(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0072875
(43) 공개일자 2019년06월26일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2006.01)
(52) CPC특허분류
G06K 9/00744 (2013.01)
G06K 9/00765 (2013.01)
(21) 출원번호 10-2017-0173999
(22) 출원일자 2017년12월18일
심사청구일자 2017년12월18일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
낭종호
서울특별시 강남구 언주로146길 18 (논현2동, 동현아파트) 1동 206호
신광수
서울특별시 관악구 신림로14길 27-9, 401호
사공락
서울특별시 송파구 문정2동 훼미리아파트 307동 1001호
(74) 대리인
이지연

전체 청구항 수 : 총 8 항

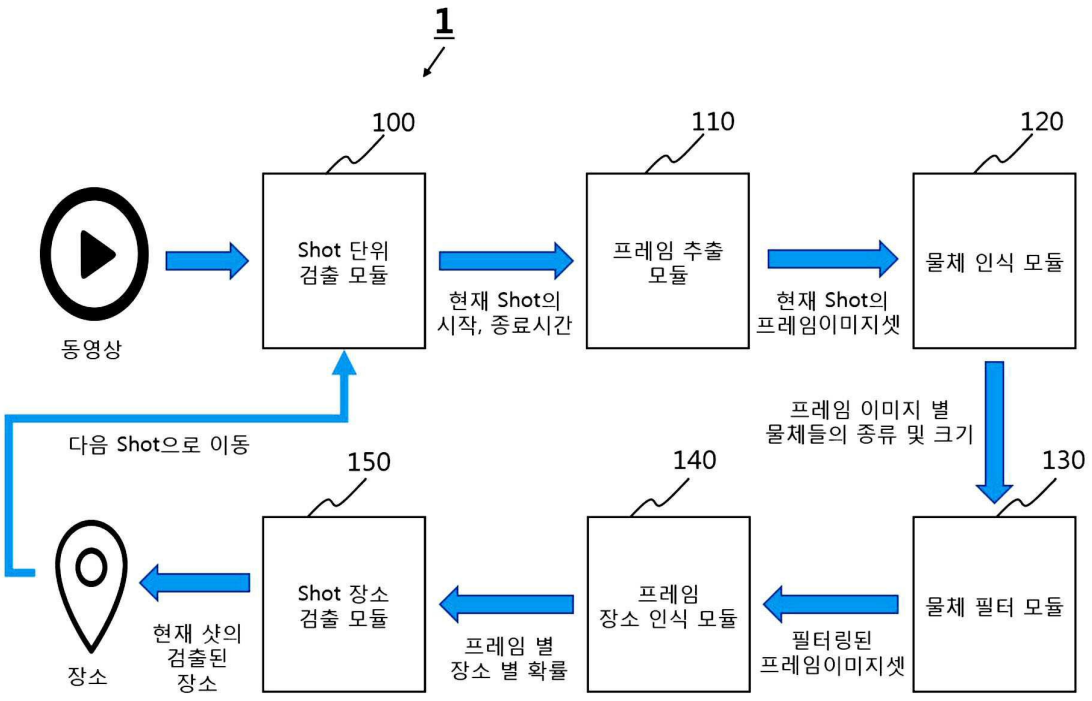
(54) 발명의 명칭 **동영상에서의 장소 추출 방법 및 장소 추출 시스템**

(57) 요 약

본 발명은 동영상 내 샷 단위의 배경 장소 추출 방법 및 시스템에 관한 것이다. 상기 배경 장소 추출 시스템은, 입력된 동영상에서 샷 단위를 검출하는 샷 단위 검출 모듈; 하나의 샷에 대하여 사전 설정된 기준에 따라 적어도 둘 이상의 프레임들을 추출하여 해당 샷에 대한 프레임 이미지 셋을 생성하는 프레임 추출 모듈; 샷에 대한 프레

(뒷면에 계속)

대 표 도 - 도3



임 이미지 셋을 구성하는 각 프레임들에 대하여, 각 프레임 이미지들에 포함된 물체들을 인식하는 물체 인식 모듈; 각 프레임 이미지들에서 인식된 물체가 사전 설정된 조건을 만족하지 못하는 경우, 해당 프레임 이미지들을 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 물체 필터 모듈; 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대하여 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 검출하는 프레임 장소 인식 모듈; 상기 프레임 장소 인식 모듈에 의해 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대해 검출된 배경 장소 확률값들을 이용하여 해당 샷에 대한 배경 장소를 추출하는 샷 장소 검출 모듈; 을 구비하여, 동영상의 샷 단위로 배경 장소를 추출한다.

(52) CPC특허분류
G06T 2207/20024 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 2017-0-00271
부처명 과학기술정보통신부
연구관리전문기관 정보통신기술진흥센터
연구사업명 정보통신·방송연구개발 사업
연구과제명 아카이브 솔루션 및 콘텐츠 플랫폼 개발
기 여 율 1/1
주관기관 (주)아이온커뮤니케이션즈
연구기간 2017.04.01 ~ 2017.12.31

명 세 서

청구범위

청구항 1

(a) 동영상에서 샷 단위를 검출하는 단계;

(b) 검출된 샷에 대하여 사전 설정된 기준에 따라 적어도 둘 이상의 프레임들을 추출하여 해당 샷에 대한 프레임 이미지 셋을 생성하는 단계;

(c) 상기 프레임 이미지 셋을 구성하는 각 프레임들에 대하여, 각 프레임 이미지들에 포함된 물체들을 인식하는 단계;

(d) 각 프레임 이미지들에서 인식된 물체가 사전 설정된 조건을 만족하지 못하는 경우, 해당 프레임 이미지들을 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 단계;

(e) 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대하여 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 검출하는 단계;

(f) 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대해 검출된 배경 장소 확률값들을 이용하여 해당 샷에 대한 배경 장소를 추출하는 단계;

를 구비하여, 동영상의 샷 단위로 배경 장소를 추출하는 것을 특징으로 하는 동영상내 배경 장소 추출 방법.

청구항 2

제1항에 있어서, 상기 (b) 단계에서 생성된 프레임 이미지 셋은 샷의 시작 프레임, 종료 프레임 및 사전 설정된 시각 간격당 하나씩 추출된 프레임들을 구비하는 것을 특징으로 하는 동영상내 배경 장소 추출 방법.

청구항 3

제1항에 있어서, 상기 (d) 단계는,

상기 물체의 종류가 사전 설정된 물체 리스트 중 하나에 대응되고, 상기 물체의 크기가 사전 설정된 기준값보다 큰 경우, 해당 프레임 이미지들은 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 것을 특징으로 하는 동영상내 배경 장소 추출 방법.

청구항 4

제항에 있어서, 상기 (f) 단계는

각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합하여 각각의 배경 장소들에 대한 확률합을 구하고, 가장 높은 확률합을 갖는 배경 장소를 해당 샷의 배경 장소로 결정하거나,

각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합한 후 각각의 배경 장소들에 대한 평균 확률값 및 각각의 배경 장소들에 대한 중간 확률값을 구하고, 평균 확률값과 중간 확률값을 이용하여 해당 샷의 배경 장소를 결정하는 것을 특징으로 하는 동영상내 배경 장소 추출 방법.

청구항 5

입력된 동영상에서 샷 단위를 검출하는 샷 단위 검출 모듈;

하나의 샷에 대하여 사전 설정된 기준에 따라 적어도 둘 이상의 프레임들을 추출하여 해당 샷에 대한 프레임 이미지 셋을 생성하는 프레임 추출 모듈;

샷에 대한 프레임 이미지 셋을 구성하는 각 프레임들에 대하여, 각 프레임 이미지들에 포함된 물체들을 인식하는 물체 인식 모듈;

각 프레임 이미지들에서 인식된 물체가 사전 설정된 조건을 만족하지 못하는 경우, 해당 프레임 이미지들을 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 물체 필터 모듈;

상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대하여 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 검출하는 프레임 장소 인식 모듈;

상기 프레임 장소 인식 모듈에 의해 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대해 검출된 배경 장소 확률값들을 이용하여 해당 샷에 대한 배경 장소를 추출하는 샷 장소 검출 모듈;

을 구비하여, 동영상의 샷 단위로 배경 장소를 추출하는 것을 특징으로 하는 동영상내 배경 장소 추출 시스템.

청구항 6

제5항에 있어서, 상기 프레임 추출 모듈에서 생성된 프레임 이미지 셋은 샷의 시작 프레임, 종료 프레임 및 사전 설정된 시각 간격당 하나씩 추출된 프레임들을 구비하는 것을 특징으로 하는 동영상내 배경 장소 추출 시스템.

청구항 7

제5항에 있어서, 상기 물체 필터 모듈은,

상기 물체의 종류가 사전 설정된 물체 리스트 중 하나에 대응되고 상기 물체의 크기가 사전 설정된 기준값보다 큰 경우, 해당 프레임 이미지들은 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 것을 특징으로 하는 동영상내 배경 장소 추출 시스템.

청구항 8

제5항에 있어서, 상기 샷 장소 검출 모듈은,

각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합하여 각각의 배경 장소들에 대한 확률합을 구하고, 가장 높은 확률합을 갖는 배경 장소를 해당 샷의 배경 장소로 결정하거나,

각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합한 후 각각의 배경 장소들에 대한 평균 확률값 및 각각의 배경 장소들에 대한 중간 확률값을 구하고, 평균 확률값과 중간 확률값을 이용하여 해당 샷의 배경 장소를 결정하는 것을 특징으로 하는 동영상내 배경 장소 추출 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 동영상에서의 샷단위의 장소 추출 방법 및 시스템에 관한 것으로서, 더욱 구체적으로는 이미지 기반의 딥러닝 모델을 이용하여, 다수 개의 프레임들로 구성되는 샷단위의 배경 장소를 추출하여 분류하는 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 대용량 멀티미디어 데이터베이스의 증가와 통신 및 디지털 미디어 처리 기술의 발달로 인하여 수많은 동영상 비디오가 출현함에 따라, 축약된 동영상 비디오의 요약 정보를 바탕으로 한 검색 서비스를 통해 사용자의 편의와 만족을 도모하기 위한 노력이 증대되고 있다.

[0003] 따라서, 동영상 비디오와 관련된 다양한 형태의 상업이 전개되면서 대량의 동영상 비디오를 자동으로 분석하고자 하는 필요성이 증대되고 있으며, 이를 해결하기 위한 많은 연구들이 활발하게 진행되고 있다.

[0004] 종래의 기술에 따르면, 동영상에 대한 장소 분류기는 동영상에 대한 장소 또는 배경 장소에 대한 정보를 추출하여 제공하는 것으로서, 동영상에 대한 각 프레임 단위로 배경 장소를 추출하게 된다.

[0005] 한편, 동영상 콘텐츠에서 샷(Shot)은 일반적으로 카메라가 한 번의 촬영을 시작하여 멈추기 전까지를 의미하므

로, 샷은 화면전환의 경계를 나타낼 수 있으며 비디오 콘텐츠를 나누는 기본적인 단위가 되기도 한다. 따라서, 하나의 샷은 일반적으로 하나의 배경 장소로 이루어지게 된다.

[0006] 종래의 장소 분류기는 동영상을 각 프레임 단위로 배경 장소를 추출하게 되는데, 이 경우 단일의 샷을 구성하는 프레임들이 동일한 배경 장소를 가짐에도 불구하고, 각 프레임들의 전경 물체 등으로 인하여 배경 장소가 일정하게 추출되지 않는 문제점이 발생한다.

[0007] 도 1은 종래의 장소 분류기가 각 프레임별로 배경 장소를 분류한 결과를 예시적으로 도시한 것이다. 도 1에 도시된 바와 같이, 하나의 샷을 구성하는 2개의 프레임들이 배경 장소에 대한 서로 다른 확률값으로 도출하게 된다. 그 결과 하나의 샷이 동일한 배경 장소임에도 불구하고 각 프레임마다 서로 다른 분류 결과를 가져오게 되는 문제점이 발생하게 된다.

[0008] 한편, 종래의 장소 분류기는 프레임의 전경 물체의 크기가 큰 경우, 배경 장소가 잘 보이지 않게 되어, 배경 장소에 대한 분류를 제대로 하지 못하는 문제점이 있다.

[0009] 도 2는 종래의 장소 분류기가 전경 물체로 인하여 배경 장소 분류를 제대로 하지 못하는 프레임들을 예시적으로 도시한 것이다. 도 2에 도시된 바와 같이, 각 프레임 이미지에 포함된 전경 물체의 크기가 크거나 프레임 이미지에 다수 개의 전경 물체가 다양하게 배치된 경우, 해당 프레임 이미지의 배경 장소를 정확하게 추출해내지 못하게 된다.

선행기술문헌

특허문헌

[0010] (특허문헌 0001) 한국공개특허

(특허문헌 0002) 한국공개특허공보 제 10-2004-0033766 호

(특허문헌 0003) 한국등록특허공보 제 10-1706365 호

발명의 내용

해결하려는 과제

[0011] 전술한 문제점을 해결하기 위한 본 발명의 목적은 동영상에서 하나의 샷에 대한 배경 장소를 정확하게 추출할 수 있는 방법 및 시스템을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0012] 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징에 따른 동영상내 배경 장소 추출 방법은, (a) 동영상에서 샷 단위를 검출하는 단계; (b) 검출된 샷에 대하여 사전 설정된 기준에 따라 적어도 둘 이상의 프레임들을 추출하여 해당 샷에 대한 프레임 이미지 셋을 생성하는 단계; (c) 상기 프레임 이미지 셋을 구성하는 각 프레임들에 대하여, 각 프레임 이미지들에 포함된 물체들을 인식하는 단계; (d) 각 프레임 이미지들에서 인식된 물체가 사전 설정된 조건을 만족하지 못하는 경우, 해당 프레임 이미지들을 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 단계; (e) 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대하여 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 검출하는 단계; 및 (f) 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대해 검출된 배경 장소 확률값들을 이용하여 해당 샷에 대한 배경 장소를 추출하는 단계;를 구비하여, 동영상의 샷 단위로 배경 장소를 추출한다.

[0013] 전술한 제1 특징에 따른 동영상내 배경 장소 추출 방법에 있어서, 상기 (b) 단계에서 생성된 프레임 이미지 셋은 샷의 시작 프레임, 종료 프레임 및 사전 설정된 시각 간격당 하나씩 추출된 프레임들을 구비하는 것이 바람직하다.

[0014] 전술한 제1 특징에 따른 동영상내 배경 장소 추출 방법에 있어서, 상기 (d) 단계는, 상기 물체의 종류가 사전 설정된 물체 리스트 중 하나에 대응되고, 상기 물체의 크기가 사전 설정된 기준값보다 큰 경우, 해당 프레임 이미지들은 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 것이 바람직하다.

[0015] 전술한 제1 특징에 따른 동영상내 배경 장소 추출 방법에 있어서, 상기 (f) 단계는 각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합하여 각각의 배경 장소들에 대한 확률합을 구하고, 가장 높은 확률합을 갖는 배경 장소를 해당 샷의 배경 장소로 결정하거나,

[0016] 각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합한 후 각각의 배경 장소들에 대한 평균 확률값 및 각각의 배경 장소들에 대한 중간 확률값을 구하고, 평균 확률값과 중간 확률값을 이용하여 해당 샷의 배경 장소를 결정하는 것이 바람직하다.

[0017] 본 발명의 제2 특징에 따른 동영상내 배경 장소 추출 시스템은, 입력된 동영상에서 샷 단위를 검출하는 샷 단위 검출 모듈; 하나의 샷에 대하여 사전 설정된 기준에 따라 적어도 둘 이상의 프레임들을 추출하여 해당 샷에 대한 프레임 이미지 셋을 생성하는 프레임 추출 모듈; 샷에 대한 프레임 이미지 셋을 구성하는 각 프레임들에 대하여, 각 프레임 이미지들에 포함된 물체들을 인식하는 물체 인식 모듈; 각 프레임 이미지들에서 인식된 물체가 사전 설정된 조건을 만족하지 못하는 경우, 해당 프레임 이미지들을 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 물체 필터 모듈; 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대하여 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 검출하는 프레임 장소 인식 모듈; 및 상기 프레임 장소 인식 모듈에 의해 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대해 검출된 배경 장소 확률값들을 이용하여 해당 샷에 대한 배경 장소를 추출하는 샷 장소 검출 모듈;을 구비하여, 동영상의 샷 단위로 배경 장소를 추출한다.

[0018] 전술한 제2 특징에 따른 동영상내 배경 장소 추출 시스템에 있어서, 상기 프레임 추출 모듈에서 생성된 프레임 이미지 셋은 샷의 시작 프레임, 종료 프레임 및 사전 설정된 시각 간격당 하나씩 추출된 프레임들을 구비하는 것이 바람직하다.

[0019] 전술한 제2 특징에 따른 동영상내 배경 장소 추출 시스템에 있어서, 상기 물체 필터 모듈은, 상기 물체의 종류가 사전 설정된 물체 리스트 중 하나에 대응되고 상기 물체의 크기가 사전 설정된 기준값보다 큰 경우, 해당 프레임 이미지들은 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하는 것이 바람직하다.

[0020] 전술한 제2 특징에 따른 동영상내 배경 장소 추출 시스템에 있어서, 상기 샷 장소 검출 모듈은, 각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합하여 각각의 배경 장소들에 대한 확률합을 구하고, 가장 높은 확률합을 갖는 배경 장소를 해당 샷의 배경 장소로 결정하거나,

[0021] 각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합한 후 각각의 배경 장소들에 대한 평균 확률값 및 각각의 배경 장소들에 대한 중간 확률값을 구하고, 평균 확률값과 중간 확률값을 이용하여 해당 샷의 배경 장소를 결정하는 것이 바람직하다.

발명의 효과

[0022] 본 발명에 따른 샷 단위의 배경 장소 추출 방법 및 시스템은 하나의 샷을 구성하는 다수 개의 프레임들에 대한 배경 장소별 확률값들을 구하고, 배경 장소들에 대한 확률합을 구하고 가장 높은 확률값을 갖는 배경 장소를 결정하거나, 배경 장소들에 대한 평균 확률값과 중간 확률값을 이용하여 해당 샷의 배경 장소를 결정하게 된다. 그 결과, 해당 샷에 대하여 항상 일정한 결과를 얻을 수 있게 된다.

[0023] 또한, 본 발명에 따른 샷 단위의 배경 장소 추출 방법 및 시스템은 샷에 대한 프레임 이미지 셋을 구성하는 프레임 이미지가 전경의 물체의 종류 및 크기 등이 사전 설정된 조건을 만족하지 못해 정확한 배경 장소를 추출하기 어려운 경우, 해당 프레임을 프레임 이미지 셋으로부터 제거하여 필터링함으로써, 보다 정확한 결과를 도출할 수 있게 된다.

도면의 간단한 설명

[0024] 도 1은 종래의 장소 분류기가 각 프레임별로 배경 장소를 분류한 결과를 예시적으로 도시한 것이다.

도 2는 종래의 장소 분류기가 전경 물체로 인하여 배경 장소를 정확하게 추출하지 못하는 프레임들을 예시적으로 도시한 것이다.

도 3은 본 발명의 바람직한 실시예에 따른 동영상 내 샷 단위의 배경 장소 추출 시스템(1)을 개략적으로 도시한 블록도이다.

도 4는 본 발명의 바람직한 실시예에 따른 동영상내 배경 장소 추출 시스템에 있어서, 프레임 추출 모듈에 의해

생성된 프레임 이미지 셋을 예시적으로 도시한 것이다.

도 5는 본 발명의 바람직한 실시예에 따른 동영상내 배경 장소 추출 시스템에 있어서, 물체 인식 모듈을 설명하기 위하여 도시된 모식도이다.

도 6은 본 발명의 바람직한 실시예에 따른 동영상 내 배경 장소 추출 시스템에 있어서, 물체 인식 모듈 및 물체 필터 모듈의 동작을 설명하는 흐름도이다.

도 7은 프레임 장소 인식 모듈인 PlaceNet Model을 개념적으로 도시한 구성도이다.

도 8은 본 발명의 바람직한 실시예에 따른 동영상 내 배경 장소 추출 시스템에 있어서, 프레임 장소 인식 모듈을 설명하기 위하여 도시한 구성도이다.

도 9는 본 발명의 바람직한 실시예에 따른 동영상 내 배경 장소 추출 시스템에 있어서, 샷 장소 검출 모듈에 의해 얻는 각 프레임들에 대하여 구한 배경 장소별 확률값들과 평균 확률값, 중간 확률값 및 결과값들을 예시적으로 도시한 도표이다.

발명을 실시하기 위한 구체적인 내용

[0025] 본 발명에 따른 동영상 내 샷(shot) 단위의 배경 장소 추출 방법 및 시스템은, 동영상으로부터 샷 단위를 검출하고, 검출된 샷 단위에 대하여 다수 개의 프레임들을 추출하여 프레임 이미지 셋을 생성하고, 프레임 이미지 셋을 구성하는 각 프레임들에 포함된 물체들에 따라 필터링하고, 필터링된 프레임 이미지 셋을 구성하는 프레임들에 대하여 배경 장소별 확률값들을 구하고, 이들을 이용하여 해당 샷에 대한 배경 장소를 결정하는 것을 특징으로 한다. 이와 같이, 본 발명에 따른 배경 장소 추출 방법은 샷을 구성하는 프레임들에 대한 배경 장소별 확률값들을 이용하여 해당 샷에 대한 배경 장소를 결정함으로써, 정확하면서도 일관된 결과를 도출할 수 있게 된다.

[0026] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 대한 동영상 내 샷 단위의 배경 장소 추출 방법 및 시스템에 대하여 구체적으로 설명한다. 본 발명에 따른 배경 장소 추출 시스템은 동영상에 대한 다양한 처리가 가능하고 CNN(Convolution Neural Network)과 같은 Deep learning 학습이 가능한 컴퓨터 등으로 구성될 수 있으며, 상기 배경 장소 추출 시스템을 구성하는 각 모듈들은 프로그램과 같은 소프트웨어로 구성될 수 있다. 따라서, 본 발명에 따른 배경 장소 추출 방법은, 상기 배경 장소 추출 시스템을 구성하는 각 모듈들을 구현하는 소프트웨어의 동작 방법들로 구성될 수 있다.

[0027] 도 3은 본 발명의 바람직한 실시예에 따른 동영상 내 샷 단위의 배경 장소 추출 시스템(1)을 개략적으로 도시한 블록도이다. 도 3을 참조하면, 본 발명에 따른 배경 장소 추출 시스템(1)은, 샷 단위 검출 모듈(100), 프레임 추출 모듈(110), 물체 인식 모듈(120), 물체 필터 모듈(130), 프레임 장소 인식 모듈(140), 및 샷 장소 검출 모듈(150)을 구비한다.

[0028] 상기 샷 단위 검출 모듈(100)은 입력된 동영상에서 샷 단위를 검출한다. 샷 단위 검출 모듈은 동영상을 샷 단위로 구분하고 해당 샷의 시작 시간과 종료 시간을 출력한다. 동영상에 대한 샷 단위 검출 모듈은 당업계에 널리 알려진 다양한 방법들 중 하나로 구현될 수 있으며, 간단한 Threshold 기반의 페이드-인(fade-in), 페이드-아웃(fade-out) 감지 방법 또는 컨텐츠 인식을 통한 샷 경계 검출 방법 등 여러 가지의 감지 방법을 통해 샷 경계를 검출하여 제공한다. 본 발명에서는 Python 기반으로 작성된 오픈소스인 PySceneDetect 모듈을 사용하였으며, PySceneDetect 모듈은 동영상을 입력값으로 제공하면, 동영상을 shot 단위로 구분하고, 각 shot의 시작 시간과 종료 시간을 출력한다.

[0029] 상기 프레임 추출 모듈(110)은 상기 샷 단위 검출 모듈에 의해 검출된 하나의 샷에 대하여 사전 설정된 기준에 따라 적어도 둘 이상의 프레임들을 추출하여 해당 샷에 대한 프레임 이미지 셋을 생성한다. 본 발명에 따른 프레임 추출 모듈(110)은, 각 샷에 대하여 샷의 시작 프레임, 종료 프레임 및 사전 설정된 시각 간격당 하나씩 프레임들을 추출하여 이들로 구성되는 프레임 이미지 셋을 생성한다.

[0030] 도 4는 본 발명의 바람직한 실시예에 따른 동영상내 배경 장소 추출 시스템에 있어서, 프레임 추출 모듈에 의해 생성된 프레임 이미지 셋을 예시적으로 도시한 것이다. 도 4를 참조하면, 본 발명에 따른 프레임 추출 모듈에 의하여, 하나의 샷에 대하여, 시작 프레임, 종료 프레임, 1초마다 추출된 8개의 프레임들을 추출하여 총 10개의 프레임들로 구성되는 프레임 이미지 셋을 생성하게 된다.

[0031] 상기 물체 인식 모듈(120)은 상기 프레임 추출 모듈에 의해 생성된 샷에 대한 프레임 이미지 셋을 구성하는 각

프레임들에 대하여, 각 프레임 이미지들에 포함된 물체들을 인식한다.

[0032] 도 5는 본 발명의 바람직한 실시예에 따른 동영상내 배경 장소 추출 시스템에 있어서, 물체 인식 모듈을 설명하기 위하여 도시된 모식도이다. 상기 물체 인식 모듈은 당업계에 널리 알려진 다양한 방법들 중 하나로 구현될 수 있으며, 본 발명에서는 CNN을 사용한 다물체 인식 모듈 중 하나인 YOLO("You Only Look Once") 모듈을 사용하였다. 도 5는 YOLO 모듈을 이용하여 프레임 이미지로부터 물체를 인식하는 과정을 도시한 것이다.

[0033] 도 5를 참조하면, 물체 인식 모듈은 먼저 프레임 이미지를 N×N 격자 모양으로 나누고, 격자들을 이용하여 bounding box 영역을 만든다. 다음, 프레임 이미지에서 bounding box 영역을 CNN Input image size 에 맞게 크기를 재조정한다. 다음, 크기를 재조정된 이미지를 입력으로 넣어, 해당 bounding box 영역에 대하여, 각각의 물체에 대한 확률을 구한다. Non-Maximum Suppression(NMS)을 이용하여 bounding box들 간의 물체에 대한 확률값에 따라 물체로 인식되는 부분을 서로 묶어 최종적으로 한개의 이미지에 대한 여러 물체의 위치와 크기, 그리고 확률값을 제공한다.

[0034] 그리고, 각 bounding box의 물체의 종류를 판단하고, 해당 물체의 크기를 측정하여 제공하게 된다.

[0035] 상기 물체 필터 모듈(130)은 각 프레임 이미지들에서 인식된 물체의 종류 및 크기가 사전 설정된 조건을 만족하지 못하는 경우, 해당 프레임 이미지들을 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링한다.

[0036] 도 6은 본 발명의 바람직한 실시예에 따른 동영상 내 배경 장소 추출 시스템에 있어서, 물체 인식 모듈 및 물체 필터 모듈의 동작을 설명하는 흐름도이다.

[0037] 도 6을 참조하면, 상기 물체 인식 모듈은 현재 샷의 프레임 이미지 셋에 대하여 각 프레임 이미지에 포함된 물체들의 종류 및 크기를 인식하고, 그 결과를 물체 필터 모듈로 제공하게 된다.

[0038] 다음, 상기 물체 필터 모듈은 물체에 대한 리스트 및 크기 기준값을 사전에 저장 및 관리하고, 프레임 이미지에 포함된 물체가 상기 물체 리스트 중의 하나에 해당하고, 물체의 크기가 기준값보다 큰 경우, 해당 프레임은 배경 장소 추출이 어려운 프레임으로 판단하여 상기 프레임 이미지 셋으로부터 제거하여 프레임 이미지 셋을 필터링하게 된다.

[0039] 다음, 상기 프레임 장소 인식 모듈(140)은 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대하여 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 검출한다. 상기 프레임 장소 인식 모듈은 당업계에서 널리 알여진 방법들 중 하나를 사용할 수 있으며, 본 발명에서는 이미지의 장소를 분류하는 딥 러닝 모델인 PlaceNet 모델을 일부 사용하였다. 상기 PlaceNet 모델은 MIT 에서 개발한 오픈 소스로서, 사전 설정된 365 개의 장소들에 대한 확률값을 제공하는 모듈이다.

[0040] 도 7은 프레임 장소 인식 모듈인 PlaceNet Model을 개념적으로 도시한 구성도이다. 도 7을 참조하면, 입력된 Image에 대하여 CNN 학습하여 각 배경 장소들에 대한 확률값(Probability)들을 출력한다.

[0041] 도 8은 본 발명의 바람직한 실시예에 따른 동영상 내 배경 장소 추출 시스템에 있어서, 프레임 장소 인식 모듈을 설명하기 위하여 도시한 구성도이다. 도 8을 참조하면, 본 발명에 따른 프레임 본 발명에서는 기존의 PlaceNet Model 에서 사용하는 365 개의 장소에 한국 미디어에 자주 나오는 장소들을 추가하여, 배경 장소들에 대한 데이터베이스를 수정하여 구성함으로써, 한국의 배경 장소들에 대한 신뢰도를 향상시켰다.

[0042] 상기 샷 장소 검출 모듈(150)은 상기 프레임 장소 인식 모듈에 의해 상기 필터링된 프레임 이미지 셋을 구성하는 프레임 이미지들의 각각에 대해 검출된 배경 장소 확률값들을 이용하여 해당 샷에 대한 배경 장소를 추출하여 제공한다.

[0043] 상기 샷 장소 검출 모듈의 일 실시형태는, 각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합하여 각각의 배경 장소들에 대한 확률합을 구하고, 가장 높은 확률합을 갖는 배경 장소를 해당 샷의 배경 장소로 결정할 수 있다.

[0044] 상기 샷 장소 검출 모듈의 다른 실시형태는, 각 프레임 이미지들에 대해 검출된 사전 설정된 다수 개의 배경 장소들에 대한 확률값들을 배경 장소별로 합한 후 각각의 배경 장소들에 대한 평균 확률값을 구하고 각각의 배경 장소들에 대한 중간 확률값을 구하고, 평균 확률값과 중간 확률값을 이용하여 해당 샷의 배경 장소를 결정할 수 있다.

[0045] 도 9는 본 발명의 바람직한 실시예에 따른 동영상 내 배경 장소 추출 시스템에 있어서, 샷 장소 검출 모듈에 의

해 얻는 각 프레임들에 대하여 구한 배경 장소별 확률값들과 평균 확률값, 중간 확률값 및 결과값들을 예시적으로 도시한 도표이다.

[0046] 도 9를 참조하면, 하나의 샷에 대해 필터링된 프레임들(#1, #2, #3, #4, #5)에 대하여 각각의 배경 장소들(교실, 병원, 사무실, 약국, 볼링장)에 대한 확률값을 알 수 있다. 각 배경 장소들에 대하여 구한 확률값들에 대하여, 평균값, 중간값을 구하고, 상기 평균값과 중간값을 합하여 구한 결과값이 가장 큰 배경 장소인 사무실을 최종적으로 해당 샷의 배경 장소로 결정하게 된다.

[0047] 이러한 결정 방법에 의하여, 어느 한 프레임에서 확률값 중 하나가 높게 나오더라도, 그것이 실제로는 잘못된 값일 가능성이 있음을 고려하여 결정할 수 있게 된다.

[0048] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0049] 1 : 배경 장소 추출 시스템

100 : 샷 단위 검출 모듈

110 : 프레임 추출 모듈

120 : 물체 인식 모듈

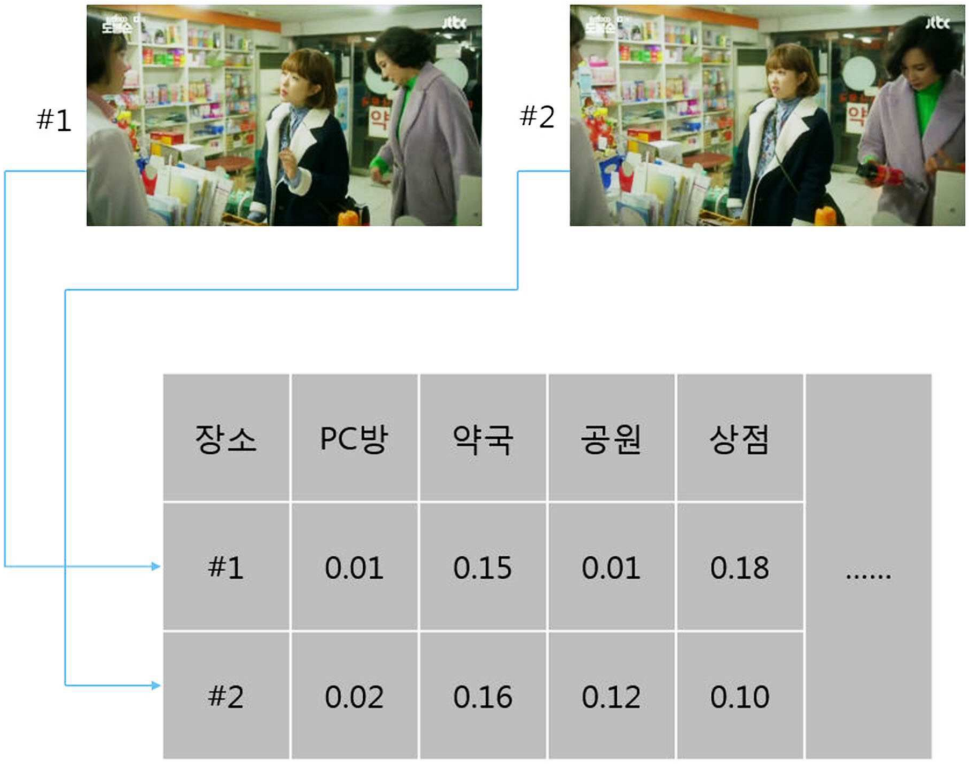
130 : 물체 필터 모듈

140 : 프레임 장소 인식 모듈

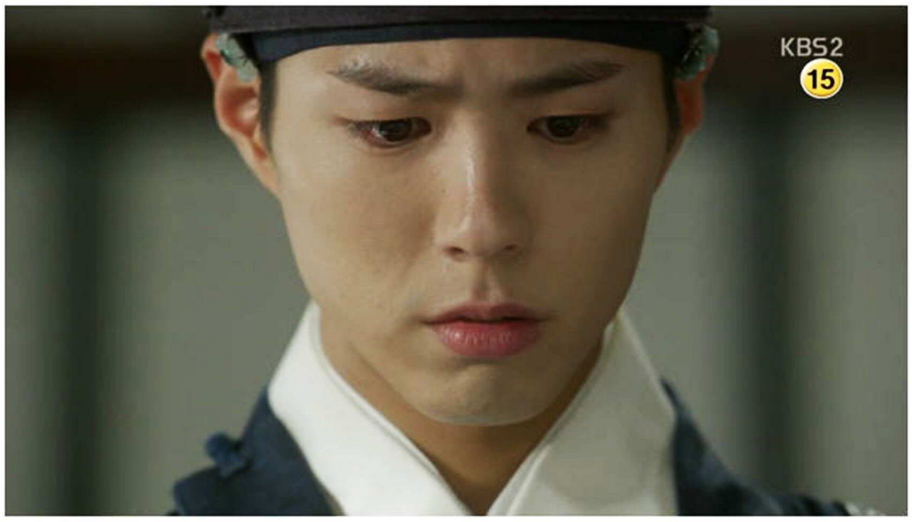
150 : 샷 장소 검출 모듈

도면

도면1



도면2

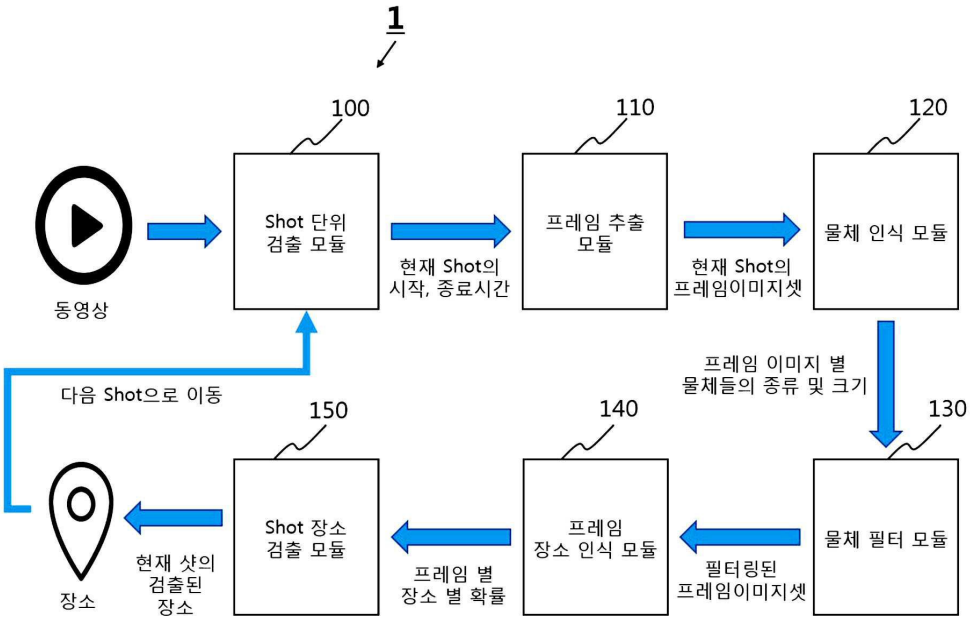


(a)

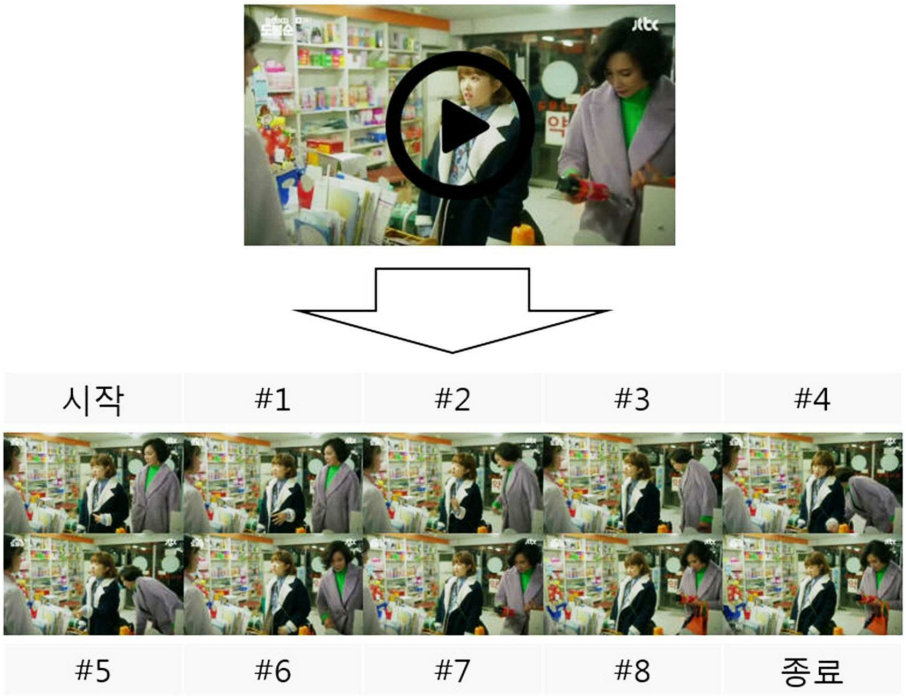


(b)

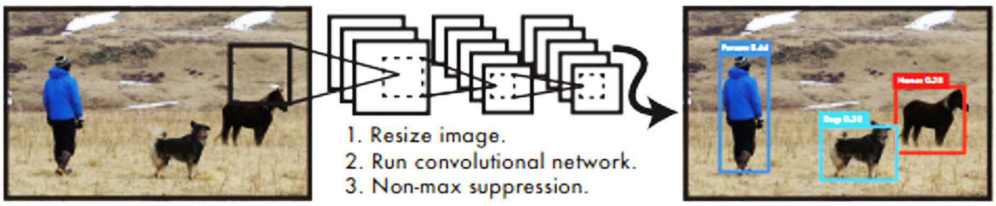
도면3


1
100
110
120
Shot 단위
검출 모듈
프레임 추출
모듈
물체 인식 모듈
동영상
현재 Shot의
시작, 종료시간
현재 Shot의
프레임이미지셋
프레임 이미지 별
물체들의 종류 및 크기
다음 Shot으로 이동
150
140
130
Shot 장소
검출 모듈
프레임
장소 인식 모듈
물체 필터 모듈
장소
현재 샷의
검출된
장소
프레임 별
장소 별 확률
필터링된
프레임이미지셋

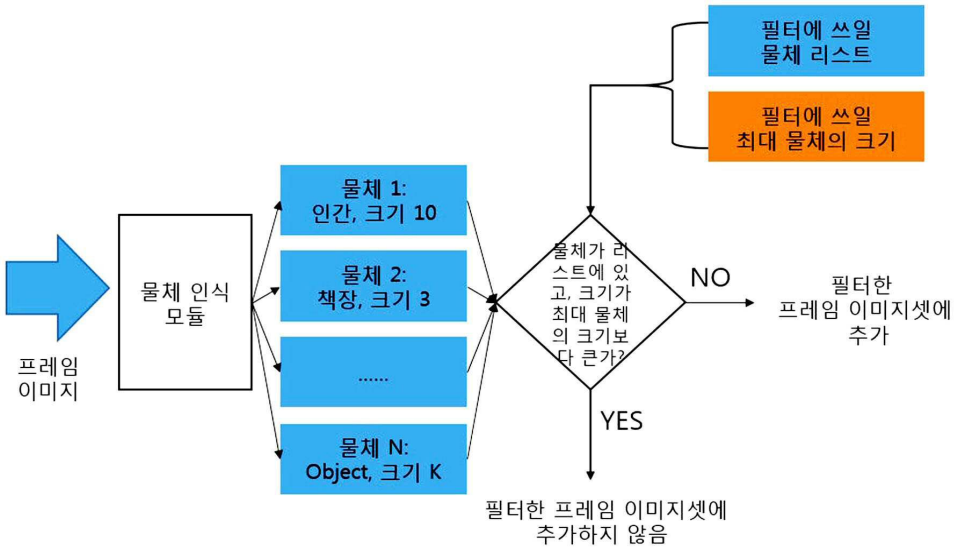
도면4


시작
#1
#2
#3
#4
#5
#6
#7
#8
종료

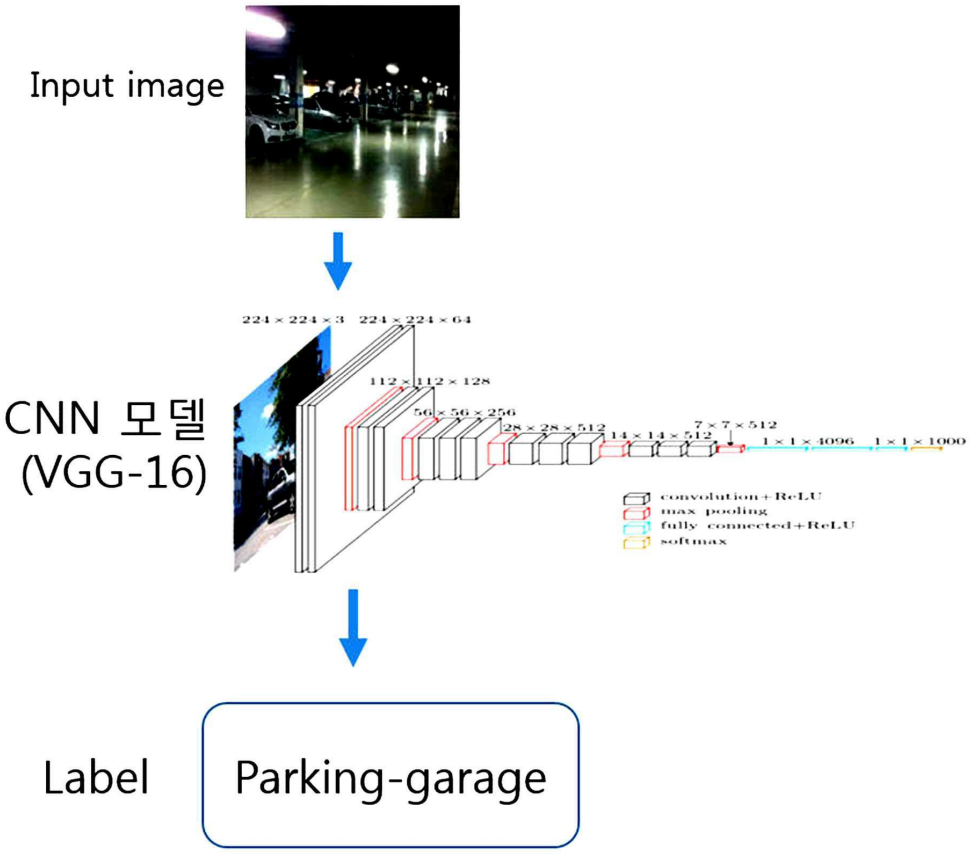
도면5


1. Resize image.
2. Run convolutional network.
3. Non-max suppression.

도면6


필터에 쓰일
물체 리스트
필터에 쓰일
최대 물체의 크기
프레임
이미지
물체 인식
모듈
물체 1:
인간, 크기 10
물체 2:
책장, 크기 3
.....
물체 N:
Object, 크기 K
물체가 리스트에 있고, 크기가 최대 물체의 크기보다 큰가?
NO
필터한
프레임 이미지셋에
추가
YES
필터한 프레임 이미지셋에
추가하지 않음

도면7


Input image
CNN 모델
(VGG-16)
224 × 224 × 3
224 × 224 × 64
112 × 112 × 128
56 × 56 × 256
28 × 28 × 512
14 × 14 × 512
7 × 7 × 512
1 × 1 × 4096
1 × 1 × 1000
convolution+ReLU
max pooling
fully connected+ReLU
softmax
Label
Parking-garage

도면8



도면9

	...	교실	병원	사무실	약국	볼링장	...
#1	...	0.10	0.50	0.10	0.30	0.05	...
#2	...	0.20	0.30	0.20	0.20	0.05	...
#3	...	0.10	0.05	0.25	0.00	0.30	...
#4	...	0.21	0.05	0.10	0.00	0.15	...
#5	...	0.15	0.02	0.20	0.01	0.30	...
평균	...	0.152	0.184	0.170	0.102	0.170	...
중간 값	...	0.150	0.050	0.200	0.010	0.150	...
결과 값	...	0.302	0.234	0.370	0.112	0.320	...