득될 수 있다.

[0072] 그리고 키 프레임 선택부(420)는 키 프레임 확률맵(Y)으로부터 입력 비디오(I)의 다수의 프레임 중 키 프레임일 확률이 높고 키 프레임이 아닐 확률이 낮은 프레임을 기지정된 방식으로 선출하여 키 프레임(F_{key})으로 선택한다.

[0073] 키 프레임 선택부(420)는 일 예로 키 프레임(F_{key})을 기지정된 개수로 선택하거나, 키 프레임일 확률이 기지정된 제1 문턱값 이상이거나, 키 프레임이 아닐 확률이 기지정된 제2 문턱값 이하인 프레임을 키 프레임(F_{key})으로 선택할 수 있다.

[0074] 한편, 그래프 컨볼루션 네트워크(GCN)를 포함하여 구현되는 그래프 상관부(200)와 키 프레임 추출부(400)는 미리 학습될 필요가 있다. 따라서 본 실시예에 따른 비디오 요약 생성 장치는 그래프 상관부(200)와 키 프레임 추출부(400)를 학습시키기 위한 학습부(미도시)를 더 포함할 수 있다.

[0075] 학습부는 기지정된 다수의 손실 함수에 따라 계산되는 손실이 최소화되도록 그래프 상관부(200)와 키 프레임 추출부(400)를 학습시킬 수 있다.

[0076] 본 실시예에서 학습부는 지도 학습(Supervised learning) 방식과 비지도 학습(Unsupervised learning) 방식 중 하나로 학습을 수행할 수 있다.

[0077] 우선 지도 학습 방식으로 학습을 수행하는 경우, 학습부는 다수의 사용자에 의해 정규화 및 평균화된 중요도 점수를 기반으로 진리 키 프레임 확률맵(Y^*)이 미리 획득된 비디오를 학습 데이터로 획득하여 입력 비디오(I)로서

입력하고, 학습 데이터의 진리 키 프레임 확률맵(Y^*)과 비디오 요약 생성 장치에서 획득된 키 프레임 확률맵(Y)을 비교하여 지도 학습 손실(L_{sup})을 계산하여 역전파하여 학습을 수행한다.

[0078] 지도 학습 방식에서 학습부는 분류 손실(L_c)과 희소성 손실(L_s), 복원 손실(L_r) 및 다양성 손실(L_d)을 각각 계산한다.

[0079] 분류 손실(L_c)은 비디오 요약 생성 장치에서 생성된 키 프레임 확률맵(Y)과 진리 키 프레임 확률맵(Y^*) 사이의 이진 크로스 엔트로피 손실(binary cross-entropy loss)로 수학적 식 6에 따라 계산될 수 있다.

수학적 식 6

[0080]
$$\mathcal{L}_c = -\frac{1}{T} \sum_{t=1}^T w_t [y_t^* \log(y_t) + (1 - y_t^*) \log(1 - y_t)]$$

[0081] 여기서 y_t^* 는 t번째 프레임의 진리 레이블이고, w_t 는 t번째 프레임의 가중치로서, $w_t = \text{median_freq}/\text{freq}(s)$ 로 계산될 수 있다. $\text{freq}(s)$ 는 키 프레임 수를 총 프레임 수로 나눈 값이고, median_freq 는 키 프레임 발생 빈도의 중간값을 나타낸다.

[0082] 희소성 손실(L_s)은 입력 영상(I)의 다수의 프레임에서 키 프레임의 수가 희소해야 한다는 가정에서 도출되는 손실로서, 수학적 식 7과 같이 인접 행렬의 각 원소($a_{ij} \in A^k$)에 L1 놈 함수를 적용하여 획득될 수 있다.

수학적 식 7

[0083]
$$\mathcal{L}_s = \sum_{i=1}^T \sum_{j=1}^T \|a_{ij}\|_1$$

[0084] 복원 손실(L_r)은 키 프레임은 시각적으로 다양하게 존재해야 한다는 가정에 따라 추가되는 손실로서, 키 프레임 확률맵(Y)을 별도로 하나의 그래프 컨볼루션 모듈을 포함하고 미리 학습된 그래프 컨볼루션 네트워크에 인가하여 초기 특징 그래프(X^0)와 동일 크기의 추가 보정 특징 그래프(\hat{Z})를 생성하고, 생성된 추가 보정 특징 그래프(\hat{Z})와 초기 특징 그래프(X^0) 사이의 평균 제곱 오차(MSE)를 수학적 식 8과 같이 계산하여 획득할 수 있다.

수학적 식 8

[0085]
$$\mathcal{L}_r = \frac{1}{2N} \sum_{i=1}^N \|x_i - \hat{z}_i\|_2^2$$

[0086] 마지막으로 다양성 손실(L_d)은 추가 보정 특징 그래프(\hat{Z})에서 키 프레임으로 선택되는 노드들 사이의 반발 규격화(repelling regularizer)를 적용하여 수학적 식 9에 따라 획득할 수 있다.

수학식 9

$$\mathcal{L}_d = \frac{1}{N(N-1)} \sum_{i \in N} \sum_{j \in N, j \neq i} f(\hat{\mathbf{z}}_i, \hat{\mathbf{z}}_j)$$

[0087]

[0088]

그리고 학습부는 수학식 6 내지 9에 따라 계산된 분류 손실(\mathcal{L}_c)과 희소성 손실(\mathcal{L}_s), 복원 손실(\mathcal{L}_r) 및 다양성 손실(\mathcal{L}_d)로부터 지도 학습 시의 총 손실(\mathcal{L}_{sup})을 수학식 10과 같이 계산하여 역전파한다.

수학식 10

$$\mathcal{L}_{sup} = \mathcal{L}_c + \lambda \cdot \mathcal{L}_s + \alpha \cdot \mathcal{L}_d + \beta \cdot \mathcal{L}_r$$

[0089]

[0090]

여기서 λ , α 및 β 는 각 손실의 중요도를 조절하기 위한 가중치이다.

[0091]

한편 학습부는 비지도 학습을 수행하는 경우, 학습 데이터가 존재하지 않으므로, 수학식 10에서 분류 손실(\mathcal{L}_c)을 제외하고, 희소성 손실(\mathcal{L}_s)과 복원 손실(\mathcal{L}_r) 및 다양성 손실(\mathcal{L}_d)로부터 비지도 학습 시의 총 손실(\mathcal{L}_{unsup})을 수학식 11과 같이 계산하여 역전파한다.

수학식 11

$$\mathcal{L}_{unsup} = \mathcal{L}_s + \alpha \cdot \mathcal{L}_d + \beta \cdot \mathcal{L}_r$$

[0092]

[0093]

도 4는 재귀 반복 횟수에 따라 추출된 키 프레임의 일례를 나타낸다.

[0094]

도 4에서 (a)는 1회 재귀 반복($K=1$)을 수행한 경우를 나타내고, (b)는 3회 재귀 반복($K=3$)을 수행한 경우를 나타내며, (c)는 5회 재귀 반복($K=5$)을 수행한 경우를 나타낸다. 그리고 (a) 내지 (c) 각각에서 좌측 상단 프레임은 입력 비디오(I)에서 선택된 키 프레임(F_{key})을 나타내고, 나머지는 키 프레임(F_{key})을 제외하고 키 프레임 확률이 높은 순서로 추출된 프레임을 나타낸다. 그리고 각 프레임과 함께 표시된 s 는 다수의 사용자들에 의해 미리 주석되고 정규화 및 평균화된 중요도 점수를 나타낸다.

[0095]

도 4에 도시된 바와 같이, 의미적으로 유사한 프레임, 즉 시멘틱 유사도가 높은 키 프레임들 사이의 연결은 재귀 반복 횟수가 증가될수록 점차로 강화됨을 알 수 있다. 다만 재귀 반복 횟수가 증가할수록 연산량과 연산 시간은 증가하는데 반해, 키 프레임들 사이의 시멘틱 유사도는 크게 강화되지 않으므로, 효율성을 위해 재귀 반복 횟수는 실험을 통해 미리 지정될 수 있다.

[0096]

도 5는 본 발명의 일 실시예에 따른 비디오 요약 생성 방법을 나타낸다.

[0097]

도 1 내지 도 3을 참조하여 도 5의 비디오 요약 생성 방법을 설명하면, 우선 비디오 요약이 생성되어야 하는 다수 프레임의 입력 비디오(I)를 인가받아 각 프레임에서 추출된 특징맵(X)을 노드 벡터로 간주하여 초기 특징 그래프(X^0)를 생성하며, 노드 벡터 사이의 유사도를 계산하여 초기 인접 행렬(A^0)을 생성한다(S10).

[0098]

초기 특징 그래프(X^0)와 초기 인접 행렬(A^0)을 생성하는 단계(S10)에서는 먼저 다수의 프레임으로 구성된 입력 비디오(I)를 획득한다(S11). 그리고 미리 학습된 패턴 추정 방식에 따라 입력 비디오(I)의 다수의 프레임 각각에 대한 특징을 추출하여 다수의 특징맵(X)을 생성한다(S12). 다수의 특징맵(X)이 생성되면, 다수의 특징맵(X) 각각을 다수의 노드와 다수의 노드를 서로 연결하는 다수의 에지로 구성되는 그래프의 노드를 나타내는 노드 벡터로 간주하여 초기 특징 그래프(X^0)를 획득한다(S13).

[0099]

그리고 다수의 특징맵(X)에 각각 대응하는 다수의 노드를 포함하는 초기 특징 그래프(X^0)가 획득되면, 다수의 노

드 사이의 유사도를 수학적 식 1에 따라 계산하여 초기 인접 행렬(A^0)을 획득한다(S14).

- [0100] 초기 특징 그래프(X^0)와 초기 인접 행렬(A^0)이 획득되면, 초기 특징 그래프(X^0)와 초기 인접 행렬(A^0)을 패턴 추정 방식이 미리 학습된 다수의 그래프 컨볼루션 네트워크에 인가하여 초기 특징 그래프(X^0)의 다수의 노드 사이의 패턴을 보정하여 보정 그래프(Z^k)를 추출한다(S20).
- [0101] 그리고 재귀 반복 횟수(k)가 기지정된 횟수(K) 이상($k \geq K$)인지 판별한다(S30).
- [0102] 만일 재귀 반복 횟수(k)가 기지정된 횟수(K) 미만이면, 보정 그래프(Z^k)를 서로 다른 선형 임베딩 공간에 투사하고, 투사된 2개의 보정 그래프($W_\theta Z^k, W_\phi Z^k$) 사이의 보정 유사도(dA^k)를 계산하여 다음 보정 그래프(Z^{k+1})를 획득하기 위한 인접 행렬(A^k)을 획득한다(S40).
- [0103] 구체적으로, 재귀 반복 횟수(k)가 기지정된 횟수(K) 미만이면, 서로 다른 기지정된 가중치 함수(W_θ, W_ϕ)를 보정 그래프(Z^k)에 가중하여, 보정 그래프(Z^k)를 서로 다른 선형 임베딩 공간에 투사한다(S41). 그리고 2개의 투영 보정 그래프($W_\theta Z^k, W_\phi Z^k$) 사이의 유사도를 계산하여 보정 유사도(dA^k)를 획득한다(S42). 보정 유사도(dA^k)가 획득되면, 이전 인접 행렬(A^{k-1})과 보정 유사도(dA^k)를 가산하여 다음 인접 행렬(A^k)을 획득하고, 보정 그래프(Z^k)를 다음 특징 그래프(X^k)로서 적용한다(S43). 획득된 다음 인접 행렬(A^k)과 다음 특징 그래프(X^k)는 패턴 추정 방식이 미리 학습된 다수의 그래프 컨볼루션 네트워크에 재귀되어 인가됨으로써, 다음 인접 행렬(A^k)과 다음 특징 그래프(X^k)로부터 다음 보정 그래프(Z^{k+1})가 추출되도록 한다(S20).
- [0104] 그러나 재귀 반복 횟수(k)가 기지정된 횟수(K) 이상이면, 최종적으로 획득된 최종 인접 행렬(A^K)과 최종 보정 그래프(Z^K)로부터 미리 학습된 패턴 추정 방식에 따라 최종 보정 그래프(Z^K)의 각 노드 사이의 시멘틱 유사도를 추정하여 키 프레임(F_{key})을 선택한다(S50).
- [0105] 키 프레임(F_{key})을 선택하기 위해서는 먼저 패턴 추정 방식이 미리 학습된 별도의 그래프 컨볼루션 네트워크에 우선 최종적으로 획득된 최종 인접 행렬(A^K)과 최종 보정 그래프(Z^K)를 인가하여, 최종 보정 그래프(Z^K)의 각 노드 사이의 시멘틱 유사도 패턴에 따라 입력 비디오(I)의 다수의 프레임 각각이 키 프레임(F_{key})일 확률을 나타내는 키 프레임 확률맵(Y)을 추출한다(S51). 그리고 추출된 키 프레임 확률맵(Y)로부터 입력 비디오(I)의 다수의 프레임 각각이 키 프레임될 확률을 분석하여 다수의 키 프레임(F_{key})을 선택한다(S52). 여기서 선택된 다수의 키 프레임(F_{key})은 입력 비디오(I)의 다수의 프레임 중 시멘틱 유사성이 높은 프레임들로서 비디오 요약으로 볼 수 있다.
- [0106] 도 6은 본 실시예에 따른 비디오 요약 생성 방법과 기존의 비디오 요약 생성 방법에 따라 생성된 비디오 요약을 비교한 결과를 나타낸다.
- [0107] 도 6에서 (a) 와 (c)는 기존에 공개된 SUM-FCN 기법에 따라 비디오 요약을 생성한 결과를 나타내고, (b)와 (d)는 본 실시예에 따른 비디오 요약 기법인 SumGraph 기법에 따라 비디오 요약을 생성한 결과를 나타낸다. 그리고 도 6의 (a) 내지 (d)에서 선택된 프레임 상부에 표시된 막대 그래프는 학습 데이터에 레이블된 중요도 값을 나타낸다. 막대 그래프에서 붉은색으로 표시된 영역이 선택된 각 기법에 따라 선택된 키 프레임을 나타낸다.
- [0108] 도 6에 도시된 바와 같이, 본 실시예에 따른 비디오 요약 생성 방법은 기존의 기법에 비해 더욱 시각적으로 다양하고, 레이블된 중요도 값의 피크에 대응하는 프레임을 키 프레임으로 선택함을 알 수 있다. 이는 본 실시예에 따른 비디오 요약 생성 방법이 입력 비디오에서 최적의 유의미한 요약을 생성하기 위해, 다수의 프레임 사이의 시멘틱 관계를 정확하게 추정할 수 있음을 나타낸다.
- [0109] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 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 비분

리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.

[0110] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

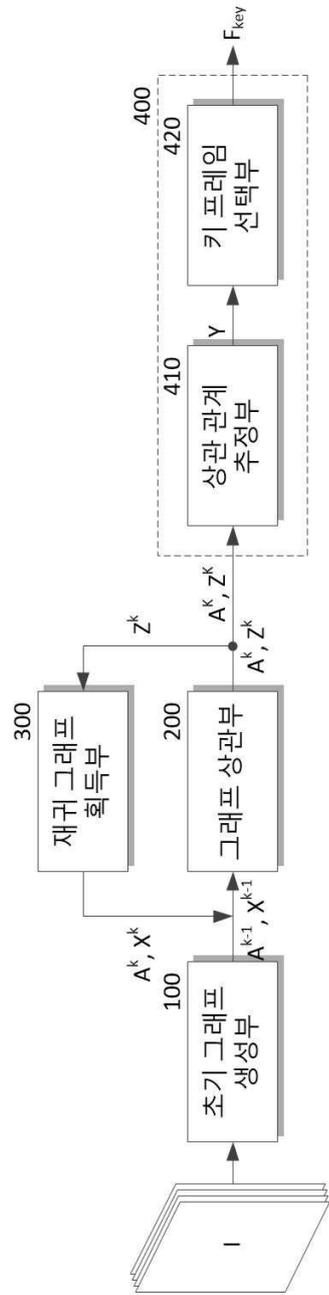
[0111] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

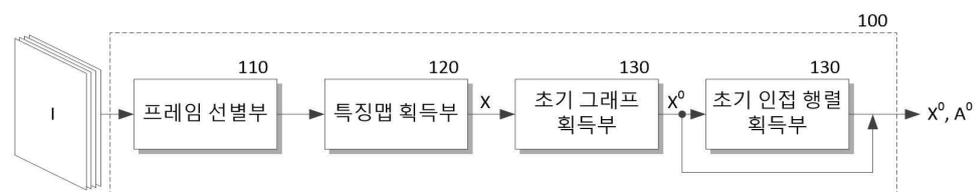
- | | | |
|--------|----------------------|-----------------|
| [0112] | 100: 초기 그래프 생성부 | 110: 프레임 선별부 |
| | 120: 특징맵 획득부 | 130: 초기 그래프 획득부 |
| | 140: 초기 인접 행렬 획득부 | 200: 그래프 상관부 |
| | 210 ~ 230: 그래프 상관 모듈 | 300: 재귀 그래프 획득부 |
| | 310: 제1 투사부 | 320: 제2 투사부 |
| | 330: 인접 보정값 획득부 | 340: 인접 행렬 획득부 |
| | 400: 키 프레임 추출부 | 410: 상관 관계 추정부 |
| | 420: 키 프레임 선택부 | |

도면

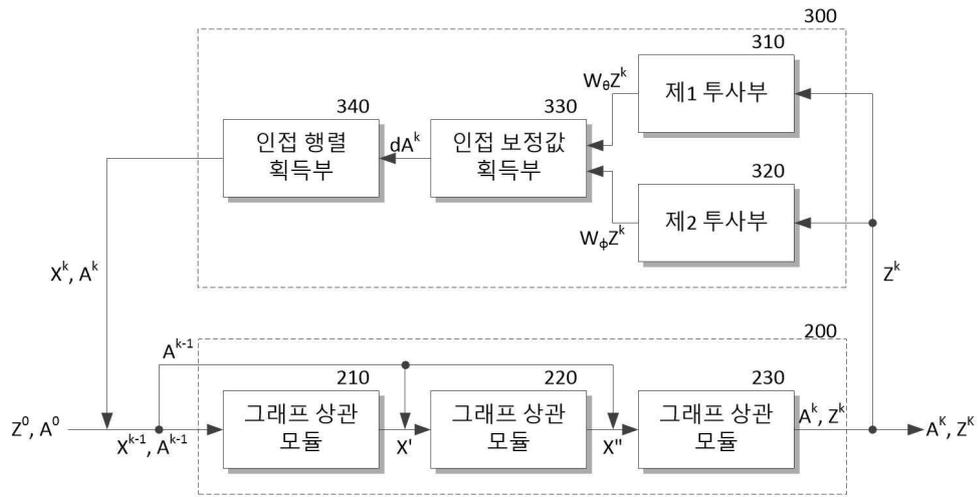
도면1



도면2



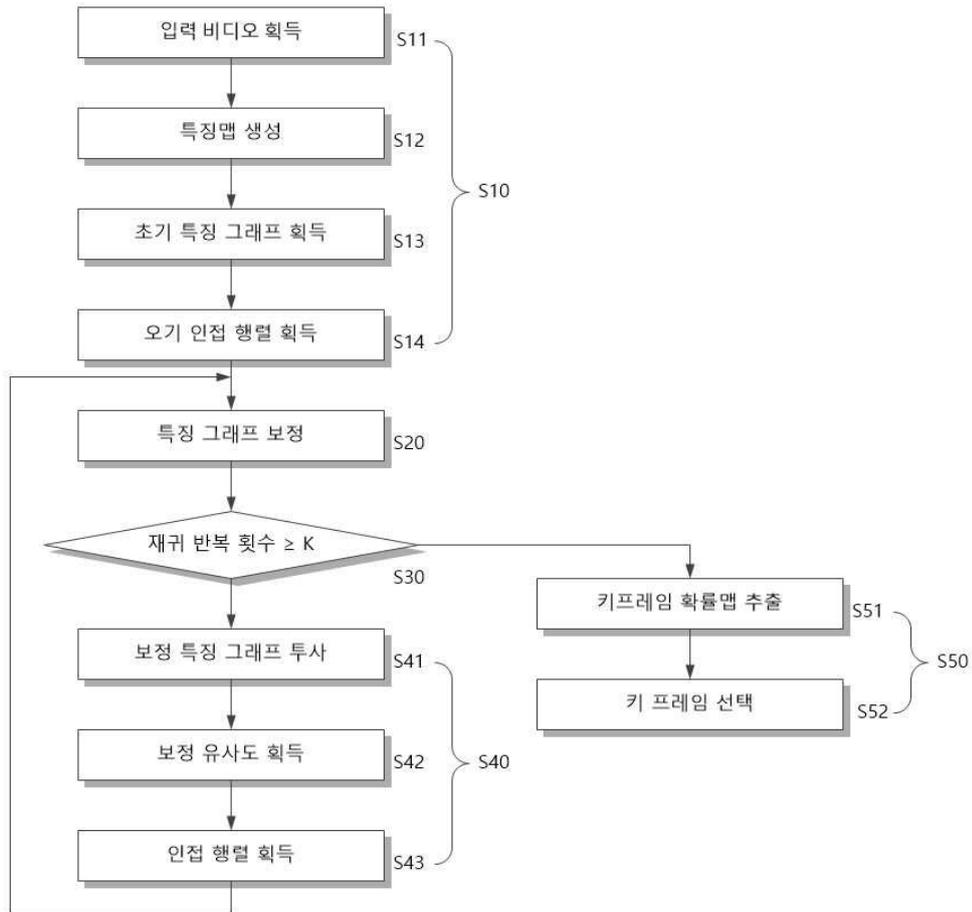
도면3



도면4



도면5



도면6

