

■■■특집■■■

자유시점 3차원 영상서비스를 위한 가상시점의 영상생성 기술

호 요 성, 이 천 (광주과학기술원)

I. 서론

최근 TV 시장이 급변하고 있다. PDP, LCD TV가 안방의 브라운관 TV를 내몰더니, 이제는 3DTV로의 세대교체가 이루어지고 있다. 실제로 일반 LCD TV와 3DTV의 가격차이가 크지 않아 이왕이면 3DTV라는 인식이 커져가고 있다. 또한, 각 영화 상영관에서도 3차원 입체 상영관을 하나 이상 확보하여 3차원 입체 영화를 줄이어 개봉하고 있다. 그야말로 3D 시대라고 해도 과언이 아니다. 이러한 시장의 변화에 발맞추어 각 방송사는 3차원 TV 기술에 많은 투자를 하고 있다. 또한 케이

블 및 위성TV 방송사들은 3차원 방송 서비스가 가능한 채널을 확보하여 실험방송을 하거나 이미 방송을 시행하고 있다. 각 가전업체들도 3차원 TV를 적극 생산하여 소비자의 관심을 이끌고 있다.

이러한 3차원 영상에 대한 연구는 사실 그 역사가 매우 깊다. 기원전 300년경 그리스의 수학자 유클리드(Euclid)의 양안 시차에 대한 설명과 15세기 이탈리아 수학자인 레오나르도 다빈치의 입체시(stereopsis) 증명으로 시작되었다 [1]. 이러한 오랜 역사에도 불구하고 3차원 영상이 일반인에게 직접 다가가지 못한 이유는 3차원 영상처리 기술이 충분히 성숙하지 못했기 때문이다. 예를 들어, 시청자가 입체

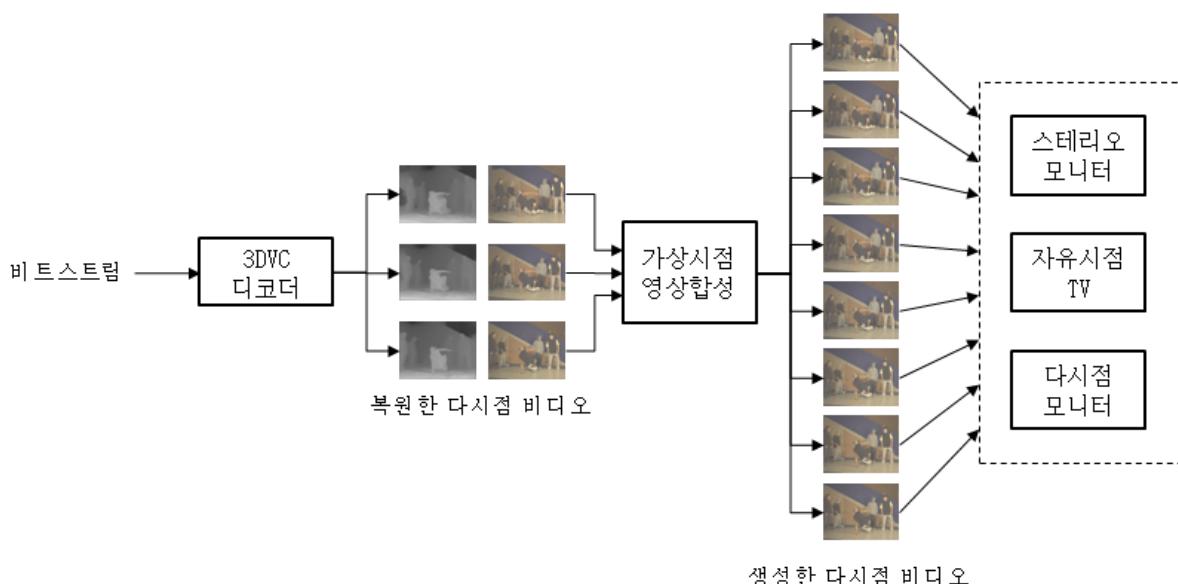


그림 1. 가상시점 영상생성 기술을 이용한 3차원 영상서비스

영상을 장시간 보면 시청자가 어지러움을 느낀다거나 구토 증상이 발생하는 문제가 있다. 이는 사람 눈의 특성을 충분히 고려하지 못한 채 3차원 영상을 제작하는 미숙함 때문이다. 또한 깊이감이 있는 컨텐츠 제작의 어려움과 저렴하고 성능이 뛰어난 3차원 재현 장치의 미비, 기존 기술과의 호환성 문제, 미약한 3차원 시장 등의 어려움이 있다. 최근 이러한 문제점들을 해결하기 위한 다양한 방법이 제안되고 있다.

최근 MPEG(moving picture experts group)의 3차원 비디오(3D video, 3DV) 표준화 그룹에서는 다양한 3차원 영상 서비스를 제공할 수 있는 시스템을 정의하고, 다시점 영상과 깊이 정보를 효율적으로 압축 부호화하는 국제 표준화 작업을 진행하고 있다 [2]. MPEG에서 제시하는 3차원 비디오 시스템은 기존의 양안식(stereoscopic) 영상을 기본적으로 이용하고, 보다 발전된 고화질 고해상도 광각의 3차원 비디오 시스템이다. 이 시스템은 다시점 카메라로 획득한 2차원 영상과 이에 대한 깊이 정보를 입력 데이터로 받아 사용자가 원하는 대로 시청 시점을 자유롭게 선택할 수 있는 다시점 3차원 재현 장치를 이용하여 다양한 형태로 입체 영상을 즐길 수 있게 한다.

그림 1에서 보인 것과 같이, MPEG 표준화 그룹에서 제시한 이 시스템은 미래의 3차원 비디오 기술의 기초기술을 제공하기 위하여 보다 미래지향적인 시스템을 구상하여 개발하고 있다. 채널을 통해 전송된 비트스트림을 디코더에서 복원한 다시점 비디오는 깊이영상을 이용하여 보다 많은 시점을 생성한 후 다양한 재현장치를 이용할 수 있게 한다. 이러한 서비스를 제공하는 재현장치는 보다 많은 시점의 영상을 효과적으로 이용한다. 이러한 장치에 공통적으로 필요한 기술은 제한된 시점의 영상을 이용하여 보다 많은 시점을 생성하는 가상시점 영상생성 기술이다.

본 논문에서는 위에서 설명한 자유시점 영상서비스를 위한 가상시점의 영상생성 기술을 간략히 설명한다. 특히, MPEG의 3DV 그룹에

서 제공하고 있는 시점합성 참조소프트웨어 (view synthesis reference software, VSRS)에 이용되고 있는 기술을 기반으로 설명한다. 먼저 3차원 비디오의 개념을 이해하기 위한 입체시와 3차원 카메라에 대해 소개하고, 이어서 가상시점의 영상생성 기술을 소개한다.

II. 3차원 비디오의 소개

1. 입체시와 3차원 카메라

사람이 시각적으로 입체감을 느끼는 원리는 양안의 시차에 따른 입체시의 개념으로 설명할 수 있다. 입체감을 느끼는 요인은 크게 생리적 요인과 경험적 요인으로 구분할 수 있다.

생리적 요인은 그림 2와 같이 양안시차(binocular disparity)에 의해 입체감을 느끼는 것이 대표적이다. 사람의 두 눈은 평균적으로 65mm 정도 떨어져 있기 때문에 같은 장면을 보더라도 망막에 맷히는 상은 다소 차이가 있다. 가까이에 있는 물체는 양안에 맷히는 위치가 크게 차이나며, 물체가 크게 맷힌다. 반면, 멀리 떨어져 있는 물체는 양안에 맷히는 위치가 크게 차이가 나지 않고, 물체가 작게 맷힌다. 이로 인해 물체의 원근을 인식할 수 있다. 경험적 요인은 관찰 대상의 움직임 속도 차이 (motion parallax)나 물체의 거리에 따른 투명도 차이(aerial perspective), 선원근법(linear perspective) 등이 있다.

3차원 카메라는 위의 생리적 요인을 반영하여 설계된다. 일반적인 스테레오 카메라는 두 카메라가 같은 방향을 향하는 평행형 (parallel) 카메라이다. 하지만 최근에는 양안

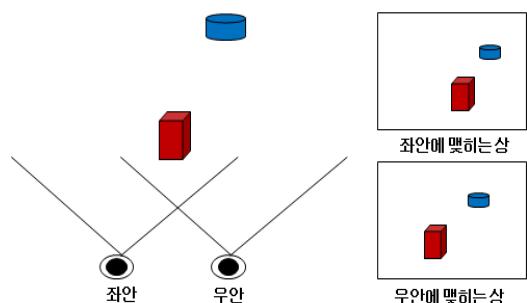


그림 2. 양안시차에 의한 입체시 원리

의 특성을 더욱 적극적으로 반영한 카메라가 개발되고 있다. 그림 3에 보인, 한국방송(KBS)에서 개발한 HD 양안식 3D 카메라는 기본적으로 두 카메라의 거리를 65mm로 두고 필요에 따라 거리를 조정한다. 또한 주시각을 조절하기 위해 카메라의 방향을 조절하는 기능도 포함한다 [3].



그림 3. 한국방송의 양안식 3D 카메라

최근 양안식 카메라보다 시점수를 늘려 다시 점의 영상을 획득하여 보다 효과적인 입체영상 을 제현하려는 시도가 진행되고 있다. 그림 4(a)과 같이 동일한 종류의 카메라를 다수 배치하고, 카메라 사이의 동기를 유지하여 다시 점 영상을 획득한다. 또한, 각 시점의 깊이데이터를 획득하기 위하여 그림 4(b)와 같이 깊이카메라를 설치하여 다시점 영상을 획득하기도 한다. 이렇게 획득된 영상은 다시점 카메라의 장면과 해상도를 동일하게 맞추기 위한 작업을 수행한다. 이와 같은 다시점 카메라를 이용하여 획득한 영상은 카메라간 불일치 요소를 제거하기 위한 처리를 수행한다. 카메라 세팅의 미세한 불일치를 조정하기 위해 다시점 영상정렬(multi-view rectification)과 다시점 색상보정(color correction)을 사용한다.

2. 3차원 비디오 시스템

현재 MPEG에서는 3차원 비디오 시스템에 대한 논의를 통해 3차원 비디오 부호화 (3D video coding) 기술에 대한 국제 표준화 작업을 진행하고 있다. 다시점 영상 기반의 3차원 정보는 시점 수에 따라 데이터양이 증가하기



(a) 평행형 다시점 카메라



(b) 복합형 다시점 카메라

그림 4. 다시점 카메라 시스템

때문에 기존의 한 시점 영상보다 많은 정보를 처리해야 한다. 이는 제한된 대역폭을 통한 3 차원 비디오 서비스에는 큰 걸림돌이 될 수 있다. 이에 MPEG에서는 3차원 비디오 부호화에 중점을 둔 국제 표준을 만들기 위해 노력하고 있다.

그림 5는 MPEG에서 고려하고 있는 다시 점 영상을 이용한 3차원 비디오 시스템의 개념도이다. 이 시스템의 특징은 다시점 비디오를 이용하여 보다 많은 시점의 영상을 합성하기 위해 깊이 정보를 이용한다는 점이다. 전송 단에서는 다시점 비디오와 깊이 정보를 압축하

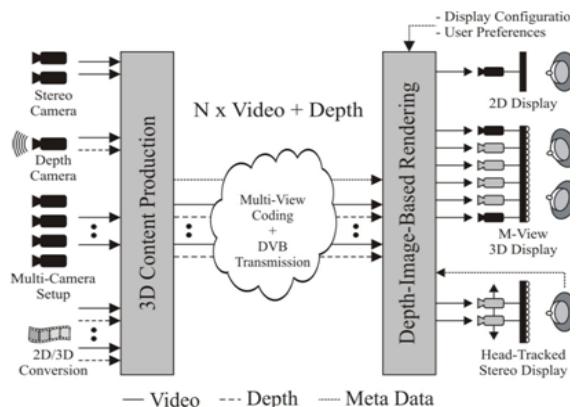


그림 5. 다시점 영상을 이용한 3차원 비디오 시스템

여 전송하고, 수신단에서는 이 신호를 받아서 복원한 뒤 재현 장치의 종류에 맞는 시점을 생성하고 재현한다.

양안식 영상은 두 시점의 영상을 이용하여 한 시점의 3차원 깊이감을 제공하지만, 다시점 영상을 이용하면 여러 시점을 이용한 3차원 영상을 재현할 수 있으며, 사용자가 원하는 임의의 시점의 영상을 만들어 보여줄 수도 있다. 3차원 입체영상을 재현하는 디스플레이 장치 중에는 안경을 사용하지 않는 9시점이나 16시점 무안경식 3차원 디스플레이 장치가 있는데, 이러한 장치를 이용하면 안경을 쓰는 불편함 없이 보다 실감나는 입체 영상을 감상할 수 있다.

3. 3차원 비디오 부호화 기술

MPEG 그룹에서 개발하고 있는 3차원 비디오 부호화 방법은 다시점 영상과 다시점 깊이 영상을 효과적으로 부호화 하는 기술이다. 보다 많은 시점의 영상을 생성하기 위하여 깊이 영상을 동시에 압축하여 전송해야 하므로, 깊이영상에 대한 부호화 방법에 대한 개발이 필요하다. 이에 MPEG 그룹에서는 그림 6과 같이 3차원 비디오 부호화의 개발 방향을 발표했다 [2].

3차원 비디오를 부호화하는 방법을 크게 두 가지로 나눌 수 있다. 첫째는, 단시점의 2차원 영상과 깊이 영상을 동시에 부호화하는 방법이다. 복호된 깊이 영상을 이용하면 임의의 시점을 합성할 수 있기 때문에 3차원 영상을 재현 할 수 있다. 그러나 참조할 수 있는 시점이 하나이므로 시점이동에 따른 빈 영역의 처리가 어려워진다. 이러한 문제점은 다시점 영상을 이용하면 해결할 수 있다. 그래서 두 번째 방법은 다시점 비디오와 깊이 영상을 부호화하는 방법이다. 다시점 비디오를 각 시점별로 개별적으로 부호화하지 않고 시점 사이의 상관도를 이용하여 부호화한다. 깊이영상도 마찬가지로 시점간 상관도를 이용하여 부호화 한다. 향후

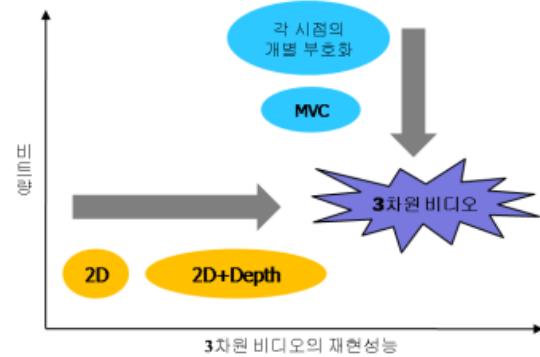


그림 6. 3차원 비디오 부호화의 개발 방향

진행될 MPEG 회의에서는 다시점 영상뿐만 아니라 다시점 깊이 영상을 효과적으로 부호화 하는 방법이 제안될 것으로 보인다.

4. 3차원 비디오 부호화를 위한 테스트 영상

MPEG의 3DV 표준화 그룹은 크게 두 번에 걸쳐 테스트 영상을 모집했다. 일차원 평행형 다시점 카메라를 이용한 다시점 비디오를 요청 했고 [4], 이어서 움직이는 다시점 카메라로 촬영된 영상을 요청했다. 이러한 요청에 대한 응답으로 여러 연구기관에서 다양한 테스트 영상을 제공했다. 모든 테스트 영상은 영상 정렬화가 완료되었으며, 대부분의 영상이 색상보정이 된 영상이다. 특히, 나고야 대학의 'Balloons', 'Kendo' 영상과 Poznan 대학의 'Poznan_Hall1', 'Poznan_Hall2' 영상은 움직이는 다시점 카메라로 촬영한 영상이다. 이에 비해 Philips에서 제공한 'Mobile' 영상과 'Beer_garden' 영상은 컴퓨터 그래픽스로 획득한 영상이기 때문에 깊이 영상의 정확도가 가장 높다.

III. 가상 시점의 영상생성 기술

시청자가 원하는 임의의 시점의 영상을 자유롭게 보여주기 위해서는 많은 수의 시점이 필요하다. 전송 채널의 대역폭 제한으로 인해 전송된 시점의 수가 충분치 않다면 임의의 중간시

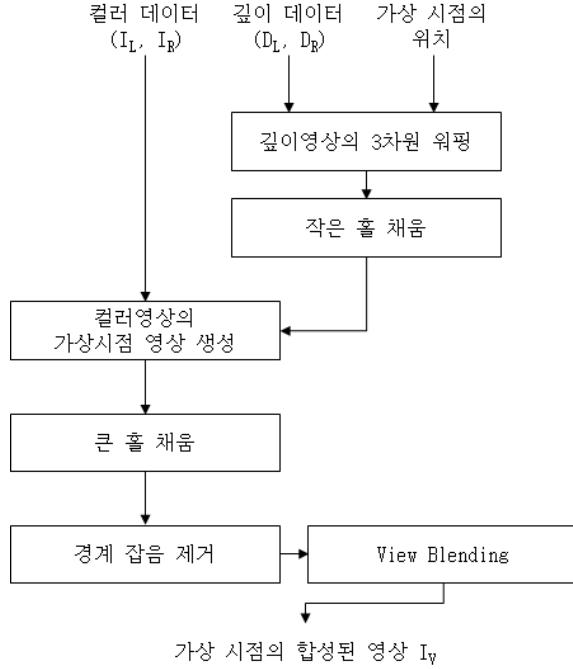


그림 7. 가상 시점 영상 생성

점 영상을 합성하여 3차원 모니터에 재현해야 한다. 이렇게 임의의 중간시점 영상을 생성하는 기술이 영상 합성(view synthesis)이라고 한다. 그림 7은 VSRS의 3차원 워핑(3D warping)을 이용한 영상합성 방법을 도시화한 것이다. 본 논문에서는 그림과 같은 순서로 설명한다.

현재 MPEG 그룹에서 고려하고 있는 3차원 비디오 시스템은 각 시점에 대한 깊이 영상과 카메라 매개 변수를 부가적인 입력 데이터로 정의하고 있다. 이러한 부가적인 데이터를 이용하면 임의의 중간시점의 영상을 자유롭게 합성할 수 있다. 이러한 기능을 수행하는 시점 합성 참조 소프트웨어 (view synthesis reference software, VSRS)가 개발되었다. 현재는 VSRS 3.5가 MPEG의 SVN 서버를 통해 배포되고 있다 [5].

1. 좌표계와 카메라 변수

자유시점의 영상을 합성하는 기술을 이해하기 위해서는 카메라의 기본 구조를 알아야 한

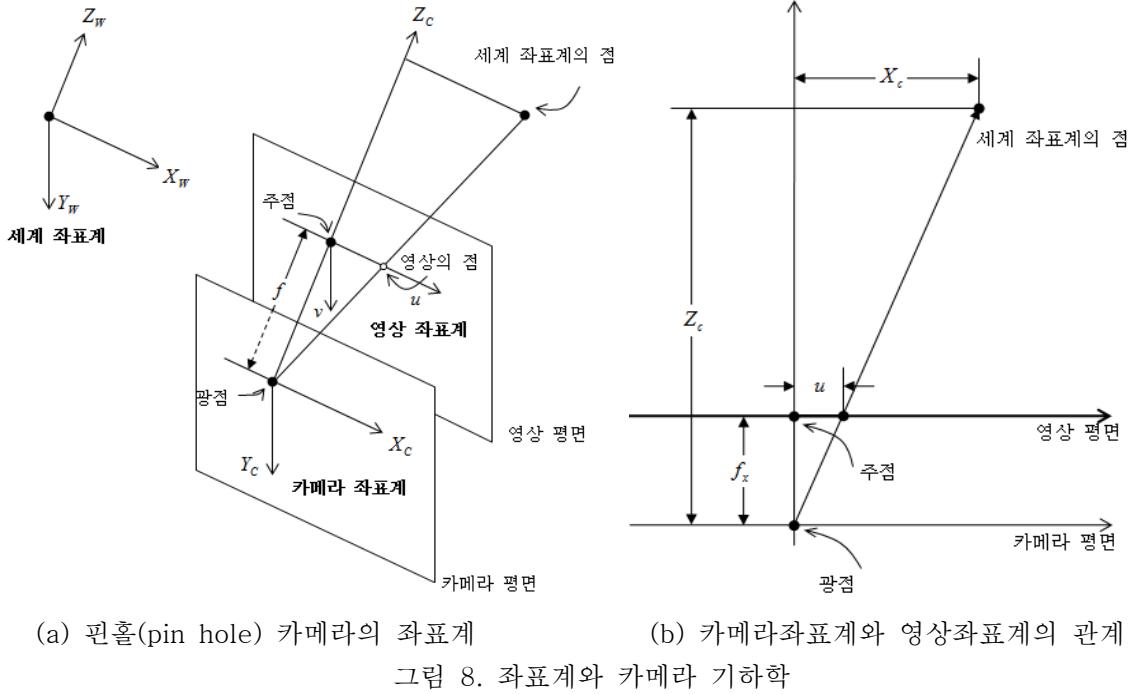
다. 카메라 구조를 설명하는 기본 모델은 핀홀 카메라(pinhole camera)이다. 3차원 공간의 한 점이 핀홀 카메라를 통하여 영상에 맷히는 과정을 세 개의 좌표계를 이용하여 설명할 수 있다. 장면이 존재하는 공간의 좌표를 표시하는 세계좌표계(world coordinate system), 카메라 내부의 공간을 표시하는 카메라좌표계(camera coordinate system)과 영상의 위치를 나타내는 영상좌표계(image coordinate system)이다. 이 세 좌표계는 서로 유기적으로 연관되어 있다. 그림 8(a)는 세계좌표계상의 한 점(3D point)이 카메라를 통해 영상에 사영되는 과정을 설명한 것이다.

단일 카메라 환경에서의 세계좌표계는 임의적이다. 그러나 다시점 비디오 시스템에서는 단일의 세계좌표계를 이용하고, 각 카메라의 위치는 이 세계좌표계를 기준으로 표시한다. 그리고 각 카메라는 자신의 고유한 카메라좌표와 영상좌표계를 가진다. 카메라좌표계는 주평면(principal plane)인 x_c-y_c 평면을 가지는 3차원 좌표계이다. 영상좌표계는 세계좌표계에서 투영된 빛이 맷히게 되는 2차원 좌표계이다. 카메라좌표계와 영상좌표계는 서로 평행한 관계를 가진다.

이와 같은 세 가지의 좌표계간의 관계를 설명하기 위해서 두 가지의 카메라 변수를 이용한다. 카메라 내부의 기하학적 구조를 설명하는 내부변수(intrinsic parameter)와 카메라의 위치와 방향을 표시하는 외부변수(extrinsic parameter)를 이용한다. 카메라 내부변수 A는 카메라좌표계에서 영상좌표계로 변환하는 과정을 설명한다. 일반적인 내부변수는 다음 식 (1)과 같이 표현한다.

$$\mathbf{A} = \begin{pmatrix} f_x & 0 & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

여기서 f_x 와 f_y 는 x-축과 y-축의 초점거리를 나타낸다. 그리고 (p_x, p_y) 는 영상좌표계의 원



점으로부터 주점(principal point)의 위치를 나타낸다. 일반적인 영상압축에서는 영상의 원점이 영상의 좌측 상단이므로 위의 값은 양수를 갖는다. 그림 8(b)과 같이, 카메라좌표계의 한 점 $\mathbf{M}_C = [x_C \ y_C \ z_C]^T$ 과 영상에 사영되는 점 $\mathbf{m} = [u \ v]^T$ 과의 관계를 다음의 식(2)와 같이 표현할 수 있다.

$$\frac{u}{x_C} = \frac{f_x}{z_C} \quad (2)$$

여기서 (p_x, p_y) 가 영벡터(zero vector)가 아니면, 영상의 u 값은 다음의 식(3)과 같이 구할 수 있다.

$$u = \frac{f_x \cdot x_C}{z_C} + p_x \quad (3)$$

위의 과정을 종합해 보면, 카메라 내부 변수와 두 점의 관계식을 다음의 식 (4)를 이용하여 표현할 수 있다.

$$z_C \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{A} \cdot \begin{pmatrix} x_C \\ y_C \\ z_C \end{pmatrix} = \begin{pmatrix} f_x & 0 & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_C \\ y_C \\ z_C \end{pmatrix} \quad (4)$$

카메라 내부 변수와 달리, 세계좌표계에서

카메라좌표계로의 변환을 나타낸 것이다. 방향변환을 나타내는 방향행렬 \mathbf{R} 과 카메라의 위치를 나타내는 위치벡터 \mathbf{t} 로 이루어진다. 이러한 변수를 이용하면, 세계좌표계의 한 점 $\mathbf{M}_W = [x_W \ y_W \ z_W]^T$ 과 카메라좌표계의 한 점 $\mathbf{M}_C = [x_C \ y_C \ z_C]^T$ 의 관계를 다음의 식 (5)로 표현할 수 있다.

$$\begin{pmatrix} x_C \\ y_C \\ z_C \end{pmatrix} = \mathbf{R} \cdot \begin{pmatrix} x_W \\ y_W \\ z_W \end{pmatrix} + \mathbf{t} \quad (5)$$

위 식 (4)와 식 (5)를 종합하면, 세계좌표계의 점과 영상좌표계의 점의 관계를 다음의 식 (6)과 같이 표현할 수 있다.

$$\tilde{\mathbf{m}} = \mathbf{A} \cdot \mathbf{R} \cdot \mathbf{M}_W + \mathbf{A} \cdot \mathbf{t} \quad (6)$$

여기서 $\tilde{\mathbf{m}}$ 는 깊이값이 곱해진 3차원 벡터이다. 그리므로 실제의 2차원 점의 위치를 구하기 위해서는 x축 요소와 y축 요소를 z축 요소로 나누면 된다.

2. 깊이값 표현 방법

깊이영상은 영상내의 각 화소의 실제 깊이

정보를 8-비트 영상으로 표현한 것이다. 식 (7)은 깊이영상의 화소값을 실제 깊이값으로 변환하는 식을 보인 것이다.

$$Z = \frac{1}{\frac{v}{255} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}}} \quad (7)$$

여기서 Z_{near} 와 Z_{far} 는 깊이값의 범위를 나타내는 값이고, v 는 깊이영상의 각 화소값을 나타낸다. 깊이영상의 각 화소값 v 가 255일 경우는 객체가 카메라에 가장 가까이 있다는 것이고, 0일 경우는 객체가 카메라와 가장 멀리 떨어져 있음을 의미한다.

3. 3차원 워핑을 이용한 영상 합성

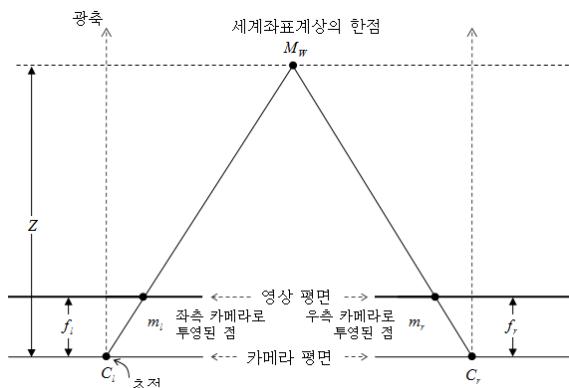


그림 9. 두 카메라에 투영된 상응화소

위와 같은 카메라 변수와 깊이영상을 이용하여 임의의 중간시점의 영상을 생성하기 위해서는 시점간의 상응화소를 찾아야 한다. 그림 9는 3차원 공간상의 한 점과 이에 상응하는 두 시점간의 관계를 나타낸 것이다. 3차원 공간의 한 점 M_w 가 두 시점의 각 점 m_l 과 m_r 로 사영되었을 때, 다음과 같은 관계가 성립한다.

$$I_l(\mathbf{m}_l) = I_r(\mathbf{m}_r) \quad (8)$$

두 점의 위치를 알기 위해서는 깊이값을 이용해야 한다. 그림 9에서 왼쪽 영상의 한 점 m_l 과 상응하는 점 m_r 을 오른쪽 영상에서 찾으려 할 경우, 두 점간의 관계식을 다음의 식

(9)와 같이 표현 할 수 있다.

$$\begin{aligned} \tilde{\mathbf{m}}_r &= \mathbf{A}_r \cdot \mathbf{R}_r \cdot \mathbf{R}_l^{-1} \cdot \mathbf{A}_l^{-1} \cdot [\mathbf{m}_l | 1]^T \cdot Z_l(\mathbf{m}_l) \\ &\quad - \mathbf{A}_r \cdot \mathbf{R}_r \cdot \mathbf{R}_l^{-1} \cdot \mathbf{t}_l + \mathbf{A}_r \cdot \mathbf{t}_r \end{aligned} \quad (9)$$

여기서 각 변수의 l, r 은 왼쪽 시점과 오른쪽 시점을 각각 가리킨다. 또한 Z_l 은 식 (7)을 이용한 왼쪽시점의 실제 깊이값을 가리킨다. 이로써, 참조시점인 왼쪽 시점의 영상을 가상시점인 오른쪽 영상의 시점으로 옮길 수 있다.

이와 같은 방법으로 가상시점의 영상을 합성하는 방법은 크게 두 가지가 있다. 참조시점의 컬러영상을 가상시점으로 직접 워핑하는 전방향 워핑(forward warping)과, 깊이영상을 가상시점으로 워핑하여 합성하는 후방향 워핑(backward warping)이 있다. VSRS의 'General Mode'는 후방향 워핑을 이용한다 [6].

4. 1차원 평행형 카메라 구조에서의 영상합성

위에서 설명한 3차원 워핑을 이용한 영상 합성은 일반적인 카메라 구조에서 사용가능한 방법이다. 만약 카메라의 구조가 1차원 평행형이면 위의 식 (9)가 단순하게 정리된다. 1차원 평행형 카메라는 카메라의 방향이 같기 때문에 $Rr \cdot Rl - 1 = I$ 이 성립한다. 또한 카메라 내부 변수의 초점거리(f)가 같다고 가정하고 주점(principal point)의 x-축 값만 다르다고 하면, 식 (9)는 다음의 식 (10)과 같이 간단하게 표현할 수 있다.

$$u_r = u_l - \frac{f \cdot (t_x^l - t_x^r)}{Z(\mathbf{m}_l)} + (p_x^l - p_x^r) \quad (10)$$

여기서 u_l 은 좌영상의 x축 화소 위치이고, u_r 은 우영상의 x축 화소 위치이다. 또한 t_x^l 과 t_x^r 은 카메라의 위치 벡터의 x축 값을 나타낸다. 이와 같은 방법은 VSRS의 '1D mode'에 구현되어 있다 [7].

5. 훌채움(Hole filling) 기술

위의 시점간 상응화소 결정 방법을 이용하여

가상시점의 영상을 합성하면, 시점 이동에 따른 빈공간(hole)이 드러나게 된다. 빈공간은 크게 두 종류로 나눌 수 있다. 한 화소 단위로 발생하는 좁은 (혹은 작은) 빈공간과 너비가 넓은 빈공간으로 나눈다. 좁은 빈공간은 시점 간 화소의 위치를 결정하는 과정에서 발생하는 반올림 오차에 의해 발생한다. 이러한 빈공간은 화소의 너비가 좁기 때문에, 인접화소의 값 을 이용하여 채울 수 있다. 가장 간단한 방법 으로 미디언 필터(median filter)를 이용한다.

좁은 빈공간은 비교적 간단한 방법으로 해결 할 수 있으나, 넓은 빈공간은 너비가 넓으므로 인접화소의 값을 이용하는데 한계가 있다. 그러나 MPEG의 3DVC 그룹에서 취하는 입력데 이터는 두개 이상의 참조시점을 이용하므로, 빈공간을 채우기 위한 정보를 다른 참조시점에 서 얻을 수 있다. 그럼 10은 두 참조시점을 이용하여 가상시점으로 합성한 두 영상을 홀채움 하는 방법을 설명한 것이다. 두 참조시점에서 각각 가상시점으로 위평을 하면 서로 다른 영 역에서 넓은 빈공간이 발생한다. 그러므로 빈 공간을 채우기 위한 참조 가능한 정보는 다른 시점에서 위평한 영상에서 찾을 수 있다. 이러한 방법을 통해 대부분의 빈공간을 채운다. 그리고 두 참조시점에서 공통으로 보이지 않는 빈공간은 인페인팅(inpainting)과 같은 방법 으로 채운다.



그림 10. 넓은 빈공간 채우기: (a) 오른쪽 참 조영상에서 위평한 가상시점 영상, (b) 왼쪽 참 조영상에서 위평한 가상시점 영상

6. 합성한 영상의 경계 잡음 제거

여러 시점에서 찍은 영상은 객체를 모두 포함하고 있기 때문에 각 시점사이의 거리 차이 를 계산하여 깊이 영상을 얻는다. 이 방법은 깊이 카메라와 같이 촬영 환경의 제약이 없으나, 색상의 유사도로 거리 값을 추정하기 때문에 비교적 오류가 크다는 단점을 지닌다. 이러한 방법으로 얻은 깊이 영상은 영상 합성 과정에 중요한 정보를 제공하지만, 잘못 추정된 깊이값은 합성 영상에서 눈에 거슬리는 잡음으로 나타난다. 특히 객체의 경계 부근에서 깊이값이 잘못 추정될 수 있는데, 이 때문에 객체의 경계에 잡음이 자주 발생한다. 이에 광주과학 기술원에서는 깊이 영상을 이용한 영상합성 방법과 깊이값이 객체의 경계 부분에서 많이 발생하는 점을 고려하여 경계 잡음을 제거하는 방법을 제안했다 [8]. 그림 11은 중간시점 영상을 합성할 때 발생하는 경계 잡음과 제안된 방법으로 경계 잡음을 줄인 결과를 비교한 것이다.

7. 시점 혼합 (View Blending) 기술

지금까지 설명한 방법을 이용하여 생성한 가상시점의 영상은 참조시점의 수만큼 얻을 수 있다. 본 논문에서 이용한 두 시점의 영상을 이용한 영상합성방법은 두 개의 가상시점의 영상을 얻을 수 있다. 이렇게 획득한 영상을 각각 $I_{v,l}$ 과 $I_{v,r}$ 이라고 하면 다음의 식(11)과 같은 방



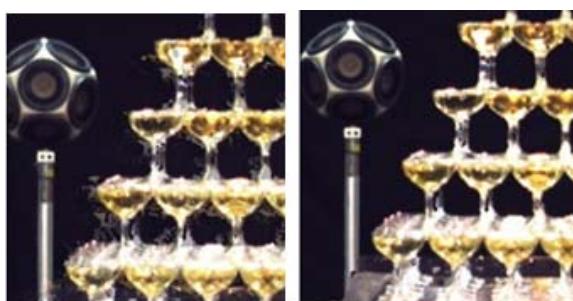
그림 11. 경계 잡음이 제거된 영상

법으로 최종 영상을 획득할 수 있다.

$$\begin{cases} I_v = \alpha \cdot I_{v,l} + (1-\alpha) \cdot I_{v,r} \\ \alpha = \frac{|t_x^v - t_x^r|}{|t_x^l - t_x^r|} \end{cases} \quad (11)$$

이때, 상수값 알파는 참조시점과 가상시점의 거리를 나타낸다.

참조시점의 깊이영상과 기타 변수가 정확할 경우에는 위의 방법이 효과적으로 이용된다. 하지만, 데이터가 부정확할 경우는 두 가상시점의 영상을 합성하는 방법이 효과적이지 않게 된다. 특히 두 깊이영상이 서로 일관성이 낮으면 가상시점의 영상에 눈에 띠는 잡음이 발생하게 된다. 이러한 영향을 줄이기 위한 방법으로 참조시점에 가까운 가상 시점의 영상을 주로 이용하는 방법이다. 예를 들어, 가상시점의 위치가 왼쪽 참조영상과 가까운 경우, 홀채움과 경계잡음이 제거된 왼쪽 가상영상을 최종 합성영상으로 이용하는 것이다. 그림 12는 이와 같은 방법을 이용하여 'Champagne_tower' 영상을 합성한 결과이다. 그림 12(a)와 같이, 가상시점으로 합성한 두 영상을 이용한 영상은 캡 주위가 지저분하지만, 가까운 시점을 이용한 방법(그림 12(b))은 깨끗한 것을 확인할 수 있다.



(a) 두 영상 합성 (b) 가까운 시점 이용
그림 12. 시점 혼합 기술을 이용한 합성결과

8. 계층적 깊이 비디오를 이용한 영상합성

MPEG 그룹에서 제시하고 있는 3차원 비디오 시스템은 다시점 비디오와 다시점 깊이 비디오를 입력데이터로 취하고 있다. 이때 N 시



(a) 중앙시점의 색차영상 (b) 중앙시점의 깊이 영상



(c) 감춰진 색차영상 (d) 감춰진 깊이 영상

그림 13. 계층적 깊이 비디오 데이터

점의 영상과 N 시점의 깊이 영상이 입력 데이터이므로 시점의 증가에 따라 데이터량도 선형적으로 증가하는 문제가 있다. 이에 Philips에서는 부호화할 시점의 수를 줄이고 압축 효율을 높이기 위하여 계층적 깊이비디오(layered depth video, LDV)를 제안하였다 [9].

LDV는 기준 시점의 영상과 시점 이동에 따라 가려지고 드러나는 영역에 대한 데이터를 이용하여 임의 시점의 영상을 생성하는 방법이다. 그림 13은 LDV를 구성하는 데이터를 보인 것이다. 다시점 비디오에서 중앙에 위치하는 시점의 영상을 그림 13(a)는 중앙 시점의 색차 영상으로 지정하고, 그 시점의 깊이 영상을 그림 13(b)의 데이터로 지정한다. 그리고 그림 13(c)와 그림 13(d)와 같이 시점 이동에 따라 새로 드러나는 정보만을 취한다. 이렇게 간소화된 데이터를 이용하면 카메라 수의 입력 데이터를 전송할 필요가 없이 줄여진 데이터만을 이용하여 3차원 비디오 시스템을 구현할 수 있다는 장점이 있다.

VI. 결론

3차원 영상 서비스가 시작되면, 시청자는 보다 입체감 있고 실감나는 영상을 체험할 수 있을 것이다. 이러한 3차원 비디오 시대를 이끌

수 있는 기술이 바로 3차원 비디오 부호화 기술이다. 본 논문에서는 이와 같은 3차원 비디오 시스템의 요소기술인 가상시점의 영상합성 방법을 설명했다. 깊이영상은 이용한 시점간의 상응화소 결정 방법, 시점 이동에 따른 빈공간을 채우는 방법, 경계 잡음을 제거하는 방법과 계층적 깊이비디오를 이용한 영상합성 방법을 설명했다. 이와 같은 방법을 이용하면 보다 선명한 화질의 가상시점 영상을 합성할 수 있다. 지금까지 MPEG에서 만든 영상 압축 표준이 우리의 생활에 큰 변화를 주었던 것처럼, 앞으로 제정될 3차원 비디오 부호화 기술도 우리 생활 가까이 다가올 것으로 기대된다.

*** 감사의 글 ***

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

*** 참고문헌 ***

- [1] 호요성, 김성열, "3DTV 3차원 입체영상 정보처리," 두양사, 2010.
- [2] ISO/IEC JTC1/SC29/WG11, "Vision on 3D Video," N10357, Feb. 2007.
- [3] 박창섭, 이준용, 함상진, 조인준, 이근식, "3DTV 방송용 카메라 기술 개발동향," 방송공학회지, 제15권, 제1호, pp. 12-23, 2010. 03.
- [4] ISO/IEC JTC1/SC29/WG11, "Call for Contributions on 3D Video Test Material," N9595, Jan. 2008.
- [5] MPEG 3DV AhG, "View Synthesis Reference Software," URL: http://wg11.scs29.org/svn/repos/MPEG-4/test/trunk/3D/view_synthesis/VSRS
- [6] ISO/IEC JTC1/SC29/WG11, "Reference Softwares for Depth Estimation and View Synthesis," M15377, April 2008.
- [7] ISO/IEC JTC1/SC29/WG11, "Simple View Synthesis," m15696, July 2008.
- [8] Cheon Lee and Yo-Sung Ho, "Boundary Filtering on Synthesized Views of 3D Video," Signal processing, Image processing and Pattern Recognition (SIP) 2008, pp. 15(1-4), Dec. 2008.
- [9] ISO/IEC JTC1/SC29/WG11, "Creation of LDV (Layered Depth Video) streams out of MV sequences," m15590, July 2008.