

AI기술이 영상 미디어 생태계에 미치는 영향과 방향

□ 함경준, 김선중, 이호재, 곽창욱 / 한국전자통신연구원

I. 서론

인공지능 기술의 발달로 영상 미디어의 생태계가 급변하고 있다. 영상 미디어 제작에서부터 소비되는 모든 환경에 AI(Artificial Intelligence)기술을 필수로 혼합현실과 같은 최신 ICT(Information & Communication Technology)기술이 본격적으로 활용되면서 기존의 미디어 형태나 소비 패턴이 급격히 변화하고 있다. 본 고에서는 현재 AI기술이 영상 미디어 생태계에 미치는 영향을 분석한다. 특히 영상 미디어가 제작되어 소비되는 각 과정에서 AI나 최신 ICT 기술에 대한 소개와 활용 사례를 통해 미디어 생태계의 앞으로의 발전방향에 대하여 모색해보도록 한다. 일반적인 영상 미디어 제작 단계는 다음과 같다. 1단계는 아이디어를 구성하고 시나리오를 완성하는 과정이며, 2단계는 감독, 스태프, 배우, 소품, 장소 등 촬영 전 모든 준비를 끝내는 단계이

다. 3단계는 실제적인 촬영이 이루어지며, 4단계는 촬영본에 대한 편집, 사운드 작업, 그래픽 효과 등 콘텐츠를 최종 완성하는 단계이다. 마지막으로 5단계는 미디어 배급 및 소비 단계로 최종 완성된 콘텐츠를 다양한 방법으로 배급하고 사용자에게 선보이는 단계이다. 본 고에서는 각 단계에서 활용되거나 영향을 미칠만한 기술을 소개하고 앞으로의 발전 방향에 대한 분석 결과를 결론에서 다루도록 한다.

II. AI 기반의 시나리오 작성 기술

인공지능이 선도하는 기술적 진보는 각 산업분야로 영향력이 파급되고 있으며, 영화 및 드라마 등 미디어 분야에서도 실질적인 적용으로 이어지고 있다. 인공지능 기술의 적용은 초기에는 영화 제작을 위한 보조(Assistance) 수단으로 사용되었으나, 최

근에는 인간 고유의 영역이라고 여겨져 왔던 스토리 창작(Storytelling)분야까지 그 범위가 확장되고 있다는 점을 주목할 필요가 있다.

초기에는 창작자의 생산성 향상을 위한 목적으로 기술이 적용되었다. 이스라엘 스타트업 업체인 ‘볼트(Vault)’에서는 2015년부터 시나리오만 보고도 영화의 흥행을 예측하는 서비스를 제공하고 있다. 이 업체는 약 30~40만개의 시나리오를 학습하여 흥행에 미치는 패턴들을 분석하였고, 새로운 시나리오의 흥행 여부에 대해 65~70%의 정확도를 보이고 있다. 캐나다에 위치한 그린라이트 에센셜(Greenlight Essentials)에서는 인공지능을 기반으로



〈그림 1〉 AI가 창작한 시나리오를 기반으로 제작된 영화 'Impossible things'

영화 제작자를 위한 솔루션을 개발하고 있다. 이 업체에서는 2016년에 세계 최초로 인공지능이 창작한 시놉시스를 기반으로 'Impossible things'라는 제목의 영화를 제작하였다. 인공지능을 통해 수 천편의 공포 영화 시놉시스를 학습하였고, 이를 기반으로 줄거리와 캐스팅의 구체적인 조합까지 생성해냈다.

위의 사례들은 AI 기술이 시놉시스 생성 등 단순히 창작자의 싱크탱크(Think tank) 역할로만 이용되었으나, 최근에는 직접 시나리오 창작에 진출함으로써 창작자의 고유 영역을 위협하고 있다. 또한 미국의 만화가 앤디 허드는 종영된 미국 드라마 '프렌즈' 시나리오를 신경망을 통해 학습하여, 후속편 내용이 담긴 시나리오 생성을 시도하기도 했다. 생성된 시나리오는 약간의 수정만 거치면 될 정도로 완성도가 높다는 평가를 받았다.

또한, 오스카 샤프 감독과 인공지능 연구자인 로스 굿윈은 '벤자민'이라는 인공지능 작가를 구현하여 시나리오를 창작하였다. AI 작가가 작성한 시나리오를 기반으로, 영국에서 열린 '48시간 내에 영화 제작' 경연에 'Sunspring'이라는 9분 내외의 단편

```
Checkpoint restored...
Sampling...

[OPENING CREDITS]

[Scene: Chandler and Joey's, Joey is waiting for the door. Ross is looking for Joey and Chandler stuck on the couch and screaming.]

Chandler: Well, I proposed to my shoe...

Joey: (laughs) This is his father... (Cave children.) Hey Pheebs?

Monica: I hate men! I hate men!

Ross: What are you gonna do?

(All the dinner enters.)

Monica: Happy Gandolf.

(Pause for a beautiful women walk into his apartment and she keeps glaring at Joey's jacket.)

Monica: Okay, I'm going to Minsk.

Rachel: Yeah, sure.
```

〈그림 2〉 인공지능이 생성한 프렌즈 후속 시나리오

SF 영화를 출품하였다. 이 영화는 Top-10에 포함되는 성과를 보였으며, SF 작품으로써는 뛰어나지만 스토리가 연결되는 부분에서 개연성이 부족하다는 평가를 받았다.

위의 사례들을 통해 시나리오 창작을 위한 여러 기술적 시도들이 다양한 목적을 위해 이루어지고 있다는 점을 확인할 수 있었다. 이러한 기술적 시도들이 도전에 그치지 않고 실제적인 효과를 보였다는 점에서 의미가 있으며, 단순히 비서 역할에 국한되지 않고 자체적인 시나리오의 창작이 가능하다는 점에서 앞으로의 행보를 주목할 필요가 있다.

Ⅲ. 촬영 준비 단계에서의 AI 기술 활용

앞서 언급한대로 인공지능 기술의 발전으로 인해 영상 미디어 제작 과정에 다양한 변화가 예상된다. 촬영 준비, 스토리보드 및 장면 혹은 샷 리스트 생성, 촬영 장소 섭외, 촬영 스케줄 조율, 촬영 비용 예측 등 여러 범주의 과정에 도움을 줄 수 있을 것으로 예상된다.

영상 미디어 사전 제작에 인공지능 기술이 적용된 사례를 보면 미국 End Cue의 Agile Producer 플랫폼이 있다. 이 플랫폼은 자체 개발한 인공지능 기반 자연어 처리 기술을 통해 각본을 분석하여 촬영에 필요한 캐릭터, 소품, 오디오/비디오 효과 등 다양한 요소들을 추출한다. 또한, 배우 스케줄, 촬영 장소와



〈그림 3〉 Agile Producer 각본 분석 화면



〈그림 4〉 Agile Producer 예산 배분 화면

장비 가용성, 날씨, 촬영 예산 제한을 고려하여 최적의 촬영 스케줄을 자동으로 관리해 준다. 따라서, 촬영 예산 배분 및 소요 비용 예측도 가능하다.

IBM Watson은 배우들의 얼굴들과 각 장면마다 배경의 영상 분석을 통해 최적의 배우를 나열하여 감독이 배우를 선정하는데 도움을 줄 수 있고, 영상의 스토리를 이해해서 영상의 촬영 구도를 도와주는 AI 조수를 개발 중인 것으로 알려진다.

이러한 플랫폼이 개발되고 기술 고도화를 통해 사전 제작 단계에서 다양한 형태의 인공지능 기술이 실제로 적용되었으며 앞으로 상당부분의 사전 제작 단계가 간소화되고 인공지능 기술로 대체될 것으로 예상된다.

Ⅳ. 촬영 단계에서의 AI 기술 활용

영상 촬영 과정에서도 인공지능 기술이 적극적인



〈그림 5〉 구글 AI 카메라, Clips

로 활용되고 있다. 구글에선 AI 카메라 ‘Clips’을 개발하여 제품으로 판매 중이다. 해당 카메라는 사용자의 얼굴을 인식하고 친하게 지내는 사람들을 학습하고 표정 변화와 색다른 움직임을 포착하여 사용자의 의미있는 순간을 촬영한다.

Amazon AWS에서는 DeepLens라는 딥러닝 기반의 프로그래밍 가능한 비디오 카메라를 개발하여 원하는 객체를 인식하고 행동을 감지하여 촬영하는 인공지능 카메라로써 올해 4월에 해당 제품을 출시할 예정이다.



〈그림 6〉 Amazon DeepLens

또한 인공지능 기술 기반의 카메라 드론이 개발되어 사용자로 등록된 사람을 추적하여 신체를 지



〈그림 7〉 Hover Camera Passport, 자체비행 카메라 드론

속적으로 촬영하거나 얼굴만 촬영이 가능하다. 카메라 감독이 직접 촬영하지 않아도 카메라 자체촬영이 가능하다. 이러한 인공지능 기술이 적용된 카메라를 이용한 새로운 촬영 기법이 사용되면 다채롭고 흥미로운 영상 미디어가 제작될 것으로 보인다.

애니메이션 연출 기법에도 인공지능 기술이 사용되고 있다. Disney에서는 사람의 말하는 모습을 학습하여 만화 캐릭터 혹은 다른 배우의 모습으로 표현할 수 있는 기술을 개발하였다. 해당 기술을 소개하고 있는 논문에 의하면 입모양, 눈썹, 이마의 움직임과 같이 사람이 말할 때 변화하는 모습에서 특징을 추출하고 학습시켜서 원하는 캐릭터가 해당 특징을 반영하여 표현하도록 하는 기술이라고 소개하였다. 현재는 4가지의 캐릭터만 표현이 가능하지만 이후엔 다른 사람의 얼굴을 통해 자연스럽게 말하는 것처럼 표현이 가능해질 전망이다.

2016년도에는 최신의 인공지능 기술을 조합하여 단편 영화 “Eclipse”가 제작된 사례가 있다. IBM Watson의 감정 분석기를 통해 가사에 담긴 감정을 분석하여 영상의 분위기를 표현하는 방법을 학습시켰으며, MUSE EE의 뇌 분석 도구를 이용해 해당 가수의 실제 감정을 분석하여 이후에 영상을 어떻게 표



〈그림 8〉 인공지능 영화 “Eclipse”의 스틸컷

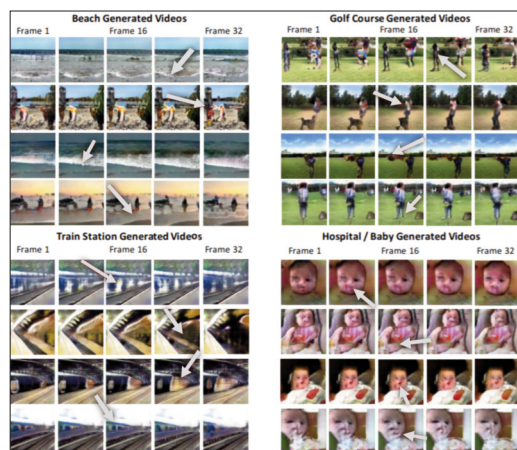
현할 지에 대한 학습을 수행하였다. 촬영의 대상도 스스로 결정을 하였는데, 이를 위해 Microsoft AI Rinna 챗봇에게 질문하여 대답을 바탕으로 촬영을 진행했다. 이후 촬영은 AI 카메라 드론을 이용해 촬영하고 감정 분석에 따른 분위기에 맞게 편집까지 수행했다. 이처럼, 인공지능 기술 고도화에 따라 새로운 형태의 영상 미디어가 제작될 것으로 예상된다.

V. AI 기반의 영상 편집 기술

인공지능 기술은 영상 편집 기술에도 적용되고 있다. 기술의 방향이 단편적으로 자르고 붙이는 수준을 넘어서, 최근에는 인공지능이 텍스트, 이미지, 영상 등 다양한 입력에 대해 포함하고 있는 맥락(Context)을 이해하고 영상 생성, 요약하는 연구가 이루어지고 있다.

미국 위비츠(Wibbitz)는 텍스트를 영상으로 변환

하는 기술을 개발 중이다. 이 회사는 영상 뉴스 제 공을 위해 뉴스 텍스트와 이미지를 분석하고 중요 부분을 추출하여 영상으로 변환한다. 한국전자통신 연구원은 한단계 더 나아가 사용자가 입력한 시나리오를 영상으로 손쉽게 제작할 수 있는 기술을 개발 중이다. 해당 기술은 사용자 시나리오에 가장 적



〈그림 10〉 입력 이미지를 기반으로 전후 영상 생성

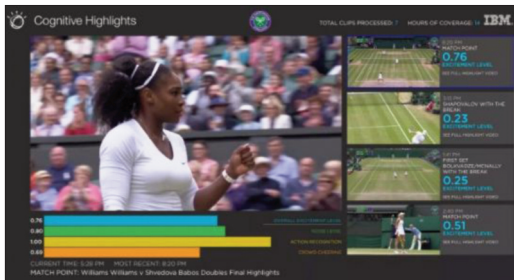


〈그림 9〉 ETRI 사용자 시나리오 기반 영상 창작 플랫폼

합한 영상을 검색하여 스토리라인을 구성하여 영상을 제작하는 기능을 제공한다.

MIT에서는 2017년 이미지를 입력으로 영상을 생성하는 연구를 수행하였다. 이 연구에 따르면, GAN(Generative Adversarial Networks) 신경망을 기반으로 주어진 입력 이미지에 대해 움직임을 반영한 짧은 영상을 생성한다. 입력된 이미지를 바탕으로 전후 영상 예측을 다루었다는 데에 의미가 있다고 볼 수 있다.

미국 IBM의 인공지능 시스템인 왓슨은 2016년

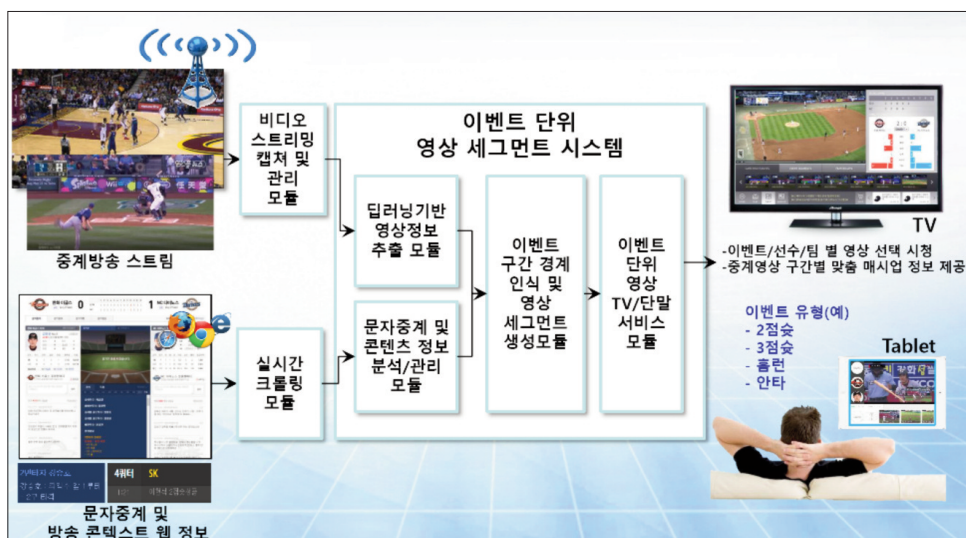


〈그림 11〉 왓슨이 생성한 테니스 하이라이트 장면

‘모건(Morgan)’이라는 SF 영화의 트레일러를 만들었다. 왓슨은 영화에서 극대화할 요소들을 자가 학습하여, 인상적인 장면들을 스스로 추출했다. 또한, 왓슨은 2017년 Wimbledon 테니스 하이라이트 영상 생성을 시도했다. 서브 에이스, 관중 환호 장면 등 영상에 등장한 요소들을 이해하고 하이라이트를 생성했다. 중요한 점은 영상뿐만 아니라, 음악, 음향 효과 등 멀티모달 측면을 반영한 예고편을 생성했다는 것이다.

한국전자통신연구원은 농구, 야구와 같은 스포츠 중계 영상을 딥러닝을 이용하여 분석하고 문자중계 데이터를 활용하여 샷, 홈런과 같은 이벤트 단위로 장면을 자동으로 편집하는 시스템을 개발하였다. 해당 시스템을 이용하면 사람이 수작업으로 편집한 영상 품질과 비슷한 수준의 하이라이트 영상을 생성하여 사용자는 자기가 좋아하는 선수나 이벤트 유형을 선택하여 시청할 수 있게 된다.

최근에는 음성이나 영상을 합성하는 연구들도 시



〈그림 12〉 ETRI 스포츠 하이라이트 자동생성 시스템

도되고 있다. 2017년 국내 데브시스터스는 음성합성 기술을 공개했다. 딥러닝을 통해 목소리 특징을 학습하고, 입력된 텍스트에 대해 학습된 음성 특징이 적용된 음성을 생성하였다.

이처럼, 최근 텍스트, 이미지 등 다양한 입력을 대상으로 영상 생성을 위한 연구들이 이루어지고 있다. 뿐만 아니라 인공지능 기술을 적용하여 영상의 의미를 이해하고 이를 기반으로 재편집하는 방법들이 산업적으로 이용되고 있다. 지금까지는 작은 단위의 입력만을 다루고 있으나, 단위가 점차 확장될 것으로 예상되며, 향후에는 입력된 시나리오를 기반으로 영상 제작 콘티 생성, 가이드 영상 제작 등 다양한 측면에서 적용될 것으로 예상된다.

VI. 배급 및 소비 단계에서의 ICT 활용

불과 10년 전만 해도 내 입맛에 맞는 콘텐츠를 찾아서 감상할 수 있는 환경이 아니었다. 영상 콘텐츠의 종류가 제한적이고 소비하는 단말의 형태도 TV 위주였기 때문에 대중이 공감할 수 있는 보편적인 소재를 담고 있는 콘텐츠를 시청할 수 밖에 없었다. 하지만 스트리밍 서비스가 활성화되고 스마트폰과 같은 IT 기기가 보급됨에 따라 이러한 환경은 급변하였다. 개인이 제작한 미디어까지 활발하게 보급이 되면서 우리가 접할 수 있는 콘텐츠의 양은 급격하게 방대해지고 있고, 콘텐츠를 소비하는 단말도 TV뿐만 아니라 모바일, HMD 등으로 다양해져서 소비 형태 또한 바뀌고 있다. CJ E&M에서 발표한 리포트에 따르면 연령대가 낮을수록 동영상, 소셜미디어 등 다양한 매체를 활용하며 TV보다는 모바일의 이용 비중이 높은 것으로 나타났다. 하지만 전 연령대에 걸

쳐서 모바일의 매일 이용 비중이 높은 것으로 알려졌다. 이러한 미디어 소비 형태의 변화는 영상 미디어의 배급 및 상영 환경에도 상당한 영향을 주고 있다.

이번 CES 2018에서는 다양한 멀티미디어 장비가 등장을 하였는데, NVIDIA는 초대형 스크린에서 PC 게이밍 경험을 위한 빅 포맷 게이밍 디스플레이를 선보였다. Pimax사는 8K 해상도까지 지원하는 HMD를 이용하여 8K VR을 시연하였는데, 기존의 HMD에 대비하여 2배의 해상도를 제공하게 된다. 삼성전자는 S-Ray 제품을 선보였는데, 사용자가 원하는 방향으로만 소리를 들려주는 포터블 지향성 스피커이다. 이러한 첨단 디스플레이 및 오디오 장비를 통해 높은 수준의 홈씨어터 환경을 구축할 수 있게 되었으며 영상 미디어를 더욱 생생하게 접할 수 있는 환경이 조성되고 있다.



〈그림 13〉 Pimax사의 세계 최초 8K VR Headset

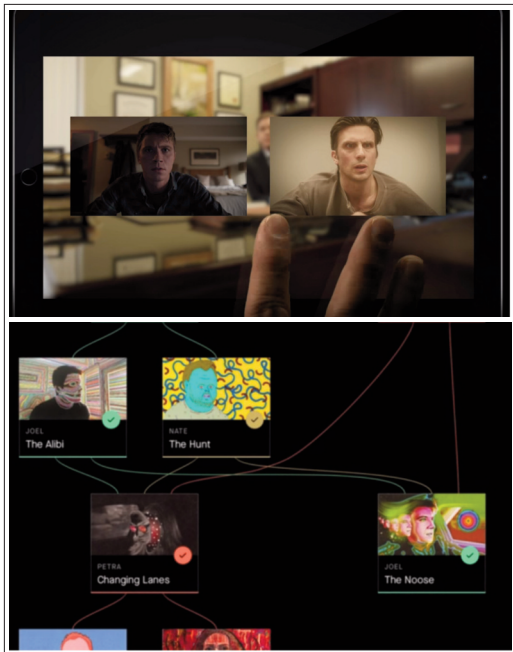


〈그림 14〉 ETRI 인터랙티브 영상 플레이어 시연

한국전자통신연구원에서는 스마트폰, 립모션 장

치, VR장치 등을 통해 미디어의 인터랙티브 요소를 자유롭게 편집할 수 있는 시스템을 개발하여 사용자가 영상 미디어를 시청하다가 인터랙티브리티를 경험할 수 있는 연구도 이루어지고 있다.

멀티미디어 기기의 발전은 미디어의 제작 단계에도 영향을 미칠 것으로 예상된다. HBO는 얼마전 새로운 미니 시리즈 'Mosaic'의 방영을 시작하였는데, 시청자가 어떠한 관점에서 이야기를 전개시킬 것인지를 선택할 수 있는 선택권을 모바일 앱을 통해 제공하여 자신만의 시나리오 전개를 가능하도록 하였다. 이처럼 소비 형태의 변화와 함께 전 세계 콘텐츠 시장에서는 인터랙티브 기술 개발이 활발히 이루어지고 있으며, 결과물로 제작된 인터랙티브 미디어는 앞으로 더욱 다양한 형태로 제공될 것으로 전망된다.



〈그림 15〉 HBO 신규 미니 시리즈 'Mosaic' 트레일러 영상 스냅샷

미디어의 상영단계에서도 AI 기술을 활용하여 시청자의 반응을 분석하여 콘텐츠의 선호도를 파악하고 향후 콘텐츠 제작에 활용하려는 시도가 이루어지고 있다. Disney Research에서 개발한 'FVAE (Factorized Variational Autoencoders)'는 얼굴을 분석하여 감정을 판별하는 기술로 카메라를 통해 수집된 영화 관객의 표정을 통해 해당 콘텐츠에 대한 선호도를 빠르게 파악할 수 있게 된다.

더 나아가 인공지능을 이용한 캐릭터 생성 및 영상 편집 기술이 더욱 고도화 된다면 내가 좋아하는 배우나 주변 인물을 등장시켜서 나만의 콘텐츠를 생성하는 인터랙티브 미디어가 등장할 것으로 예상된다. 한편, 혼합현실(Mixed Reality) 단말이 보편화되고 스마트홈 환경이 구축된다면 인공지능 기술을 활용한 다양한 형태의 미디어 소비 경험을 제공할 수 있다. 인공지능이 사용자의 인터랙티브 디바이스와 스크린 장치 등을 인지하고 사용자가 영상을 시청하고 있을때에 사용자의 홈 씨어터 환경에 적합한 인터랙션을 제공함으로써 나만의 인터랙티브 미디어 경험이 가능해질 전망이다. 또한 '포켓몬고'처럼 위치기반으로 사용자가 특정 랜드마크에 가면 스토리 전개 방식을 취하는 인터랙티브 미디어 콘텐츠의 출현도 기대된다. 즉 인공지능 기술을 활용한 콘텐츠 제작이 활성화되어 기존의 수동적 참여 유도 인터랙티브리티를 넘어 능동적 참여 유도 인터랙티브리티 콘텐츠의 보급이 활발해 질 것으로 전망되며 이를 위해서는 미디어 제작의 전 단계에 걸쳐 많은 변화가 예상된다.

Ⅶ. 결 론

앞에서 살펴본 것처럼 4차 산업혁명의 핵심 요소인 인공지능 기술은 산업계 전반에 활용되고 있으며, 영상 미디어 생태계에도 이미 많은 분야에서 활용되고 있고 상당한 변화를 가져왔다. 본 고에서는 다루진 않았지만 실감형 미디어 기술의 발전도 이러한 변화를 더욱 가속시킬 것으로 전망된다.

이미 방대하게 제작되어 있는 영상 미디어를 AI 기술을 활용하여 분절(fragmentation), 조합, 편집을 통해 새로운 미디어로 자동 생성하려는 시도도 이루어지고 있다. 영상 미디어를 기획하고 생성하는 작업이 그동안은 전문가 집단에서 주로 이루어졌지만, AI 기술을 활용한다면 누구나 손쉽게 영상 콘텐츠를 제작하고 공유할 수 있는 환경이 조성되

고 있다.

한편, 본 연구에서 살펴보았듯이, AI의 기술은 미디어 제작 단가를 낮추고, 제작 기간을 획기적으로 축소시킬 것이다. 현재의 미디어 소비 시장은 인공지능기반 추천 기술을 활용한 정확한 큐레이션 서비스가 핵심 경쟁력으로 꼽혔다. 하지만 현재의 AI 기술이 더욱 고도화 된다면 콘텐츠 추천 서비스를 넘어서 완전한 개인 맞춤형 미디어가 생성되어 나만의 콘텐츠를 감상할 수 있는 날이 머지 않았다. 즉 인공지능 기술이 미디어 비서의 역할을 그동안 수행하였다면 이제는 미디어 친구의 역할을 할 것으로 전망된다. 다양한 사용자 경험이 가능한 지능형 미디어의 본래 모습을 이제는 실제로 경험할 수 있는 시대를 맞이한 것이다.

참 고 문 헌

- [1] STRABASE 보고서, "HBO의 신규 미니 시리즈 'Mosaic', 양방향 시청 앱 제공 개시... TV 콘텐츠의 진화 방향 시사", 2017.11.
- [2] 박영수, 정원식, 허남호, 인공지능을 활용한 미디어 제작의 오늘과 내일, 주간기술동향, 2017.01
- [3] 이현정, 인터랙티브 미디어에 적용되는 인터랙션 의미의 범주화, 한국콘텐츠학회논문지, 2015.08
- [4] 미디어 이슈&트렌드, 방송산업의 인공지능(AI) 활용 사례 및 전망, 한국방송통신전파진흥원, 2017.10
- [5] Agile Producer, End cue, <https://www.endcue.com/artificial>
- [6] As Hollywood taps into A.I., what will you build with IBM Watson?, <https://www.ibm.com/blogs/watson/2017/04/hollywood-taps-build-ibm-watson/>
- [7] Google Clips, <https://support.google.com/googleclips/answer/7545440?hl=en>
- [8] Amazon AWS DeepLens, <https://aws.amazon.com/ko/deeplens/>
- [9] Hover Camera Passport, <https://gethover.com/?d=pc&c=kr>
- [10] 'Cut!' the AI director, <http://www.bbc.com/news/technology-36608933>
- [11] S. Taylor, T. Kim, Y. Yue and et al., A Deep Learning Approach for Generalized Speech Animation, ACM Transactions on Graphics, Vol. 36, No. 4, Article 93. July 2017.
- [12] Vondrick, C., Pirsiavash, H., & Torralba, A. (2016). Gen-erating videos with scene dynamics. In Advances In Neural Information Processing Systems (pp. 613-621).

필자소개



함 경 준

- 2005년 : 숭실대학교 산업정보시스템공학과 학사
- 2014년 : 한국과학기술원 산업및시스템공학과 박사
- 2015년 ~ 현재 : 한국전자통신연구원 스마트미디어연구그룹 선임연구원
- 주관심분야 : 정보검색, 기계학습, 온톨로지



김 선 중

- 1989년 : 충남대학교 계산통계학과 학사
- 2000년 : 충남대학교 컴퓨터과학과 석사
- 1989년 ~ 현재 : 한국전자통신연구원 스마트미디어연구그룹 그룹장
- 주관심분야 : 융합서비스제어, 지능형미디어, 콘텐츠 마이닝



이 호 재

- 2010년 : Iowa State University, Electrical Engineering 학사
- 2014년 : 과학기술연합대학원대학교, 이동통신 및 디지털 방송공학 석사
- 2014년 ~ 현재 : 한국전자통신연구원 스마트미디어연구그룹 연구원
- 주관심분야 : 컴퓨터 비전, 기계학습, 영상 부호화



곽 창 욱

- 2013년 : 동국대학교 컴퓨터멀티미디어학부 학사
- 2015년 : 경북대학교 컴퓨터학부 석사
- 2016년 ~ 현재 : 한국전자통신연구원 스마트미디어연구그룹 연구원
- 주관심분야 : 텍스트 마이닝, 기계학습