Vision API를 활용한 영상 속 랜드마크 검색 시스템

장지훈O, 문미경*
OO동서대학교 컴퓨터공학부,
*동서대학교 컴퓨터공학부
e-mail: xxwlgns@gmail.comO, mkmoon.dongseo.ac.kr*

Landmark Retrieval System in-Video Using Vision API

Jihoon JangO, Mikyeong Moon*
ODivision of Computer Engineering, Dongseo University, Korea,
*Division of Computer Engineering, Dongseo University, Korea

● 요 약 ●

인터넷의 발전으로 인터넷을 통해 미디어 콘텐츠를 제공하는 OTT 서비스의 이용자가 점점 증가하고 있다. 본 논문에서는 OTT 서비스 이용자의 검색 편의성을 위한 Vision API를 활용한 영상 속 랜드마크 검색 시스템을 제안한다. 이는 영상을 시청하는 사용자가 영상을 보다가 궁금한 장소가 있으면 따로 검색하지 않고, 바로 해당 장소의 정보를 얻도록 하는 것이다. 본 논문에서 개발한 영상 속 랜드마크 검색 시스템은 사용자가 원하는 영상을 검색 및 재생할 수 있고 재생 중 바로 화면을 캡처할 수 있다. 해당 캡처 이미지를 랜드마크를 기계 학습한 Vision API를 통해 어떤 랜드마크인지 파악하고 결과를 표시한다, 또한 동일 영상에서 다른 사용자가 검색한 이력이 있는 경우 이력을 표시하여주며, 검색한 랜드마크가 나온 다른 영상을 검색하여 보여준다.

키워드: OTT 서비스(OTT service), 기계 학습(Machine Learning), Vision API

I. Introduction

OTT(Over The Top) 서비스는 인터넷을 통해 방송 프로그램·영화·교육 등 각종 미디어 콘텐츠를 제공하는 서비스를 말한다 [1]. OTT 서비스 이용자들은 영상을 시청하다가 영상 속 장소가 궁금한 경우, 새로운 창을 연 뒤 검색해서 정보를 얻어야 하고 만약 영상에서 간단한 정보조차 알려주지 않았다면, 장소의 정보를 얻기가 힘들다. 본 논문에서는 OTT 서비스 이용자의 검색 편의성을 위한 영상 속 랜드마크 검색 시스템을 제안한다. 이 시스템을 통해 사용자들은 영상을 시청하다가 궁금한 장소가 있으면 즉시 정보를 얻음으로써 영상시청과 검색의 편의성이 향상될 것으로 기대된다.

II. Preliminaries

Vision API는 REST 및 RPC API를 통해 선행 학습된 강력한 머신러닝 모델을 제공한다. 이미지에 라벨을 할당하고 사전 정의된 수백만 개의 카테고리로 빠르게 분류할 수 있으며 객체와 얼굴을 감지하고 인쇄 및 필기 텍스트를 읽으며 유용한 메타 데이터를 이미지 카탈로그에 구축해 준다. Vision API의 기능은 다음과 같이 7가지가 있다.

- •LABEL_DETECTION : 이미지에 포함된 사물 인식
- •FACE_DETECTION : 얼굴 인식
- •TEXT_DETECTION : 문자 인식
- •LANDMARK_DETECTION : 지형지물 인식
- •LOGO_DETECTION : 회사 로고 인식
- •SAFE_SEARCH_DETECTION : 성인 컨텐츠 인식
- •IMAGE_PROPERTIES : 이미지의 주요 특성 인식

Vision API를 활용한 연구로는 Vision API를 활용한 스마트 헬스케어가 있다. 저울에 음식을 놓으면 음식의 무게 측정과 동시에 사진을 찍고 사진을 LABEL_DETECTION을 통해 무슨 음식인지를 유추한다. 그리고 유추한 음식의 이름을 미리 저장해놓은 음식의 칼로리와 영양성분 데이터와 비교하여 유추한 음식의 칼로리와 영양성분을 앱으로 전송해 사용자의 식단을 관리해준다 [2].

III. Landmark Retrieval System In Video Using Vision API

Fig 1은 본 논문에서 개발하는 시스템의 개념도이다.

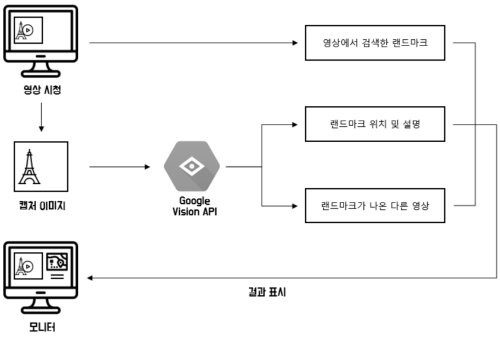

Fig. 1. Conceptual Diagram

사용자는 영상을 시청하다가 궁금한 장소가 있으면 캡처 버튼을 눌러 영상을 캡처한다. 캡처된 이미지는 Vision API를 통해 지형지물을 파악하고 해당 장소의 이름과 위치를 모니터에 표시한다, 또한 동일 영상에서 다른 사용자가 검색한 이력이 있는 경우 이력을 표시하여준다.

1. 동영상 재생 및 캡처

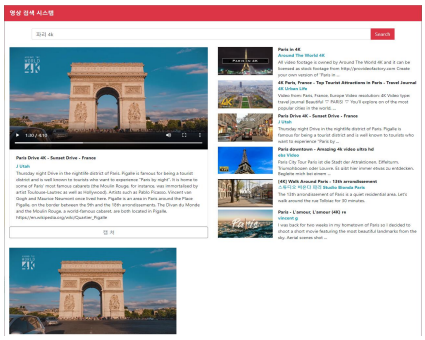

Fig. 2. Video Play and Capture

Fig 2는 결과화면이다. 본 논문에서 영상은 OTT 서비스 중 가장 많이 사용하는 YouTube 영상을 검색 및 재생한다. 영상 아래편의 캡처 버튼을 누르면 현재 장면을 Canvas에 그린다.

2. Vision API

본 논문에서는 Vision API의 LANDMARK DETECTION을 사용하여 지형지물을 파악한다. 영상시청 중에 캡처한 이미지를 Vision API를 통해 인식 시키면 결과값으로 위도, 경도, 랜드마크 이름, 일치율, 사진에서의 위치가 반환된다. Fig 3은 LANDMARK_DETECTION을 통해 Fig 2의 캡처 이미지를 파악한 결과이다.

```
Landmarks:
{
  locations: [ { latLng: [Object] } ],
  properties: [],
  mid: '/m/0zv_',
  locale: '',
  description: 'Arc de Triomphe',
  score: 0.88954758644104,
  confidence: 0,
  topicality: 0,
  boundingPoly: {
    vertices: [ [Object], [Object], [Object], [Object] ],
    normalizedVertices: []
  }
}
```

Fig. 3. LANDMARK_DETECTION Result

IV. Conclusions

본 논문에서는 OTT 서비스 이용자의 검색 편의성을 위한 영상 속 랜드마크 검색 시스템에 관하여 기술하였다. 본 시스템을 통해 사용들은 영상을 시청하다가 궁금한 장소가 있으면 즉시 정보를 얻음으로써 영상시청과 검색의 편의성이 향상될 것으로 기대된다.

ACKNOWLEDGEMENT

본 연구는 과학기술정통신부 및 정보 통신기획평가원의 SW중심 대학지원사업의 연구결과로 수행되었음(2019-0-01817)

REFERENCES

[1] Wikipedia, OTT
[2] Lim chong bin, Joe gun hee, Kim myung seok, “Smart Healthcare with Vision API”, The Korean Institute of Electrical Engineers, 207-208(2 pages), 2019