

### 3차원 비디오의 중간시점 영상합성을 위한 공통 빈공간 채움 방법

이천<sup>o</sup>, 호요성

광주과학기술원 정보통신공학부

leecheon@gist.ac.kr, hoyo@gist.ac.kr

#### 요 약

3차원 비디오 시스템은 다시점 비디오를 기반으로 보다 실감나는 입체영상을 제공한다. 중간시점 영상합성 기술은 다시점 카메라로 획득하지 못한 임의의 중간시점의 영상을 생성하는 방법이다. 영상합성의 중요한 기술은 참조 영상에 존재하지 않는 정보를 효과적으로 채우는 방법이다. 본 논문에서는 고화질의 중간시점 영상합성을 위한 공통 빈공간 채움 방법을 제안한다. 공통 빈공간을 배경의 영상으로 채우기 위해 인접 영역의 깊이값 중 가장 작은 값으로 깊이값을 결정하고, 참조하는 인접 화소의 거리와 깊이값을 고려하는 양방향 필터(bilateral filter)를 이용하여 공통 빈공간을 채운다. 이 방법을 이용하면 전경의 객체가 공통 빈공간에 영향을 주지 않아 배경 객체의 컬러가 자연스럽게 연결된 중간영상을 생성할 수 있었다.

#### 1. 서론

3차원 비디오 시스템(3D Video System)은 실감나는 입체영상 재현을 위해 고안된 차세대 멀티미디어 시스템이다. 이 시스템은 다시점 카메라로 획득한 영상의 보정, 압축, 복원, 재현 등의 모든 관련 기술을 포함한다. 이 시스템의 가장 큰 장점은 사용자의 위치나 선택에 맞추어 가장 효율적으로 입체감을 제공할 수 있다는 것이다. 이러한 기술을 뒷받침하는 핵심 기술은 제한된 시점의 영상으로 임의의 시점의 영상을 생성하는 중간시점 영상합성 기술이다 [1].

국제 표준화 기구인 MPEG (moving picture experts group)의 3차원 비디오 부호화(3D Video Coding) 그룹에서는 이러한 시스템의 요소기술을 개발하고 비디오 데이터를 효율적으로 압축하기 위한 표준화 작업이 진행되고 있다. 이 그룹에서 제안하는 미래의 3차원 비디오 시스템은 다시점의 영상과 깊이영상을 이용하여 다양한 3차원 디스플레이 장치에 이용할 수 있는 입체영상을 효과적으로 재현할 수 있는 장치이다. 이때, 다시점의 영상을 전송해야 하므로 다시점 영상의 시점간 중복성을 이용하는 다시점 비디오와 깊이 데이터의 압축-복원 방법을 개발하고 있다 [2].

중간시점의 영상합성은 복원된 다시점 비디오와 깊이 데이터를 이용하여 필요한 가상의 시점의 영상을 생성하는 방법이다 [3]. 깊이 데이터는 장면을 구성하는 객체의 3차원 거리를 나타낸다. 이 정보를 이용하면 참조영상의 각 화소를 3차원 공간으로 환원하고, 이를 다시 가상 시점으로 투영하면 두 시점의 위치관계를 알 수 있다. 이를 이용하면 참조시점의 영상을 가상시점의 영상으로 재구성할 수 있다.

이 과정에서 발생하는 문제는 시점 이동으로 인해 발생하는 빈공간(hole region) 문제이다. 빈공간은 참조시점에는 존재하지 않던 영역이 가상시점에 새로이 들어나는 영역을 일컫는다. 다시점을 이용한 영상합성에서 빈공간을 채우는 일반적인 방법은, 빈공간을 채울 수 있는 가용 영역을 다른 참조 시점에서 찾아 이용하는 방법이다. 대부분의 빈공간은 다른 참조시점에 존재하므로 빈공간 문제를 해결할 수 있다. 하지만, 그림 1과 같은 구조의 장면에서는, 가상시점에서는 보이지만 참조시점에서는 보이지 않는 공통 빈공간(common-hole region)이 발생한다. 일반적인 빈공간 채움 방법은 손상된 영상을 복원하는 인페인팅(inpainting) 방법을 이용하는데 [4], 이는 객체의 위치관계를 고려하지 않으므로 채워진 빈공간 영역이 부자연스럽게 보일 수 있다.

본 논문에서는 공통 빈공간을 보다 자연스럽게 채우기 위하여 객체의 깊이값을 이용하는 빈공간 채움 방법을 제안했다.

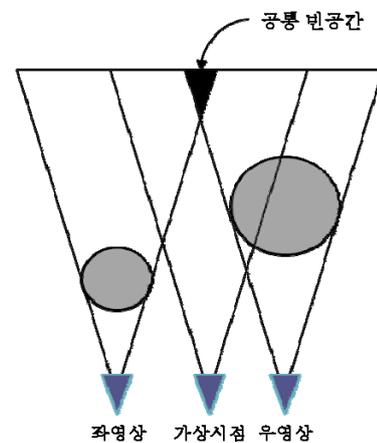


그림 1. 공통 빈공간 문제

2. 3차원 비디오 시스템을 위한 영상합성

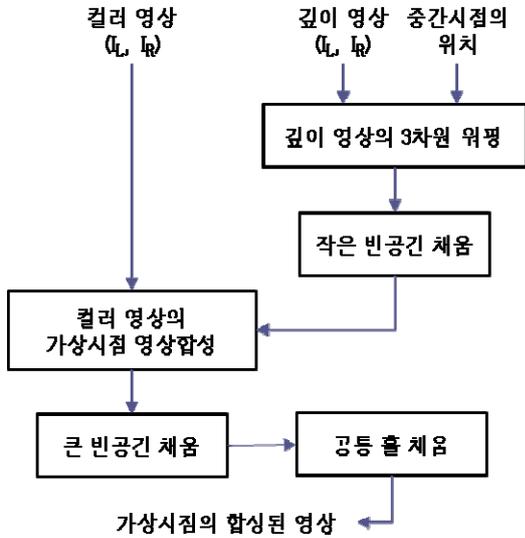


그림 2. 가상시점 영상합성

중간시점 영상합성 기술은 자유롭게 시청하는 시점을 변경하거나, 효율적인 입체영상 재현을 위해 서 임의의 시점의 영상을 생성하는데 필수적인 기술이다. 예를 들면, 복원된 두 영상의 시차가 사용자에 부적절할 경우, 이 기술을 이용하여 최적의 시점의 영상을 합성하여 가장 편안한 입체영상을 재현할 수 있다. 그림 2는 두 시점의 컬러 영상과 깊이영상을 이용하여 임의의 중간시점의 영상을 합성하는 방법을 보인 것이다.

입력 데이터는 두 시점의 컬러 영상과 깊이영상, 중간시점을 포함한 각 시점의 카메라 변수로 구성 된다. 먼저 두 시점의 깊이영상을 3차원 워핑 기술을 이용하여 중간시점으로 각각 시점 이동한다. 이 과정에서 발생하는 빈공간 중에 점과 같은 작은 빈공간은 미디언 필터를 이용하여 채운다. 그리고 워핑한 두 시점의 깊이영상과 참조 시점의 컬러 영상을 이용하여 중간시점의 영상을 각각 합성한다. 이 과정을 거치면 왼쪽 참조 시점에서 합성한 중간시점 영상  $I_{C,L}$ 과 오른쪽 참조 시점에서 합성한 중간시점 영상  $I_{C,R}$ 을 얻을 수 있다. 이후 시점이동에 따른 큰 빈공간을 채우기 위하여 상호 참조 가능한 빈공간 영역 정보를 이용한다. 마지막 단계로 두 참조시점에 존재하지 않는 공통 빈공간을 인접 화소의 정보를 이용하여 채우게 된다.

3. 공통 빈공간 채움 방법

앞에 설명한 중간시점 영상합성 방법을 이용하면 고화질의 중간시점의 영상을 생성할 수 있다. 중간시점 영상합성을 위하여 다시점 영상을 이용하게 되면 빈공간 문제는 대부분 해결할 수 있다. 예를 들어, 왼쪽 참조 시점의 영상을 오른쪽으로 이동했

을 경우 빈공간은 객체의 왼쪽에 발생하는데, 이 빈공간의 정보를 오른쪽 참조영상에서 참조할 수 있다. 반대로 오른쪽에서 왼쪽으로 시점 이동하여 발생한 빈공간은 왼쪽 참조영상을 이용할 수 있다.

하지만, 장면을 구성하는 객체의 위치가 복잡하거나 카메라간 간격이 넓은 경우 그림 1에서 보인 것과 같은 공통 빈공간 문제가 자주 발생할 수 있다. 이러한 공통 빈공간을 채우는 일반적인 방법은 선형 보간법, 배경의 값의 외삽법 및 인페인팅 방법 등이 있다. 최근 MPEG에서 가장 많이 사용하고 있는 VSRS (view synthesis reference software) 소프트웨어는 인페인팅(inpainting) 방법을 이용하고 있다 [4][5].

3.1 인페인팅을 이용한 공통 빈공간 채움 방법

영상 인페인팅 기술은 그림 3에 보인 것처럼 손상된 영상을 복원하기 위해 사용되는 기술이다. 손상된 영상과 손상된 영역을 나타내는 마스크 영상을 입력 데이터로 주면 손상된 영역의 정보를 복원한 영상을 얻을 수 있다. 식 (1)은 손상된 화소인  $p$ 의 정보를 복원하기 위한 인페인팅 방법을 보인 것이다.

$$I(p) = \frac{\sum_{q \in B_\epsilon(p)} \omega(p, q) |I(q) + \nabla I(q)(p - q)|}{\sum_{q \in B_\epsilon(p)} \omega(p, q)} \quad (1)$$

여기서  $B_\epsilon$ 와  $q$ 는 손상되지 않은 참조 가능한 인접 화소이고,  $\omega(p, q)$ 는 참조 화소의 가중치이다. 이와 같은 방법을 반복적으로 수행하여 손상된 화소 영역의 외부로부터 조금씩 복원하게 된다.



그림 3. 인페인팅을 이용한 영상 복원

중간 시점의 영상을 합성하는 과정에서 발생하는 대부분의 공통 빈공간은 그 영역의 크기가 크지 않다. 그러므로 인페인팅과 같은 방법을 이용하여 빈공간을 채우더라도 입체영상을 관람하는 사용자는 크게 불편함을 느끼지 않을 수 있다. 하지만, 그 객체의 위치가 복잡하여 공통 빈공간이 넓게 발생하게 될 경우에는 인페인팅을 이용한 방법은 비효율적이다.

공통 빈공간이 넓은 경우는 인페인팅에 의해 채워진 영역이 매우 눈에 거슬리는 영상을 생성할 수 있다. 왜냐하면, 인페인팅 방법은 각 객체의 깊이값을 고려하지 않고 인접 영상을 모두 이용하므로 자연스럽지 않은 값으로 빈공간이 채워지기 때문이다.

### 3.2 제안하는 공통 빈공간 채움 방법

본 논문에서 제안하는 공통 빈공간 채움 방법은 객체의 깊이값을 고려하여 참조할 화소를 선별적으로 이용하는 방법이다. 이를 위하여 공통 빈공간에 대한 몇 가지 가정을 세울 수 있다. 첫째, 공통 빈공간은 참조 시점에서 전경 객체에 의해 가려진 영역이므로, 이 영역은 전경 객체 뒤에 있는 배경 영역일 확률이 높다. 둘째, 그러므로 공통 빈공간의 깊이값은 인접한 화소의 깊이값 중에 가장 작은 값을 가질 확률이 높다. 셋째, 공통 빈공간의 영역은 비슷한 깊이값을 갖는 인접 영역과 비슷한 값을 가질 확률이 높다. 그림 4는 이와 같은 가정을 바탕으로 고안한 공통 빈공간 채움 방법을 나타낸 것이다.

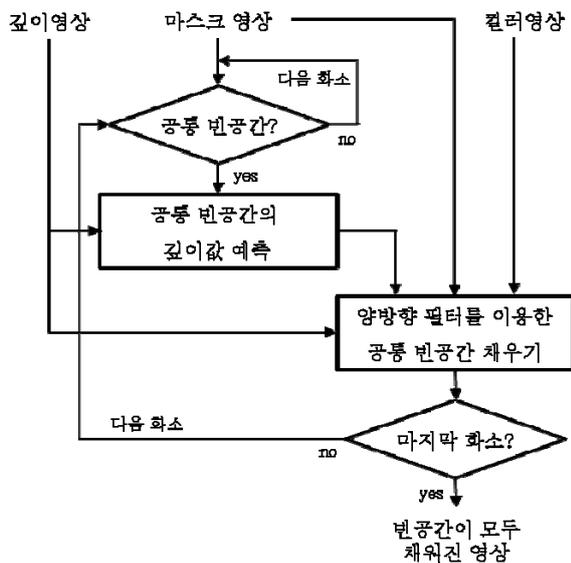


그림 4. 제안하는 공통 빈공간 채움 방법

위의 가정의 첫 번째와 두 번째 가정을 이용하면 공통 빈공간에 대한 깊이값을 식 (1)과 같이 예측할 수 있다. 충분한 크기를 갖는 윈도우  $W$  내의 빈공간이 아닌 화소 중에서 가장 작은 깊이값을 갖는 깊이값을 채워야 할 공통 빈공간의 깊이값으로 취한다. 그림 5는 공통 빈공간의 깊이값을 결정하기 위한 예를 보인 것으로, 가장 어두운 화소 값이 가장 멀리 떨어져 있는 배경에 해당하는 깊이값이므로 이 값을 공통 빈공간의 예측 깊이값으로 취한다.

$$\hat{d} = \min D(u, v), \quad D(u, v) \in W \quad (2)$$

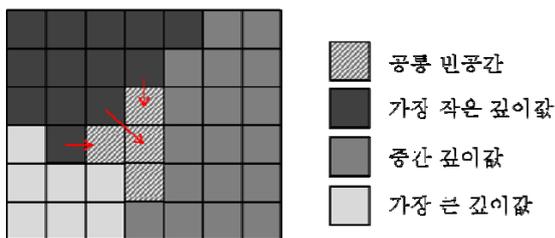


그림 5. 공통 빈공간의 깊이값 예측

이렇게 결정된 깊이값은 식 (3)과 같이 인접 화소의 깊이값을 고려하여 공통 빈공간의 색상 정보를 예측하게 된다. 식 (3)에 보인 것과 같이 공통 빈공간을 채우기 위한 방법으로 양방향 필터 (bilateral filter)를 이용한다.

$$C(x, y) = \frac{\sum_{u \in \hat{u}_p} \sum_{v \in \hat{v}_p} W(u, v, \hat{d}) C(u, v)}{\sum_{u \in \hat{u}_p} \sum_{v \in \hat{v}_p} W(u, v, \hat{d})} \quad (3)$$

$$W(u, v, \hat{d}) = \exp\left(-\frac{\|\hat{d}, D(u, v)\|^2}{2\sigma_D}\right) \exp\left(-\frac{(x-u)^2 + (y-v)^2}{2\sigma_r}\right) \quad (4)$$

$$C(u, v) = \alpha(u, v) I(u, v) \quad (5)$$

여기서  $(x, y)$ 는 참조할 인접화소의 위치이고,  $(u, v)$ 는 채워야 할 공통 빈공간의 위치를 말한다. 공통 빈공간의 값을 채우기 위한 인접 화소의 가중치 값은 식 (4)와 같이 결정한다. 깊이값과 비슷한 깊이값을 갖는 화소에 더 큰 가중치를 할당하고, 공통 빈공간 화소와 거리가 가까울수록 더 큰 가중치를 준다.  $\sigma_D$ 는 각각 깊이값의 유사도를 반영하는 상수이고,  $\sigma_r$ 은 인접 참조화소의 거리에 따른 참조 정도를 나타내는 상수이다. 식 (5)의  $\alpha$ 는 공통 빈공간을 표시하는 마스크 정보이다. 빈공간은 0으로 나머지 영역은 1의 값을 갖는다. 다시 말해, 빈공간의 화소는 고려하지 않는다.

### 4. 실험 결과 및 분석

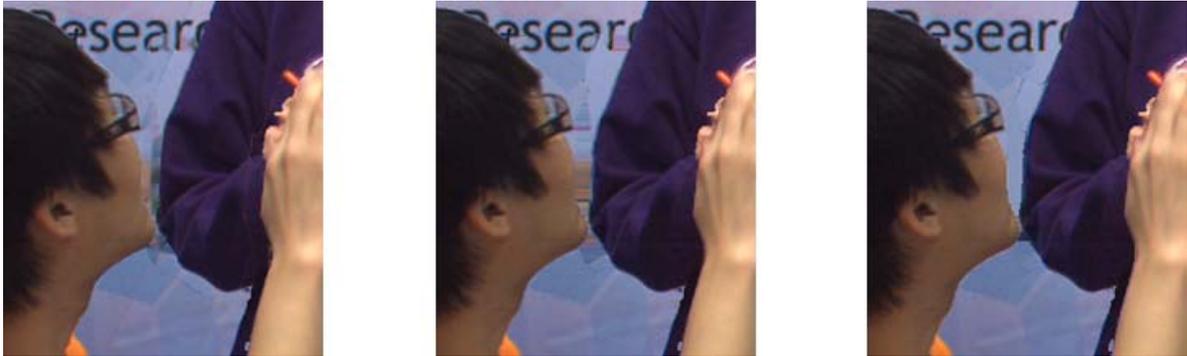
본 논문에서 제시한 방법을 평가하기 위해서는 그림 1에 보인 것과 같은 공통 빈공간이 존재하는 테스트 영상을 이용해야 한다. 그리고 중간시점 영상합성을 위해 이용 가능한 깊이영상이 있어야 하므로, 현재 MPEG 3DV 그룹에서 이용하고 있는 Café 영상을 이용했다 [6]. Café 영상은 1920x1080 해상도의 5시점으로 구성된 영상이다. 카메라 간격은 65mm이고 깊이값의 변화가 크고 장면을 구성하는 객체의 수가 많은 영상이다. 깊이영상은 MPEG 3DV 그룹에서 배포하고 있는 영상을 이용했다. 2번 4번 시점의 영상을 참조 시점으로 이용하고, 3번 시점의 영상을 합성했다.

중간시점의 영상을 합성하기 위한 참조 소프트웨어는 VSRS 3.5를 이용했다. 공통 빈공간의 깊이값을 예측과 컬러 값을 결정하기 위한 윈도우( $W$ )는 10x10의 마스크를 이용했다. 그리고 식 (4)의 표준편차 상수는  $\sigma_D = \sigma_r = 10$ 로 이용했다.

제안한 알고리즘과 기존의 방법을 비교하기 위하여 VSRS의 General Mode에 구현되어 있는 Inpainting 방법과, BoundaryNoiseRemoval (BNR) [7]에 구현되어 있는 같은 높이의 이용 가능한 깊이값 중



(a) 81 번 화면: (좌) 인페인팅 (중)BNR의 공통 빈공간 채움 (우) 제안한 방법



(b) 165 번 화면: (좌) 인페인팅 (중)BNR의 공통 빈공간 채움 (우) 제안한 방법

그림 6. 'Café' 영상을 이용한 중간시점 합성 결과

낮은 깊이값을 갖는 화소의 컬러를 이용하는 방법을 이용했다. 그림 6은 공통 빈공간 문제가 두드러지게 나타나는 81번째 화면과 165번째 화면을 비교한 것이다.

그림 6에서 보인 것처럼 인페인팅을 이용한 방법은 전경과 배경의 값을 구분하지 않고 참조 가능한 인접 화소를 모두 이용하기 때문에 눈에 거슬리는 결과를 보였다. 그리고 앞에 설명한 BNR에 구현되어 있는 방법은 수평방향으로 참조 가능한 깊이값 중 낮은 깊이값의 컬러 정보를 이용하기 때문에 줄무늬 모양의 영상이 생성되었다. 이에 비해 제안한 방법은 공통 빈공간의 영역이 배경의 컬러로 자연스럽게 연결된 것을 알 수 있다.

## 5. 결론

본 논문에서는 3차원 비디오 시스템을 구성하는 중요한 기술인 중간시점 영상합성 방법을 개선하기 위한 공통 빈공간 채움 방법을 제안했다. 공통 빈공간을 채우기 위하여 인접 화소의 정보를 고려하여 채우는 인페인팅 방법을 이용하면, 전경 객체의 컬러 정보가 공통 빈공간에 영향을 주게 되므로 부자연스러운 합성영상을 생성하게 된다. 본 논문에서 제안하는 방법은 공통 빈공간의 깊이값을 예측하고, 예측한 깊이값과 인접한 화소들의 깊이값을 고려하여 공통 빈공간을 채운다. 이 방법을 이용하면 공통 빈공간 근처의 가장 깊이값이 낮은 컬러값으로 공통 빈공간을 채우게 된다. 이 방법을 이용하면 객체의 구성이 복잡하여 공통 빈공간이 많이 발생할 수 있는 영상에서도 자연스러운 합성 영상을 생성할 수 있다.

## 감사의 글

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

## 참고문헌

- [1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, "3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards," IEEE ICME, pp.2161-2164, July 2006.
- [2] ISO/IEC JTC1/SC29/WG11 "Report on Experimental Framework for 3D Video Coding," N11631, Oct. 2010.
- [3] C. Fehn, "Depth-Image-Based Rendering (DIBR), compression and transmission for a new approach on 3D-TV," Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XI, pp.93-104, Jan. 2004.
- [4] ISO/IEC JTC1/SC29/WG11, "View Synthesis Algorithm in View Synthesis Reference Software 2.0 (VSRS2.0)", m16090, Feb. 2009.
- [5] A. Telea, "An image inpainting technique based on the fast marching method", J. Graphics Tools, vol.9, no.1, pp.25-36, May 2004.
- [6] ISO/IEC JTC1/SC29/WG11 "3D Video Test Sequence and Camera Parameters" m16949, Oct. 2009.
- [7] C. Lee and Y.S. Ho, "View synthesis using depth map for 3D video", Proc. 2009 APSIPA Summit and Conference, Oct. 2009