

# 3D 딥러닝 기술에 대한 표준화 동향

이승욱 한국전자통신연구원 책임연구원

## 1. 머리말

컴퓨팅 성능의 향상과 알고리즘 및 데이터의 발전에 따라 인공지능 기술이 다양한 응용 분야에서 활용되고 있다. 영상, 오디오 등의 멀티미디어 데이터뿐만 아니라 고객의 구매 성향, 가정의 에너지 소비 패턴 등의 다양한 데이터에 대해서도 인공지능 기술이 적용되고 있다. 인공지능은 자연어 처리, 음성인식, 시각인지 등의 기술을 이용하여 안전, 의료, 국방, 금융, 복지 등 사회의 전 분야에 대해서 적용되는 추세이다.

인공지능은 1956년 다트머스 회의에서 처음으로 용어가 등장한 이후 많은 기대와 실패가 반복되었고, 최근 5~6년 사이에 급격히 기술이 발전하였다. 특히 2016년 구글 딥마인드가 개발한 인공지능 바둑 프로그램인 알파고가 인간 프로기사에게 승리한 것을 계기로 국내에서 인공지능에 관한 관심이 급증하였다. 최근 컴퓨터 비전 등의 영역에서는 기존의 수작업 기반 특징점(Hand Crafted Feature)에 비해 인공지능이 집단지성 기반의 고성능을 제공하기에 많은 연구가 활발히 진행되고 있다.

그러나 인공지능 기술의 역사에 비해 인공지능 기술의 표준화에 대한 논의는 최근에 시작되었다. 기술의 발전 방향에서 보면 음성 및 영상 관련 기술이 먼저 개발된 후 3D 콘텐츠 분야에 대한 인공지능 기술이 발전하였으나, 표준화 측면에서 보면 데이터 형태나 응용 분야에 대한 영역의 구분 없이 동시에 표준화가 시작되고 있다.

본고에서는 ISO/IEC JTC 1/SC 42와 3D 콘텐츠 생성 및 서비스 프로토콜을 정의하는 OMASpec Works, 그리고 추론을 위한 학습모델을 압축하는 ISO/IEC JTC 1/SC 29/WG 11에서의 표준화 동향을 3D 콘텐츠의 관점에서 살펴본다. 그리고 국내 표준화 단체인 TTA의 디지털콘텐츠 프로젝트그룹(PG 610)에서 진행된 표준화를 3D 콘텐츠 위주로 살펴보고자 한다.

## 2. 3D 딥러닝 표준화 영역

표준화에서 중요한 부분은 '무엇을 표준화하는가?'이다. 일반적으로 표준 문서에서 'Scope'이라는 제목으로 작성된다. 인공지능의 표준화 영역은 요구사항, 파일포맷, 참조구현, 프레임워크 등 아주 다양하다. 본 장에서는 3D 딥러닝에 관련된 분야 중 파일포맷과 참조구현을 중점적으로 살펴볼 것이다.

먼저 파일포맷은 학습데이터를 위한 파일포맷과 학습모델을 위한 파일포맷으로 나뉜다. 특히

3D 콘텐츠의 경우 학습데이터 확보가 어렵고, 학습데이터의 형태가 다양하여(3D 데이터 표현 형식이 다양함에 따라 학습데이터의 형태도 다양함) 학습데이터를 재사용하기 위한 표준포맷은 아주 중요하다. 학습모델의 경우는 모델 압축이 중요한 이슈이다. 모델의 크기가 크면 추론시간이 오래 걸리고 모바일 단말 등에서 사용하기 어려워 학습모델 자체를 압축하는 것이 중요하다. 이는 ISO/IEC JTC 1/SC 29/WG 11(MPEG) 및 TTA PG 610에서 진행하고 있다.

다른 분야로는 서비스를 위한 참조구현이 있다. 인공지능의 일반적인 서비스 모델을 생각하면, 추론모델을 원격지에 설치하고 사용자 단말에서 추론할 입력값을 전송하여 결과를 받는 형태로 구현된다. 이때 통신 프로토콜, 참조구현 방법 등을 표준화할 수 있다. 이는 OMASpecWorks와 TTA에서 표준화가 진행된다.

## 2.1 3D 딥러닝을 위한 데이터 표현 방법

3D 딥러닝 기술이 다른 멀티미디어 데이터의 인공지능 기술과 가장 구별되는 특징은 데이터 포맷의 차이다[1]. 이는 현재 가장 많이 사용하는 CNN(합성곱 기반 학습네트워크)에 대한 적합 여부 때문이다. 일반적으로 음성, 영상 등은 합성곱 기반의 학습네트워크에 적용하는 격자구조의 표현이 가능하다. 3D 콘텐츠의 영역에서 보면 이는 유클리디언 데이터인지, 년 유클리디언 데이터인지로 구별된다. 유클리디언 기반의 3D 데이터 표현은 볼륨, RGB-D 등의 형태로 표현되어 기존의 합성곱 기반의 네트워크에 적합하다. 그러나 년 유클리디언 기반의 3D 데이터는 메시, 그래프 등의 형태로 구성되어 기존의 합성곱 네트워크에 적합하지 않다. 이처럼 데이터의 표현 방식에 따라 표준의 형태가 달라질 수 있다. 예를 들어 MPEG에서 진행하는 학습네트워크를 압축하는 표준인 NNR(Compression of Neural Networks for Multimedia Contents Description and Analysis)의 경우, 3D 데이터의 표현 방법에 따라 네트워크의 구조가 변경될 수 있어 압축 시에 고려할 사항이 추가된다. 그리고 TTA의 디지털콘텐츠표준화 프로젝트에서 진행한 학습데이터 파일포맷 표준의 경우에는 3D 학습데이터 표현 방법이 더욱 중요하게 적용된다. 본고에서는 3D 딥러닝 표준에 사용하는 학습데이터의 다양한 표현 형태, 즉 일종의 산업계 표준인 파일 구조에 대해 먼저 논의한다.

3D 데이터는 유클리디언 기반과 년 유클리디언 기반의 데이터로 나눌 수 있다. 유클리디언 구조는 정방형의 격자구조와 같이 정형화된 2차원 표현 방식으로, 예를 들어 매트릭스 형태로 표현될 수 있고, 년 유클리디언의 경우는 정형화되지 않은 방식으로 표현된다. 프리미티브 구조는 중의적인 방식으로 표현될 수 있는데, 정해진 구조(예를 들어 컨트롤 포인트의 개수 고정 등)의 방식이면 유클리디언 방식의 표현도 가능하다. 이 경우 매트릭스 형태에 정해진 컨트롤 포인트의 개수와 프리미티브 함수를 정의하는 가중치값을 매트릭스에 삽입할 수 있다. 그렇지 않을 경우는 년 유클리디언 표현으로 분류한다.

- 프리미티브: 수학적 선, 면 등의 조합으로 데이터를 생성하는 방법으로 선, 면 등을 이루는 함수 정의. 각 프리미티브 데이터를 수학적으로 변형하는 컨트롤 포인트로 제어하여 형상 변화
- 프로젝션: 3D 콘텐츠 데이터를 특별한 좌표계(구좌표계 혹은 실린더 좌표계 등)로 맵핑하여 표현
- RGB-D: 3D 데이터를 2차원 영상으로 표현하기 위해 컬러영상과 깊이 정보로 분리하여 표

현. 깊이 정보와 컬러영상을 합성하여 3D 데이터 생성 가능. 일반적으로 360도 전체를 표현하기보다는 특정한 시점(뷰) 표현

- 볼륨: 2차원의 픽셀과 비슷한 개념으로 부피를 가지는 픽셀인 복셀을 정의하여 연속된 복셀 값으로 3D 데이터 표현
- 다시점: 3D 딥러닝에 가장 많이 사용하는 표현으로 3D 데이터를 다시점으로 프로젝션하여 각 시점에서의 컬러 정보로 3D 표현. 이 경우 다시점 영상을 합성하여 3D 데이터를 생성하여 (일반적으로 매시로 생성) 서비스
- 포인트 클라우드: 3차원 공간 위치 및 색상 정보를 포함하는 포인트 정보. 3D 스캐너의 일반적인 출력 형태로 사용되며계측, 시각화 등에서 사용
- 메시: 일반적인 3차원 데이터 응용에서 가장 많이 사용하는 표현기법으로 3D 정점(Vertex), 다각형(삼각 혹은 사각형)의 면과 각각의 연결정보 등으로 구성된 3D 데이터
- 그래프: 가장 많이 사용하는 메시의 다른 형태의 표현으로, 그래프의 노드는 메시의 정점과 대응하고, 그래프의 에지는 메시의 연결정보에 대응하여 표현하는 방법

### 3. 국내외 표준화 기구의 3D 딥러닝 표준 동향

#### 3.1 국제표준화기구/국제전기기술위원회 합동 기술위원회(ISO/IEC JTC1)의 표준화 동향

JTC 1은 국제표준화기구와 국제전기기술위원회가 협력하여(JTC, Joint Technical Committee) 상호 간의 표준 통일 및 업무협조를 통해 정보기술 분야의 국제표준을 제정하기 위해 1987년에 설치한 조직이다. 현재 3,200개 이상의 국제표준을 발간하였다. 전 세계적으로 가장 큰 규모의 표준화 단체이며 34개국의 정식회원국과 66개국의 참관회원국으로 이루어져 있다. 우리나라는 1987년부터 정회원국으로 활동 중이다. JTC1 산하에는 22개의 기술 소위원회(SC)와 19개의 작업반(워킹그룹)이 있다.

##### 3.1.1 기술 소위원회 29(ISO/IEC JTC 1/SC 29)의 표준화 동향

기술 소위원회 29는 오디오, 비디오 등의 멀티미디어와 하이퍼미디어에 대한 압축 및 표현 표준을 만드는 기구이다. 586개 이상의 표준이 개발되었고, 현재 88개의 표준화 작업이 진행 중이다. 정지영상을 다루는 WG 1(JPEG)과 동영상, 음성, 3D 그래픽 데이터 등을 다루는 WG 11(MPEG)이 작업반으로 구성된다.

MPEG에서는 2019년 125차 회의에서 인공지능 기술을 이용하여 만들어진 추론 학습모델을 압축하는 NNR에 대한 최종 기술제안서 fCfP(Final Call for Proposal)을 공식 발표했다. 2020년 10월에 표준초안(DIS, Draft International Standard)을 공표하고, 2021년 10월에 최종 국제표준(IS, International Standard)을 공표할 예정이다. NNR은 다양한 플랫폼(Pytorch, Tensorflow)에서 생성된 추론모델의 상호 호환을 지원하며, 학습모델을 압축하여 모바일같이 컴퓨팅 파워가 한정된 상태에서도 학습 추론을 원활히 하도록 지원한다. 추론모델의 하이퍼파라미터가 압축(일반적으로 양자화)되어 추론 결과에 영향을 미칠 수 있지만, 일반적인 비디오 압축과 같은 방법으로 압축 효율과 추론 정밀도를 최적화해 압축방법 및 압축률을 정의한다. 영상 분류, 영상 압축 등에 대한 다양한 사용 시나리오가 있으며, 3D 콘텐츠에 대한 인식, 분

류도 그 대상이 될 수 있다. 이후 126차 회의부터 실질적인 기술기고가 시작되었고, 9개의 기술기고가 제출되었다. 기술제안에는 모바일 단말기 등을 위한 메모리 사용량과 성능 자체를 판단하는 사용자 시나리오에서의 성능, 그리고 압축률, 실행시간, LSTM(Long short-term memory)과 같은 다양한 형태의 네트워크를 압축하는 방법 등이 주요 이슈이다. 크게 3가지 CE(Core Experiment)를 진행 중이며 관련된 CE를 평가하기 위해 CTC(Common Test Conditions)관련 문서를 개정하였다. 현재 제안된 기술의 성능은 각 사용자 시나리오별로 30% 이내의 성능저하를 보이며, 인공지능 기술 선두권에 있는 중국과 유럽에서 적극적으로 표준화에 참여하고 있다. 국내에서는 한국항공대학교와 인시그널이 표준화 회의에 적극적으로 참여하고 있다. NNR에 제안된 방법은 대표적으로 LB(Local Binary)방법[3]이 논의되고 있다. 이는 CNN의 커널을 중간값(3X3 커널의 경우 정가운데값)을 기준으로 이진화하여 전송하고, 왜곡값의 크기에 따라 LB를 적용할지에 대한 플래그 정보를 보내는 방법으로 네트워크 압축을 진행한다. 제안된 방법에 따르면 8비트 유니폼 양자화 기법을 사용한 경우 25%의 압축률에 정확도는 88% 정도의 효율을 나타낸다. 향후 표준의 논의과정에서 최종 표준 문서의 내용은 변경될 수 있다.

### 3.1.2 기술 소위원회 42(ISO/IEC JTC 1/SC 42)의 표준화 동향

SC 42는 2017년 11월에 설립되어 인공지능의 용어·시스템 프레임워크·빅데이터 참조모델 등에 대한 표준 작업을 수행한다. 한국을 포함해 29개국의 정회원과 12개국의 준회원이 인공지능에 대한 국제표준을 만들어가고 있다. 현재 4개의 표준이 완성되었으며, 2018년에 빅데이터의 용어정리 및 사용자 시나리오에 대한 표준이 완성되었다. 표준화 기구 자체의 성숙도 측면에서 아직 초기 표준단체이다. 5개의 작업반과 1개의 합동 작업반 그리고 3개의 애드혹으로 구성되며 한국은 독일과 함께 합동 작업반을 이끌고 있다. 한국은 2019년 4월 아일랜드 더블린 회의에 15명의 대표단을 파견하여 다양한 산업 분야에 적용 가능한 인공지능 데이터의 생태계를 구축하는 표준을 제안하였다. 이를 통해 가공되지 않은 일반 데이터를 인공지능 기술에 바로 적용이 가능한 고품질 데이터로 만들 수 있게 하여 3D 콘텐츠를 포함하는 산업 전반에 걸쳐 적용하는 것을 목표로 한다.

### 3.2 오픈모바일연합(OMASpecWorks) 콘텐츠전송 작업반(CD)의 표준화 동향

오픈모바일연합은 2002년 6월에 모바일 이동통신 산업을 위한 개방형 표준을 만드는 기구로 설립되었다. 모바일 단말의 통신 및 서비스 프로토콜 표준을 위해 AT&T, 퀄컴, 에릭슨, 노키아 등이 참여하고 있다. 현재 정회원 27개사와 참관회원 16개사가 참여 중이다. 모바일 환경에서의 콘텐츠 서비스 작업반인 CD(Contents Delivery) 그룹은 2018년 10월부터 인공지능을 이용한 지능형 3D 콘텐츠 생성 및 서비스 프레임워크 표준인 3DCAPI(3D Contents Creation API) 표준을 진행 중이다. 한국전자통신연구원과 한성대학교가 주축이 되어 2D 영상 입력을 이용하여 3D 콘텐츠를 생성하는 서비스 API를 정의한다. 현재 전체 아키텍처와 각 시나리오별 사용자 시나리오를 정의하였으며, 콘텐츠 전송에 대한 기본 프로토콜을 정의하였다. 2020년 하반기 내에 최종표준안 후보(Candidate) 작업을 완료할 예정이다. 2019년 10월 분당에서

개최된 회의에서 산업화 행사(Industrial Day) 세미나를 통해 3DCAPI에 대한 시연을 진행하였고, 실제 산업환경에서 지능형 콘텐츠 제작 서비스의 가능성을 검증하였다.

### 3.3 TTA의 국내 표준화 동향

3D 딥러닝 관련된 국내표준은 TTA의 디지털콘텐츠 프로젝트 그룹(PG 610)에서 진행하고 있다. 학습데이터 파일포맷 표준과 지능형 콘텐츠 생성 아키텍처 등 두 가지 표준을 만들었다. 3D 학습데이터는 2D 영상과는 다르게 획득하기가 매우 어렵다. 2D의 경우는 사진 촬영을 통해 쉽게 획득할 수 있지만, 3D의 경우는 전문가가 전용 도구로 제작하거나, 스캐너와 같은 고가의 장비로 만들어야 하기에, 학습데이터를 재사용하기 위해 표준화된 파일포맷의 정의는 아주 중요하다. TTA에서 진행하는 다시점 영상을 기반으로 하는 표준(multiViewBasedProject)과 일반적인 3D 데이터를 직접 다루는 3DBasedProject로 나누어진다. 다시점 영상 기반의 경우 다시점을 표현하기 위한 카메라 구성에 대한 파라미터도 함께 표준화하여 실제 복원까지 가능한 형태로 정의되었다. 3D 모델 기반의 경우는 추후 개발될 수 있는 다양한 학습 방법에서의 확장을 위해 임의의 3D 파일포맷을 지원한다.

TTA에서는 2D 영상을 3D 콘텐츠로 만들기 위한 아키텍처 표준을 공표하였다[6]. 이는 카메라 등의 입력단에서부터 전처리를 통해 표준화된 데이터 포맷으로 변환하고, 온라인에 있는 데이터 생성단을 통해 3D 모델을 추론하여 서비스 단(사용자의 단말)으로 전송하는 구조를 정의한다.

## 4. 맺음말

최근 인공지능 기술의 발전이 가속화되면서 실감콘텐츠와 밀접한 관련이 있는 3D 분야도 시장이 확대되고 관련 표준화가 함께 진행되고 있다. 이런 일련의 활동에서 TTA와 국가기술표준원이 중추적인 역할을 수행하고 있다. 최근에 시작된 표준인 만큼 정부, 기업, 대학, 연구소 등에서 협력하면 우리나라가 주도적으로 표준화를 진행할 수 있는 분야이다.

### 참고문헌

- [1] Eman et al, 'A survey on Deep Learning Advances on Different 3D Data Representations,' arXiv 1808.01462(2018)
- [2] 이승욱, '인공지능을 이용한 3D 콘텐츠 기술 동향 및 향후 전망', 전자통신동향분석 Vol. 34 No. 4 15-22, 2019.
- [3] 문현철 외, 'MPEG-NNR의 Local Binary 방법을 이용한 CNN 압축', 한국통신학회 학술대회 논문집, 270-271, 2019.11.
- [4] 이승욱 외, 'Change of architectural diagram', OMASpecWorks OMA-CD-3DCAPI-2019-0016-CR, 2019.10.
- [5] TTA PG610, TTAK.KO-10.1189, '2D 이미지를 3D 모델로 생성하기 위한 딥러닝 학습 데이터 포맷', 2019.12.
- [6] TTA PG610, TTAK.KO-10.1190, '2D 이미지를 이용한 휴먼 캐릭터 생성 아키텍처', 2019.12.