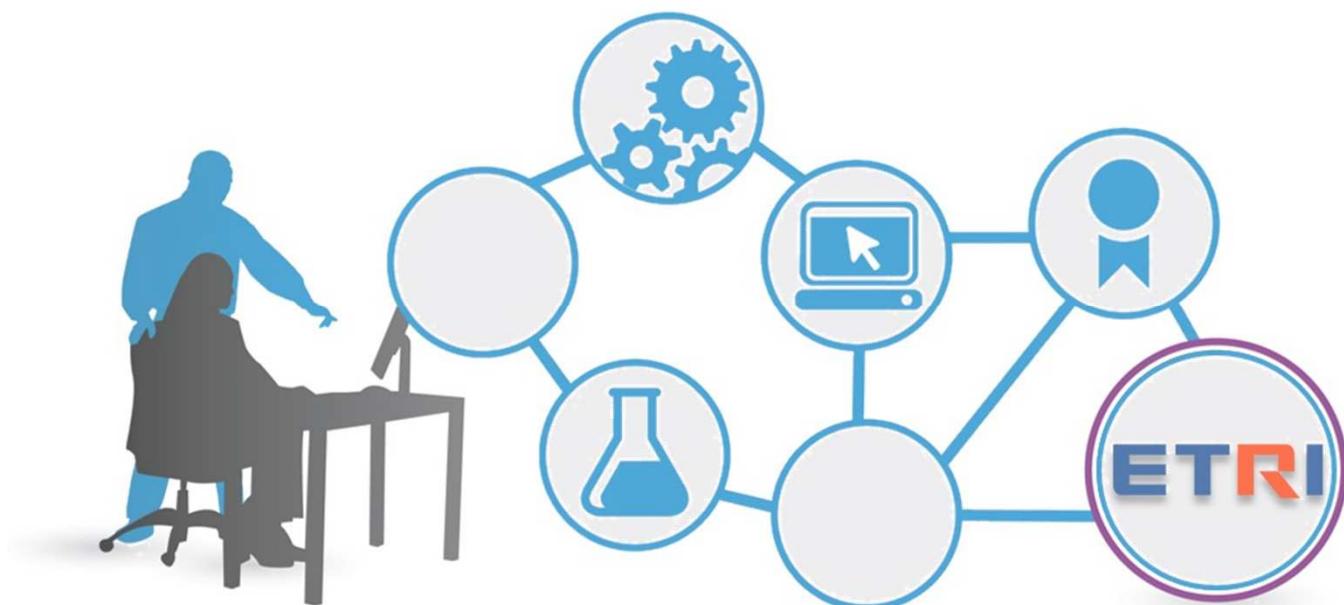


딥러닝 기반 영상 메타데이터 생성 기술



목 차

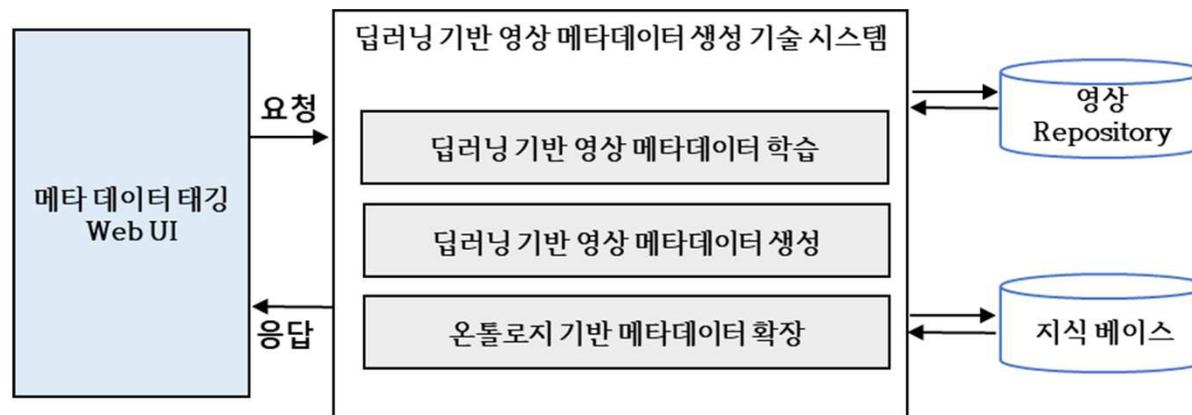
1. 기술의 개요
2. 기술이전 내용 및 범위
3. 기술 현황
4. 기술의 사업성
5. 국내외 시장 동향

1. 기술의 개요



✓ 딥러닝 기반 영상 메타데이터 생성 기술

- 본 기술은 사용자가 업로드한 영상에 대해, 시스템이 해당 영상을 분석하여 임베딩 벡터 기반의 메타데이터를 생성하는데 목적이 있음.
- 본 기술은 영상에서 특징 정보 (객체, 행위, 장소, 시간)를 추출하여 영상 메타데이터 생성 모델을 학습하는 기능을 제공함.
- 본 기술은 학습된 모델을 기반으로 임베딩 벡터 기반의 메타데이터와 영상을 설명하는 지문(영상 description)을 생성하는 기능을 제공함.
- 본 기술은 온톨로지 기반으로 클라우드 소싱을 활용하여 메타데이터 정보를 확장할 수 있는 기능을 제공함.



딥러닝 기반 영상 메타데이터 생성 기술 구조도

2. 기술이전 내용 및 범위 (1/3)



☑ 기술이전 내용

- 딥러닝 기반 영상 메타데이터 생성 모델 학습 기능
- 딥러닝 기반 영상 메타데이터 생성 기능
- 도메인 온톨로지 기반 영상 메타데이터 확장 기능
- 도메인 지식 구축 기능
- 도메인 지식 제공 기능

장면 지식 확인 영상



영상 Label 내게남은사랑을-100
지식생성일자 2020-06-17 17:40:22
Duration 19.89 (Sec.)

카탈로그 이미지



장면 임베딩 벡터 차트



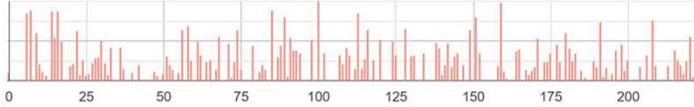
Conceptual 정보 기반 장면 벡터

지문

DESCRIPTION

사람이 사람에게 이야기한다

사람이 길거리에서 있던 사람이 일어난다



add 추가

PROPERTY	OBJECT
장소	DESCRIPTION: 사람이 사람에게 이야기한다
	DESCRIPTION: 사람이 길거리에서 있던 사람이 일어난다
	OBJECT: place/room

초기화 등록

- 마포실, 이말소
- 바다
- 박물관
- 발코니, 배관다
- 방
- 방송국
- 방파제
- 발
- 배
- 백사장, 모래사장
- 백의집
- 버스
- 밭
- 밭길
- 밭정
- 메이커리, 뽕집, 재과점
- 병원
- 보육원, 고아원
- 보후소
- 복스
- 부검실
- 분수대
- 분장실
- 바닐하우스
- 비행기
- 빌라
- 사무실
- 사무나
- 사무실 대표실
- 사형장

2. 기술이전 내용 및 범위 (2/3)



기술이전 범위

- 딥러닝 기반 영상 메타데이터 생성 기술 시스템 요구사항 정의서 1종
- 딥러닝 기반 영상 메타데이터 생성 기술 상세설계서 1종
- 딥러닝 기반 영상 메타데이터 생성 기술 시험절차 및 결과서 1종
- 딥러닝 기반 영상 메타데이터 생성 기술 프로그램 3종
 - 딥러닝 기반 영상 메타데이터 학습기 1.0
 - 딥러닝 기반 영상 메타데이터 생성기 1.0
 - 도메인 지식 기반 장면 지식 생성 1.0

2. 기술이전 내용 및 범위 (3/3)



☑ 기술 개발 현황

기술성숙도(TRL : Technology Readiness Level) :6단계

구분	단계	정의	세부 설명
기초 연구 단계	1	기초 이론/실험	◦ 기초이론 정립 단계
	2	실용목적 아이디어, 특허 등 개념정립	◦ 기술개발개념정립및아이디어에대한특허출원단계
실험 단계	3	실험실 규모의 기본성능 검증	◦ 실험실 환경에서 실험 또는 전산 시뮬레이션을 통해 기본성능이 검증될 수 있는 단계 ◦ 개발하려는부품/시스템의기본설계도면을확보하는단계
	4	실험실 규모의 소재/부품/시스템 핵심성능 평가	◦ 시험샘플을제작하여핵심성능에대한평가완료된단계 ◦ 3단계에서 도출된 다양한 결과 중에서 최적의 결과를 선택하려는 단계 ◦ 컴퓨터 모사가 가능한 경우 최적화를 완료하는 단계
시작품 단계	5	확정된 소재/부품/시스템시작품제작 및 성능 평가	◦ 확정된 소재/부품/시스템의 실험실 시작품 제작 및 성능 평가가 완료된 단계 ◦ 개발 대상의 생산을 고려하여 설계하나 실제 제작한 시작품 샘플은 1~수개 미만인 단계 ◦ 경제성을 고려하지 않고 기술의 핵심성능으로만 볼 때, 실제로 판매가 될 수 있는 정도로 목표 성능을 달성한 단계
	6	파일럿 규모 시작품 제작 및 성능 평가	◦ 파일럿 규모(복수 개~양산규모의 1/10정도)의 시작품 제작 및 평가가 완료된 단계 ◦ 파일럿규모생산품에대한생산량,품질,비용,납기,물류등에시 ◦ 파일럿 생산을 위한 대규모 투자가 동반되는 단계 ◦ 생산기업이 수요기업 적용환경에 유사하게 자체 현장테스트를 실시하여 목표 성능을 만족시킨 단계 ◦ 성능평가결과에대한가능하면공인인증기관의장확보
실용화 단계	7	신뢰성평가 및 수요기업 평가	◦ 실제 환경에서 성능 검증이 이루어지는 단계 ◦ 부품 및 소재개발의 경우 수요업체에서 직접 파일럿 시작품을 현장 평가(성능 및 신뢰성 평가) ◦ 가능하면 인증기관의 신뢰성 평가 결과 제출
	8	시제품 인증 및 표준화	◦ 표준화 및 인허가 취득 단계
사업화	9	사업화	◦ 본격적인 양산 및 사업화 단계 ◦ 6-시그마 등 품질관리가 중요한 단계

3. 기술 동향



☑ 기술 현황

국외 기술현황

- 미국의 Clarifai는 Deep learning 기반의 이미지 태깅 기술을 기반으로, 영상 콘텐츠에서 추출된 각 프레임에 대한 태깅 톨 및 API를 제공하고 있음.
 - Clarifai의 태깅 기술은 Convolution Neural Network를 기반으로 구현되었으며, 개, 의자, 책상, 다리 등의 객체부터 도시, 아침, 밤, 즐거움 등 추상적인 개념에 이르기까지 수 천개의 컨셉에 대한 학습 결과를 바탕으로 이미지에 대한 태깅을 제공함.

국내 기술현황

- KBS는 영상에 대해 세그멘테이션을 수행한 후, 수동 태깅/인덱싱 관리 저작도구를 개발하였음.
- 한밭대 김수경 교수팀은 동영상 콘텐츠의 장면 검색을 위한 장면에 내포된 개체(인물, 사건, 사물 등)의 메타체계를 온톨로지로 구성하는 연구를 수행한바 있음.
- 영화 메타데이터의 경우, 국내에서는 현재 메타데이터 정보 체계를 확립하는 수준이며 영상 검색 등 상용화 서비스에는 적용된 바 없음.

4. 기술의 사업성 (1/2)



예상 응용 제품 및 서비스

예상 제품/서비스	예상 수요자(층)
VOD 클립 서비스	VOD 콘텐츠 사업자, 포탈 사업자 등
자동 캡션 배포 서비스	방송 사업자
e-Learning 서비스	인터넷 교육 서비스 사업자

사업성

예상 제품 /서비스	예상단가 (천원)	이전기술의 비중(%)	잠재적/현재적 경쟁자와 가격,시장 등에서 경쟁상 유리한 점	판매 가능 시기
딥러닝 기반 영상 메타데이터 생성 서비스	-	50%	a. 가격경쟁력면: 신규 서비스로 가격형성 b. 시장환경면: 2021년 이후 연간 10억 예상	2021년

4. 기술의 사업성 (2/2)



☑ 상용화까지 단계별 주요 일정

- 1단계: 기술 이해 및 서비스를 위한 서버 구축 (약 6개월 소요 추정)
- 2단계: 콘텐츠 확보 및 메타데이터 생성 (초기 6개월 소요 및 지속적 유지/보수)
- 3단계: 상용화 (약 1년 후 상용화 가능 추정)

☑ 상용화를 위한 추가비용

- 영상 콘텐츠에 대한 영상 메타데이터 학습 및 생성 시 콘텐츠 확보 및 DB 관리에 따른 추가비용이 필요함.
- 자체 콘텐츠 적용에서의 성능 향상을 위해서는 학습 데이터의 추가적인 확보가 필요함.
- 영상 메타데이터 학습 및 생성 작업을 지원하기 위한 서버 구축 비용이 발생함.
- 도메인 지식 관리 및 저장을 위한 온톨로지 저장소 구입 및 수동 메타데이터 생성을 위한 인력이 필요할 수 있음.

5. 국내외 시장 동향



☑ 관련 제품/서비스 국내외 시장 규모 (추정)

관련 제품 /서비스	시장	2021	2022	2023	2024	2025
딥러닝 기반 영상 메타데이터 생성 서비스	해외 (백만불)	20	30	50	75	110
	국내 (억원)	10	15	25	40	60

☑ 예상 제품/서비스의 예상 매출액 (추정)

관련 제품 /서비스	시장	2021	2022	2023	2024	2025
딥러닝 기반 영상 메타데이터 생성 서비스	해외 (백만불)	사업준비	0.6	1	2.25	4.4
	국내 (억원)	1	3	5	12	21

감사합니다



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0122496
(43) 공개일자 2021년10월12일

(51) 국제특허분류(Int. Cl.)
HO4N 21/435 (2011.01) G06N 20/00 (2019.01)
HO4N 21/43 (2011.01) HO4N 21/854 (2011.01)
(52) CPC특허분류
HO4N 21/435 (2013.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2020-0039705
(22) 출원일자 2020년04월01일
심사청구일자 없음

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
(72) 발명자
곽창욱
대전광역시 유성구 가정로 270, B동 907호
김선중
세종특별자치시 남세종로 357, 103동 1001호
(뒷면에 계속)
(74) 대리인
한양특허법인

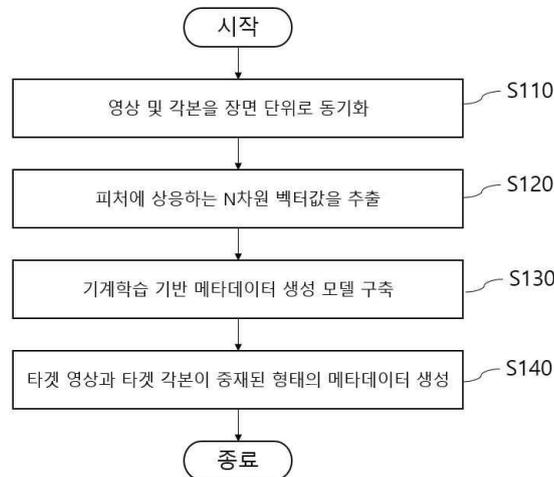
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 **벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법 및 이를 위한 장치**

(57) 요약

벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법 및 이를 위한 장치가 개시된다. 본 발명의 일실시예에 따른 메타데이터 생성 방법은 입력 영상에서 추출된 적어도 하나의 특징을 기반으로 상기 입력 영상을 입력 각본과 장면 단위로 동기화하고, 동기화된 장면 단위의 영상 및 각본 각각에 대해 상기 적어도 하나의 특징에 상응하는 N차원의 벡터값을 추출하고, 상기 N차원의 벡터값을 기반으로 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델에 대한 기계학습을 수행하여 메타데이터 생성 모델을 구축하고, 타겟 영상 및 타겟 각본 중 어느 하나만 입력된 경우, 상기 메타데이터 생성 모델을 기반으로 상기 타겟 영상과 상기 타겟 각본이 중재된 형태의 메타데이터를 생성한다.

대표도 - 도1



(52) CPC특허분류

H04N 21/4302 (2020.08)

H04N 21/854 (2013.01)

(72) 발명자

손정우

대전광역시 유성구 봉명로 93, 608동 105호

이호재

대전광역시 유성구 지족로 362, 307동 402호

한민호

대전광역시 유성구 엑스포로123번길 46-15, 202동 303호

함경준

대전광역시 유성구 어은로 57, 120동 105호

김상권

대전광역시 유성구 어은로 57, 133동 505호

박중현

대전광역시 유성구 대덕대로 617, 101동 501호

이 발명을 지원한 국가연구개발사업

과제고유번호 1711101947

과제번호 20ZH1200

부처명 과학기술정보통신부

과제관리(전문)기관명 한국전자통신연구원

연구사업명 한국전자통신연구원연구운영비지원(R&D)(주요사업비)

연구과제명 오픈시나리오 기반 프로그래머블 인터랙티브 미디어 창작 서비스 플랫폼 개발

기 여 율 1/1

과제수행기관명 한국전자통신연구원

연구기간 2020.01.01 ~ 2020.12.31

명세서

청구범위

청구항 1

입력 영상에서 추출된 적어도 하나의 특징을 기반으로 상기 입력 영상을 입력 각본과 장면 단위로 동기화하는 단계;

동기화된 장면 단위의 영상 및 각본 각각에 대해 상기 적어도 하나의 특징에 상응하는 N차원의 벡터값을 추출하는 단계;

상기 N차원의 벡터값을 기반으로 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델에 대한 기계학습을 수행하여 메타데이터 생성 모델을 구축하는 단계; 및

타겟 영상 및 타겟 각본 중 어느 하나만 입력된 경우, 상기 메타데이터 생성 모델을 기반으로 상기 타겟 영상과 상기 타겟 각본이 중재된 형태의 메타데이터를 생성하는 단계

를 포함하는 것을 특징으로 하는 메타데이터 생성 방법.

발명의 설명

기술 분야

[0001] 본 발명은 메타데이터를 자동으로 생성하는 기술에 관한 것으로, 특히 영상을 기반으로 장면을 묘사하는 정보를 추출하여 의미적 메타데이터를 생성하는 기술에 관한 것이다.

배경 기술

[0002] 방송국이 주도했던 영상 콘텐츠 시장은 온라인의 영상 플랫폼으로 중심이 이동되고 있다. 과거 텔레비전을 통해 수동적인 소비가 이루어졌다면, 최근에는 기술의 진보와 변화에 따라 생산, 유통, 소비의 방법이 다양해졌다.

[0003] 초기 영상 콘텐츠 스트리밍 서비스는 1시간 ~ 2시간 길이의 콘텐츠 전체를 활용했다. 하지만 최근 주요 OTT 업체에서는 3~4분 내외의 장면 클립으로 영상을 분할하여 온라인 플랫폼에서 서비스를 하고 있다. 이러한 클립 단위의 서비스는 빠르고 짧은 것을 선호하는 젊은 층의 소비 성향과 일치하기도 하지만, 주요 수익수단이 미디어 콘텐츠 재생 사이에 노출되는 광고라는 점을 볼 때, 온라인 콘텐츠 서비스에도 최적화된 단위이다. 영상의 단위가 축소되면서 짜깁기, 재편집과 같은 기존 영상을 재활용하는 산업적 시도도 발생하고 있다.

[0004] 이처럼 서비스에 활용되는 영상의 단위가 장면으로 축소되고, 기존 영상의 재활용이 확대됨에 따라 정확한 정보가 포함된 영상을 검색하는 것이 중요해졌다. 기존에는 콘텐츠 제목, 배우와 같은 큰 규모의 영상을 대상으로 사용자의 검색이 이루어졌다면, 장면 단위에서는 영상의 장소, 시간, 등장인물, 행동, 객체 등과 같은 다양한 정보들의 검색 요구가 발생하고 있다.

[0005] 현재, 영상 서비스 업체에서는 해시태그를 기반으로 주요한 정보들을 5개 내외의 키워드로 메타데이터를 태깅하고 이를 활용한 영상 검색이 이루어진다. 이러한 정보들은 사람들이 직접 태깅하고 있다는 점에서 시간적, 물리적 비용이 크기 때문에, 자동화된 영상 분석 및 태깅 시스템이 필요한 실정이다. 뿐만 아니라, 생성되는 키워드의 기준이 불명확하기 때문에 누락되는 정보들이 대부분이다. 따라서, 영상에 포함된 다양한 정보들을 메타데이터로 생성할 수 있는 시스템이 필요하다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국 공개 특허 제10-2018-0087969호, 2018년 8월 3일 공개(명칭: 동영상 장면과 메타데이터 저작 방법)

발명의 내용

해결하려는 과제

- [0007] 본 발명의 목적은 영상의 의미적 정보들을 효과적으로 벡터 형식으로 압축 표현함으로써 영상에 포함된 내용들을 효과적으로 나타낼 수 있는 의미적 메타데이터를 생성하여 제공하는 것이다.
- [0008] 또한, 본 발명의 목적은 영상이나 각본 중 어느 하나의 리소스의 부재에도 두 리소스 전체를 반영한 메타데이터를 생성하는 것이다.
- [0009] 또한, 본 발명의 목적은 벡터 기반의 메타데이터를 제공함으로써 벡터 사이의 유사도 계산을 통해 보다 효율적인 영상 검색 서비스를 제공하는 것이다.

과제의 해결 수단

- [0010] 상기한 목적을 달성하기 위한 본 발명에 따른 메타데이터 생성 방법은 입력 영상에서 추출된 적어도 하나의 특징을 기반으로 상기 입력 영상을 입력 각본과 장면 단위로 동기화하는 단계; 동기화된 장면 단위의 영상 및 각본 각각에 대해 상기 적어도 하나의 특징에 상응하는 N차원의 벡터값을 추출하는 단계; 상기 N차원의 벡터값을 기반으로 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델에 대한 기계학습을 수행하여 메타데이터 생성 모델을 구축하는 단계; 및 타겟 영상 및 타겟 각본 중 어느 하나만 입력된 경우, 상기 메타데이터 생성 모델을 기반으로 상기 타겟 영상과 상기 타겟 각본이 중재된 형태의 메타데이터를 생성하는 단계를 포함한다.
- [0011] 이 때, 영상 잠재 벡터 모델은 상기 장면 단위의 영상에 대한 메타데이터를 생성하고, 상기 각본 잠재 벡터 모델은 상기 장면 단위의 각본에 대한 메타데이터를 생성할 수 있다.
- [0012] 이 때, 구축하는 단계는 손실 함수를 기반으로 상기 영상 잠재 벡터 모델 및 상기 각본 잠재 벡터 모델 각각에 포함된 인코더 및 디코더를 상기 영상 잠재 벡터 모델과 상기 각본 잠재 벡터 모델 사이의 차이가 감소하도록 학습시키는 단계를 포함할 수 있다.
- [0013] 이 때, 학습시키는 단계는 상기 영상 잠재 벡터 모델 및 상기 각본 잠재 벡터 모델 각각에 상응하는 적어도 하나의 입력 벡터 및 출력 벡터를 획득하는 단계; 상기 영상 잠재 벡터 모델과 상기 각본 잠재 벡터 모델 간 상기 적어도 하나의 입력 벡터 및 출력 벡터의 오차가 감소하도록 상기 인코더 및 디코더의 변수를 조정하는 단계를 포함할 수 있다.
- [0014] 이 때, 메타데이터는 실수형의 벡터 형식에 상응할 수 있다.
- [0015] 이 때, 동기화하는 단계는 상기 적어도 하나의 특징을 기반으로 상기 입력 영상을 샷 단위로 분할하는 단계; 및 상기 샷 단위로 분할된 복수개의 샷 영상들을 의미적 연속성을 갖는 장면 단위로 병합하는 단계를 포함할 수 있다.
- [0016] 또한, 본 발명의 일실시예에 따른 메타데이터 생성 장치는, 입력 영상에서 추출된 적어도 하나의 특징을 기반으로 상기 입력 영상을 입력 각본과 장면 단위로 동기화하고, 동기화된 장면 단위의 영상 및 각본 각각에 대해 상기 적어도 하나의 특징에 상응하는 N차원의 벡터값을 추출하고, 상기 N차원의 벡터값을 기반으로 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델에 대한 기계학습을 수행하여 메타데이터 생성 모델을 구축하고, 타겟 영상 및 타겟 각본 중 어느 하나만 입력된 경우, 상기 메타데이터 생성 모델을 기반으로 상기 타겟 영상과 상기 타겟 각본이 중재된 형태의 메타데이터를 생성하는 프로세서; 및 상기 메타데이터 생성 모델을 저장하는 메모리를 포함한다.
- [0017] 이 때, 영상 잠재 벡터 모델은 상기 장면 단위의 영상에 대한 메타데이터를 생성하고, 상기 각본 잠재 벡터 모델은 상기 장면 단위의 각본에 대한 메타데이터를 생성할 수 있다.
- [0018] 이 때, 프로세서는 손실 함수를 기반으로 상기 영상 잠재 벡터 모델 및 상기 각본 잠재 벡터 모델 각각에 포함된 인코더 및 디코더를 상기 영상 잠재 벡터 모델과 상기 각본 잠재 벡터 모델 사이의 차이가 감소하도록 학습시킬 수 있다.
- [0019] 이 때, 프로세서는 상기 영상 잠재 벡터 모델 및 상기 각본 잠재 벡터 모델 각각에 상응하는 적어도 하나의 입력 벡터 및 출력 벡터를 획득하고, 상기 영상 잠재 벡터 모델과 상기 각본 잠재 벡터 모델 간 상기 적어도 하나의 입력 벡터 및 출력 벡터의 오차가 감소하도록 상기 인코더 및 디코더의 변수를 조정할 수 있다.

[0020] 이 때, 메타데이터는 실수형의 벡터 형식에 상응할 수 있다.

[0021] 이 때, 프로세서는 상기 적어도 하나의 특징을 기반으로 상기 입력 영상을 샷 단위로 분할하는 단계; 및 상기 샷 단위로 분할된 복수개의 샷 영상들을 의미적 연속성을 갖는 장면 단위로 병합할 수 있다.

발명의 효과

[0022] 본 발명에 따르면, 영상의 의미적 정보들을 효과적으로 벡터 형식으로 압축 표현함으로써 영상에 포함된 내용들을 효과적으로 나타낼 수 있는 의미적 메타데이터를 생성하여 제공할 수 있다.

[0023] 또한, 본 발명은 영상이나 각본 중 어느 하나의 리소스의 부재에도 두 리소스 전체를 반영한 메타데이터를 생성할 수 있다.

[0024] 또한, 본 발명은 벡터 기반의 메타데이터를 제공함으로써 벡터 사이의 유사도 계산을 통해 보다 효율적인 영상 검색 서비스를 제공할 수 있다.

도면의 간단한 설명

[0025] 도 1은 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법을 나타낸 동작 흐름도이다.

도 2는 본 발명에 따른 메타데이터 생성 과정을 나타낸 도면이다.

도 3은 본 발명에 따른 N차원 벡터값을 추출하는 과정을 나타낸 도면이다.

도 4는 본 발명에 따른 메타데이터 생성 모델을 구축하는 과정을 나타낸 도면이다.

도 5는 본 발명에 따른 메타데이터 생성 모델 구축 과정을 상세하게 나타낸 동작 흐름도이다.

도 6 내지 도 7은 본 발명에 따른 타겟 영상만으로 메타데이터를 생성하는 과정의 일 예를 나타낸 도면이다.

도 8 내지 도 9는 본 발명에 따른 타겟 각본만으로 메타데이터를 생성하는 과정의 일 예를 나타낸 도면이다.

도 10은 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 장치를 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0026] 본 발명을 첨부된 도면을 참조하여 상세히 설명하면 다음과 같다. 여기서, 반복되는 설명, 본 발명의 요지를 불필요하게 흐릴 수 있는 공지 기능, 및 구성에 대한 상세한 설명은 생략한다. 본 발명의 실시형태는 당 업계에서 평균적인 지식을 가진 자에게 본 발명을 보다 완전하게 설명하기 위해서 제공되는 것이다. 따라서, 도면에서의 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.

[0027] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0029] 도 1은 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법을 나타낸 동작 흐름도이다.

[0030] 본 발명은 기존의 키워드 기반의 메타데이터 생성 기술에서 진보하여, 영상에서 추출 가능한 여러 특징 정보들이 통합된 벡터 기반의 메타데이터를 생성하는 것이다. 기존의 일반적인 영상 메타데이터 생성 방법들이 영상에 등장하는 키워드들을 태깅하는 것이었다면, 본 발명에서는 영상의 의미적 요소들을 포함한 벡터 기반의 메타데이터를 생성하는 방법을 제공한다. 이를 위해, 기계학습 기반의 영상 분석을 통해 장소, 시간, 행위, 지문 등의 특징 정보들을 추출할 수 있다. 각각의 특징 정보들은 N차원 벡터 형식의 메타데이터로 표현할 수 있으며, 벡터로 변환하는 인코더 및 디코더의 학습을 통해 벡터들은 다시 텍스트로 재현할 수 있다.

[0031] 이러한 개념을 확장하여, 학습을 통해 영상에서 추출한 특징 정보들을 통합해 하나의 벡터로 표현하고, 통합된 벡터를 기반으로 다시 정보들을 재현할 수 있다. 즉, 통합된 벡터는 추출된 특징 정보들이 축약된 것으로써 영상의 정보들을 나타내는 메타데이터로서 사용될 수 있다는 것을 의미한다.

[0032] 본 발명은 크게 메타데이터를 생성하는 메타데이터 생성 모델을 학습하는 단계와 학습된 메타데이터 생성 모델을 기반으로 메타데이터를 생성하는 단계로 구분할 수 있는데, 이하에서는 먼저 메타데이터 생성 모델을 학습하

는 단계를 설명하도록 한다.

- [0033] 도 1을 참조하면, 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법은 입력 영상에서 추출된 적어도 하나의 특징을 기반으로 입력 영상을 입력 각본과 장면 단위로 동기화한다(S110).
- [0034] 예를 들어, 도 2를 참조하면, 입력된 각본(210)에 대해 각본 분할(S210)을 수행하고, 입력된 영상(220)에 대해 특징 분석(S220)후 샷 분할(S230) 및 장면 분할(S240)을 수행한 뒤 각각 분할된 영상과 각본에 대해 동기화를 수행할 수 있다(S250).
- [0035] 이 때, 입력 각본은 입력 영상이 제작되기에 앞서 기록된 텍스트 형식의 문서에 상응하는 것으로, 장소, 시간, 장면 등을 설명하는 지문이 기록될 수 있다.
- [0036] 이 때, 입력된 영상(220)에 대한 특징 분석(S220)은 메타데이터 생성 장치에 포함된 별도의 영상 특징 분석 모듈을 통해 수행될 수 있다.
- [0037] 이 때, 적어도 하나의 특징을 기반으로 입력 영상을 샷 단위로 분할할 수 있다.
- [0038] 이 때, 샷 단위로 분할된 복수개의 샷 영상들을 의미적 연속성을 갖는 장면 단위로 병합할 수 있다.
- [0039] 예를 들어, 영상에 대한 특징 분석(S220)을 통해 추출된 특징 정보들을 별도의 샷 분할 모듈을 이용하여 1~3초 내외의 샷으로 분할할 수 있다. 이렇게 분할된 샷은 장면 분할 모듈을 통해 의미적으로 연속성을 갖는 장면으로 병합되어 장면 단위의 영상으로 생성될 수 있다.
- [0040] 이와 같이 장면 단위의 영상은 샷 분할, 장면 분할 과정을 통해 생성될 수 있으며, 영상을 장면 단위로 분할하는 방법은 특정한 방법에 한정되지 않는다.
- [0041] 이 때, 입력 각본은 별도의 각본 분할 모듈을 통해서 영상과 동일한 장면 단위로 분할되어 장소, 시간, 지문 형식의 인스턴스로 변화될 수 있다.
- [0042] 따라서, 이와 같이 분할된 각본은 별도의 동기화 모듈을 기반으로 장면 단위로 분할된 영상과 동기화될 수 있다.
- [0043] 또한, 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법은 동기화된 장면 단위의 영상 및 각본 각각에 대해 적어도 하나의 특징에 상응하는 N차원의 벡터값을 추출한다(S120).
- [0044] 예를 들어, 도 2를 참조하면, 장면 분할(S240) 단계를 통해 장면 단위로 분할된 영상을 기반으로 적어도 하나의 특징에 상응하는 N차원 벡터값을 추출할 수 있다(S260).
- [0045] 이 때, 장면 단위의 영상에서는 영상의 행위 정보, 장소 정보, 시간 정보, 객체 정보, 지문 정보에 상응하는 특징 정보들이 추출될 수 있다.
- [0046] 예를 들어, 행위 정보는 '달린다', '때린다', '이야기한다' 등과 같이 인물의 행동을 기준으로 생성되는 정보에 상응할 수 있다. 장소 정보는 영상의 배경이 되는 장소에 상응하는 것으로 '공원', '학교', '도로' 등이 될 수 있다. 시간 정보는 '밤', '낮', '아침' 같이 영상의 시간적 배경이 되는 정보에 상응할 수 있다. 객체 정보는 '전화기', '컵', '벤치' 와 같이 영상에 나타난 객체들에 대한 정보에 해당할 수 있다. 지문 정보는 영상 캡셔닝의 결과로써 '남자와 여자가 이야기하고 있다' 또는 '도시의 야경이 펼쳐지고 있다' 등 영상을 설명하는 정보에 상응할 수 있다.
- [0047] 이 때, 도 3을 참조하면, 본 발명에서는 메타데이터 생성 장치에 포함된 정보 추출 모듈(310)로 장면 영상을 입력함으로써 적어도 하나의 특징에 상응하는 N차원 벡터값(320)을 추출할 수 있다.
- [0048] 이 때, 각각의 벡터값들은 정보 추출 모듈(310)에 포함된 각 인식 모델을 통해 추출된 벡터에 상응하는 것으로써 분류 태스크에서 Softmax 함수를 통해 클래스 정보를 텍스트로 표현할 수 있다. 이 때, 본 발명에서는 Softmax 함수를 사용하기 전의 잠재 벡터를 장면에 대한 특징 정보로 사용할 수 있다.
- [0049] 또한, 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법은 N차원의 벡터값으로 기반으로 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델에 대한 기계학습을 수행하여 메타데이터 생성 모델을 구축한다(S130).
- [0050] 이 때, 영상 잠재 벡터 모델은 장면 단위의 영상에 대한 메타데이터를 생성하고, 각본 잠재 벡터 모델은 장면 단위의 각본에 대한 메타데이터를 생성할 수 있다.

[0051] 이 때, 손실 함수를 기반으로 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델 각각에 포함된 인코더 및 디코더를 영상 잠재 벡터 모델과 각본 잠재 벡터 모델 사이의 차이가 감소하도록 학습시킬 수 있다.

[0052] 이 때, 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델 각각에 상응하는 적어도 하나의 입력 벡터 및 출력 벡터를 획득할 수 있다.

[0053] 예를 들어, 도 4를 참조하면, 영상 잠재 벡터 모델(410)은 정보 추출 모듈을 통해 입력되는 장면 단위의 영상에 대한 N차원 벡터값을 입력값으로 학습될 수 있고, 각본 잠재 벡터 모델(420)은 각본 분할 모듈을 통해 입력되는 장면 단위의 각본에 대한 N차원 벡터값을 입력값으로 학습될 수 있다. 즉, 영상과 각본에서 각각 추출된 인스턴스 정보를 이용하여 잠재 벡터로 표현할 수 있는 인코더 및 디코더를 학습시킬 수 있다.

[0054] 이 때, 영상 잠재 벡터 모델(410)과 각본 잠재 벡터 모델(420)에 각각 포함된 잠재 벡터는 입력된 벡터값들을 인코더를 통해 압축 표현한 것으로써 다시 디코더를 통해 각각의 벡터로 재현할 수 있다.

[0055] 예를 들어, 장면 단위의 영상에서 추출된 장소 정보, 시간 정보, 행위 정보, 객체 정보 및 지문 정보들은 영상 잠재 벡터 모델(410)의 인코더를 통해 N차원의 영상 잠재 벡터로 표현될 수 있으며, 영상 잠재 벡터는 디코더를 통해 다시 장소 정보, 시간 정보, 행위 정보, 객체 정보 및 지문 정보로 재현될 수 있다.

[0056] 이와 마찬가지로 장면 단위의 각본에서 추출된 장소, 시간, 지문 인스턴스는 워드 임베딩 모델을 통해 N차원 벡터값으로 변환될 수 있고, 벡터값으로 변환된 각각의 인스턴스 정보들은 각본 잠재 벡터 모델(420)을 통해 N차원의 각본 잠재 벡터로 표현될 수 있다. 또한, 디코더를 통해 각본 잠재 벡터를 다시 벡터값으로 변환된 각각의 인스턴스 정보들로 재현할 수도 있다.

[0057] 이 때, 영상 잠재 벡터 모델과 각본 잠재 벡터 모델 간 적어도 하나의 입력 벡터 및 출력 벡터의 오차가 감소하도록 인코더 및 디코더의 변수를 조정할 수 있다.

[0058] 예를 들어, 손실함수는 [수학식 1]과 같이 영상 잠재 벡터 모델과 각본 잠재 벡터 모델 사이의 차이가 작아지도록 학습함으로써, 메타데이터 생성 모델이 입력된 영상 정보들과 각본 정보가 중재된 벡터를 가지도록 할 수 있다. 이는 추후에 영상이나 각본 중 하나의 리소스의 부재에도 두 리소스 전체를 반영한 메타데이터를 생성할 수 있는 효과가 있다.

[0059] [수학식 1]

$$\text{Loss} = \sum_{i=1}^m (\text{영상잠재벡터}_i - \text{각본잠재벡터}_i)^2$$

[0060]

[0061] 상기와 같이 메타데이터 생성 모델을 학습시키는 과정을 상세하게 나타내면 도 5와 같이 나타낼 수 있다.

[0062] 도 5를 참조하면, 영상과 각본이 입력되면(S510), 입력된 데이터를 구분하여(S515) 영상과 각본을 구별할 수 있다.

[0063] 이 후, 영상의 특징을 분석하여 추출하고(S520), 추출된 특징을 기반으로 샷 분할과 장면 분할을 수행하여(S530) 입력된 영상에 대한 장면 단위의 영상을 생성할 수 있다. 각본 또한 영상에서 추출된 특징을 기반으로 분할하고(S540), 장면 단위의 각본에서 정보를 추출할 수 있다(S550).

[0064] 이 후, 장면 단위의 영상과 각본을 동기화할 수 있다(S560).

[0065] 이 후, 또 다시 데이터를 영상과 각본으로 구분하여(S565), 장면 단위의 영상에 대해 추출된 특징에 상응하는 N차원 벡터값을 추출한 뒤 영상 잠재 벡터 모델을 학습시키고(S570), 각본에 대해 추출된 특징에 상응하는 N차원 벡터값을 추출한 뒤 각본 잠재 벡터 모델을 학습시킬 수 있다(S580).

[0066] 이 때, 손실 함수를 이용하여 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델 각각에 포함된 인코더 및 디코더를 영상 잠재 벡터 모델과 각본 잠재 벡터 모델 사이의 차이가 감소하도록 학습시킬 수 있다.

[0067] 이와 같은 학습을 통해 메타데이터 생성 모델을 구축할 수 있다(S590).

[0068] 이하에서는 학습된 메타데이터 생성 모델을 기반으로 메타데이터를 생성하는 단계를 설명하도록 한다.

[0069] 또한, 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법은 타겟 영상 및 타겟

각본 중 어느 하나만 입력된 경우, 메타데이터 생성 모델을 기반으로 타겟 영상과 타겟 각본이 중재된 형태의 메타데이터를 생성한다(S140).

- [0070] 예를 들어, 도 6 내지 도 7을 참조하면, 타겟 영상(610)만 입력되었다고 가정하면(S710), 타겟 영상에 대한 특징을 분석 및 추출할 수 있다(S720). 이 후, 추출된 특징을 기반으로 타겟 영상에 대한 샷 분할 및 장면 분할을 수행하고(S730), N차원 벡터값을 추출하여 학습된 메타데이터 생성 모델(620)로 입력할 수 있다(S740). 이 후, 메타데이터 생성 모델(620)이 영상 및 각본이 중재된 메타데이터(630)를 생성하여 제공하면(S750), 별도의 저장소에 장면 단위의 영상에 대해 생성된 메타데이터(630)를 저장할 수 있다(S760).
- [0071] 다른 예를 들어, 도 8 내지 도 9를 참조하면, 타겟 각본(810)만 입력되었다고 가정하면(S910), 타겟 각본에 대한 정보를 추출할 수 있다(S920). 이 후, 추출된 정보를 기반으로 N차원 벡터값을 추출하여 학습된 메타데이터 생성 모델(820)로 입력할 수 있다(S930). 이 후, 메타데이터 생성 모델(820)이 영상 및 각본이 중재된 메타데이터(830)를 생성하여 제공하면(S940), 별도의 저장소에 장면 단위의 각본에 대해 생성된 메타데이터(830)를 저장할 수 있다(S950).
- [0072] 이 때, 메타데이터는 실수형의 벡터 형식에 상응할 수 있다.
- [0073] 예를 들어, 메타데이터는 아래와 같이 실수형의 벡터 형식으로 표현될 수 있다. '-0.3188 -0.456 0.1647 -0.2495 -0.9385 -0.3563 -0.0732 0.4643 0.2975 0.5199 -0.393 -0.0246 0.3141 -0.2959 0.4304 -0.471 0.3575 -0.0843 -0.213 -0.1231 0.0305 -0.1432 0.1642 -0.3709 -0.3492 0.3128 -0.3125 0.197 -0.3428 0.0803 0.7957 0.2245 -0.1008 -0.2649 -0.2901 0.1465 -0.2685 0.0587 0.3844 -0.0293 -0.0189 0.352 0.058 0.9351 0.0297 -0.5063 -0.256 0.6041 -0.1265 0.1885 -0.0964 -0.3505 -0.2208 -0.6876 -0.4202 0.0777 -0.0117 0.6559 0.3044 0.1288 -0.0751 -0.8191 -0.2499 -0.3286 -0.0498 -0.0646 0.2347 -0.0313 -0.1417 0.0131 -0.0557 0.4898 0.2188 -0.4096 -0.0245 -0.4827 0.0847 0.2517 0.203 0.2854 -0.0851 -0.2978 -0.0002 -0.4072 0.4154 0.4654 -0.1107 0.7675 0.2345 - 썸'
- [0074] 또한, 도 1에는 도시하지 아니하였으나, 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법은 상술한 메타데이터 생성 과정에서 발생하는 다양한 정보를 별도의 저장 모듈에 저장할 수 있다.
- [0075] 이와 같은 메타데이터 생성 방법을 이용함으로써 영상을 기반으로 장면을 묘사하는 정보를 추출하고, 영상에 대한 의미적 메타데이터를 생성 및 제공할 수 있다.
- [0076] 또한, 기계학습 기반의 영상 분석을 통해 영상을 묘사하는 장소, 시간, 행동 등과 같은 의미적인 정보들을 추출할 수 있으며, 이러한 정보들에 기반한 학습을 통해 통합된 하나의 벡터 형식의 의미적 메타데이터를 생성할 수도 있다.
- [0077] 또한, 각본의 수급이 비교적 어려운 환경에서 각본이 없이도 영상만으로 각본을 충분히 고려한 메타데이터를 제공하는 효과를 얻을 수도 있다.
- [0079] 도 10은 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 장치를 나타낸 도면이다.
- [0080] 본 발명은 기존의 키워드 기반의 메타데이터 생성 기술에서 진보하여, 영상에서 추출 가능한 여러 특징 정보들이 통합된 벡터 기반의 메타데이터를 생성하는 것이다. 기존의 일반적인 영상 메타데이터 생성 방법들이 영상에 등장하는 키워드들을 태깅하는 것이었다면, 본 발명에서는 영상의 의미적 요소들을 포함한 벡터 기반의 메타데이터를 생성하는 방법을 제공한다. 이를 위해, 기계학습 기반의 영상 분석을 통해 장소, 시간, 행위, 지문 등의 특징 정보들을 추출할 수 있다. 각각의 특징 정보들은 N차원 벡터 형식의 메타데이터로 표현할 수 있으며, 벡터로 변환하는 인코더 및 디코더의 학습을 통해 벡터들은 다시 텍스트로 재현할 수 있다.
- [0081] 이러한 개념을 확장하여, 학습을 통해 영상에서 추출한 특징 정보들을 통합해 하나의 벡터로 표현하고, 통합된 벡터를 기반으로 다시 정보들을 재현할 수 있다. 즉, 통합된 벡터는 추출된 특징 정보들이 축약된 것으로서 영상의 정보들을 나타내는 메타데이터로써 사용될 수 있다는 것을 의미한다.
- [0082] 본 발명은 크게 메타데이터를 생성하는 메타데이터 생성 모델을 학습하는 단계와 학습된 메타데이터 생성 모델을 기반으로 메타데이터를 생성하는 단계로 구분할 수 있는데, 이하에서는 먼저 메타데이터 생성 모델을 학습하는 단계를 설명하도록 한다.
- [0083] 도 10을 참조하면, 본 발명의 일실시예에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 장치는 통신부

(1010), 프로세서(1020) 및 메모리(1030)를 포함한다.

- [0084] 통신부(1010)는 네트워크와 같은 통신망을 통해 정보 공유를 위해 필요한 정보를 송수신하는 역할을 한다.
- [0085] 프로세서(1020)는 입력 영상에서 추출된 적어도 하나의 특징을 기반으로 입력 영상을 입력 각본과 장면 단위로 동기화한다.
- [0086] 예를 들어, 도 2를 참조하면, 입력된 각본(210)에 대해 각본 분할(S210)을 수행하고, 입력된 영상(220)에 대해 특징 분석(S220)후 샷 분할(S230) 및 장면 분할(S240)을 수행한 뒤 각각 분할된 영상과 각본에 대해 동기화를 수행할 수 있다(S250).
- [0087] 이 때, 입력 각본은 입력 영상이 제작되기에 앞서 기록된 텍스트 형식의 문서에 상응하는 것으로, 장소, 시간, 장면 등을 설명하는 지문이 기록될 수 있다.
- [0088] 이 때, 입력된 영상(220)에 대한 특징 분석(S220)은 메타데이터 생성 장치에 포함된 별도의 영상 특징 분석 모듈을 통해 수행될 수 있다.
- [0089] 이 때, 적어도 하나의 특징을 기반으로 입력 영상을 샷 단위로 분할할 수 있다.
- [0090] 이 때, 샷 단위로 분할된 복수개의 샷 영상들을 의미적 연속성을 갖는 장면 단위로 병합할 수 있다.
- [0091] 예를 들어, 영상에 대한 특징 분석(S220)을 통해 추출된 특징 정보들을 별도의 샷 분할 모듈을 이용하여 1~3초 내외의 샷으로 분할할 수 있다. 이렇게 분할된 샷은 장면 분할 모듈을 통해 의미적으로 연속성을 갖는 장면으로 병합되어 장면 단위의 영상으로 생성될 수 있다.
- [0092] 이와 같이 장면 단위의 영상은 샷 분할, 장면 분할 과정을 통해 생성될 수 있으며, 영상을 장면 단위로 분할하는 방법은 특정한 방법에 한정되지 않는다.
- [0093] 이 때, 입력 각본은 별도의 각본 분할 모듈을 통해서 영상과 동일한 장면 단위로 분할되어 장소, 시간, 지문 형식의 인스턴스로 변화될 수 있다.
- [0094] 따라서, 이와 같이 분할된 각본은 별도의 동기화 모듈을 기반으로 장면 단위로 분할된 영상과 동기화될 수 있다.
- [0095] 또한, 프로세서(1020)는 동기화된 장면 단위의 영상 및 각본 각각에 대해 적어도 하나의 특징에 상응하는 N차원의 벡터값을 추출한다.
- [0096] 예를 들어, 도 2를 참조하면, 장면 분할(S240) 단계를 통해 장면 단위로 분할된 영상을 기반으로 적어도 하나의 특징에 상응하는 N차원 벡터값을 추출할 수 있다(S260).
- [0097] 이 때, 장면 단위의 영상에서는 영상의 행위 정보, 장소 정보, 시간 정보, 객체 정보, 지문 정보에 상응하는 특징 정보들이 추출될 수 있다.
- [0098] 예를 들어, 행위 정보는 '달린다', '매린다', '이야기한다' 등과 같이 인물의 행동을 기준으로 생성되는 정보에 상응할 수 있다. 장소 정보는 영상의 배경이 되는 장소에 상응하는 것으로 '공원', '학교', '도로' 등이 될 수 있다. 시간 정보는 '밤', '낮', '아침' 같이 영상의 시간적 배경이 되는 정보에 상응할 수 있다. 객체 정보는 '전화기', '컵', '벤치' 와 같이 영상에 나타난 객체들에 대한 정보에 해당할 수 있다. 지문 정보는 영상 캡셔닝의 결과로써 '남자와 여자가 이야기하고 있다' 또는 '도시의 야경이 펼쳐지고 있다' 등 영상을 설명하는 정보에 상응할 수 있다.
- [0099] 이 때, 도 3을 참조하면, 본 발명에서는 메타데이터 생성 장치에 포함된 정보 추출 모듈(310)로 장면 영상을 입력함으로써 적어도 하나의 특징에 상응하는 N차원 벡터값(320)을 추출할 수 있다.
- [0100] 이 때, 각각의 벡터값들은 정보 추출 모듈(310)에 포함된 각 인식 모델을 통해 추출된 벡터에 상응하는 것으로써 분류 태스크에서 Softmax 함수를 통해 클래스 정보를 텍스트로 표현할 수 있다. 이 때, 본 발명에서는 Softmax 함수를 사용하기 전의 잠재 벡터를 장면에 대한 특징 정보로 사용할 수 있다.
- [0101] 또한, 프로세서(1020)는 N차원의 벡터값을 기반으로 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델에 대한 기계학습을 수행하여 메타데이터 생성 모델을 구축한다.
- [0102] 이 때, 영상 잠재 벡터 모델은 장면 단위의 영상에 대한 메타데이터를 생성하고, 각본 잠재 벡터 모델은 장면 단위의 각본에 대한 메타데이터를 생성할 수 있다.

[0103] 이 때, 손실 함수를 기반으로 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델 각각에 포함된 인코더 및 디코더를 영상 잠재 벡터 모델과 각본 잠재 벡터 모델 사이의 차이가 감소하도록 학습시킬 수 있다.

[0104] 이 때, 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델 각각에 상응하는 적어도 하나의 입력 벡터 및 출력 벡터를 획득할 수 있다.

[0105] 예를 들어, 도 4를 참조하면, 영상 잠재 벡터 모델(410)은 정보 추출 모듈을 통해 입력되는 장면 단위의 영상에 대한 N차원 벡터값을 입력값으로 학습될 수 있고, 각본 잠재 벡터 모델(420)은 각본 분할 모듈을 통해 입력되는 장면 단위의 각본에 대한 N차원 벡터값을 입력값으로 학습될 수 있다. 즉, 영상과 각본에서 각각 추출된 인스턴스 정보를 이용하여 잠재 벡터로 표현할 수 있는 인코더 및 디코더를 학습시킬 수 있다.

[0106] 이 때, 영상 잠재 벡터 모델(410)과 각본 잠재 벡터 모델(420)에 각각 포함된 잠재 벡터는 입력된 벡터값들을 인코더를 통해 압축 표현한 것으로써 다시 디코더를 통해 각각의 벡터로 재현할 수 있다.

[0107] 예를 들어, 장면 단위의 영상에서 추출된 장소 정보, 시간 정보, 행위 정보, 객체 정보 및 지문 정보들은 영상 잠재 벡터 모델(410)의 인코더를 통해 N차원의 영상 잠재 벡터로 표현될 수 있으며, 영상 잠재 벡터는 디코더를 통해 다시 장소 정보, 시간 정보, 행위 정보, 객체 정보 및 지문 정보로 재현될 수 있다.

[0108] 이와 마찬가지로 장면 단위의 각본에서 추출된 장소, 시간, 지문 인스턴스는 워드 임베딩 모델을 통해 N차원 벡터값으로 변환될 수 있고, 벡터값으로 변환된 각각의 인스턴스 정보들은 각본 잠재 벡터 모델(420)을 통해 N차원의 각본 잠재 벡터로 표현될 수 있다. 또한, 디코더를 통해 각본 잠재 벡터를 다시 벡터값으로 변환된 각각의 인스턴스 정보들로 재현할 수도 있다.

[0109] 이 때, 영상 잠재 벡터 모델과 각본 잠재 벡터 모델 간 적어도 하나의 입력 벡터 및 출력 벡터의 오차가 감소하도록 인코더 및 디코더의 변수를 조정할 수 있다.

[0110] 예를 들어, 손실함수는 [수학식 1]과 같이 영상 잠재 벡터 모델과 각본 잠재 벡터 모델 사이의 차이가 작아지도록 학습함으로써, 메타데이터 생성 모델이 입력된 영상 정보들과 각본 정보가 중재된 벡터를 가지도록 할 수 있다. 이는 추후에 영상이나 각본 중 하나의 리소스의 부재에도 두 리소스 전체를 반영한 메타데이터를 생성할 수 있는 효과가 있다.

[0111] [수학식 1]

$$\text{Loss} = \sum_{i=1}^m (\text{영상잠재벡터}_i - \text{각본잠재벡터}_i)^2$$

[0112]

[0113] 상기와 같이 메타데이터 생성 모델을 학습시키는 과정을 상세하게 나타내면 도 5와 같이 나타낼 수 있다.

[0114] 도 5를 참조하면, 영상과 각본이 입력되면(S510), 입력된 데이터를 구분하여(S515) 영상과 각본을 구별할 수 있다.

[0115] 이 후, 영상의 특징을 분석하여 추출하고(S520), 추출된 특징을 기반으로 샷 분할과 장면 분할을 수행하여(S530) 입력된 영상에 대한 장면 단위의 영상을 생성할 수 있다. 각본 또한 영상에서 추출된 특징을 기반으로 분할하고(S540), 장면 단위의 각본에서 정보를 추출할 수 있다(S550).

[0116] 이 후, 장면 단위의 영상과 각본을 동기화할 수 있다(S560).

[0117] 이 후, 또 다시 데이터를 영상과 각본으로 구분하여(S565), 장면 단위의 영상에 대해 추출된 특징에 상응하는 N차원 벡터값을 추출한 뒤 영상 잠재 벡터 모델을 학습시키고(S570), 각본에 대해 추출된 특징에 상응하는 N차원 벡터값을 추출한 뒤 각본 잠재 벡터 모델을 학습시킬 수 있다(S580).

[0118] 이 때, 손실 함수를 이용하여 영상 잠재 벡터 모델 및 각본 잠재 벡터 모델 각각에 포함된 인코더 및 디코더를 영상 잠재 벡터 모델과 각본 잠재 벡터 모델 사이의 차이가 감소하도록 학습시킬 수 있다.

[0119] 이와 같은 학습을 통해 메타데이터 생성 모델을 구축할 수 있다(S590).

[0120] 이하에서는 학습된 메타데이터 생성 모델을 기반으로 메타데이터를 생성하는 단계를 설명하도록 한다.

[0121] 또한, 프로세서(1020)는 타겟 영상 및 타겟 각본 중 어느 하나만 입력된 경우, 메타데이터 생성 모델을 기반으

로 타겟 영상과 타겟 각본이 중재된 형태의 메타데이터를 생성한다.

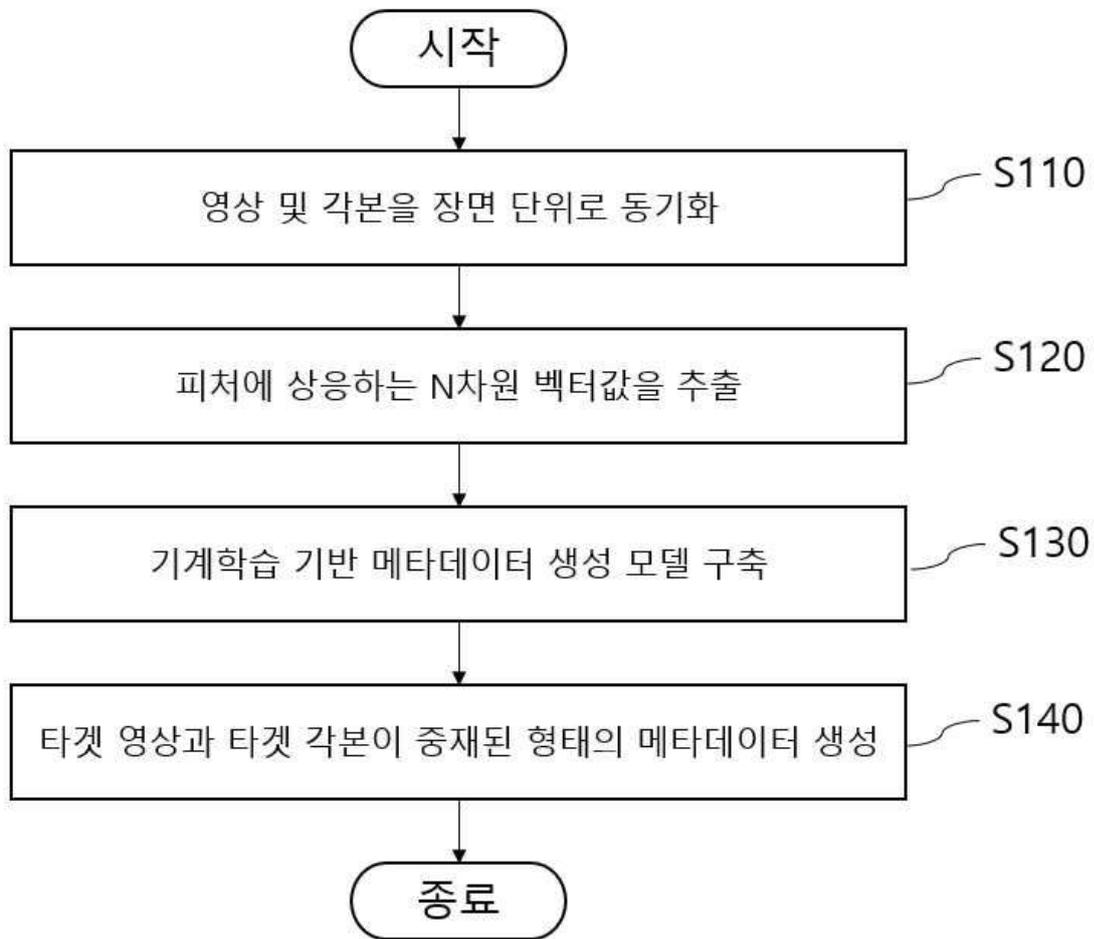
- [0122] 예를 들어, 도 6 내지 도 7을 참조하면, 타겟 영상(610)만 입력되었다고 가정하면(S710), 타겟 영상에 대한 특징을 분석 및 추출할 수 있다(S720). 이 후, 추출된 특징을 기반으로 타겟 영상에 대한 샷 분할 및 장면 분할을 수행하고(S730), N차원 벡터값을 추출하여 학습된 메타데이터 생성 모델(620)로 입력할 수 있다(S740). 이 후, 메타데이터 생성 모델(620)이 영상 및 각본이 중재된 메타데이터(630)를 생성하여 제공하면(S750), 별도의 저장소에 장면 단위의 영상에 대해 생성된 메타데이터(630)를 저장할 수 있다(S760).
- [0123] 다른 예를 들어, 도 8 내지 도 9를 참조하면, 타겟 각본(810)만 입력되었다고 가정하면(S910), 타겟 각본에 대한 정보를 추출할 수 있다(S920). 이 후, 추출된 정보를 기반으로 N차원 벡터값을 추출하여 학습된 메타데이터 생성 모델(820)로 입력할 수 있다(S930). 이 후, 메타데이터 생성 모델(820)이 영상 및 각본이 중재된 메타데이터(830)를 생성하여 제공하면(S940), 별도의 저장소에 장면 단위의 각본에 대해 생성된 메타데이터(830)를 저장할 수 있다(9560).
- [0124] 이 때, 메타데이터는 실수형의 벡터 형식에 상응할 수 있다.
- [0125] 예를 들어, 메타데이터는 아래와 같이 실수형의 벡터 형식으로 표현될 수 있다. '-0.3188 -0.456 0.1647 -0.2495 -0.9385 -0.3563 -0.0732 0.4643 0.2975 0.5199 -0.393 -0.0246 0.3141 -0.2959 0.4304 -0.471 0.3575 -0.0843 -0.213 -0.1231 0.0305 -0.1432 0.1642 -0.3709 -0.3492 0.3128 -0.3125 0.197 -0.3428 0.0803 0.7957 0.2245 -0.1008 -0.2649 -0.2901 0.1465 -0.2685 0.0587 0.3844 -0.0293 -0.0189 0.352 0.058 0.9351 0.0297 -0.5063 -0.256 0.6041 -0.1265 0.1885 -0.0964 -0.3505 -0.2208 -0.6876 -0.4202 0.0777 -0.0117 0.6559 0.3044 0.1288 -0.0751 -0.8191 -0.2499 -0.3286 -0.0498 -0.0646 0.2347 -0.0313 -0.1417 0.0131 -0.0557 0.4898 0.2188 -0.4096 -0.0245 -0.4827 0.0847 0.2517 0.203 0.2854 -0.0851 -0.2978 -0.0002 -0.4072 0.4154 0.4654 -0.1107 0.7675 0.2345 - 썸'
- [0126] 메모리(1030)는 메타데이터 생성 모델을 저장할 수 있다.
- [0127] 또한, 메모리(1030)는 메타데이터 생성 과정에서 발생하는 다양한 정보를 저장할 수 있다.
- [0128] 이와 같은 메타데이터 생성 장치를 통해 영상을 기반으로 장면을 묘사하는 정보를 추출하고, 영상에 대한 의미적 메타데이터를 생성 및 제공할 수 있다.
- [0129] 또한, 기계학습 기반의 영상 분석을 통해 영상을 묘사하는 장소, 시간, 행동 등과 같은 의미적인 정보들을 추출할 수 있으며, 이러한 정보들에 기반한 학습을 통해 통합된 하나의 벡터 형식의 의미적 메타데이터를 생성할 수도 있다.
- [0130] 또한, 각본의 수급이 비교적 어려운 환경에서 각본이 없이도 영상만으로 각본을 충분히 고려한 메타데이터를 제공하는 효과를 얻을 수도 있다.
- [0132] 이상에서와 같이 본 발명에 따른 벡터를 이용한 장면 묘사 기반의 메타데이터 생성 방법 및 이를 위한 장치는 상기한 바와 같이 설명된 실시예들의 구성과 방법이 한정되게 적용될 수 있는 것이 아니라, 상기 실시예들은 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.

부호의 설명

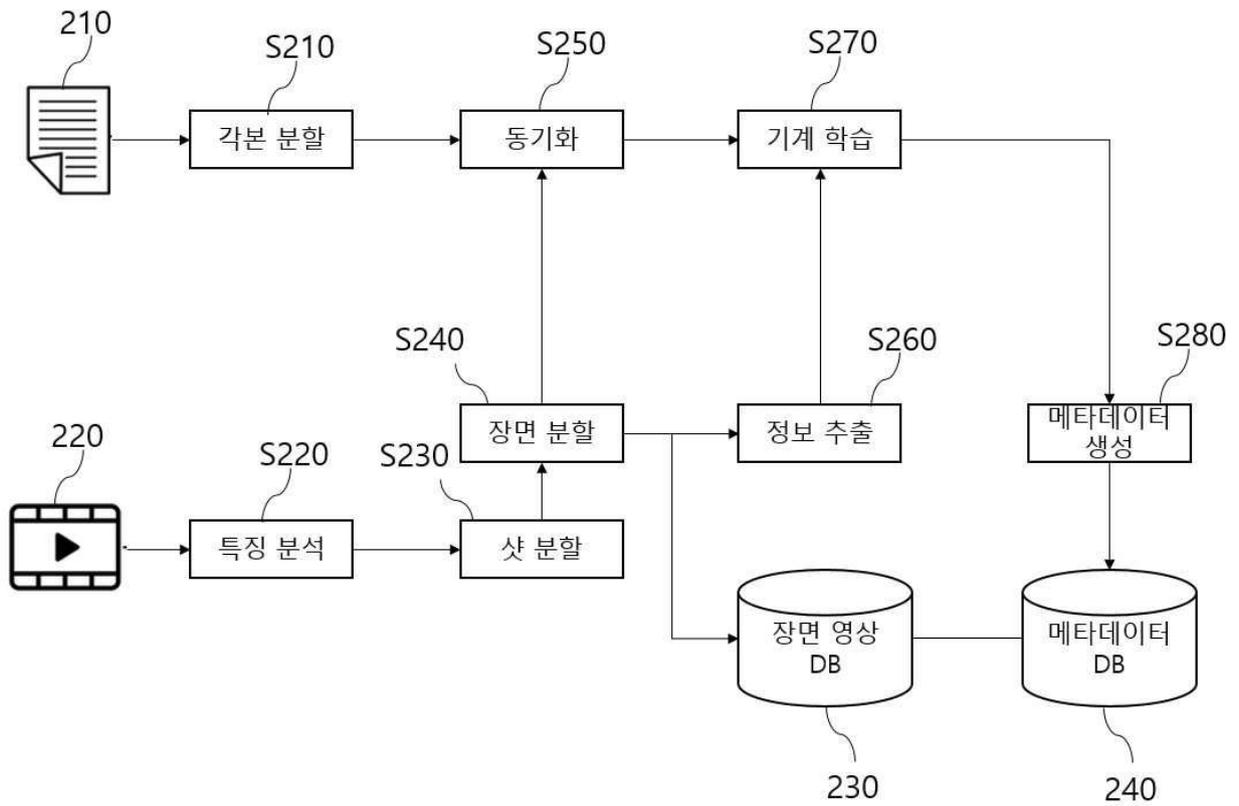
- [0133] 210: 각본 220: 영상
- 230: 장면 영상 데이터베이스 240: 메타데이터 데이터베이스
- 310: 정보 추출 모듈 320: N차원 벡터값
- 410: 영상 잠재 벡터 모델 420: 각본 잠재 벡터 모델
- 610: 타겟 영상 620, 820: 메타데이터 생성 모델
- 630, 830: 메타데이터 810: 타겟 각본
- 1010: 통신부 1020: 프로세서
- 1030: 메모리

도면

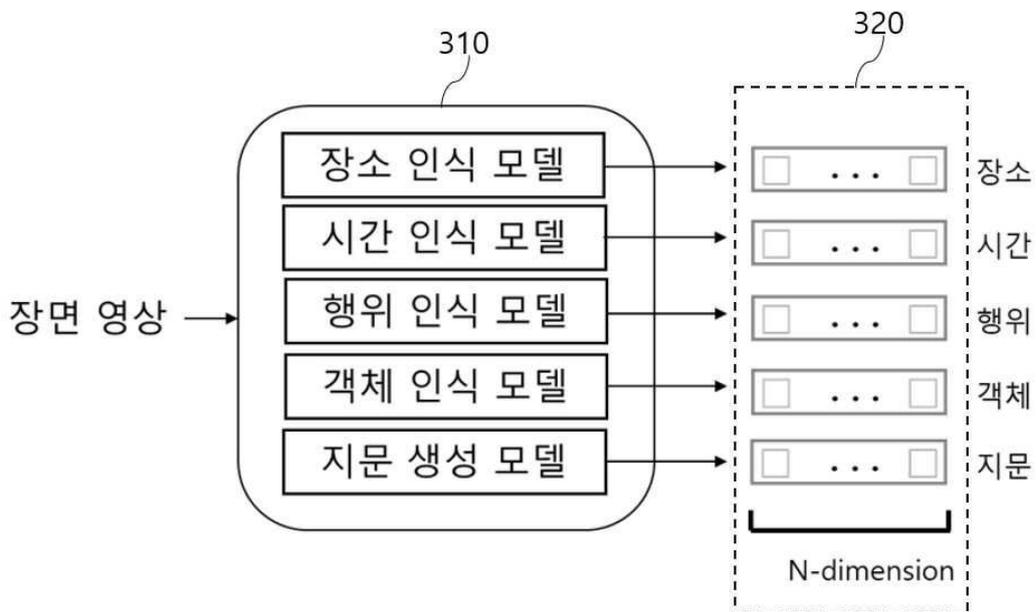
도면1



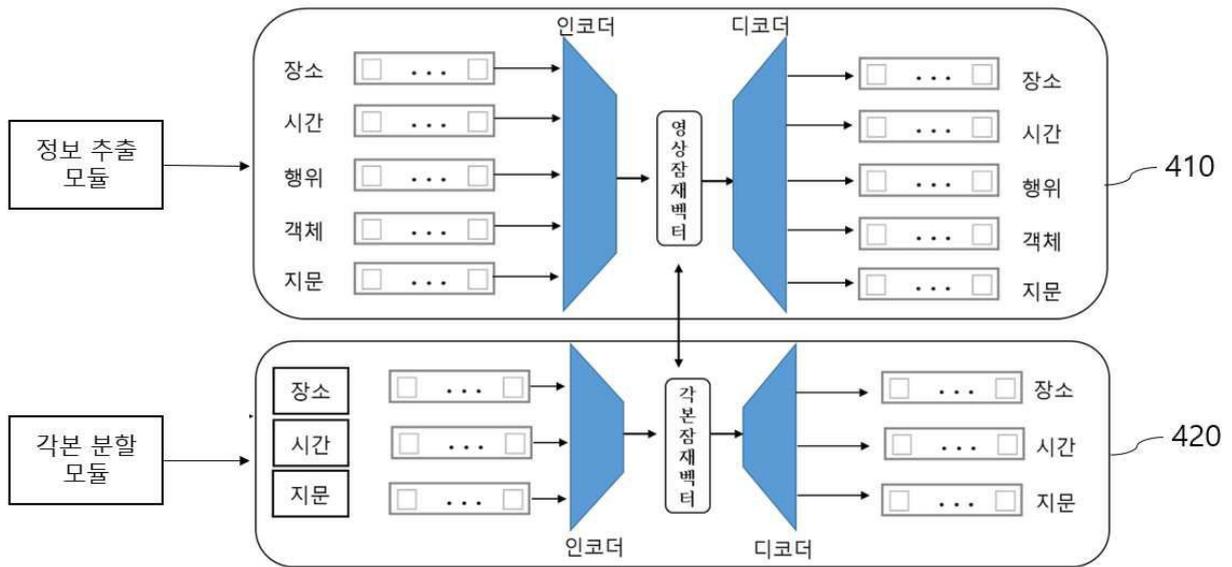
도면2



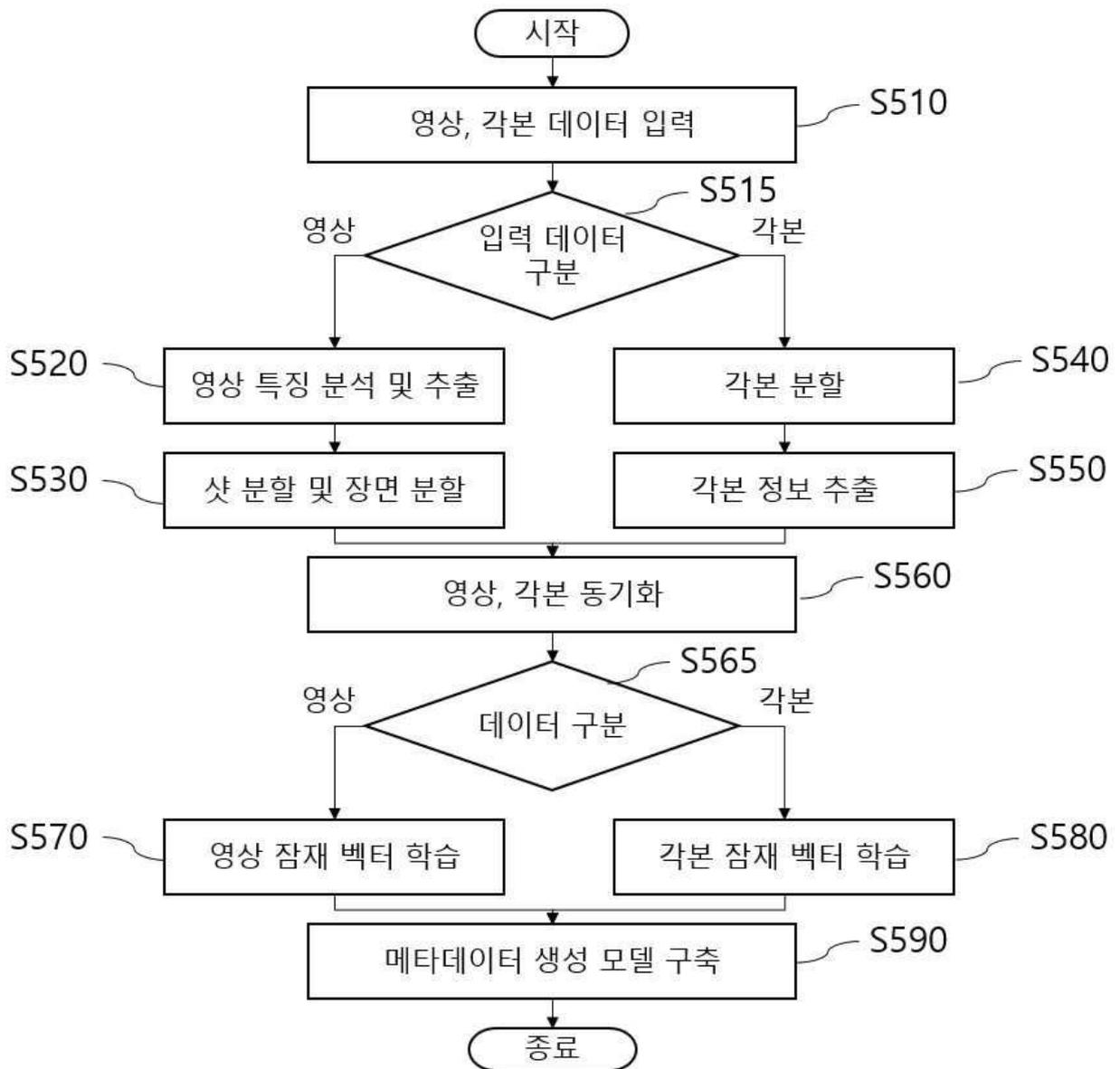
도면3



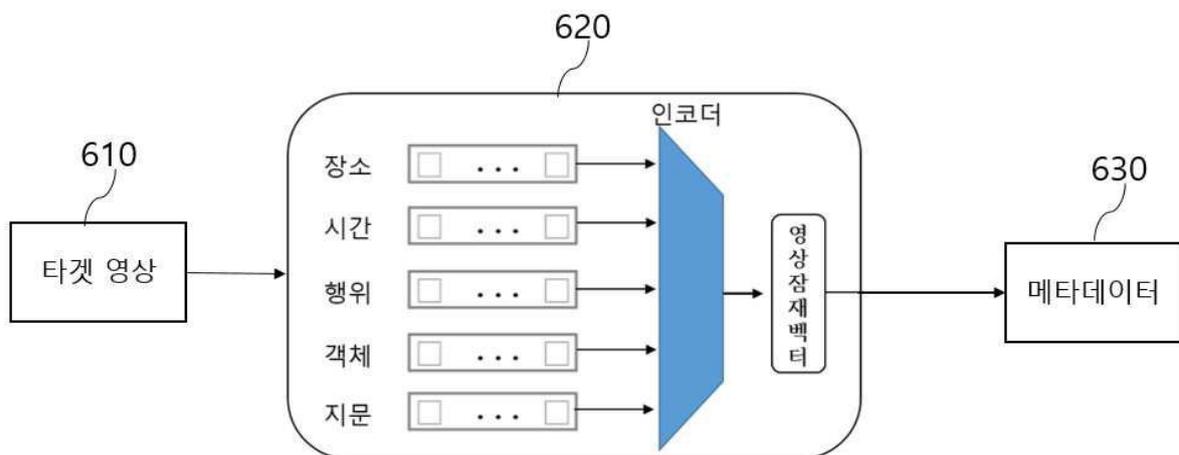
도면4



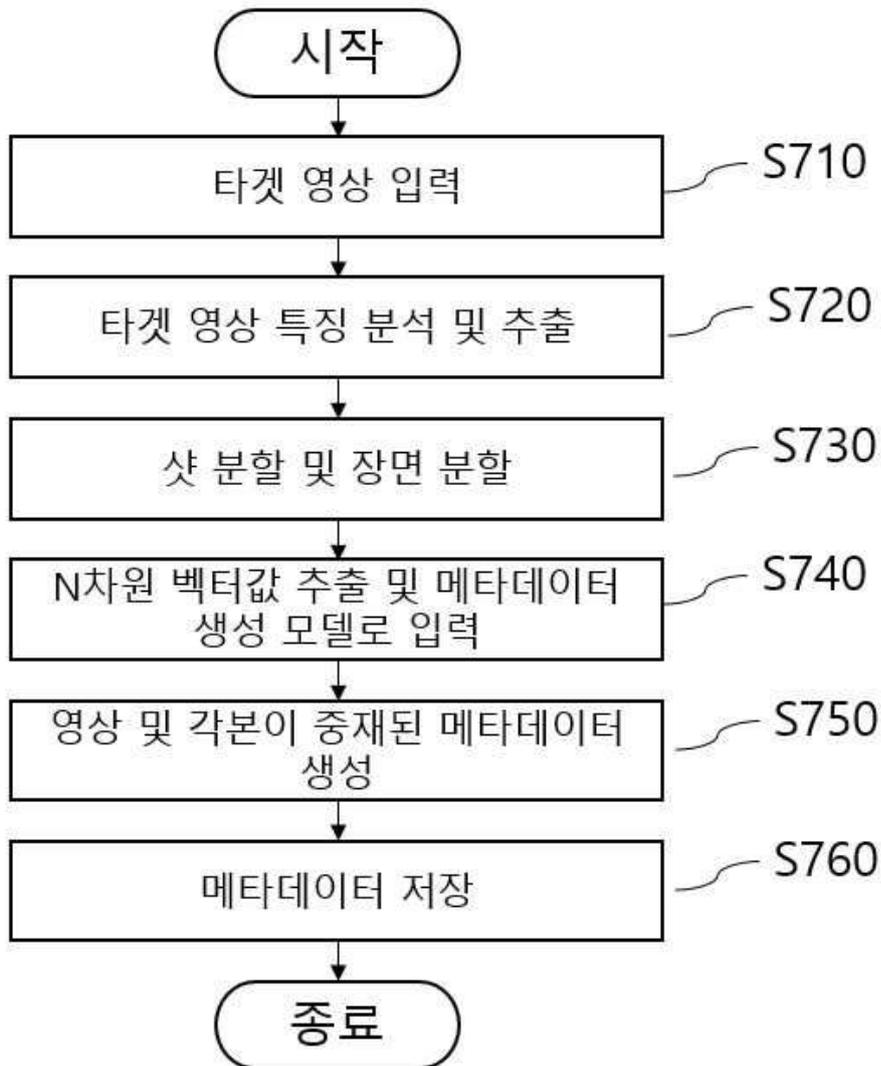
도면5



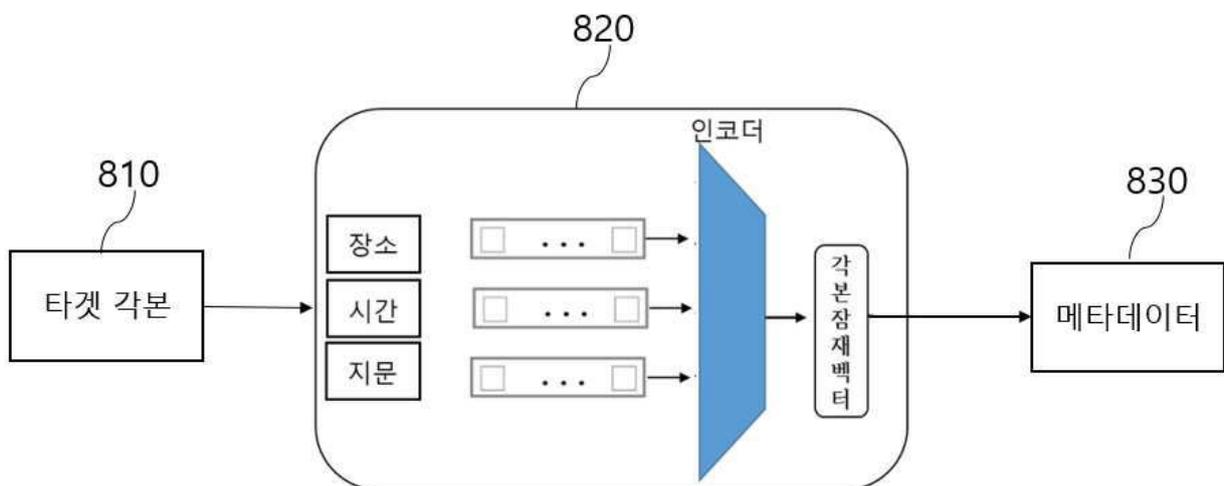
도면6



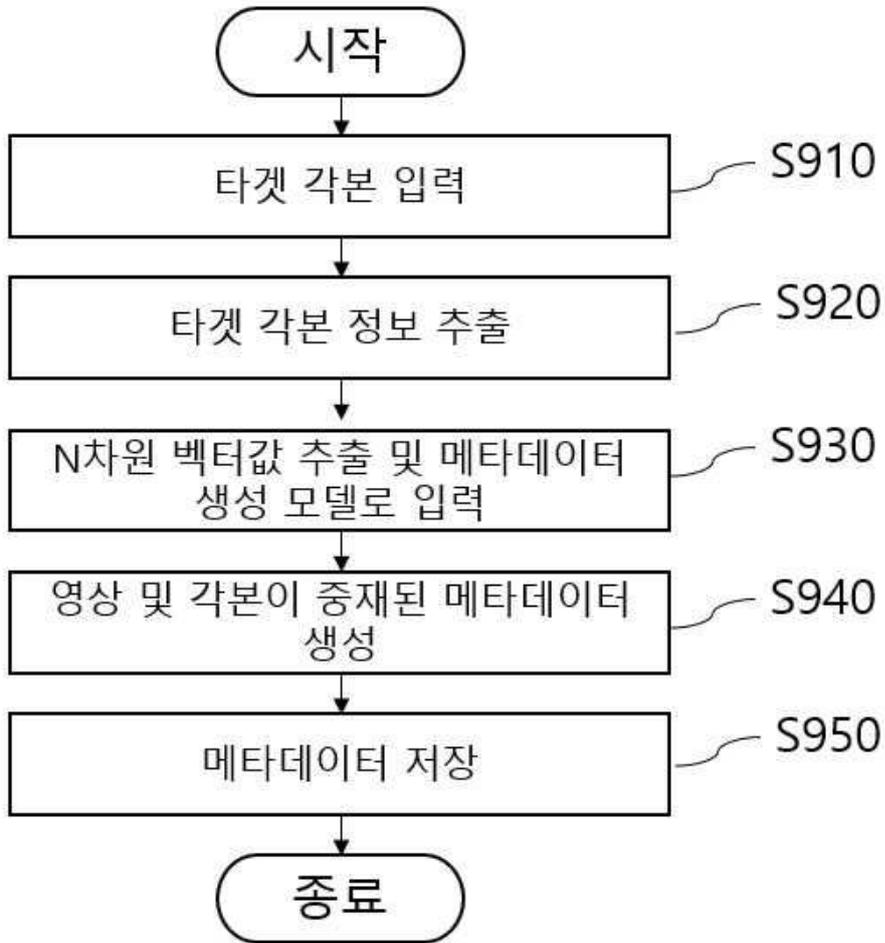
도면7



도면8



도면9



도면10

