

## 깊이 맵의 재배열을 통한 개선된 영상 합성 방법

\*김태우 \*\*박진현 \*\*원석호 \*\*\*신지태

성균관대학교 정보통신공학부

{twkim2, innosoulx, jtshin}@skku.edu, shwon@etri.re.kr

### Improved Video Synthesis Method by Depth Map Rearrangement

\*Kim, Tae-Woo \*\*Park, Jin-Hyun \*\*Won, Seok-Ho \*\*\*Shin, Jitae

Sungkyunkwan University

#### 요약

본 논문에서는 깊이 맵의 재배열 과정을 통해서, 보다 개선된 영상을 합성하는 방법을 제안한다. 제안하는 방법은 전체 깊이 맵을 여러 그룹(Group)으로 나누고, 각각의 그룹에 서로 다른 가중치를 주어 가까운 물체에 좀 더 많은 깊이 값을 가질 수 있도록 조절하였다. 깊이 맵 추정(Depth Estimation) 및 중간 시점 영상의 합성(View Synthesis)을 통하여 기존 방식과의 비교를 진행하였고 그 결과, 전체적인 비디오 시퀀스(Video Sequence)에 대한 PSNR은 유지하면서, 보다 시각적으로 자연스러운 영상을 얻을 수 있었다.

#### 1. 서론

최근 “아바타”와 같은 3D 영화와 3차원 TV에 대한 관심이 증가하고, 이를 위한 3차원 콘텐츠를 생성하려는 많은 노력이 전개되고 있다. 가까운 미래에는 사람들이 3차원 깊이 값을 체험할 수 있을 뿐만 아니라, 다시점 비디오를 이용해 시청자가 원하는 임의 시점을 선택하여 볼 수도 있을 것이다. 한 장면을 여러 시점에서 시청할 수 있는 자유 시점 기능을 구현하기 위해서는 다시점 영상을 이용하여 시점 사이에 존재하는 가상의 중간 시점 영상을 합성하는 기술이 필요하다.

중간 시점 영상을 합성하기 위해서는 깊이 맵을 사용하는데, 깊이 맵이란 영상 내에 존재하는 객체들의 3차원 거리 정보를 나타내는 영상으로서, 각 화소 값은 해당 화소의 깊이 정보를 나타낸다. 깊이 맵의 정확도는 합성된 중간 시점 영상의 화질을 좌우하기 때문에 정확한 깊이 맵을 생성하는 것이 관건이다.<sup>[1]</sup>

본 논문의 구성은 다음과 같이 이루어진다. 2장에서는 관련 연구와 기존 방법들을 알아보고, 3장에서 제안하는 깊이 맵 생성의 새로운 방법에 대해 설명한다. 4장에서는 실험을 통해 제안된 방법의 성능을 평가하며, 마지막으로 5장에서는 결론과 앞으로의 연구 방향에 대해서 모색해 본다.

#### 2. 관련 연구

변이-깊이 변환 방법에는 두 가지 방법이 있다. 첫 번째는 8비트의 깊이 맵 값의 크기(intensity)  $v$ 에 따라서 깊이 레벨을 선형적으로 나누는 방법으로서 변이-깊이 변환 식은 다음과 같다. 여기서  $Z_{far}$ 와  $Z_{near}$ 는 최대 깊이 정보와 최소 깊이 정보를 나타내고,  $z$ 는 현재

화소에 대한 깊이 정보를 의미한다.

$$v = 255 - \frac{255 \cdot (z - Z_{near})}{Z_{far} - Z_{near}} \quad (1)$$

위의 방법은  $Z_{far}$ 와  $Z_{near}$  사이의 3차원 공간을  $z$  축을 중심으로 256개의 동일한 간격을 갖도록 분할한다. 하지만, 대부분의 3차원 장면에서 객체는 화면 앞쪽으로 몰려 있도록 촬영하기 때문에, 공간을 동일하게 분할하는 것은 비효율적이다.

두 번째 방법은 선형 방법의 비효율성을 극복하기 위해 객체가 몰려 있는 앞쪽에서는 좁은 간격으로, 배경으로 갈수록 넓은 간격으로 3차원 공간을 분할하는 비선형 방법으로 식은 다음과 같다.

$$v = 255 \cdot \frac{Z_{near}}{z} \cdot \frac{Z_{far} - z}{Z_{far} - Z_{near}} \quad (2)$$

카메라들이 3차원 좌표계의  $x$ 축과 Parallel하게 놓여 있다고 가정하면 다음의 관계식들을 유도할 수 있다. 여기서  $f$ ,  $b$ 는 각각 카메라 초점거리, 카메라 간의 거리이고  $d$ ,  $\Delta d$ 는 각각 그래프 컷 알고리즘을 통해 계산된 변이 값과 카메라 오프셋(offset) 값이다.<sup>[2]-[3]</sup>

$$z = \frac{f \cdot b}{d + \Delta d}; Z_{near} = \frac{f \cdot b}{d_{max} + \Delta d}; Z_{far} = \frac{f \cdot b}{d_{min} + \Delta d} \quad (3)$$

깊이 맵에서의 크기 값(intensity)을 나타내는  $v$ 는 다음 식과 같이 0부터 255사의 값으로 나타낼 수 있다.

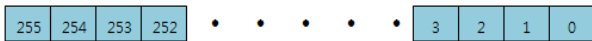
$$v = 255 \cdot \frac{d - d_{\min}}{d_{\max} - d_{\min}} \quad (4)$$

이미지(image) 전체의 각 픽셀(pixel)에 대한  $v$  값의 set이 깊이 맵을 구성하게 된다. 또한, 식 (2)를  $v$ 에 관한 관계식으로 정리하면 다음의 일반화된 식을 얻을 수 있다.

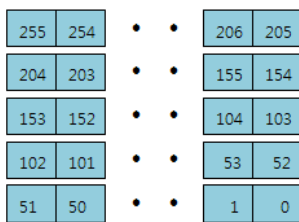
$$z = \frac{1}{\frac{v}{255} \cdot \left( \frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}}} \quad (5)$$

### 3. 제안된 방법

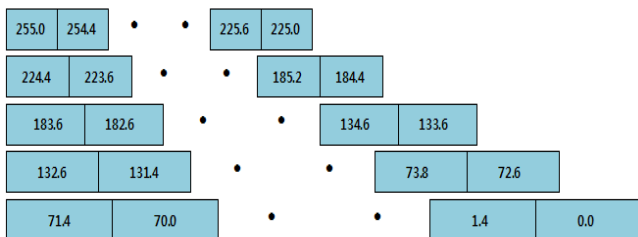
본 논문에서 제안하고자 하는 방법은 현재 MPEG에서 배포한 깊이 맵 추정 소프트웨어 *DERs* (Depth Estimation Reference Software)<sup>[4]</sup>에서 사용되고 있는 변이-깊이 변환 방법 중 비선형 공간 분할 방식은 그대로 유지하되, 깊이 맵(즉,  $v$  값들의 set)을 여러 개의 *Groups*으로 나누고 여기에 각기 다른 가중치를 두어서  $v$  값들을 재배열 시켜서 가까운 물체에 좀 더 많은 레벨을 주어서 깊이 감을 향상 시키는 데 있다.



(a) 기존의 깊이 맵에서의  $v$ 간의 배열(255-0)



(b) 그룹화( $v$ 간의 간격은 1로 동일)



(c) 가중치 적용(그룹별로  $v$ 간의 간격(=가중치  $\lambda$ )을 조절)

그림 1. 깊이 맵 도식화

우리는 여러 실험을 통해서 깊이 맵을 나누는 그룹의 수를 5개로 ( $G1, G2, G3, G4, G5$ ) 정하였고, 또한 카메라에 가장 가까운 그룹( $G1$ )부터 각 그룹에 가중치( $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ )를 각각 (0.6, 0.8, 1.0, 1.2, 1.4)로 적용하였다. 여기서 각 그룹에 적용되는 가중치는  $v$ 간의 간격을 의미한다.

식 (4)에서 나온  $v$  값은 그 크기에 따라 그룹이 정해지고 아래와 같은 의사 코드(pseudo code)를 통해 새로운 깊이 맵 크기 값인  $v'$ 으로 바뀌게 된다.

한편, 의사 코드 내의 [224.4, 183.6, 132.6, 71.4] 값들은 새롭게 재배열된 깊이 맵의 그룹  $G2', G3', G4', G5'$ 에서의  $v'$ 의 최대 값(maximum)을 의미한다. 여기서  $f(x) = \lfloor x + 0.5 \rfloor$  라고 규정하면 의사코드(pseudo code)는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} & \text{for all } v \\ & \begin{cases} \text{if}(v \in G1) & v' = f(224.4 + (v - 204) \cdot \lambda_1) \\ \text{else if}(v \in G2) & v' = f(183.6 + (v - 153) \cdot \lambda_2) \\ \text{else if}(v \in G3) & v' = f(132.6 + (v - 102) \cdot \lambda_3) \\ \text{else if}(v \in G4) & v' = f(71.4 + (v - 51) \cdot \lambda_4) \\ \text{else} & v' = f(v \cdot \lambda_5) \end{cases} \\ & \text{if}(f(v) = f(v-1)) \quad f(v) = f(v) + 1 \\ & \begin{cases} G1 = \{v | 205 \leq v \leq 255\}, G2 = \{v | 154 \leq v \leq 204\}, \\ G3 = \{v | 103 \leq v \leq 153\}, G4 = \{v | 52 \leq v \leq 102\}, \\ G5 = \{v | 0 \leq v \leq 51\} \end{cases} \end{aligned}$$

그림 2. 재배열 의사 코드(pseudo code)

이후,  $v'$ 는 식 (5)의 입력이 되어 변이-깊이 변환이 이루어지게 된다.

예를 들어, 깊이 맵 추정 소프트웨어인 *DERs* 내에서 변이 탐색 범위를 1부터 55, 1/2화소 단위로 설정하면,  $2 \cdot (55 - 1) + 1$ , 총 109개의  $v$  값이 변이-깊이 변환이 이루어진다. 위의 설정에 대해서 전체  $z$  범위를  $(Z_{far} - Z_{near})/5$ 의 동일한 간격으로 5개의 그룹으로 나누었을 때(카메라로부터 가장 가까운 그룹  $Z_1$ , 가장 먼 그룹  $Z_5$ ) 각 그룹 내에 존재하는 깊이 값의 개수를 기존 방식과 비교를 하면 다음과 같은 결과를 얻을 수 있다.

표 1. 모델 별 깊이 그룹에 따른 레벨의 개수

깊이 모델	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	Total
기존의 방법	53	25	15	9	7	109
제안된 방법	66	21	10	7	5	109

사람의 눈은 먼 물체에 비해 가까운 물체에 대해서 보다 민감하게 반응한다. 여기서 기존 방식에 비해  $(Z_{near}, (Z_{far} - Z_{near})/5)$ 의 깊이 범위에서 13개의 깊이 레벨을 더 갖는 제안된 방식이 가까운 물체에 대해 좀 더 세밀한 깊이 감을 나타낼 수가 있다.

### 4. 실험 방법 및 결과

#### 4.1 깊이 맵 성능 평가 및 PSNR 측정

본 논문에서 제안하는 깊이 맵의 재배열 방법의 성능을 평가하기 위해 MPEG의 3차원 비디오 부호화 그룹에서 테스트 영상으로 사용 중인, "kendo", "champagne\_tower", "lovebird1", "pantomime", "newspaper" 등의 영상을 사용하였다. [5]-[6] 깊이 맵 추정용 MPEG 3차원 비디오 부호화 그룹에서 제공한 깊이 추정 참조 소프트웨어 (Depth Estimation Reference Software, DERS) 버전 5.1를 사용하였다.

그림 3은 "lovebird1" 영상에 대한 깊이 맵을 나타낸다. 물체가 가까이 위치했을 때 기존 방식보다 제안하는 방식에서 좀 더 자연스러운 깊이 맵이 생성되는 것을 볼 수 있다.

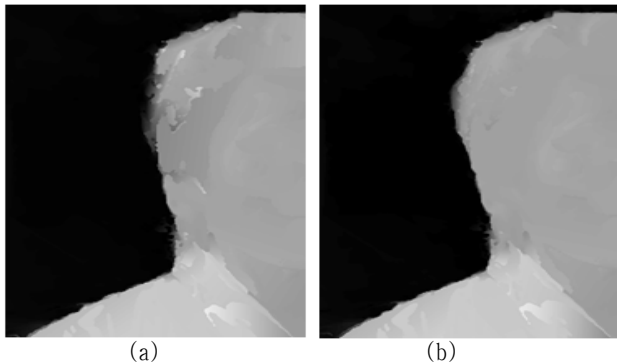


그림 3. "lovebird1" 깊이 맵 추정 결과(시점 : 7)  
(a) 기존 방식, (b) 제안하는 방식

표 2. 합성 영상의 평균 화질 비교

테스트 영상	시점	프레임 수	기존의 방법 (PSNR)	제안하는 방법 (PSNR)	$\Delta PSNR$
kendo	2	300	32.94	32.94	0
champagne_tower	42	200	31.04	30.90	-0.14
lovebird1	6	300	29.61	29.68	+0.07
pantomime	32	300	32.70	32.80	+0.1
newspaper	5	300	31.86	31.79	-0.07
balloons	2	200	34.91	34.78	-0.13

표 2는 각각의 테스트 영상에 대해서, 원본 영상과의 PSNR을 측정하여 기존 방식과 제안하는 방식과의 비교를 나타내었다. 결과를 보면 제안하는 방식은 기존 방식의 PSNR 값을 큰 차이 없이 유지하는 것을 확인할 수 있다.

#### 4.2 합성 영상의 성능 평가

여러 합성된 영상의 비교를 다음과 같이 나타내었다.

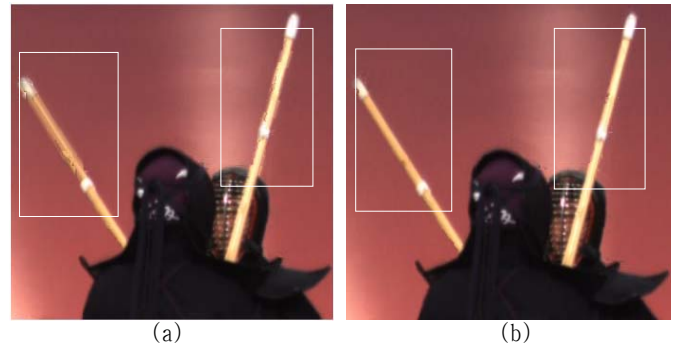


그림 4. "kendo" 영상 (a) 기존 방식, (b) 제안하는 방식

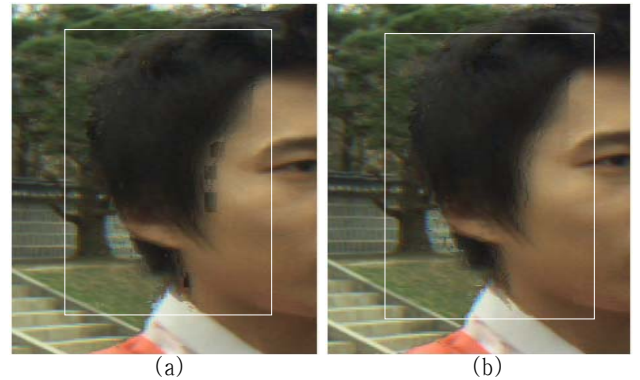


그림 5. "lovebird1" 영상 (a) 기존 방식, (b) 제안하는 방식

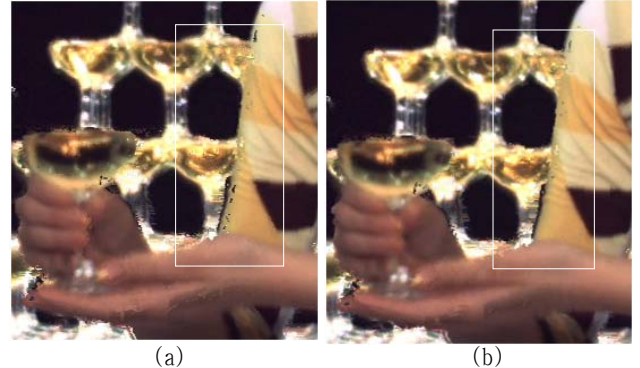


그림 6. "champagne tower" 영상 (a) 기존 방식, (b) 제안하는 방식

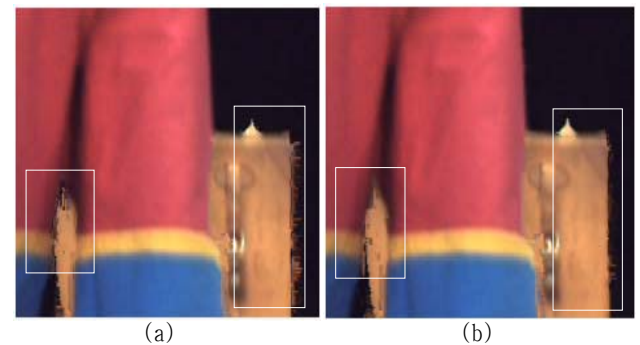


그림 7. "pantomime" 영상 (a) 기존 방식, (b) 제안하는 방식

여러 테스트 영상들에 대한 합성 영상을 분석한 결과 횡적인 움직임(좌/우) 보다 종적인 움직임(앞/뒤)이 있는 영상에 대해서 제안된 방식이 효과를 나타낼 수 있었으며, 특히 카메라 쪽으로 다가오는 객체들에 대해서는 시각적으로도 보다 선명하고 자연스러운 영상을 획득할 수 있었다.

## 5. 결론

본 논문에서는 개선된 영상 합성을 위한 깊이 맵의 재배열 방법을 제안하였다. 비선형 공간 분할 방식은 그대로 유지하되, 깊이 맵을 그룹화하여서 각기 다른 가중치를 주어 깊이 맵을 재배열 하였다. 그 결과, 기존 방식과 큰 차이 없이 전체적인 PSNR 값은 유지하면서, 가까운 물체에 대해 좀 더 자연스럽게 선명한 영상을 얻을 수 있었다. 또한, 사람의 눈이 보다 민감한(sensitive) 가까운 거리(전체 깊이의  $1/5$ )에서 기존 방식보다 약 20% 많은 깊이 레벨을 갖는 효과를 가져 올 수 있었다.

## 감사의 글

본 연구는 지식경제부 산업기술이사회 정부출연금사업의 일환으로 수행하였음. [유비쿼터스 인프라 연구, 고품질 방송통신융합 서비스를 위한 Enhanced QoS/QoE 핵심기술 개발]

## 참고문헌

- [1] 호요성, 이천, "3차원 비디오 부호화를 위한 국제 표준화 기술," 한국방송공학회 학술지, 제14권, pp. 031-044, 2009. 06
- [2] 유지상, 최미남 "다시점 영상을 위한 중간 시점 영상 합성", 한국통신학회논문지 제24권 제4호, 2007
- [3] ISO/IEC JTC1/SC29/WG11/"Report on Experimental Framework for 3D Video Coding", N11631, October 2010
- [4] ISO/IEC JTC1/SC29/WG11 "Reference Softwares for Depth Estimation and View Synthesis," M15377, April 2008.
- [5] ISO/IEC JTC1/SC29/WG11 "Contribution for 3D Video Test Material of Outdoor Scene.", M15371, April 2008
- [6] ISO/IEC JTC1/SC29/WG11 "1D Parallel Test Sequence for MPEG-FTV.", M15378, April 2008
- [7] ISO/IEC JTC1/SC29/WG11 "HHI Test Material for 3D Video," M15413, April 2008.
- [8] ISO/IEC JTC1/SC29/WG11 "Multiview Video Test Sequence and Camera Parameters," M15419, April 2008.

[9] MPEG-FTV Test Sequence Download Page,

[http://www.tanimoto.nuee.nagoya-u.ac.jp/mpeg/mpeg\\_ftv.html](http://www.tanimoto.nuee.nagoya-u.ac.jp/mpeg/mpeg_ftv.html)