

깊이 영상의 화면 내 예측을 위한 구면 모델링 모드

이동석, 권순각

동의대학교 컴퓨터소프트웨어공학과

e-mail : ulsan333@gmail.com, Soon-Kak Kwon

Spherical Modeling Mode in Intra Prediction for Depth Picture

Dong-Seok Lee, Soon-Kak Kwon

Dept. of Computer Software Engineering, Dong-Eui University

요 약

본 논문은 깊이 영상의 부호화의 화면 내 예측에서 구면 추정을 이용한 구면 모델링 모드를 제안한다. 블록 내 화소들의 영상 좌표를 깊이 값을 이용하여 3차원 카메라 좌표로 변환한 후, 최소자승법을 통해 최적의 구면을 추정한다. 그 후 추정된 구면을 바탕으로 블록 내 화소들에 대한 깊이 값을 예측한다. 실험 결과 깊이 영상에 대해 영상 부호화 표준인 H.264의 기존 화면 내 예측 모드만을 사용했을 때 보다 화면 내 예측 성능이 증가함을 보인다.

1. 서론

거리 정보로 이루어진 깊이 영상을 이용한 객체 인식 및 행동 인식 기술의 응용이 증가함에 따라 해당 깊이 영상을 전송하는 기술의 필요성이 증대되고 있다. 깊이 영상을 전송하기 위해 깊이 영상에 포함된 중복성을 제거하여 압축하는 여러 방법이 연구되었다. 깊이 영상을 부호화하는 방법으로는 깊이 영상을 3차원 점군으로 변환하여 처리하는 방법[1-5]과 기존 색상 영상 부호화 방법을 적용하는 방법[6-8]으로 구분할 수 있다. 점군을 이용한 방법은 깊이 정보를 이용하여 깊이 영상의 각 화소를 3차원 점군으로 바라본다. 그 후 형성된 3차원 점군들로부터 메쉬 구조를 형성한다. 그 후 주변의 먼저 예측된 점군들로 이루어진 메쉬를 이용하여 현재 점군의 위치를 예측한다. 하지만 점군을 통한 압축 알고리즘은 정적으로 스캔된 점군의 위치적 중복성을 제거하는데 중점을 두었지, 영상 내 프레임 간의 공간적인 중복성을 제거하지 못한다는 단점이 있다. 또한 메쉬를 추출하는 데 추가 계산이 필요하므로 부호화 복잡성이 크다는 단점이 있다. Grewatsch와 Muller[6]는 몇 가지 깊이 영상의 부호화 알고리즘을 조사하고 평가하며, 표준 H.264/AVC 부호화 방법을 적용하여 깊이 영상을 압축 할 때 메쉬 기반 코딩보다 성능이 우수함을 입증하였다. 하지만 기존 색상 영상 부호화 방법은 8bit 단위의 화소로 구성된 영상에 적용이 되는 방법이기 때문에, 8bit 범위를 초과하는 깊이 영상 내 화소에 대해서는 적용하기 어렵다는 단점이 있다. 특히 깊이 영상의

화면 내 예측은 표면의 종류에 따라 깊이 화소 값의 분포가 비선형으로 분포되는 등 색상 화면과 다른 특성을 가지고 있기 때문에 기존 화면 내 예측을 적용하기 힘들다. 이러한 깊이 화면의 특성을 이용하여 깊이 영상의 화면 내 예측 알고리즘이 여러 제안되었다. Liu[9]는 깊이 블록의 회소 표현으로 깊이 화면을 재구성하여 화면 내 예측을 하는 방법을 제안하였다. Lan[10]는 인트라 DC 예측 방향 하의 HEVC의 예측 단위에서 컨텍스트 기반 공간도메인의 화면 내 예측 모드를 도입하였다. 이 때 화면 내 예측 모드에서 같은 객체에 속한 이미 예측된 픽셀만이 예측 단계에 이용된다. Shen[11]은 edge위치 기반의 예측자를 결정함으로써 예측을 수행하는 Edge-aware 방법을 제안했다. 하지만 이러한 연구들은 객체 기반의 예측 방법으로, 객체 정보를 얻기 위해 필요한 전처리 과정에서 많은 연산이 필요하다는 단점이 있다. 이에 깊이 영상에서 많은 부분을 차지하는 평면을 블록 단위로 추정하여 화면 내 예측을 하는 평면 추정 모드가 제안되었다[12]. 이 방법을 통해 깊이 화면의 화면 내 예측이 상당량 개선되었지만, 평면이 아닌 면으로 이루어진 영역에 대해서는 예측의 정확도가 떨어진다는 단점이 있다.

본 논문에서는 깊이 영상의 화면 내 예측에 적용할 수 있는 구면 모델링 모드를 제안한다. 깊이 영상 내 블록에 대해 제일 근접한 구면을 최소자승법을 이용하여 추정한 후, 추정된 구면을 기반으로 화면을 예측한다. 이를 통해 구면으로 이루어진 영역에 대해서 화면 내 예측을 좀 더 효율적으로 할 수 있다.

2. 영상 부호화 표준에서의 화면 내 예측 방법

H.264/AVC[13]에서의 화면 내 예측은 블록은 이전에 예측된 주변 블록에 기반한다. 화면 내 예측 모드는 명도 매크로 블록에 대해서는 4×4, 8×8 블록에 대해 9가지의 모드와 16×16 블록에 대한 4가지 모드, 그리고 채도 블록에 대한 4가지 모드가 있다. 4×4 블록 모드에서는 그림 1과 같은 8개의 방향 예측 모드와 평균을 이용한 예측 모드(DC 모드)를 합한 총 9개의 화면 내 예측 모드를 지원한다. 16×16 블록에 대해서는 4×4 블록의 방향 예측 모드의 일부인 vertical 및 horizontal 예측 모드와 DC모드와 더불어 이미 부호화된 이웃한 화소들을 전부 사용하여 현재 블록을 선형 예측하는 plane 모드를 지원한다.

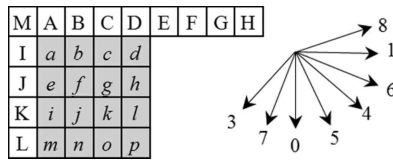


그림 1. H.264/AVC 4×4 luma 블록에서의 화면 내 예측 모드

H.265/HEVC[14]내 화면내 예측에서는 H.264/AVC와 유사하게 방향 예측 모드와 평균 예측 모드를 지원하는데, 방향 예측 모드는 H.264/AVC보다 좀 더 많은 방향을 지원하여 좀 더 정밀한 예측이 가능하다. 또한 H.265/HEVC는 planar 모드를 지원하는데, 이는 H.264/AVC의 plane 모드에서의 단점이었던 블록 경계에서 초래되는 불연속성 문제를 해결하였다.

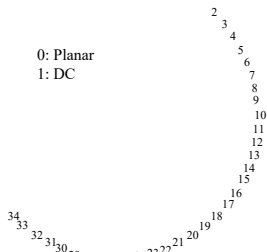


그림 2. H.265/HEVC에서의 화면 내 예측 모드

3. 깊이 영상의 화면 내 예측을 위한 구면 모델링 모드

3.1 깊이 화소의 좌표 변환 방법

핀 홀 카메라 모델에서는 3차원 좌표계에서의 한 점이 카메라로 정의되는 원점으로부터 f 만큼 떨어진 가상의 영상 평면으로 투영되는 위치는 그림 3에서 볼 수 있듯이 그 점과 원점을 잇는 선에 위치한다. 여기서 원점과 영상 평면간의 거리 f 는 초점 거리로 정의된다. 깊이 영상에서의 한 화소의 깊이는 카메라의 촬영 축 방향으로의 거리로 정의된다. 이는 3차원 좌표계에서의 Z축의 좌표와 일치한다. 따라서 깊이 값 d 를 통해 2차원 영상 좌표로부터 카메라를 원점으로 하는 3차원 카메라 좌표로 식 (1)을 통해 변환을 할 수 있다.

$$\begin{aligned} x_c &= xd/f \\ y_c &= yd/f \\ z_c &= d \end{aligned} \quad (1)$$

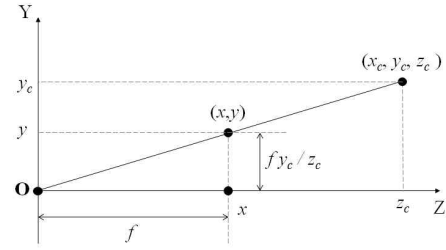


그림 3. 핀홀 카메라 모델에서의 영상 투영

3.2 구면 추정을 통한 깊이 영상의 화면 내 예측

3차원 좌표계에서의 구면은 식 (2)으로 표현된다.

$$(x_c - a)^2 + (y_c - b)^2 + (z_c - c)^2 = r^2 \quad (2)$$

식 (2)에서 a, b, c, r 은 구를 결정하는 인자이다. 이 때 주어진 n 개의 점 (x_{ci}, y_{ci}, z_{ci}) ($1 \leq i \leq n$)을 식 (2)에 대입하면 식 (3) 형태의 행렬식이 구해진다.

$$\begin{aligned} AR &= B \\ A &= \begin{bmatrix} x_{c_1}^2 + y_{c_1}^2 & x_{c_1} & y_{c_1} & z_{c_1} & 1 \\ x_{c_2}^2 + y_{c_2}^2 & x_{c_2} & y_{c_2} & z_{c_2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{c_n}^2 + y_{c_n}^2 & x_{c_n} & y_{c_n} & z_{c_n} & 1 \end{bmatrix}, B = \begin{bmatrix} z_{c_1}^2 \\ z_{c_2}^2 \\ \vdots \\ z_{c_n}^2 \end{bmatrix} \\ R &= \begin{bmatrix} -1 \\ 2a \\ 2b \\ 2c \\ -(a^2 + b^2 + c^2 - r^2) \end{bmatrix} \end{aligned} \quad (3)$$

이 때 주어진 점들이 5개를 초과하는 경우에는 식 (2)을 만족하는 하나의 인자, 즉 식 (3)에서의 R 을 결정할 수 없다. 이러한 경우 식 (2)을 변형한 각 점의 식 (4)으로 표현되는 오차의 제곱 합이 최소가 되는 인자를 구하는 문제를 생각할 수 있다.

$$(x_{c_i} - a)^2 + (y_{c_i} - b)^2 + (z_{c_i} - c)^2 - r^2 \quad (4)$$

오차의 제곱 합이 최소가 되는 인자의 행렬 R 은 식 (5)으로 정의되는 A 의 유사역행렬을 통해 식 (6)을 통해 구할 수 있다.

$$A^+ = (A^T A)^{-1} A^T \quad (5)$$

$$R = BA^+ \quad (6)$$

블록 내 깊이 화소에 대해 구면을 추정하기 위해서는 블록 내 각 화소에 대해 식 (1)을 통해 영상 좌표에서 3차원 카메라 좌표로 변환한 후, 좌표들을 식 (3)에 대입한 후, 식 (6)을 통해 추정된 곡면의 인자를 계산한다. 그 후 추정된 구면을 통해 블록 내 화소에 대해 깊이 값을 예측하기 위해 식 (2)의 식에 3차원 좌표를 영상 좌표와 깊이 값으로 변환하는 식 (7)을 대입한 후 d_i 로 정리하면 식 (8)을 얻을 수 있다.

$$\begin{aligned}x_i &= x_c f / z_i \\y_i &= y_c f / z_i \\d_i &= z_i\end{aligned}\quad (7)$$

$$\begin{aligned}&\left(-\left[\left(\frac{x_i}{f}\right)^2 + \left(\frac{y_i}{f}\right)^2\right] - 1\right)d_i^2 + \\&\left(2a\frac{x_i}{f} + 2b\frac{y_i}{f} + 2c\right)d_i - \\&(a^2 + b^2 + c^2 - r^2) = 0\end{aligned}\quad (8)$$

블록 내 영상 좌표와 식 (6)을 통해 추정된 곡면의 인자 a, b, c, r 를 식 (8)에 대입한 후, d_i 에 대한 이차방정식의 해를 계산한다. 이 때 d_i 의 해는 두 개가 될 수 있는데, 이는 인자 a, b, c, r 를 가지는 구가 영상 평면의 (x_i, y_i) 로 투영 되는 면이 상대적으로 거리가 가까운 면과 먼 면 두 개가 된다는 것을 의미한다. 따라서 구면 모델링 모드는 안쪽 면 모드와 바깥쪽 면 모드로 구분하여 추정해야 한다. 이 때 가까운 면으로 추정되는 값이 값이 음수가 나오는 화소가 발생할 경우 해당 블록에서는 내부면 추정 모드를 수행하지 않는다.

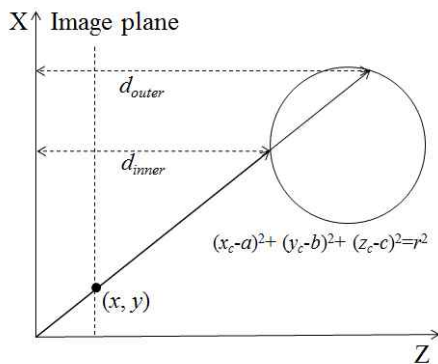


그림 4. 구면 모델링 모드의 깊이 값 예측

4. 실험결과

본 논문에서 제안된 구면 모델링 모드의 성능을 측정하기 위해 영상 부호화 표준인 H.264/AVC의 화면 내 예측 방법을 이용하여 성능을 측정하였다. 이 때 실험으로 쓰일 영상은 그림 5와 같이 구면을 가지는 객체가 포함된 깊이 영상과 통상의 장면을 촬영한 깊이 영상 두개를 이용하였다. 이 때 그림 5 (a)는 깊이 카메라로 Kinect V2를 이용하였으며, 해상도는 512×424이고, 식 (1), (7)에서 쓰이는 깊이 카메라의 f 는 365.456이다. 또한 그림 5 (b)[15]는 깊이 카메라로 Kinect를 이용하였으며, 해상도는 640×480이고, 식 (1), (7)에서 쓰이는 깊이 카메라의 f 는 585.6이다.

4×4 luma 모드에서의 화면 내 예측 모드에 제안된 구면 모델링 모드를 추가하여 화면 내 예측을 수행하였다. 이 때 기존 9개 모드만을 사용했을 때와, 제안된 구면 모델링 모드를 추가했을 때의 성능의 비교는 표 1과 같다. 구면 모델링 모드가 추가되었을 때 화면 내 예측 성능이 개선되었다.



(a) (b)

그림 5. 실험 영상

표 1. 구면 모델링 모드를 포함한 화면 내 예측

영상	MSE	
	기존 모드 사용	기존 모드 + 구면 모델링 모드
그림 5 (a)	897.009	113.262
그림 5 (b)	527.915	146.835

그림 6은 기존 모드만을 이용하여 화면 내 예측을 수행했을 때와 제안된 모드를 추가했을 때의 각 모드가 선택된 비율을 나타낸 것이다. 이 때 구면 모델링 모드가 추가되었을 때 구면 모델링 모드의 선택 비율이 높았는데, 이는 깊이 영상의 분포가 비선형으로 나타나는 영역이 많다는 것을 의미한다.

(a)

(b)

그림 6. 화면 내 예측에서의 모드 선택 비율. (a) 기존 모드 사용, (b) 구면 모델링 모드 추가

5. 결론

본 논문에서는 깊이 영상의 부호화에서 화면 내 예측에 적용할 수 있는 구면 모델링 모드를 제안하였다. 깊이 화면 내 블록에 대해 제일 인접한 구면을 가지는 구의 인자를 최소자승법을 통해 구한 후, 인자를 통해 안쪽 면 모드와 바깥쪽 면 모드에서의 깊이 값을 예측했다. 실험 결과 기존 H.264/AVC의 화면 내 예측과 비교했을 때 MSE가 확연하게 개선됨을 보였다. 본 논문에서 제안된 방법을 통해 깊이 영상을 더욱 효율적으로 부호화할 수 있을 것으로 기대된다.

감사의 글

이 논문은 2018년도 BB21+ 사업으로 지원되었음.

참고문헌

- [1] S. Gumhold, Z. Karni, M. Isenburg, M., and H. Seidel, "Predictive Pointcloud Compression," *Proceeding of ACM SIGGRAPH*, pp. 137-141, 2005.
- [2] J. Kammerl, N. Blodow, R. B. Rusu, S. Gedikli, M. Beetz, and E. Steinbach, "Real-time Compression of Point Cloud Streams," *Proceeding of IEEE International Conference on Robotics and Automation*, pp. 778-785, 2012.
- [3] O. Devillers and P. Gandoin, "Geometric Compression for Interactive Transmission," *Proceeding of International Conference on Information Visualization*, pp. 319-326, 2000.
- [4] J. Peng and C. Kuo, "Octree-based Progressive Geometry Encoder," *Proceeding of SPIE*, pp. 301-311, 2003.
- [5] Y. Huang, J. Peng, C. Kuo, and M. Gopi, "Octree-based Progressive Geometry Coding of Point Clouds," *Proceeding of Eurographics Symposium on Point-Based Graphics*, pp. 103-110, 2006.
- [6] S. Grewatsch and E. Muller, "Evaluation of Motion Compensation and Coding Strategies for Compression of Depth Map Sequences," *Proceeding of 49th SPIE's Annual Meeting, Proceeding of Mathematics of Data/Image Coding, Compression, and Encryption VII, with Applications*, pp. 117-125, 2004.
- [7] Y. Morvan, D. Farin, and P. H. N. Dewith, "Depth-image Compression Based on An R-D Optimized Quadtree Decomposition for The Transmission of Multiview Images," *Proceeding of IEEE International Conference on Image Processing*, pp. V105-V108, 2007.
- [8] S. Milani and G. Calvagno, "A depth image coder based on progressive silhouettes," *IEEE Signal Processing Letters*, Vol. 17, No. 8, pp. 711-714, 2010.
- [9] S. Liu, P. Lai, D. Tian, and C. W. Chen, "New Depth Coding Techniques with Utilization of Corresponding Video," *IEEE Transactions on Broadcasting*, Vol. 57, No. 2, pp. 551-561, 2011.
- [10] C. Lan, J. Xu, and F. Wu, "Object-based Coding for Kinect Depth and Color Videos," *Proceeding of Visual Communications and Image Processing*, pp. 1-6, 2012.
- [11] G. Shen, W. S. Kim, A. Ortega, J. Lee, and H. Wey, "Edge-aware Intra Prediction for Depth-map Coding," *Proceeding of IEEE International Conference on Image Processing*, pp. 3393-3396, 2010.
- [12] D. S. Lee and S. K. Kwon, "Intra Prediction of Depth Picture with Plane Modeling," *Symmetry*, Vol. 10, No. 12, pp. 1-16, 2018.
- [13] S. K. Kwon, A. Tamhankar, and K. R. Rao, "Overview of H.264/MPEG-4 part 10," *Journal of Visual Communication and Image Representation*, Vol. 17, pp. 186-216, 2006.
- [14] G. J. Sullivan, J. Ohm, W. J. Han, and T. Wiegand, "Overview of The High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 12, pp. 1649-1668, 2012.
- [15] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor Segmentation and Support Inference from RGBD Images," *Proceeding of European Conference on Computer Vision*, pp. 746-760, 2012.