

[멀티미디어응용] 초다시점 3차원 동영상 서비스를 위한 MPEG 표준화 동향

FTV 3차원 비디오 국제 표준화 작업

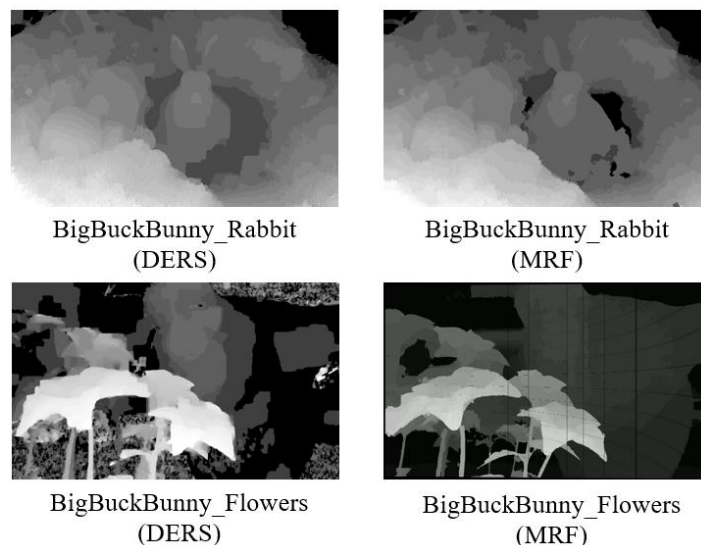
최근 방송장비의 발달과 실시간 영상 데이터 전송이 가능해짐에 따라 2차원 영상 콘텐츠보다 3차원 영상 콘텐츠 제작에 많은 관심이 쏟아지고 있다. 2차원 영상은 일반적인 카메라로 촬영해 획득한 영상을 의미하며, 3차원 영상은 2차원 영상에 깊이 정보를 더함으로써 시청자에게 입체감을 제공할 수 있다. 2차원 영상으로부터 입체감을 느끼기 위해서는 색상 영상 시점과 일치하는 깊이 정보가 필요하다. 하지만 색상카메라와 깊이 카메라의 배치 구조 문제와 기하학적 문제로 인해 정확한 깊이정보를 사용하여 3차원 영상을 생성하는데 한계가 있다. 이러한 문제를 해결하기 위해 초다시점(super multi-view, SMV)과 자유 탐색(free-navigation, FN) 카메라 시스템을 이용한 영상 표준화 작업이 MPEG회의에서 진행되고 있다.

초다시점 시스템은 이웃한 카메라들이 짧은 간격으로 배치되어 있으며, 이로 인해 자연스러운 중간시점 영상을 생성할 수 있다. 하지만 많은 수의 영상 데이터를 보내야 하기 때문에 전송 대역폭 문제를 갖는다. 자유 탐색 시스템은 이웃한 카메라들의 거리가 길며 초다시점 시스템에 비해 적은 수의 카메라를 사용하기 때문에 부자연스러운 중간 시점 영상을 생성하게 된다. 하지만 적은 수의 카메라를 사용하기 때문에 전송 대역폭의 문제는 발생하지 않는다. 초다시점 영상은 자유 탐색 영상에 비해 상대적으로 제한된 시점의 영상만을 제공하기 때문에 다양한 콘텐츠 제작에 제약이 있다. 최근 진행된 MPEG회의에서는 초다시점 영상에 대한 CfE(call for evidence)를 마쳤으며, 깊이정보 생성을 위해 기존에 사용하던 DERS 대신 새로운 깊이지도 생성 방법에 대한 논의가 이루어졌다. 차후 진행될 회의에서는 자유 탐색 영상의 객관적 평가를 위한 평가 방법과 알고리즘 개선 방안에 대해 논의하며, 영상 평면 기반 깊이 추정을 통해 정확한 가상시점 영상 생성 방법에 대한 논의를 계속할 예정이다.

중간시점 영상 합성을 위한 깊이지도 생성

초다시점 영상과 자유 탐색 영상을 생성하기 위해서는 카메라로 촬영한 영상이나 컴퓨터 그래픽을 통해 생성된 영상을 사용한다. 시청자에게 다양한 시점의 영상을 자유롭게 시청하게

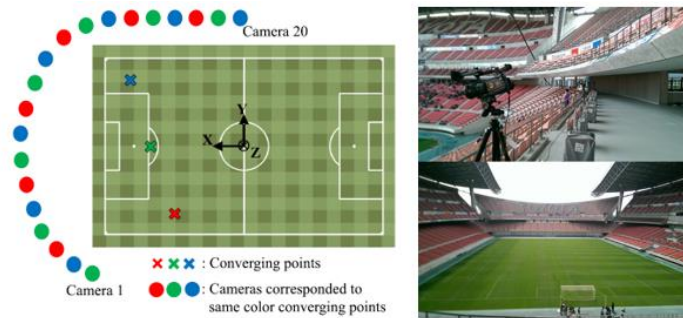
하기 위해서는 많은 수의 중간시점 영상을 생성할 필요가 있는데, 중간시점의 영상은 깊이 지도의 정확도에 따라 생성된 영상의 화질이 달라진다. 카메라를 통해 촬영할 경우 각각의 색상 카메라 시점에 해당하는 깊이영상 카메라를 할당하기에는 물리적 제약이 존재한다. 이러한 경우에는 좌, 우영상 기반 스테레오 정합을 사용하여 깊이 지도를 생성할 수 있다. MPEG FTV그룹에서 제작한 깊이 지도 추정소프트웨어(Depth Estimation Reference Software, DERS)를 사용하여 깊이 지도를 생성할 수 있지만, 입력 매개변수값에 따라 생성되는 깊이 지도의 정확성이 변하게 되는 문제가 있다. DERS를 사용하는 대신 스테레오 정합 방법을 사용하여 매개변수에 독립적인 깊이 지도를 생성하는 방법에 대한 논의가 이루어졌다. 스테레오 정합을 위해 마코브 랜덤 필드(Markov Random Field, MRF)를 통해 스테레오 영상의 정합 비용을 계산한다. MRF는 전역기반 매칭 방법이므로 최종 깊이 지도를 생성하기 위해서 신뢰확산(Belief Propagation, BP) 혹은 그래프 절단(Graph Cut)과 같은 최적화를 수행한다. <그림 1>은 DERS를 사용하여 생성한 깊이 지도와 MRF기반 스테레오 정합을 수행하여 획득한 깊이 지도를 나타낸다. 깊이 지도 결과 영상에서 알 수 있듯이 기존 DERS를 통해 획득한 깊이 지도 보다 높은 정확도를 갖는 깊이 지도를 스테레오 정합을 통해 획득했다. 자연스운 초다시점 영상을 생성하기 위해서는 MRF뿐만 아니라 다양한 스테레오 정합 방법을 사용한 깊이 지도 획득 방법 관련 연구가 지속적으로 진행되어야 한다.



<그림 1> MRF 를 이용한 정확한 깊이 지도 생성 결과(출처: ISO/IEC JTC1/SC29/WG11

자유 탐색 영상 연구 동향과 적용분야

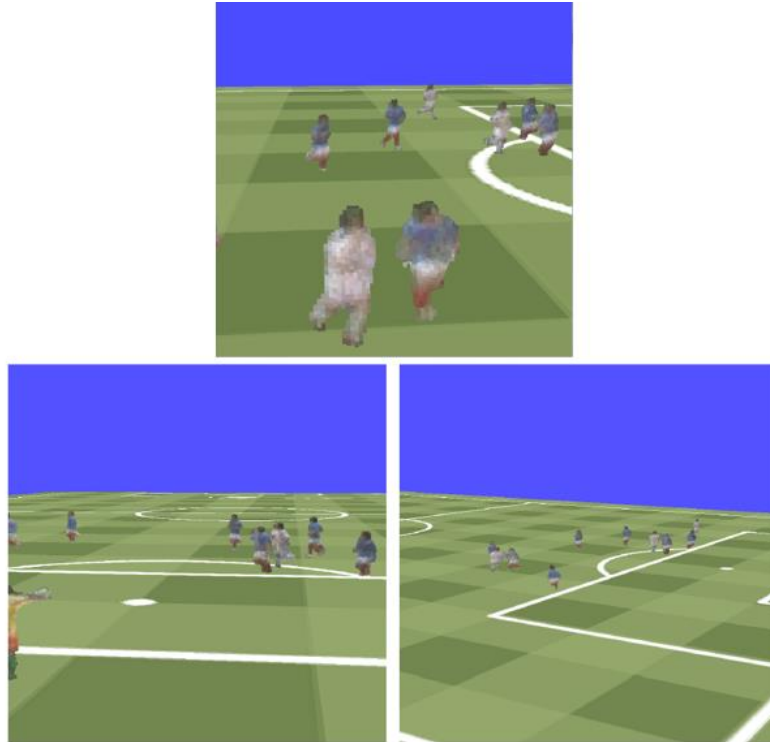
기존 자유 탐색 영상의 한계점을 해결하기 위해 다양한 해결 방안이 최근 MPEG미팅에서 논의되었다. 최근 제안된 방법 중, 자유 탐색 영상 시스템의 구조를 일부 변경하며 가상 시점 영상을 렌더링하여 생성하는 기술이 소개되었다. 이 방법은 자유 탐색 영상을 획득하기 위해 기존 카메라 시스템 보다 많은 수의 카메라를 사용하여 다양한 시점에서의 영상을 획득한다. <그림 2>는 자유 탐색 영상의 한계를 극복하기 위해 새롭게 제안된 카메라 배열 구조를 나타낸다.



<그림 2> 가상시점 렌더링을 위한 자유 탐색 영상 카메라 배열(출처: ISO/IEC JTC1/SC29/WG11 MPEG2016/M37985)

제안된 카메라 구조는 축구장이나, 농구장과 같이 넓은 영역에서 촬영된 영상을 사용하여 생성한 가상 시점 영상을 시청자에게 제공하기 적합한 방법이다. 각각의 카메라는 경기장의 일부 영역을 전담하여 촬영함으로써 다양한 시점의 영상을 획득할 수 있다. 영상 촬영에 사용된 카메라 캘리브레이션을 수행하기 위해 내부 파라미터는 일반적으로 사용하는 체스보드 패턴을 사용하였으며, 외부 파라미터는 경기장 내부에 있는 경계선들을 특징점으로 사용하여 획득한다. 자유 탐색 영상 시스템을 사용하여 획득한 영상을 사용하여 실제와 같은 새로운 시점의 영상을 생성할 수 있지만, 컴퓨터 그래픽을 사용하여 다양한 가상 시점의 영상을 생성하는 방법을 적용시켰다. 촬영 영상을 사용하여 실제 영상과 유사한 가상시점 영상을 생성할 수 있지만, 컴퓨터 그래픽을 사용하여 렌더링을 수행할 경우 정확하지는 않지만 실제와 매우 유사한 구조를

갖는 영상을 제공함으로써 더 높은 시점의 자유도를 갖게 된다. <그림 3>은 20대의 카메라를 통해 획득한 영상을 기반으로 렌더링한 일부 결과를 보이고 있다.



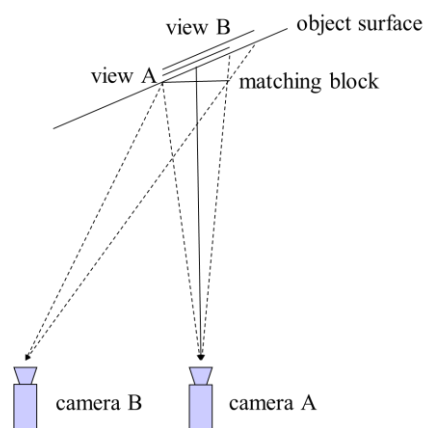
<그림 3> 가상시점 영상 렌더링 결과 (출처: ISO/IEC JTC1/SC29/WG11 MPEG2016/M37985)

카메라로 촬영한 영상과 같이 객체가 정확하게 표현되지는 못하지만 촬영 당시 상황과 구조를 정확하게 표현할 수 있다는 점으로 미루어보아 움직임이 많은 실외 운동 경기 촬영에 가장 적합한 것을 알 수 있다. 이 기술은 자유 탐색 영상 시스템의 한계점을 극복하기에 적합한 카메라 구조를 사용하고 있으며, 렌더링을 사용하여 최종 가상 시점 영상을 제공한다는 점에서 새로운 가상 시점 영상 제공 방법으로 높은 가능성을 지니고 있다. 이와 같이 MPEG FTV 표준화 회의에서 진행되는 내용을 통해 앞으로 개최될 다양한 스포츠 경기를 실감나고 현장감 있게 즐길 수 있을 것으로 예상된다.

효율적인 깊이 지도 생성을 위한 새로운 FTV 시스템

현재 FTV는 기본적으로 다시점 영상과 깊이 영상 (Multiview plus depth, MVD)을 사용하여

3차원 장면을 생성하며, 가상 시점 영상 또한 깊이 정보를 기반으로 3차원 워핑을 수행함으로써 만들어진다. 이러한 영상들은 필수적으로 깊이 정보를 사용하기 때문에 DERS (Depth Estimation Reference Software)와 VSRS(View Synthesis Reference Software)를 통해 정확한 깊이 영상을 생성하고 워핑을 수행한다. DRES의 경우 스테레오 입력 영상을 사용하며, 블록 기반 스테레오 매칭을 통해 깊이 지도를 생성한다. 하지만 <그림 4>와 같이 다른 시점에서 촬영된 영상 내부에 있는 객체의 기울기 정보가 다를 경우, 스테레오 매칭을 통해 정확한 깊이 지도를 생성할 수 없는 문제가 발생한다.

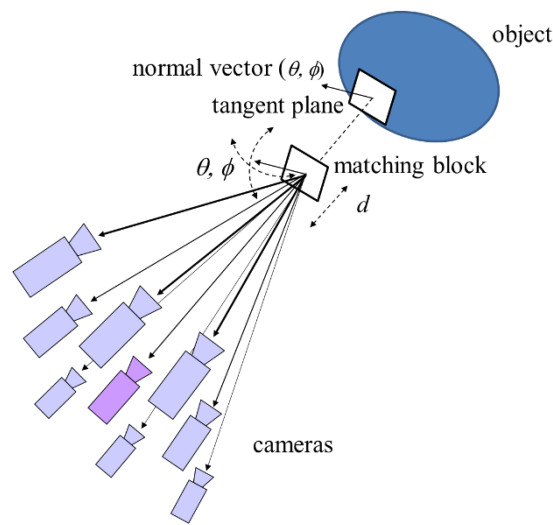


<그림 4> 객체의 기울기가 있는 영상에서의 매칭 영역(출처: ISO/IEC JTC1/SC29/WG11 MPEG2016/M38547)

<그림 4>와 같은 상황의 스테레오 매칭 과정에서 발생하는 문제를 해결하기 위해, FTV그룹에서는 깊이 정보와 함께 평면(surface)의 정보를 고려한 MVS (Multiview plus surface)기반 FTV 시스템에 대한 논의가 진행되고있다. MVS는 크게 3가지 요소로 구성되는데, 기존에 사용된 깊이정보(d)와 객체 평면의 방향성을 나타내는 법선 벡터 정보(θ, ϕ)를 통해 정확한 깊이 정보를 생성한다. 일반적으로 색상 정보는 밝기(Luminance: Y)와 색상(Chrominance: U, V)으로 표현되는데, 평면의 정보도 이와 유사하게 깊이(Depth: d)와 방향성(Direction: θ, ϕ)으로 구성된 요소들을 통해 나타낼 수 있다. 깊이 정보와 함께 방향성 정보를 사용함으로써 다른 시점에서 촬영된 영상 기반 스테레오 매칭을 수행할 때 발생하는 블록 영역 부정합 문제를

해결할 수 있다. 이 과정에서 발생하는 블록 영역 부정합은 실제 촬영된 객체와 카메라가 사이에 존재하는 경사도(slant) 때문에 발생한다.

MVS구조를 통해 다양한 시점에서 촬영한 것과 유사한 영상을 여러 깊이 정보와 이에 해당하는 방향성 정보를 사용함으로써 얻을 수 있다. 깊이 정보와 방향성 정보를 혼합하여 얻은 다양한 시점의 영상을 기반으로 블록 매칭을 수행함으로써 가장 정확한 깊이 정보를 획득하게 되는데, 이 시스템 구조는 <그림 5>에 나타나 있다.



<그림 5> 객체의 기울기가 있는 영상에서의 매칭 영역 (출처: ISO/IEC JTC1/SC29/WG11 MPEG2016/M38547)

새롭게 제안한 MVS구조 기반 3차원 영상은 MVD구조에서 생성하는 깊이 정보보다 정확하기 때문에 MVD를 통해 획득한 3차원 영상 또는 가상 시점 영상보다 높은 정확도를 갖게 된다. 하지만 깊이와 방향 정보에 따라 블록 매칭을 수행함으로써 시간 복잡도가 증가하게 되고, 객체와 카메라간의 경사도를 정확하게 측정할 수 있는 방법은 제안되지 않았다. 정확한 깊이 정보를 이용하여 3차원 영상 콘텐츠를 제작하기 위해 MPEG-FTV 그룹은 MVS시스템에 대한 다양한 내용을 추후 MPEG미팅에서 계속 논의할 예정이다.

호요성 (광주과학기술원 실감방송연구센터 교수, hoyo@gist.ac.kr)