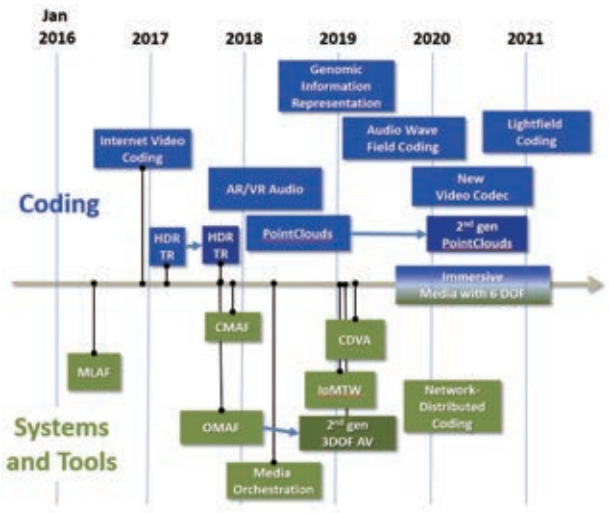
MPEG-I 표준과 360도 비디오 콘텐츠 생성

호요성
GIST 전기전자컴퓨터공학부

Ⅰ. 서 론

2010년대 초반에는 3차원 기술을 앞세워 여러 연구기관과 기업들이 3차원 영상을 디스플레이할 수 있는 장치를 개발하기 위해 많은 시간과 노력을 들였다. 3차원 영상처리 하드웨어의 발전에 따라 영화, 게임, 광고 등 다양한 분야에서 3차원 영상 기술을 이용한 콘텐츠들이 제작되었다. 대표적인 3차원 기술 기반 콘텐츠 성공 사례로 영화 '아바타'를 들 수 있다. 이러한 성공에 힘입어 각국의 기업들은 고화질, 고해상도의 3차원 영상을 시청할 수 있는 디스플레이 장치를 개발했다. 하지만 3차원 영상을 시청하기 위해서는 시청자가 특수 안경을 착용해야 한다는 큰 불편함을 감수해야 했다. 또한, 입체 영상을 시청할 때 발생하는 어지럼증과 두통 현상이 3차원 영상 기술의 큰 한계로 지적되었다. 이러한 문제를 해결하지 못함에 따라 3차원 디스플레이 시장은 점차 위축되었고, 기존의 문제를 해결하기 위해 다른 방법으로 기술 접근을 해야 한다는 인식을 갖게 되었다. 2010년도 중반에 접어들었을 때에도 여전히 소비자들은 3차원 영상 콘텐츠 시청을 원했기 때문에 기존의 3차원 영상 디스플레이의 문제를 해결하기 위한 방법들이 개발되었다.

미국의 Oculus사는 2014년 후반기에 Oculus Rift 버전 1을 출시하며 사용자의 머리에 장착하여 시청할 수 있는 디스플레이(Head Mount Display, HMD) 시장을 개척했다[1]. 당시 HMD는 고가의 제품이었기 때문에 일반 소비자들이 쉽게 접근하기 어려웠지만, 미국의 Google에서 종이상자와 휴대전화로 만들 수 있는 Cardboard를 시장에 판매하여 일반 소비자들도 손쉽게 HMD 기술을 접할 수 있게 되었다. 잇따라 삼성전자에서는 Oculus사와 협업하여 GearVR을 제작하여 판매했으며, GearVR은 삼성에서 제작한 고급 휴대폰을 디스플레이 장치로 사용하기 때문에 QHD(Quad High Definition) 2560×1440 해상도의



〈그림 1〉 MPEG 표준화 로드맵

VR(Virtual Reality) 영상 시청이 가능하게 되었다.

하지만 이러한 VR 장치들은 대부분 모바일 환경에서 구동되기 때문에 일반 PC나 대형 디스플레이 장치에 내장되어 있는 그래픽 처리 장치를 통해 제공되는 영상에 비해 화질이 떨어지게 된다. 또한, 360° 전방향 영상 콘텐츠를 제작하기 위해 CG(Computer Graphics) 영상이 아닌 실제 촬영된 영상 사용 할 경우 모바일 장치에서 영상을 처리하기는 더욱 어렵다.

실제 촬영된 영상을 VR 장치를 통해 고품질의 전방향 파노라마 영상을 제작하기 위해서는 4K UHD 해상도 정도 되어야 하는데, 이 경우 처리해야 할 데이터 량이 급격히 증가하게 된다. 또한, 막대한 양의 영상 콘텐츠는 대역폭과 전송 속도의 한계로 인해 네트워크를 통해 전송하기 쉽지 않다[2]. VR장치를 통해 전방향 영상을 시청할 경우 일반적으로 사용자 머리의 움직임을 3방향으로 정의하는데, 이를 3DoF(Degree of Freedom)이라 한다. 3DoF는 Yaw, Pitch, 그리고 Roll을 의미하며, 사용자 머리의 움직임을 트래킹하는 방향축이 기준이 된다. 하지만 VR 영상을 시청할 경우 객체 사이의 가려짐 현상이나 촬영 카메라의 시야각 문제로 인해 3DoF만을 통해 현실감 있는 VR 영상을 시청하기는 어렵다

MPEG 표준화 그룹에서는 116차 모임에서 MPEG-I 그룹을 만들어 몰입형, 전방향 비디오를 위한 포맷과 포인트 클라우드 등 기존의 기술적 문제를 해결하기 위한 표준화 로드맵을 제시하고 논의를 진행하고 있다[3]. 2017년부터 표준화 작업을 시작하여 2021년까지 표준화 작업을 완료하는 것을 목표로 하고 있으며, 내부적으로 5개의 서브파트가 존재하며, 각 파트별 표준화 작업은 〈그림 1〉에 나타나 있듯이 연도별로 세분화된 MPEG 로드맵을 따르고 있다[4].

Ⅱ. MPEG-I Phase의 분류

MPEG-I 표준에서는 시청자에게 현재의 3차원 영상 콘텐츠보다 더 자유롭고 현실감있는 전방향 영상을 제공하기 위한 표준화 작업을 진행하고 있다. 이러한 목적을 위해 MPEG-I는 서브파트 그룹을 만들어 각 기술개발 작업을 세분화하여 진행하고 있으며, 부호화 및 복호화 되는 데이터의 종류와 양에 따라 Phase 1.a, Phase 1.b, 그리고 Phase 2와 같이 3단계로 나누어 기술 개발을 진행할 예정이다.

1. Phase 1

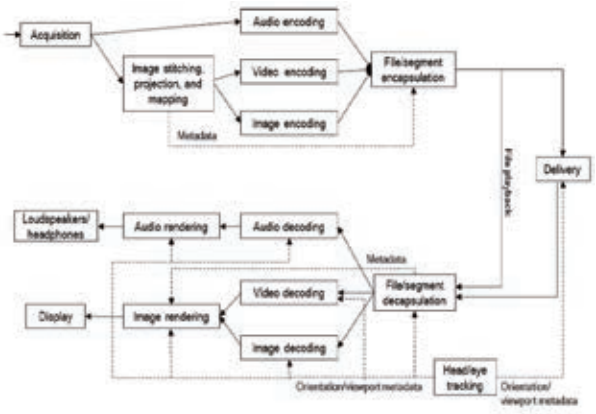
표준화 단계(Phase)는 크게 Phase 1과 Phase 2로 구분되는데, Phase 1은 2개의 서브파트로 구성되어 있으며, Phase 2는 3개의 서브파트로 구성된다. Phase 1의 서브파트 1은 "Technical Report on Immersive Media"로 몰입형 미디어 기술에 대한 구조와 기술을 다룬다. Phase 1의 서브파트 2인 "Omnidirectional MediA Format (OMAF)"는 360° 카메라로 촬영한 영상 콘텐츠를 네트워크를 통해 전송하기 위한 부호화 및 복호화를 위한 기술 개발을 주로 진행, 최종 수신단에서 6DoF로 복원된 영상을 제공하는 것을 목표로 하고 있다. Phase 1.a의 전체적인 목표는 전방향 VR 영상을 네트워크를 통해 저장 및 전송하도록 하는 것이다. 단, Phase 1.a는 시야각이 3DoF로 한정되어 있는데, 3DoF는 〈그림 2〉와 같이 시청자가 고정된 위치에서 영상을 감상할 때 시청자 머리의 Yaw, Pitch, 그리고 Roll에 대한 움직임이 전방 시야각 X, Y, Z축에 대해 한정되어 있는 상황을 의미한다[5]. Phase 1.a에서는 최대 360° 구형 영상에 대한 영상 및 비디오 콘텐츠를 제공하는 것을 목표로 하고 있으

〈그림 2〉 3DoF의 시야각 및 자유도

며, 네트워크 환경이 지원 가능한 경우 4K 60fps의 영상을 부호화, 복호화가 가능하도록 2017년 후반기까지 표준화 작업을 진행할 예정이다.

〈그림 3〉과 같이, Phase 1.a는 단일 영상의 스티칭, 프로젝션, 그리고 매핑 정보들을 기반으로 영상 및 비디오 부호화를 수행하며, 동시에 오디오 부호화도 진행한다. 복호화 부분에서도 마찬가지로 동일한 데이터에 대해 복호화를 진행하며 추가적으로 전방향 영상을 지원하기 위해 사용자가 바라보는 시점에 대한 트래킹 정보가 전송 데이터에 포함된다.

Phase 1.b의 경우 Phase 1.a의 시청 가능 시야각의 자유도가 일부 증가한 3DoF+ 콘텐츠 제공을 목표로 하고 있다. 〈그림 4〉에 보인 것처럼, 3DoF+는 3DoF에서 후



〈그림 3〉 MPEG-I Phase 1.a

〈그림 4〉 3DoF+의 시야각 및 자유도

방으로의 Yaw, Pitch, 그리고 Roll에 대한 움직임이 일부 제한적으로 추가된다[6]. 즉, 3DoF에서 제한적으로 시점의 자유도가 증가한 것이다.

고품질의 3DoF+ 환경을 구성하기 위해서는 사람이 바라보는 시점을 트래킹하여 영상을 자연스럽게 생성할 수 있도록 데이터 전송 지연 문제가 없어야 한다. 그리고 HMD 장치를 이용하여 사용자간 실시간 영상을 주고받기 위해서는 데이터 전송에 대한 최적화 기술도 Phase 1.b에서 고려해야 할 점이다. Phase 1.b는 깊이정보가 포함된 여러 영상들과 비디오 데이터들이 부호화 입력 정보로 사용된다. 복호화단에서는 전송된 영상의 비디오 데이터와 깊이 정보를 함께 사용하여 렌더링을 수행함으로써 최종 데이터를 생성한다.

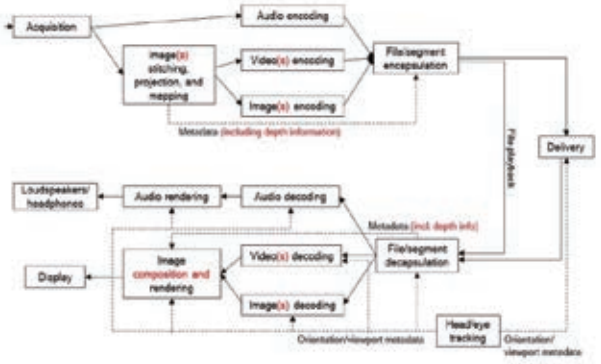
Phase 1.b에서는 최종 데이터에 깊이정보가 포함되어 전송되기 때문에 Phase 1.a보다 높은 시점의 자유도를 갖게 된다. Phase 1.b의 시스템 구성도는 〈그림 5〉에 나타나 있다.

2. Phase 2

Phase 2는 Phase 1의 서브파트 2에 이어 서브파트 3부터 시작하게 되는데, 이는 "Immersive Video"라고 부르며, 전방향 비디오 콘텐츠를 제작하고 개발하는 작업을 중점적으로 진행한다. 서브파트 4는 "Immersive Audio"이며, 전방향 비디오 콘텐츠에 사용되는 오디오 데이터를 제작하는 작업을 수행한다. 마지막 서브파트 5인 "Point Cloud Compression"는 라이트필드 영상과 함께 전방향 비디오 콘텐츠 제작을 위해 사용될 기술이다. 〈그림 6〉에 보인 것처럼, Phase 2의 목표는 시청자가 자유롭게 움직일 수 있는 환경에서 시점의 제한이 없는 6D0F 영상을 제공하는 것이다. Phase 2의 가장 중요한 개발 요소는 6DoF 영상 전송이 가능한 비디오 코덱의 개발이다.

Phase 2에 사용되는 데이터의 종류는 비디오, 정지영상, 오디오, 라이트필드 및 포인트 클라우드 등 다양한 데이터가 처리되고 전송되어야 하기 때문에 이를 안정적으로 지원하기 위한 차세대 비디오 코덱의 개발은 Phase 2에서 매우 중요한 요소이다.

Phase 2 시스템의 경우 입력으로 사용되는 영상정보



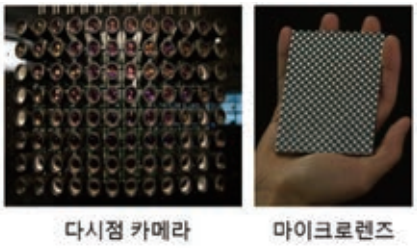
〈그림 5〉 MPEG-I Phase 1.b

〈그림 6〉 6DoF의 시야각 및 자유도

증가로 인해 부호화해야 하는 데이터의 종류가 Phase 1.b와 비교하여 많이 증가하게 된다. 원활한 데이터 전송을 위해 지금보다 빠르고 넓은 대역폭을 갖는 네트워크 환경이 필요하기 때문에 5G의 개발이 완료되는 시점에서 MPEG-I의 전체적인 시스템 구성이 완료되도록 로드맵이 설계되어 있다.

III. MPEG-I Visual 기술 동향

MPEG-I는 자연스러운 전방향 영상 콘텐츠를 오디오와 함께 제공하는 것에 초점을 두고 표준화 작업을 진행하고 있다. 그 중 서브파트 3인 "Immersive Video"는 현재 3DoF+와 6DoF를 위한 실험영상 제작에 많은 노력을 기울이고 있다[7]. 최근 라이트필드 카메라가 상용화되어 여러 분야에 사용되고 있으며, 라이트필드 영상을 이용한 깊이지도 생성 및 중간 시점 영상 합성 등 다양한 알고리즘이 개발되고 있다. MPEG-I도 Phase 2에서 라이트필드 영상을 사용하여 전방향 영상 콘텐츠를 제작하는 것을 목표로 하고 있으며, 현재 MPEG 미팅에서 라이트필드 카메라를 이용한 다시점 영상 제작 방법에 대한 논의가 활발하게 이루어지고 있다. 특히, 라이트필드 카메라



〈그림 7〉 다시점 카메라 구조와 마이크로 렌즈 배열

영상을 이용하여 정확한 깊이지도를 생성하는 방법에 대한 연구의 필요성이 증대되고 있다.

〈그림 7〉에 나타나 있듯이, 라이트필드 카메라는 일반 카메라에서는 사용되지 않는 마이크로 렌즈 배열을 사용하기 때문에 다시점 카메라로 촬영한 영상과 동일한 결과를 얻을 수 있다.

하지만 라이트필드 영상은 다시점 카메라로 획득한 영상보다 시점간 거리가 매우 좁다는 특징을 가지고 있다. MPEG-I에서는 획득한 깊이정보를 기반으로 중간시점 영상을 생성하거나 3차원 모델링을 수행해야 하기 때문에 정확한 깊이정보를 획득하는 것이 중요한 문제이다. 일반적으로 라이트필드 영상으로부터 깊이영상을 획득하기 위해 스테레오 정합 방법을 사용한다. 최근 MPEG 미팅에서 라이트필 영상의 원본 깊이정보가 없을 경우 생성된 깊이지도의 정확성을 평가하기 위해 포인트 클라우드로 생성된 깊이지도를 시각화하여 평가하는 방법들이 논의되었다[7-8].

라이트필드 영상을 사용하여 스테레오 정합을 수행할 때, 일반적으로 비용 함수는 식 (1)과 같이 Zero Mean Normalized Cross-Correlation (ZNCC) 또는 식 (2)처럼 Sum of Absolute Differences (SAD)와 영상의 기울기 항을 동시에 사용하는 등 다양한 비용 함수를 모델링하여 사용한다. 라이트필드 영상의 스테레오 정합 역시 일반적인 양안 영상의 정합 방법에 사용되는 후처리 과정이 적용된다. 가이드 영상 필터링[9] 또는 계층적 해상도 최적화[10] 방법을 적용하여 고품질의 깊이지도를 생성할 수 있게 된다.

$$C(x,y,u,v)=\frac{I(u+x,v+y)-\mu(I(u,v),n)}{\sigma(\bar{I}(u,v),n)}$$
$$\bar{I}(u,v)=I(u,v)-\mu(I(u,v),n) \quad (1)$$

$$C(x) = \alpha \cdot C_{SAD} + (1-\alpha) \cdot C_{Grad}(x)$$
$$C_{SAD}(x) = \sum_{u \in Vx \in W} \sum \arg min(|L(u,x) - L_{\alpha}(u,x)|)$$
$$C_{Grad}(x) = \sum_{u \in Vx \in W} \sum [\arg min(\Delta_{x}(L, L_{\alpha}, x, u) + \arg min(\Delta_{y}(L, L_{\alpha}, x, u))] \quad (2)$$

라이트필드 영상은 마이크로렌즈 배열을 통해 촬영된 영상만을 통칭하지 않는다. 〈그림 8〉과 같이 일반 카메라를 사용하여 상하 시차가 존재하며 좁은 베이스라인을 갖도록 촬영된 영상도 라이트필드 영상이라고 말할 수 있다. 이렇게 얻은 실험 영상을 기반으로 식 (1)을 사용하여 획득한 깊이지도와 DERS (Depth Estimation Reference Software)를 통해 획득한 깊이지도 결과를 〈그림 9〉에 나타내었다[7]. DERS는 스테레오 영상을 이용하여 깊이지도를 생성할 수 있도록 MPEG 그룹에서 제작하여 공개한 소프트웨어이다.

〈그림 9〉의 좌측은 식 (1)을 이용하여 획득한 깊이지도를, 우측은 DERS를 이용하여 획득한 깊이지도를 나타낸다. 이 실험 영상의 경우 원본 깊이지도가 존재하지 않기 때문에 객관적으로 평가할 수 있는 방법이 없다. 깊이지도의 정확성을 평가하기 위해 생성된 깊이지도를 기반으로 3차원 공간상에 포인트 클라우드를 생성하여 정확성을 주관적으로 평가한다. 라이트필드 실험 영상에 대한 원본 깊이지도가 존재하지 않을 경우에는 생성된 깊이지도를 사용하여 중간시점 영상을 생성하거나 3차원 객체 모델링을 수행함으로써 생성된 깊이지도의 정확성을 평가할 수 있다. 이 실험 영상은 포인트 클라우드를 생성하여 깊이지도를 평가하는 방법을 채택했기 때문에 라이트필드 카메라 시스템에 대한 파라미터를 알아야 하기 때문에 미리 카메라 캘리브레이션이 수행되어야 한다.

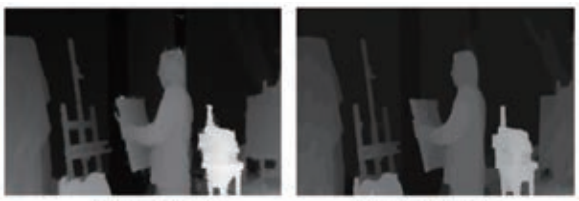
〈그림 10〉은 〈그림 9〉와 같이 획득한 깊이지도를 기반으로 포인트 클라우드를 생성하고, 원본 라이트필드 영상의 텍스쳐 정보를 사용하여 Meshlab tool[11]을 기반으로 시각화한 결과를 나타내고 있다.

〈그림 10〉의 결과에서 알 수 있듯이, 식(1)을 이용한 깊이지도 기반 포인트 클라우드 시각화 결과가 DERS 깊이지도 기반 시각화 결과에 비해 정확하게 표현된 것을 확인할 수 있다. 하지만 DERS는 현재 라이트필드 영상에 대한 깊이지도를 생성하는 알고리즘이 아니기 때문에 두 결과 중 어느 것이 우수하다고 객관적으로 평가할 수는 없다. 이러한 문제를 해결하기 위해 라이트필드 영상에 대해 원본 깊이지도를 포함하고 있는 실험 영상을 제작하여 제공하는 연구를 지속적으로 진행해야 한다. 또한 라이트필드 영상으로 생성된 포인트 클라우드 결과를 객관적으로 평가할 수 있는 평가 방법에 대한 연구도 같이 진행되어야 한다.

Ⅳ. 전망 및 전망

최근 3차원 영상을 시청할 때 발생하는 문제점에 착안하여 다양한 연구들이 진행되고 있는데, 대표적으로 HMD를 이용한 VR 전방향 콘텐츠 시청을 들 수 있다. 초기 VR 영상은 사용자에게 다양한 시점의 영상을 제공함으로써 기존 3차원 디스플레이를 통해 느낄 수 있는 영상에 비해 보다 많은 시점의 자유도를 제공했다. 하지만

〈그림 8〉 라이트필드 영상



〈그림 9〉 스테레오 정합을 이용한 깊이지도 비교

〈그림 10〉 Meshlab기반 깊이지도 시각화

사용자들은 시점의 자유도뿐만 아니라 VR을 통해 고화질의 풍부한 영상 콘텐츠를 감상하기를 바랐고, 기존의 VR보다 덜 제한적인 시점의 자유도를 원하게 되었다. 이에 따라 MPEG 표준화 그룹에서는 MPEG-I를 조직하여 표준화 작업을 진행하고 있다. 특히, OMAF 개념과 3DoF로부터 시작하여 이상적인 6DoF VR영상을 제작하기 위한 장기적인 로드맵을 작성하고, 5개의 서브파트로 나누어 표준화 연구를 진행하고 있다. 특히, 이번 118차 MPEG 미팅에서 MPEG-I VIsual 그룹은 3DoF+, Omnidirectional 6DoF, Windowed 6DoF 그리고 6DoF와 같이 시청자가 경험할 수 있는 시각적 자유도에 대한 개념을 세분화했다. 119차 MPEG 미팅에서는 시각적 자유도에 대한 세부 예시들을 정의하고 다양한 실험 영상들을 제공하는 것을 목표로 하고 있다. 앞으로 몰입형 비디오의 표준화 작업을 성공적으로 수행하기 위해서는 MPEG-I의 주제에 많은 관심을 가져야 하며, 각 세부 파트별로 관련있는 연구 성과를 기고서로 제출하여 표준화 작업에 많은 기여를 해야 할 것으로 보인다.

감사의 글

본 연구는 미래창조과학부 '범부처 Giga KOREA 사업'의 일환으로 수행하였음. [GK16C0100, 기가급 대용량 양방향 실감 콘텐츠 기술 개발]

참고문헌

[1] M. A. Conn, and S. Sharma, "Immersive Telerobotics using the Oculus Rift and the 5DT Ultra Data Glove," CTS, pp. 387–391, Nov. 2016.

[2] R. Kijima, and K. Yamaguchi, "VR device time-Hi-precision Time Management by Synchronizing Times Between Devices and Host PC Through USB," IEEE Virtual Reality(VR), DOI, 10.1109/VR.2016.7504723, March 2016.

[3] "New Work Item Proposal on Coded Representation of Immersive Media," ISO/IEC JTC1/SC29/WG11, N16541, Jan. 2017.

[4] "MP20 Roadmap," ISO/IEC JTC1/SC29/WG11, N16719, Jan. 2017.

[5] M.L. Champel and R. Dore, "Quality Requirements for VR," ISO/IEC JTC1/SC29/WG11, M39979, Jan. 2017.

[6] "MPEG-I Use Cases for omnidirectional 6DoF, windowed 6DoF, and 6DoF," N16767, April 2017.

[7] D. Doyen, G. Boisson, N. Sabater, and V. Dreyfus, "Estimation and Visualization of Depth from Light Field Content," M40597, April 2017.

[8] J.H. Mun, and Y.S. Ho "Light-Field Depth Map Generation and Visualization," M40289, April 2017.

[9] K. He, J. Sun, and X. Tang, "Guided Image Filtering," IEEE Trans. on PAMI, Vol. 35, No. 6, June 2013.

[10] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with and Application to Stereo Vision," IJCAI, vol. 2, pp. 674–679, Aug. 1981.

[11] www.meshlab.net/#download

호요성

- 1981년 서울대학교 공과대학 전자공학과 학사
- 1983년 서울대학교 공과대학 전자공학과 석사
- 1989년 Univ. of Califonia, SB 전기컴퓨터공학과 박사
- 1983년~1995년 한국전자통신연구원 선임연구원
- 1990년~1993년 미국 Philips 연구소 선임연구원
- 1995년~현재 광주과학기술원 교수
- 1995년~현재 실감방송연구센터 센터장
- 2016년~한국방송미디어공학회 회장
- 2017년~현재 IEEE Fellow

〈관심분야〉
Digital Signal and Image Processing, Image and VIdeo Data Compression, Digital Television and High Definition Television System, 3DTV and Free-viewpoint Video System