

3차원 합성영상의 화질 개선을 위한 깊이 경계 선명화

송 윤 석*, 이 천*, 호 요 성°

Depth Boundary Sharpening for Improved 3D View Synthesis

Yunseok Song*, Cheon Lee*, Yo-Sung Ho°

요 약

본 논문에서는 3차원 합성영상의 화질 개선을 위하여 깊이영상의 경계를 개선하는 방법을 제안한다. 깊이영상을 부호화할 때 객체의 경계부근에서 왜곡이 발생하게 되는데, 이러한 왜곡은 합성영상의 화질을 떨어뜨리는 원인이 된다. 본 논문에서 제안하는 방법은 경계 주변만을 필터링하기 위해 에지 맵을 추출하고, 이를 이용하여 객체의 경계를 포함하는 매크로블록에만 제안하는 방법을 적용한다. 여기서 경계주변의 참조 화소의 빈도, 유사성, 거리 등 세 가지 요소를 고려하는 윈도우-기반의 경계 선명화 필터를 적용하여 중앙의 대상 화소를 최적의 화소값으로 대체한다. 제안한 방법은 다시점 비디오 부호화의 참조 소프트웨어인 JMVC 8.3에 디블록킹 필터의 후처리로 구현했다. 기존 방법들과 비교하여 평균 0.49 dB의 화질 개선과 16.58%의 비트율 감소가 있었고, 주관적으로도 개선된 부분을 확인하였다.

Key Words : 3D video coding, depth coding, view synthesis

ABSTRACT

This paper presents a depth boundary sharpening method for improved view synthesis in 3D video. In depth coding, distortion occurs around object boundaries, degrading the quality of synthesized images. In order to encounter this problem, the proposed method estimates an edge map for each frame to filter only the boundary regions. In particular, a window-based filter is employed to choose the most reliable pixel as the replacement considering three factors: frequency, similarity and closeness. The proposed method was implemented as post-processing of the deblocking filter in JMVC 8.3. Compared to the conventional methods, the proposed method generated 0.49 dB PSNR increase and 16.58% bitrate decrease on average. The improved portions were subjectively confirmed as well.

I. 서 론

근대 방송 기술사에서 주목할 만한 혁신을 두가지 꼽는다면 흑백TV에서 컬러TV로 변환되는 기술과 2000년대 초반에 디지털 영상 기술을 통해 고해상도의 영상을 보이는 기술이다. 앞으로는 단순히 사용자가 영상을 눈으로 감상하기보다는 입체감을

주는 실감방송이 방송 기술을 이끌 것으로 기대된다.

3차원 영상은 기존 2차원 영상과는 다르게 색상 정보만 이용하는 것이 아니라, 깊이 정보도 함께 사용하여 가까운 물체가 튀어나오게 보이는 입체감을 제공한다. 하지만 3차원 영상에서의 깊이 정보는 시점수의 증가로 인한 입력 데이터양 증가로 이어지

• 주저자 : 광주과학기술원 정보통신공학과, ysong@gist.ac.kr, 정회원

° 교신저자 : 광주과학기술원 정보통신공학과, hoyo@gist.ac.kr, 종신회원

* 광주과학기술원 정보통신공학과, leecheon@gist.ac.kr, 정회원

논문번호 : KICS2012-03-137, 접수일자 : 2012년 3월 26일, 최종논문접수일자 : 2012년 9월 24일

기도 한다.

3차원 영상을 처리, 전송 등을 수행하려면 효과적인 부호화 기술이 필수적이다. 효과적인 데이터양의 감축을 위해 존재하지 않는 시점은 임의의 시점 합성 기술을 이용하여 다시점 영상을 획득한다^{1,2}.

고화질의 합성영상을 획득하기 위해서는 정확한 깊이 정보가 요구된다. 깊이영상을 부호화할 때, 블록 기반 부호화의 특성상 경계에서 블록 왜곡(blocking artifact)이 발생한다. 이러한 블록 왜곡들은 영상 합성한 결과에서 왜곡된 영상을 생성한다.

본 논문에서는 경계 선명화를 이용한 합성영상의 화질을 개선하는 방법을 제안한다. 먼저 Canny 에지 검출 방법을 이용해 영상 내에서 필터링할 부분을 찾는다. 매크로블록(MB) 기반 부호화 특성을 이용하여 매크로블록 단위로 에지가 주변에 존재하는지 탐색하고, 매크로블록 내부에서 에지가 존재하는 영역을 결정하여 필터링이 필요한 세부 블록을 탐색한다. 사용하는 필터는 주변 화소의 발생빈도, 유사성, 거리 등 세 가지 요소를 바탕으로 하여 경계가 선명화될 수 있도록 깊이값을 대체한다.

2장에서는 깊이영상의 경계 부분 왜곡이 합성영상에 미치는 영향에 대한 내용을 살펴본다. 3장에서는 이를 해결하기 위한 제안 방법을 세 가지 단계로 나누어 구체적으로 설명한다. 4장에서 참조 소프트웨어와의 비교를 포함한 실험 결과를 보이고, 5장에서 결론을 맺어 본 논문을 마무리한다.

II. 왜곡과 합성영상

일반적으로 가상 시점의 영상을 만들려면 인접하는 시점들의 색상영상과 그에 해당하는 깊이영상을 필요로 한다. 깊이영상의 깊이정보를 이용하여 3차원 구조를 정의하고, 이를 토대로 시점 간의 화소들을 정합할 수 있다^{3,4}.

단일시점의 비디오는 H.264/AVC로 높은 부호화 성능을 얻었지만, 최근 3차원 영상에 대한 관심이 높아지면서 다시점 영상을 위한 부호화 방법이 개발되고 있다. Moving picture experts group(MPEG)에서는 H.264/AVC의 부호화 방법과 호환성을 지니고 다시점 비디오를 효과적으로 부호화할 수 있는 다시점 비디오 부호화(multi-view video coding, MVC) 방법을 개발하고 이를 바탕으로 Joint multi-view video coding(JMVC)라는 참조 소프트웨어를 개발했다. JMVC는 각 프레임이 16×16 크기의 매크로블록 단위로 부호화되며 사용하는 디블

록킹 필터는 H.264/AVC의 것과 같다.

깊이영상에서 왜곡이 많이 발생할수록 합성영상의 화질을 떨어뜨린다. 그림 1은 ‘Cafe’ 깊이영상을 JMVC를 이용하여 두 가지의 QP를 이용하여 부호화한 깊이영상의 결과이다. 디블록킹 필터를 적용시켰음에도 불구하고 배경과 전경의 경계 부근에서 왜곡이 많이 발생하는 것을 볼 수 있다. 이는 에지 부근에서의 블로킹 현상이 영향을 준 것이다. 그림 2는 앞에서 부호화한 깊이영상을 이용하여 임의의 시점의 영상을 합성한 결과를 보인 것이다. 합성된 영상들을 보면 깊이영상에서처럼 경계 부근에서 정확하지 않은 값들이 존재하는 것을 확인할 수 있으며 이 현상은 결과적으로 합성영상의 화질에 영향을 준다. 여기서 합성에 사용된 소프트웨어는 View synthesis reference software(VSRS)이며 표준화 그룹에 제안된 기술들을 바탕으로 만들어졌다.

III. 깊이영상의 경계 선명 방법

제안하는 방법은 JMVC에서 디블록킹 필터에 후처리 필터를 추가적으로 수행하는 것이다. 먼저 이진화된 에지 맵을 추출하고 이 정보를 바탕으로 부호화할 매크로블록이나 그것에 인접하는 매크로블록들이 에지를 포함하는지 여부를 고려하여 후처리 필터를 적용할 지 여부를 결정한다. 최종적으로 필터링을 수행할 화소와 주변의 참조 화소들의 빈도, 유사성, 거리 등 세 가지 요소를 고려하여 깊이영상의 경계를 선명화한다. 그림 3은 제안 방법의 전체적인 흐름도이다.



그림 1. ‘Cafe’ 깊이영상 부호화 결과
Fig. 1. ‘Cafe’ coded depth maps



그림 2. 그림 1의 깊이영상을 이용해 합성한 결과
Fig. 2. Synthesis results using Fig. 1 depth maps

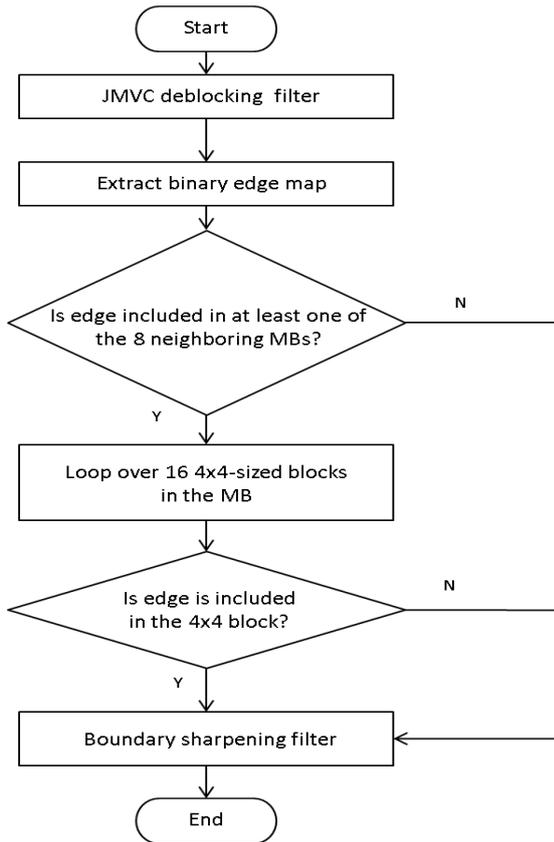


그림 3. 제안 방법의 전체 흐름도
Fig. 3. Flowchart of the proposed algorithm

3.1. 이진 에지 맵 추출

이진 에지 맵을 추출하는 것은 깊이영상의 에지 정보를 획득하기 위한 것이다. 이는 경계 부근의 영상 합성 왜곡을 줄이고 선명하게 교정해야 하므로, 필터링이 필요한 부분에만 제안하는 방법을 적용시키는 것이 디블록킹 필터처럼 영상 전체에 적용시키는 것보다 효율적이다.

제안 방법에서는 에지 맵을 추출해내기 위해 Canny 에지 검출 방법을 사용한다⁵⁾. 각 프레임마다 Canny 방법을 이용해 에지 맵을 추출해 다음 과정에서 사용하게 된다. 제안 방법에서는 실험할 시퀀스에 임계값을 실험적으로 넣고 제일 좋은 성능을 보이는 값들을 선택하였다.

3.2. 매크로블록 특성에 따른 필터링 선택

제안하는 깊이영상 화질 개선 방법은 깊이영상 전체에 필터링을 이용하는 것이 아니라 영상 내의 경계 부근만 선명하게 하는 것이다. 매크로블록 단위로 깊이영상의 특성에 따라 필터링을 할 지 생략할 지 등을 결정한다.

앞의 과정에서 획득한 이진 에지 맵을 이용하여

매크로블록 단위로 좌, 우, 위, 아래로 인접하는 여덟 가지 매크로블록에 에지 성분을 포함하고 있는지 체크한다. 여기서 인접하는 매크로블록에 중에서 모두 에지가 포함되지 않았다면 일반적인 영상의 특성상 중앙 매크로블록에 에지를 포함할 확률은 극히 적다고 할 수 있다. 그러므로 이 경우는 필터링을 하지 않고 다음 매크로블록으로 넘어간다. 그림 4는 이진 에지 맵에서 필터링을 수행할 매크로블록을 보인 것이다. 인접 매크로블록에 에지가 없는 매크로블록들은 짙은 회색으로 처리되어 필터링 없이 바로 생략한다. 짙은 회색 이외의 매크로블록들에서는 최소한 한 개의 4x4 블록이라도 에지가 포함되어 있는 경우 필터링을 거치는 매크로블록들은 밝은 회색, 에지가 포함되지 않아서 생략되는 블록들은 중간 회색으로 표시되었다.

3.3. 경계 선명화 필터

제안하는 방법에서 이용하는 경계 선명화 필터는 윈도우를 기반으로 주변 화소 정보를 이용하는 방법이다. 화소 정보를 이용할 때 세 가지의 요소를 고려하는데 이는 인접 참조 화소의 빈도, 유사성, 거리다⁶⁾.

빈도는 윈도우 내의 주변 화소들 중에서 중앙 화소와 같은 값을 갖는 화소의 수를 나타낸다. 이러한 빈도가 높을수록 신뢰성 있는 화소값이 되는 것이다. 먼저 k 는 주변 화소의 값이다. 예를 들어 5x5 사이즈의 윈도우 안에 25개 화소 중 24개가 주변 화소이고 서로 다른 화소 값이 총 6개가 존재한다고 하면 6이 변수 k 의 값이 된다. 이어서 n 은 윈도우의 크기이고, $W_{n \times n}$ 은 윈도우가 정사각형이란 전제하에 $n \times n$ 만큼의 화소가 있다 하는 것이다. k 에 따른 빈도 값(cost)은 식 (1)과 같이 $F(k)$ 로 정의한다.

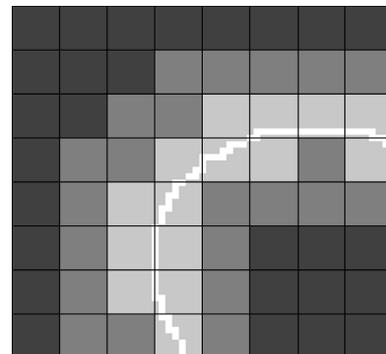


그림 4. MB에 에지 분포에 따른 영역 구분
Fig. 4. Region separation based on MB edge information

$$F(k) = \sum_{i=0}^{n \times n - 1} \delta[k - W_{n \times n}(i)] \quad (1)$$

유사성 정보는 중앙 화소값과 주변 화소값의 절대차를 뜻한다. 절대차가 작을수록 주변 화소 값이 중앙 화소 값에 가깝다는 것을 의미하므로 높은 신뢰도를 갖는다고 가정할 수 있다. k 에 따른 유사성 값은 식 (2)와 같이 $S(k)$ 로 정의한다. I_{cur} 는 중앙 화소값을 뜻한다.

$$S(k) = |I_{cur} - k| \quad (2)$$

거리 정보는 중앙 화소값과 주변 화소값의 Euclidean 거리를 뜻한다. 거리가 가까울수록 중앙 화소와 비슷하다고 할 수 있으므로, 신뢰도가 높아진다. x_{cur} 와 x_i 는 각각 중앙 화소의 수평 위치를 나타내고 y_{cur} 와 y_i 는 수직 위치를 뜻한다. 각 거리를 모두 더한 다음 평균 거리를 구하기 위해 빈도 값인 $F(k)$ 로 나누어준다. k 에 따른 거리 값은 식 (3)과 같이 $C(k)$ 로 정의한다.

$$C(k) = \frac{1}{F(k)} \sum_{i=0}^{F(k)-1} \sqrt{(x_{cur} - x_i)^2 + (y_{cur} - y_i)^2} \quad (3)$$

각 요소 값을 정규화(normalize)하기 위해 최대값과 최소값을 이용하여 값을 조정한다. 식 (4)-(6)은 정규화된 요소 비용 식들이다. 비용값을 가장 작게 만드는 화소값 k 가 가장 신뢰도가 높은 화소이다.

$$J_F(k) = \frac{F(k) - F(\min)}{F(\max) - F(\min)} \quad (4)$$

$$J_S(k) = \frac{S(\max) - S(k)}{S(\max) - S(\min)} \quad (5)$$

$$J_C(k) = \frac{C(\max) - C(k)}{C(\max) - C(\min)} \quad (6)$$

이렇게 구한 각 요소 별로 따로 실험한 결과 빈도-유사성-거리 순으로 기여도 차이가 조금씩 발생한다. 효율적으로 세 가지 요소 모두를 고려하기 위해 서로 다른 가중치를 주어 식 (7)과 같이 종합적인 비용값을 구한다. 종합 비용값을 가장 작게 만드는 화소값이 중앙 화소값을 대체하게 된다.

$$J(k) = 3 \times J_F(k) + 2 \times J_S(k) + J_C(k) \quad (7)$$

정리하면, 이 윈도우-기반의 경계 선명화 필터는 주변의 정보를 이용하여 최적의 깊이값을 찾아 중앙 화소값을 대체한다. 주변 화소값의 신뢰도를 평가하기 위해 빈도, 유사성, 거리 등 세 가지의 요소를 고려한다. 윈도우 내에서 화소값의 빈도가 높을수록, 중앙 화소값과 유사할수록, 주변 화소값을 가지는 화소들과 중앙 화소와의 거리 평균이 작을수록 그 화소값의 신뢰도를 높게 설정한다. 각 요소의 신뢰도값을 기여도에 따라 서로 다른 가중치를 둔다. 이는 각 요소 하나씩만 고려했을 때의 성능을 평가한 것을 바탕으로 했다. 이 필터는 [6]에서의 필터의 기본 구조를 갖지만 각 요소마다 서로 다른 가중치를 설정하여 더 효율적으로 하였다. 또한 필터링이 영상 전체에 사용하지 않고 에지를 이용하여 MB를 탐색하여 화소값 대체가 필요한 곳에만 필터링을 하게 된다.

IV. 실험 결과

본 논문에서는 제안 방법을 JMVC 8.3의 더블록킹 필터의 후처리로 구현해 성능을 평가한다. 실험에 이용한 시퀀스는 Microsoft Research에서 제공한 ‘Ballet’와 ‘Breakdancers’이며 모두 8-시점으로 구성되어 있다^[7]. 실험 환경 설정을 정리하여 표 1에 포함했다. 표 2와 표 3은 각각 ‘Ballet’와 ‘Breakdancers’의 부호화 결과를 나타냈으며 그림 5와 그림 6은 이들의 RD 커브이다. VSRS를 통해 합성하는 과정에서 원본 색상영상을 사용하고 깊이 영상만 부호화한 결과를 이용했다.

제안 방법은 특히 높은 비트율에서 효과적이었다. 이는 고화질의 깊이영상일수록 제안 방법을 적용할 때 신뢰도 있는 화소값을 정확히 찾을 확률이 높아지기 때문이다.

표 1. 실험 환경 설정
Table 1. Experimental conditions

| Sequence | Ballet, Breakdancers | |
|--------------------|---|----------|
| QP | 32, 37, 42, 47 | |
| Number of frames | 100 | |
| View to synthesize | Synthesize view 4 using view 3 and view 5 | |
| Canny thresholds | Ballet | 170, 180 |
| | Breakdancers | 100, 130 |

표 2. 'Ballet' 부호화 결과
Table 2. 'Ballet' coding results

| | JMVC | | Proposed method | |
|-------|-----------|-------------------|-----------------|-------------------|
| | PSNR (dB) | Bitrate (kbits/s) | PSNR (dB) | Bitrate (kbits/s) |
| QP 32 | 35.43 | 368.55 | 36.53 | 389.11 |
| QP 37 | 33.95 | 197.89 | 34.84 | 218.99 |
| QP 42 | 32.42 | 110.53 | 33.20 | 122.71 |
| QP 47 | 29.78 | 67.46 | 30.29 | 73.76 |

표 3. 'Breakdancers' 부호화 결과
Table 3. 'Breakdancers' coding results

| | JMVC | | Proposed method | |
|-------|-----------|-------------------|-----------------|-------------------|
| | PSNR (dB) | Bitrate (kbits/s) | PSNR (dB) | Bitrate (kbits/s) |
| QP 32 | 35.33 | 345.25 | 36.35 | 356.17 |
| QP 37 | 33.96 | 180.87 | 34.63 | 187.37 |
| QP 42 | 32.47 | 100.81 | 32.86 | 103.75 |
| QP 47 | 30.97 | 61.35 | 31.15 | 62.12 |

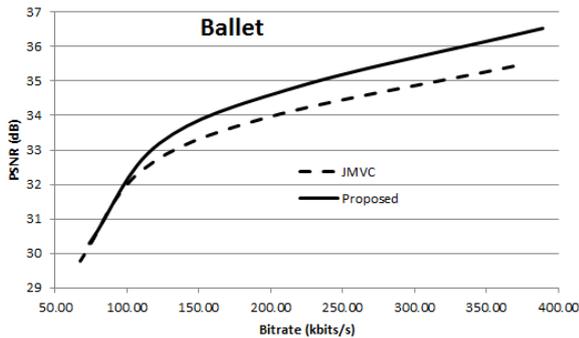


그림 5. 'Ballet' RD 커브
Fig. 5. 'Ballet' RD curve

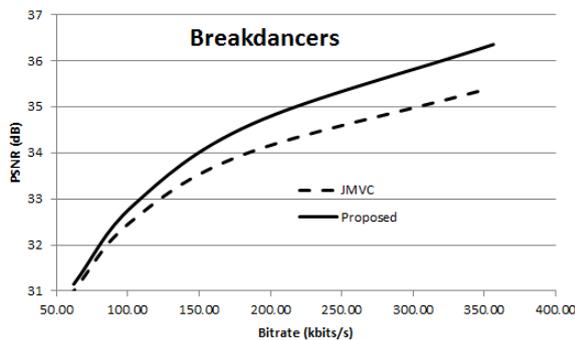


그림 6. 'Breakdancers' RD 커브
Fig. 6. 'Breakdancers' RD curve

JMVC에 비교했을 때 평균 0.49 dB의 화질 개선과 16.58%의 비트율 감소가 발생했다. 제안 방법은 에지 맵 추출과 화소별 필터링 등으로 인한 복잡도가 소폭 증가한다. JMVC에 비해 부호화 시간이 5.7% 증가했는데 이는 16.58%의 비트율 감소에 따른 트레이드오프가 된다.

그림 7와 'Ballet'의 결과 깊이영상, 그림 8은 그림 7의 깊이영상과 원본색상영상을 통해 생성한 합성영상을 나타낸다. 이 영상들은 QP 37을 사용했을 때의 결과 영상들이다. JMVC와 제안 방법의 결과 영상 화질을 비교하면 경계 부근에서 차이가 나타난다. 제안 방법을 사용했을 때는 JMVC 결과에서 발생하는 상당수 왜곡된 화소값들이 정확한 화소값들로 대체된 것을 확인할 수 있다.

결과 합성영상을 비교하면 깊이영상에서 선평해진 경계가 합성영상의 경계 개선으로 이어짐을 볼 수 있다. 경계선이 JMVC를 사용했을 때 울퉁불퉁했던 것이 제안 방법을 사용했을 때는 더 자연스럽게 나타난다. 결론적으로 객관적 화질 평과와 주관적 화질 평가 모두 제안 방법이 기존 방법을 사용했을 때보다 더 좋은 화질의 합성영상을 만들어내는 것을 확인했다.

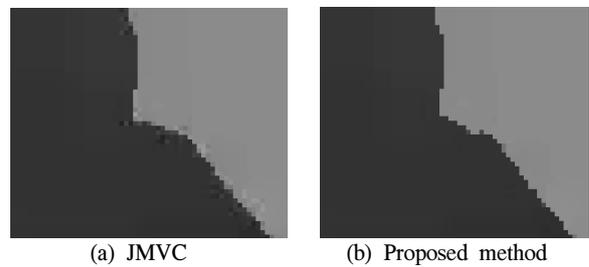


그림 7. 'Ballet' 결과 깊이영상
Fig. 7. 'Ballet' depth map results

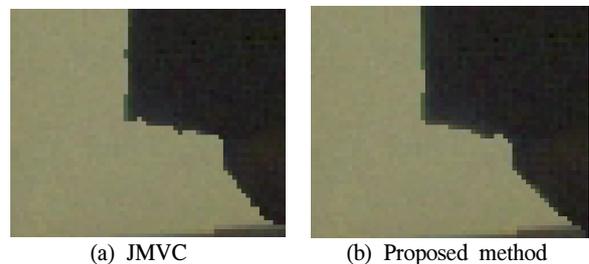


그림 8. 'Ballet' 결과 합성영상
Fig. 8. 'Ballet' synthesis results

VI. 결 론

본 논문에서는 3차원 비디오 부호화에서 합성영상의 화질을 개선하기 위한 깊이영상 경계 선명화 기술을 제안했다. 제안 방법에서는 Canny 에지 검출 방법을 통해 에지 맵을 얻은 뒤 매크로블록에 에지 포함 여부를 구분하여 필터링할지를 결정했다. 경계 선명화 필터는 빈도, 유사성, 거리 등 세 가지 요소를 고려해 중앙 화소값을 최적의 화소값으로 대체하였다. 빈도가 높을수록, 유사성이 높을수록, 거리가 짧을수록 화소가 대체 화소로써 높은 신뢰도를 갖는다. 제안한 방법은 JMVC 8.3에서 디블록킹 필터의 후처리로 구현해 성능을 평가했다. 평균 0.49 dB의 화질 개선과 16.58%의 비트율 감소 결과를 얻었으며, 주관적으로도 경계 부근에서의 화질이 개선되었음을 확인하였다.

참 고 문 헌

- [1] W.S. Jang, C. Lee, and Y.S. Ho, "Efficient depth map generation for various stereo camera arrangements," *J. Kor. Info. Commun. Soc. (J-KICS)*, vol. 37A, no. 6, pp. 458-463, May 2012.
- [2] S. Yea and A. Vetro, "View synthesis prediction for multiview video coding," *Signal Proc.: Image Commun.*, vol. 24, pp. 89-100, Jan. 2009.
- [3] C. Lee, B.H. Choi, and Y.S. Ho, "Efficient multiview depth video coding using depth synthesis prediction," *Optical Eng.*, vol. 50, no. 7, pp. 1-14, July 2011.
- [4] J.F. Canny, "A computation approach to edge detection," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 8, no. 6, pp. 769-798, Nov. 1986.
- [5] K.J. Oh., A. Vetro, and Y.S. Ho, "Depth coding using a boundary reconstruction filter for 3-D video systems," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 350-359, Mar. 2011.
- [6] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered

representation," *ACM SIGGRAPH and ACM Trans. on Graphics*, pp. 600-608, Aug. 2004.

송 윤 석 (Yunseok Song)



2008년 Illinois Institute of Technology Electrical Engineering 졸업 (학사)
2009년 Univ. of Southern California Electrical Engineering 졸업 (석사)
2010년-현재 광주과학기술원

정보통신공학과 박사과정

<관심분야> 디지털 영상처리, 실감방송, 3차원 비디오 부호화

이 천 (Cheon Lee)



2005년 한국항공대학교 항공 전자공학과 졸업 (학사)
2007년 광주과학기술원 정보통신공학과 졸업 (석사)

2007년-현재 광주과학기술원 정보통신공학과 박사과정
<관심분야> 3차원 TV, 실감방송, 3차원 비디오 부호화, MPEG 표준화

호 요 성 (Yo-Sung Ho)



1981년 서울대학교 공과대학 전자공학과 졸업 (학사)
1983년 서울대학교 대학원 전자공학과 졸업 (석사)
1989년 Univ. of California, Santa Barbara, Dept. of Electrical and Computer Engineering (박사)

1983년~1995년 한국전자통신연구소 선임연구원

1990년~1993년 미국 Philips 연구소, Senior research member

1995년~현재 광주과학기술원 정보통신공학과 교수
<관심분야> 디지털 신호처리, 영상 신호 처리 및 압축, MPEG 표준, 3차원 TV, 실감방송