

깊이 얼굴 영상 부호화에서의 양자화 인자 결정 방법⁺

(Quantization Parameter Determination Method for Face Depth Image Encoding)

박 동 진¹⁾, 권 순 각^{2)*}
(Dong-Jin Park and Soon-Kak Kwon)

요 약 본 논문에서는 얼굴 인식 정확도에 미치는 영향을 최소화하면서 효율적으로 깊이 얼굴 영상을 압축하기 위한 양자화 변수 결정 방법을 제안한다. H.264/AVC의 양자화를 적용하여 깊이 얼굴 영상을 압축 할 때 얼굴 특징을 최대한 유지할 수 있도록 타원체 모델링의 예측 정확도와 각각의 양자화 단위 블록의 얼굴 인식에서의 중요도를 이용하여 양자화 인자를 차등적으로 부여한다. 모의실험 결과 제안된 방법을 통해 같은 압축율에서 얼굴 인식 성공률이 최대 6% 개선되었다.

핵심주제어 : 깊이 영상 압축, 얼굴 인식, 타원체 모델링

Abstract In this paper, we propose a quantization parameter determination method for face depth image encoding in order to minimize an impact on a face recognition accuracy. When a face depth image is compressed through quantization in H.264/AVC, differential quantization parameters are assigned according to an accuracy of ellipsoid modeling prediction and an importance degree of a unit block in extracting facial features. The simulation results show that the face recognition success rates are improved by up to 6% at the same compression rates through the proposed compression rate determination method.

Keywords : Depth video compression, Face recognition, Ellipsoid modeling

1. 서 론

얼굴 인식은 생체 인식 기술 분야 중의 하나로, 촬영된 얼굴에서 신원을 확인하는 인식 기술이다. 얼굴 인식은 비접촉식으로 사용자가 의식하지 않고도 자연스럽게 진행된다는 장점이 있다. 색상 영상을 이용한 얼굴인식 방법은 많은

연구가 진행되었다. 하지만 색상 영상을 이용한 얼굴인식 방법은 조명과 외부요인에 취약하다는 단점이 있다. 이러한 문제점을 해결하기 위해 Kwon (2019)가 화소 값이 거리로 이루어진 깊이 영상을 이용하여 얼굴을 인식하는 방법을 제안하였다. 이 방법은 깊이 영상의 지역이진패턴을 계산하여 얼굴인식을 하는 방법으로, 기존 색상 영상 기반 방법에 비해 어두운 환경에서 얼굴 인식 정확도가 개선되었다.

실시간의 얼굴 인식 시스템 구현을 위해 깊이 얼굴 영상을 고성능의 서버로 전송하는 것이 효율적이다. 이 때 영상이 효율적으로 압축될수록

* Corresponding Author: skkwon@deu.ac.kr

+ 이 논문은 2019년도 BB21+사업에 의하여 지원되었음.

Manuscript received January 10, 2020 / revised January 31, 2020 / accepted February 02, 2020

1) 동의대학교 컴퓨터소프트웨어공학과, 제1저자

2) 동의대학교 컴퓨터소프트웨어공학과, 교신저자

전송에 드는 시간이 줄어들어 실시간 시스템 구현이 용이해진다.

깊이 영상을 압축하는 방법으로는 3D 메쉬 (Wu et al., 2019), 옥트리 (Schnabel and Klein, 2006), 쿼드트리 (Morvan et al., 2007) 등으로 변환하여 그 구조를 압축하는 방법과 색상 영상 부호화 방법을 응용한 방법 (Gumhold et al., 2005; Milani and Calvagno, 2010; Shen et al., 2010; Maitre and Do, 2010; Fu et al., 2013)으로 분류된다. 3차원 구조를 통해 깊이 영상을 압축하기 위해 깊이 영상은 3차원 공간상에서의 점군 (Point cloud)으로 변환된 후 점군으로 부터 3차원 구조를 모델링하여 그 구조를 압축하는 방법이다. 하지만 이 방법은 구조를 모델링하기 위한 계산 복잡도가 크고, 프레임 간의 중복성을 제거하지 못한다는 단점이 있다. 색상 영상 부호화 방법을 응용한 방법은 기존 색상 영상의 화면 내 및 화면 간 예측 방법을 사용하는 방법이다. 하지만 색상 영상 부호화를 적용한 방법은 현재까지 객체와 배경의 경계의 예측 개선에 국한되어 있고 객체 내 예측 방법에 대해서는 아직까지 연구가 부족하다.

최근 연구로는 Wang et al. (2016)의 3D 워핑을 적용한 움직임 추정을 통해 깊이 영상을 압축하는 연구와, Wu et al. (2019)의 깊이 영상에 3D 메쉬 모델링을 수행하여 그 구조를 압축하는 연구 등이 있다. Lee and Kwon (2018)은 깊이 영상을 압축하는 방법으로 깊이 영상이 물체의 표면을 표현한다는 것을 이용하여 평면모델링을 통해 영상을 예측하여 압축하는 방법을 제안함으로써 평면으로 이루어진 객체 내의 예측 정확도를 대폭 개선하였다.

얼굴 인식의 특성에 맞춰 깊이 얼굴 영상을 압축한다면, 인식률을 유지하면서 압축율을 최대한으로 유지할 수 있을 것이다. 본 논문에서는 깊이 영상을 H.264/AVC을 통해 압축할 때, 얼굴 인식률을 일정 수준 이상으로 유지할 수 있는 최적의 양자화 인자를 결정하는 방법을 제안한다. 이를 위해 타원체 모델링을 통해 깊이 얼굴 영상을 예측하는 방법을 제안한다. 기존 얼굴의 3차원 모델은 MPEG-4 표준에서 제안하는 FAMC (Mamou et al., 2008), 자세에 강인한 BFM 모델

(Paysan et al., 2009) 등의 정교한 모델이 제안되었지만, 이들 모델은 얼굴을 모델링하기 위해 필요한 인자가 많아 주어진 얼굴 영상에 대한 인자 결정에 필요한 계산 복잡도가 크기 때문에 실시간 시스템에 적합하지 않다.

반면 본 논문에서 제안하는 타원체 모델링은 기존 방법에 비해 정확도가 비교적 떨어지지만, 모델을 결정하기 위한 인자가 2개 뿐으로, 신속하게 모델링 인자를 계산할 수 있다. 그 후 타원체 모델링을 통해 계산된 예측 영상과 주어진 영상의 잔차를 H.264/AVC의 양자화 과정 및 엔트로피 부호화를 통해 깊이 얼굴 영상을 압축한다. 압축 과정에서의 최적의 양자화 인자를 구하기 위해, 타원체 모델링을 통한 예측 정확도가 낮을수록 얼굴을 인식하는 데에 이용되는 특징이 잘 나타난다는 것을 고려하여, 예측 정확도가 높을수록 복원 영상의 정확도를 더 높이도록 양자화 계단 크기를 작게 설정하여 양자화를 수행한다. 또한 깊이 얼굴 영상 내 눈, 코, 입 등의 얼굴 인식에 중요하다고 여겨지는 부위가 차지하는 영역과 그렇지 않은 영역에 가중치를 차등적으로 부여하여 양자화 인자를 결정한다.

2. 깊이 얼굴 영상 부호화에서의 양자화 인자 결정 방법

2.1 타원체 모델링을 통한 깊이 얼굴 영상의 예측 방법

깊이 얼굴 영상을 얼굴 인식을 처리하는 서버로 전송하기 위해 깊이 영상을 효율적으로 압축하는 것이 중요하다. 이 때 깊이 영상에서 평면 또는 구면으로 예측하여 압축할 때 기존 화면 내 예측 방법보다 효율적이다 (Lee and Kwon, 2018). 사람의 얼굴은 타원형을 가지므로 깊이 얼굴 영상을 타원형으로 예측한다면 정확하게 사람의 얼굴을 예측할 수 있다. 이를 위해 Kwon (2019)가 제안한 얼굴 인식 방법을 적용하여 깊이 영상에서 얼굴을 검출한 후 크기 정규화를 거친 56×56 크기의 얼굴 영상을 생성한 후, 얼굴 영역에서 코 지점의 깊이 값으로 전체 화면의 깊이

화소 값을 뺀다. 이 때 코를 중심으로 하고, 깊이 값이 타원형으로 분포되는 모델을 생성할 수 있다. 이러한 타원면의 표현식은 식 (1)로 표현할 수 있다. 식 (1)에서 a, b, c 는 타원의 인자이다.

$$x^2/a^2 + y^2/b^2 + (z-c)^2/c = 1 \quad (1)$$

식 (1)에서 z 는 x, y 에서의 깊이 값이라는 것을 이용한다면, 식 (2)와 같은 2개의 해를 얻을 수 있다.

$$d_{x,y} = -c(1 - \sqrt{1 \pm 4(x^2/a^2 + y^2/b^2)}) \quad (2)$$

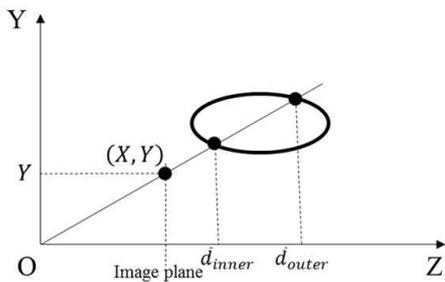


Fig. 1 Prediction through Ellipsoid Modeling

식 (2)에서의 2개의 해의 의미는 Fig. 1과 같이 안쪽면과 바깥면의 두 깊이가 예측이 된다는 의미이다. 하지만 얼굴의 형태는 볼록면밖에 없기 때문에, 오목한 면을 의미하는 바깥면의 깊이 예측값은 필요가 없다. 따라서 식 (3)과 같이 안쪽면을 해당 타원체 모델링을 통한 예측 깊이값으로 한다.

$$d_{x,y} = -c(1 - \sqrt{1 - 4(x^2/a^2 + y^2/b^2)}) \quad (3)$$

얼굴 영상에 대해 식 (1)에서의 타원 인자를 찾아 식 (3)으로 깊이를 예측함에 있어, 그 예측 오차가 최소가 되는 인자들을 찾는 타원체 모델링이 필요하다. 이 때 식 (1)의 c 는 해당 타원에서 z 축에서의 반지름을 나타내는 의미인데, 이는 얼굴에서 코의 깊이 값 d_{nose} 와 코와 수평위치에 있는 얼굴의 좌측 또는 우측의 경계점의 깊이인 d_l 또는 d_r 간의 거리로 근사할 수 있다. 본 논문

에서는 c 를 d_{nose} 와 d_l 또는 d_r 간의 차이 중 그 값이 큰 값으로 설정했다. 이를 통해 찾아야 할 타원의 인자 개수를 줄일 수 있다. 식 (1)의 타원 인자 a, b 를 찾기 위해서는 얼굴 전체 영상에서 코를 중심으로 하는 좌표를 좌표 원점으로 설정하고, 화소 값이 있는 얼굴 영역에서 화소 값에서 코의 화소 값을 뺄므로써, 코 점에서의 화소 값이 0이 되도록 한다. 그 후 얼굴 영역의 화소들에 대해 해당 좌표들을 식 (1)에 대입하여 식 (4)의 행렬식을 얻는다.

$$AR = B$$

$$A = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \dots & \dots \end{bmatrix} B = \begin{bmatrix} -\left(\frac{d_1^2}{\delta^2} + \frac{2d_1}{\delta}\right) \\ -\left(\frac{d_2^2}{\delta^2} + \frac{2d_2}{\delta}\right) \\ \dots \end{bmatrix} \quad (4)$$

$$R = \begin{bmatrix} 1/a^2 \\ 1/b^2 \end{bmatrix}$$

$$\delta = \max(|d_{nose} - d_l|, |d_{nose} - d_r|)$$

식 (3)을 통해 예측된 깊이 값과 실제 깊이 값의 차이인 오차를 최소화 하는 a, b 를 구하기 위해 최소제곱법을 사용한다. 이 때 식 (4)를 통해 구해진 행렬 A 에 대한 의사역행렬 A^+ 를 식 (5)를 통해 계산한 후, $R=A^+B$ 를 통해 R 의 원소인 a, b 를 계산함으로써 최소제곱법을 적용할 수 있다 (Lee and Kwon, 2018). Fig. 2는 얼굴 깊이 화면과 타원체 모델링을 통해 예측된 깊이 값, 그리고 실제로 전송될 예측 화면과의 차이를 보인다.

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

2.2 깊이 얼굴 영상 부호화를 위한 양자화 인자 결정 방법

본 논문에서는 깊이 얼굴 영상의 압축을 위해 타원체 모델링을 통해 예측된 깊이 얼굴 영상의 잔차를 H.264/AVC에 도입된 정수 변환 방법과 양자화 방법을 통해 양자화를 수행한다. 정수 변환은 공간 영역에 분포된 값들을 주파수 영역으

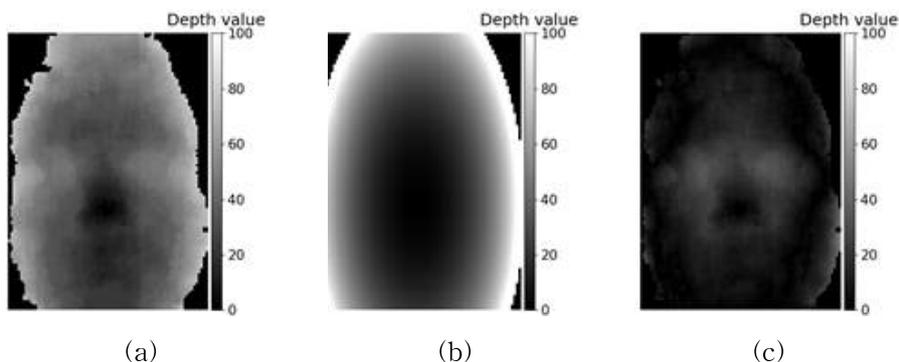


Fig. 2 Predicting Face Depth Image through Ellipsoid Modeling. (a) Original Face Depth Image, (b) Prediction Image through Ellipsoid Modeling, and (c) Residual between Original and Prediction Images

로 변환하는 방법으로, 기존 MPEG에서 쓰이는 DCT와 유사하다. 하지만 정수 변환은 DCT와 달리 변환 수행 중 모든 연산이 덧셈과 시프트 연산으로만 할 수 있기 때문에 연산 복잡도를 줄일 수 있고, 양자화 과정에서 발생하는 손실도 적다는 장점이 있다.

H.264/AVC에서의 정수 변환은 4×4 블록단위로 이루어진다. 이 때 4×4 블록 영상 X 에 대한 정수 변환은 식 (6)을 통해 이루어진다. 식 (6)에서 \times 연산자는 행렬 곱을 뜻하며 \circ 연산자는 두 행렬 간 요소를 곱하는 것을 의미한다. 식 (6)에서의 C_f 과 R_f 는 식 (7)로 정의된다.

$$W = (C_f \times X \times C_f^T) \circ R_f \tag{6}$$

$$C_f = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \tag{7}$$

$$R_f = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/\sqrt{10} & 1/\sqrt{10} & 1/\sqrt{10} & 1/\sqrt{10} \\ 1/2 & 1/2 & 1/2 & 1/2 \\ 1/\sqrt{10} & 1/\sqrt{10} & 1/\sqrt{10} & 1/\sqrt{10} \end{bmatrix}$$

그 후 정수변환을 수행한 4×4 블록에 대해서 H.264/AVC에서 정의된 양자화 방법을 적용한다. H.264/AVC에서는 양자화 계단 크기를 직접 적용하는 것이 아니고, 양자화 인자 QP를 이용하여 양자화 계단크기를 계산하도록 하였다. QP는 6단계마다 양자화 계단 크기가 두 배가 되도록 설정된다. 양자화 인자 QP와 양자화 계단 크기

Q_{step} 의 관계는 식 (8)로 정의된다. 식 (8)에서 %는 나머지를 의미하고, floor(x)는 x를 정수자리에서 버린다는 것을 의미한다.

$$Q_{step}(QP) = 2^{\text{floor}(QP/6)} \cdot Q_{step}(QP \% 6) \tag{8}$$

양자화를 위한 QP에 따른 양자화 행렬 M_f 는 QP가 0~5일 때 식 (9)로 정의된다. 식 (9)에서 정의된 $m(QP, QP \% 6)$ 은 Table 1과 같다. QP가 6이상일 경우, QP를 6으로 나눈 나머지에 해당하는 행렬에 $2^{\text{floor}(QP/6)}$ 를 나누어 사용한다.

$$M_f = \begin{bmatrix} m(QP,0) & m(QP,2) & m(QP,0) & m(QP,2) \\ m(QP,2) & m(QP,1) & m(QP,2) & m(QP,1) \\ m(QP,0) & m(QP,2) & m(QP,0) & m(QP,2) \\ m(QP,2) & m(QP,1) & m(QP,2) & m(QP,1) \end{bmatrix} \tag{9}$$

Table 1 Quantization Matrix M_f

QP	$m(QP,0)$	$m(QP,1)$	$m(QP,2)$
1	13107	5243	8066
2	11916	4660	7490
3	10082	4194	6554
4	9362	3647	5825
5	8192	3355	5243

이 때 정수 변환을 수행한 4×4 블록을 W 라 했을 때, M_f 를 이용한 양자화 블록 Y 는 식 (10)을 통해 구한다. 식 (10)에서 round(x)는 x를 반

올림한 정수이다. Fig. 3은 이러한 양자화의 결과를 보인다. 양자화 결과, 부호화해야 할 계수 값의 크기가 확연히 줄어들었음을 보인다.

$$Y = \text{round} (W \cdot R_f / 2^{15 + \text{floor}(QP)}) \quad (10)$$

58	64	51	58
52	64	56	66
62	63	61	64
59	51	63	69

961	-41	15	-48
-34	72	-30	-104
-15	3	15	24
13	81	-5	8

(a) (b)

48	-1	0	-1
-1	1	-1	-2
0	0	0	0
0	1	0	0

(c)

Fig. 3 Quantization through H.264/AVC for 4×4 Block. (a) Original Block, (b) after Integer Transformation, and (c) Quantization through H.264/AVC

양자화 인자인 QP를 결정하기 위해 타원체 모델링의 예측 정확도를 이용한다. 깊이 얼굴 영상의 타원체 모델링 예측 정확도가 높다는 것은 얼굴의 굴곡이 비교적 적다는 의미로, 얼굴 인식에서의 얼굴의 특징이 비교적 부각되지 않는다는 의미이다. 따라서 이러한 경우에는 QP를 낮게 설정해서 세밀한 얼굴의 특징을 추출하여 얼굴 인식의 정확도를 높여야 한다. 반면에 타원체 모델링 예측 정확도가 낮다면 얼굴의 굴곡이 두드러진다는 뜻으로, QP를 높게 설정하여 어느 정도 영상의 손실이 일어나더라도 얼굴을 인식하기 위한 특징 정보를 얻을 수 있다는 의미이다.

타원체 모델링의 예측 정확도를 측정하기 위해 정규화된 56×56 화소 크기의 얼굴 영상에서 타원체 모델링을 통해 예측된 영상과 실제 얼굴 영상

간의 SSD (Sum of square difference)를 식(11)과 같이 계산한다. 식 (11)에서 $I(i,j)$ 와 $P(i,j)$ 는 각각 깊이 얼굴 영상과 예측 영상의 (i,j) 좌표에서의 화소 값을 의미하며, $SSD(I,P)$ 는 입력 영상과 예측 영상간의 SSD를 의미한다. $SSD(I,P)$ 는 타원체 모델링을 통한 예측 정확도가 높을수록 낮아지며, 완벽하게 타원체 모델링을 통해 영상을 예측했을 경우 $SSD(I,P)$ 는 0이다.

$$SSD(I,P) = \sum_{i=1}^{56} \sum_{j=1}^{56} (I(i,j) - P(i,j))^2 \quad (11)$$

H.264/AVC의 양자화 과정을 통해 부호화된 영상의 PSNR은 양자화 인자 QP와 비례 관계가 있다는 것이 실험적으로 증명되었다 (Takagi et al., 2003; Ma et al., 2005). 이를 식으로 나타내면 식 (12)와 같다. 식 (12)에서 PSNR은 원 영상과 양자화 되어 부호화된 영상 간의 PSNR이다. 또한 l 과 b 는 영상의 특성에 따라 달라지는 상수이다.

$$PSNR = l \times QP + b \quad (12)$$

Fig. 4는 실제 깊이 얼굴 영상에 대해 타원체 모델링을 통한 얼굴 영상 압축 방법을 적용했을 때의 QP와 PSNR간의 관계를 측정된 결과이다. Fig. 4에서 실선은 실제로 측정된 값이고, 점선은 측정된 값들에 대해 선형 회귀 분석을 한 결과이다. 이 때 제안된 타원체 모델링을 통한 얼굴 영상 압축 방법에서도 식 (12)가 성립함을 보인다. 이 때 l 은 -0.6104, b 는 81.25의 값을 가진다.

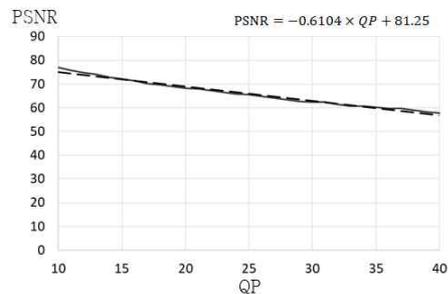


Fig. 4 Relationship between QP and PSNR through Compression of Face Depth Image by Ellipsoid Modeling

이 때 식 (12)를 원 영상과 양자화 되어 부호화된 영상 간의 MSE (Mean of square difference)로 나타내어 QP에 대한 식으로 나타내면 식 (13)과 같다. 이 때 MSE는 식 (14)를 통해 구할 수 있다. 식 (13)에서 max 는 해당 영상에서 화소의 최대값을 의미하며, 색상 영상에서는 그 최대값이 255지만, 깊이 영상에서는 깊이 카메라가 측정할 수 있는 최대 거리를 의미한다. 본 논문에서는 max 를 4000으로 설정한다. 식 (14)에서 $E(i, j)$ 는 양자화를 통해 압축된 영상을 다시 복원한 영상의 (i, j) 좌표에서의 화소값을 의미하고, $MSE(I, E)$ 는 입력 영상과 압축된 영상을 복원한 영상간의 MSE를 의미한다.

$$QP = \text{round} \left(\frac{10}{l} \log \frac{max^2}{MSE(I, E)} - b \right) \quad (13)$$

$$MSE(I, E) = \sum_{i=1}^{56} \sum_{j=1}^{56} (I(i, j) - E(i, j))^2 \quad (14)$$

여기서 $MSE(I, E)$ 는 부호화를 통해 기대되는 손실 정도로 볼 수 있다. 이 때 타원체 모델링의 예측 정확도인 $SSD(I, P)$ 가 높을수록 손실량인 $MSE(I, E)$ 을 낮춰야 한다. 이를 위해 $SSD(I, P)$ 의 평균인 $MSE(I, P)$ 를 이용하여 $MSE(I, P)$ 와 $MSE(I, E)$ 를 식 (15)과 같이 반비례 관계로 둔다. 식 (15)에서 ω 는 상수로써, 예측정확도 $SSD(I, P)$ 의 평균인 $MSE(I, P)$ 에 따라 예측되는 손실 $MSE(I, E)$ 에 대한 가중치이다.

$$MSE(I, P) = \frac{\omega}{MSE(I, E) + 1} \quad (15)$$

식 (15)를 통해 식 (13)을 식 (16)으로 정리할 수 있다. 식 (16)에서 a 는 예측 정확도에 따른 QP의 증가율을 결정하는 인자이며, β 는 초기 QP값과 관련된 인자이다. 식 (16)을 통해 타원체 모델링의 예측정확도에 따라 QP를 결정한다.

$$QP = \text{round} (\alpha \times \log (MSE(I, P) + 1) + \beta) \quad (16)$$

또한 양자화 과정에서 각 블록의 얼굴 인식에서의 중요도에 따라 차등적인 QP를 적용할 수 있다. 이는 얼굴 인식에서 더 중요하게 여겨지는 특징인 눈, 코, 입 등이 포함된 블록에 대해서는 상

대적으로 더 낮은 QP로 양자화를 수행하고, 얼굴 경계부분 등의 얼굴 인식에 있어 중요하지 않는 부분인 블록에 대해서는 높은 QP로 양자화를 수행한다. 이를 통해 가중치가 낮은 블록에 대해서는 높은 블록에 비해 상대적으로 손실을 더 크게 허용함으로써 얼굴 인식의 성능은 유지시키면서 영상 압축을 더 효율적으로 수행할 수 있다. 이 때 얼굴 인식의 중요도에 따라 가중치를 부여하여 차등적인 QP를 부여한다. 얼굴 인식에서 중요한 부위에는 덜 중요한 부위에 비해 더 낮은 QP를 부여하여 얼굴 인식의 정확도를 높인다. 이를 위해 블록의 가중치 w 에 대해 QP를 차등적으로 부여한다. 블록에 부여된 가중치 w 에 대한 차등적인 QP인 QP_w 는 식 (17)을 통해 제공한다. 식 (17)에서의 QP는 식 (16)을 통해 결정된 타원체 모델링의 예측정확도에 따른 QP이다.

$$QP_w = QP + 2 * (3 - w) \quad (17)$$

본 논문에서는 Kwon (2019)가 제안한 얼굴 인식 방법에서 얼굴 인식을 위해 부여한 Fig. 5의 7×7의 가중치를 적용한다. 본 논문에서는 정규화된 56×56 크기의 영상에 대하여 이를 7×7의 영역으로 분할하여 각 영역마다 해당하는 가중치를 적용한다.

0	1	1	1	1	1	0
1	3	3	1	3	3	1
1	1	2	2	2	1	1
1	1	2	2	2	1	1
1	2	2	2	2	2	1
1	2	2	2	2	2	1
0	1	1	1	1	1	0

Fig. 5 Weights for QP Determination

식 (17)를 통해 양자화 과정에서 한 블록 B_i 에 최종적으로 결정되는 $QP(B_i)$ 는 식 (18)과 같다. 식 (18)에서 w_i 는 B_i 에서의 가중치이다.

$$QP(b_i) = \text{round} (\alpha \log (MSE(I, P) + 1) + \beta + 2(3 - w_i)) \quad (18)$$

3. 실험결과

본 논문에서 제안한 깊이 얼굴 영상의 압축 방법과 압축율 결정 방법을 통한 얼굴 영상의 압축률과 얼굴 인식 정확도의 관계를 측정하였다. 모의실험에서 데이터 셋은 Hg et al. (2012)가 사람의 얼굴을 Kinect로 촬영한 깊이 영상들을 사용하였다. 해당 데이터 셋은 30명에 대한 얼굴이 포함된 깊이 영상이 있으며, 각 사람마다 Fig. 7와 같이 얼굴을 바라보는 8가지 방향과 정면에 대한 총 9가지 얼굴 자세에 대해 각각 3번 촬영한 영상이 포함되어 있다. 실험 영상은 Kinect를 통해 촬영되었으며, 영상의 해상도는 640×480이다. 실험에서 얼굴을 인식하는 알고리즘은 Kwon (2019)의 알고리즘을 적용하였다. 본 논문에서 제안된 방법은 깊이 영상에 대해 LBP를 추출하여 얼굴을 인식하는 방법으로, 기존 방법에 비해 깊이 영상에 대한 얼굴 인식률과 인식 처리 시간이 개선된 방법이다.

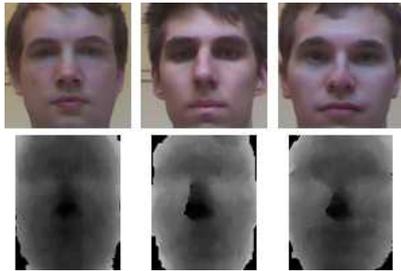


Fig. 6 Face Depth Images for Simulation

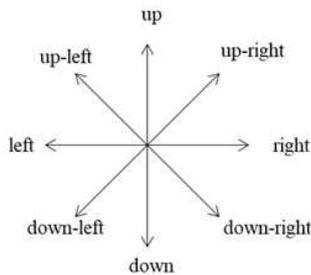


Fig. 7 Face Poses for Simulation

타원체 모델링을 통해 깊이 얼굴 영상을 압축했을 때의 압축율은 QP에 따라 Fig. 8과 같이 나타났다. 이 때 QP는 10~40범위에서 측정하였

다. 영상의 압축율은 식 (19)로 정의된다. 식 (19)에서 $R_c(QP)$ 는 해당 QP에서의 영상의 압축율이며, $b_c(QP)$ 는 해당 QP에서의 비트수이다. b_o 는 원래 영상의 비트수로써, 여기서는 깊이 얼굴 영상으로 쓰이는 56×56 크기에 깊이 영상의 화소의 최대치인 4,000을 표현할 수 있는 비트인 각각 12비트씩을 할당했을 때의 크기, 즉 37,632 비트를 뜻한다.

$$R_c(QP) = \frac{b_o - b_c(QP)}{b_o} \quad (19)$$

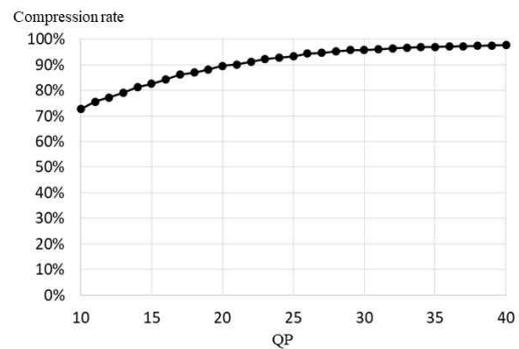


Fig. 8 Compression Rate through Ellipsoid Modeling according to QP

또한 타원체 모델링을 통해 깊이 얼굴 영상을 압축했을 때 생성된 비트수와 원 영상과 복원 영상간의 MSE의 관계는 Fig. 9와 같다. MSE가 낮을수록 압축으로 인한 영상의 손실이 적어지며, 이는 향후 복원을 통한 얼굴 인식에 있어 얼굴 인식 정확도에 영향을 끼치는 정도이다.

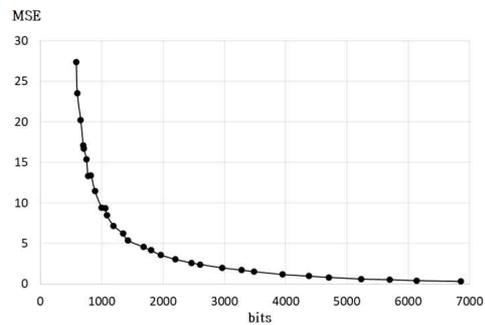


Fig. 9 Relationship between MSE and the Number of Generated Bits

다음은 정면 자세의 깊이 얼굴 영상에 대해 타원체 모델링을 통해 깊이 얼굴 영상을 압축하여 다시 복원했을 때의 얼굴 인식 성공률에 대한 관계를 Fig. 10과 같이 측정하였다. 이 실험에서 얼굴 인식 성공률은 깊이 얼굴 영상이 입력했을 때, 해당 얼굴을 정확한 신원으로 인식을 성공한 비율로 정의한다. QP가 15일 때까지는 얼굴 인식 성공률이 약 90%로 일정하게 유지되었고, QP가 25일 때까지는 얼굴 인식 성공률이 완만하게 감소하였다.

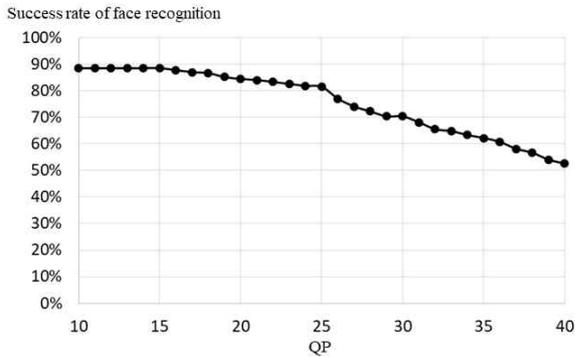


Fig. 10 Success Rate of Face Recognition for Depth Image according to QP

제안한 타원체 모델링의 예측 정확도를 이용한 QP를 차등적으로 부여하는 방법을 적용했을 때에 대해 압축율과 얼굴 인식 성공률을 a 와 β 를 달리하면서 Table 2-3과 같이 측정하였다. 이 때 a 가 아주 크거나 작을 경우, β 가 클 경우에는 QP를 지정해서 양자화하는 경우와 차이가 나지 않았다. 하지만 a 와 β 를 적절한 값으로 설정할 경우, 동 압축율에서 얼굴 인식 성공률이 대폭 개선됨을 보인다. a 가 13, β 가 3일 때 얼굴 인식 성공률이 기존 방법보다 동 압축율 기준 약 3% 개선되어 88%의 얼굴 인식 성공률을 보였다. 또한 a 와 β 를 달리해서 측정한 얼굴 인식 성공률과 압축율 간의 관계를 QP를 달리해서 측정한 관계와 Fig. 11과 같이 비교하였다. 제안하는 QP 결정 방법이 기존 양자화 방법보다 더 효율적임을 보인다.

Table 2 Success Rate of Face Recognition and Compression Rate according to a ($\beta = 2$)

a	Recognition success rate	Compression rate
6	90.0%	72.8%
7	89.3%	77.2%
8	88.9%	79.1%
9	88.5%	82.4%
10	88.1%	84.3%
11	87.0%	88.1%
12	85.6%	90.0%
13	85.2%	90.4%
14	84.1%	91.3%
15	83.3%	92.3%
16	82.2%	93.5%
17	77.8%	94.2%
18	74.4%	94.5%
19	72.6%	95.5%
20	71.1%	95.7%

Table 3 Success Rate of Face Recognition and Compression Rate according to β ($a = 10$)

β	Recognition success rate	Compression rate
0	88.9%	81.6%
1	88.5%	82.8%
2	88.1%	84.3%
3	87.4%	86.3%
4	87.0%	87.0%
5	85.6%	88.5%
6	84.8%	89.9%
7	84.4%	90.5%
8	83.7%	91.4%

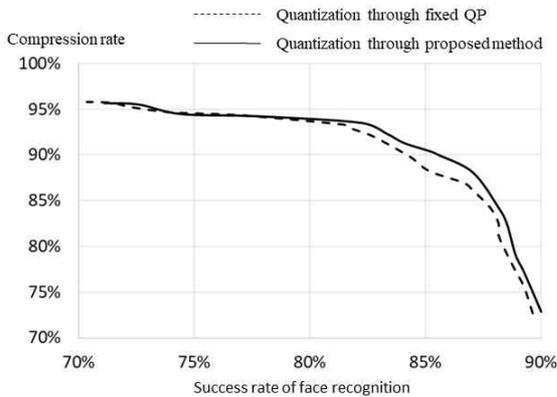


Fig. 11 Comparison between Proposed QP Determination Method and Fixed QP Method

4. 결론

본 논문에서는 깊이 얼굴 영상을 부호화할 때 적용할 수 있는 양자화 인자 결정 방법을 제안하였다. 얼굴 검출 과정에서 검출된 깊이 얼굴 영상에 대해 타원체 모델링을 통한 예측을 하고, 원 영상과의 잔차를 H.264/AVC의 양자화 방법 및 부호화 방법을 적용하여 깊이 얼굴 영상을 압축하였다. 이 때 양자화 과정에서 양자화 인자인 QP는 타원체 모델링의 예측 정확도와 화면 내 얼굴 인식의 중요도에 따라 차등적으로 결정했다.

실험 결과 타원체 모델링을 통한 깊이 얼굴 영상의 예측은 기존 구면 모델링을 통한 예측보다 예측 정확도가 약 35% 개선되었다. 또한 타원체 모델링의 예측 정확도와 화면 내 얼굴 인식 중요도에 따른 차등 양자화를 적용했을 때 깊이 얼굴 영상의 압축률은 얼굴 인식 성공률 85%를 기준으로 할 때 약 88%로 측정되어 양자화 인자를 고정했을 때보다 더욱 효율적인 압축이 수행되었다. 또한 타원체 모델링의 예측 정확도에 따른 양자화 인자 결정에 쓰이는 인자로 α 를 13, β 를 3으로 했을 때 얼굴 인식 성공률이 약 3% 개선되었다. 또한 α , β 를 직접 결정할 수도 있지만, α , β 에 따른 인식 성공률 및 압축율의 관계를 통해 원하는 인식 성공률이나 압축율에서의 최적의 α , β 를 찾을 수 있다.

본 논문에서 제안된 양자화 인자 결정 방법을 통해 깊이 얼굴 영상을 손실이 더 적게 압축함으로써 얼굴 인식 정확도를 개선할 수 있다. 또한 신원 인식을 위해 촬영된 얼굴 영상을 제안된 방법을 통해 압축하여 고성능의 기기로 신속하게 전송하여 처리함으로써 실시간 시스템의 구현이 용이하다. 본 논문에서 제안된 방법은 H.264/AVC의 양자화 과정과 부호화 방법을 이용하였지만, 해당 과정은 H.265/HEVC등의 부호화 표준에 적용된 양자화 및 부호화 과정을 이용할 수도 있다.

향후 얼굴 검출 및 인식 방법 (Lee et al., 2018)에 대해 좀 더 개선된 방법을 적용하더라도, 제안된 깊이 얼굴 영상 압축 방법을 적용하여 실시간 시스템을 유지할 수 있다. 본 논문에서 제안된 방법을 적용한 깊이 영상을 이용한 신원 인식 시스템을 통해 정확하고 거짓 인증 시도에 안전한 신원 인식을 함으로써 신원 인식 시스템이나 출입 관리 시스템 등에 적용할 수 있을 것으로 기대된다.

References

Kwon, S. K. (2019). Face Recognition Using Depth and Infrared Pictures, *Nonlinear Theory and Its Applications, IEICE*, 10(1), 2-15. <https://doi.org/10.1587/nolta.10.2>.

Wu, H., Sun, X., Yang, J., and Wu, F. (2019). 3D Mesh Based Inter-Image Prediction for Image Set Compression, *Proceeding of the IEEE International Conference on Multimedia and Expo*, July 8-12, Shanghai, China.

Schnabel, R., and Klein, R. (2006). Octree-based Point-cloud Compression. *Proceeding of the Eurographics Symposium on Point-Based Graphics*, July 29-30, Boston, Massachusetts, USA.

Gumhold, S., Karni, Z., Isenburg, M., and Seidel, H. (2005) Predictive Pointcloud Compression, *Proceeding of ACM SIGGRAPH*, Jul. 31-Aug. 4, Los Angeles, California, USA.

Morvan, Y., Farin, D., and deWith, P.H.N. (2007).

- Depth-image Compression Based on An R-D Optimized Quadtree Decomposition for The Transmission of Multiview Images, *Proceeding of the IEEE International Conference on Image Processing*, Sep. 16-Oct. 19, San Antonio, Texas, USA.
- Milani, S., and Calvagno, G. (2010). A Depth Image Coder Based on Progressive Silhouettes, *IEEE Signal Process. Letters*, 17(8), 711-714. <https://doi.org/10.1109/LSP.2010.2051619>.
- Shen, G., Kim, W., Narang, S., Orterga, A., Lee, J., and Wey, H. (2010). Edge Adaptive Transform for Efficient Depth Map Coding, *Proceeding of Picture Coding Symposium*, Dec. 8-10, Nagoya, Japan.
- Maitre, M., and Do, M. (2010). Depth and Depth-Color Coding Using Shape Adaptive Wavelets, *Journal of Visual Communication and Image Representation*, 21(5-6), 513-522. <https://doi.org/10.1016/j.jvcir.2010.03.005>.
- 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), 1340-1352. <https://doi.org/10.1109/TMM.2013.2247584>
- Wang, X., Şekercioğlu, Y. A., Drummond, T., Natalizio, E., Fantoni, I., and Frémont, V. (2016). Fast Depth Video Compression for Mobile RGB-D Sensors, *IEEE Transactions on Circuits and Systems for Video Technology*, 26(4), 673-686, <https://doi.org/10.1109/TCSVT.2015.2416571>
- Mamou, K., Zaharia, T., and Preteux, F. (2008). FAMC: The MPEG-4 Standard for Animated Mesh Compression, *Proceeding of the IEEE International Conference on Image Processing*, Oct. 12-15, San Diego, California, USA.
- Paysan, P., Knothe, R., Amberg, B., Romdhani, S., and Vetter, T. (2009). A 3D Face Model for Pose and Illumination Invariant Face Recognition, *Proceeding of IEEE International Conference on Advanced Video and Signal Based Surveillance*, Sep. 2-4, Genova, Italy.
- Lee, D. S. and Kwon, S. K. (2018). Intra Prediction of Depth Picture with Plane Modeling, *Symmetry*, 10, 1-16. <https://doi.org/10.3390/sym10120715>
- Takagi, K., Takishima, Y., and Nakajima, Y. (2003). A Study on Rate Distortion Optimization Scheme for JVT Coder, *Proceeding of SPIE*, Apr. 21, Orlando, Florida, USA.
- Ma, S., Wen, G., and Yan, L. (2005). Rate-Distortion Analysis for H.264/AVC Video Coding and its Application to Rate Control, *IEEE Transactions on Circuits and Systems for Video Technology*, 15(12), 1533-1544. <https://doi.org/10.1109/TCSVT.2005.857300>
- Hg, R. I., Jasek, P., Rofidal, C., Nasrollahi, K., and Moeslund, T. B. (2012). An RGB-D Database using Microsoft's Kinect for Windows for Face Detection, *Proceeding of the IEEE 8th International Conference on Signal Image Technology & Internet Based Systems*, Nov. 25-29, Sorrento, Naples, Italy.
- Lee, D. S., Han D. H., and Kwon, S. K. (2018). Face Recognition Method by Using Infrared and Depth Images, *Journal of the Korea Industrial Information Systems Research*, 23(2), 1-9. <http://dx.doi.org/10.9723/jksiiis.2018.23.2.001>.



박 동 진 (Dong-Jin Park)

- 정회원
- 동의대학교 컴퓨터소프트웨어공학과
공학석사
- 동의대학교 컴퓨터소프트웨어공학과
박사과정

• 관심분야 : 멀티미디어 신호처리, 영상 인식



권 순 각 (Soon-Kak Kwon)

- 정회원
- 경북대학교 전자공학과 공학사
- KAIST 전기및전자공학과 공학
석사
- KAIST 전기및전자공학과 공
학박사

• 동의대학교 컴퓨터소프트웨어공학과 교수

• 관심분야 : 멀티미디어처리, IoT