

Chapter
04비디오의 시각적 요소 그래프 구축
파이프라인

김주호_한국과학기술원 부교수

I. 결과물 개요

개발목표시기	2021. 12.	기술성숙도 (TRL)	개발 전	개발 후
			3	5
결과물 형태	SW-System	검증방법	자체 검증	
Keywords	비디오 그래프, 비디오 모델링, 그래프 모델링, 비디오 데이터셋 Video graph, Video modeling, Graph modeling, Video datasets			
외부기술요소	Open source 사용 (공개 비디오 데이터셋)	권리성	SW, 특허	

II. 기술의 개념 및 내용

1. 비디오 그래프 생성을 위한 데이터 수집

- 비디오 그래프 생성을 위한 데이터 수집
- 비디오 그래프 개발을 위해 주요 객체 1개 및 부객체 여러 개를 담은 비디오 수집
- 비디오 분야별 사용자의 필요에 맞는 비디오 그래프 구성 요소 구축을 위해 비디오 정보를

* 본 내용은 김주호 부교수(☎ 042-350-3570, juhokim@kaist.ac.kr)에게 문의하시기 바랍니다.

** 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

***정보통신기획평가원은 현재 개발 진행 및 완료 예정인 ICT R&D 성과 결과물을 과제 종료 이전에 공개하는 "ICT R&D 사업화를 위한 기술예고"를 2014년부터 실시하고 있는 바, 본 칼럼에서는 이를 통해 공개한 결과물의 기술이전, 사업화 등 기술 활용도 제고를 위해 매주 1~2건의 관련 기술을 소개함

담아내는 분류 체계(taxonomy) 설정

- ▶ 분류 체계에 따라 비디오 내 객체(인물 및 물체), 객체 사이의 관계를 레이블하는 웹기반 크라우드소싱 인터페이스 개발

2. 비디오 시각적 요소 그래프 생성 모델

- ▶ 분류 체계에 따라 비디오 내 등장인물 및 주요 물체 등의 객체의 시공간적 정보(위치 및 지속시간) 추출
- ▶ 비디오 내에서 탐지된 객체들 사이의 관계 정보 및 그 관계의 지속 시간 추출
- ▶ 비디오의 시각적 요소를 담은 객체를 그래프의 형태로 의미적으로 연결함으로써 비디오 시각적 요소 그래프 구축



[그림 1] 기술개념도

III. 국내외 기술 동향 및 경쟁력

1. 기술의 특성 및 성능

- 비디오 그래프 구축 파이프라인은 비디오 내의 중요 객체와 객체 간의 관계를 파악하여 이 정보를 자동으로 그래프로 표현
- 1차년도 예상 성능: 100개의 종류의 객체를 인식, 0.92의 정확도로 객체를 인식

2. 경쟁기술/대체기술 동향 및 현황

- 스탠포드 대학 연구진은¹⁾ 사건의 원형 단위가 행동-객체 상호작용이라는 연구로부터 행동을 시공간 장면 그래프로 분해하여 표현한 Action Genome을 제시함
 - Action Genome은 작업이 발생하는 동안 객체와 다른 객체의 관계 간의 변화를 포착하고, 이를 활용하여 장면 그래프를 시공간 특징 बैं크로 통합해서 Charades 데이터셋에서 더 나은 성능을 보임
- CMU 연구진은²⁾ 인간이 행동을 인식하는 형태로부터 시간적 형태 역학 모델링과 인간과 물체 간의 기능적 관계 모델링을 주장하며, 비디오에 대해 이 두 의미를 포함한 시공간 영역 그래프를 제안함
 - 그래프 노드는 긴 범위의 비디오에서 다른 프레임의 객체 영역 제안으로 정의되며, 노드의 연결은 객체 간의 유사성 관계와 주변 객체 간의 상호작용을 캡처하는 공간-시간 관계로 정의되는데, 이 방식으로 2018년 Charades 데이터셋에 대해 최고 성능을 보임
- 암리타 공과대학 연구진은³⁾ 비디오 내의 음성 정보를 사용하여 객체와 객체 간의 관계를

1) Ji, Jingwei, Ranjay Krishna, Li Fei-Fei, and Juan Carlos Niebles. "Action genome: Actions as compositions of spatio-temporal scene graphs," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10236-10247. 2020.

2) Wang X, Gupta A. "Videos as space-time region graphs," In Proceedings of the European conference on computer vision(ECCV) 2018.

3) Shanmukhaa GS, Nandita SK, Kiran MV. Construction of Knowledge Graphs for video lectures. In 2020 6th International Conference on Advanced Computing and Communication Systems(ICACCS) 2020 Mar 6(pp. 127-131). IEEE.

추출하여 비디오 강의 내용을 개념을 노드로 하는 knowledge graph로 구성하는 기술을 제안함

3. 우수성 및 차별성

경쟁기술	본 기술의 우수성/차별성
<ul style="list-style-type: none"> - Action Genome¹⁾: 주객체와 부객체 간의 관계를 크게 attention(주목), spatial(공간), contact(접촉)으로 분류해서 레이블을 확보함 - CMU 연구진²⁾: 비디오 내 연속적인 장면에 대해서 동일 물체에 대한 시간 정보를 추출함 	<ul style="list-style-type: none"> - 비디오 도메인(예; 요리, 스포츠, 일상 대화 및 행동)의 특성에 맞는 레이블 분류체계를 설정하여 각 도메인에 특화된 비디오 정보 데이터셋을 확보할 예정임 - 연속적인 장면뿐만이 아닌, 비디오 전반에 걸쳐서 등장하는 객체의 지속시간 및 동일 객체에 대한 정보를 추출할 예정임

4. 표준화 및 특허 동향

➤ 표준화 동향

- OASIS(Organization for the Advancement of Structured Information Standards)에서는 자연 언어 문서, 이메일, 발화, 이미지 및 비디오 등 비정형 데이터를 위한 데이터 표현 방식 및 분석을 위한 인터페이스 표준인 비구조 정보처리 아키텍처(UIMA - Unstructured Information Management Architecture)를 제시함⁴⁾
- MPEG-4 Part 11에서는 영상을 효율적으로 표현하기 위해 영상 내 물체 간의 관계를 그래프 형태로 표현하기 위한 표준 Scene Description and Application Engine을 제시함⁵⁾

➤ 관련 보유특허

No.	국가	출원·등록번호(출원·등록일)	상태	명칭
1	대한민국	출원: 1020180048667(2018.04.26) 등록: 1021704160000(2020.10.21)	등록 완료	컴퓨터와 클라우드 소싱을 이용한 비디오 사전 처리 방법

4) OASIS(2009). Unstructured Information Management Architecture(UIMA) Version 1.0.
<http://docs.oasis-open.org/uima/v1.0/uima-v1.0.html>

5) MPEG. MPEG-4 Scene Description and Application Engine.
<https://mpeg.chiariglione.org/standards/mpeg-4/scene-description-and-application-engine>

IV. 국내외 시장 동향 및 전망

1. 국내외 시장 동향 및 전망

- ▶ 비디오는 엔터테인먼트와 학습의 주요 매체로 활용되고 있으며, 이로 인해 YouTube, Twitch, 카카오TV, 네이버TV 등 다양한 비디오 플랫폼이 활성화되고 있는데, 이 중 가장 사용자가 많은 플랫폼인 유튜브의 월간 사용자는 20억 명에 달하고 매분 총 500시간의 비디오가 업로드됨
- ▶ 최근에는 Facebook이나 Instagram 같은 소셜 미디어 서비스도 콘텐츠의 형식으로서 비디오에 더 집중하고 있고, 비디오 기반인 TikTok 같은 새로운 소셜 미디어 서비스도 빠르게 인기를 얻고 있음
- ▶ IPTV 서비스 및 Netflix, 웨이브, 티빙, 왓챠, 시즌 등을 비롯한 OTT 서비스의 증가는 고객들이 기존의 TV 시청에서 온라인 비디오 시청으로 전환하고 있다는 것을 보여줌
- ▶ 온라인 동영상 시장 중, K-MOOC, 클래스101, 멀티캠퍼스 등 온라인 기반 교육 플랫폼의 경우 본 연구와 크게 관련이 있을 것으로 예상됨

2. 제품화 및 활용 분야

활용 분야(제품/서비스)	제품 및 활용 분야 세부내용
온라인 비디오 플랫폼	비디오 그래프 구축 파이프라인을 통해 비디오 플랫폼이 비디오의 시각적 요소를 파악하고 유해한 요소를 제거하는 한편 비디오 내의 시각적 요소와 이 요소들의 관계를 고려하는 비디오 추천 알고리즘을 고도화할 수 있을 것으로 예상함
OTT 서비스	파이프라인을 통해 각 비디오의 그래프를 비교하여 사용자에게 더 알맞는 추천을 제공하는 알고리즘을 개발할 수 있을 것으로 예상함
온라인 교육	파이프라인을 통해 학습자가 어려워하는 비디오 장면의 시각적 요소와 이 요소들의 관계를 인식하고 강의자는 이 정보를 사용하여 학습자들의 어려움에 대해 더 자세히 이해하고 대처할 수 있음

V. 기대효과

1. 기술도입으로 인한 경제적 효과

- ▶ 비디오 그래프 모델링을 통한 비디오 저작, 시청, 분석의 토대를 마련함으로써 2025년까지 6,687억 달러에 이를 것으로 전망되는 비디오 시장에서 비디오 생산 및 소비의 규모 확대 및 촉진을 가져올 것으로 예상됨
- ▶ 2019년 기준 3조 8,000억 원 규모로 추산되는 비디오 기반 온라인 교육 시장에 사업의 확장 및 촉진을 가져올 것으로 예상됨
- ▶ 2020년 기준 7,801억 원 규모로 추산되는 국내 OTT 시장에 문화 및 오락 사업 관련 확장 및 촉진을 가져올 것으로 예상됨
- ▶ 비디오 그래프 기반 저작을 통해 2023년까지 8조 원 규모로 성장할 것으로 예상되는 국내 1인 미디어 시장의 확장 및 촉진이 기대됨
- ▶ 현재 공개되어 있는 연구용 비디오 데이터셋에 깊이 있는 레이블을 더함으로써 비디오 관련 연구 및 산업에 발전을 가져올 것으로 기대됨

2. 기술사업화로 인한 파급효과

- ▶ 비디오 그래프 모델링을 통해 비디오 저작자는 시청자가 즐기는 콘텐츠에 대한 깊이 있는 이해를 바탕으로 더 나은 콘텐츠를 저작하여 시청률을 높일 수 있을 것으로 예상됨
- ▶ 비디오 플랫폼 측면에서 비디오 그래프 모델링을 통해 대규모 비디오 콘텐츠를 분석하여 플랫폼 사용자에게 더 적절한 추천을 제공함으로써 플랫폼의 사용률을 증가시킬 수 있을 것으로 예상됨
- ▶ 비디오 그래프의 풍부한 정보가 비디오 저작 소프트웨어에 통합될 수 있는 새로운 저작 인터랙션 연구와 개발이 가능할 것으로 예상됨