

+ 호요성, 이상범 · 광주과학기술원 실감방송연구센터

TECH&TREND

3차원 비디오 부호화를 위한 MPEG 국제 표준화 작업

1. 서론

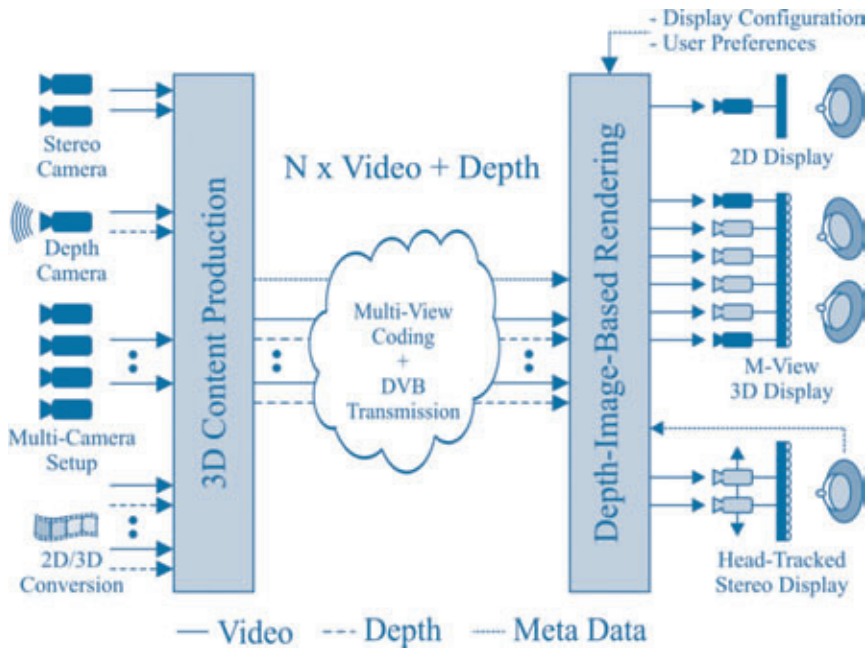
“3D가 디지털 미디어의 르네상스를 만들어냈다. 3D는 엔터테인먼트의 획기적 사건이자 전 세계적 변화이다.” 이는 3차원 영화 “아바타”로 전 세계적인 관심을 이끌어 낸 제임스 카메론 감독이 제7회 서울 디지털 포럼에서 발표한 내용이다[1]. 최근 폭발적인 증가세를 보이고 있는 3차원 시장의 점화제 역할을 한 제임스 카메론 감독의 이런 발언은, 방송 시스템이 아날로그에서 디지털로 전환되면서 영상 산업 전반에 큰 변화를 가져 왔듯이, 가까운 미래에 디지털 미디어가 2차원에서 3차원으로 전환되는 영상 산업의 또 다른 혁명을 불러일으킬 것이라는 전망과 일치한다[2]. 실제로, 세계 선진국에서는 3차원 방송을 위한 3차원 영상 콘텐츠를 제작하고 있으며, 국내에서도 여러 방송사업자들을 중심으로 3D 시험방송을 준비하고 있다[3]~[6].

이러한 산업 동향과 더불어, 최근 MPEG(moving picture experts group) 국제 표준화 기구에서는 3차원 비디오 시스템을 정의하고, 입력 데이터를 효과적으로 부호화하는 방법을 개발하고 있다[7]~[9]. MPEG에서 정의한 3차원 비디오 시스템은 기본적으로 양안식(stereoscopic) 영상을 포함하고, 3시점 혹은 그 이상의 넓은 시야각을 제공하는 고해상도의 3차원 비디오 시스템을 말한다. 3차원 비디오 시스템은 다수의 카메라로 획득한 넓은 시야각의 다시점 영상과 장면의 3차원 거리 정보를 표현하는 깊이영상이 기본 입력 데이터로 취한다. 이러한 시스템을 구현하기 위해서는 각 시점의 깊이영상을 획득하는 깊이영상 추정기술과 깊이영상을 이용하여 사용자가 원하는 임의의 시점에서 장면을 생성하는 중간시점 영상합성 기술이 필요하다. 본 논문에서는 3차원 비디오 부호화를 위한 MPEG 국제 표준화 작업을 기술한다.

2. 3차원 비디오 시스템

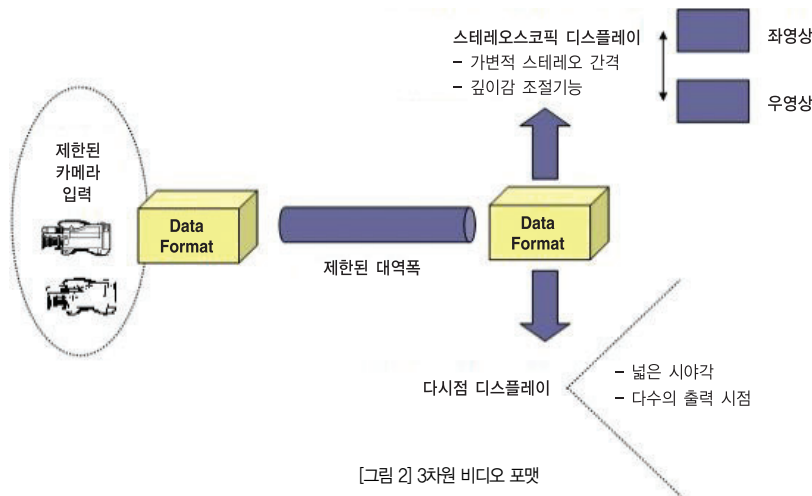
2-1. 3차원 비디오 시스템 개요

[그림 1]은 3차원 비디오 시스템의 개념도를 나타낸 것이다[9]. 이 시스템은 N시점의 다시점 색상영상과 이에 상응하는 깊이영상을 이용한 3차원 비디오의 획득/처리/전송/재생에 대한 모든 과정을 포함한다. 3차원 비디오는 스테레오 카메라나 다시점 카메라를 이용하여 직접 얻을 수 있다. 깊이 카메라를 이용하면 깊이영상을 직접 획득할 수 있지만, 스테레오/다시점 카메라의 경우와 기존에 제작된 2차원 영상의 경우에는 깊이영상을 직접 구할 수 없기 때문에 촬영된 장면으로부터 3차원 깊이 정보를 추정하여 깊이영상을 획득한다.



[그림 1] 3차원 비디오 시스템의 개념도

수신단으로 전송된 색상영상과 깊이영상은 디스플레이 장치에 따라 여러 가지 방법으로 재생될 수 있다. 기본적으로 기존의 3차원 디스플레이 장치에서 재생이 가능해야 하므로, 전송된 N시점 중에서 정해진(혹은 선택 가능한 시점) 시점의 영상을 시청할 수 있다. 또한, 스테레오 디스플레이 혹은 무안경식 3차원 디스플레이 등 다양한 종류의 3차원 디스플레이 장치를 이용하여 재생할 수 있다. 만약, 디스플레이 장치에서 필요한 시점의 수가 전송된 시점의 수보다 클 경우에는 필요한 시점만큼의 중간시점 영상을 합성해야 한다. 이러한 기능을 시점 합성(view synthesis)이라 한다.



[그림 2] 3차원 비디오 포맷

2-2. 3차원 비디오에 대한 비전

2009년 2월에 열린 MPEG 표준화 회의에서는 3차원 비디오에 대한 비전이 제시되었다[8]. 이 비전에서는 향후 제정될 표준 기술이 기존의 스테레오스코픽 디스플레이 장치에 호환될 뿐만 아니라, 향후 개발될 다시점 입체 디스플레이 장치에도 호환성을 갖도록 하는 기술이 언급됐다. [그림 2]에서 알 수 있듯이, 3차원 비디오 포맷은 제한된 수의 카메라와 제한된 대역폭의 한계를 뛰어 넘어 스테레오 디스플레이와 다시점 디스플레이에 모두 사용될 수 있어야 한다.

3. 깊이영상 추정 기술

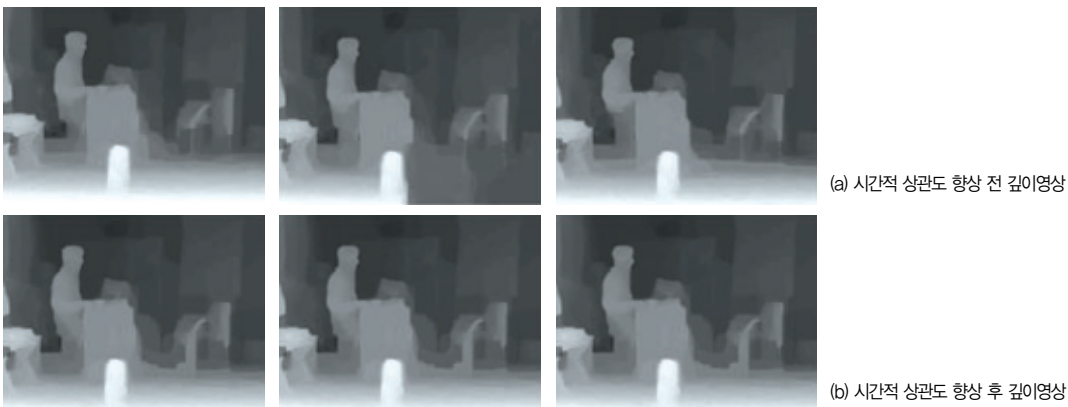
깊이영상(depth image 혹은 depth map)이란 영상 내에 존재하는 객체의 3차원 거리 정보를 8비트로 표현한 영상을 말하며, 깊이영상의 화소 값은 각 해당 화소의 깊이 정보를 나타낸다. 깊이영상은 깊이 카메라를 이용해 직접 획득할 수 있지만, 스테레오 카메라 및 다시점 카메라를 이용하는 경우에는 깊이영상을 계산적으로 추정해서 얻을 수 있다. 보통 깊이영상을 얻기 위해서는 이웃하는 영상의 상관관계를 이용하여 변이(disparity) 정보를 탐색하는 스테레오 정합 기술이 가장 많이 이용된다. 스테레오 정합 기술은 인접한 두 시점 영상에서 객체가 수평으로 이동한 정도, 즉, 변이를 계산해서 깊이 정보를 획득하는 방법이다[10]. 이 방법은 특별한 센서 없이도 깊이 정보를 획득할 수 있기 때문에 비용이 적게 들고, 이미 촬영된 영상에 대해서도 깊이 정보를 획득할 수 있다는 장점을 지닌다.

MPEG 3차원 비디오 부호화 그룹에서는 3차원 비디오 부호화 기술의 국제 표준을 만들기 위해 깊이영상을 추정하는 소프트웨어를 개발했다. 깊이 추정 참조 소프트웨어(Depth Estimation Reference Software, DERS)는 2008년 4월에 배포되었고, 이후 여러 실험을 통해 다양한 기술들이 소프트웨어에 추가적으로 구현되었다. 현재 DERS 버전 5.1이 배포된 상황이며, 이 소프트웨어는 MPEG 3차원 비디오 부호화 그룹에서 관리하는 SVN 서버에서 다운로드 받을 수 있다[11]. 본 장에서는 해당 소프트웨어에 구현된 기술들 가운데 최근 중요하게 부각된 기술들을 소개한다.

3-1. 깊이영상의 시간적 상관도 향상 기술

깊이영상의 시간적 상관도 저하 문제는 DERS가 매 화면(frame)마다 독립적으로 깊이영상을 탐색하면서 발생하는 문제를 말한다. 움직임이 작은 객체의 경우에 매 화면마다 비슷한 깊이 값을 가져야 하지만, 깊이영상 탐색 방법의 오차로 인해 화면마다 깊이 값이 약간씩 달라질 수 있다. 이러한 깊이영상을 이용하여 중간시점 영상을 합성했을 때에는 물체의 배경 부분이 조금씩 바뀌는 깜빡임(flickering) 현상이 발생하여 시청자에게 시각적 불편함을 준다. 또한, 깊이영상 부호화에서 시간적 예측 효과를 저하시키기 때문에 부호화 효율이 많이 떨어진다.

최근 깊이영상의 시간적 상관도 저하 문제를 해결하기 위해, 이전 화면에서의 깊이 정보를 참조하는 새로운 정합 함수를 이용하는 방법이 제안되었다[12]. 새로운 정합 함수에는 이전 화면에서의 깊이 정보를 사용하는 가중치 함수가 추가된다. [그림 3]에서 알 수 있듯이, 처음에는 동일한 장면임에도 불구하고 배경 부분의 깊이 값이 약간 다르지만, 시간적 상관도를 향상시키는 기술을 적용한 후에는 깊이 값이 거의 비슷한 것을 확인할 수 있다.



[그림 3] 깊이영상의 시간적 상관도 향상

3-2. 깊이영상의 시점간 상관도 향상 기술

깊이영상의 시점간 상관도 저하 문제는 다시점 영상의 각 시점마다 독립적으로 깊이영상을 탐색하는 과정에서 발생한다. 이론적으로 다시점 영상의 여러 시점에 존재하는 객체는 동일한 깊이 값을 가져야 하지만, 깊이영상 탐색 방법의 오차로 인해 각 시점마다 깊이 값이 달라질 수 있다. 이러한 차이는 영상합성 과정의 오차로 그대로 전파될 뿐만 아니라, 다시점 영상 부호화에서 시점간 예측의 성능을 저하시킨다.

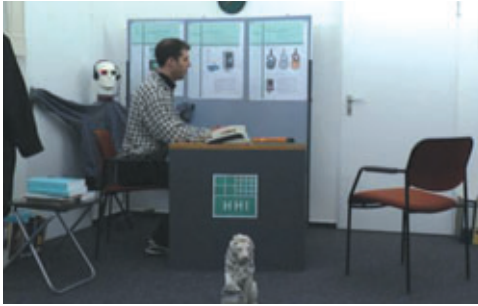
최근 깊이영상의 시점간 오차를 줄이기 위해, 깊이영상을 탐색할 때 인접한 시점의 깊이 값을 참조하여 시점간 상관도를 향상시키는 기술이 제안되었다[13]. 이 방법은 목표 시점의 깊이영상을 탐색하는 과정에서 이미 구한 참조 시점의 깊이영상을 고려한 새로운 정합 함수를 사용한다. [그림 4]에서 알 수 있듯이, 동일한 객체임에도 불구하고 각 시점마다 구한 깊이 값이 서로 다르지만, 시점간 상관도 향상 기술을 적용한 후에는 깊이 값이 거의 비슷해짐을 확인할 수 있다.



[그림 4] 시점간 상관도가 향상된 깊이영상

3-3. 반자동 깊이영상 추정 기술

반자동 깊이영상 추정 기술은 자동 추정 기술을 수행할 때 외부에서 수동으로 구한 특정 화면의 부가적인 정보들을 이용해 깊이영상의 정확도를 높이는 기술이다. 최근 한국전자통신연구원(ETRI)과 일본 Nagoya 대학에서는 반자동으로 깊이영상을 추정하는 기술을 제안했다[14][15]. Nagoya 대학에서 제안한 반자동 깊이영상 추정 기술은 [그림 5]와 같이 수동 깊이영상, 수동 경계영상, 수동 정지영상을 이용한다. 수동 깊이영상은 해당 화면의 특정한 영역에서 미리 계산된 깊이 값을 알려 주고, 수동 경계영상은 객체의 경계를 표시한다. 수동 정지영상은 구하고자 하는 영상 전체 화면에 대해서 움직임이 없는 영역을 나타낸다. 이러한 세 가지 수동 영상은 모두 사용자에게 의해 수작업으로 얻어진 영상이다.



(a) 색상영상



(b) 수동 깊이영상



(c) 수동 경계영상



(d) 수동 정지영상

[그림 5] 반자동 깊이 추정 기술을 위한 수동 영상

국내의 전자통신연구원(ETRI)에서 제안한 반자동 깊이영상 추정 기술은 수동으로 구한 깊이영상을 단서 화면(key frame)으로 정하고, 다음 화면부터 블록 단위의 움직임 예측을 통해 움직임이 있는 영역과 그렇지 않은 영역을 분류한 다음, 움직임이 없는 영역에 대해 이전 화면의 깊이 값을 이용한다. [그림 6]은 블록 단위의 움직임이 있는 영역을 검출한 결과를 나타낸 것이다. 반자동 깊이영상 추정 기술의 전체적인 성능은 수동 영상의 정확도에 따라 크게 좌우되기 때문에, 깊이영상의 정확도를 향상시키기 위해서는 수동 영상의 정확도를 높이는 작업이 매우 중요하다.



(a) 색상영상



(b) 움직임이 없는 영역

[그림 6] 움직임이 없는 영역 검출



4. 3차원 비디오 부호화 기술

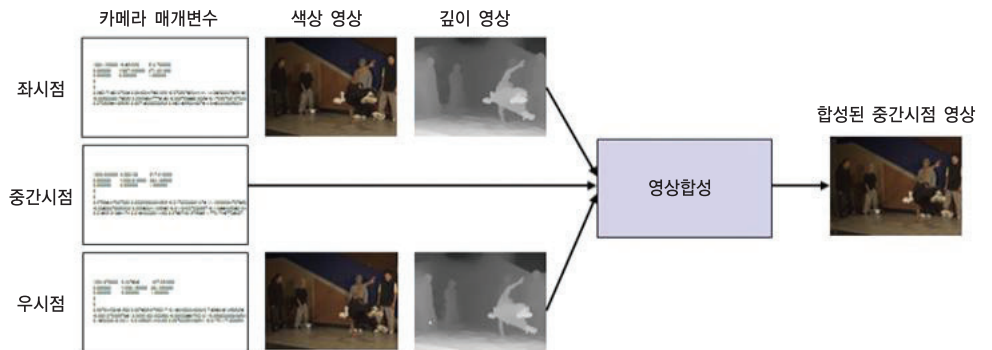
다시점 색상영상과 깊이영상을 포함한 3차원 비디오는 카메라의 수에 비례하여 대용량의 정보를 가지는데, 이러한 대용량의 3차원 정보를 효과적으로 전송하거나 저장하기 위해서는 효율적인 부호화 과정이 필요하다. MPEG에서는 3차원 비디오 부호화의 국제 표준을 제정하기 위해 2008년 1월 회의에서 3차원 비디오 부호화를 위한 테스트 영상을 요청하여, 지금까지 여러 번의 선행 실험과 분석을 통해 3차원 비디오 부호화를 위한 실험 환경을 구축하고 있다. 오는 2010년 10월에는 3차원 비디오를 효과적으로 부호화하는 방법에 대한 제안요청서(Call for Proposal)를 배포할 예정이다[16].

효과적인 3차원 비디오 부호화를 위해서는 이미 표준화가 완료된 다시점 비디오 부호화 표준인 MVC(Multiview Video Coding)를 이용할 수 있으나, MVC는 색상영상의 특징을 고려해서 개발된 H.264/AVC 기술을 기반으로 하기 때문에 깊이영상의 부호화에는 최적화 되어 있지 않다. 그러므로 깊이영상을 효과적으로 부호화할 수 있는 기술이 요구된다. 깊이영상은 색상영상에 비해 값의 변화가 작고, 객체의 경계에서 불연속성이 두드러진다. 또한, 깊이 값의 오차는 합성영상의 화질에 큰 영향을 미치기 때문에, 깊이영상의 특성을 이용한 효율적인 부호화 기술이 필요하다.

최근 발표된 깊이영상 부호화 기술들은 이러한 깊이영상의 특성을 이용한다. 그 예로, 깊이영상에서 객체의 경계에 대한 효과적인 부호화를 위해 기하학적 블록 분할을 이용한 깊이영상 부호화 기술이 제안됐고[17], 기존의 부호화 방법에 사용됐던 보간 필터의 문제점을 보완하기 위해 미디언 필터를 이용하는 기술이 제안됐다[18]. MPEG에서 3차원 비디오 부호화 국제 표준을 만들기 위한 제안요청서가 배포되면 많은 연구기관에서 다양한 부호화 기술을 제안할 것으로 예상된다.

5. 중간시점 영상합성 기술

앞서 언급한 것처럼, 다양한 비디오 재현 장치에 여러 시점의 3차원 영상을 만들어 보여주기 위해서는 임의 시점의 영상을 합성하는 기능을 포함해야 한다. 중간시점 영상합성 기술이란 다시점 색상영상과 깊이영상 및 카메라 매개변수를 이용하여 시청자가 원하는 임의의 가상시점에 대한 영상을 생성하는 기술을 말한다.

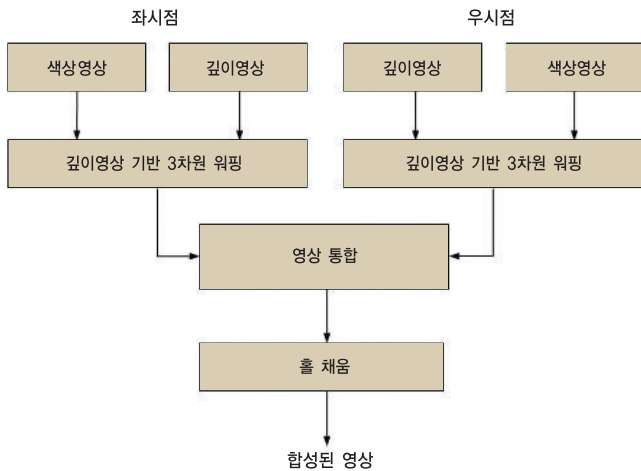


[그림 7] 중간시점 영상합성에 대한 개념도

MPEG 3차원 비디오 부호화 그룹에서는 깊이영상 추정 기술과 함께 중간시점 영상합성 기술을 개발했다. 각 기관에서 제안된 기술을 종합하여 중간시점 영상합성 참조 소프트웨어(View Synthesis Reference Software, VSRS)를 배포하고 있다. 현재 VSRS 버전 3.5가 배포된 상황이며, 이 소프트웨어는 MPEG 3차원 비디오 부호화 그룹에서 관리하는 SVN 서버에서 다운로드 받을 수 있다[19]. 본 장에서는 해당 소프트웨어에 구현된 중요한 기술을 소개한다.

5-1. 중간시점 영상합성 과정

중간시점 영상합성 과정은 [그림 8]과 같이 크게 3단계로 나눌 수 있다. 첫 번째는 3차원 워핑(3D warping) 동작이다. 3차원 워핑이란 깊이영상과 카메라 매개변수를 이용하여 색상영상을 3차원 공간상에 역투영한 후, 이를 다시 목표 시점에 투영하는 과정을 말한다[20]. 이러한 과정에서 시점 이동으로 인해 참조 시점에서 존재하지 않았던 영역은 홀(hole)로 나타나는데, 이러한 빈 곳은 좌우 참조 화면으로부터 중간시점에 워핑된 두 영상을 하나로 합치는 영상 통합 기술을 이용하여 대부분 채워진다. 마지막으로 영상 통합 과정에서 남아 있는 빈 곳을 영상 보간법이나 인페인팅(inpainting) 기술을 이용하여 채운다.



[그림 8] 중간시점 영상합성 과정

5-2. 3차원 워핑을 이용한 영상합성 기술

중간시점 영상을 합성하기 위해서는 3차원 워핑 기술을 이용한다. 하지만, 기하학적인 구조의 차이로 인해 시점 이동 과정이 정확하게 일대일 사상(mapping)이 되지 않기 때문에, 홀(hole)이 발생하게 된다. 다시점 영상을 이용하는 중간시점 영상합성 과정에서는 가상시점을 기준으로 가장 가까운 영상을 이용하여 홀을 채운다. [그림 9(a)]와 같이 좌영상만 이용하여 중간영상을 합성할 경우에 홀이 객체의 왼쪽 영역에 존재하고, [그림 9(b)]와 같이 우영상만 이용할 경우에는 홀이 객체의 오른쪽 영역에 존재하게 된다. 따라서, 일반적으로는 [그림 9(c)]와 같이 가상시점 주위의 좌우영상을 모두 이용하여 홀을 채운다.



(a) 좌영상만 이용



(b) 우영상만 이용



(c) 양쪽시점 모두 이용

[그림 9] 참조시점에 따른 중간시점 영상합성 결과

5-3. 합성영상의 경계 잡음 제거 기술

깊이영상은 중간시점 영상합성을 위해 필수적이지만, 깊이영상의 오차는 합성된 영상에서 눈에 거슬리는 잡음으로 나타난다. 특히, 색상영상과 깊이영상의 객체의 경계가 일치하지 않는 경우에는 합성된 영상에서 배경에 전경의 색상이 포함되는 경계 잡음이 발생한다. 최근 합성된 중간시점 영상의 경계 잡음을 제거하기 위해, 깊이영상을 이용한 영상합성 방법을 통해 경계 잡음을 제거하는 새로운 기술이 제안되었다[21]. 이 기술에서는 객체의 경계에 잡음이 발생하는 영역을 미리 선택한 다음, 해당 영역에 대해서 참조 시점의 배경이 사상되도록 한다.



(a) 경계 잡음 제거 기술 적용 전 합성영상



(b) 경계 잡음 제거 기술 적용 후 합성영상

[그림 10] 합성영상의 경계 잡음 제거



6. 결론

본 논문에서는 최근 큰 이슈가 되고 있는 3차원 비디오 압축 부호화의 국제 표준화 동향 및 주요 기술을 살펴보았다. 최근까지 깊이영상 추정 기술과 중간시점 영상합성 기술이 주로 논의되었으며, 앞으로는 다시점 색상영상과 깊이영상의 부호화 기술이 본격적으로 논의될 예정이다. 이러한 기술은 최근 급격히 증가하고 있는 3차원 방송 및 입체영상 산업에 커다란 촉진제 역할을 할 것으로 기대된다. 따라서, 3차원 비디오 부호화에 대한 국내 연구기관과 산업체의 지속적인 연구와 정부 차원의 적극적인 지원이 뒷받침 된다면, 우리나라가 3차원 방송의 핵심 기술을 확보하여 차세대 실감미디어 산업을 주도해 나갈 것이다.

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었다.(NIPA-2010-(C1090-1011-0003))

참고 문헌

- [1] "제임스캐머런, 상상력 기술의 新 르네상스", <http://www.mt.co.kr/view/mtview.php?type=1&no=2010051308483641428&outlink=1>.
- [2] 호요성, 오관정, "다시점 비디오 기반의 3차원 실감방송", 방송과 기술, 제161권, pp.136-143, 2010. 05.
- [3] "방통위, 3차원방송 등 집중 육성키로", http://www.hani.co.kr/arti/economy/economy_general/419977.html.
- [4] "스카이라이프 3D영화 방영", http://www.fnnews.com/view?ra=Sent0901m_View&corp=fnnews&arcid=0921987669&cDateYear=2010&cDateMonth=05&cDateDay=17.
- [5] "[또 하나의 세계, 실감미디어] <1부-9> 3D 방송서비스", <http://www.etnews.co.kr/news/detail.html?id=201003170020>.
- [6] "[글로벌리포트] 영국 디지털방송 시장", <http://www.etnews.co.kr/news/detail.html?id=201005140029>.
- [7] 호요성, 이상범, "3차원 비디오 부호화 기술의 국제 표준화 동향", TTA Journal, No. 123, pp.77-83, 2009. 06.
- [8] ISO/IEC JTC1/SC29/WG11, "Vision on 3D Video," N10357, Feb. 2009.
- [9] ISO/IEC JTC1/SC29/WG11, "Applications and Requirements on FTV," N9466, Oct. 2007.
- [10] D. Sharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms", Proc. of IEEE Workshop on Stereo and Multi-Baseline Vision, pp.131-140, April 2001.
- [11] "Depth Estimation Reference Software", http://wg11.sc29.org/svn/repos/MPEG-4/test/trunk/3D/depth_estimation/DEFS/DEFS.
- [12] ISO/IEC JTC1/SC29/WG11 "Experimental Results on Improved Temporal Consistency Enhancement", M16063, Feb. 2008.
- [13] ISO/IEC JTC1/SC29/WG11 "Depth Estimation Reference Software(DEFS) 5.0", M16923, Oct. 2009.
- [14] ISO/IEC JTC1/SC29/WG11, "Semi-automatic Depth Estimation for FTV", M16391, April 2009.
- [15] ISO/IEC JTC1/SC29/WG11, "Depth Estimation algorithm in SADERS1.0", M16411, April 2009.
- [16] 호요성, 이천, "MVC 다시점 비디오 부호화 기술", 방송과 기술, 2010. 06.
- [17] M. Kang, J. Lee, J. Y. Lee, and Y. Ho, "Geometry-based Block Partitioning for Efficient Intra Prediction in Depth Video Coding," Proc. of the SPIE, vol.7543, pp.75430A-75430A-11, Jan. 2010.
- [18] 이천, 허진, 호요성, "미디어 필터를 이용한 깊이영상 부호화의 인터라 예측모드 설계", 한국통신학회 추계종합학술대회, pp.29(1-2), 2009. 11.
- [19] "View Synthesis Reference Software", http://wg11.sc29.org/svn/repos/MPEG-4/test/trunk/3D/view_synthesis/VSRs.
- [20] W. R. Mark, L. McMillan, G. Bishop, "Post-Rendering 3D Warping", Proc. of Symposium on Interactive 3D Graphics, pp.7-16, April 1997.
- [21] C. Lee and Y. Ho, "Boundary Filtering on Synthesized Views of 3D Video", in Proc. of Signal processing, Image processing and Pattern Recognition (SIP) 2008, pp.15(1-4), Dec. 2008.