

# 멀티카메라 동영상의 실시간 스티칭 기술을 이용한 실감형 스포츠 및 문화공연 중계 적용

임성용 \_ 한국전자통신연구원 책임연구원  
 이회경 \_ 한국전자통신연구원 책임연구원  
 김현철 \_ 한국전자통신연구원 책임연구원  
 석주명 \_ 한국전자통신연구원 책임연구원  
 정순홍 \_ 한국전자통신연구원 책임연구원  
 서정일 \_ 한국전자통신연구원 실감미디어연구실 실장



## 1. 머리말

평창올림픽을 기점으로 UHD, 가상현실, IoT, 5G, AI 등 우리나라가 사랑하는 ICT 기술을 스포츠 경기와 접목하기 위한 다양한 시도가 있었으며, 특히, 360VR 서비스의 보편화로 인하여 넓은 시야각을 제공하는 광시야각 동영상 기반의 스포츠 경기 애플리케이션이 상용화되었다[1]. 이러한 넓은 시야 영역을 소비하기 위해서는 제한된 시청 환경을 제공하는 데

내 TV가 아닌 이동 단말을 활용한 헤드 마운트 스크린 혹은 광장에 설치한 초대형 스크린 등이 적용되었으며, 일례로는 [그림 1]에서와 같이, 고화질 광시야각 동영상을 활용하여 쇼트트랙 경기를 넓은 대화면에서 소비하는 새로운 몰입형 미디어 서비스가 시도되었다.

본고에서는 이러한 광시야각 동영상 서비스를 스포츠 경기와 문화 공연에 적용하여 수백 km 떨어진 원격지에서 실시간으로 현장감을 그대로 느낄 수 있



※ 출처: 평창ICT올림픽 가이드북

[그림 1] 평창올림픽 고품질 광시야각 대화면 시범 서비스(2018. 인천공항)



[그림 2] 광시야각 동영상 서비스를 위한 핵심 기술 구성

는 적용 사례를 소개하고자 한다. 고품질 광시야각 동영상은 360VR, 180VR 그리고 UWV(Ultra Wide Vision)과 같이 8Kx4K, 4Kx4K, 12Kx2K 등의 해상도를 가지면서도 고품질 화소밀집도를 제공하는 동영상을 모두 포함한다.

## 2. 고품질 광시야각 동영상 서비스를 위한 핵심 기술

일반적인 중계, 방송 서비스와는 달리 고품질 광시야각 동영상을 위해서는 멀티카메라를 사용하거나, 멀티 센서와 렌즈를 통합한 일체형 카메라를 활용해야 FullHD 이상의 균일한 화소밀집도를 유지할 수 있다. [그림 2]에서와 같이 멀티카메라를 활용하기 위해서는 안정적으로 카메라를 정렬할 수 있어야 하고, 시간적으로도 동기가 제공되어야 한다. 또한, 동시에 멀티카메라의 입력을 실시간으로 확인할 수 있어야 최대한 하나의 카메라처럼 촬영할 수 있고, 촬영 시 품질을 보장할 수 있다. 최근 360VR을 지원하는 상용 일체형 카메라의 경우 자동화된 도구들, 특히 자동 색상 보정, 실시간 모니터링 등을 제공하고 있어 촬영 편의성 측면에서 큰 장점을 가지고 있으

나, 제한된 화각과 화소밀집도 그리고 제한된 품질 측면에서 단점이 있다. 반대로 방송용 멀티카메라와 카메라 정렬 구조체 등을 활용하는 경우, 화각 설정이 자유롭고, 고급 카메라와 렌즈를 활용할 수 있는 장점이 있으나, 전체 멀티카메라 세트의 크기가 매우 크고 복잡하기 때문에 촬영 편의성은 떨어지는 단점이 있다.

멀티카메라로 촬영된 영상들을 하나의 광시야각 동영상으로 생성하기 위해서 여러 동영상을 하나의 균일한 동영상으로 변환해 주는 실시간 스티칭 기술이 필요하다. 멀티카메라의 광학 원점을 물리적으로 일치시킬 수 없기 때문에 각 카메라로부터 획득한 동영상으로부터 카메라의 3차원 자세와 방향을 추측하고, 최대한 기하 오류와 색상 차이를 줄일 수 있도록 경계면을 부드럽게 이어 붙이는 역할을 한다. 특히 실시간 중계, 방송 서비스를 위해서는 이 처리 과정을 실시간, 즉 8Kx4K 해상도 혹은 12Kx2K 해상도 동영상을 최소 초당 30프레임 이상 처리하고 출력할 수 있어야 한다.

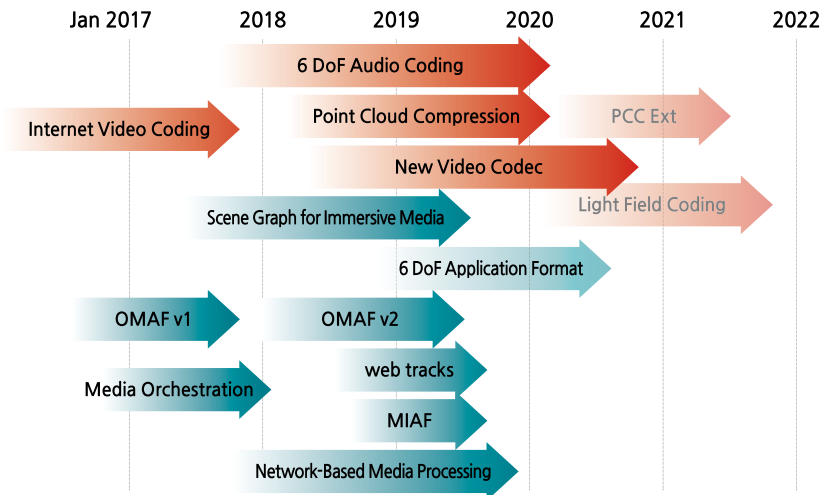
한편, 최종 사용자에게 연결되는 네트워크 대역폭을 효율적으로 활용하기 위해 동영상 부복호화 및 전송 스트리밍 방법이 필요하다. 일반적인 상용 부

호화기는 FullHD 혹은 UHD급을 제공하고 있으나, 8K급 혹은 12Kx2K급은 상용화 제품을 활용할 수 없기 때문에 기존 UHD 급 부호화기를 다수로 묶어 활용하거나 혹은 별도의 부호화 방법을 적용해야 한다. 특히 사용자의 관심 영역 혹은 시선 정보를 바탕으로 전체 동영상을 다수의 영역, 즉 타일로 나누고 이를 필요한 타일들만 전달하는 타일기반 적응적 스트리밍 및 복호화 방법을 활용하는 추세에 있다.

마지막으로 사용자는 멀티패널 혹은 멀티프로젝션 디스플레이 시스템을 활용하여 전체 광시야각 동영상을 넓은 대화면에 재생하거나, 개인 HMD(Head Mounted Display)를 활용하여 사용자의 시선에 해당하는 영역을 이동하면서 광시야각 동영상을 소비할 수 있다. 이를 위해서는 UHD 이상의 해상도를 복호화 할 수 있어야 하고, 재생 환경에 따라 실린더 혹은 구형 공간에 동영상을 프로젝션 할 수 있는 실시간 재생기술을 제공해야 한다. 또한, 사용자의 시선 혹은 관심영역에 따라 동영상의 일부영역은 고품질, 이외의 영역은 저품질 스트림을 수신하여 적응적 타일 기반 복호화 및 재생 기술을 필요로 한다.

### 3. 광시야각 동영상 관련 MPEG 표준화 동향

360VR을 비롯한 몰입형 미디어와 관련하여 국제 표준화 단체인 MPEG에서는 MPEG-I (Coded Representation of Immersive Media) 라는 표준화 프로젝트명으로 [그림 3]에서와 같이 부호화 분야와 시스템 분야에서 여러 가지 표준화 활동을 진행하고 있다. 이 중 광시야각 동영상과 관련된 부분만 살펴보면, 부호화 분야에서는 HEVC(High Efficiency Video Coding)에서 지원하는 타일링 부호화 방법을 확장하여, 슬라이스 간 의존성을 가지지 않는 방법을 지원하는 MCTS(Motion-Constrained Tile Set)을 추가하여 주변 타일이 수신되지 않더라도 수신된 타일 영역만으로 동영상의 복호화가 가능하도록 함으로써, 사용자의 시선에 적합한 타일만을 수신하여 대역폭을 효율적으로 활용할 수 있다[3]. 시스템 분야에서는 전방위 미디어를 저장할 수 있는 미디어 포맷인 OMAF(Omnidirectional Media Format) 버전 1 표준을 완료하여, 기본적으로 ISO/BMFF(ISO Base Media File Format) 확장, 360VR 미디어 표현



[그림 3] 몰입형 미디어 관련 MPEG 표준화 로드맵[2]

에 필요한 프로젝션 방법, 패키징 방법, 사용자 시선 제공 방법 등을 추가하였다. 현재 진행 중인 버전 2에서는 복수의 시점(Viewpoint), 오버레이 방법, 복수의 360VR 미디어 등을 제공하여 3DoF+ 미디어를 제공하는 것을 목표로 하고 있다.

한편, 멀티카메라와 멀티스크린을 사용하는 분야에서는 MORE(Media ORchEstration) 표준을 완료하였으며, 멀티 소스와 싱크의 위치정보, 방향정보 그리고 서로 다른 시간축을 동기화 할 수 있는 시간 동기화 신호를 표준화하였다[4]. 이를 확장하여 복수의 미디어 개체와 네트워크상의 클라우드 혹은 원격 프로세서를 활용하는 표준 또한 NBMP(Network-Based Media Processing)라는 이름으로 진행 중이다[5]. NBMP에서는 복수의 미디어 소스가 제공하는 복수의 미디어 스트림을 입력으로 하는 네트워크 상에 존재하는 미디어 처리 엔진을 활용하도록 워크플로우를 기술하고, 각 미디어 처리 기능을 가진 태스크를 다수 생성하여 소비자가 원하는 이머시브 미디어를 제공하는 것을 목표로 한다.

#### 4. 고품질 광시야각 동영상을 이용한 실감형 서비스 적용 사례

지금까지 기술한 고품질 광시야각 동영상을 위한 핵심기술과 표준화 현황을 바탕으로 원격지에서도 현장감을 느낄 수 있는 실감형 문화공연 및 스포츠 경기 중계 시범을 수행하였다. 먼저, 2018년 2월 평창올림픽 문화공연 행사의 일환으로 펼쳐진 난타공연을 인천공항 내 대형 멀티패널 환경에서 소비하는 사례와 2018년 9월 네덜란드 축구 국가대표팀의 경기를 암스테르담과 대한민국 대전을 실시간 연결하여 멀티패널과 HMD를 활용하여 소비하는 것을 목표로 하였다.

#### 4.1 촬영, 재생 환경과 요구사항

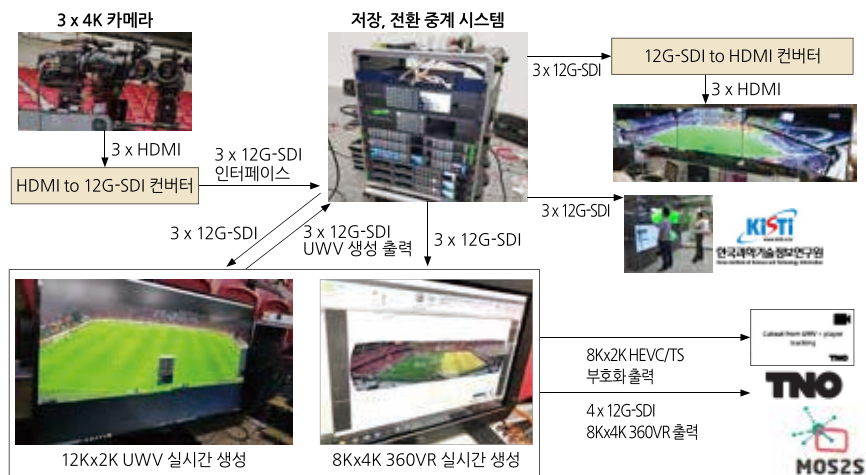
문화공연 광시야각 동영상 서비스의 경우 평창올림픽 조직위와 과학기술정보통신부의 협조를 통해 강릉아트센터에서 펼쳐진 문화공연을 인천공항 공항철도 라운지에 설치된 48개의 멀티패널에서 실시간 재생하는 것을 목적으로 하였다. 방송용 4K 카메라 6대를 활용하여 3대는 하나의 광시야각 동영상을 생성하고, 다른 3대는 개별 출연자를 확대 촬영하는 용도로 사용하며, 이를 중계서버를 통해서 전환하도록 하였다. 스포츠 경기의 광시야각 서비스는 유럽공동연구과제인 MOS2S(Media Orchestration from Sensors to Screens)의 2차년도 현장 시험의 일환으로 수행되어 네덜란드, 벨기에, 터키 그리고 대한민국의 16개 기관이 함께 참여하였다. 암스테르담의 Johan Cruyff Arena 경기장에서 펼쳐지는 네덜란드 축구 국가대표팀의 경기를 촬영하여, 대한민국 대전 KISTI에 설치된 멀티패널 혹은 HMD에서 재생하는 것을 목표로 하였다. <표 1>에서는 두가지 적용 사례가 가지는 주요 환경을 비교하고 있다.

#### 4.2 촬영 시스템 및 송수신 구성

문화공연과 스포츠 경기 촬영, 스티칭(생성), 중계를 위한 시스템 구조는 카메라 대수, 구조체 종류, 스티칭 서버의 종류, 네트워크 환경 인터페이스가 일부 상이한 점을 제외하고는 전체적으로 유사하다. [그림 4]에서는 공통된 멀티카메라 촬영, 스티칭 및 중계를 위한 시스템 구조를 설명하고 있다. 4K급 방송용 카메라의 광학원점 사이의 거리를 최소화하기 위하여 방사형이 아닌 집중형으로 배치를 하였고, 미세조정이 가능한 고니어미터와 실시간 모니터링 시스템을 통해서 3차원 공간, 실린더 혹은 구형 공간에서 멀티카메라 동영상 간의 기하보정, 색상보정을 수행하였다. 이렇게 촬영된 3개의 4K스트림은 12G-SDI 인터

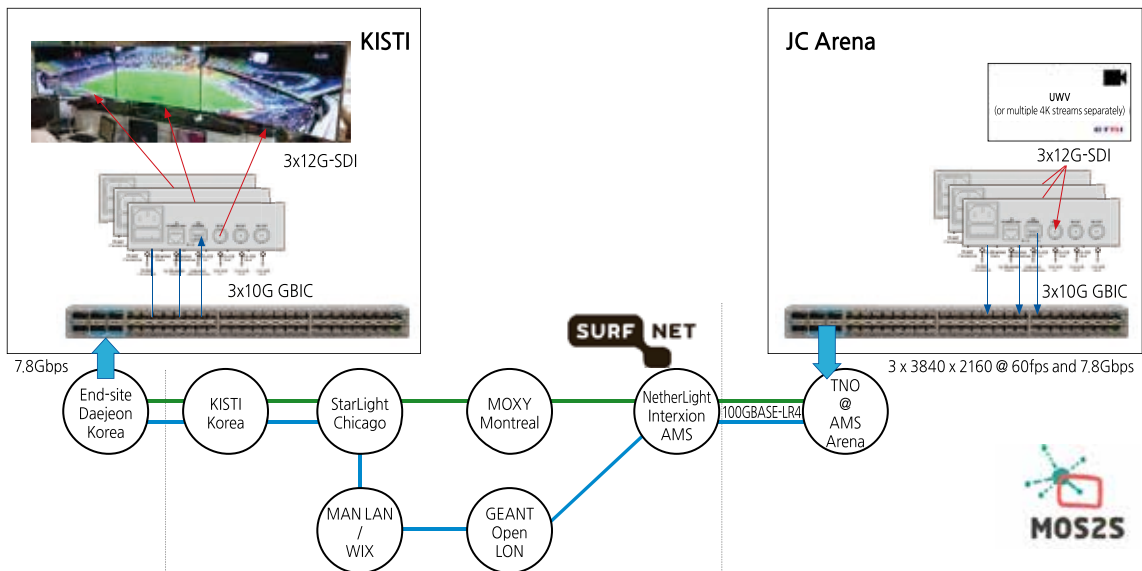
<표 1> 문화공연 및 스포츠 경기 환경 요약

	평창올림픽 문화공연	네덜란드 축구 국가대표팀 경기
촬영 환경	강릉아트센터 소공연장 	암스테르담 Johan Cruyff Arena 
재생 환경	인천공항 공항철도라운지 	대전 KISTI 멀티패널 
멀티카메라 구성	전동증 지원 멀티카메라 	수동식 멀티카메라 



[그림 4] 멀티카메라 촬영, 스티칭(생성) 및 중계 시스템 구조도





[그림 5] 네덜란드 암스테르담과 대한민국 대전 간의 연구망 활용

페이스를 통하여 손실 없이 저장, 전환 중계시스템으로 전달되어 카메라 원본 영상의 저장, 스티칭 서버로의 전송 및 최종 영상 출력이 이루어진다. 스티칭 서버로는 12Kx2K 영상 생성을 위한 UWV 스티칭 서버와 8Kx4K 360VR 영상 생성을 위한 8KVR 스티칭 서버를 동시에 구성하였다. 카메라 입력에서 스티칭 완료 시점까지 걸리는 시간 지연은 300ms를 넘지 않았다.

최종 출력 영상은 크게 부호화된 출력스트림과 무손실 원본스트림 두 가지 형태로 제공되었으며, 문화공연의 경우 KT의 협조를 통하여 평창올림픽 전용회선을 활용하여 강릉과 인천 간 전용회선을 사용하였다. 12Kx2K 영상을 원격지로 서비스하기 위하여 3채널의 12G-SDI 인터페이스로 전달되는 영상들을 3대의 4K급 부호화기를 이용하여 약 60Mbps 압축률로 부호화하였으며, 수신단에서는 3채널의 동기를 해결하기 위하여 하나의 복호화 서버에서 개별 4K채널을 복호화 한 후 동기를 조정하는 방식

을 사용하였다. 스포츠 경기의 경우 유럽공동연구과제의 일환으로 수행된 관계로 초고속 글로벌 연구망을 담당하는 대한민국의 KISTI(한국과학기술정보연구원)와 네덜란드의 SurfNet(surfnet.nl), KPN(Koninklijke PTT Nederland)의 협조를 받아 양국 간에 100Gbps 대역폭의 전용 VLAN 네트워크를 구성하였다. 최대한 무손실 스트림을 전달하기 위해 12G-SDI를 10G 광통신모듈로 전환하면서 무손실 부호화 도구인 TiCO(Tiny Codec)를 지원하는 장비들과 100Gbps 스위칭 장비를 사용하였으며, VLAN Tag를 구성하여 인접 트래픽에 영향을 받지 않도록 구성하였다. 10G 광통신 장비는 PTP(Precision Time Protocol)을 사용하여 양국 간의 시간 동기가 보장되도록 구성하였으며, 양국 간의 전송 지연이 250ms를 넘지 않는 초고품질 연구망임을 확인할 수 있었다.

<표 2> 고품질 광시야각 동영상을 이용한 실시간 중계 서비스 결과


문화공연	강릉			
	인천			
스포츠경기	네덜란드 Johan Cruyff Arena SkyBox		KISTI 멀티패널	
	8Kx4K 360VR 동영상			
				

#### 4.3 적용 결과

<표 2>에서는 최종 적용 결과물을 보여주고 있다. 문화공연의 경우 인천공항을 사용하는 고객들을 대상으로 실시간 공연 중계를 시연하였다. 스포츠 경기의 경우 네덜란드 축구협회, 미디어 관련 기술 관계자, SurfNet, KISTI 연구원을 대상으로 실증하였으며 실황 방송 신호와의 비교, 양국 간의 실시간 통화를 통한 검증, HMD 환경에서의 해상도 확인을 수행하였다.

#### 5. 맺음말

본 적용사례를 통하여 12Kx2K UWV, 8Kx4K 360VR 등 멀티카메라를 이용한 고품질 광시야각 동영상 서비스가 차세대 몰입형 서비스로서 충분한 부가가치를 가져올 수 있음을 확인할 수 있었고, 부호화 방법을 사용하는 경우 5G 네트워크가 보장할 수 있는 100Mbps 이내에서 충분히 서비스 가능함

을 실증하였다. 이 적용사례를 바탕으로 국내 프로 스포츠 경기 중계와 해외 문화공연 이벤트에 적용하기 위한 논의가 지속적으로 진행 중이다. 하지만, 소비자가 가용한 광시야각 단말의 저변확대가 부족하고, 서비스 사업자에 의존적인 방송, 중계 서비스라는 부분은 해결해야 할 문제이다. 이를 위해 지속적인 표준화, 표준특허 발굴 노력을 비롯하여 국내외 서비스 사업자와의 현장시험을 통한 서비스 실증을 수행할 필요가 있다. 

※ 본 연구는 산업통산자원부 국제공동연구협력사업의 지원을 받아 수행됨[8K급 360VR 동영상 적응형 라이브 시스템 개발, N0002304].

#### [참고문헌]

- [1] Intel True View, retrieved from <https://www.intel.com/content/www/us/en/sports/olympic-games/overview.html>
- [2] Rob Koenen, 'MPEG Roadmap', MPEG Workshop on Immersive Services Roadmap, Gwangju, 2018.
- [3] R. Ghaznavi-Youvalari et al., 'Comparison of HEVC coding schemes for tile-based viewport-adaptive streaming of omnidirectional video', 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Luton, 2017, pp. 1-6.
- [4] MORE(Media ORchEstration), ISO/IEC 23001-13:2018
- [5] NBMP vision paper, 2018, retrieved from <https://mpeg.chiariglione.org/standards/mpeg-i/network-based-media-processing/nbmp-vision-paper>