

움직임 객체 기반의 하이브리드 깊이 맵 다운샘플링 기법

김태우*, 김정훈*, 박명우*, 신지태^o

Hybrid Down-Sampling Method of Depth Map Based on Moving Objects

Tae-Woo Kim*, Jung Hun Kim*, Myung Woo Park*, Jitae Shin^o

요약

3차원 비디오 전송에 있어서, 렌더링에 사용되는 깊이 맵은 압축 효율을 위하여 일반적으로 낮은 해상도로 다운샘플링 되어 전송된다. 이때, 다운샘플링 과정에서 발생하는 오류는 디코딩 후의 적절한 업샘플링 기법을 통하여 효과적으로 복구할 수 있다. 하지만 기존의 연구들은 이러한 오류를 줄이기 위하여 여러 업샘플링 기법에 대해서만 초점이 맞춰져 있었다. 본 논문에서는 사람의 인지적인 화질을 보다 제고시키기 위하여 동적인 객체와 정적인 배경에 서로 다른 다운샘플링 비율을 적용시키는 새로운 하이브리드 깊이 맵 다운샘플링 기법을 제안한다. 실험적인 결과는 제안하는 방법이 시각적인 품질과 PSNR 측면에서의 이득이 있음을 보여준다. 또한 제안하는 방법은 여러 업샘플링 기법들과 호환이 가능하다.

Key Words : 3D Video Transmission, Depth Map, Depth-Based Image Rendering (DIBR), Down-sampling, Moving Objects, 3차원 비디오 전송, 깊이 맵, 깊이 맵 기반 렌더링, 움직임이 있는 객체

ABSTRACT

In 3D video transmission, a depth map being used for depth image based rendering (DIBR) is generally compressed by reducing resolution for coding efficiency. Errors in resolution reduction are recovered by an appropriate up-sampling method after decoding. However, most previous works only focus on up-sampling techniques to reduce errors. In this paper, we propose a novel down-sampling technique of depth map that applies different down-sampling rates on moving objects and background in order to enhance human perceptual quality. Experimental results demonstrate that the proposed scheme provides both higher visual quality and peak signal-to-noise ratio (PSNR). Also, our method is compatible with other up-sampling techniques.

I. 서론

영화 아바타의 성공 이후로 3D 산업은 영화뿐만 아니라 게임, 방송, 홈 엔터테인먼트, 스마트 폰에 이르기까지 다양한 산업에서 새로운 서비스 모델로서의

가치를 확고히 하고 있다. 특히, 사용자에게 현장감과 몰입감을 제공하는 다양한 매체들을 기반으로 한 방송 서비스, 즉 실감 방송 영역은 전 세계적으로 여러 컨소시엄과 3D 관련 과제들을 통해서 연구되어지고 있다^[1]. 3차원 비디오는 2D 비디오 영상과 그에 따른

* 주저자 : 성균관대학교 정보통신대학 전자전기공학부, twkim2@skku.edu, 정회원

^o 교신저자 : 성균관대학교 정보통신대학 전자전기공학부, jtshin@skku.edu, 정회원

* 성균관대학교 정보통신대학 전자전기공학부, kimenjoy@skku.edu, parkmw@skku.edu

논문번호 : KICS2012-07-339, 접수일자 : 2012년 7월 31일, 최종논문접수일자 : 2012년 10월 30일

깊이 맵 영상, 그리고 다 시점 비디오 영상과 그에 따른 깊이 맵 영상으로 표현될 수 있다^[2]. 3차원 상에서 카메라와 객체간의 상대적인 거리를 나타내는 깊이 맵은 2D 비디오(texture) 영상과 비교하여 다른 특징들을 보인다. 깊이 맵 영상은 회색 톤으로(gray-scale) 표현되고 날카로운 경계면(edge) 부분을 제외하고 대부분 평탄한 값들로 분포된다^[3]. 이러한 특징들로 인해 깊이 맵 압축 시 다운샘플링 과정이 일반적으로 이루어지는데 이를 통한 여러 효과들을 가져올 수 있다. 즉, 압축 효율 측면에서 전송되는 데이터 양을 줄일 수가 있고, 평탄한 값으로 이루어져 있는 객체 내부의 세부 정보들을 줄일 수가 있으며 깊이 맵 상의 비정상 부분(outliers)을 제거할 수가 있다^[4]. 디코딩 후의 깊이 맵은 전송 전의 원래 크기대로 업샘플링 되는데 이때, 적절한 업샘플링 기법들을 통하여 다운샘플링 시의 오류를 효과적으로 복구할 수 있다. 깊이 맵 다운/업샘플링을 이용한 3차원 스테레오스코픽(stereoscopic) 시스템에 대한 흐름도는 아래 그림 1과 같다.

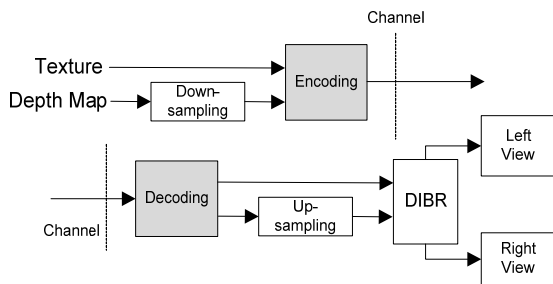


그림 1. 깊이 맵 다운/업샘플링을 이용한 스테레오스코픽 시스템 흐름도
Fig. 1. System diagram of a stereoscopic system using depth map down/up-sampling

깊이 맵 기반의 렌더링(DIBR : Depth Image Based Rendering) 측면에서 깊이 맵의 정확성은 매우 중요하다. 이는 가상 시점에서의 합성 영상의 품질과 직결되기 때문이다^[5]. 즉, 텍스처 상의 오류는 해당 픽셀의 약간의 밝기 차이만을 가져오지만, 깊이 맵 상의 오류는 해당 픽셀의 이동되는(shift) 위치가 달라진다는 점이다^[6]. 따라서 깊이 맵 전송에 있어서 세련된 다운/업샘플링 기법들이 요구된다. 그러나 기존의 연구에서는 다운샘플링 과정에서 발생된 오류에 대한 적절한 회복을 목적으로 하여 업샘플링 기법에 대해서만 초점이 맞추어져 있다^[7-11]. 최근의 다운샘플링 기법들은 전체 해상도(full resolution) 상의 여러 픽셀들을 다운샘플링된 하나의 해상도 상의 픽셀로 어떻게 매핑

(mapping) 시키느냐에 대한 단순한 과정이다^[7,9,12]. 또한 이러한 기법들에서는 사람이 느끼는 인지적인 특징(Human Visual Perception(HVP))을 고려하지 않았다.

본 논문에서는 깊이 맵을 이용한 스테레오스코픽 시스템 하에서의 새로운 깊이 맵 다운샘플링 방법을 제안한다. 이는 정적인 배경(static background) 보다 움직임을 있는 동적인 객체(dynamic object)에 보다 민감하게 반응하는 사람의 인지적인 특성을 보다 반영시키기 위하여 배경과 객체에 서로 다른 다운샘플링 비율(down-sampling ratio)을 적용시키는 하이브리드 깊이 맵 다운샘플링 기법을 말한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 깊이 맵 다운샘플링 기법들을 소개하고, 3장은 제안하는 하이브리드 다운샘플링 기법에 대해서 설명한다. 4장에서는 제안하는 방법을 다양한 여러 업샘플링 기법에 적용하여 그 성능을 기존의 방법들과 비교 실험하였으며, 5장에서는 본 논문에 대한 결론을 맺는다.

II. 하이브리드 깊이 맵 다운샘플링

2.1. 기존의 깊이 맵 다운샘플링 기법들

깊이 맵은 일반적으로 전송되는 데이터의 양을 줄이기 위해서 다운샘플링 과정을 거친다. $f(i, j)$ 를 이미지(image)상의 (i, j) 위치에서의 픽셀 값이라고 한다면, 기존의 깊이 맵 다운샘플링 기법들로는 원래의 다운샘플링^[9](original down-sampling), 평균 값 다운샘플링^[12](average down-sampling), 중간 값 다운샘플링^[7](median down-sampling)들로 다음과 같이 표현될 수 있다.

$$\begin{aligned}
 \text{original} : f_{\text{down}}(i/DR, j/DR) &= f(i, j) \\
 \text{average} : f_{\text{down}}(i/DR, j/DR) &= \text{average}(W_{DR \times DR}) \\
 \text{median} : f_{\text{down}}(i/DR, j/DR) &= \text{median}(W_{DR \times DR}) \\
 \text{For } i &= 0, 1 \times DR, 2 \times DR, 3 \times DR, \dots \text{Width} \\
 j &= 0, 1 \times DR, 2 \times DR, 3 \times DR, \dots \text{Height}
 \end{aligned} \tag{1}$$

여기서 DR 은 다운샘플링 비율(down-sampling ratio)이고, $W_{DR \times DR}$ 은 다운샘플링이 적용되는 이미지 상의 윈도우(window)를 나타낸다. f_{down} 은 다운샘플링이 적용된 이미지 상의 픽셀 값을 의미한다. 기존의 방식들은 즉, $W_{DR \times DR}$ 윈도우 내의 픽셀들

을 어느 하나의 대표 값 f_{down} 으로 매핑(mapping)을 시키는지에 대한 방법들로서 윈도우 내의 첫 번째 픽셀 값을 대표 값으로 정하는 경우(original down-sampling), 윈도우 내의 픽셀 값들의 평균값으로 대표 값을 정하는 경우(average down-sampling), 윈도우 내의 픽셀 값들 중 중간값으로 대표 값을 정하는 경우(median down-sampling)가 있다. 여기서 주목할 점은 기존의 깊이 맵 다운샘플링 기법들은 모든 이미지 상의 단 한 개의 다운샘플링 비율로 균일하게(uniformly) 적용시킨다는 점이다.

2.2. 제안하는 하이브리드 깊이 맵 다운샘플링 방법

기존 방식들과 비교하였을 때, 제안하는 다운샘플링 방법은 사람의 인지적 특성을 보다 반영시키기 위하여 하이브리드 방식을 채택하여 움직임이 있는 객체 부분과 움직임이 없는 배경 부분에 서로 다른 다운샘플링 비율을 적용 시킨다.

$$DR_O (= UR_O) < DR_R (= UR_R) < DR_B (= UR_B) \quad (2)$$

여기서 DR_O 와 DR_B 는 각각 객체와 배경에 적용되는 다운샘플링 비율이고 UR_O 와 UR_B 는 각각 객체와 배경에 적용되는 업샘플링 비율(up-sampling ratio)이다. 이때, 객체와 배경에 적용되는 다운/업샘플링 비율은 서로 같다. 기존 방식(하나의 다운샘플링 비율을 이미지 전체에 적용시키는 방식)과의 비교를 위하여 DR_R 와 UR_R 을 각각 기존 방식에서의 기준점(reference) 다운/업샘플링 비율이라고 한다. 수식 (2)와 같이 기존 방식에서 사용되는 다운샘플링 비율 보다 작은 값(비율)을 객체 부분에 적용시키고 보다 큰 값(비율)을 배경 부분에 적용시킨다. 이는 보다 낮은 다운샘플링 비율을 객체 부분에 적용시킴으로서 객체 부분에서 시각적 품질 측면에서의 개선을 얻을 수 있다. 반대로 배경 부분에서는 시각적 품질 측면에서 저하가 나타나는데, 이는 고정된 비트 제한량(constant bit budget) 때문이다. 즉, 사람의 인지적 특성을 보다 반영시키기 위하여 본 논문에서는 사람이 보다 민감하게 반응하지 않는 정적인 배경 부분에서의 저하 효과를 감수하되, 보다 민감하게 반응하는 움직임이 많은 객체 부분에서의 개선 효과를 증대시키는데 있다.

그림 2은 제안하는 하이브리드 깊이 맵 다운샘플링 방식에 대한 전체 흐름도이다.

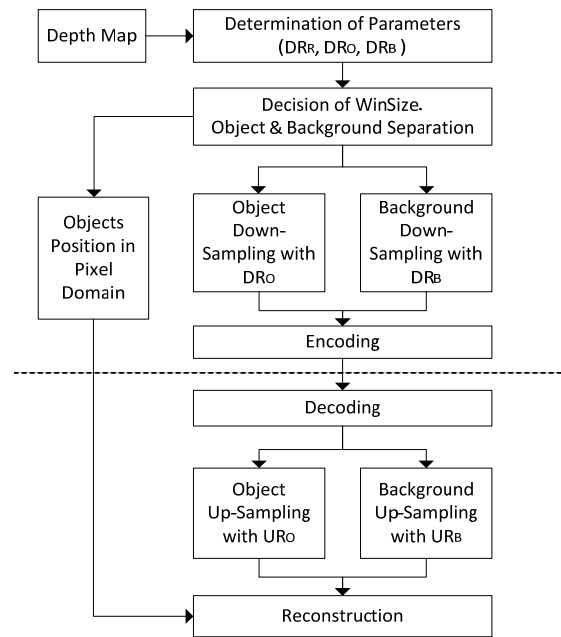


그림 2. 제안하는 하이브리드 다운샘플링 기법에 대한 전체 흐름도
Fig. 2. Flow chart of the proposed adaptive down-sampling method

먼저 깊이 맵 상의 객체와 배경에 적용될 다운샘플링 비율과 기준점 다운샘플링 비율이 정해지면 이들 비율간의 관계식에 의해 객체에 적용되는 윈도우 사이즈(Win.Size)가 결정된다. 이후 객체와 배경의 분리 과정이 이루어지고, 각각 서로 다른 다운샘플링 비율로 다운샘플링이 진행된다. 수신 단에서는 디코딩 이후 위 과정들에 대한 역과정이 진행된다. 객체에 적용되는 윈도우의 위치는 매 프레임 별로 움직임을 추적하여 바뀌게 되는데, 해당 객체의 위치 정보에 대한 전송은 오류 없는 채널 상태를 가정한다. 이는 수신 단에서의 올바른 객체와 배경과의 결합을 위함이다.

2.2.1. 다운샘플링 비율(DR_O, DR_B, DR_R) 및 객체 윈도우의 크기(WinSize)

본 논문에서 제안하는 하이브리드 방식과 기존 방식과의 공정한 비교를 위해서 전송되어지는 데이터의 양은 같다는 전제 조건이 필요하다. 이를 위해 수식 (3)과 같이 이미지 상의 픽셀의 개수를 같게 함으로서 위의 조건을 만족시킬 수 있다.

$$\underbrace{\frac{W \times H}{(DR_R)^2}}_{reference} = \underbrace{\frac{W_O \times H_O}{(DR_O)^2}}_{object} + \underbrace{\frac{W \times H - W_O \times H_O}{(DR_B)^2}}_{background} \quad (3)$$

객체와 배경에 적용될 다운샘플링 비율(DR_O , DR_B)과 그리고 기준점 다운샘플링 비율(DR_R)이 결정되면, 다운 샘플링 되기 전 원래의 이미지 사이즈($W \times H$)와 객체 부분에 적용될 윈도우 사이즈($W_O \times H_O$)간의 비율이 결정된다. 배경에 적용되는 다운샘플링 비율(DR_B)은 객체 부분을 제외한 나머지 부분($W \times H - W_O \times H_O$)에 적용된다. 위의 관계식은 다음과 같이 간략화 할 수 있다.

$$\frac{(W_O \times H_O)}{(W \times H)} = \frac{(DR_O)^2 \cdot (DR_B^2 - DR_R^2)}{(DR_R)^2 \cdot (DR_B^2 - DR_O^2)} \quad (4)$$

해당 과정은 즉, 결정된 다운샘플링 비율들을 기반으로 하여 수식 (4)와 같이 객체 부분에 적용시킬 윈도우의 크기를 결정하는 단계이다. 여기서 주목할 점은, 객체 윈도우의 너비(W_O)와 높이(H_O)의 곱의 값이 정해진 것이지 너비와 높이가 정해진 것은 아니라는 점이다. 이는 다음 과정인 객체/배경 분리 과정에서 객체의 움직임을 고려하여 결정된다.

2.2.2. 객체/배경 분리 방법

수식 (4)로부터 $(W_O \times H_O) / (W \times H)$ 의 비율이 정해졌다 할지라도, 여기에는 여러 다른 경우가 존재한다. 예를 들어 $(W_O, H_O) = (200, 400), (400, 200), (100, 800), (800, 100)$ 은 같은 $(W_O \times H_O)$ 값을 갖는 서로 다른 여러 윈도우를 나타낸다.

이를 위해 본 과정에서는 객체의 움직임을 최대한 포함할 수 있도록 객체 윈도우의 너비와 높이가 결정된다. 객체/배경 분리 과정에 대한 흐름도는 위의 그림 3과 같다.

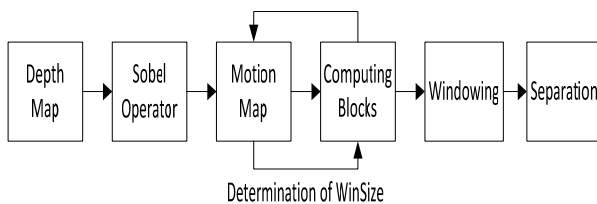


그림 3. 객체/배경 분리 과정에 대한 흐름도
Fig. 3. Flow chart of the object/background separation process

객체 윈도우의 너비와 높이인 W_O, H_O 는 다음과 같은 과정으로 결정된다. 입력된 깊이 맵에 대한 소벨 연산자(Sobel operator)를 통하여 그림 4.(a)와 같이 경계선 맵(edge map)이 생성이 되고, 생성된 경계선 맵에 대하여 이전 프레임과 현재 프레임에

서의 블록 단위 차 영상(MAD : Mean Absolute Difference)을 구하여 임의의 임계값보다 클 경우 해당 블록(8x8)의 밝기 값을 255로 설정하고 작을 경우 0값으로 설정한다. 즉, 경계선 맵에 대한 모션 맵(motion map)은 그림 4.(b)와 같이 움직임이 있는 흰색(밝기 값 255)의 블록과 움직임이 없는 검은색(밝기 값 0)의 블록으로 구성된다. 여기서 소벨 연산자를 사용하는 이유는 모션 정보(motion information)뿐만 아니라 객체의 경계선 정보(edge information)도 객체/배경 분리 과정에 포함시키기 위함이다. W_O, H_O 에 대한 결정 과정은 다음 수식 (5)와 같다.

$$\underset{(w \times h) \in (W_O \times H_O)}{\operatorname{argmax}} \operatorname{Count}(w \times h) \quad (5)$$

여기서 $(w \times h)$ 는 위의 예처럼 같은 $(W_O \times H_O)$ 값을 갖는 윈도우 집합 중의 하나의 후보 윈도우이다. 또한 $\operatorname{Count}(w \times h)$ 는 모션 맵 상의 $(w \times h)$ 의 윈도우 내의 255의 밝기 값을 가지는 블록의 개수를 카운팅(counting)하는 함수이다. 즉, 수식 (5)를 통해서 같은 $(W_O \times H_O)$ 값을 갖는 여러 후보 윈도우 중에서 모션 맵 상의 움직임이 있는 블록들을 가장 많이 포함 시킬 수 있는 윈도우를 탐색하게 된다. $\operatorname{Count}(w \times h)$ 의 값이 최대값이 될 경우, 매 프레임마다 가장 움직임이 활발한 객체의 위치가 결정된다. 객체 윈도우의 너비 및 높이는 배경 부분에 적용되는 다운 샘플링 비율(DR_B)에 대해 배수 관계가 성립하여야 하는데, 이는 디코딩 후의 올바른 복원을 위함이다. 따라서 $(w \times h)$ 후보군은 그리 많은 경우가 발생하지 않는다. $\operatorname{WinSize}$ 가 최종적으로 결정이 되면, 객체와 배경은 그림 4.(c)와 같이 객체 부분에 윈도우를 적용하는 과정을 거쳐 분리가 된다.

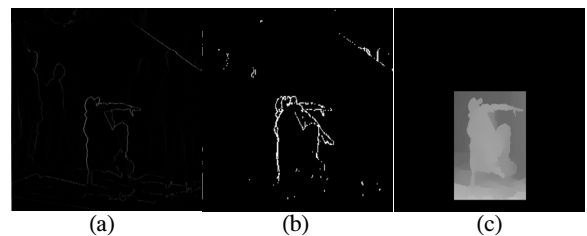


그림 4. 객체와 배경 분리 과정의 예시 (a) 경계선 맵 (b) 모션 맵 (c) 윈도우링
Fig. 4. An example of the objects/background separation process (a) edge map (b) motion map (c) windowing

2.2.3. 파라미터 전송 방법

객체/배경 분리 과정은 그림 5과 같이 매 프레임 별로 객체의 움직임을 추적하여 객체에 적용되는 윈도우의 위치는 바뀌게 된다. 수신 단에서는 올바른 업샘플링을 위하여 이와 같은 프레임 별 객체의 위치 정보와 객체와 배경에 적용되는 샘플링 비율에 대한 정보가 필요하다. 객체의 위치 정보의 경우 객체에 적용되는 윈도우의 한 포인트에 대한 픽셀 도메인 상의 위치 정보(=최대 3바이트)와 해당 윈도우의 너비와 높이 사이즈(=최대 3바이트)로 프레임 마다 객체의 위치 정보를 추정할 수 있고, 샘플링 비율 정보는 최대 1바이트면 충분히 표현 가능하다. 예를 들어 (1024×768) 사이즈의 해상도인 경우, 해상도의 높이와 너비는 각각 10비트($2^{10} = 1024$)면 충분히 표현할 수 있다. 즉, N 개의 프레임에 대한 파라미터의 데이터 크기는 최대 $(3 \times N + 3 + 1)$ 바이트 정도가 필요한데, 이는 표 1과 같이 본 실험에서 얻을 수 있는 데이터 이득에 비하면 무시할만하다. 이러한 파라미터 데이터들은 오류 없이 전송됨을 가정하고, 예를 들어, 해당 데이터는 오류 시 재전송이 가능한 TCP 기반 프로토콜을 사용하여 동영상 데이터 전송 시작 전에 초기 연결 시 주고 받는 제어패킷 내에 콘텐츠 서술자(Content Descriptors)로 포함되어 전송 될 수 있다. 디코딩 이후, 객체와 배경은 각각의 업샘플링 비율(UR_O , UR_B)대로 업샘플링 되고, 오류 없이 전송된 객체 윈도우의 위치 정보를 참조하여 결합되어 원래의 사이즈대로 리스케일링(rescaling)된다.

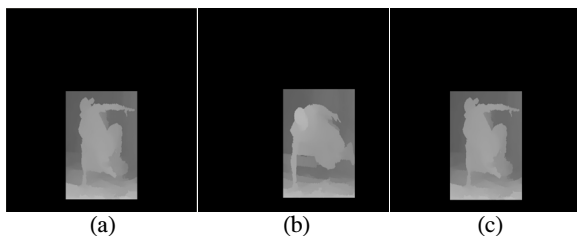


그림 5. 시간에 따른 객체 윈도우의 위치 변화 (a) 첫 번째 프레임 (b) 열 번째 프레임 (c) 스무 번째 프레임
Fig. 5. Tracking of the moving objects as time advanced (a) the first frame (b) the tenth frame (c) the twentieth frame

III. 실험 결과

그림 1과 같은 깊이 맵을 이용한 스테레오스코픽 시스템을 기반으로 하여 한 시점에 대한 텍스처 시퀀스(texture video)와 그에 따른 깊이 맵(depth

map) 시퀀스를 이용하여 렌더링(rendering) 된 영상상의 품질을 측정하였다^[13]. 제안하는 방식은 기존의 하나의 다운샘플링 비율로 균일하게 다운샘플링 하는 방식과의 비교를 나타내었다. 참고로 현재 MPEG-3DV ATM 코덱 내에서는 깊이 맵은 1/4 해상도로 기본 셋팅이 되어 있다. 즉, $DR_R (= UR_R) = 2$ 이다. 하지만 본 실험에서는 제안하는 하이브리드 방식의 효율성을 검증하고자 표 1과 같이 객체와 배경에 적용되는 다운/업 샘플링 비율을 각각 2와 8로 설정하였고, 이를 다운/업샘플링 비율을 4로 설정한 기존의 방식과의 비교를 나타내었다. 또한 수신 단에서의 업샘플링 과정에 대해서, OpenCV에서 제공하는 대표적인 리스케일링(rescaling) 알고리즘(이웃 보간법(nearest-neighbor (NN) interpolation), 선형 보간법(bilinear (LINEAR) interpolation), 지역 관계를 이용한 리샘플링 기법(re-sampling using area relation (AREA)), bicubic 보간법(bi-cubic (CUBIC) interpolation))을 이용하여 제안하는 다운샘플링 방식과의 호환성을 검토하였다^[14]. 실험에 사용된 파라미터들은 표 1과 같다.

표 1. 실험에 사용된 파라미터
Table 1. Simulation Parameters

DR_O	DR_B	DR_R	W	H	W_O	H_O
2	8	4	1024	768	384	408
No. of pixels per frame in the conventional scheme						49152
No. of pixels per frame in the proposed scheme						49008

표 1과 같이 제안하는 방식은 DR_R 을 사용한 기존의 방식보다 좀 더 적은 데이터를 사용했음을 알 수 있다(대략 10.5 KB/ 50프레임). 대표적인 3D 비디오 시퀀스인 1024x768 사이즈의 'Ballet', 'Breakdancer'를 사용하였고, 압축 과정은 전송되는 픽셀의 개수를 갖게 함으로서 생략하였다. 객체의 개수가 1개인 경우와 그 이상일 경우에 대해서 각각 나누어 실험하였다. 즉, 'BALLET' 시퀀스의 경우 두 개의 객체(좌측의 여자와 우측의 남자)에 윈도우를 나눠서 실험하였고($Win.Size(2 \times 192 \times 408)$), 'BREAK DANCER' 시퀀스는 화면 가운데 객체에 대해 하나의 윈도우($Win.Size(384 \times 408)$)를 적용시켰다.

표 2. PSNR을 이용한 객관적 화질 평가 비교
Table 2. Objective visual quality comparison using PSNR metric

Sequence	Up-sampling	Ref	Prop	$\Delta PSNR$
BALLET	NN	38.06	38.80	+0.74
	LINEAR	33.58	34.89	+1.31
	AREA	36.51	37.86	+1.35
	CUBIC	32.51	33.31	+0.8
BREAK DANCER	NN	40.96	41.17	+0.21
	LINEAR	38.39	39.12	+0.73
	AREA	40.19	40.87	+0.68
	CUBIC	38.10	38.51	+0.41

표 2에서 기존 방식 (Ref)은 하나의 다운샘플링 비율($DR=4$)을 사용한 경우이고, 제안하는 방식 (Prop)은 두 개의 다운샘플링 비율($DR=2$ and 8)을 혼용한 방식임을 주의한다. 표 2는 좀 더 적은 데이터를 사용했음에도 불구하고, PSNR(Peak-Signal to Noise Ratio) 측면에서의 개선 효과가 있음을 보여준다. 또한 그림 6과 8은 각각 'Ballet'/'Breakdancer' 시퀀스에 대해서 윈도우가 적용된 움직임이 있는 객체 부분에서 제안하는 방식이 기존의 방식보다 명확하고 깨끗한 시각적 효과가 있음을 보여준다. 마찬가지로 그림 7과 9는 'Ballet'/'Breakdancer' 시퀀스에 대해서 기존 방식과 제안하는 방식과의 전체 영상(full resolution)에 대한 시각적 품질 비교를 나타낸다. 좀 더 낮은 다운샘플링 비율이 적용된 객체 부분에서의 개선 효과가 명확하고, 좀 더 높은 다운샘플링 비율이 적용된 배경 부분에서의 저하 효과 역시 사람의 인지적인 특성(HVS)에 크게 눈에 띄지(noticeable) 않는 것을 확인할 수 있다.

본 논문에서 제안하는 하이브리드 깊이 맵 다운샘플링 알고리즘은 고정된 카메라(Fixed Camera)로 촬영된 영상에 대해서만 적용이 가능하다는 점을 언급하고자 한다. 즉, 움직이는 카메라(Moving Camera)로 촬영된 영상(MPEG 테스트 시퀀스 중 'Kendo', 'Balloons' 등)의 경우, 현재 프레임과 이전 프레임과의 차 영상을 이용한 본 알고리즘의 객체/배경의 구분 과정에서 배경 부분도 움직임이 있는 블록으로 간주될 수 있기 때문에 본 알고리즘과 적용 시 객체/배경의 적절한 구분이 어려울 수 있다. 또한, 객체의 움직임 정도와 윈도우 내의 움직이는 객체 포함비율이 본 알고리즘의 성능과 비례적인

관계가 있음을 알 수 있다. 즉, 본 실험에서와 같이 객체의 움직임 정도가 크고 윈도우 내의 움직이는 객체포함 비율이 높은 시퀀스의 경우('Ballet') 성능이 높게 나타나지만, 그 반대 특성의 시퀀스의 경우('Lovebird1')는 이득 정도(PSNR 기준. $-0.2dB$ 에서 $+0.2dB$)가 낮고 시각적 효과 차이는 크게 차이 나지 않는 것을 확인하였다.

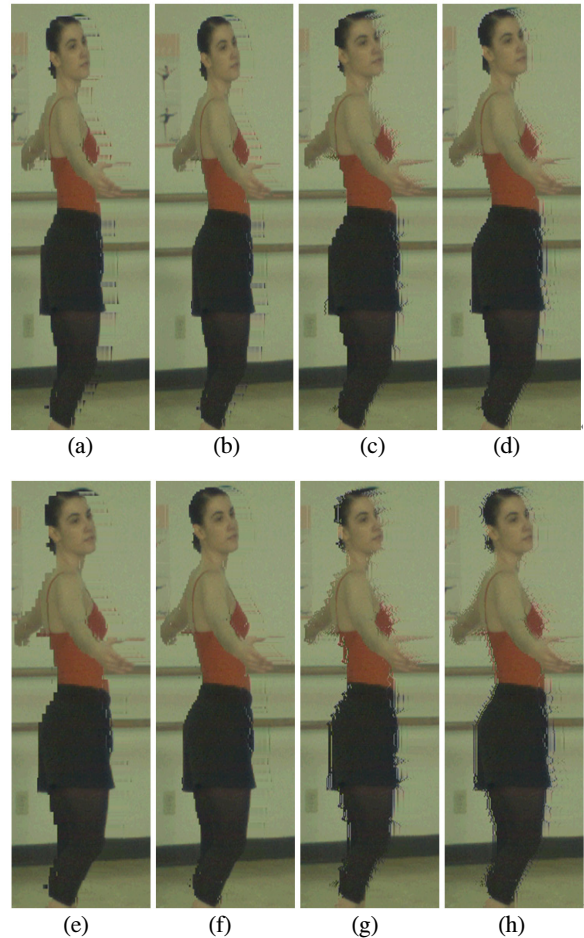
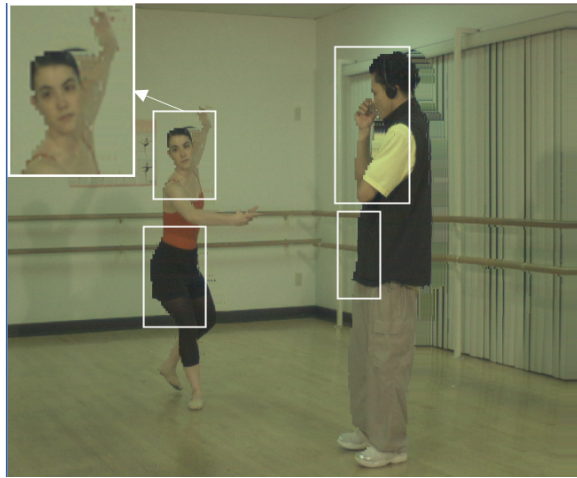
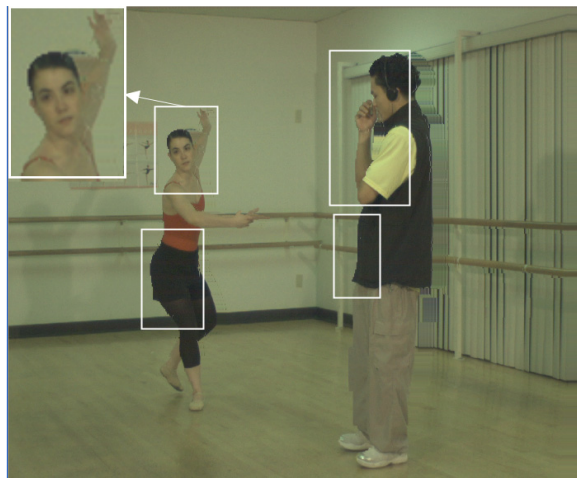


그림 6. Ballet 영상을 이용한 시각적 품질 비교;
(a), (c), (e), (g): 기존 방식을 이용하여 렌더링한 우영상. 각각 NN, LINEAR, AREA, CUBIC 업샘플링 기법 사용.
(b), (d), (f), (h): 제안하는 방식을 이용하여 렌더링한 우영상. 각각 NN, LINEAR, AREA, CUBIC 업샘플링 기법 사용.

Fig. 6. Visual quality comparison using Ballet;
(a), (c), (e), (g): conventional scheme with NN, LINEAR, AREA, or CUBIC
(b), (d), (f), (h): proposed scheme with NN, LINEAR, AREA, or CUBIC



(a)



(b)

그림 7. Ballet 영상을 이용한 렌더링 효과 비교(전체 영상) (a) 기존 방식으로 렌더링한 전체 영상 (b) 제안하는 방식으로 렌더링한 전체 영상

Fig. 7. Visual quality comparison using Ballet(full resolution) (a) with the conventional scheme (b) with the proposed scheme



(a)

(b)

(c)

(d)



(e)

(f)

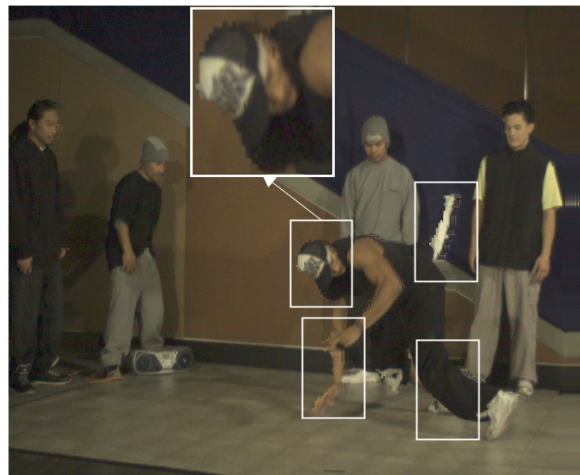
(g)

(h)

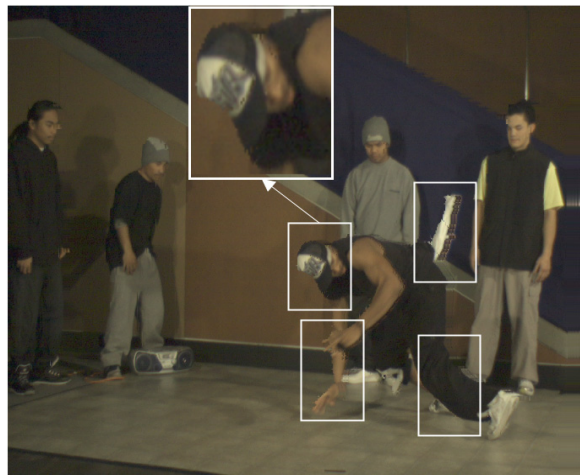
그림 8. Breakdancer 영상을 이용한 시각적 품질 비교; (a), (c), (e), (g): 기존 방식을 이용하여 렌더링한 우영상. 각각 NN, LINEAR, AREA, CUBIC 업샘플링 기법 사용. (b), (d), (f), (h): 제안하는 방식을 이용하여 렌더링한 우영상. 각각 NN, LINEAR, AREA, CUBIC 업샘플링 기법 사용.

Fig. 8. Visual quality comparison using Breakdancer; (a), (c), (e), (g): conventional scheme with NN, LINEAR, AREA, or CUBIC

(b), (d), (f), (h): proposed scheme with NN, LINEAR, AREA, or CUBIC



(a)



(b)

그림 9. Breakdancer 영상을 이용한 렌더링 효과 비교(전체 영상) (a) 기존 방식으로 렌더링한 전체 영상 (b) 제안하는 방식으로 렌더링한 전체 영상

Fig. 9. Visual quality comparison using Breakdancer (full resolution) (a) with the conventional scheme (b) with the proposed scheme

IV. 결론 및 추후 연구

본 논문에서는 객체와 배경에 서로 다른 다운샘플링 비율을 적용시키는 새로운 하이브리드 깊이 맵 다운샘플링 방법을 제안하였다. 움직임이 있는 객체 부분을 강조함으로써, 정적인 배경보다 동적인 객체에 보다 민감하게 반응하는 사람의 인지적 특성을 반영시켜 시각적 품질향상을 제고하고자 하였다. 실험적인 결과는 PSNR 측면에서 시퀀스 별로 평균적으로 0.5dB, 1dB 정도의 이득을 얻을 수 있었고, 시각적 품질 측면에서도 보다 자연스러운 렌더링 영상을 얻을 수 있었다. 또한 제안하는 방식은 여러 업샘플링 기법들과도 호환성을 유지한다.

추후 연구로는 움직이는 카메라로 촬영된 영상에도 적용 가능한 객체/배경 분리 알고리즘에 대한 연구와 본 알고리즘과의 호환성 여부에 대한 검증이 필요하다.

참 고 문 헌

- [1] K. M. Eom, "Three-dimensional image processing techniques for multi-view video," *Information and Communications Magazine, KICS*, vol. 28. pp. 21-29. 2011. 6
- [2] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV", in *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI. San Jose, CA, USA, 2004*, vol. 5291, pp. 93-104
- [3] R. Cheng, K. Nahrstedt, "Empirical study of 3d video source coding for autostereoscopic displays", in *MM*, ACM, 2007, pp 573-576
- [4] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Müller, M. Lang, "Three-Dimensional Video Postproduction and Processing", *Proceeding of the IEEE*, vol. 99, no. 4, April 2011
- [5] K. Muller, P. Merkle, T. Wiegand, "3-D Video Representation Using Depth Maps", *Proc. of the IEEE*, Apr. 2011, vol. 99, issue. 4
- [6] K. Muller, P. Merkle, T. Wiegand, "3-D Video Representation Using Depth Maps", *Proc. of the IEEE*, April 2011, vol. 99, issue. 4
- [7] Kwan-Jung Oh, Sehoon Yea, A. Vetro, Yo-Sung Ho "Depth Reconstruction Filter and Down/Up Sampling for Depth Coding in 3-D Video", *Signal Processing Letters, IEEE*, Sept. 2011, vol. 16, issue. 9, pp. 747-750
- [8] E. Ekmekcioglu, M. Mrak, S.T. Worrall, A.M. Kondo, "Edge adaptive upsampling of depth map videos for enhanced free-viewpoint video quality", *Electronics Letters*, 26th March 2009, vol. 45, issue.7
- [9] L. Yanjie, S. Lifeng, "A novel upsampling scheme for depth map compression in 3DTV system", *28th Picture Coding Symposium (PCS)*, Nagoya, Japan, Dec. 2010,
- [10] M.O. Wildeboer, T. Yendo, M. Panahpour Tehrani, T. Fujii, M. Tanimoto, "Depth up-sampling for depth coding using view information", *3DTV Conference : The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2011
- [11] Z. Qiuwen, A. Ping, Z. Yan, Z. Qian, Z. Zhaoyang, "Reduced resolution depth compression for multiview video plus depth coding", *Signal Processing (ICSP), 2010 IEEE 10th International Conference*, pp.1145-1148
- [12] H.A. Karim, N.S.M.A. Shah, N.M. Arif, A. Sali, S. Worrall, "Reduced resolution depth coding for stereoscopic 3D video", *Consumer Electronics, IEEE Transaction*, 2010, vol. 56, issue. 3
- [13] ISO/IEC JTC1/SC29/WG11, Committee Draft of ISO/IEC 23002-3 Auxiliary Video Data Representation. WG11 Doc. N8038 Apr. 2006
- [14] OpenCV 2.x C++ Reference Manual Release 2.3, June 2011. <http://opencv.itseez.com/trunk/>

김 태 우 (Tae-Woo Kim)



2010년 2월 광운대학교 전파공학과 졸업
2011년 3월~현재 성균관대학교 IT융합학과 석사 과정
<관심분야> 3차원 영상처리, 깊이 맵

김 정 훈 (Jung Hun Kim)



2012년 2월 성균관대학교 전자전기공학과 졸업
2012년 3월~현재 성균관대학교 이동통신공학과 석사 과정
<관심분야> 3차원 영상처리, 화질 평가

박 명 우 (Myung Woo Park)



2012년 2월 광운대학교 정보제어공학과 졸업
2012년 3월~현재 성균관대학교 이동통신공학과 석사 과정
<관심분야> 3차원 영상 코덱

신 지 태 (Jitae Shin)



1986년 2월 서울대학교 전기공학과 학사
1988년 2월 KAIST 원자력공학과 석사
2001년 5월 Univ. of Southern California (USC), Dept. of Electrical Eng. 석사 및 박사

1988년~1991년 한국전력공사 고리원자력 발전소
1991년~1996년 한국원자력연구소 계측제어설계부
2001년 8월~2002년 2월 경희대학교 정보통신전문대학원 교수
2008년 8월~2009년 8월 Univ. of Toronto, Canada 방문교수
2002년 3월~현재 성균관대학교 정보통신대학 교수
<관심분야> 멀티미디어 통신/네트워크, 이동휴대 방송기술, 유무선 통신네트워크 시스템, 영상처리