

# 깊이 영상의 이차 곡면 모델링을 통한 화면 내 예측 방법<sup>+</sup>

## (Intra Prediction Method by Quadric Surface Modeling for Depth Video)

이 동 석<sup>1)</sup>, 권 순 각<sup>2)\*</sup>  
(Dong-seok Lee and Soon-kak Kwon)

**요 약** 본 논문은 깊이 영상의 부호화를 위해 이차 곡면 모델링 방법을 통한 화면 내 예측 방법을 제안한다. 깊이 영상 내 깊이 화소는 거리 정보를 통해 3차원 좌표로 변환한다. 화면 내 예측을 위한 참조 화소들에 대해 최소자승법을 통해 오차가 제일 작은 이차 곡면을 찾는다. 참조 화소는 상단의 화소들 또는 좌단의 화소들 중 하나가 될 수 있다. 이차 곡면을 통한 화면 내 예측에서, 한 화소에 대해 두 개의 예측 값이 계산된다. 각각의 참조 화소에 대해 예측 값들과 참조 화소의 차이의 제곱합으로 두 오차 값을 계산한다. 계산된 총 4개의 오차 중 제일 작은 오차를 가지는 참조 화소 선택 방법과 예측 화소 선택 방법이 선택되고, 이를 통해 블록 내 화소를 예측한다. 실험 결과는 최신 영상 부호화 방법과 비교하여 왜곡과 비트율이 각각 최대 5.16%과 5.12% 개선됨을 보인다.

**핵심주제어:** 깊이 영상 부호화, 화면 내 예측, 이차 곡면 모델링

**Abstract** In this paper, we propose an intra-picture prediction method by a quadratic surface modeling method for depth video coding. The pixels of depth video are transformed to 3D coordinates using distance information. A quadratic surface with the smallest error is found by least square method for reference pixels. The reference pixel can be either the upper pixels or the left pixels. In the intra prediction using the quadratic surface, two prediction values are computed for one pixel. Two errors are computed as the square sums of differences between each prediction values and the pixel values of the reference pixels. The pixel of the block are predicted by the reference pixels and prediction method that they have the lowest error. Comparing with the-state-of-art video coding method, simulation results show that the distortion and the bit rate are improved by up to 5.16% and 5.12%, respectively.

**Keywords:** Depth Video coding, Intra Prediction, Quadric Surface Modeling

\* Corresponding Author: skkwon@deu.ac.kr

+ 본 논문은 2021년도 BB21+ 사업으로 지원되었으며, 또한 과학 기술정보통신부 및 정보통신기획평가원의 Grand ICT연구센터 지원사업의 연구결과로 수행되었으며 (IITP-2022-2020-0-017 91), 또한 정부(과학기술정보통신부)의 재원으로 한국연구재단

의 지원을 받아 수행된 연구임(NRF-2021R1F1A1062131).

Manuscript received February 14, 2022 / revised April 01, 2022 / accepted April 19, 2022

1) 동의대학교 인공지능그랜드ICT연구센터, 제1저자  
2) 동의대학교 컴퓨터소프트웨어공학과, 교신저자

## 1. 서론

깊이 영상은 카메라로부터 피사체까지 거리가 화소 값으로 저장된 영상이다. 화소가 가지는 피사체까지의 거리들과 좌표값을 통해 촬영 영역의 3차원 표면에 대한 정보를 계산할 수 있다. 따라서 깊이 영상은 해당 촬영영역에 대한 3차원 표면 정보를 가진다고 볼 수 있다. 깊이 영상이 가지는 표면 정보를 통해 특정 형태를 가지는 객체를 검출하고 추적하여 특정 행동을 인식할 수 있다(Zhao et al., 2017; Ren et al., 2017; Jiang et al., 2019). 또한 깊이 영상을 통해 지표면이나 벽 등을 검출하여 동시적 위치추정 및 SLAM(Simultaneous localization and mapping)에 적용할 수 있다(Sun et al., 2017; Aguilar et al., 2017). 이들 응용들을 실시간으로 처리하기 위해서는 깊이 영상을 고성능의 장치로 신속하게 전송하는 것이 필요하다. 신속한 전송을 위해서 깊이 영상의 효율적인 부호화 방법이 필요하다.

깊이 영상의 부호화를 위하여 기존 영상 부호화 표준인 H.264/AVC, H.265/HEVC, VVC (Versatile video coding) 등의 기존 영상 부호화 표준을 적용할 수 있다. 하지만 영상 부호화 표준의 시간적 또는 공간적 예측 기법은 깊이 영상에서는 부정확하다는 문제가 있다. 예측 기법을 통한 부호화에서는 화소 간 유사성을 이용한다. 색상 영상의 유사성은 색상의 유사성 즉 색상 화소의 유사성으로 표현될 수 있다. 이는 부호화를 위해 시간적 또는 공간적으로 인접한 색상 화소 값으로 예측할 수 있다는 의미이다. 하지만 깊이 영상의 유사성은 표면의 유사성으로 표현된다. 달리 말하면 같은 객체 내에 있는 인접한 두 색상 화소의 값은 거의 비슷하지만, 같은 표면에 위치한 인접한 두 깊이 화소의 값은 서로 다를 수 있다. 하지만 해당 화소가 속한 3차원 표면의 정보를 예측할 수 있다면 이를 수학적으로 모델링하여 좌표에 따른 깊이 화소 값을 예측할 수 있다.

기존 영상 부호화 표준을 보완하여 깊이 영상을 화면 간 예측하는 연구가 있었지만(Fu et al., 2013; Wang et al., 2016, 2018), 깊이 영상

의 화면 내 예측에 대한 연구는 거의 이루어지지 않고 있다. Lee et al.(2021)의 연구는 평면 표면의 화면 내 예측을 위한 평면 예측을 통한 화면 내 예측 방법을 제안했지만, 해당 연구는 평면 표면에서만 효율적으로 예측하고, 곡면에 대해서는 예측의 효율이 떨어진다는 단점이 있다. Lee et al.(2019)의 연구에서는 깊이 영상의 구면 모델링을 다뤘지만, 해당 연구는 깊이 영상 부호화를 부호화할 때 사용한 구면의 모델링 계수 부호화를 고려하지 않았다.

본 논문에서는 깊이 영상의 화면 내 예측을 위한 이차 곡면 표면 모델링을 통한 화면 내 예측 방법을 제안한다. 참조 화소들을 통해 이차 곡면 표면을 예측하고, 이를 수학적으로 모델링한다. 이 때 사용되는 참조 화소들은 블록 상단에 있는 화소들 또는 블록 좌단에 있는 화소들 중 하나를 사용한다. 참조 화소의 선택을 위해 블록 상단과 좌단에 있는 화소들을 먼저 모델링 한 후, 해당 모델링 정확도를 측정하여 정확도가 높은 참조 화소를 사용하여 블록을 예측한다. 이차 곡면 모델링을 통해 구면, 타원면 등의 곡면을 가지는 표면위의 깊이 화소 값을 예측할 수 있다.

## 2. 관련 연구

### 2.1 영상 부호화를 위한 예측 기법

영상은 신호의 중복성을 제거함으로써 영상을 압축할 수 있다. 영상 신호의 중복성은 시간적 중복성, 공간적 중복성, 통계적 중복성으로 분류된다. 시간적 중복성과 공간적 중복성은 각각 시간적, 공간적으로 이웃한 화소들 간의 유사성을 의미하고, 통계적 중복성은 영상을 표현하기 위한 기호의 반복적인 등장을 의미한다. 시간적 중복성과 공간적 중복성은 각각 시간적, 공간적으로 이웃한 화소를 통해 예측하여, 그 잔차만을 전송함으로써 제거하고, 통계적 중복성은 기호의 등장 빈도에 따라 차등적인 부호를 할당하여 제거한다. H.264/AVC(Kwon et al. 2006), H.265/HEVC(Sullivan et al., 2012), VVC(Bross et al., 2021) 등의 영상 부호화 표준에서는 시간

적 중복성을 제거하기 위해서 인접한 프레임으로부터 부호화하려는 블록 내 화소들과 제일 유사한 블록을 검출한다. 공간적 중복성을 제거하기 위해서는 부호화하려는 블록과 공간적으로 인접하면서 이미 부호화된 화소를 이용하여 블록 내 화소들을 예측한다. 이 때 인접한 참조 화소들로부터 블록 내 화소를 예측하는 방법을 의미하는 다수의 인트라 모드 중 하나를 사용한다. 이 때 압축된 영상 신호를 복원하기 위해서 사용된 참조 프레임의 정보와 참조 프레임 내에서 사용된 블록의 위치, 또는 사용된 인트라 모드에 대한 정보가 추가적으로 제공되어야 한다.

## 2.2 기존 깊이 영상 부호화를 통한 깊이 영상 부호화 방법

Shen et al.(2010)는 참조 블록의 정보를 통해 에지(edge) 정보를 예측하여 깊이 영상을 부호화 하는 연구를 수행하였다. Liu et al.(2011)는 깊이 영상의 블록의 희소 표현을 적용하여 깊이 영상을 부호화하는 방법을 제안했다. Nenci et al.(2014)는 최대 8bit 범위의 화소 값들을 부호화할 수 있는 H.264/AVC를 깊이 영상 부호화에 적용하기 위해 단일 채널 깊이 영상을 8bit의 화소 값을 가지는 다 채널로 분할하여 영상 부호화 표준을 적용하는 방법을 제안하였다. Stankiewicz et al. (2013)는 H.264/AVC의 확장을 통한 깊이 영상 부호화를 위해 깊이 데이터의 비선형 변환을 제안하였다. 하지만 이러한 방법들은 깊이 화소들이 이루는 표면의 특징을 고려하지 않은 방법이다. Fu et al. (2013)는 인접한 화면 간의 표면 유사도를 통해 화면 간 예측을 하는 방법을 제안한다. 블록의 깊이 화소들은 체적 적분을 통해 표면으로 표현한다. 깊이 화소들은 인접한 화면에서 유사한 표면을 가지는 영역의 깊이 화소 값들로 예측된다. Wang et al. (2016, 2018)은 시간상으로 인접한 깊이 화면 간 카메라의 움직임을 계산하여 깊이 화소 값을 예측하는 방법을 제안했다. Lee et al.(2021)는 평면 추정을 통해 깊이 영상을 화면 내 예측하는 방법인 평면 모델링을 제안했다. 깊이 화면 내 한 블록을 대표하는 평면은 블록

내 깊이 화소 값들을 통해 추정된다. 추정된 평면을 결정하는 평면 모델링 계수를 통해 블록 내 깊이 화소를 예측한다.

## 3. 이차 곡면 모델링을 통한 깊이 영상의 화면 내 예측 방법

### 3.1 깊이 화면의 이차 곡면 모델링

깊이 화면은 깊이 영상 내 특정 시간의 프레임으로 정의된다. 깊이 화면의 한 깊이 화소의 값은 해당 깊이 화소에 대응되는 3차원 공간의 한 점이 가지는 z좌표, 즉 카메라의 촬영 방향에 대응하는 3차원 좌표계에서의 좌표로 정의된다.

핀홀 카메라 모델을 통해 깊이 화소의 3차원 좌표를 구할 수 있다(Hartley et al., 2000). 핀홀 카메라 모델은 촬영 화면을 원점에서 거리가 카메라의 초점거리  $f$ 만큼 떨어진 가상 영상 평면으로 설명한다. 3차원 공간상의 한 점은 Fig. 1과 같이 영상 평면의 깊이 화소로 투영되는 것으로 간주된다. 핀홀 카메라 모델을 통해 깊이 화소의 영상 좌표는 식 (1)과 같이 3차원 좌표로 변환된다. 식 (1)에서  $p(x, y)$ 는 영상 좌표  $(x, y)$ 에서의 깊이 화소 값이고,  $(X, Y, Z)$ 는  $(x, y)$ 에 대응하는 3차원 좌표이다.

$$\begin{aligned} X &= \frac{x}{f}p(x, y) \\ Y &= \frac{y}{f}p(x, y) \\ Z &= p(x, y) \end{aligned} \quad (1)$$

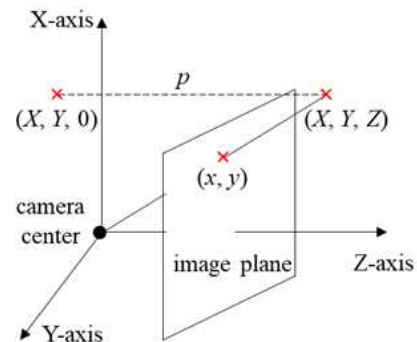


Fig. 1 Pinhole camera model

이차 곡면은 식 (1)과 같이  $X, Y, Z$ 에 대한 2차 다항식으로 표현되는 표면이다. 식 (2)에서  $c_1-c_3$ 은  $x, y, z$  각 축에 대한 스케일을 의미하며,  $c_4-c_6$ 은  $z$ 축,  $x$ 축,  $y$ 축에 대한 회전을,  $c_7-c_9$ 는 각 축에 대한 너비를, 그리고  $c_{10}$ 은 곡면의 전체적인 크기를 의미한다.

$$c_1X^2 + c_2Y^2 + c_3Z^2 + c_4XY + c_5YZ + c_6ZX + c_7X + c_8Y + c_9Z + c_{10} = 0 \quad (2)$$

이차곡면을 나타내는 식 (2)에서 각 계수의 값에 따라 구면, 타원면, 쌍곡포물면 등 다양한 곡면의 형태가 나온다.

이차 곡면 모델링은 식 (2)를 이루고 있는 계수  $c_1-c_{10}$ 를 찾는 것으로 정의한다. 이 때 참조 화소들이 이상적인 이차 곡면을 구성한다면 정확한 계수를 찾을 수 있다. 하지만 실제로는 깊이 측정의 오차나 카메라 렌즈 곡률 등의 다양한 요소들로 인해 식 (2)를 통한 모델링은 실제 표면과 차이가 발생한다. 따라서 목표는 실제 표면과 제일 오차가 작은 이차 곡면을 구성하는 계수를 찾는 것이다. 이 때 부호화할려는 블록의 크기는 충분히 작기 때문에  $c_1-c_3$ 을 1로 두어 식 (3)으로 변형한다. 이는 찾으려는 계수의 개수를 줄여서 오차의 영향을 적게 받고, 좀 더 정확한 계수를 찾기 위함이다.

$$X^2 + Y^2 + Z^2 + k_1XY + k_2YZ + k_3ZX + k_4X + k_5Y + k_6Z + k_7 = 0 \quad (3)$$

식 (3)에 이차 곡면을 모델링하기 위해 사용되는 한 점의 3차원 좌표 ( $X_i, Y_i, Z_i$ )을 대입하면 다음과 같은 식을 구할 수 있다. 식 (4)에서  $e_i$ 는 이차 곡면에 대한 점 ( $X_i, Y_i, Z_i$ )과의 오차이며, 만약 이상적인 상황일 경우  $e_i=0$ 이다.

$$X_i^2 + Y_i^2 + Z_i^2 + k_1X_iY_i + k_2Y_iZ_i + k_3Z_iX_i + k_4X_i + k_5Y_i + k_6Z_i + k_7 = e_i \quad (4)$$

제일 오차가 작은 이차 곡면, 즉 모든  $i$ 에 대한 식 (4)에서의  $e_i$ 의 합을 최소화하는 계수

$k_1-k_7$ 를 찾기 위해서는, 식 (4)가 다변수 비선형 식이기 때문에 가우스 뉴턴법을 사용할 수 있다. 하지만 해당 방법의 성능은 초기 추정 계수에 따라 성능이 좌우된다는 문제가 있다. 즉 초기 추정 계수에 따라 정확한 계수를 찾을 수도 있고, 값이 발산하여 계수를 찾지 못할 수 있다는 문제가 있다. 따라서 식 (4)에 대해 최소자승법을 적용한다. 식 (3)에서 오차를  $-(X^2+Y^2+Z^2)$ 와 나머지 항 간의 차이로 정의한다면 식 (3)을 다음과 같이 변형할 수 있다.

$$k_1XY + k_2YZ + k_3ZX + k_4X + k_5Y + k_6Z + k_7 - (-X^2 - Y^2 - Z^2) = 0 \quad (5)$$

식 (5)에 대해 참조 화소의 3차원 좌표 ( $X_i, Y_i, Z_i$ ) ( $1 \leq i \leq n$ ,  $n$ 은 참조 화소 개수) 들을 대입하면 식 (6)과 같은 행렬식을 구할 수 있다.

$$\begin{aligned} \mathbf{AR} &= \mathbf{B} \\ \mathbf{A} &= \begin{bmatrix} X_1Y_1 & Y_1Z_1 & Z_1X_1 & X_1 & Y_1 & Z_1 & 1 \\ X_2Y_2 & Y_2Z_2 & Z_2X_2 & X_2 & Y_2 & Z_2 & 1 \\ & & \dots & & & & \\ X_nY_n & Y_nZ_n & Z_nX_n & X_n & Y_n & Z_n & 1 \end{bmatrix} \\ \mathbf{R} &= [k_1 \ k_2 \ k_3 \ k_4 \ k_5 \ k_6 \ k_7]^T \\ \mathbf{B} &= \begin{bmatrix} -(X_1^2 + Y_1^2 + Z_1^2) \\ -(X_2^2 + Y_2^2 + Z_2^2) \\ \dots \\ -(X_n^2 + Y_n^2 + Z_n^2) \end{bmatrix} \end{aligned} \quad (6)$$

여기서 식 (7) 최소자승법을 적용하여 계수의 행렬  $\mathbf{R}$ 의 항목 값들을 구한다. 식 (7)에서  $\mathbf{A}^+$ 는  $\mathbf{A}$ 의 의사역행렬로, 식 (8)을 적용하여 구한다.

$$\begin{aligned} \mathbf{R} &= \mathbf{A}^+ \mathbf{B} \quad (7) \\ \mathbf{A}^+ &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (8) \end{aligned}$$

### 3.2 이차 곡면 모델링을 통한 화소 예측

식 (3)의 계수  $k_1-k_7$ 를 가지는 식 (3)을, 식 (1)을 통해 2차원 영상 ( $x, y$ )에 대한 식으로 변환하면 식 (9)와 같다.

$$\begin{aligned} & \left(\frac{p(x,y)}{f}x\right)^2 + \left(\frac{p(x,y)}{f}y\right)^2 + p(x,y)^2 + \\ & k_1\left(\frac{p(x,y)}{f}x\right)\left(\frac{p(x,y)}{f}y\right) + k_2\left(\frac{p(x,y)}{f}y\right)p(x,y) + \\ & k_3\left(\frac{p(x,y)}{f}x\right)p(x,y) + k_4\left(\frac{p(x,y)}{f}x\right) + \\ & k_5\left(\frac{p(x,y)}{f}y\right) + k_6p(x,y) + k_7 = 0 \end{aligned} \quad (9)$$

식 (9)를  $p(x, y)$ 에 대해 정리하면 식 (10)과 같은  $p(x, y)$ 에 대한 이차식임을 알 수 있다.

$$\begin{aligned} & p(x,y)^2[(x/f)^2 + (y/f)^2 + \\ & k_1xy/f^2 + k_2y/f + k_3x/f + 1]^2 + \\ & p(x,y)(k_4x/f + k_5y/f + 1) + k_7 = 0 \end{aligned} \quad (10)$$

이차 곡면의 계수  $k_1-k_7$ 을 이차 곡면 모델링을 통해 찾았을 때, 식 (10)에 예측하고자 하는 화소의 좌표  $(x, y)$ 를 대입한 후  $p(x, y)$ 에 대한 이차식의 해가 해당 좌표에서 예측된 깊이 화소 값이다. 이 때 식 (10)은  $p(x, y)$ 에 대한 이차식 이므로 두 해가 나온다. 이는 계수  $k_1-k_7$ 의 계수를 가지는 이차 곡면에서 영상 좌표  $(x, y)$ 에 대해 3차원 좌표  $(X_1, Y_1, Z_1)$ ,  $(X_2, Y_2, Z_2)$ 가 될 수 있다는 의미이다. 예를 들어 구체의 경우에는 Fig. 2와 같이 안쪽 면과 바깥쪽 면의 깊이가 예측될 수 있다. 이 때 상대적으로 작은 해는 안쪽 면을 의미하고 큰 해는 바깥쪽 면을 의미한다. 따라서 식 (10)을 통해 하나의 깊이 화소 값을 예측하기 위해서는 이차 곡면의 계수들의 정보뿐만 아니라 해당 표면이 안쪽 또는 바깥쪽의 표면이라는 것에 대한 정보도 추가적으로 필요하다.

### 3.3 이차 곡면 모델링을 통한 화면 내 예측 방법

블록 단위 예측 대상이 되는  $m \times n$  사각형 블록에 대해 식 (6)과 식 (7)을 통해 식 (5)의 계수  $k_1-k_7$ 를 찾는다. 이 때 식 (6)을 적용하기 위한 참조 화소로는 높이 또는 너비가 4인 인접한 좌단 화소들 또는 상단 화소들을 Fig. 2와 같이 사용한다. Fig. 3은  $16 \times 16$  블록에 대한 참조 화

소의 선택 방법들을 보인다.

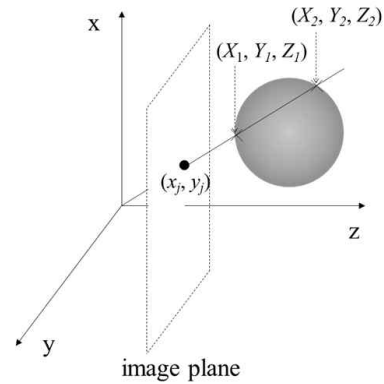


Fig. 2 Two possible prediction cases by quadric surface modeling

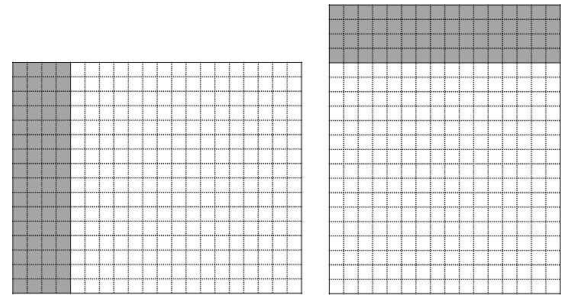


Fig. 3 Selections of reference pixels for quadric surface modeling

두 참조 화소 선택 방법 중 참조 화소로 이루어진 표면과 이차 곡면간의 오차가 적은 참조 화소들을 사용한다. 이차 곡면간의 오차가 큰 참조 화소들의 경우 해당 화소들로 이루어진 표면이 하나의 이차 표면이 아닌 다른 종류의 표면이거나, 여러 표면의 경계가 포함된다는 의미이다. 또한 하나의 이차 곡면 계수에서 두 개의 깊이 화소 예측 값이 계산되기 때문에 안쪽 표면의 예측 값과 바깥쪽 표면의 예측 값 중 하나를 선택하여야 한다. 즉 하나의 블록을 예측하기 위해 2개의 참조 화소 선택 방법과 2개의 예측 화소 선택 방법 중 각각 하나를 선택해야 한다.

이를 위해 각각의 참조 화소 선택 방법에서 사용되는 참조 화소들을 식 (6)에 대입하여 식 (7)을 통해 이차 곡면의 계수를 찾은 후, 해당

계수와 참조 화소들의 좌표를 식 (10)에 대입하여 각 참조 화소들에 대한 깊이 화소 예측을 한다. 이 때 한 참조 화소  $(x_i, y_i)$  들에 대해 식 (10)의 해를 구하면 두개의 깊이 화소 값이 예측된다. 두 개의 화소 값  $\hat{p}_i(x_j, y_j)$ ,  $\hat{p}_o(x_j, y_j)$  ( $\hat{p}_i(x_j, y_j) < \hat{p}_o(x_j, y_j)$ )에 대해 각각 아래식과 같이 오차를 계산한다. 식 (11)에서  $N$ 은 참조 화소의 개수를 의미한다.

$$E_i = \frac{1}{N} \sum_{j=1}^N (\hat{p}_i(x_j, y_j) - p(x_j, y_j))^2$$

$$E_o = \frac{1}{N} \sum_{j=1}^N (\hat{p}_o(x_j, y_j) - p(x_j, y_j))^2 \quad (11)$$

식 (11)의 계산을 인접한 좌단 화소들에 수행하여 안쪽 화소에 대한 오차  $E_{li}$ 와  $E_{lo}$ 를 구하고, 좌단 화소들에 대해 수행하여  $E_{ur}$ 와  $E_{uo}$ 를 계산한다. 이렇게 구한 4개의 오차 중 제일 작은 오차를 가지는 참조 화소 선택 방법 및 예측 화소 선택 방법을 선택한 후, 해당 참조 화소 선택 방법을 통해 계산된 이차 곡면 계수와, 예측하고자 하는 영상 좌표들을 식 (10)에 대입하여 해를 구하고, 계산된 두 개의 해 중 선택된 예측 화소 선택 방법을 통해 화소들을 예측한다.

제안된 방법을 실제 부호화 과정에 적용할 때, 화면 내 예측 방법이 추가됨에 따라 화면

내 예측 방법을 의미하는 인트라 모드의 수가 하나 더 추가된다. 해당 인트라 모드를 나타내는 기호의 개수가 하나 증가함에 따라 부호화에 따른 영향을 미칠 수도 있다. 하지만, 예측 화소의 선택과 예측 값 선택은 참조 화소의 이차 곡면 모델링 정확도에 따라 선택되어 지기 때문에 추가적인 비트스트림의 변경이 일어나지 않는다.

#### 4. 모의실험

본 논문에서는 최신의 영상 부호화 표준인 VVC와 비교하여 제안된 방법을 모의실험한다. 제안된 방법의 실험을 위해 VVC의 테스트모델인 VVC Test Model(VTM)(Fraunhofer, 2022)에 제안된 화면 내 예측 방법을 추가한다. QP(Quantization parameter) 범위는 VTM이 지원하는 전 범위인 0-51에서 실험되었다. 모든 프레임은 화면 내 예측을 통한 부호화가 되는 I 프레임으로 설정되었다. 화면 내 예측에서 발생한 신호들을 부호화하기 위해 VTM에 내장된 산술부호화 방법인 CABAC(Context adaptive binary arithmetic coding)을 적용한다. 제안된 예측 방법을 포함한 화면 내 예측 방법의 선택 방법으로는 VTM에 적용된 율-왜곡 최적화 기법 (Rate-distortion optimization)를 적용한다.

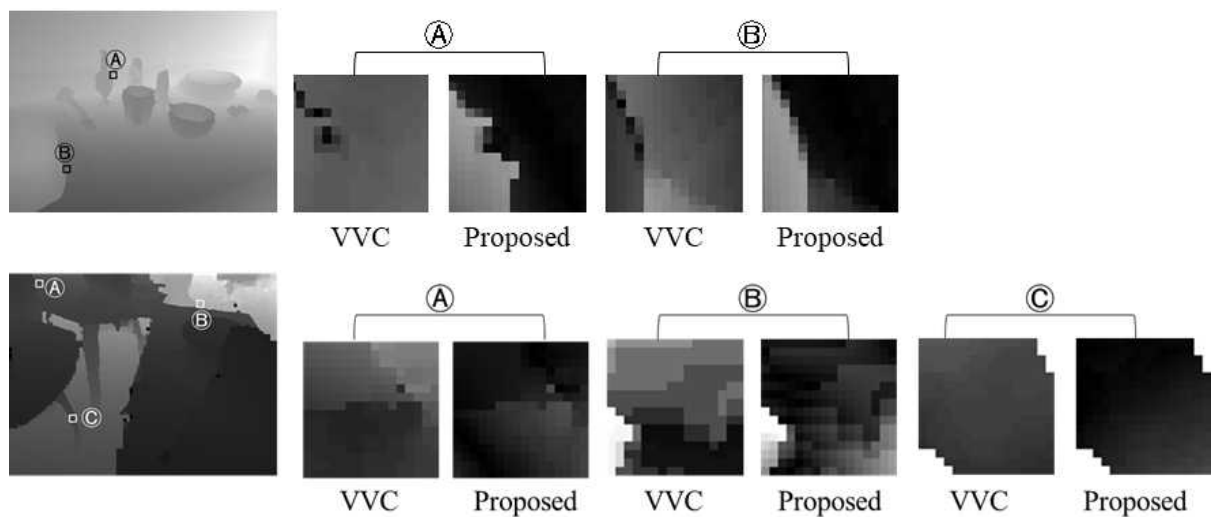


Fig. 4 Comparison of errors between previous and proposed intra prediction methods

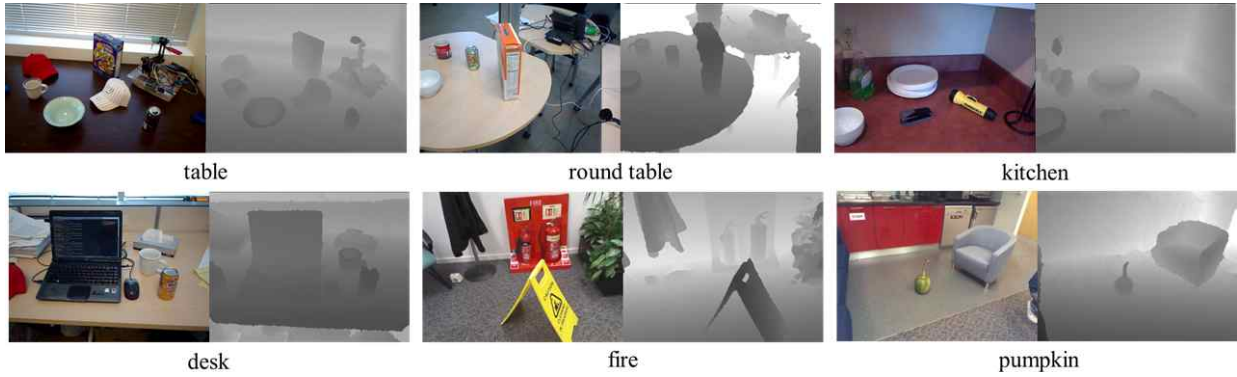


Fig. 5 Depth videos for simulation

Fig. 4는 VVC의 기존 화면 내 예측과 제안된 인트라 모드가 포함된 화면 내 예측을 비교한 것이다. Fig. 4의 왼쪽 이미지는 원래의 깊이 화면을 가시화 한 것이고, 오른쪽 이미지는 화면 내 예측의 오차를 표현한 것으로 밝을수록 오차가 크다는 의미이다. 구면이 포함된 영역에 대해서 기존 화면 내 예측 방법에 비해 제안된 화면 내 예측 방법이 더 예측 정확도가 높다는 것을 확인 할 수 있다.

통상의 깊이 영상들에 대해 제안된 방법을 적용한 깊이 영상 부호화 성능의 개선을 측정하기 위해 Fig. 5의 깊이 영상들(Lai et al., 2011; Silberman et al., 2012)을 이용한다. 깊이 영상들은 Kinect로 촬영되었으며, 해상도는 640x480이다. 깊이 카메라인 Kinect의 초점거리인 식 (1)의  $f$ 는 526.370이며, 각 영상들의 최초 60장의 프레임을 부호화한다. 왜곡 측정에서 척도인 PSNR을 사용하지 않는다. PSNR은 보여지는 화면에 대한 왜곡의 척도로써, 실제 거리 정보를 저장하는 깊이 영상에 대해서는 적절하지 않다. PSNR을 대신해서 3차원 공간상에서의 왜곡을 그 기준으로 잡는 것이 합당하다. 본 모의실험에서 왜곡의 척도로써 영상 내 각 화소들의 3차원 공간상 위치를 비교하여 계산된 RMSE를 척도로 삼는다. 식 (12)에서  $d_i$ 는  $(x_i, y_i)$  영상 좌표가 가지는 원래의 3차원 좌표  $(X_i, Y_i, Z_i)$ 와 제안된 이차 곡면 모델링을 통해 예측된 3차원 좌표  $(\hat{X}_i, \hat{Y}_i, \hat{Z}_i)$ 간의 유클리드 거리를 의미한다. RMSE의 단위는 거리를 의미하는 화소의 단위와 같으며, 모의실험의 영상에서는 mm 단위이다.

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N ((d_i))^2} \\
 d_i^2 &= (\hat{X}_i - X_i)^2 + (\hat{Y}_i - Y_i)^2 + (\hat{Z}_i - Z_i)^2 \\
 &= ((p(x_i, y_i)x_i - \hat{p}(x_i, y_i)x_i)^2 f^{-2} \\
 &\quad + (p(x_i, y_i)y_i - \hat{p}(x_i, y_i)y_i)^2 \\
 &\quad + (p(x_i, y_i) - \hat{p}(x_i, y_i))^2 \\
 &= f^{-2} [(p(x_i, y_i) - \hat{p}(x_i, y_i))^2 (x_i^2 + y_i^2 + f^2)
 \end{aligned}
 \tag{12}$$

Fig. 6은 실험 영상에 대한 비트-왜곡 곡선을 보인다. 제안된 방법은 VVC의 화면 내 예측을 통한 부호화에 비해 성능이 향상되었음을 보인다. Table 1-2는 각각 RMSE와 비트율의 개선을 보인다. 이 때 비트율이 각각 500kbps, 1000kbps로 고정했을 때, RMSE는 2.11%-5.16%의 개선을 보인다. RMSE를 각각 10mm, 15mm으로 고정했을 때, 비트율은 2.15%-5.12%의 개선을 보인다.

Table 1 RMSE reduction by proposed method

Depth Video	RMSE reduction rate (%)	
	500kbps	1000kbps
table	4.13	5.16
round table	2.11	2.63
kitchen	2.85	2.11
desk	2.01	2.29
fire	3.14	2.51
pumpkin	1.97	2.01

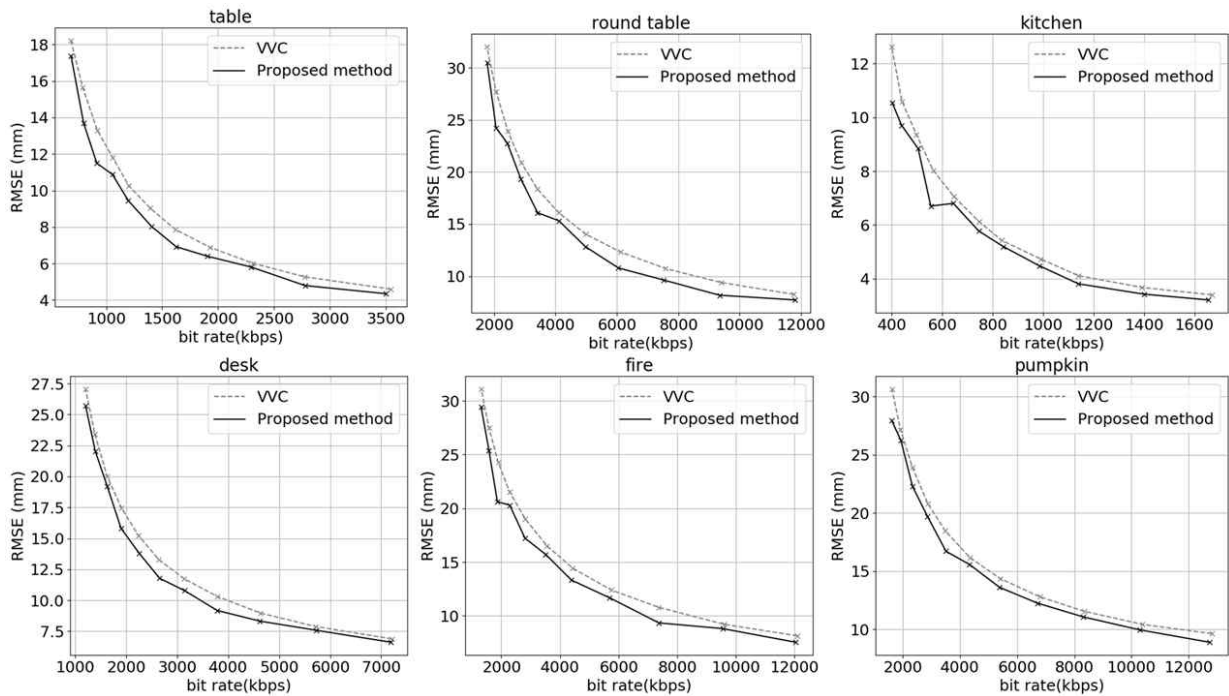


Fig. 6 Rate-distortion curve

Table 2 Bit rate reduction by proposed method

Depth Video	Bitrate reduction rate (%)	
	10mm	15mm
table	4.21	3.37
round table	4.34	2.84
kitchen	3.41	2.15
desk	3.14	2.49
fire	4.23	3.61
pumpkin	3.07	3.18

### 5. 결론

본 논문에서는 깊이 영상의 화면 내 예측에서 이차 곡면 모델링을 통한 인트라 모드를 제안하였다. 각각의 참조 화소들의 후보에 대해 깊이 화면 내 블록의 깊이 값과 제일 오차가 작은 이차 곡면을 모델링하여 이차 곡면의 계수를 구하였다. 이차 곡면의 계수를 통해 참조 화소들의 모델링 오차를 계산하여 참조 화소로 쓰일 화소들을 선택했다. 그 후 선택된 참조 화소들과 예

측 화소 선택 방법을 통해 블록 내 화소들을 예측하였다. 기존 VVC의 화면 내 예측과 비교했을 때 RMSE는 최고 5.16%, 비트율은 최고 5.12% 개선되었다. 이러한 실험 결과에서 제안된 인트라 모드를 기존 영상 부호화 표준의 깊이 영상 부호화에 적용한다면 성능이 개선될 것으로 예측할 수 있다. 최근 스마트 폰의 모바일 장치에 깊이 센서의 장착이 증가하는 추세이며, 자율주행차량에서 주변의 객체들의 정보를 획득하기 위해 거리를 측정하는 라이다 센서의 사용도 증가하고 있는데, 본 연구를 통해 이러한 거리 정보들을 효율적으로 압축할 수 있을 것으로 예측한다.

### References

Aguilar, W. G., Rodriguez, G. A., Alvarez, L., Sandoval, S., Quisaguano, F., and Limaico, A. (2017). Visual SLAM with a RGB-D camera on a quadrotor UAV using on-board processing, *International*



- Work-Conference on Artificial Neural Networks*, Jun. 14-16, Cadiz, Spain, pp. 596-606.
- Bross, B., Wang, Y. K., Ye, Y., Liu, S., Chen, J., Sullivan, G. J., and Ohm, J. R. (2021). Overview of the versatile video coding (VVC) standard and its applications, *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10), pp. 3736-3764.
- Fraunhofer Heinrich Hertz Institute (2022). VVC Test Model, [https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware\\_VTM](https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM) (Accessed on Apr. 14th, 2022)
- Fu, J., Miao, D., Yu, W., Wang, S., Lu, Y., and Li, S. (2013). Kinect-like depth data compression, *IEEE Transactions on Multimedia*, 15(6), pp. 1340-1352.
- Hartley, R. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*, New York, USA, Cambridge University Press.
- Jiang, M., Luo, X., Hai, T., Wang, H., Yang, S., Abdalla, A. N. (2019). Visual Object Tracking in RGB-D Data via Genetic Feature Learning, *Complexity*, 2019(2019), pp. 1-8.
- Kwon, S. K., Tamhankar, A., and Rao, K. R. (2006). Overview of H.264/MPEG-4 part 10, *Journal of Visual Communication and Image Representation*, 17(2), pp. 186-216.
- Lai, K., Bo, L., Ren, X., and Fox, D. (2011). A large-scale hierarchical multi-view RGB-D object dataset, *IEEE International Conference on Robotics and Automation*, May 9-13, Shanghai, China, pp. 1817-1824.
- Lee, D. S. and Kwon, S. K. (2019). Prediction method for depth picture through spherical modeling mode, *Journal of Korea Multimedia Society*, 22(12), pp. 1368-1375.
- Lee, D. S., Kim, B. K., and Kwon, S. K. (2021). Efficient depth data coding method based on plane modeling for intra prediction, *IEEE Access*, 9, pp. 29153-29164.
- Liu, S., Lai, P., Tian, D., and Chen, C. W. (2011). New depth coding techniques with utilization of corresponding video, *IEEE Transactions on Broadcasting*, 57(2), pp. 551-561.
- Nenci, F., Spinello, L., and Stachniss, C. (2014). Effective compression of range data streams for remote robot operations using H.264, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 14-18, Chicago, USA, pp. 3794-3799.
- Ren, C. Y., Prisacariu, V. A., Kahler, O., Reid, I. D., and Murray, D. W. (2017). "Real-time Tracking of Single and Multiple Objects from Depth-colour Imagery Using 3D Signed Distance Functions, *International Journal of Computer Vision*, 124(1), pp. 80-95.
- Shen, G., Kim, W. S., Ortega, A., Lee, J., and Wey, H. (2010). Edge-aware intra prediction for depth-map coding. *IEEE International Conference on Image Processing*, Sep. 26-29, Hong Kong, pp. 3393-3396.
- Silberman, N., Hoiem, D., Kohli, P., and Fergus, R. (2012). Indoor segmentation and support inference from rgb-d images, *European Conference on Computer Vision*, Oct. 7-13 Firenze, Italy, pp. 746-760.
- Stankiewicz, O., Wegner, K., and Domanski, M. (2013). Nonlinear depth representation for 3D video coding, *IEEE International Conference on Image Processing*, Sep. 15-18, Melbourne, Australia, pp. 1752-1756.
- Sullivan, G. J., Ohm, J., Han, W. J., and Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard, *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), pp. 1649-1668.
- Sun, Y., Liu, M., and Meng, M. Q. H. (2017). Improving RGB-D SLAM in dynamic

environments: A motion removal approach, *Robotics and Autonomous Systems*, 89, pp. 110-122.

Wang, X., Sekercioglu, Y., Drummond, T., Fremont, V., Natalizio, E., and Fantoni, I. (2018). Relative pose based redundancy removal: collaborative RGB-D data transmission in mobile visual sensor networks, *Sensors*, 18(8), pp. 1-23.

Wang, X., Sekercioglu, Y., Drummond, T., Natalizio, E., Fantoni, I., and Fremont, V. (2016). Fast depth video compression for mobile RGB-D sensors, *IEEE Transactions on Circuits and Systems for Video Technology*, 26(4), pp. 673-686.

Zhao, Y., Carraro, M., Munaro, M., and Menegatti, E. (2017). Robust Multiple Object Tracking in RGB-D Camera Networks, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 24-28, Vancouver, Canada, pp. 6625-6632.



**이 동 석 (Dong-seok Lee)**

- 정회원
- 동의대학교 컴퓨터소프트웨어공학과 공학석사
- 동의대학교 컴퓨터소프트웨어공학과 공학박사
- 동의대학교 인공지능그랜드

ICT연구센터 연구교수

- 관심분야 : 멀티미디어 신호처리, 영상 인식



**권 순 각 (Soon-kak Kwon)**

- 정회원
- 경북대학교 전자공학과 공학사
- KAIST 전기및전자공학과 공학석사
- KAIST 전기및전자공학과 공학박사

- 동의대학교 컴퓨터소프트웨어공학과 교수
- 관심분야 : 멀티미디어처리, IoT