

# 깊이영상 부호화를 위한 양방향 예측 방법

오관정<sup>o</sup>, 호요성

광주과학기술원 정보통신공학과

{kjoh81, hoyo}@gist.ac.kr

## 요약

깊이는 객체와 카메라 간의 거리를 나타내는 3차원 데이터로 3차원 비디오 시스템에서 가상영상 생성 시 이용된다. 일반적으로 전체 깊이 공간을 256개의 양자화 된 값으로 나타내는 8비트의 영상으로 표현되며, 이 경우 깊이와 변이의 상관관계를 고려하여 변이 및 화소 영역에서는 균등하게 양자화되고, 깊이 영역에서는 비균등 방법으로 양자화된다. 따라서, 기존의 양방향 예측 방법은 선형적으로 움직이는 깊이영상에 대해 정확한 값을 예측할 수 없다. 본 논문에서는 이러한 문제점을 해결하기 위해 깊이영상 부호화를 위한 양방향 예측 방법을 제안한다. 제안하는 방법은 깊이와 변이간의 변환을 통해 깊이 영역에서 양방향 예측을 수행함으로써 기존의 양방향 예측의 문제점을 해결했다. 실험을 통해 제안하는 방법이 0.37 dB의 부호화 효율 향상과 0.02 dB 정도의 합성 영상 화질 개선 효과가 있음을 확인했다.

## 1. 서론

3차원 비디오는 입체감과 자유시점을 동시에 제공할 수 있는 비디오 기술로 일반적으로 고품질 비디오와 함께 차세대 비디오 기술로 각광받고 있다. 3차원 비디오는 흔히 다시점 비디오와 다시점 깊이 영상으로 구성되고, 향후 3DTV, 자유시점 TV, 실감형 원격 회의 등에 활용될 것으로 기대된다.

깊이영상은 색상 영상에서 찍힌 3차원 공간을 양자화하여 이미지 형태로 표현한 것이다. 깊이영상의 주된 목적은 3차원 비디오 시스템의 수신단에서 송신단에서 전송하지 않은 시점에 대한 재현이 필요한 경우 다시점 비디오 영상과 함께 가상 시점 영상을 생성할 때 3차원 정보로 활용되는 것이다. 일반적으로 깊이영상은 다시점 비디오로부터 스테레오 정합을 통해 계산적으로 구해지거나 혹은 깊이 카메라를 이용하여 획득된다.

최근 MPEG의 3DV [1] 그룹에서 3차원 비디오에 관한 국제 표준화 작업과 함께 깊이영상 부호화를 포함한 다양한 3차원 비디오 처리 기술에 대한 연구가 활발하게 진행되고 있다. 현재는 다양한 3차원 비디오 데이터의 모집, 깊이영상 획득, 가상 시점 영상 생성에 관한 연구가 활발히 진행되고 있고, 향후 깊이영상 및 이를 이용한 다시점 비디오 부호화에 대한 연구가 진행될 예정이다.

깊이영상의 부호화 방식은 크게 다시점 비디오와의 상관관계를 이용하는 방법과 독립적으로 깊이 영상만을 부호화 하는 방법이 있다. 전자의 경우 깊이영상과 다시점 비디오 사이의 데이터의 중복성에 기반하여 움직임 벡터를 공유하거나 깊이영상을 이용하여 가상영상을 생성한 후 이를 다시점 비디오 부호화에 이용하는 방법등이 있다.

깊이영상을 독립적으로 부호화 하는 경우는 깊이영상의 부호화 및 가상영상 생성 시 민감한 영역인 객체의 경계 부분을 효율적으로 부호화하기 위해 쿼드트리를 이용하여 객체의 경계를 표현함으로써 합성 영상의 화질을 향상 시키는 방법이 제안되었다 [2].

본 논문에서는 기존의 방법과 달리 깊이영상에서 깊이의 고유한 특성을 고려한 부호화 방법을 제안한다. 제안하는 방법은 비디오 부호화 시 이용되는 양방향 예측 기술이 깊이영상에 적용되는 경우 깊이 영역에서 비균등하게 양자화되는 특성 때문에 예측 효율이 떨어지는 점에 착안하여 이를 보완하기 위해 깊이 영역에서 양방향 예측을 수행하는 방법이다. 이를 위해 깊이와 변이사이의 변환을 이용하여 선형적으로 움직이는 객체에 대해 양방향 예측이 적용되는 경우 그 정확도를 향상시켜 부호화 효율을 증진시킨다.

## 2. 깊이영상의 표현

깊이영상은 색상 영상에 대해 각 객체와 카메라 간의 거리를 8 비트의 이미지 형태로 표현한 데이터이다. 그림 1은 "Mobile" 영상에 대한 색상 영상과 깊이영상을 보여주고 있다.



그림 1. "Mobile" 영상의 색상 영상과 깊이영상

깊이영상은 3차원 공간을 양자화하여 256개의 값으로 표현한 것인데, Chai[3]가 제안한 plenoptic sampling 이론에 기반한 최적의 비균등화 양자화 방법이 통용되고 있다. 이 방법은 스테레오 영상에서 변이와 깊이의 상관관계를 기반으로 최적의 양자화 방법을 유도한 것이다. 먼저 식 (1)과 같이 양자화하고자 하는 3차원 공간의  $Z_{near}$ 와  $Z_{far}$ 에 대한 최대 변이와 최소 변이를 계산하여 전체 변이의 영역을 계산한다. 그림 2는 3차원 공간상에서 깊이에 대해 비균등화 양자화를 한 경우 깊이와 변이의 상관관계를 보여주고 있다.

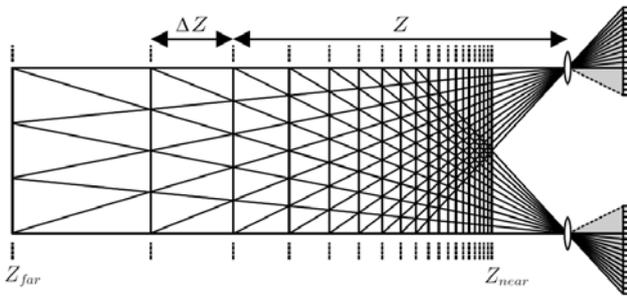


그림 2. 깊이에 대한 비균등 양자화

$$\Delta d_{total} = d(Z_{near}) - d(Z_{far}) \quad (1)$$

여기서  $d(Z_{near})$ 와  $d(Z_{far})$ 는 각각 최대 변이와 최소 변이값이다. 식 (1)을 바탕으로 단위 변위는 다음과 같이 계산된다.

$$\Delta d = \frac{\Delta d_{total}}{N-1} \quad (2)$$

깊이영상의 화소값  $v$ 에 대한 변이값은 식 (3)과 같이 계산된다.

$$d_v = d(Z_{near}) - (N-1-v) \cdot \Delta d \quad (3)$$

식 (3)의  $d_v$ 를 식 (1)과 식 (2)를 이용하여 정리하면, 식 (4)와 같다.

$$d_v = d(Z_{far}) - \frac{v}{N-1}(d(Z_{near}) - d(Z_{far})) \quad (4)$$

식 (4)를 변이와 깊이의 상관관계인 식 (5)를 이용하여  $Z_v$ 에 관해서 풀면, 식 (6)과 같다. 식 (5)에서  $f$ 는 초점 거리를  $B$ 는 두 카메라 사이의 거리를 의미한다.

$$d = \frac{f \cdot B}{Z} \quad (5)$$

$$Z_v = \frac{1}{\frac{v}{N-1} \cdot \left( \frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}}} \quad (6)$$

식 (5)와 (6)으로부터 깊이영상의 화소값  $v$ 는 식 (7)이나 식 (8)과 같이 정의된다.

$$v = (N-1) \cdot \frac{d_v - d(Z_{far})}{d(Z_{near}) - d(Z_{far})} \quad (7)$$

$$v = (N-1) \cdot \frac{\frac{1}{Z_v} - \frac{1}{Z_{far}}}{\frac{1}{Z_{near}} - \frac{1}{Z_{far}}} \quad (8)$$

## 3. 깊이영상을 위한 양방향 예측 방법

다시점 비디오 부호화는 시간 방향에 대해 그림 3과 같은 계층적 B화면 구조를 이용하여 부호화 한다. 계층적 B화면 구조는 다수의 B화면을 이용하고, 또한 B화면을 참조 화면으로 이용할 수 있도록 허용하여 큰 부호화 효율을 보인다 [4].

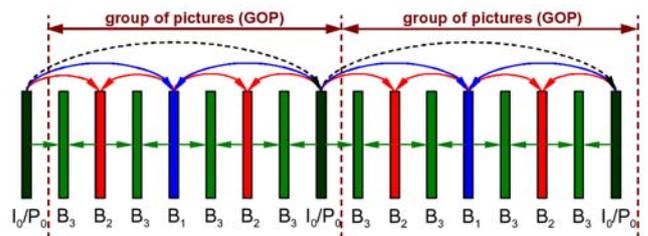


그림 3. 계층적 B화면 구조

일반적으로 B화면 부호화는 I화면이나 P화면 부호화에 비해 복잡도는 높으나 부호화 효율은 좋은 특성을 보인다. 이는 B화면 부호화가 다이렉트 예측, 리스트 0 예측, 리스트 1 예측, 양방향 예측과 같이 다양한 예측 모드를 이용하기 때문이다. 이중 양방향 예측은 B화면 부호화 효율 향상에 크게 기여하는데, H.264/AVC에 이용되는 양방향 예측은 식 (9)와 같이 리스트 0에서 예측한 값과 리스트 1에서 예측한 값의 평균값을 이용한다. 여기서  $pred0$ 와  $pred1$ 은 각각 리스트 0와 리스트 1에서의 최적의 예측값이다.

$$pred(i, j) = (pred0(i, j) + pred1(i, j) + 1) \gg 1 \quad (9)$$

하지만 현재의 양방향 예측 방법은 깊이영상에서 객체의 선형적인 움직임이 발생하는 경우 올바른 값을 예측할 수 없다. 그림 4는 깊이영상 부호화 시 양방향 예측을 이용하는 경우에 대한 예를 보여주고 있다.

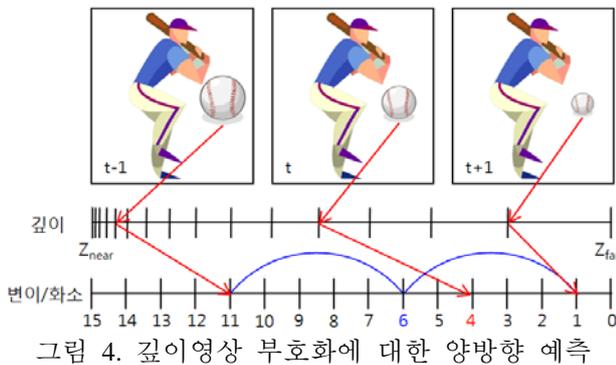


그림 4. 깊이영상 부호화에 대한 양방향 예측

그림 4에서 시간  $t-1$ 과 시간  $t+1$ 의 화면을 이용하여 시간  $t$ 의 화면을 양방향 예측을 이용하여 부호화 하고자 할 때, 파란선은 기존의 양방향 예측을 이용하여 예측된 값이고, 빨간선은 실제 정확한 화소값을 보여주고 있다. 그림에서 보듯이 야구공이 깊이 영역에서 선형적으로 변하여 그에 상응하는 변이/화소값이 11과 1인 경우 기존의 양방향 예측을 이용하면 그 예측값은 6이 된다. 그러나 시간  $t$ 에서의 야구공에 대한 정확한 변이/화소값은 6이 아닌 4이다. 이러한 오류가 발생하는 원인은 깊이 영상 표현 시 변이/화소 영역에서 균등 양자화를 하기 때문에 깊이 영역에서는 비균등 양자화가 적용되기 때문이다.

이를 해결하기 위해 본 논문에서는 깊이 영역에서 양방향 예측을 수행한다. 이를 위해 깊이와 깊이영상의 화소값 사이의 상호변환을 이용한다. 먼저, 식 (9)의 리스트 0와 리스트 1에 대한 예측값인  $pred0$ 와  $pred1$ 에 대한 깊이값을 식 (10)과 식 (11)을 이용하여 구한다.

$$Z_{pred0} = \frac{1}{\frac{v_{pred0}}{N-1} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) + \frac{1}{Z_{far}}} \quad (10)$$

$$Z_{pred1} = \frac{1}{\frac{v_{pred1}}{N-1} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) + \frac{1}{Z_{far}}} \quad (11)$$

다음은 식 (12)를 이용하여 식 (10)과 식 (11)에서 얻어진  $Z_{pred0}$ 와  $Z_{pred1}$ 을 이용하여 깊이 영역에서 양방향 예측을 수행한다.

$$Z_{pred}(i, j) = (Z_{pred0}(i, j) + Z_{pred1}(i, j) + 1) \gg 1 \quad (12)$$

마지막으로 식 (12)에서 얻어진  $Z_{pred}$ 에 상응하는 변이/화소값을 식 (13)을 이용하여 계산한다.

$$Pred_v(i, j) = (N-1) \cdot \frac{\frac{1}{Z_{pred}(i, j)} - \frac{1}{Z_{far}}}{\frac{1}{Z_{near}} - \frac{1}{Z_{far}}} \quad (13)$$

#### 4. 실험 결과 및 분석

본 논문에서 제안하는 깊이영상을 위한 양방향 예측 기술에 대한 성능을 평가하기 위해 MPEG 3DV 그룹에서 사용되는 "Mobile" 영상을 이용했다. 제안하는 방법은 JM14.0에 구현되었으며, 계층적 B화면 구조를 이용하여 QP 22, 25, 28, 31에 대해 깊이영상을 부호화 했다. GOP 길이는 15였으며, 움직임 탐색 범위는 96으로 설정됐다. 계층에 따른 양자화 값의 차는 0으로 설정했다. 4번과 7번 시점을 부호화하고 5번 시점을 합성하여 합성 영상에 대한 화질도 비교했다. 영상 합성에서 MPEG 3DV 그룹에서 사용하는 VSRS 2.0 [5]을 이용했다. 색상 영상은 부호화되지 않은 원영상을 이용했다.

표 1은 "Mobile" 영상에 대한 깊이영상 부호화 효율을 보여주고 있고, 표 2는 "Mobile" 영상에 대해 부호화 한 영상을 이용하여 가상영상을 생성한 경우에 대해 깊이의 비트율 대비 합성 영상의 화질에 대한 결과를 보여주고 있다.

표 1. "Mobile" 영상에 대한 깊이영상 부호화 결과

QP	비트율 (kbps)		PSNR (dB)	
	기존 방법	제안하는 방법	기존 방법	제안하는 방법
22	237.63	230.37	55.94	56.19
25	190.82	186.87	54.07	54.14
28	153.67	151.00	51.98	52.16
31	125.83	122.52	49.78	50.13

표 2. "Mobile" 영상에 대한 깊이영상 비트율 대비 합성 영상의 화질에 대한 실험 결과

QP	비트율 (kbps)		PSNR (dB)	
	기존 방법	제안하는 방법	기존 방법	제안하는 방법
22	237.63	230.37	41.18	41.18
25	190.82	186.87	41.14	41.17
28	153.67	151.00	41.18	41.20
31	125.83	122.52	41.07	41.07

그림 5와 그림 6은 각각 표 1과 표 2에 대한 비트율 왜곡 곡선을 보여주고 있다.

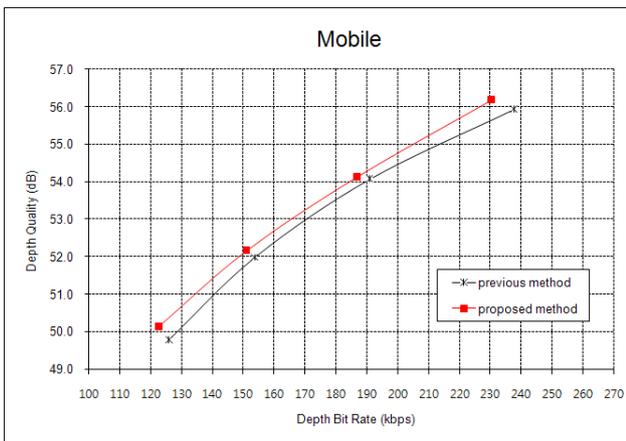


그림 5. "Mobile" 영상에 대한 비트율-왜곡 곡선

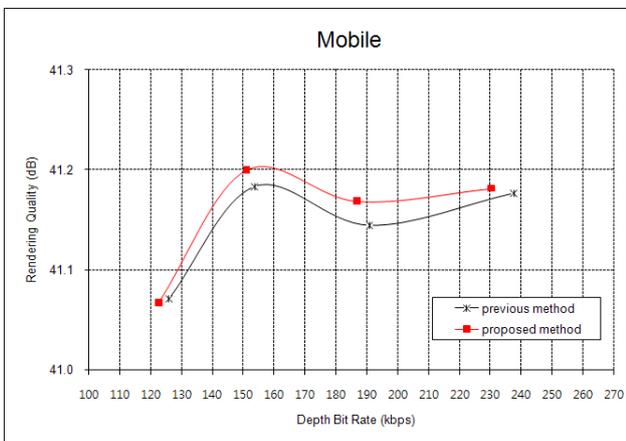


그림 6. "Mobile" 영상에 대한 깊이영상 비트율 대비 합성 영상의 화질에 대한 비트율-왜곡 곡선

위 실험결과에서 보듯이 제안하는 양방향 예측 방법은 "Mobile"에 대해 깊이영상의 부호화 측면에서 기존의 양방향 예측 방법에 비해 0.37 dB의 정도의 화질 개선효과나 3.73% 정도의 비트율 절감효과를 보였다. 또한 합성 영상의 화질을 비교했을 경우 0.02 dB 정도의 합성 영상 화질의 개선 효과를 보였다.

### 5. 결론

본 논문에서는 깊이 데이터가 깊이영상으로 표현되는 과정에서 깊이 영역에서는 비균등 양자화되고, 변이/화소 영역에서는 균등 양자화되는 특성을 이용하여 깊이영상에 적합한 양방향 예측 방법을 제안한다. 기존의 양방향 예측을 이용하는 경우 객체의 선형적인 움직임에 대해 잘못된 값을 예측하므로 제안하는 방법은 변이/화소와 깊이 사이의 상호 변환을 이용하여 변이/화소 영역의 예측값을 깊이 영역으로 변환한 후 양방향 예측을 적용한다. 그 후, 양방향 예측된 깊이값을 다시 변이/화소 영역으로 변환하여 올바르게 양방향 예측된 변이/화소값을 얻는다. 실험을 통해 제안하는 양방향 예측 방법이 깊이영상 부호화 시 3.73%의 비트율 절감효과와 가상영상 합성 시 0.02 dB의 화질 개선효과가 있음을 확인했다.

### 감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2010-(C1090-0902-0017)).

### 참고문헌

- [1] A. Smolic, H. Kimata, and A. Vetro, "Development of MPEG standards for 3D and free viewpoint video," Proc. of SPIE Optics East, Three-Dimensional TV, Video, and Display IV, Boston, MA, Oct. 2005.
- [2] Y. Morvan, D. Farin, and P.H.N. de With, "Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview images," Proc. of International Conference on Image Processing (ICIP 2007), vol. 5, pp. 105-108, Oct. 2007.
- [3] J. Chai, X. Tong, S. Chan, and H. Shum, "Plenoptic sampling," Proc. of ACM SIGGRAPH, pp. 307-318, 2000.
- [4] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," Proc. of International Conferences on Multimedia & Expo, pp. 1929-1932, 2006.
- [5] ISO/IEC JTC1/SC29/WG11 "Reference Softwares for Depth Estimation and View Synthesis," M15377, April 2008.