



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0071056
(43) 공개일자 2024년05월22일

(51) 국제특허분류(Int. Cl.)
G06F 16/783 (2019.01) G06F 17/16 (2006.01)
G06V 10/82 (2022.01) G06V 20/40 (2022.01)
(52) CPC특허분류
G06F 16/783 (2019.01)
G06F 17/16 (2013.01)
(21) 출원번호 10-2022-0152725
(22) 출원일자 2022년11월15일
심사청구일자 2022년11월15일

(71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
최유경
서울특별시 광진구 능동로 209, 세종대학교 대양 AI센터 601호(군자동)
조원
서울특별시 광진구 광나루로14길 26, 아젠시아 503호(화양동)
(74) 대리인
송인호, 최관락

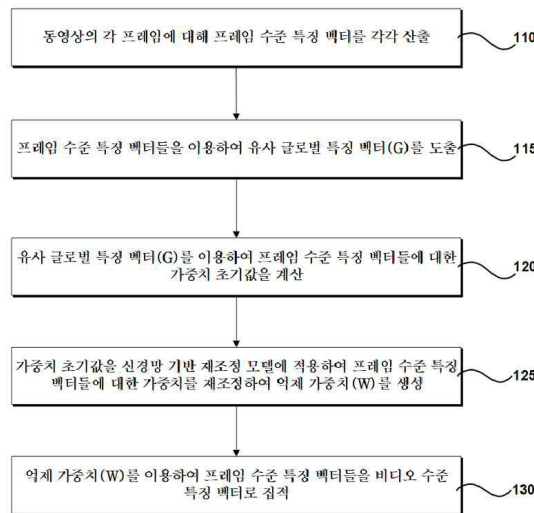
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 동영상 검색을 위한 동영상 특징 서술자 추출 방법 및 그 장치

(57) 요약

동영상 검색을 위한 동영상 특징 추출 방법 및 그 장치가 개시된다. 동영상 검색을 위한 동영상 특징 추출 방법은 동영상에 대해 T개의 프레임 각각에 대한 프레임 수준 특징 벡터들을 각각 추출하는 단계; 상기 추출된 T개의 프레임 수준 특징 벡터를 이용하여 유사 글로벌 특징 벡터(G)를 계산하는 단계; 상기 유사 글로벌 특징 벡터(G)를 이용하여 상기 프레임 수준 특징 벡터의 가중치 초기값을 계산하는 단계; 상기 가중치 초기값을 신경망 기반 재조정 모델에 적용하여 상기 프레임 수준 특징 벡터의 가중치를 재조정하는 단계; 및 상기 재조정된 가중치를 이용하여 상기 프레임 수준 특징 벡터를 비디오 수준 특징 벡터로 집계하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06V 10/82 (2022.01)

G06V 20/46 (2022.01)

G06V 20/49 (2022.01)

(72) 발명자

임근택

서울특별시 광진구 아차산로21길 86(화양동)

이광진

서울특별시 광진구 천호대로137길 9, 한울타워 A동
502호(구의동)

김현우

서울특별시 송파구 올림픽로43길 34, 105동 1001
호(풍납동, 잠실 올림픽공원 아이파크)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711159902
과제번호	2021-0-02067-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대인공지능핵심원천기술개발
연구과제명	다목적 비디오 검색을 위한 차세대 인공지능경망 기술 개발
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2022.04.01 ~ 2023.03.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711152786
과제번호	2021-0-00755-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발
연구과제명	데이터 규모 확장과 정확성 향상을 위한 다크 데이터 분석 기술 개발
기 여 율	1/2
과제수행기관명	(주)와이즈넷
연구기간	2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

동영상에 대한 T개의 프레임 각각에 대해 프레임 수준 특징 벡터들을 각각 추출하는 단계;

상기 추출된 T개의 프레임 수준 특징 벡터를 이용하여 유사 글로벌 특징 벡터(G)를 계산하는 단계;

상기 유사 글로벌 특징 벡터(G)를 이용하여 상기 프레임 수준 특징 벡터의 가중치 초기값을 계산하는 단계;

상기 가중치 초기값을 신경망 기반 재조정 모델에 적용하여 상기 프레임 수준 특징 벡터의 가중치를 재조정하는 단계; 및

상기 재조정된 가중치를 이용하여 상기 프레임 수준 특징 벡터를 비디오 수준 특징 벡터로 집계하는 단계를 포함하는 동영상 검색을 위한 동영상 특징 추출 방법.

청구항 2

제1 항에 있어서,

상기 유사 글로벌 특징 벡터(G)는 상기 T개의 프레임 수준 특징 벡터를 시간적-공간적 평균 풀링(spatio-temporal average pooling) 연산하여 계산되는 것을 특징으로 하는 동영상 검색을 위한 동영상 특징 추출 방법.

청구항 3

제1 항에 있어서,

상기 가중치 초기값은 하기 수학식을 이용하여 계산되는 것을 특징으로 하는 동영상 특징 추출 방법.

$$w_t^{\text{initial}} = \frac{G \cdot \beta_s(f_t)}{\|G\| \|\beta_s(f_t)\|}$$

β_s

여기서, G는 유사 글로벌 특징 벡터를 나타내고, β_s 는 프레임 수준 특징 벡터(f_t)의 공간적 평균 풀링

연산(spatial average pooling)을 나타내고, $\| \ \|$ 는 벡터 놈(vector norm)을 나타냄.

청구항 4

제1 항에 있어서,

상기 비디오 수준 특징 벡터는 하기 수학적식을 이용하여 산출되는 것을 특징으로 하는 동영상 특징 추출 방법.

$$V = \alpha \left(\{f_t \otimes w_t\}_{t=1}^T \right)$$

$$w_t \quad f_t$$

여기서, w_t 는 재조정 가중치를 나타내고, f_t 는 t번째 프레임 수준 특징 벡터를 나타내며, T는 전체

프레임 수를 나타내며, α 는 미리 정의된 집적 함수를 나타내고, $\otimes$ 는 브로드캐스팅(broadcasting) 곱 연산을 나타냄.

청구항 5

제1 항 내지 제4 항 중 어느 하나의 항에 따른 방법을 수행하기 위한 프로그램 코드를 기록한 컴퓨터로 판독 가능한 기록매체.

청구항 6

동영상의 각 프레임을 백본 네트워크 모델에 적용하여 프레임 수준 특징 벡터를 각각 추출하는 특징 추출부;

상기 프레임 수준 특징 벡터를 이용하여 유사 글로벌 특징 벡터(G)를 계산하고, 상기 유사 글로벌 특징 벡터(G)를 이용하여 상기 프레임 수준 특징 벡터의 가중치 초기값을 도출한 후 상기 가중치 초기값을 신경망 기반 재조정 모델에 적용하여 상기 프레임 수준 특징 벡터의 가중치를 재조정하는 가중치 생성부; 및

상기 재조정된 가중치를 이용하여 상기 프레임 수준 특징 벡터를 비디오 수준 특징 벡터로 집적하는 임베딩부를 포함하는 동영상 검색을 위한 동영상 특징 서술자 추출 장치.

청구항 7

제6 항에 있어서,

상기 유사 글로벌 특징 벡터(G)는 상기 T개의 프레임 수준 특징 벡터를 시간적-공간적 평균 풀링(spatio-temporal average pooling) 연산하여 계산되는 것을 특징으로 하는 동영상 검색을 위한 동영상 특징 서술자 추출 장치.

청구항 8

제6 항에 있어서,

상기 가중치 생성부는,

상기 각 프레임 수준 특징 벡터와 상기 유사 글로벌 특징 벡터(G)와의 유사도를 계산하여 상기 가중치 초기값을 각각 계산하는 것을 특징으로 하는 동영상 검색을 위한 동영상 특징 서술자 추출 장치.

청구항 9

제6 항에 있어서,

상기 신경망 기반 재조정 모델은 특정 프레임에 연속된 인접 프레임의 가중치들을 고려하여 가중치를 재조정하는 것을 특징으로 하는 동영상 검색을 위한 동영상 특징 서술자 추출 장치.

발명의 설명

기술 분야

[0001] 본 발명은 동영상 검색을 위한 동영상 특징 서술자 추출 방법 및 그 장치에 관한 것이다.

배경 기술

[0003] 정보 검색은 대규모 컬렉션에서 가장 관련성이 높은 정보를 찾는 것으로 정의된다. 정보 검색은 문서 내에서 텍스트를 찾는 것에서 영상 세트내에서 영상 클립을 찾는 것으로 발전했다. 최근 동영상 스트리밍 시장이 급성장하면서 동영상 세트내에서 원하는 동영상 클립을 찾기 위한 CBVR(content-based video retrieval)에 대한 연구가 진행되고 있다.

[0004] 콘텐츠 기반 동영상 검색 기반의 핵심 기술은 자르지 않은 동영상을 포함하여 길이가 다른 동영상간의 유사도를 측정하는 것이다. 두 영상간의 유사도를 측정하는 기본 단위에 따라 프레임 수준 특징 기반 접근 방법과 동영상 수준 특징 기반 접근 방식의 두 가지 스트림으로 나뉜다.

[0005] 프레임 수준 기능 기반 접근 방식은 두 비디오의 프레임 수준 기능 간의 유사성을 집계하여 비디오 간 유사성을 계산하는 반면, 비디오 수준 기능 기반 접근 방식은 각 비디오를 단일 기능으로 설명하고 이를 기반으로 비디오 간 유사성을 계산한다.

[0006] 프레임 수준 기능 기반 접근 방식은 각 프레임을 직접 비교하며 비디오 지속 시간 및 스트리밍 여부와 같은 요소에 덜 의존하여 비교적 정확한 검색이 가능하나 처리 속도와 메모리가 비싸다. 이에 비해 비디오 수준의 기능 기반 접근 방식은 단일 기능 쌍 간에 한 번의 작업만 필요하므로 처리 속도와 메모리 측면에서 더 효율적이다. 그러나, 하나의 비디오에서 하나의 기능으로 많은 콘텐츠를 줄이기가 어렵기 때문에 이 접근 방식은 일반적으로 부정확한 검색 기능을 보이는 단점이 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 동영상 검색을 위한 동영상 특징 서술자 추출 방법 및 그 장치를 제공하기 위한 것이다.

[0009] 또한, 본 발명은 동영상 프레임들로부터 산출되는 프레임 수준 특징 벡터들을 비디오 수준 특징 벡터로 집계되, 동영상 전체 주제를 나타내는 글로벌 특징 벡터를 이용하여 전체 비디오 맥락에 기반하여 비디오 수준 서술자의 표현력을 개선할 수 있는 동영상 검색을 위한 동영상 특징 서술자 추출 방법 및 그 장치를 제공하기 위한 것이다.

[0010] 또한, 본 발명은 프레임 수준 서술자들의 중요도 조정을 통해 전체 비디오 맥락에 상응하도록 비디오 수준 서술자의 표현력을 개선할 수 있는 동영상 검색을 위한 동영상 특징 서술자 추출 방법 및 그 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0012] 본 발명의 일 측면에 따르면 동영상 검색을 위한 동영상 특징 서술자 추출 방법이 제공된다.

[0013] 본 발명의 일 실시예에 따르면, 동영상에 대한 T개의 프레임 각각에 대해 프레임 수준 특징 벡터들을 각각 추출하는 단계; 상기 추출된 T개의 프레임 수준 특징 벡터를 이용하여 유사 글로벌 특징 벡터(G)를 계산하는 단계;

상기 유사 글로벌 특징 벡터(G)를 이용하여 상기 프레임 수준 특징 벡터의 가중치 초기값을 계산하는 단계; 상기 가중치 초기값을 신경망 기반 재조정 모델에 적용하여 상기 프레임 수준 특징 벡터의 가중치를 재조정하는 단계; 및 상기 재조정된 가중치를 이용하여 상기 프레임 수준 특징 벡터를 비디오 수준 특징 벡터로 집계하는 단계를 포함하는 동영상 검색을 위한 동영상 특징 추출 방법이 제공될 수 있다.

[0014] 상기 유사 글로벌 특징 벡터(G)는 상기 T개의 프레임 수준 특징 벡터를 시간적-공간적 평균 풀링(spatio-temporal average pooling) 연산하여 계산될 수 있다.

[0015] 상기 가중치 초기값은 하기 수학식을 이용하여 계산되되,

$$w_t^{\text{initial}} = \frac{G \cdot \beta_s(f_t)}{\|G\| \|\beta_s(f_t)\|}$$

[0016]

[0017] 여기서, G는 유사 글로벌 특징 벡터를 나타내고, β_s 는 프레임 수준 특징 벡터(f_t)의 공간적 평균 풀링 연산(spatial average pooling)을 나타내고, $\| \cdot \|$ 는 벡터 놈(vector norm)을 나타낸다.

[0018] 상기 비디오 수준 특징 벡터는 하기 수학식을 이용하여 산출되되,

$$V = \alpha \left(\{f_t \otimes w_t\}_{t=1}^T \right)$$

[0019]

[0020] 여기서, w_t 는 재조정 가중치를 나타내고, f_t 는 t번째 프레임 수준 특징 벡터를 나타내며, T는 전체 프레임 수를 나타내며, α 는 미리 정의된 집계 함수를 나타내고, $\otimes$ 는 브로드캐스팅(broadcasting) 곱 연산을 나타낸다.

[0022] 본 발명의 다른 측면에 따르면 동영상 검색을 위한 동영상 특징 서술자 추출 장치가 제공된다.

[0023] 본 발명의 일 실시예에 따르면, 동영상의 각 프레임을 백본 네트워크 모델에 적용하여 프레임 수준 특징 벡터를 각각 추출하는 특징 추출부; 상기 프레임 수준 특징 벡터를 이용하여 유사 글로벌 특징 벡터(G)를 계산하고, 상기 유사 글로벌 특징 벡터(G)를 이용하여 상기 프레임 수준 특징 벡터의 가중치 초기값을 도출한 후 상기 가중치 초기값을 신경망 기반 재조정 모델에 적용하여 상기 프레임 수준 특징 벡터의 가중치를 재조정하는 가중치 생성부; 및 상기 재조정된 가중치를 이용하여 상기 프레임 수준 특징 벡터를 비디오 수준 특징 벡터로 집계하는 임베딩부를 포함하는 동영상 검색을 위한 동영상 특징 서술자 추출 장치가 제공될 수 있다.

[0024] 상기 유사 글로벌 특징 벡터(G)는 상기 T개의 프레임 수준 특징 벡터를 시간적-공간적 평균 풀링(spatio-temporal average pooling) 연산하여 계산될 수 있다.

[0025] 상기 가중치 생성부는, 상기 각 프레임 수준 특징 벡터와 상기 유사 글로벌 특징 벡터(G)와의 유사도를 계산하여 상기 가중치 초기값을 각각 계산할 수 있다.

[0026] 상기 신경망 기반 재조정 모델은 특정 프레임에 연속된 인접 프레임의 가중치들을 고려하여 가중치를 재조정할

수 있다.

발명의 효과

[0028] 본 발명의 일 실시예에 따른 동영상 검색을 위한 동영상 특징 서술자 추출 방법 및 그 장치를 제공함으로써, 동영상 프레임들로부터 산출되는 프레임 수준 특징 벡터들을 비디오 수준 특징 벡터로 집적하되, 전체 비디오의 맥락 및 프레임 수준 서술자들의 중요도에 기반하여 비디오 수준 서술자의 표현력을 개선할 수 있는 이점이 있다.

도면의 간단한 설명

[0030] 도 1은 본 발명의 일 실시예에 따른 동영상 검색을 위한 동영상 특징 서술자 추출 방법을 나타낸 순서도.
 도 2는 본 발명의 일 실시예에 따른 비디오 수준 특징 벡터를 생성하는 전체 구조를 도시한 도면.
 도 3은 본 발명의 일 실시예에 따른 신경망 기반 재조정 모델의 상세 아키텍처를 도시한 도면.
 도 4는 종래와 본 발명의 일 실시예에 따른 비행기 추락 사고에 대한 비디오 수준 특징 벡터를 생성한 결과를 비교한 도면.
 도 5는 종래와 본 발명의 일 실시예에 따른 건물 화재 사고에 대한 비디오 수준 특징 벡터를 생성한 결과를 비교한 도면.
 도 6은 본 발명의 일 실시예에 따른 동영상 검색을 위한 동영상 특징 서술자 추출 장치의 구성을 개략적으로 도시한 블록도.

발명을 실시하기 위한 구체적인 내용

[0031] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0032] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.

[0034] 도 1은 본 발명의 일 실시예에 따른 동영상 검색을 위한 동영상 특징 서술자 추출 방법을 나타낸 순서도이고, 도 2는 본 발명의 일 실시예에 따른 비디오 수준 특징 벡터를 생성하는 전체 구조를 도시한 도면이고, 도 3은 본 발명의 일 실시예에 따른 신경망 기반 재조정 모델의 상세 아키텍처를 도시한 도면이며, 도 4는 종래와 본 발명의 일 실시예에 따른 비행기 추락 사고에 대한 비디오 수준 특징 벡터를 생성한 결과를 비교한 도면이고, 도 5는 종래와 본 발명의 일 실시예에 따른 건물 화재 사고에 대한 비디오 수준 특징 벡터를 생성한 결과를 비교한 도면이다.

[0035] 단계 110에서 동영상 특징 서술자 추출 장치(100)는 동영상의 각 프레임에 대해 프레임 수준 특징 벡터를 각각 산출한다.

[0036] 예를 들어, 동영상 특징 서술자 추출 장치(100)는 동영상 X의 각 프레임들에 대해 시각적 특징 정보를 요약하는 특징 벡터를 각각 추출한다. 여기서, 특징 벡터는 프레임 수준 특징 벡터로 동영상 내 국소 특징점 서술자(예를 들어, SIFT)를 집적하거나 신경망 기반 백본 네트워크 모델을 통해 산출될 수도 있다.

[0037] 프레임 수준 특징 벡터는 복수의 프레임을 참조하여 산출될 수 있다. 예를 들어, 동영상 특징 서술자 추출 장치(100)는 프레임 수준 특징 벡터(f_t)를 추출하는 경우, t번째 프레임 전후 N개의 프레임을 참조하여 추출할 수 있다.

[0038] 이하에서는 이해와 설명의 편의를 도모하기 위해 프레임 수준 특징 벡터(f_t)가 단일 프레임(X_t)으로부터 추출되는 L_N -iMAC 특징 벡터인 것을 가정하여 설명하기로 한다. 여기서, L_N -iMAC 특징 벡터는 백본 네트워크 모델의 K개 중간 레이어 출력에 대해 미리 정해진 $S \times S$ 개 구획별로 영역별 최대 활성화 풀링(regional maximum activation pooling)을 수행한 뒤, 풀링된 MAC 특징 벡터를 채널 방향으로 결합(concatenation)하는 방식으로 산출된다. 따라서, 특징 벡터의 차원은 $S \times S \times C$ 가 된다. T개 프레임에 대한 특징 벡터 모음을 F로 표기하며, 그 차원은 $S \times S \times C \times T$ 로 표기될 수 있다.

[0039] 단계 115에서 동영상 특징 서술자 추출 장치(100)는 프레임 수준 특징 벡터들을 이용하여 유사 글로벌 특징 벡터(G)를 도출한다(도 2 참조).

[0040] 유사 글로벌 특징 벡터(G)는 초기에 비디오의 전체적인 특징을 일차적으로 요약하기 위해 구성된다. 평균 연산은 의사 글로벌 특징 벡터를 생성하는데 사용된다. 즉, 유사 글로벌 특징 벡터(G)는 프레임 수준 특징 벡터 모음 F를 기초적인 시공간 풀링(spatio-temporal pooling)을 통해 집적한 차원 C의 특징 벡터이다.

[0041] 예를 들어, 동영상 특징 서술자 추출 장치(100)는 특징 벡터들에 대한 시공간 평균 풀링(spatio-temporal average pooling)을 수행하여 유사 글로벌 특징 벡터(G)를 산출할 수 있다.

[0042] 동영상 특징 서술자 추출 장치(100)는 수학적 1을 이용하여 유사 글로벌 특징 벡터(G)를 계산할 수 있다.

수학적 1

$$G = \beta_t(\beta_s(F)) \in \mathbb{R}^C$$

[0043]

[0044] 여기서, β_s 는 프레임의 특징 로컬 영역에 대한 편향을 방지하는 공간 축의 글로벌 평균 풀링을 나타내고, β_t 는 편향을 방지하는 시간 축의 글로벌 평균 풀링을 나타낸다.

[0045] 이와 같이, 산출된 유사 글로벌 특징 벡터(G)는 각 프레임 수준 특징 벡터들을 이용하여 산출되므로, 비디오 전체 주제를 모사하는 비디오 주제 대표 정보일 수 있다. 이러한 유사 글로벌 특징 벡터(G)는 미리 정해진 커널 함수에 기반하여 프레임 수준 특징 벡터(f_t)의 가중치 초기값을 결정하는데 사용될 수 있다.

[0046] 단계 120에서 동영상 특징 서술자 추출 장치(100)는 유사 글로벌 특징 벡터(G)를 이용하여 프레임 수준 특징 벡터들에 대한 가중치 초기값을 계산한다.

[0047] 예를 들어, 커널 함수가 공간 평균 풀링(spatial average pooling)과 결합된 코사인 유사도인 경우, 수학적 2를 이용하여 가중치 초기값이 도출될 수 있다.

[0048]

수학적 2

$$w_t^{\text{initial}} = \frac{G \cdot \beta_s(f_t)}{\|G\| \|\beta_s(f_t)\|}$$

[0048]

$$\beta_s$$

[0049] 여기서, β_s 는 공간 평균 풀링 연산을 나타내고, $\| \cdot \|$ 는 벡터 놈(vector norm)을 나타낸다.

[0050] 가중치 초기값은 각 프레임 수준 특징 벡터와 유사 글로벌 특징 벡터(G)간의 유사도를 나타내며, 가중치 초기값은 결과적으로 각 프레임이 비디오 주제에 얼마나 유사한지에 따라 결정될 수 있다. 따라서, 가중치 초기값은 공간 축을 따라 풀링된 프레임 수준 특징 벡터와 유사 글로벌 특징 벡터(G)간의 코사인 유사도와 같이 도출될 수 있다.

[0051] 단계 125에서 동영상 특징 서술자 추출 장치(100)는 계산된 가중치 초기값을 신경망 기반 재조정 모델에 적용하여 프레임 수준 특징 벡터들에 대한 가중치를 재조정하여 억제 가중치(W)를 생성한다(도 2 참조).

[0052] 신경망 기반 재조정 모델은 도 3에 도시된 바와 같다. 신경망 기반 재조정 모델은 계층적 연결 구조를 가지는 복수의 컨볼루션 레이어를 포함할 수 있다.

[0053] 예를 들어, 신경망 기반 재조정 모델은 T개의 가중치 초기값을 벡터로 입력받은 후 T개의 재조정된 억제 가중치(W)를 출력할 수 있다. 가중치 재조정은 합성곱 신경망의 수용 필드(receptive field) 범위에 들어오는 가중치들의 변화 패턴 인식을 통해 수행되며, 다양한 시간 스케일의 패턴을 종합적으로 고려하기 위해 도 3에 도시된 바와 같이 계층적 연결 구조를 가질 수 있다.

[0054] 보다 상세하게, 가중치 초기값이 입력되면, 신경망 기반 재조정 모델의 시간적 수용 필드(temporal receptive field)는 복제 패턴 레이어와 컨볼루션 레이어를 통해 3배로 확장될 수 있다. 이어, 최종 컨볼루션 레이어는 이전 작업의 출력으로부터 억제 가중치(W)를 생성한다. 모든 활성화 함수는 컨볼루션 레이어 이후에 전달되며, 억제 가중치를 안정적으로 학습하기 위해 강화된 시그모이드가 수학식 3과 같이 공식화될 수 있다.

수학식 3

$$\text{sigmoid}(h) = \frac{\rho}{1 + e^{-h/\tau}} - o$$

[0055]

[0056] 여기서, ρ 는 스케일 인자를 나타내고, τ 는 온도(temperature)를 나타내고, o는 오프셋을 나타낸다.

$$\rho$$

와 o는 0과 1 사이의 억제 가중치 값을 획득하기 위해 1과 0으로 각각 설정될 수 있다.

[0057] 상술한 계층적 연결 구조를 통해 특정 프레임의 주변 정보를 취합하여 억제 가중치(W)를 재조정할 수 있다. 즉, 계층적 연결 구조를 통해 계층적으로 단기 및 장기 정보를 모두 사용하여 특정 프레임 주변에 연속 프레임에 나타나는 다양한 길이의 콘텐츠를 식별하도록 재조정 모델을 지원할 수 있다.

[0058] 활성화 함수를 건너뛰는 제1 컨볼루션 레이어, 제2 컨볼루션 레이어, 제3 컨볼루션 레이어의 3개의 특징맵은 시그모이드 연산 없이 최종 컨볼루션 레이어의 출력과 채널별로 연결되고, 입력 크기가 255 x T인 집계 컨볼루션 레이어를 통해 융합될 수 있다.

[0059] 결과적으로 매초 하나의 프레임이 샘플링되면, 연결에 사용된 3개의 특징맵은 3, 9, 27초의 계층적 적용 범위를 가지며 억제 가중치(W)는 이러한 특징 맵을 결합하는 집계 컨볼루션 레이어에 의해 생성될 수 있다.

[0060] 단계 130에서 동영상 특징 서술자 추출 장치(100)는 억제 가중치(W)를 이용하여 프레임 수준 특징 벡터들을 비디오 수준 특징 벡터로 집계한다.

[0061] 이를 수학식으로 나타내면 수학식 4와 같다.

수학식 4

$$V = \alpha \left(\{f_t \otimes w_t\}_{t=1}^T \right)$$

[0062]

[0063]

여기서, w_t 는 재조정 가중치를 나타내고, f_t 는 t번째 프레임 수준 특징 벡터를 나타내며, T는 전체 프레임 수를 나타내며, α 는 미리 정의된 집적 함수를 나타내고, $\otimes$ 는 브로드캐스팅(broadcasting) 곱 연산을 나타낸다.

[0064]

본 발명의 일 실시예에 따른 비디오 수준 특징 벡터를 추출하는 모델은 비디오 검색의 관점에서 학습될 수 있다. 예를 들어, 수학식 5와 같은 삼중항 손실 함수에 기반하여 거리 학습을 수행할 수 있다.

수학식 5

$$L = \max \left\{ 0, \frac{V_a \cdot V_n}{\|V_a\| \|V_n\|} - \frac{V_a \cdot V_p}{\|V_a\| \|V_p\|} + \gamma \right\}$$

[0065]

[0066]

여기서, V_a , V_p , V_n 은 각각 앵커, 포지티브, 네거티브 동영상 샘플에 대한 비디오 수준 특징 벡터를 나타내며, γ 는 마진 파라미터를 나타낸다.

[0067]

도 4는 종래와 본 발명의 일 실시예에 따른 비행기 추락 사고에 대한 비디오 수준 특징 벡터를 생성한 결과를 비교한 도면이다.

[0068]

도 4의 경우 '비행기 추락 사고'에 대한 비디오에서 제1 프레임과 제5 프레임은 영상은 사고를 설명하는 텍스트만 나오거나 이 영상을 처리한 회사의 로고만 등장하고, 실질적으로 비행기 추락 사고와는 관련이 없는 것을 알 수 있다. 제1 프레임의 경우, 사람의 입장에서는 영상 콘텐츠의 이해를 돕기 위한 것일 뿐, 실제 비디오 주제와는 관련이 없으며, 이와 같은 주제의 다른 비디오에서도 동일한 장면이 존재할지 여부는 알 수 없다. 또한, 이러한 프레임은 다른 장면과의 관계를 추론하여 단일 비디오 수준의 특징을 유추하는데도 유용하지 않다. 따라서, 이러한 프레임들은 가중치 값을 낮출 필요가 있다.

[0069]

이에, 본 발명의 일 실시예에서는 주변 콘텐츠를 활용하는 계층적 연결 구조를 통해 가중치 초기값을 적절하게 조정할 수 있다. 프레임이 주제와 관련이 있는 경우(예를 들어, 도 4에서 제2, 제3 및 제4 프레임), 주변 프레임들과의 관계를 이용하여 높은 가중치값을 가지도록 조정할 수 있다.

[0070]

도 5의 경우, '건물 화재 사고'라는 주제를 나타내는 동영상의 경우, 제3 프레임과 제4 프레임에 소셜 미디어 반응을 묘사하는 프레임이 포함되어 있다.

[0071]

이는 실제 '건물 화재 사고'주제와는 무관하므로, 본 발명의 일 실시예에서는 가중치 재조정을 통해 해당 프레임의 가중치를 조정할 수 있다.

- [0072] 도 4 및 도 5에서 보여지는 바와 같이, 본 발명의 일 실시예에 따른 유사 글로벌 특징 벡터 및 가중치 재조정을 통해 종래에 비해 성능 개선 효과가 좋은 것을 알 수 있다.
- [0074] 도 6은 본 발명의 일 실시예에 따른 동영상 검색을 위한 동영상 특징 서술자 추출 장치의 구성을 개략적으로 도시한 블록도이다.
- [0075] 도 6을 참조하면, 본 발명의 일 실시예에 따른 동영상 검색을 위한 동영상 특징 서술자 추출 장치(100)는 영상 획득부(610), 특징 추출부(620), 가중치 생성부(630), 임베딩부(640), 메모리(650) 및 프로세서(660)를 포함하여 구성된다.
- [0076] 영상 획득부(610)는 동영상을 획득하기 위한 수단이다.
- [0077] 특징 추출부(620)는 동영상의 각 프레임에 대해 프레임 수준 특징 벡터를 각각 추출한다. 전술한 바와 같이, 특징 추출부(620)는 백본 네트워크 모델을 가지되, 해당 백본 네트워크 모델에 각 프레임들을 입력하여 프레임 수준 특징 벡터를 각각 추출할 수 있다(도 2 참조).
- [0078] 가중치 생성부(630)는 프레임 수준 특징 벡터들을 이용하여 유사 글로벌 특징 벡터(G)를 도출하고, 해당 유사 글로벌 특징 벡터(G)를 이용하여 억제 가중치(W)를 생성한다.
- [0079] 이에 대해 보다 상세히 설명하면 다음과 같다.
- [0080] 추출된 T개의 프레임 수준 특징 벡터를 이용하여 유사 글로벌 특징 벡터(G)를 계산할 수 있다. 이어, 가중치 생성부(630)는 프레임 수준 특징 벡터와 유사 글로벌 특징 벡터(G)간의 유사도를 계산하여 가중치 초기값을 각각 계산할 수 있다.
- [0081] 가중치 생성부(630)는 가중치 초기값을 신경망 기반 재조정 모델에 적용하여 가중치를 재조정하여 억제 가중치(W)를 생성한다. 전술한 바와 같이, 신경망 기반 재조정 모델에 가중치 초기값을 적용한 후 특정 프레임의 주변 정보들을 취합하여 억제 가중치(W)를 생성할 수 있다.
- [0082] 임베딩부(640)는 재조정된 억제 가중치(W)를 이용하여 프레임 수준 특징 벡터를 비디오 수준 특징 벡터로 집계한다.
- [0083] 메모리(650)는 본 발명의 일 실시예에 따른 동영상 검색을 위한 동영상 특징 추출 방법을 수행하기 위한 프로그램 코드들을 저장한다.
- [0084] 프로세서(660)는 본 발명의 일 실시예에 따른 동영상 검색을 위한 동영상 특징 서술자 추출 장치(100)의 내부 구성 요소들(예를 들어, 영상 획득부(610), 특징 추출부(620), 가중치 생성부(630), 임베딩부(640), 메모리(650) 등)을 제어하기 위한 수단이다.
- [0086] 본 발명의 실시 예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0087] 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0088] 이제까지 본 발명에 대하여 그 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야

한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

부호의 설명

[0090]

100: 동영상 특징 서술자 추출 장치

610: 영상 획득부

620: 특징 추출부

630: 가중치 생성부

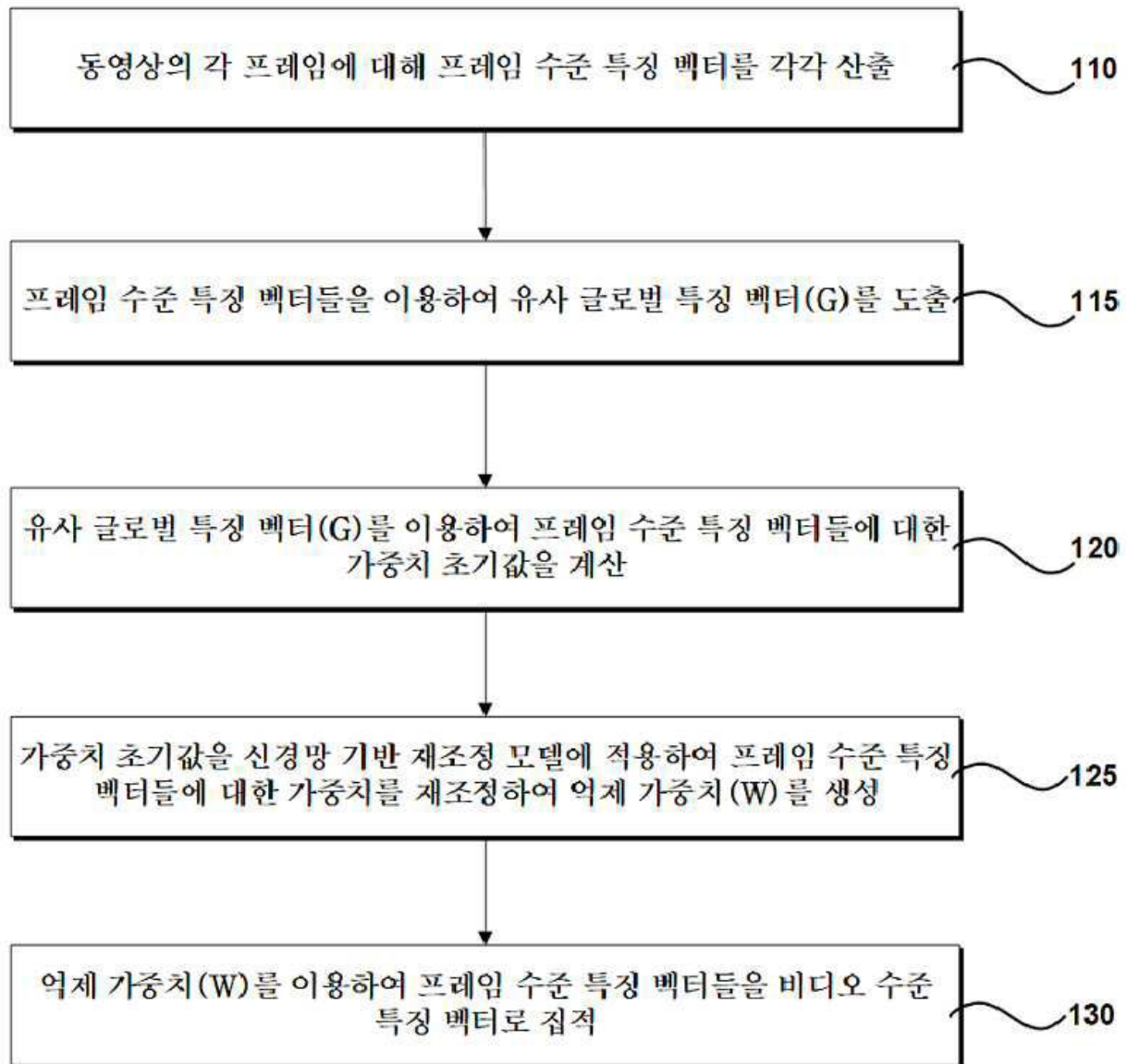
640: 임베딩부

650: 메모리

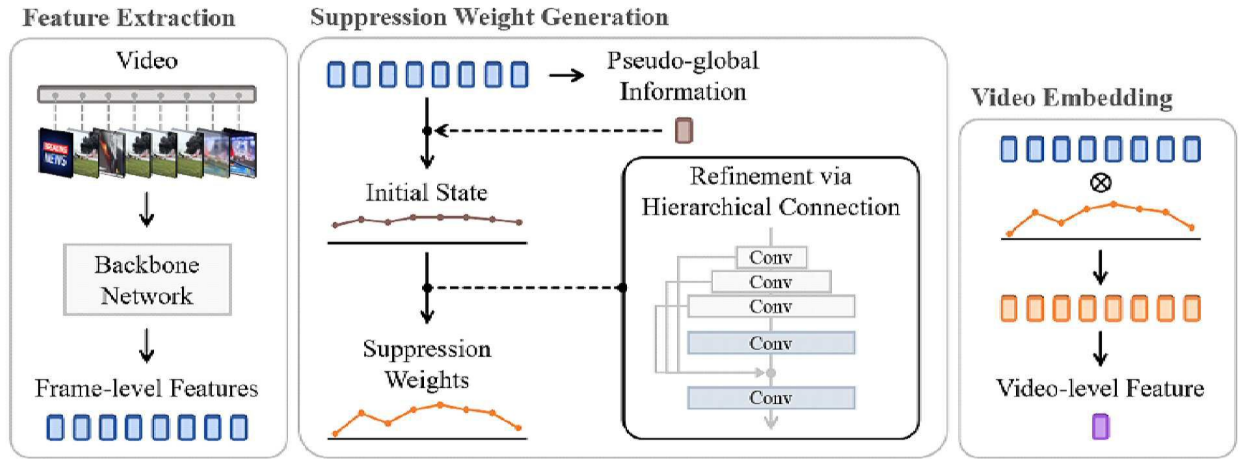
660: 프로세서

도면

도면1



도면2

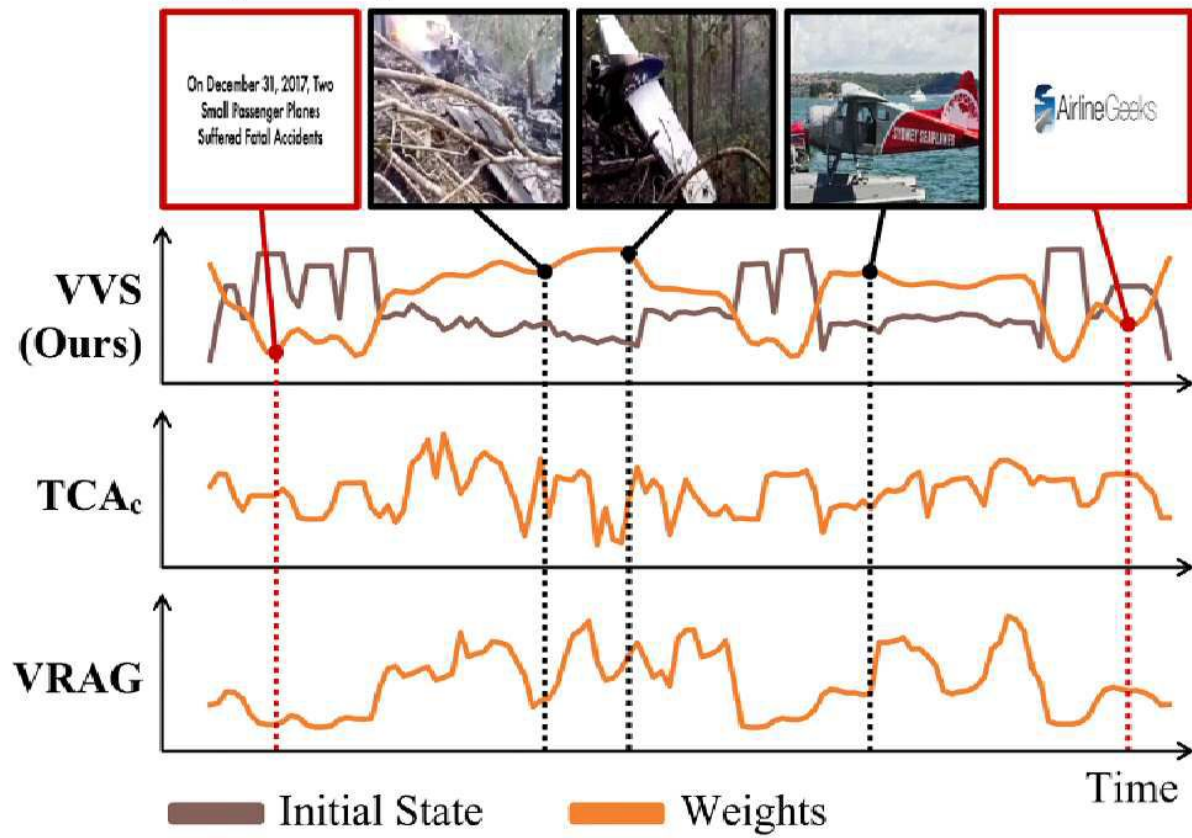


도면3

layer	k	st	d	act	in size	out size
<i>data</i>		Initial State I			-	$1 \times T$
<i>rpadd1</i>			1		$1 \times T$	$1 \times (T+2)$
<i>conv1</i>	3	1	32	<i>relu</i>	$1 \times (T+2)$	$32 \times T$
<i>rpadd2</i>			1		$32 \times T$	$32 \times (T+2)$
<i>conv2</i>	3	1	64	<i>relu</i>	$32 \times (T+2)$	$64 \times T$
<i>rpadd3</i>			1		$64 \times T$	$64 \times (T+2)$
<i>conv3</i>	3	1	128	<i>relu</i>	$64 \times (T+2)$	$128 \times T$
w/o Hierarchical Connection						
<i>fconv</i>	1	1	1	<i>sigmoid</i>	$128 \times T$	$1 \times T$
w/ Hierarchical Connection						
<i>fconv</i>	1	1	1	-	$128 \times T$	$1 \times T$
<i>aconv</i>	1	1	1	<i>sigmoid</i>	$225 \times T$	$1 \times T$

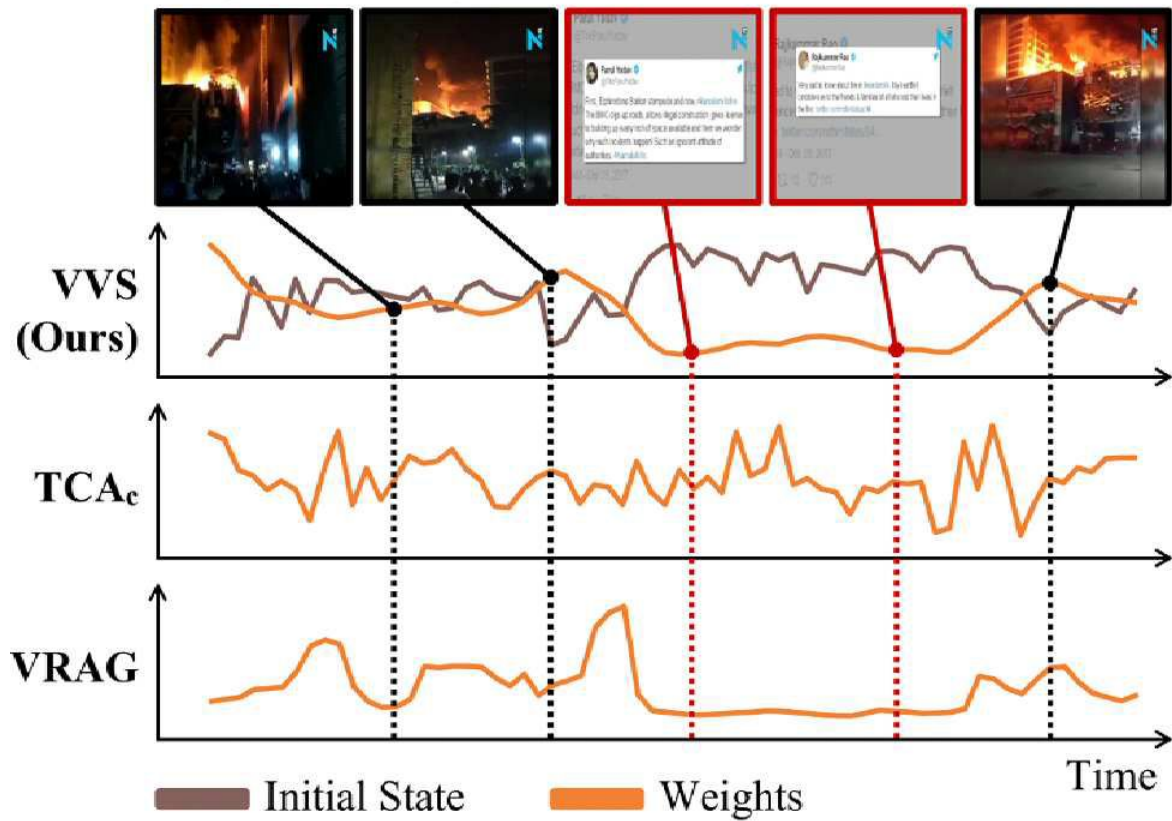
도면4

Topic: Airplane Crash Accident



도면5

Topic: Building Fire Accident



도면6

