

깊이 카메라 시스템을 이용한 삼차원 실사 동영상 표현방법

김성열* · 호요성*

1. 서론

우리는 정보혁명과 디지털시대로 대변되는 21세기에 살고 있다. 대형 스크린과 입체 음향기기를 갖는 고성능 텔레비전을 통해 영화관에서의 감동을 그대로 느낄 수 있으며, 영상전화를 이용하여 멀리 떨어져 있는 사람들과 통화할 수 있다. 또한, 초고속 네트워크가 가정에 있는 고성능 PC에 연결이 되고 3차원 디스플레이 기술이 발전함에 따라, 3차원 TV나 3차원 게임과 같은 다차원 정보를 이용한 다양한 멀티미디어 응용들이 출현하고 있다. 이러한 다차원 디지털 멀티미디어 응용은 고화질의 영상 서비스를 제공할 뿐 만 아니라, 다시점 장면 재현이나 촉각 체험과 같은 상호작용 기술도 제공한다.

디지털 시대의 핵심은 다양한 멀티미디어 정보를 디지털 형태로 변환하여 전송하는 디지털 방송에 있다고 해도 과언이 아니다. 우리나라에서는 2002년도에 디지털 위성방송이 개시된 이후에 본격적인 디지털 방송 시대로 돌입하게 되었다. 2003년부터 일부 지상파 방송을 디지털화했으며, 현재는 디지털 멀티미디어 방송을 실행하고 있다.

더 나아가 2010년도에는 모든 방송시스템을 아날로그 형태에서 디지털 형태로 전환하는 계획을 세우고 있어 그 어느 때보다도 디지털 방송에 대한 관심이 높아지고 있다.

디지털 방송은 기존의 아날로그 방식에 비해 고화질 방송에 적합하고 다채널 방송이 가능하다. 또한 디지털 방송은 데이터 방송에 유리할 뿐만 아니라, 효율적으로 주파수 대역을 활용할 수 있는 장점이 있다. 그림 1은 2003년에 과학기술부에서 만든 국가과학기술지도(national technology roadmap, NTRM)[1]에서 발췌한 것으로, 방송기술의 전체적인 발전 추세를 나타내고 있다.

그림 1의 그래프에서 보는 것과 같이, 방송기술의 발전 추세는 단순히 시청자가 일방적으로 방송을 시청하는 단방향 형태에서 입체 TV, 실감방송[2]처럼 시청자가 정보를 선택하여 재창조할 수 있는 양방향 또는 대화형 방송으로 변한다.

현재의 오디오 비디오 콘텐츠가 대부분 2차원 정보미디어 형태로 이루어져 있어, 차세대 방송의 요구사항들을 충족시키기에는 무리가 있다. 따라서 우리는 3차원 정보미디어를 효과적으로 표현하는 방법을 개발해야 한다. 일반적으로 3차원 정보미디어를 표현하기 위해서, 기하학과 연결성 정보로 이루어진 메쉬 구조(mesh structure)[3]를 사용한다. 특히, 동적 메쉬 (dynamic meshes)[4], 애니메이션 보간법(animation interpolation)[5],

* 광주과학기술원

※ 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업 (IITA-2005-C1090-0502-0022)인 광주과학기술원 실감방송연구센터의 연구결과로 수행되었음.

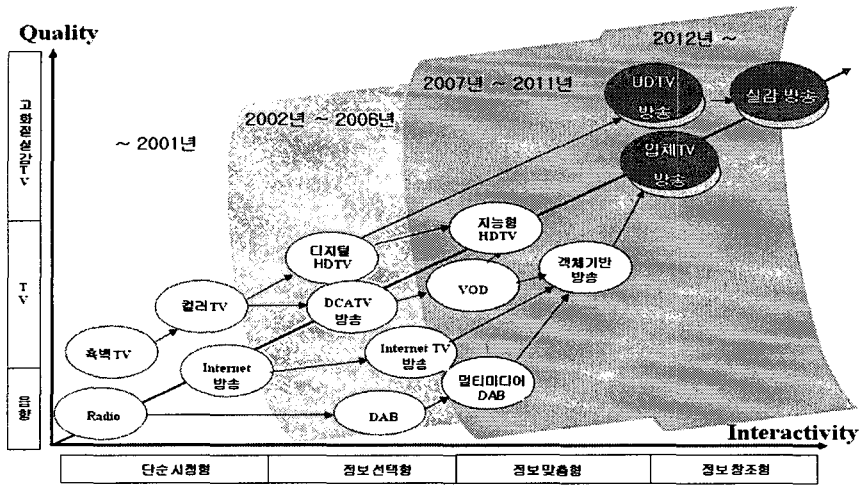


그림 1. 방송기술의 발전 추세

메쉬 격자 표현법(mesh-grid representation)[6]을 이용하여 3차원 동적 장면을 재현할 수 있다.

기존의 3차원 표현기법들은 장면 기반의 표현 방법이 아닌 객체 기반의 생성방법이기 때문에, 3차원 실사 동영상(3-D natural videos)을 연속적으로 구성하기가 상당히 어렵다. 여기서, 3차원 실사 동영상이라 함은 기존의 2차원 동영상과 같이, 자연 영상을 기반으로 하여 프레임마다 연속적으로 3차원 장면을 갱신한 것을 의미한다. 3차원 실사 동영상은 3차원 TV나 실감방송과 같은 차세대 방송 시스템의 방송콘텐츠로 이용가능하다[7].

최근에 3차원 깊이 정보(depth information)를 획득하는데 필요한 광학적 센서 기술이 발달하여, 깊이 카메라(depth camera)[8]를 통해 실제 객체를 직접 촬영하고 픽셀 단위로 깊이 정보를 표현할 수 있게 되었다. 획득한 깊이영상을 이용하여 3차원 전체 모델링(3-D full modeling)을 하기에는 아직까지 기술적인 어려움이 많다. 그럼에도 불구하고, 깊이 카메라로부터 획득한 깊이 정보는 2차원 영상 신호처리의 편리성과 3차원 모델링의 장점을 고루 갖추고 있어, 3차원 실사 동영상을 표현하는데 다른 3차원 신호보다 적합하다.

본 논문에서는 깊이 카메라로부터 획득한 색상 및 깊이 영상 시퀀스를 이용하여 3차원 실사 동영상을 재현하는 새로운 구조를 제안한다. 본 논문에서는 3차원 영상의 품질을 향상시키기 위해, 깊이 영상에 존재하는 잡음과 왜곡을 제거하고 3차원 장면을 계층적 구조로 표현하여 3차원 실사 동영상을 점진적으로 렌더링한다.

본 논문의 2장에서는 깊이 카메라 시스템을 이용하여 색상 및 깊이 영상 시퀀스를 획득하는 과정을 설명한다. 3장에서는 제안한 3차원 실사 동영상 표현방법을 소개한다. 그리고 4장에서는 제안한 방법을 이용하여 실험한 결과를 보여주고, 5장에서 본 논문의 결론을 맺는다.

2. 색상 및 깊이 영상 시퀀스 생성

2.1 깊이 카메라 시스템

본 논문에서는 실제 장면으로부터 색상 및 깊이 영상을 얻기 위해서, 깊이 센서가 부착된 깊이 카메라를 사용한다. 깊이 카메라는 실제 장면의 깊이 정보를 바로바로 포착하여 그레이 레벨로 깊이 영상을 생성한다.

깊이 카메라의 원리는 TOF(time-of-flight) 기술에 기반을 둔다. 적외선 레이저가 링조명기를 통해서 FOI(field-of-illumination)을 형성하고 획득 렌즈(taking lens)가 FOV(field-of-view) 안에 있는 객체의 반사 광선을 모은다. 또한, D 모듈 장치가 들어오는 광선을 가시광선과 적외선으로 나누어 분사시킨다. 가시광선은 RGB 카메라로 보내져 색상 영상을 만들게 되고, 적외선 광선은 D 카메라로 반사되어서 깊이 영상을 생성한다. 깊이 영상을 생성하기 위해서, D 카메라는 실제의 깊이 값을 256 레벨, 즉 8 비트로 양자화하여 그레이 영상 형태로 저장한다. 그림 2는 깊이 카메라의 구성요소를 보여준다.

2.2 색상 및 깊이 영상 시퀀스

본 논문에서는 깊이 카메라로부터 얻은 색상 및 깊이 영상 시퀀스를 이용하여 3차원 실사 동영상을 생성한다. 깊이 영상의 가로 세로축 좌표 값과 그 위치에서의 픽셀 값은 3차원 공간상에서의 기하학 정보로 간주할 수 있다. 또한, 색상 영상 시퀀스는 텍스처 매핑(texture mapping)에서 사용하는 텍스처 영상 시퀀스로 간주하여, 깊이 영상의 기하학 정보로부터 만들어진 3차원 표면을 실제와 같이 렌더링한다.

색상 및 깊이 영상 시퀀스를 이용한 3차원

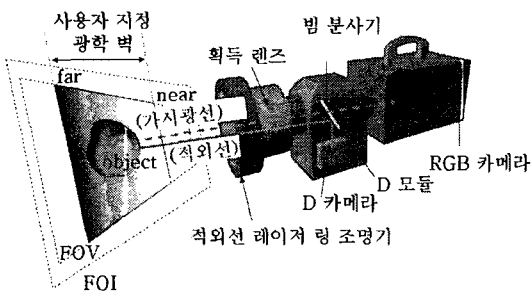


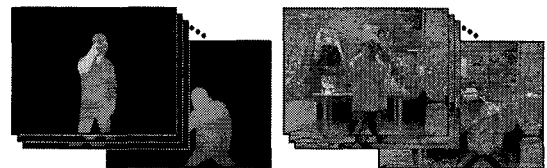
그림 2. 깊이 카메라 시스템 구성도

모델링 기술은 복잡한 자연 영상도 그대로 3차원으로 재현할 수 있다. 또한, 텍스처 매핑을 수행할 때, 깊이 영상의 가로 세로축 좌표 값을 텍스처 좌표(texture coordinate)로 이용하기 때문에, 별도로 텍스처 좌표를 고려하지 않아도 된다. 그림 3은 깊이 카메라로부터 획득한 색상 및 깊이 영상 시퀀스의 예를 보여준다.

2.3 색상 정보를 이용한 깊이 정보 향상

깊이 카메라로부터 획득한 깊이 영상은 8비트의 제한적인 해상도를 갖기 때문에, 기본적으로 양자화 오류를 포함한다. 게다가 깊이 영상은 적외선 센서를 통해서 깊이 정보를 획득하는 방식을 채택하기 때문에, 객체의 반사광, 색상 편차 등의 고주파 잡음과 형태 왜곡도 포함한다. 따라서 이러한 잡음이나 왜곡을 제거하여 깊이 정보의 품질을 향상시켜야 한다.

깊이 정보의 품질을 향상시키기 위해서, 깊이 영상기반 렌더링(depth image-based rendering, DIBR)[9]과 색상 정보를 이용한 경계 추출법을 사용한다. DIBR를 이용하여 깊이 정보를 향상시킨 후[10], 깊이 카메라로부터 획득한 색상 정보를 이용하여 깊이 영상 안에 있는 객체의 형태를 보정한다. 광학적 렌즈의 물리적 한계로, 깊이 카메라에서 정확한 객체의 경계를 추출하지 못하는 경우가 발생한다. 이러한 문제를 해결하기 위해서, 본 논문에서는 색상 영상의 객체 경계를 추출



(a) 깊이 영상 (b) 색상 영상

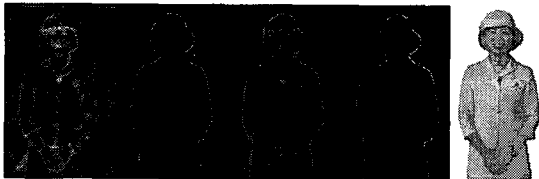
그림 3. 색상 및 깊이 영상 시퀀스의 예

하여 깊이 영상을 향상시키는 방법을 제안한다.

그림 4에서와 같이, 먼저 색상 영상과 깊이 영상에 경계 필터인 Sobel 필터를 수직 수평방향으로 적용하여 객체 경계를 찾는다. 그림 4(a)와 그림 4(b)는 색상 영상과 깊이 영상에 Sobel 필터를 적용한 결과이다. 또한, 그림 4(c)와 같이, 색상 영상으로부터 얻은 객체 경계를 보정하기 위해서, 깊이 영상으로부터 얻은 객체 경계를 색상 영상의 초기 객체 경계로 간주한다. 그런 다음, 식 1로 정의된 에너지 함수를 이용하여 초기 객체 경계를 변화시킴으로써 정확한 객체 경계를 찾아낸다.

$$E_i = \frac{1}{10} \sum_{j=i-5}^{i+4} (P_j - P_{j+1})^2 \quad (1)$$

식 1은 초기 객체 경계 위에 있는 픽셀에 대해 주변 10개의 이웃 픽셀 값을 사용하여 최적의 경계 위치를 구하는 것을 의미한다. E_i 는 i 픽셀의



(a) (b) (c) (d) (e)

그림 4. 깊이 정보 향상

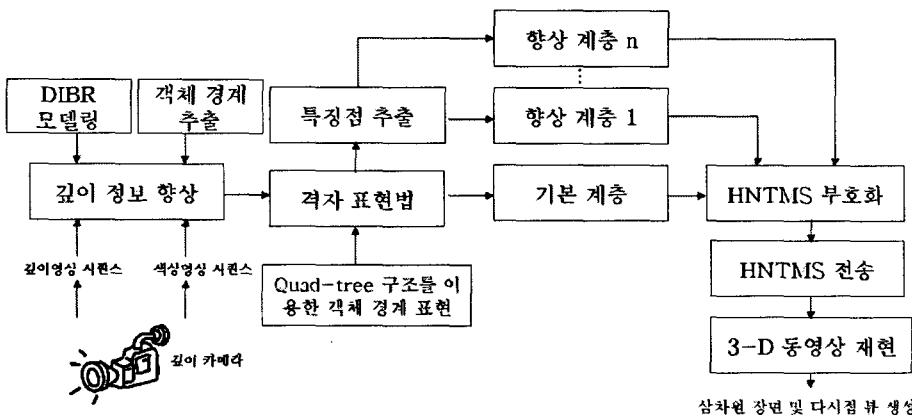
에너지 값을 의미하고 P_j 와 P_{j+1} 은 i 픽셀이거나 이웃 픽셀을 나타낸다. 이와 같은 과정을 통해 에너지가 큰 픽셀의 위치를 찾아 객체의 경계로 간주한다. 그림 4(d)와 그림 4(e)는 찾아낸 객체 경계와 텍스처 매핑한 결과를 보여준다.

3. 계층적 자연 텍스처 메쉬 스트림

본 논문은 깊이 카메라 시스템을 기반으로 하여 3차원 실사 동영상을 표현하는 새로운 구조인 계층적 자연 텍스처 메쉬 스트림 (hierarchical natural-textured mesh stream, HNTMS)을 제안한다. 그림 5는 HNTMS의 전체적인 시스템 흐름도를 보여준다.

기본적으로 HNTMS는 깊이 카메라로부터 얻은 색상 및 깊이 영상 시퀀스를 입력으로 받는다. 그런 다음, 격자 표현법(grid representation)과 특징점 추출(feature points extraction) 과정을 통해서 3차원 장면을 구조적으로 표현한다. HNTMS는 기본계층과 향상계층을 갖는 다해상도의 계층적 구조로, 3차원 실사 동영상을 주변의 상황에 따라 적응적으로 렌더링할 수 있다.

HNTMS는 기존의 3차원 다해상도 표현법과 구별된다. 먼저 Hoppe가 제안한 점진적 메쉬



삼차원 장면 및 다시점 뷰 생성

그림 5. HNTMS의 전체 흐름도

(progressive mesh)[11]와 HNTMS는 구조적인 차이를 지닌다. 점진적 메쉬가 다해상도의 3차원 장면을 표현하는 개념적인 내용은 유사하지만, 실사 기반의 3차원 모델링을 위해 깊이 및 색상 영상 시퀀스를 사용하지 않았다. 또한, HNTMS는 구조적으로 계층의 수(number of layer, NOL), 기본계층, 향상계층을 두어 점진적 렌더링이 가능하도록 하였다. 또한, HNTMS는 계층적 깊이 영상(layered depth image, LDI)[12]과도 구별된다. LDI는 영상 기반의 다시점 생성방법이지만, HNTMS는 메쉬 기반의 생성방법이다.

3.1 격자 표현법

본 논문에서는 색상 및 깊이 영상으로부터 격자 표현법을 이용하여 초기의 3차원 표면을 정의한다. 사용자가 정의한 격자 셀(grid cell)의 크기(8×8 또는 16×16)에 따라, 3차원 기하학 정보를 깊이 영상으로부터 찾아내고 메쉬 삼각화 과정을 거쳐서 초기 3차원 장면을 구성한다.

우리는 격자 표현법을 3차원 실사 동영상에 필요한 최소한의 정보라고 간주하여 이를 기본계층(base layer)으로 한다. 격자 표현법은 3차원 영상을 단순하고 편리하게 처리할 수 있도록 규칙성(regularity)을 제공한다. 일반적으로 3차원 장면을 표현하기 위해 불규칙한 3차원 그래픽 정보로부터 규칙성을 찾고 최소한의 표현 정보를 정하는 것은 여간 어려운 것이 아니다. 그러나 깊이 카메라로부터 얻은 색상 및 깊이 영상은 3차원 특성과 일반적인 2차원 영상의 특성을 동시에 포함하고 있어 균등하게 영상 영역을 나눌 수 있다.

격자 셀은 격자 표현법의 기본 단위이다. 따라서 $x_{res} \times y_{res}$ 의 해상도를 갖고 $p \times q$ 픽셀을 격자 셀의 크기로 정의한 색상 및 깊이 영상에 대해,

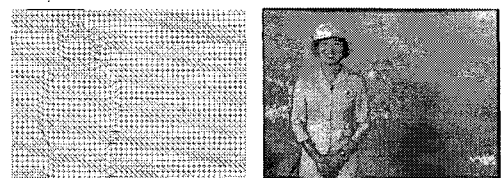
총 격자 셀의 개수는 $x_{res}/p + y_{res}/q$ 이고, 삼각형 총 개수는 $2 \times (x_{res}/p + y_{res}/q)$ 이다. 기본적으로 격자 표현법은 깊이 영상을 격자 셀의 크기에 따라 블록 형태로 나눈다. 그런 다음, 반시계 방향으로 격자 셀의 모서리 픽셀 4개를 찾는다. 따라서, 찾아낸 모서리 픽셀 4개에 대해 다음과 같이 3차원 꼭지점을 정의할 수 있다.

$$\begin{aligned} v_1 &= \{x_1, y_1, z_1, v_{1index}\}, v_2 = \{x_2, y_2, z_2, v_{2index}\} \\ v_3 &= \{x_3, y_3, z_3, v_{3index}\}, v_4 = \{x_4, y_4, z_4, v_{4index}\} \end{aligned} \quad (2)$$

여기서, $v_{1index}, v_{2index}, v_{3index}, v_{4index}$ 는 해당 꼭지점의 꼭지점 인덱스를 나타낸다. 새로운 꼭지점이 꼭지점 정보를 저장해놓은 메모리 버퍼에 입력되면, 중복되어지는 꼭지점 인덱스와 위치 정보를 피하기 위해 해당 버퍼를 갱신한다. 격자 표현법은 정의한 v_1, v_2, v_3, v_4 을 이용하여 2개의 삼각형을 만든다. 한 삼각형은 꼭지점 $\{v_1, v_2, v_3\}$ 로 구성되고, 다른 삼각형은 꼭지점 $\{v_1, v_3, v_4\}$ 로 이루어진다. 마지막으로 해당 꼭지점 정보와 연결성 정보를 HNTMS 구조에 할당한다. 그림 6은 3차원 장면을 격자 표현법으로 3차원 장면을 모델링한 예를 보여준다.

3.2 특징점 추출

격자 표현법만으로 고품질의 영상 서비스를 충분히 제공할 수 없기 때문에, 본 논문에서는 격자 셀에 존재하는 특징점 정보를 추가하여 3차원 실사 동영상을 표현한다. 여기서, HNTMS에 삽입



(a) 와이어프레임 (b) 텍스처를 입힌 결과

그림 6. 격자 표현법

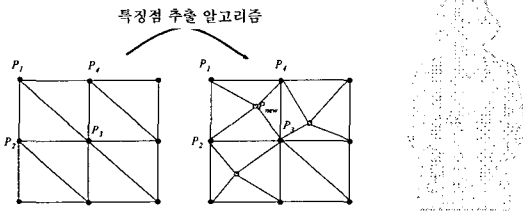


그림 7. 특징점 추출 및 메쉬 삼각화

된 특징점을 향상계층(enhancement layer)으로 간주한다. 이와 같이, 매 프레임마다 특징점을 추출하여 정밀한 삼각형 메쉬를 생성함으로써 HNTMS의 구조를 완성할 수 있다.

특징점을 추출하기 위해, 다음과 같은 최대 거리 알고리즘을 수정하여 사용한다.

1. 격자 셀에 초기 부모 픽셀 집합 $\{P\}$ 를 선택한다. 처음 격자 셀로부터 특징점을 추출하면 초기 부모 픽셀 집합 $\{P\}$ 는 그 셀에 있는 4개의 모서리 픽셀이다.
2. 격자 셀의 모든 픽셀에 대해서 깊이값 편차를 계산한다. 만약 깊이값 편차가 임계값 T 보다 작으면, 단계 1로 간다. 그렇지 않으면 단계 3로 간다.
3. 식 3를 이용하여 $\{P\}$ 와 격자 셀에 있는 픽셀 $x_{i(<N)}$ 의 거리를 계산한다. 이 과정은 격자 셀에 있는 픽셀 개수인 N 번만큼 반복 수행한다.

$$D_i = \frac{1}{n\{P\}} \sum_{j=0}^{n\{P\}} (x_i - p_j)^2 \quad (3)$$
4. D_i 중에 최대거리를 갖는 픽셀 p_{max} 를 선택하고, 이를 부모 픽셀 집합 $\{P\}$ 에 추가하여 새로운 집합 $\{P'\}$ 를 만든다.
5. $\{P'\}$ 를 이용하여 메쉬 삼각화 과정을 다시 수행한다.

추가한 특징점 p 도 깊이 영상에 있는 정보이기 때문에, 이를 3차원 공간에서의 꼭지점으로 간주한다. 결과적으로, 추출된 특징점 p_i 는 다음과 같이 정의할 수 있다.

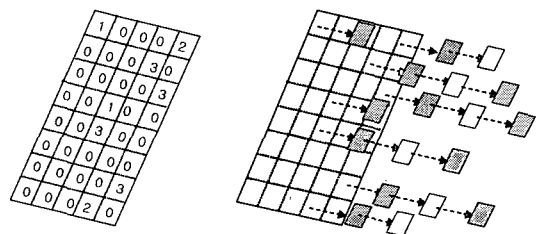
$$p_i = \{x_i, y_i, z_i, p_{iIndex}\} \quad (4)$$

여기서, p_{iIndex} 는 해당 꼭지점의 인덱스 정보를 의미한다. p_i 가 향상계층을 표현하는 큐 구조에 입력되면, 추가된 꼭지점 p_i 로 인하여 큐가 갱신된다. 그런 다음, 큐 안에서 p_i 로부터 가장 가까운 4개의 꼭지점 v_1, v_2, v_3, v_4 을 찾는다. 그리고 v_1, v_2, v_3, v_4 와 p_i 를 이용하여 4개의 삼각형을 구성한다. 생성되는 삼각형은 각각 $\{p_i, v_i, v_2\}, \{p_i, v_2, v_3\}, \{p_i, v_3, v_4\}, \{p_i, v_4, v_1\}$ 이다.

3.3 HNTMS의 구조

HNTMS는 프레임마다 NOL, 기본계층, 향상계층으로 이루어진 계층적 구조 (hierarchical structure)를 갖는다. 3.1절의 격자 표현법을 이용하여 매 프레임마다 기본계층을 만든다. 그런 다음, 격자 표현법의 격자 셀 안에서 특징점을 계산하여 여러 개의 향상계층을 만든다. 그림 8은 한 프레임에 대한 HNTMS 구조를 도시화한 것이다.

NOL은 격자 셀에 포함된 향상계층의 개수를 나타낸다. 만약 격자 셀의 NOL 정보가 n 이면, 해당 격자 셀은 하나의 기본계층과 n 개의 향상계층을 가진다. 마찬가지로 격자 셀의 NOL 정보가 0이면, 단지 기본계층만으로 해당 격자 셀을 표현한다. 한 프레임에서의 HNTMS의 구조는 다음과 같은 개념적 표현으로 요약할 수 있다.



(a) NOL (b) 기본계층과 향상계층

그림 8. 한 프레임에서의 HNTMS 구조

항상계층 =

깊이 영상의 특징점 X, Y 좌표값: 32 비트 정수형

깊이 영상의 특징점 깊이 값: 8 비트 정수형

격자 셀 =

NOL : 8 비트 정수형

4개의 모서리 픽셀 위치: 32 비트 정수형

4개의 모서리 깊이 값: 8 비트 정수형

제층[0 ... NOL-1] : 항상계층의 배열

HNTMS =

격자 셀 크기 : 8 비트 정수형

격자[0...xres/격자 셀 크기-1, 0...yres/격자 셀

크기-1] : 격자 셀의 배열

여기서 xres와 yres는 깊이 영상의 x축 해상도와 y축 해상도를 나타낸다.

그림 9는 모든 프레임에서의 HNTMS의 구조를 개념적으로 도시화한 것이다.

3.4 HNTMS의 품질 평가

HNTMS의 시각적 품질을 평가하기 위해서, 본문에서는 수정된 Hausdroff 거리를 제안한다. 일반적으로, Hausdroff 거리[13]는 3차원 모델사이의 왜곡 정도를 측정하는데 널리 이용된다. 수정된 Hausdroff 거리는 HNTMS에 의해 재구성된 3차원 장면과 해당 깊이 영상으로부터 3차원 전체 모델링한 결과와의 왜곡정도를 구한다.

프레임 0	프레임 1	프레임 2	프레임 3	...	프레임 N
NOL	NOL	NOL	NOL	...	NOL
기본	기본	기본	기본	...	기본
항상 1	항상 1	항상 1	항상 1	...	항상 1
항상 2	항상 2	항상 2	항상 2	...	항상 2
...
항상 k	항상 m	항상 k	항상 n	...	항상 k

그림 9. 전체적인 HNTMS 구조

깊이 영상의 한 픽셀의 깊이 값과 HNTMS에 의해 재구성한 3차원 장면의 깊이 값을 비교하여, 가장 가까운 거리에 위치한 깊이 픽셀은 식 5에 의해서 결정된다.

$$d(p, S') = \min_{p' \in S'} \|p - p'\| \quad (5)$$

여기서, $d(p, S')$ 는 재구성된 3차원 장면 S' 와 깊이 픽셀 p 사이의 거리를 의미한다. 또한, p' 는 재구성된 3차원 장면 S' 에 포함하는 깊이 픽셀을 나타낸다. 결과적으로, 식 5를 이용하여 식 6과 같이, 수정된 Hausdroff 거리를 정의할 수 있다.

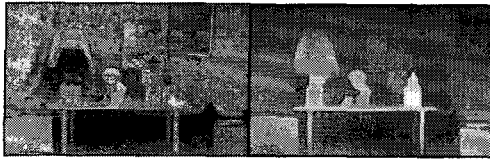
$$d(S, S') = \max_{p \in S} d_z(p, S') \quad (6)$$

여기서, S 는 깊이 영상을 이용하여 전체 모델링한 3차원 장면이고, $d_z(p, S')$ 는 S' 에 속한 깊이 픽셀 p' 와 선택된 최소거리의 깊이 픽셀 사이의 깊이 거리를 나타낸다.

4. 실험 결과 및 분석

제안한 방법을 실험하기 위해서, ZcamTM이라고 하는 깊이 카메라로 실험 시퀀스를 획득하여 사용하였다. 실험에 사용한 시퀀스는 그림 10(a)의 Background 영상과 그림 10(b)의 Home-shopping 시퀀스이고, 모두 색상 및 깊이 영상으로 이루어져 있다. Home-shopping 시퀀스는 총 100 프레임으로 720×486의 영상 해상도로 갖는다. Background는 한 프레임으로 이루어졌으며 해상도는 720×486이다.

Home-shopping 시퀀스는 실감방송을 위한 방송콘텐츠를 제작하기 위해 2005년에 광주과학기술원 실감방송연구센터(Realistic Broadcasting Research Center, RBRC)[14]에서 제작한 것이다. Home-shopping 시퀀스는 2005년 한국정보통신진흥원에서 지원한 ITRC 포럼에 전시되었다. 그



(a) Background



(b) Home-shopping

그림 10. 실험데이터

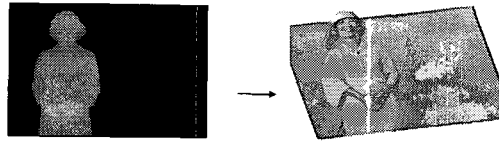
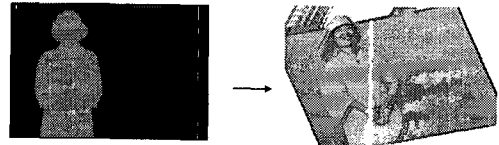
리고 Background 영상은 한국전자통신연구원에서 3차원 TV 콘텐츠 연구를 위해 제작한 것이다.

4.1 깊이 정보 향상

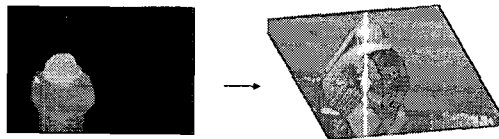
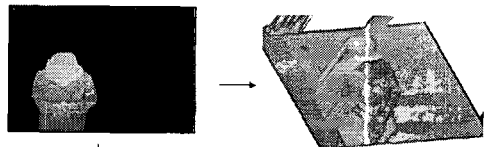
그림 11는 원 깊이 영상과 향상된 깊이 영상을 보여준다. 그림 11(a)와 그림 11(b)는 향상된 깊이 영상의 결과로, Home-shopping 시퀀스의 35번째 프레임과 64번째 프레임을 보여준다. 원 깊이 영상을 이용하여 3차원 렌더링을 한 결과와 향상된 깊이 영상을 이용하여 3차원 전체 렌더링을 한 결과를 비교하면, 원 깊이 영상에 존재한 광학적 잡음이 감소되었다는 것을 알 수 있다. 그림 11에서 볼 수 있듯이, 고주파 성질을 갖는 광학적 잡음이 감소하였고, 전체적으로 3차원 표면이 부드러워졌음을 알 수 있다.

4.2 HNTMS 표현

그림 12는 HNTMS의 기본계층을 의미하는 격자 표현법을 보여준다. 격자 셀의 크기는 16×16 픽셀로 정하였다. 격자 표현법 후에 실험 시퀀스에 대해서 두 개의 향상계층을 생성하였다. 격자



(a) Home-shopping의 35번째 프레임



(b) Home-shopping의 64번째 프레임

그림 11. 향상된 깊이 영상의 실험 결과

표현법으로 3차원 실사 동영상을 표현하는 최소한의 정보를 제공할 수 있었다. 즉, 멀티미디어 플랫폼의 디스플레이 크기가 작거나 컴퓨팅 파워가 낮아 연산속도가 제한되어 고품질의 서비스가 힘들 경우, 격자 표현법으로 최소한의 품질을 유지할 수 있었다. 그림 13은 기본계층과 두개의 향상계층으로 3차원 장면을 표현한 결과를 보여준다. 전송되는 향상계층의 수에 따라 시각적 품질을 조절할 수 있음을 알 수 있었다.

4.3 다시점 영상 생성

3차원방송 시스템에서 시청자들은 다양한 시점에서의 영상을 보기를 원한다. 본 실험에서는 HNTMS 구조를 이용하여 재구성한 3차원 실사

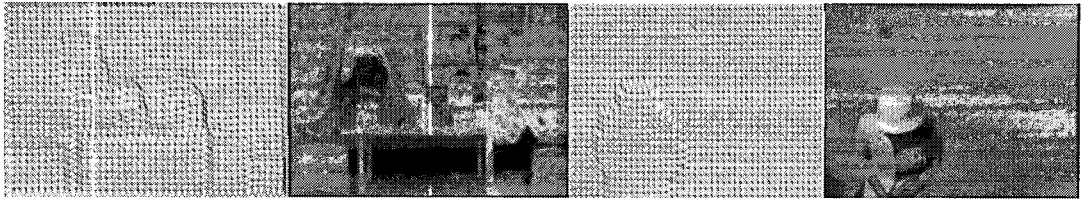


그림 12. HNTMS의 기본 계층 표현

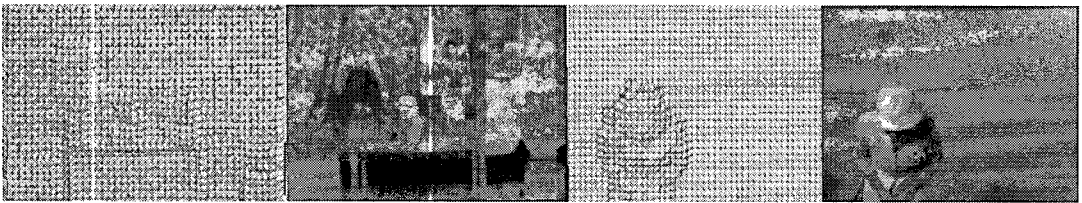


그림 13. HNTMS의 향상 계층 표현

동영상에 대해, 가상 카메라의 위치를 조정하여 다시점 영상을 생성하였다. 사용자가 볼 수 있는 시점영역에 한계가 어느 정도 있지만, 자연 영상으로부터 다시점 영상을 생성할 수 있었다. 그림 14는 HNTMS를 통해 렌더링한 후, 다시점 영상을 생성한 결과를 보여준다. 3차원으로 재구성된 장면으로부터 가상 카메라를 X축 방향으로 -10도에서 +10도까지 이동하여 다시점 영상을 생성하였다. 그림 14(a), 그림 14(b), 그림 14(c)는 각각

-10도, 0도, +10도의 결과를 보여준다.

4.4 렌더링 속도 및 시각 품질 분석

3차원방송 시스템에서 3차원 실사 동영상을 이용하기 위해서는 실시간 렌더링이 필수적이다. 본 실험에서는 HNTMS의 렌더링 속도와 시각 품질을 분석하였다. 표 1은 실험 시퀀스를 가지고 3차원 전체 모델링한 결과와 HNTMS를 이용한 모델링한 결과의 렌더링 시간을 비교한 것이다.

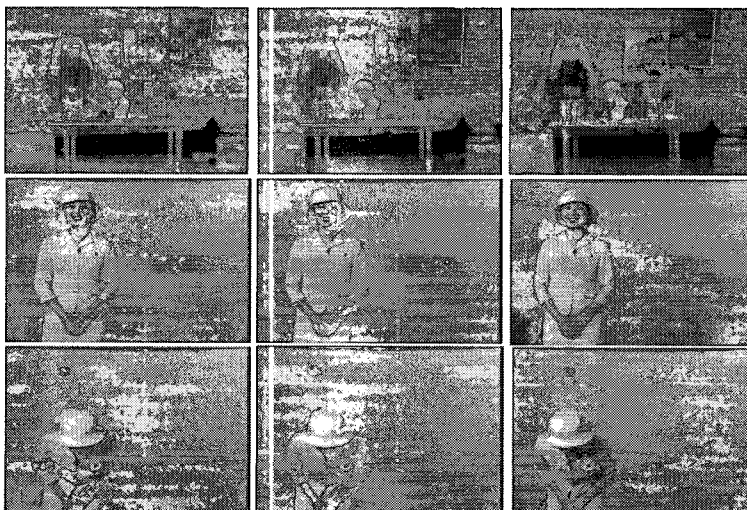


그림 14. 다시점 생성

표 1에서 볼 수 있듯이, 3차원 전체 모델링을 이용하면 고품질의 시각 품질은 제공할 수 있지만, 3차원 동적 장면을 실시간으로 렌더링하는 것은 불가능하다. 하지만, HNTMS를 사용하면 신뢰할 만한 시각 품질을 제공하면서 실시간 렌더링이 가능하다. 특히, Home-shopping 시퀀스의 경우, 호스트를 제외한 대부분의 배경부분은 비슷한 깊이 값을 나타내기 때문에, HNTMS를 이용하면 보다 효과적으로 렌더링할 수 있다.

본 실험에서는 완전 모델링한 결과와 HNTMS를 이용한 결과의 시각 품질도 비교하였다. 왜곡 정도를 실험 평가하기 위해서, Home-shopping 시퀀스로 실험하였고 수정된 Hausdroff 거리를 사용하였다. 그림 15는 수정된 Hausdroff 거리를 이용하여 시각 품질을 평가한 그래프를 보여준다.

그래프 수치는 Hausdroff 거리 값으로 깊이영상의 최대 픽셀 값 차이를 의미한다. 그림 15에서 보는 것과 같이, 기본계층과 2개의 확장계층을 보내면 완전 모델링과 거의 같은 품질을 가졌다. 기본계층만으로 표현한 3차원 동적 장면은 왜곡도가 심하지만 향상계층이 추가하면 할수록 시각 품질의 결과가 좋았다.

또한, Home-shopping 시퀀스의 3차원 시각 품

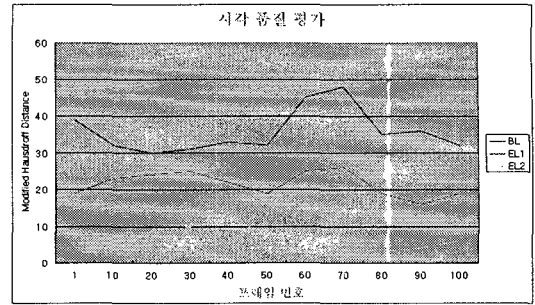


그림 15. 시각 품질 평가

질은 프레임마다 다양하였다. 특히, 60~75 프레임에서는 더 많은 3차원 장면 왜곡이 일어났다. 그 이유는 해당 프레임에서 깊이 값의 편차가 컸기 때문이다. 결과적으로, 렌더링 속도와 시각 품질 평가를 통해서 적은 정보량으로 고품질의 3차원 실사 동영상을 실시간 렌더링할 수 있었다.

5. 결론

차세대 방송 시스템에서 다시점 영상 서비스와 같은 다양한 상호작용을 지원하기 위해서는 3차원 실사 동영상을 쉽게 표현하고 생성해야 한다. 본 논문에서는 깊이 카메라 시스템을 기반으로 한 계층적 자연 텍스처 매쉬 스트림(HNTMS)을 제안하였다. 실사의 색상 및 깊이 영상 시퀀스를 얻기 위해 깊이 카메라 시스템을 이용하였고, 객체 경계 추출법을 제안하여 깊이 정보를 향상시켰다. 또한, 격자 표현법과 특징점 추출법을 통해서 HNTMS 구조를 완성하여 3차원 실사 동영상을 효과적으로 표현하였다. 뿐만 아니라, HNTMS를 통해 고품질의 3차원 실사 동영상을 실시간으로 렌더링하였고 다시점 영상도 쉽게 생성할 수 있었다. 제안한 방법은 규칙성을 제공하면서 3차원 실사 동영상을 구조적으로 쉽게 표현할 수 있을 뿐만 아니라, 사용자의 수용 능력과 네트워크 상황에 따라 시각 품질을 적응적으로 조절하였다.

표 1. HNTMS와 완전 모델링 렌더링 속도 비교

실험 데이터		삼각형 수	한 장면 렌더링 시간(초)	프레임율 (프레임/초)
BG.	완전 모델링	344,922	1.532	0.65
	기본계층	3202	0.027	37.04
	기본+확장 1	13802	0.065	18.18
	기본+확장 2	14204	0.078	12.82
H.S.	완전 모델링	344,922	1.532	0.65
	기본계층	2699	0.022	45.45
	기본+확장 1	3065	0.035	28.57
	기본+확장 2	3966	0.042	23.81

참 고 문 헌

[1] Ministry of Science and Technology in Korea, National Technology Roadmap, 2003.

[2] S.Y. Kim, S.U. Yoon, Y.S. Ho, "Realistic Broadcasting Using Multi-modal Immersive Media," Lecture Notes in Computer Science, vol. 3768, pp. 164-175, 2005.

[3] G. Taubin, J. Rossignac, "Geometry Compression through Topological Surgery," SIGGRAPH, pp. 84-115, 1998.

[4] J.H. Ahn, C.S. Kim, C.C. Kro, Y.S. Ho, "Motion-compensated Compression of 3D Animation Models," Electronics Letters, vol. 37, pp. 1445-1446, 2001.

[5] E.S. Jang, D.K. Kim, S.Y. Jung, M.J. Han, S.O. Woo, S.J. Lee, "Interpolator Data Compression for MPEG-4 Animation," IEEE Trans. Circuits and Systems for Video Technology, vol. 14, no. 7, pp. 989-1008, 2004.

[6] ISO/IEC JTC1/SC29/WG11 m12061, "Encoding of Dynamic Meshes with MeshGrid," 2005.

[7] J. Cha, S. M Kim, S. Y. Kim, S. Kim, S. U. Yoon, I. Oakley, J. Ryu, K. H. Lee, W. Woo, Y. S. Ho, "Client System for Realistic Broadcasting: A First Prototype" Lecture Notes in Computer Science, vol. 3768, pp. 176-186, 2005.

[8] ZCam™ <http://www.3dvsystems.com>, 2005

[9] H. Shum, S. Kang, "A Review of Image-based Rendering Techniques," Proc. SPIE Visual Communication and Image Processing, pp. 2-13, 2000.

[10] S.M Kim, J. Cha, J. Ryu, K.H. Lee, "Depth Video Enhancement of Haptic Interaction Using a Smooth Surface Reconstruction," IEICE Trans. Information and Systems, vol. E89-D, no. 1, pp. 37-44, 2006.

[11] H. Hoppe, "Progressive Meshes," SIGGRAPH, pp. 99-108, 1996.

[12] J. Shade, S. Gotler, R. Szeliski, "Layered Depth Images," SIGGRAPH, pp.291-298, 1998.

[13] M. Eck, T. DeRose, T. Duchamp, H. Hoppe, M. Lounsbery, W. Stuetzle, "Multi-resolution Analysis of Arbitrary Meshes," SIGGRAPH, pp. 173-182, 1995.

[14] Realistic Broadcasting Research Center, <http://rbrc.gist.ac.kr>, 2003.



김 성 열

- 2001년 강원대학교 정보통신공학 학사
- 2003년 광주과학기술원 정보통신공학 석사
- 2001년~현재 광주과학기술원 정보통신공학 박사과정
- 관심분야 : 디지털 신호처리, 3-D 메쉬 압축, 3-D TV



호 요 성

- 1981년 서울대학교 전자공학 학사
- 1983년 서울대학교 전자공학 석사
- 1990년 Univ. of California at Santa Barbara, 전기컴퓨터공학 박사
- 1990~1993 미국 Philips 연구소 선임연구원
- 1983~1995 한국전자통신연구원 선임연구원
- 1995~현재 광주과학기술원 교수
- 관심분야 : 디지털 신호처리 및 방송, 영상 압축, 멀티미디어 시스템, 3-D TV 및 실감방송