

키넥트 카메라를 이용한 실시간 가상 시점 영상 생성 기법

이 규 철*, 유 지 상^o

Real-Time Virtual-View Image Synthesis Algorithm Using Kinect Camera

Gyu-cheol Lee*, Jisang Yoo^o

요 약

키넥트 카메라는 마이크로소프트사에서 2010년 11월에 출시한 xbox360의 움직임 감지 카메라로 깊이 영상과 색상 영상을 획득할 수 있다. 하지만 적외선 패턴을 이용한 깊이 영상의 획득 방법의 한계로 인해 객체의 경계 주변으로 홀(hole) 및 잡음이 생기고 영상으로 재생 시 경계 주변에서 흔들림(flickering) 현상이 발생한다. 본 논문에서는 이러한 흔들림 현상을 보정하여 화질이 좋은 가상 시점 영상을 실시간으로 생성하는 기법을 제안한다. 제안하는 기법에서는 먼저 결합형 양방향 필터를 이용하여 경계 주변의 홀을 채운다. 경계 주변의 흔들림 현상은 화소를 탐색하여 처리하는 기법을 적용하여 보정한다. 향상된 깊이 영상과 색상 영상에 3D 워핑(3D warping) 기법을 적용하여 가상 시점 영상을 획득한다. 획득된 영상에서 가려짐 영역(occlusion region)으로 인하여 생기는 홀은 블록기반의 기울기 탐색 기법과 블록의 신뢰도를 이용하여 채우게 된다. 실험을 통해 제안하는 시스템이 가상 시점 영상을 실시간으로 합성하는 것을 확인하였다.

Key Words : Kinect, Virtual viewpoint image, Depth image enhancement, Hole filling, 3D Warping

ABSTRACT

Kinect released by Microsoft in November 2010 is a motion sensing camera in xbox360 and gives depth and color images. However, Kinect camera also generates holes and noise around object boundaries in the obtained images because it uses infrared pattern. Also, boundary flickering phenomenon occurs. Therefore, we propose a real-time virtual-view video synthesis algorithm which results in a high-quality virtual view by solving these problems. In the proposed algorithm, holes around the boundary are filled by using the joint bilateral filter. Color image is converted into intensity image and then flickering pixels are searched by analyzing the variation of intensity and depth images. Finally, boundary flickering phenomenon can be reduced by converting values of flickering pixels into the maximum pixel value of a previous depth image and virtual views are generated by applying 3D warping technique. Holes existing on regions that are not part of occlusion region are also filled with a center pixel value of the highest reliability block after the final block reliability is calculated by using a block based gradient searching algorithm with block reliability. The experimental results show that the proposed algorithm generated the virtual view image in real-time.

※ 이 논문은 2012년도 광운대학교 교내 학술연구비 지원에 의해 연구되었음.

♦ 주저자 : 광운대학교 전자공학과 디지털미디어 연구실, gyucheol0116@gmail.com, 학생회원

o 교신저자 : 광운대학교 전자공학과 디지털미디어 연구실, jsyoo@kw.ac.kr, 종신회원

논문번호 : KICS2013-02-092, 접수일자 : 2013년 2월 14일, 최종논문접수일자 : 2013년 5월 1일

I. 서 론

2009년 개봉한 3D 영화 아바타의 흥행 성공 이후 지난 3년 동안 3D 입체영상은 영화, 애니메이션, 게임, 스포츠 중계 등 다양한 분야의 미디어 서비스로 활용되고 있으며, 3D 기능을 탑재한 TV, 노트북, 휴대전화, 게임기 등의 제품이 연이어 출시되면서 3D 시장은 양적 성장을 거듭하고 있다. 하지만 안경을 써야 하는 불편함, 다양한 콘텐츠의 부족으로 인하여 아직까지 일반화되지 못하고 있다. 실제 시장 조사기관 Retrevo가 발표한 자료에 따르면, 2012년 HDTV 구매 계획을 가진 소비자 가운데 55%가 3D 기능의 필요성에 대해 의문을 제기했으며, 그 이유로는 안경을 써야 하는 번거로움과 콘텐츠의 부족으로 나타났다¹⁾.

또 다른 이유로는 현재 3D 디스플레이의 방식은 입체감을 느낄 수 있는 시점이 하나인 단시점 방식이기 때문에 시점을 벗어나면 물체의 실감과 생동감이 떨어지기 때문이다. 이에 대한 대안으로 최근 주목 받고 있는 기술이 다시점 무안경 디스플레이 기술이다. 다시점 무안경 디스플레이는 스테레오 디스플레이보다 시점의 개수가 증가하기 때문에 안경을 착용하지 않고 관찰자가 시점을 바꾸는 경우에도 시점에 맞는 영상이 재생되어 보다 자연스럽게 실감나는 3차원 영상의 감상이 가능하다²⁾.

다시점 영상을 획득하는 방법 중 하나가 가상 시점 영상을 생성하는 것이다. 가상 시점 영상을 생성하기 위해서는 일반적으로 색상 영상과 색상 영상에 대응하는 깊이 영상을 획득해야 한다. 깊이 영상을 획득하는 방법에는 스테레오 정합을 이용하는 방법과 TOF 카메라 등 깊이 카메라를 이용하는 방법이 있다³⁾. 스테레오 정합은 환경의 제약이 적게 받고 효과적이지만 수행시간이 길고 깊이 영상의 정확도가 떨어지며 TOF 카메라는 높은 정확도의 깊이 영상을 획득할 수 있는 대신 해상도가 낮고 장비가 고가라는 단점이 있다. 최근에 키넥트(Kinect) 카메라와 같이 가격 대비 높은 성능을 보여주는 카메라가 시중에 출시되면서 일반인들도 손쉽게 깊이 영상을 획득할 수 있는 카메라를 접할 수 있게 되었다⁴⁾. 하지만 키넥트 카메라로 획득된 깊이 영상은 카메라 내외의 문제점들로 인해 홀이 생기고 각각의 깊이 영상을 동영상으로 연속 재생할 경우 경계 주변의 흔들림(flickering) 현상이 발생하기 때문에 현재 이를 보정하는 연구가 진행되고 있다⁵⁾.

가상 시점 영상 합성을 위한 방법에는 크게 스테레오 영상의 변이 정보를 이용하는 양방향 선형 보간법

(bidirectional linear interpolation)과 깊이 정보를 이용하는 3D 워핑(3D warping) 기법 등이 있다. 양방향 선형 보간법은 변이 정보를 이용하여 기준 시점 영상 사이의 제한된 시점의 중간 시점 영상을 생성하는 방법이다⁶⁾. 3D 워핑은 카메라의 기하학적 구조를 기반으로 깊이 정보와 카메라의 내, 외부 파라미터를 이용하여 영상의 화소들에 대한 실제 좌표(세계 좌표계; world coordinate)를 산출하고 다시 가상 시점의 영상으로 재투영하는 기법이다. 중간 시점뿐만 아니라 임의의 가상 시점 영상을 생성함으로써 보다 자유로운 시점을 제공한다⁷⁾.

본 논문에서는 키넥트 카메라를 이용하여 가상 시점 영상을 생성하는 기법을 제안한다. 우선 키넥트 카메라에서 획득된 깊이 영상은 카메라 내외적인 문제로 인하여 홀이 생기기 때문에 홀을 채워야 한다. 기존의 방법으로는 인페인팅(inpainting) 기법을 이용한 홀 채움 방법과 결합형 양방향 필터(joint bilateral filter)를 이용하여 참조 영상의 윤곽선대로 홀을 채우는 방법이 대표적이다⁸⁾. 인페인팅 방식은 실사 이미지에서 제거된 특정 영역을 주변 배경과 어울리게 채워 넣는 기술로 깊이 영상의 홀 채움에 적용하여 사용될 수 있다^{9,10)}. 하지만 인페인팅 기법은 배경 영역과 객체 영역을 구분하지 못하기 때문에 홀 영역을 흐릿하게 채워 기존 배경과의 연속성을 떨어뜨린다. 또한 이 같은 방법들은 한 장의 영상에 적용하는 기법이기 때문에 홀은 채울 수 있지만 경계 흔들림 현상을 보정할 수 없다는 단점이 있다. 이러한 문제를 해결하기 위해 깊이 영상의 흔들리는 화소를 탐색하여 처리하는 기법을 제안한다. 키넥트 카메라에서 획득된 색상 영상을 HSI 색상모델의 명도(intensity) 영상으로 변환하고, 이전 프레임의 명도 영상과 깊이 영상의 값을 분석하여 흔들리는 화소를 찾는다. 흔들리는 화소는 이전 프레임 중 최대 화소 값으로 변환한다. 하지만 흔들리는 화소를 잘못 찾는 경우, 객체가 움직일 때 잔상이 생긴다. 이를 보정해 주기 위해 키넥트 카메라에서 획득된 깊이 영상에서 홀 영역과 홀의 주변 영역을 변환한 값으로 대체해 줌으로써 객체의 움직임을 보존하면서 잔상을 줄일 수 있다¹¹⁾.

향상된 깊이 영상과 키넥트 카메라에서 획득된 색상 영상에 3D 워핑 기법을 적용하여 가상 시점을 생성한다¹²⁾. 먼저 기준 영상 화소들의 실제 좌표를 산출하고 다시 원하는 가상 시점으로 재투영한다. 이때 기준 시점 영상에서는 존재하지 않는 가려짐(occlusion) 영역이 홀의 형태로 영상에 나타나게 된다¹³⁾. 가려짐 영역을 채우는 대표적인 기법에는 인페인팅 기법, 선

형보간법, 나선형 가중 평균 기법 등이 있다^{7,9,10}. 인페인팅 기법은 블러 현상(blur effect)이 발생하며 수행시간이 길기 때문에 실시간 구현이 어렵다. 선형보간법은 실시간으로 구현이 가능하나 영상의 연속성을 유지하기 힘들고 경계 잡음을 제거하지 않기 때문에 홀 채움 시 잘못된 색상 정보를 사용하여 영상이 부자연스러워진다. 나선형 가중 평균 기법은 영상의 깊이 정보를 이용하여 홀을 채우는 방법으로 객체와 영상을 구분하여 홀을 채우기 때문에 채워진 영상이 자연스러워지지만 실시간 구현이 어렵다. 본 논문에서는 실시간으로 구현이 가능하고 영상의 연속성을 유지하면서 채워주는 블록기반의 기울기 탐색 기법을 제안한다. 홀 주변의 블록들의 기울기를 탐색하여 가장 높은 기울기 값을 가지고 있는 블록을 선정한다. 홀이 있는 블록을 참조할 가능성이 있기 때문에 블록의 신뢰도를 계산하여 블록의 최종 신뢰도를 산출하고 가장 높은 신뢰도를 가지는 블록의 중앙값을 이용하여 홀을 채운다. 기존 기법들과의 비교를 통하여 제안하는 기법의 성능이 우수하다는 것을 확인하였다.

본 논문의 2장에서는 키넥트 카메라에서 획득된 깊이 영상을 이용한 가상 시점 영상 생성 기법에 대해 설명하고, 3장에서는 가상 시점 영상 생성 시 생기는 홀을 채우는 후처리 기법을 제안한다. 4장에서는 실험을 통하여 제안한 기법의 성능을 평가하며, 5장에서 결론을 맺는다.

II. 키넥트 카메라의 깊이 영상을 이용한 가상 시점 영상 생성기법

키넥트(Kinect) 카메라는 깊이 영상 생성을 위해 물체에 반사되어 돌아오는 적외선의 패턴을 수광센서로 검출한다. 하지만 발광장치와 수광센서간의 거리 차이로 인한 사각 지대(blind zone) 발생, 매끄러운 물체 표면으로 인하여 적외선의 난반사가 적게 일어나 수광센서로 레이저가 돌아오지 못하는 경우에 적외선 패턴을 감지할 수 없기 때문에 검출되지 않은 위치에 대한 값은 깊이 영상에서 홀(hole)의 형태로 나타난다. 또한 깊이 정보의 불연속점과 적외선 송출부가 수직일 경우에는 객체 경계 주변에 홀이 발생한다⁵. 따라서 각각의 깊이 영상을 동영상으로 연속 재생 할 경우 경계 주변의 흔들림(flickering) 현상을 확인할 수 있으며 가상 시점 영상을 생성하기 위해서는 이러한 흔들림 현상을 반드시 제거하여야 한다. 그림 1은 제안하는 키넥트 카메라의 깊이 영상 향상 기법의 흐름도이다.

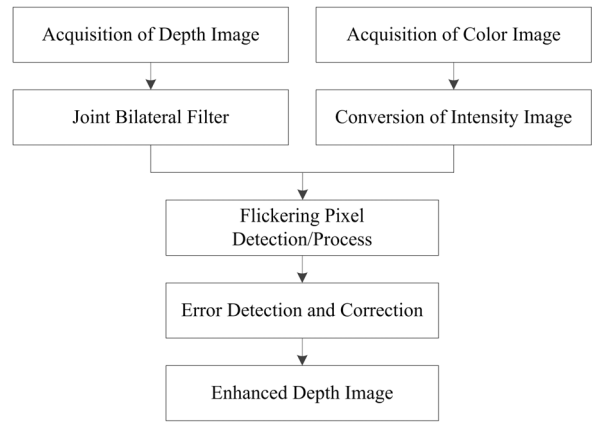


그림 1. 제안하는 깊이 영상 향상 기법
Fig. 1. Proposed depth image enhancement method

2.1. 결합형 양방향 필터

그림 2(a)는 키넥트 카메라에서 획득된 깊이 영상이다. 획득된 깊이 영상에서는 홀이 발생하기 때문에 가상 시점 합성을 위해서는 우선 홀을 채워야 한다. 본 논문에서는 그림 2(b)와 같이 참조 영상의 경계를 보존하여 채우는 결합형 양방향 필터(joint bilateral filter)를 이용하여 홀을 채운다¹⁴. 결합형 양방향 필터는 두 장의 영상에 대하여 인접 화소와의 거리 차, 경계를 보존하려는 참조 영상의 색상 차에 두 개의 가우시안 함수를 사용한다. 결합형 양방향 필터는 식 (1)과 식 (2)와 같이 정의된다.

$$D'_p = \frac{1}{W_p} \sum_{q \in s} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(I_p - I_q) D_q \quad (1)$$

$$W_p = \sum_{q \in s} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(I_p - I_q) \quad (2)$$

여기서 D'_p 는 깊이 영상 D 와 명도(intensity) 영상 I 에 결합형 양방향 필터를 적용하여 생성한 화소 값이다. $G_{\alpha}(\|p - q\|)$ 는 화소 p 와 화소 q 사이의 유클리디안 거리(Euclidean distance)에 따른 가우시안 가중치이고 $G_{\alpha}(I_p - I_q)$ 는 명도 영상 I 의 p 화소 값과 q 화소 값 차에 따른 가우시안 가중치이다. s 는 영상에서 공간상의 모든 위치를 나타내며, α 와 σ_r 는 필터를 적용할 범위이다. W_p 는 가중치의 합을 1.0으로 만드는 정규화 인자이다.

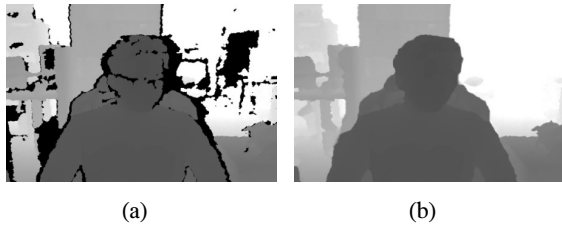


그림 2. (a) 키넥트 카메라에서 획득된 깊이 영상 (b) 결합형 양방향 필터를 적용한 결과
 Fig. 2. (a) Depth image acquired by Kinect camera (b) Result of joint bilateral filtering

2.2. 경계 흔들림 처리 기법

키넥트 카메라에서 획득한 깊이 영상은 적외선 패턴을 이용한 감지방법의 오차로 인하여 깊이 값의 변동 폭이 심하다. 따라서 연속해서 획득된 깊이 영상을 동영상으로 연속 재생할 경우 경계 주변의 흔들림 현상을 쉽게 관측할 수 있다. 결합형 양방향 필터는 각 영상에 독립적으로 적용되기 때문에 이러한 흔들림 현상을 해결할 수 없다.

그림 3은 흔들리는 화소를 탐색하는 방법이다. 먼저 결합형 양방향 필터로 홀을 채운 N 개의 깊이 영상의 평균 영상을 구한다. 모든 화소에 대하여 평균 영상의 화소 값과 현재 깊이 영상 화소 값의 차이 값 $D'(i,j)$ 이 임계값 α_1 보다 크다면 흔들림이나 움직임이 있는 화소이다. 또한 색상 영상을 명도 영상으로 변환한 후 평균을 구한다. 평균 영상의 화소 값과 현재 명도 영상의 화소 값의 차이 값 $I'(i,j)$ 이 임계값 α_2 보다 작다면 움직임이 없는 화소로 판단하여 최종적으로 경계 흔들림이 있는 화소로 판단하게 된다. 임계값에 의해 성능이 좌우되기 때문에 α_1 , α_2 의 값은 실험을 통해 최적의 값을 결정한다. α_1 의 값이 작을수록 흔들리거나 움직이는 화소로 판단하는 화소의 범위가 넓어지며 α_2 의 값이 작을수록 흔들리는 화소로 판단하는 화소의 범위가 좁아진다. 식 (3)과 식 (4)는 흔들리는 화소의 조건을 표현한 식이다.

$$D_N - \frac{1}{N} \sum_{n=1}^N D_n > \alpha_1 \quad (3)$$

$$I_N - \frac{1}{N} \sum_{n=1}^N I_n < \alpha_2 \quad (4)$$

여기서 D_n 은 n 번째 결합형 양방향 필터를 적용하여 홀을 채운 깊이 영상의 화소 값이고 I_n 은 n 번째 명도 영상의 화소 값이다.

흔들리는 화소는 배경으로 판단하고 N 개의 깊이 영상의 같은 위치에서 가장 큰 값을 갖는 화소로 대체한다. 그림 4는 현재 프레임과 이전 프레임의 차영상이다. 흔들리는 화소를 보정한 후 차이가 많이 줄어든 것을 확인할 수 있다.

객체가 움직이는 사물일 경우에는 경계 흔들림이 있는 화소를 탐색하는 과정에서 오류가 발생하여 모션 블러(motion blur) 현상이 발생할 수 있다. 모션 블러 현상을 제거하면서 객체의 움직임을 보존하기 위해 깊이 영상에서 홀 영역과 주변 영역은 처리한 흔들리는 화소 값으로 변환하고 나머지 영역은 원본 깊이 영상을 그대로 사용한다.

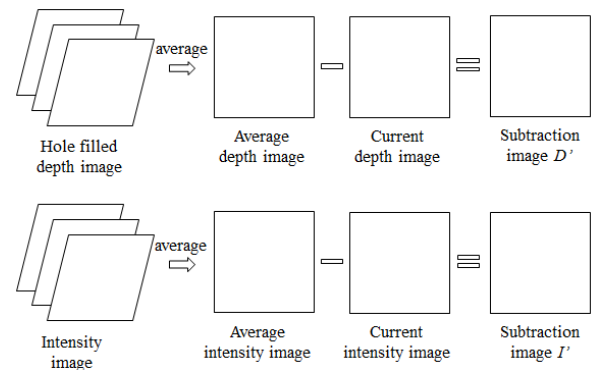


그림 3. 흔들리는 화소 탐색 방법
 Fig. 3. Detection of edge flickering pixels

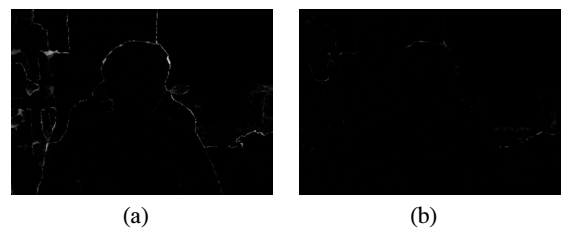


그림 4. 현재 프레임과 이전 프레임의 차영상
 (a) 흔들리는 화소 보정 전 (b) 흔들리는 화소 보정 후
 Fig. 4. Difference image between current and previous frames
 (a) Before flickering pixel processing (b) After flickering pixel processing

2.3. 3D 워핑을 이용한 가상 시점 영상 생성

3D 워핑은 카메라의 기하학적 구조를 기반으로 깊이 정보와 카메라의 내, 외부 파라미터를 이용하여 영상의 화소들에 대한 실세계 좌표(세계 좌표계, world coordinate)를 산출하고 다시 가상 시점의 영상으로 재투영하는 기법이다⁷⁾. 3D 실세계 좌표계와 카메라 좌표계의 관계식은 식 (5)와 같다.

III. 후처리 기법

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K[R|T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (5)$$

여기서 x, y 는 영상에 투영된 물체의 2차원 좌표이고, K 는 카메라 내부 파라미터 (intrinsic parameter), R 과 T 는 각각 회전(rotation)과 이동(translation) 행렬을 나타내며, X, Y, Z 는 영상에 투영된 물체의 실제 좌표를 의미한다. 가상 시점 영상을 생성하기 위해서는 영상의 실제 좌표 X, Y, Z 를 산출해야 한다. 영상의 실제 좌표 X, Y 는 식 (5)에 역 행렬과 전치 행렬을 적용하여 식 (6)와 같이 유도할 수 있다.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K[R|T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow K^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} + T \quad (6)$$

$$\Rightarrow R^T K^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} - R^T T = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

실제 좌표 Z 를 산출하기 위해서는 식 (7)을 이용한다.

$$Z(i,j) = \frac{1.0}{\left(\frac{D(i,j)}{255.0} \left(\frac{1.0}{MinZ} - \frac{1.0}{MaxZ} \right) + \frac{1.0}{MaxZ} \right)} \quad (7)$$

여기서 $Z(i,j)$ 는 좌표 (i,j) 에서 카메라와 객체간의 거리이고, $D(i,j)$ 는 깊이 영상의 화소 값을 나타낸다. $MinZ$ 와 $MaxZ$ 는 Z 가 가지는 최소, 최대값을 의미한다. 가상 시점의 좌표는 가상 시점 카메라의 내부, 외부 파라미터를 식 (8)에 적용하여 구할 수 있다.

$$\begin{bmatrix} x_v \\ y_v \\ 1 \end{bmatrix} = K_v[R_v|T_v] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (8)$$

여기서 x_v, y_v 는 생성된 가상 시점 영상의 2차원 좌표를 의미하며, K_v, R_v, T_v 는 각각 가상 시점 카메라의 내부 파라미터, 회전 행렬, 이동 벡터를 나타낸다.

기준 시점 영상과 깊이 영상에 3D 워핑(3D warping) 기법을 적용하여 가상 시점 영상을 생성하면 그림 5(c)와 같이 기준 시점 영상에서는 존재하지 않는 가려짐 영역(occlusion region)이 홀(hole)의 형태로 나타나게 된다⁷⁾. 가려짐 영역을 채우는 방법은 크게 두 가지로 나눌 수 있다. 첫째는 워핑을 적용하기 전에 깊이 영상에 저역 통과 필터(low pass filter)를 적용하여 평활화(smoothing)함으로써 가려짐 영역을 줄이는 방법이고 둘째는 워핑 기법을 적용한 후에 나타나는 가려짐 영역을 주변 색상 정보를 이용해 채우는 방법이다¹⁵⁾. 본 논문에서는 깊이 영상의 화질을 향상시킬 때 적용한 결합형 양방향 필터가 이미 평활화 효과를 가지고 있기 때문에 워핑 기법 후 생성된 가상 영상의 주변 정보를 이용하여 홀을 채우는 후처리 기법을 제안한다. 그림 6은 제안하는 홀 채움 기법의 흐름도이다.

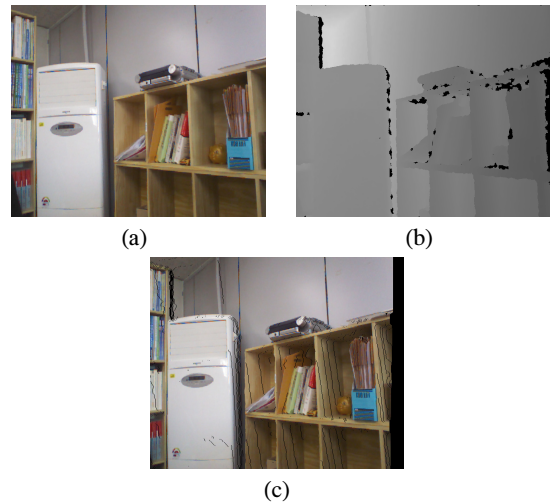


그림 5. 3D 워핑을 이용하여 생성된 가상 시점 영상
(a) 기준 시점 영상 (b) 깊이 영상 (c) 생성된 가상 시점 영상
Fig. 5. Virtual view-point image synthesized by using 3D warping
(a) Reference image (b) Depth image (c) Generated virtual view-point image

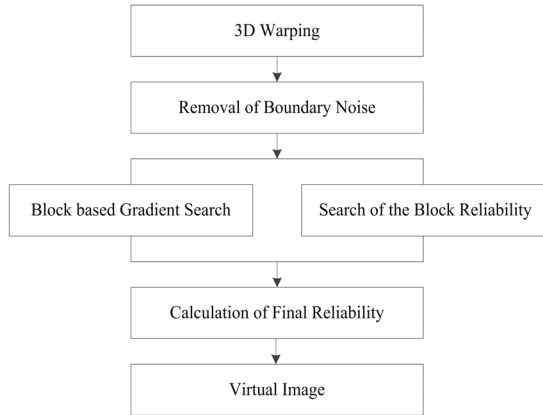


그림 6. 제안하는 홀 채움 기법
Fig. 6. Proposed hole filling method

3.1. 블록 기반의 기울기 탐색

주변 색상 정보를 이용하여 가려짐 영역을 채우는 기존의 방법들은 홀 주변의 영역들을 배경 영역과 객체 영역으로 분류한 후, 배경 영역에서 일부분을 복사하여 홀을 채운다. 가장 간단한 방식으로는 배경 영역이라 판단되는 영역이 시작되는 화소를 수평으로 복사하는 방식이 있다. 하지만 이러한 방식은 배경 영역이 특정한 구조이거나 수직의 윤곽선을 가질 경우 영상이 부자연스러워진다.

제안하는 기법에서는 기존 배경과 홀을 채우는 영역의 연속성을 높이기 위해 블록 기반의 기울기 탐색 기법을 적용한다. 그림 7은 홀을 중심으로 8개의 블록을 생성한 모습이다. 생성된 각각의 블록 내부에서 화살표 방향으로 기울기의 변화량을 조사하여 변화량이 큰 블록을 우선적으로 고려한다. 식 (9)는 블록의 기울기 변화량을 구하는 식이다.

$$\Delta G_B = \frac{1}{K} \sum_{\epsilon \in B} \| I_c - I_\epsilon \| \quad (9)$$

여기서 블록의 기울기 변화량인 ΔG_B 는 블록의 중앙값인 I_c 에서 블록 내부의 화소 값 I_ϵ 와의 차를 구하고 평균값을 구한 것이다. K 는 최대 기울기 변화량을 나타낸다.

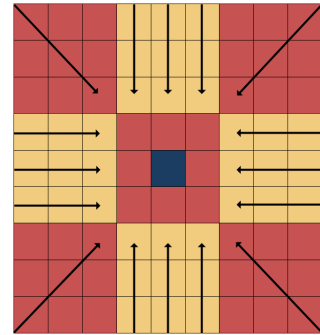


그림 7. 블록 기반의 기울기 탐색
Fig. 7. Block based gradient searching algorithm

3.2. 블록의 신뢰도

블록 기반의 기울기 탐색 방법은 가려짐 영역으로 인해 생기는 홀을 중심으로 주변의 블록들 중 가장 기울기 변화량이 큰 블록을 찾는다. 따라서 홀이 기울기가 큰 방향의 화소 값으로 채워지게 되어 영상의 연속성을 유지할 수 있다. 하지만 다수의 홀이 존재하는 블록을 참조하여 홀이 채워지지 않는 경우가 발생할 수 있다. 이러한 문제를 해결하기 위하여 블록의 신뢰도를 측정한다. 블록의 신뢰도는 블록 내부의 홀이 아닌 화소 수 대 블록 내부의 화소의 비로 구한다. 홀이 많은 블록은 신뢰도가 낮아 참조할 가능성을 낮춘다. 식 (10)은 블록의 신뢰도를 구하는 식이다.

$$C = \frac{\sum_{q \in B} D(q)}{|n(B)|} \quad (10)$$

$$\text{where, } D(q) = \begin{cases} 0, & q \in \text{hole} \\ 1, & \text{otherwise} \end{cases}$$

여기서 블록의 신뢰도를 나타내는 C 는 블록 내부의 홀 유무를 파악하는 $D(q)$ 의 합을 블록의 화소 수 $n(B)$ 로 나눈 값으로 표현된다.

3.3. 최적의 블록 탐색 및 홀 채움 과정

앞에서 구한 블록 기반의 기울기와 블록의 신뢰도를 곱하여 블록 내부의 기울기 성분을 최대한 고려하되 그중에서 높은 신뢰도를 가지는 블록을 홀 채움에 사용한다. 식 (11)은 블록의 최종 신뢰도를 나타내는 식이다.

$$T_n = \Delta G_{B_n} C_n, n \in N(B) \quad (11)$$

$$T = \text{argmax}(T_n)$$

여기서 T_n 은 n 번째 블록의 최종 신뢰도이고 n 번째 블록의 기울기 변화량 ΔG_{B_n} 과 신뢰도 C_n 의 곱으로 표현된다. $N(B)$ 는 블록의 개수이다. 홀 화소를 중심으로 $N(B)$ 개 블록 각각의 최종 신뢰도를 구한 다음 가장 높은 신뢰도를 갖는 블록을 찾는다. 가장 높은 신뢰도를 갖는 블록의 중앙 화소 값을 홀 화소의 최종 값으로 결정하게 된다.

3.4. 경계 잡음 제거

경계 잡음(boundary noise)은 그림 8(a)와 같이 기준 시점에서 색상 영상의 경계와 그에 대응하는 깊이 영상의 경계가 일치하지 않기 때문에 발생하는 잡음이다³⁾. 경계 잡음을 제거하지 않으면 잘못된 색상 정보로 홀을 채우기 때문에 그림 8(b)와 같이 자연스럽게 못한 결과를 얻게 된다. 본 논문에서는 홀에서 수평방향으로 화소의 값을 분석하여 값의 변동 폭이 큰 방향의 화소를 경계 잡음으로 검출한다. 객체 영역을 경계 잡음으로 판단하는 경우를 배제하기 위해 홀이 아닌 화소가 연속적으로 이어지는 영역은 경계 잡음 검출 범위에서 제외시킨다. 검출된 경계 잡음은 홀 영역으로 포함시켜 제거하게 된다. 그림 8(c)는 경계 잡음을 제거한 후 홀을 채운 영상이다. 그림 8(b)보다 자연스럽게 채워진 것을 확인할 수 있다.

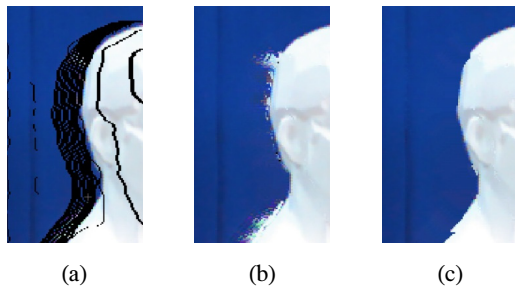


그림 8. (a) 경계 잡음 (b) 경계 잡음 제거 없이 홀 채운 한 결과 (c) 경계 잡음 제거 후 홀 채운 한 결과
Fig 8. (a) Boundary noise (b) Hole filling with boundary noise (c) Hole filling after removing the boundary noise

IV. 실험 및 결과

본 논문에서는 키넥트(Kinect) 카메라에서 획득된 깊이 영상과 색상 영상을 실험 영상으로 사용한다. 제안하는 기법의 흔들리는(flickering) 화소 탐색의 프레임 수는 실험을 통해 가장 좋은 결과가 나온 다섯 장으로 하였다. a_1 의 값은 작을수록 흔들리거나 움직이는 화소로 판단하는 화소의 범위가 넓어지며, a_2 의 값은 작을수록 흔들리는 화소로 판단하는 화소의 범위가

가 좁아진다. a_1, a_2 의 값에 따라 흔들림 현상 제거 효과가 발생하지 않거나 블러 현상(blur effect)이 발생한다. 각각의 값은 실험을 통해 주관적 화질이 가장 좋게 나오는 2와 6으로 설정하였다. 블록기반의 기울기 탐색 기법의 블록 크기는 3X3으로 하였다. 제안하는 기법의 성능을 기존의 선형 보간법(linear interpolation)¹¹⁾, 인페인팅(in-painting) 기법^{9, 10)}과 비교하였다.

그림 9와 그림 10은 흔들리는 화소 보정 후 3D 워핑(3D warping) 기법을 적용하여 가상 시점 영상을 생성한 후 각각의 홀 채움 방법을 적용한 결과 영상이다. 기존의 기법들은 경계 잡음을 제거하지 않고 홀 채움을 하기 때문에 잘못된 색상 정보를 참조하여 홀 채움 영역이 부자연스러워진다. 또한 배경 영역과 객체 영역을 구분하지 않기 때문에 영상에 흐려짐 현상(blurring effect)이 발생한다. 제안하는 기법은 경계 잡음을 제거하고 홀 채움을 하기 때문에 비교적 자연스러운 영상을 획득할 수 있다.

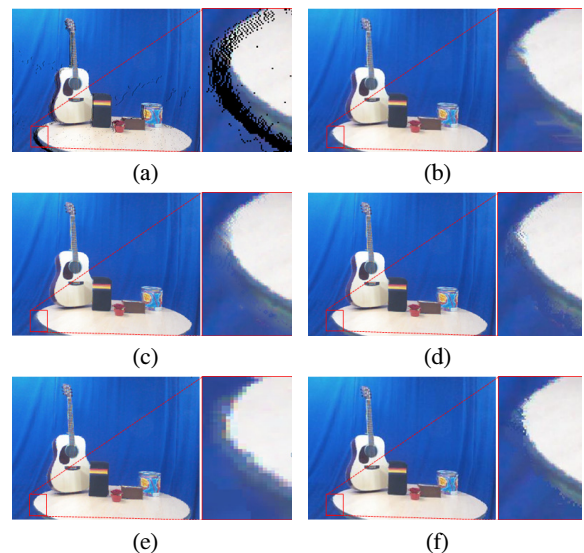


그림 9. 후처리된 가상 시점 영상
(a) 후처리 전 영상 (b) 선형 보간법 (c) Telea의 인페인팅 기법 (d) Criminisi의 인페인팅 기법 (e) 나선형 가중 평균 기법 (f) 제안하는 기법
Fig. 9. Result of virtual view-point image after post-processing
(a)Before processing (b) Linear interpolation (c) Telea's inpainting (d) Criminisi's inpainting (e) Spiral weighted average method (f) Proposed method

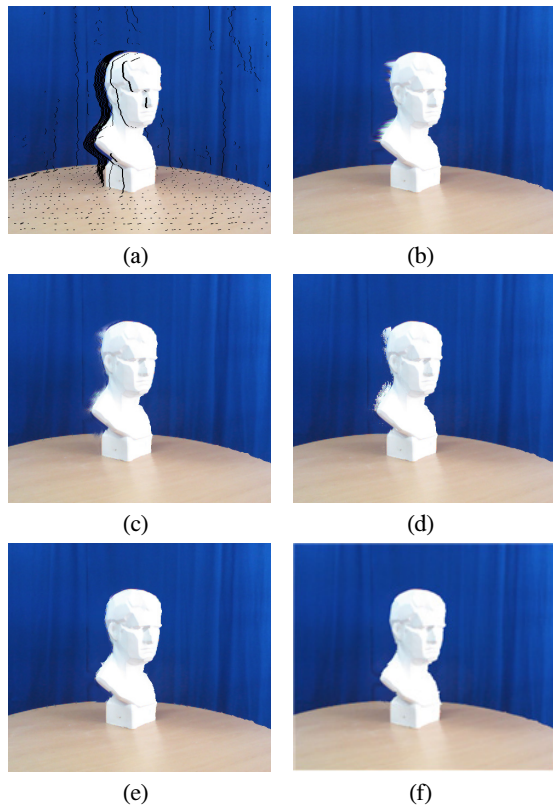


그림 10. 후처리된 가상 시점 영상
 (a) 후처리 전 영상 (b) 선형 보간법 (c) Telea의 인페인팅 기법
 (d) Criminisi의 인페인팅 기법 (e) 나선형 가중 평균 기법 (f) 제
 안하는 기법
 Fig. 10. Result of virtual view-point image after
 post-processing
 (a)Before processing (b) Linear interpolation (c) Telea's
 inpainting (d) Criminisi's inpainting (e) Spiral weighted
 average method (f) Proposed method

그림 11은 홀 채움 후 현재 프레임에서 이전 프레임을 뺀 차영상이다. 그림 11(a)는 흔들리는 화소를 보정하지 않았기 때문에 영상의 흔들림 현상이 나타난다. 그림 11(b)는 흔들리는 화소를 탐색하여 보정한 영상이다. 그림 11(a)에 비해 흔들림 현상이 줄어들었음을 확인할 수 있다.

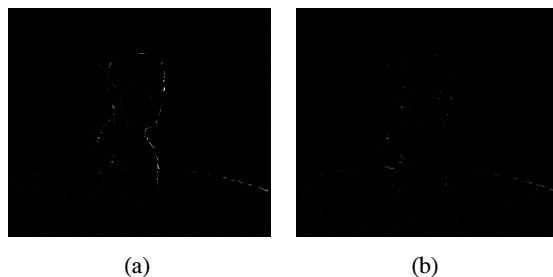


그림 11. 홀 채움 후 현재 프레임과 이전 프레임의 차영상
 (a) 흔들리는 화소 처리 전 (b) 흔들리는 화소 처리 후
 Fig. 11. Difference image between current and previous
 frames after hole filling
 (a) Before flickering pixel processing (b) After flickering pixel
 processing

그림 12는 움직이는 객체에 대하여 제안하는 기법을 적용한 영상이다. 가로는 순서대로 깊이 영상 워핑 결과, 컬러 영상 워핑 결과, 제안하는 홀 채움 기법을 적용한 결과, 세로는 순서대로 1, 2, 3, 4, 5번째 프레임을 각각 나타낸다. 객체가 심하게 움직일 경우 깊이 영상이 정확하게 추출되지 않아 생성된 가상 시점의 품질이 떨어지지만 객체의 움직임이 제한적일 경우 비교적 정확한 깊이 영상의 추출이 가능하며 제안하는 시스템을 통해 실시간으로 가상 시점을 생성하는 것이 가능하다.

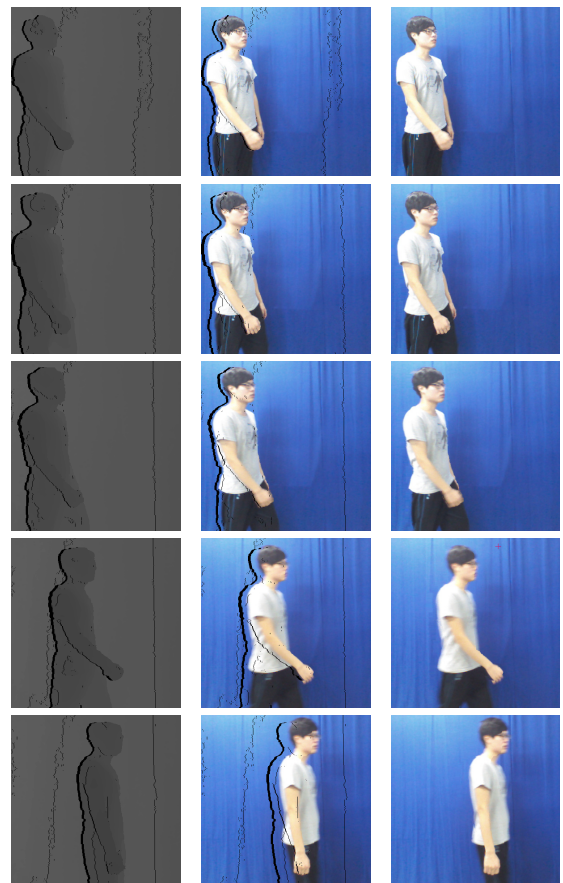


그림 12. 움직이는 객체에 제안하는 기법을 적용한 결과. 왼쪽부터 깊이 영상 워핑 결과, 컬러 영상 워핑 결과, 제안하는 홀 채움 기법을 적용한 결과. 위쪽부터 첫 번째 프레임, 두 번째 프레임, 세 번째 프레임, 네 번째 프레임, 다섯 번째 프레임.
 Fig. 12. Results of 1st, 2nd, 3rd, 4th and 5th frames. From left to right: Result of depth image warping ; Result of color image warping; Result of hole filling result.

표 1은 각 홀 채움 기법들의 수행 속도를 나타내는 표이다. 표 1에서 보는 바와 같이 경계 잡음을 제거하고 기울기를 이용하여 홀을 채우는 본 기법이 기존의 방법들에 비해 빠른 처리속도를 보인다. 기존의 몇몇 방법들에 비해 굴곡진 형태의 홀 채움에는 조금 떨어지는 성능을 보이나, 실시간 처리에 있어서 뛰어난 성

능을 보임을 알 수 있다.

표 1. 성능비교(처리속도)
Table 1. Performance comparison (processing speed)

Method	Computational time(ms)
Linear interpolation	0.3
Inpainting(Telea)	100.4
Inpainting(Criminisi)	4596.7
Spatial weighted average method	12.0
The proposed method	8.0

표 2는 제안하는 시스템의 각 과정 별 수행속도를 측정하여 표이다. 결합형 양방향 필터는 실시간 처리를 위해 GPU를 이용하여 구현하였다. 측정 결과 32fps의 프레임 율을 보여주었으며 제안하는 시스템이 실시간으로 다시점 영상을 생성할 수 있음을 확인할 수 있다. ftp 서버를 통해 제안하는 기법의 결과 영상을 볼 수 있다(ftp://128.134.65.1, ID : guest).

표 2. 과정별 처리 속도
Table 2. Processing speed

Process	Time(ms)
Joint bilateral filtering	11.5
Flickering pixel processing	6.5
3D warping	5.0
Boundary noise removal	2.0
Block based gradient search	6.0
Total time consumption	31.0(32.2fps)

V. 결 론

본 논문에서는 키넥트(Kinect)에서 획득된 깊이 영상과 색상 영상을 이용하여 실시간으로 가상 시점 영상을 생성하는 방법을 제안하였다. 키넥트의 깊이 영상은 카메라의 내외적인 문제로 인하여 홀(hole)이 생기고 경계 주변에 잡음이 생겨 동영상으로 재생할 경우 경계 흔들림(flickering) 현상이 발생한다. 제안한 기법에서는 흔들리는 화소를 찾아 처리함으로써 경계 흔들림 문제를 해결하였다. 하지만 움직이는 객체의 경우에는 흔들리는 화소 검색

과정에서 오류가 발생하여 잔상(afterimage)이 발생한다. 이러한 문제는 획득된 깊이 영상을 이용하여 움직이는 객체의 깊이 값을 보정하였다. 향상된 깊이 영상과 대응되는 색상 영상에 3D 워핑(3D warping) 기법을 적용하여 가상 시점을 생성한다. 가려짐 영역(occlusion region)에 의해 나타나는 홀은 블록 기반 기술기 탐색 과정과 블록의 신뢰도를 조합하여 블록의 최종 신뢰도를 산출한 뒤 블록의 중앙값으로 채운다.

실험을 통해 제안하는 시스템이 영상의 흔들림을 보정해주고 자연스럽게 홀이 채워지며 실시간으로 처리되는 것을 확인하였다. 이러한 결과를 얻을 수 있었던 이유는 제안하는 기법에서는 흔들리는 화소를 보정하여 화면의 흔들림 현상을 줄이고, 후처리 기법에서 홀 채움 하기 전에 경계 잡음을 먼저 제거하여 잘못된 색상 정보를 사용하는 것을 방지하였고, 배경영역과 객체영역을 분리하여 효과적으로 홀을 채울 수 있었으며, 영상의 기술기 정보를 이용하여 영상의 연속성을 유지할 수 있었기 때문인 것으로 판단된다.

References

- [1] Retrevo Corporation, *Could low interest in 3DTV hurt the TV business?*, Retrieved Nov., 2011, from <http://www.retrevo.com/content/node/1915>.
- [2] G. M. Um, G. H. Cheong, W. S. Cheong, and N. H. Hur, "Technical development and standardization trends of multi-view 3D and free-viewpoint video," *IEEK Mag.*, vol. 38, no. 2, pp. 18-23, Feb. 2011.
- [3] T. J. Kim and J. S. Yoo, "Hierarchical stereo matching with color information," *J. KICS*, vol. 34, no. 3, pp. 279-287, Mar. 2009.
- [4] T. Leyvand, C. Meekhof, Y. C. Wei, J. Sun, and B. Guo, "Kinect identity: technology and experience," *Computer (IEEE Comput. Soc.)*, vol. 44, no. 4, pp. 94-96, Apr. 2011.
- [5] Y. S. Park, S. M. Yun, and C. S. Won, "Hole filling for kinect depth image according to the causes of the holes," in *Proc. Conf. Korean Soc. Broadcast Eng. (KOSBE)*, pp. 75-80, Jeju Island, Korea, July 2012.
- [6] C. J. Park, J. H. Ko, and E. S. Kim, "A new intermediate view reconstruction scheme

- based-on stereo image rectification algorithm,” *J. KICS*, vol. 29, no. 5C, pp. 632-641, May. 2004.
- [7] M. S. Ko and J. S. Yoo, “Boundary noise removal and hole filling algorithm for virtual viewpoint image generation,” *J. KICS*, vol. 37, no. 8, pp. 679-688, Aug. 2012.
- [8] L. Zhao and H. Wang, “Image denoising using trivariate shrinkage filter in the wavelet domain and joint bilateral filter in the spatial Domain,” *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2364-2369, Oct. 2009.
- [9] A. Telea, “An image inpainting technique based on the fast marching method,” *J. Graphics Tools*, vol. 9, no. 1, pp. 25-36, Dec. 2004.
- [10] A. Criminisi, P. Perez, and K. Toyama, “Region filling and object removal by exemplar-based image in-painting,” *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200-1212, Sept. 2004.
- [11] G. C. Lee, Y. H. Seo, and J. S. Yoo, “GPGPU-based multiview synthesis using kinect depth image,” in *Proc. Conf. KICS*, pp. 152-154, Yongpyong, Korea, Jan. 2012.
- [12] Y. Mori, N. Fukushima, T. Yendoa, T. Fujii, and M. Tanimotoa, “View generation with 3D warping using depth information for FTV,” *Signal Processing : Image Communication (ELSEVIER)*, vol. 24, no. 1-2, pp. 65-72, Jan. 2009.
- [13] Y. J. Kim, S. H. Lee, and J. I. Park, “A high-quality occlusion filling method using image inpainting,” *J. Korean Soc. Broadcast Eng. (KOSBE)*, vol. 15, no. 1, pp. 3-13, Jan. 2010.
- [14] S. B. Lee and Y. S. Ho, “Real time eye contact system using a kinect depth camera for realistic telepresence,” *J. KICS*, vol. 37, no. 4, pp. 277-282, Apr. 2012.
- [15] W. J. Tam, G. Alain, L. Zhang, T. Martin, and R. Renaud, “Smoothing depth maps for improved stereoscopic image quality,” in *Proc. SPIE Conf. Three-Dimensional TV, Video, Display III*, vol. 5599, pp. 162-172, Philadelphia, U.S.A., Oct. 2004.
- [16] T. Y. Kim, Y. G. Jeon, and J. C. Jeong, “Adaptive linear interpolation using the new distance weight and local patterns,” *J. KICS*, vol. 31, no. 12C, pp. 1184-1193, Dec. 2006.
- [17] G. C. Lee and J. S. Yoo, “Kinect depth map enhancement using boundary flickering compensation,” in *Proc. Conf. Korean Soc. Broadcast Eng. (KOSBE)*, pp. 25-28, Seoul, Korea, Nov. 2012.
- [18] T. J. Kim, E. Y. Chang, N. H. Hur, J. W. Kim, and J. S. Yoo, “Virtual viewpoint image synthesis algorithm using multi-view geometry,” *J. KICS*, vol. 34, no. 12, pp. 1154-1166, Dec. 2009.
- [19] J. H. Park and C. G. Song, “Effective shadow removal from aerial image of golf course to extract components,” *J. Korean Inst. of Inform. Sci. Eng. (KIISE)*, vol. 39, no. 7, pp. 577-582, July 2012
- [20] N. E. Yang, Y. G. Kim, and R. H. Hong, “Depth hole filling using the depth distribution of neighboring regions of depth holes in the kinect sensor,” in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, pp. 658-661, Hong Kong, China, Aug. 2012
- [21] S.-B. Lee and Y.-S. Ho, “Real-time stereo view generation using kinect depth camera,” in *Proc. Asia-Pacific Signal Inform. Process. Assoc. Annu. Summit Conf. (APSIPA ASC) 2011*, pp. 1-4, Xi'an, China, Oct. 2011.

이 규철 (Gyu-cheol Lee)



2013년 2월 광운대학교 전자공학과 학사

2013년 3월~현재 광운대학교 전자공학과 석사과정

<관심분야> 3D 입체영상처리/ 압축, 스테레오 매칭, 영상 신호 처리

유 지 상 (Jisang Yoo)



1985년 2월 서울대학교 전자
공학과 학사

1987년 2월 서울대학교 전자
공학과 석사

1993년 5월 Purdue Univ.
EE, Ph.D.

1997년 9월~현재 광운대학교

전자공학과 교수

<관심분야> 3D 입체영상처리/압축, 웨이블릿 기반
영상처리, 비선형 디지털 신호처리