

3차원 영상 생성을 위한 깊이맵 추정 및 중간시점 영상합성 방법

정회원 이상범*, 이 천**, 종신회원 호요성***

Depth Estimation and Intermediate View Synthesis for Three-dimensional Video Generation

Sang-Beom Lee*, Cheon Lee** *Regular Members*, Yo-Sung Ho*** *Lifelong Member*

요 약

본 논문에서는 3차원 영상 생성을 위한 깊이맵 추정 및 중간시점 영상합성 방법을 제안한다. 제안하는 방법은 깊이맵의 시간적 상관도를 향상시키기 위해 깊이값을 추정하는 과정에서 기존의 정합 함수에 이전 프레임에서 추정된 깊이값을 고려하는 가중치 함수를 추가한다. 또한, 중간시점 영상을 합성하는 과정에서 발생하는 경계 잡음을 제거하는 방법을 제안한다. 중간시점 영상을 합성할 때, 비폐색 영역을 합성한 후 경계 잡음이 발생할 수 있는 영역을 비폐색 (disocclusion) 영역을 따라 구별한 다음, 잡음이 없는 참조 영상을 이용하여 경계 잡음을 처리한다. 컴퓨터 모의실험 결과를 통해 깊이맵의 시간적 상관도를 향상시켜서 사용자의 시각적 피로감을 줄일 수 있었고, 배경 잡음이 사라진 자연스러운 중간시점 영상을 생성할 수 있었다.

Key Words : Multi-View Video, Depth Map, View Synthesis, Temporal Consistency, Boundary Noise

ABSTRACT

In this paper, we propose new depth estimation and intermediate view synthesis algorithms for three-dimensional video generation. In order to improve temporal consistency of the depth map sequence, we add a temporal weighting function to the conventional matching function when we compute the matching cost for estimating the depth information. In addition, we propose a boundary noise removal method in the view synthesis operation. After finding boundary noise areas using the depth map, we replace them with corresponding texture information from the other reference image. Experimental results showed that the proposed algorithm improved temporal consistency of the depth sequence and reduced flickering artifacts in the virtual view. It also improved visual quality of the synthesized virtual views by removing the boundary noise.

I. 서 론

다양한 영상처리 및 방송통신 기술의 발전으로 인하여, 우리는 컴퓨터를 이용하여 현실 세계를 재창조하고 이를 경험할 수 있게 되었다. 이 가운데 다시점 영상을 이용한 3차원 TV(Three-dimensional

TV, 3DTV)는 현실 세계를 재구성한 콘텐츠와 3차원 입체 모니터를 통해 실감나는 영상을 사용자에게 제공하기 때문에 차세대 방송기술 시스템으로 주목받고 있다^[1]. 여기서 다시점 영상이란 3차원 장면을 동일 시간대에서 여러 시점에 위치한 여러 대의 카메라로 획득한 영상을 말한다.

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 대학IT연구센터 지원사업으로 수행되었음. (NIPA-2009-(C1090-0902-0017))

* 광주과학기술원 정보통신공학과 ({sblee, leecheon, hoyo}@gist.ac.kr)

논문번호 : KICS2009-07-314, 접수일자 : 2009년 7월 31일, 최종논문접수일자 : 2009년 9월 23일

다시점 영상을 획득하기 위해 스테레오 (stereoscopic) 카메라, 다시점 (multi-view) 카메라, 깊이 카메라 (depth camera) 등과 같은 장치들이 개발되었으며, 이를 이용하여 얻은 영상을 처리하고 부호화하는 기술들에 대한 연구가 활발히 진행 중이다. 우리는 다시점 영상을 이용하여 사용자가 원하는 시점에서의 영상을 선택적으로 재생하거나, 인접한 시점의 두 영상 혹은 다시점 영상 전체를 이용하여 사용자에게 현실감 있는 3차원 장면을 재현할 수 있다^[2].

한 장면을 여러 시점에서 시청할 수 있는 자유시점 기능을 구현하기 위해서는 다시점 영상을 이용하여 시점 사이에 존재하는 가상의 중간시점 영상을 합성하는 기술이 필요하다. 중간시점 영상을 합성하기 위해서는 깊이맵을 사용하는데, 깊이맵이란 영상 내에 존재하는 객체들의 3차원 거리 정보를 나타내는 영상으로, 각 화소값은 해당 화소의 깊이 정보를 나타낸다. 깊이맵의 정확도는 합성된 중간시점 영상의 화질을 좌우하기 때문에 정확한 깊이맵을 생성하는 것이 매우 중요하다.

최근, 국제 표준화 그룹인 Moving Picture Experts Group(MPEG)에서는 다시점 영상의 중요성을 인지하고, 다시점 영상 부호화 및 다양한 기술들에 대한 표준화를 진행하고 있다. 2008년 1월 MPEG 표준화 회의에서는 3차원 비디오 부호화 표준화를 위한 테스트 영상을 요청하였고 (call for test material, CFT)^[3], 이에 대한 응답으로 2008년 4월 미팅에서 4개 연구기관에서 총 10개의 테스트 영상이 기고되었다. 그리고 MPEG 3차원 비디오 부호화 그룹에서는 깊이 영상 탐색 및 중간시점 합성을 위한 소프트웨어를 구현 및 배포하여 깊이 영상과 시점합성의 정확도 테스트를 시작했다^[4].

표준화 회의가 거듭될수록 소프트웨어의 성능은 향상되어 왔으나, 소프트웨어의 성능을 저하시키는 요인들, 예를 들면, 폐색 영역(occlusion)의 처리, 깊이영상의 시간적 상관도 저하, 중간시점 합성영상의 화질 저하 등의 문제점들이 여전히 존재한다. 본 논문에서는 3차원 영상 생성을 위한 깊이맵 추정 및 중간시점 영상합성 방법을 제안한다. 제안하는 방법은 깊이맵을 추정할 때 이전 프레임에서 추정한 깊이값을 고려하도록 하여 깊이맵 추정 과정에서 발생하는 시간적 상관도 저하 문제를 해결하는 방법을 제안한다. 또한, 중간시점 영상합성 과정에서 색상 영상과 깊이맵의 객체 모양이 정확하게 일치하지 않아 발생하는 문제가 영상합성 과정에서도 그대로 반영되는데, 본 논문에서는 이러한 잡음을 경계 잡음(boundary noise)이라

정의하고 이를 해결하기 위한 방안을 제시한다.

본 논문에서는 3차원 영상을 생성하기 위한 기술을 전체적으로 살펴보고자 한다. 2장에서는 3차원 비디오 시스템에 대한 일반적인 내용을 소개하고, 최근에 주목받고 있는 기술을 설명한다. 3장에서는 제안하는 3차원 영상생성 방법에 대해 설명하고, 4장에서는 다양한 실험을 통해서 제안하는 방법의 성능을 비교하고, 5장에서 결론을 맺는다.

II. 3차원 비디오 시스템

그림 1은 3차원 비디오 시스템의 개념도를 나타낸 것이다. 3차원 비디오 시스템은 각 시점의 깊이맵을 포함하는 3차원 비디오의 획득/처리/전송/재생에 대한 모든 과정을 포함하고 있고, 기본적으로 N 시점의 영상과 함께 N시점의 깊이맵을 동시에 획득/처리/전송/재생하도록 설계되었다. 그리고 3차원 비디오는 스테레오 카메라, 깊이 카메라, 다시점 카메라 등 다양한 저작 도구들을 이용하여 제작할 수 있는데, 깊이맵을 직접적으로 획득할 수 있는 깊이 카메라를 제외한 나머지 저작 도구들을 이용할 경우, 장면의 3차원 정보를 계산적으로 추정하는 방법을 이용하여 깊이맵을 획득한다.

송신단에서 압축하여 수신단으로 전송된 색상 영상과 깊이맵은 디스플레이 장치에 따라 여러 가지 방법으로 재생될 수 있다. 기본적으로 기존에 이용된 2차원 디스플레이 장치에서 재생할 수 있으며, 전송된 N 시점 중에서 원하는 시점의 2차원 영상을 자유롭게 선택하여 시청할 수 있다. 또한, 다양한 종류의 3차원 디스플레이 장치를 이용하여 재생할 수 있는데, 장치에 따른 여러 시점의 3차원 비디오를 재생하기 위해

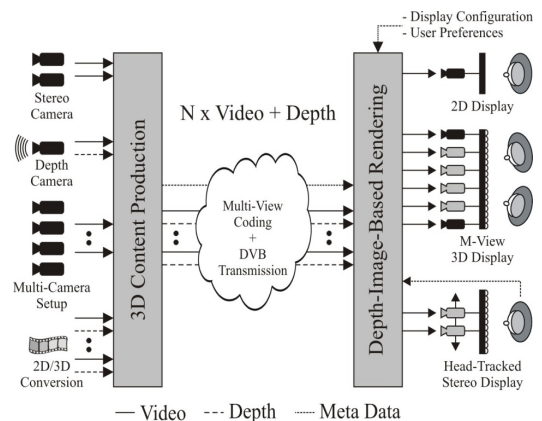


그림 1. 3차원 비디오 시스템의 개념도

서는 복원한 N시점 영상보다 많은 수의 영상 혹은 보다 좁은 간격의 시점 영상이 필요한 경우를 대비하여, 임의의 시점 영상에 대한 중간시점 영상합성 기능을 포함해야 한다.

III. 제안하는 3차원 영상 생성 방법

3.1 시간적 상관도를 고려한 깊이맵 추정 방법

제안하는 방법은 현재 프레임의 깊이맵을 추정할 때, 새로운 정합 함수를 사용하여 깊이맵의 시간적 상관도를 향상시킨다⁵⁾. 새로운 정합함수에는 이전 프레임에서의 깊이값을 이용하는 가중치 함수가 더해진다. 시간적 상관도를 고려한 새로운 정합 함수는 다음과 같이 나타낼 수 있다.

$$E_{new}(x,y,d) = E_{sim}(x,y,d) + E_{temp}(x,y,d) \quad (1)$$

$$E_{temp}(x,y,d) = \lambda|d - D_{prev}(x,y)| \quad (2)$$

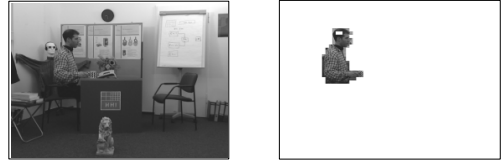
여기서 $E_{sim}(x,y,d)$ 은 기존의 정합 함수, λ 는 가중치 함수의 기울기, $D_{prev}(x,y)$ 는 이전 프레임에서의 깊이값을 나타낸다.

앞서 언급한 가중치 함수는 이전 프레임의 정확히 같은 위치에서의 깊이값을 참조하기 때문에 움직임이 없는 영역에 대해서만 유효하다. 그러므로 제안하는 방법은 움직이는 객체를 추정된 다음, 이를 제외한 영역에 대해서 가중치 함수를 적용한다. 사용자들은 움직임이 있는 영역에서는 떨림 현상을 쉽게 느끼지 못하고 움직임이 없는 배경 영역에 대해서 이를 느낄 수 있기 때문에, 배경 영역에 대해서만 가중치 함수를 추가하여도 무관하다.

따라서, 제안하는 방법은 움직이는 객체 분리를 위해서 영상 내의 16x16 크기의 블록에 대해서 평균 절대차 (mean absolute difference, MAD)를 계산한 다음, 임계값을 이용하여 해당 블록이 움직임이 있는지 없는지 여부를 판단한다. 그러므로 앞서 언급한 가중치 함수의 기울기 λ 는 다음과 같이 정의될 수 있다.

$$\lambda = \begin{cases} 1 & \text{if } MAD_k < Th \\ 0 & \text{if } MAD_k \geq Th \end{cases} \quad (3)$$

여기서 MAD_k 는 좌표 (x,y) 를 포함하는 k 번째 블록의 MAD를 나타내고 Th 는 임계값을 나타낸다. 그림 2는 움직이는 객체 분리 결과를 나타낸다.



(a) 원영상

(b) 객체 분리 결과

그림 2. "Book Arrival"의 객체 분리 과정

3.2 경계 잡음을 제거하기 위한 영상합성 방법

3차원 워핑을 이용하여 참조영상의 시점을 이동하고 비폐색 영역을 다른 참조영상으로부터 채우면 완벽하게 가상시점의 영상을 얻을 수 있다⁶⁾. 그러나 이 방법은 깊이값이 정확할 때에만 가능한 방법이다. 깊이맵은 컴퓨터 비전의 기술을 이용한 스테레오 정합 방법을 이용하거나⁷⁾, 깊이 카메라를 이용하여 획득할 수 있다⁸⁾. 현재 MPEG 3차원 비디오 부호화의 표준화에서 고려하고 있는 기술은 스테레오 정합 방법 기반의 깊이맵 추정 방법이다. 이 방법은 지금까지 많은 발전을 이루어 왔지만 여전히 문제점들이 존재한다. 이러한 문제점들로 인해 깊이맵은 오차를 갖게 되고, 깊이맵의 오차는 중간시점 영상 생성 과정에서도 잡음을 발생시킨다.

깊이맵의 오차는 깊이맵에서의 객체와 색상 영상에서의 객체의 경계를 불일치시키고, 이것은 중간시점 영상합성 과정에 그대로 반영된다. 그 결과, 전경에 위치한 객체의 일부분이 배경에 위치하게 된다. 그림 3은 중간시점 영상합성 과정에서 발생하는 경계 잡음을 나타낸 것이다. 그림 3(a)는 깊이맵에서의 객체와 색상 영상에서의 객체 경계의 불일치를 나타내고 있다. 그림 3(b)에서 알 수 있듯이, 객체의 의 파란 네모를 살펴보면, 전경 객체의 일부 정보들이 배경에 위치해 있음을 알 수 있다. 이러한 잡음을 본 논문에서는 경계 잡음(boundary noise)이라 정의하고 그 해결방법을 제시한다.

경계 잡음은 객체의 경계 부분과 비폐색 영역의 배경 부분에서 주로 발생한다. 이러한 가정에 착안하여



(a) 객체의 경계 불일치

(b) 경계 잡음

그림 3. 경계 잡음

비폐색 영역의 경계를 따라 배경에 해당하는 부분을 구별하고, 그 주변의 화소값을 대체하면 경계 잡음을 해결할 수 있다. 비폐색 영역은 3차원 워핑을 통해 쉽게 알 수 있다. 그렇다면 어떤 경계 부분이 배경에 해당하는지 구분해야 한다. 그림 4는 경계 잡음 처리 방법의 순서도를 나타낸다.

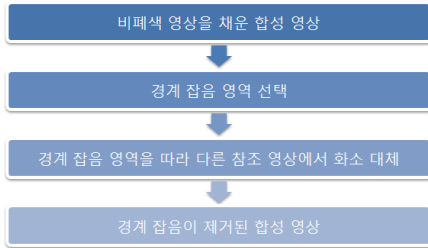


그림 4. 경계 잡음 처리 방법

그림 5를 살펴보면 비폐색 영역에는 깊이값이 할당되어 있지 않다. 반면 전경과 배경에는 각각 깊이값이 할당되어 있다. 이를 이용하여 비폐색 영역의 경계를 따라 비폐색 영역의 경계에서 배경과 맞닿아 있는 부분을 찾아낸다. 만약 동일한 y좌표에서 비폐색 영역과 닿아있는 배경의 깊이값이 101이고 전경의 깊이값이 155라면, 우리는 155가 전경임을 알 수 있다. 비폐색 영역 좌우의 깊이값이 미리 지정한 문턱치를 넘어가면 작은 깊이값을 가지는 부분이 배경이라고 설정한다.

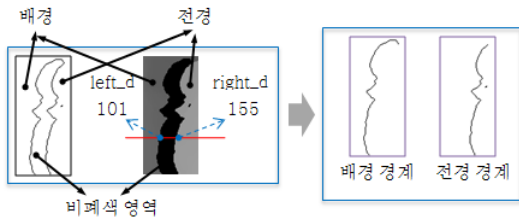


그림 5. 비폐색 영역의 전경/배경 구분

그림 6은 위의 과정에서 결정한 배경 경계를 이용하여 경계 잡음을 제거하는 과정을 나타낸 것이다. 이전 단계에서 구분한 경계를 따라서 다른 시점에서 합성된 영상의 부분을 경계 잡음이 있는 부분에 대체한다. 즉, 왼쪽 참조영상에서 가상시점으로 시점 이동한 영상의 배경 잡음 영역을 오른쪽 참조영상에서 가상시점으로 시점 이동한 영상으로 대체한다. 다른 시점의 합성영상은 경계 잡음이 존재하지 않으므로 깨끗한 합성영상을 생성할 수 있다.

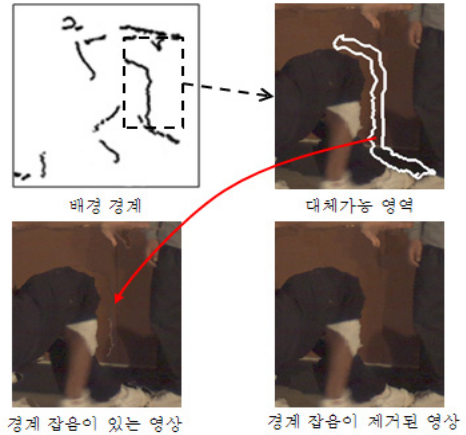


그림 6. 경계 잡음 처리 과정

IV. 실험 결과 및 분석

4.1 깊이맵의 성능 평가

본 논문에서 제안하는 깊이맵 추정 방법의 성능을 평가하기 위해 MPEG의 3차원 비디오 부호화 그룹에서 테스트 영상으로 사용 중인 "Book Arrival", "Alt Moabit", "Newspaper", "Lovebird1" 영상을 사용했다^{[9]-[12]}. 깊이맵 추정은 MPEG 3차원 비디오 부호화 그룹에서 제공한 깊이 추정 참조 소프트웨어 (Depth Estimation Reference Software, DERS) 버전 3.0을 사용했다. 실험에 사용한 임계값은 표 1과 같다.

표 1. 객체 분리를 위한 임계값

테스트 영상	Th
Book Arrival	2.50
Alt Moabit	2.50
Newspaper	1.50
Lovebird1	1.50

그림 7은 "Book Arrival" 영상에 대한 깊이맵을 나타낸다. 그림 7(a)에서 알 수 있듯이, 움직임이 없는 배경은 같은 깊이값을 가져야 함에도 불구하고 다른 깊이값을 가지는 것을 확인할 수 있다. 반면에, 시간적 상관도 향상 방법을 적용하게 되면, 그림 7(b)에서와 같이 움직임이 없는 물체는 동일한 깊이값을 가지는 것을 확인할 수 있다.

깊이맵의 정확도 평가를 위해서, 각 영상의 깊이맵을 탐색하였고, 이를 이용하여 영상을 시점 이동하여

합성한 후, 원본 영상과 비교하였다. 프레임 수는 100 프레임으로 정하였다. 표 2는 합성영상의 PSNR을 나타낸다. 표 2를 통해 제안하는 방법이 깊이맵의 정확도를 그대로 유지시키며 시간적 상관도를 향상시켰다는 것을 확인하였다.

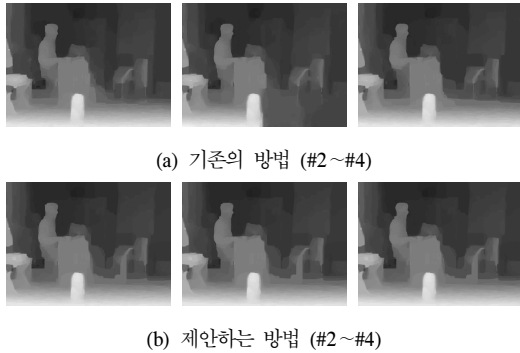


그림 7. "Book Arrival" 깊이맵 추정 결과

표 2. 합성영상의 평균 화질 비교

테스트 영상	시점	기존의 방법 (dB)	제안하는 방법 (dB)	Δ dB
Book Arrival	8	34.40	34.48	+0.08
Alt Moabit	8	35.15	35.34	+0.19
Newspaper	4	24.37	24.37	-
Lovebird1	6	30.99	31.00	+0.01

4.2 합성영상의 성능 평가

제안하는 경계 잡음 제거 방법을 평가하기 위해 Microsoft Research에서 제공한 "Breakdancers" 영상과 "Ballet" 영상을 이용하여 실험했다^[13]. "Breakdancers" 영상과 "Ballet" 영상은 영역 기반으로 측정된 깊이맵이 제공되므로, 이를 이용하여 중간시점 영상을 합성했다. 그림 8의 결과에서 확인할 수 있듯이, 기존의 방법은 남자의 팔 주변에 경계 잡음이 생기지만, 제안하는 방법을 사용한 결과에서는 잡음이 사라졌음을 확인할 수 있었다.

표 3은 합성된 영상과 원본 영상 100프레임의 평균 PSNR을 계산한 결과를 나타낸다. Δ PSNR을 보면 알 수 있듯이, 제안하는 방법이 합성영상의 화질을 떨어뜨리지 않고도 경계 잡음을 효과

적으로 처리했음을 알 수 있었다.

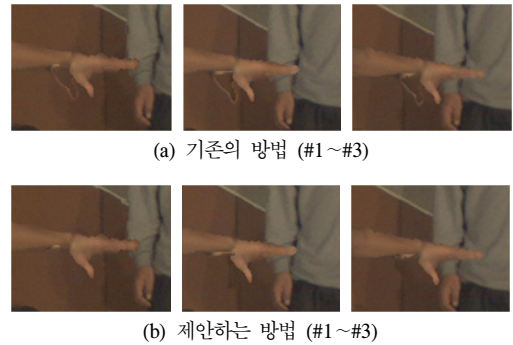


그림 8. "Breakdancers" 중간시점 영상합성 결과

표 3. 합성영상의 평균 화질 비교

테스트 영상	기존의 방법 (dB)	제안하는 방법 (dB)	Δ dB
Breakdancers	30.36	30.43	+0.07
Ballet	25.06	25.10	+0.04

V. 결 론

본 논문에서는 3차원 영상 생성을 위한 깊이맵 추정 방법과 중간시점 영상합성 방법을 제안했다. 제안한 방법은 기존의 정합함수에 가중치 함수를 추가하여 가중치 함수가 효과를 볼 수 있는 배경 영역에 대해서 시간적 상관도를 높였다. 또한, 경계 잡음 제거를 위해 영상합성 과정에서 배경에 해당하는 경계를 구분하고, 경계 주변 정보를 다른 참조영상의 정확한 정보를 이용했다. 그 결과, 제안한 방법을 통해 평균 약 0.07dB의 화질을 향상시켰고, 합성된 시점의 시간적 상관도를 향상시켜 사용자의 시각적 피로감을 줄일 수 있었으며 영상합성 과정에서 발생하는 경계 잡음을 제거하여 사용자에게 보다 자연스러운 다시점 영상을 제공할 수 있었다.

참 고 문 헌

- [1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards," IEEE International Conference on Multimedia and

- Expo, pp. 2161-2164, July 2006.
- [2] ISO/IEC JTC1/SC29/WG11, "Vision on 3D Video," N10357, Feb. 2009.
 - [3] ISO/IEC JTC1/SC29/WG11 "Call for Contributions on 3D Video Test Material," N9595, Jan. 2008.
 - [4] ISO/IEC JTC1/SC29/WG11 "Reference Softwares for Depth Estimation and View Synthesis," M15377, April 2008.
 - [5] S.-B. Lee, C. Lee, and Y.-S. Ho, "Temporal Consistency Enhancement of Background for Depth Estimation," in Proc. of 3D Systems and Applications, pp. S08-01.1~4, April, 2009.
 - [6] W. R. Mark, L. McMillan, G. Bishop, "Post-Rendering 3D Warping," in Proc. of Symposium on Interactive 3D Graphics, pp. 7-16, April 1997.
 - [7] D. Sharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," Proc. of IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 131-140, Dec. 2001.
 - [8] C. Fehn, "Depth-image-based Rendering (DIBR), Compression and Transmission for a New Approach on 3DTV," Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 5291, pp. 93-104, 2004.
 - [9] ISO/IEC JTC1/SC29/WG11 "Contribution for 3D Video Test Material of Outdoor Scene," M15371, April 2008.
 - [10] ISO/IEC JTC1/SC29/WG11 "1D Parallel Test Sequences for MPEG-FTV," M15378, April 2008.
 - [11] ISO/IEC JTC1/SC29/WG11 "HHI Test Material for 3D Video," M15413, April 2008.
 - [12] ISO/IEC JTC1/SC29/WG11 "Multiview Video Test Sequence and Camera Parameters," M15419, April 2008.
 - [13] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, "High-quality Video View Interpolation Using a Layered Representation," in Proc. of SIGGRAPH'04, pp. 600-608, Aug. 2004.

이 상 범 (Sang-Beom Lee)

정회원



2004년 경북대학교 전자전기공학부 학사
2006년 광주과학기술원 정보통신공학과 석사
2006년~현재 광주과학기술원 정보통신공학과 박사과정
<관심분야> 3차원 TV, 실감방송, 3차원 비디오 부호화

이 천 (Cheon Lee)

정회원



2005년 한국항공대학교 항공전자 및 정보통신공학부 학사
2007년 광주과학기술원 정보통신공학과 석사
2007년~현재 광주과학기술원 정보통신공학과 박사과정
<관심분야> 다시점 비디오 부호화, 영상신호 처리 및 압축, 3차원 비디오 부호화

호 요 성 (Yo-Sung Ho)

종신회원



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

1983년~1995년 한국전자통신연구원 선임연구원

1990년~1993년 미국 Philips 연구소, Senior

Research Member

1995년~현재 광주과학기술원 정보통신공학과 교수

<관심분야> 디지털 신호처리, 영상 신호 처리 및 압축, 디지털 TV와 고선명 TV, 멀티미디어 시스템, MPEG 표준, 3차원 TV, 실감방송