

깊이 카메라를 이용한 객체 분리 및 고해상도 깊이 맵 생성 방법

강 윤 석*, 호 요 성^o

Foreground Segmentation and High-Resolution Depth Map Generation Using a Time-of-Flight Depth Camera

Yun-Suk Kang*, Yo-Sung Ho^o

요 약

본 논문에서는 색상 카메라와 Time-of-Flight (TOF) 깊이 카메라를 이용해 촬영된 장면에서 전경 영역을 분리하고 영상의 고해상도 깊이 정보를 구하는 방법에 대해 제안한다. 깊이 카메라는 장면의 깊이 정보를 실시간으로 측정할 수 있는 장점이 있지만 잡음과 왜곡이 발생하고 색상 영상과의 상관도도 떨어진다. 따라서 이를 색상 영상과 함께 사용하기 위한 색상 영상의 영역화 및 깊이 카메라 영상의 3차원 투영(warping) 작업, 깊이 경계 영역 탐색 등을 진행한 후, 전경의 객체를 분리하고, 객체와 배경에 대하여 깊이 값 계산한다. 깊이 카메라로부터 얻은 초기 깊이 정보를 이용하여 색상 영상에서 구해진 깊이 맵은 기존의 방법인 스테레오 정합 등의 방법보다 우수한 성능을 나타내었고, 무늬가 없는 영역이나 객체 경계 영역에서도 정확한 깊이 정보를 구할 수 있었다.

Key Words : TOF Depth Camera, Object Extraction, Depth Generation, Sensor Fusion, 3DTV

ABSTRACT

In this paper, we propose a foreground extraction and depth map generation method using a time-of-flight (TOF) depth camera. Although, the TOF depth camera captures the scene's depth information in real-time, it has a built-in noise and distortion. Therefore, we perform several preprocessing steps such as image enhancement, segmentation, and 3D warping, and then use the TOF depth data to generate the depth-discontinuity regions. Then, we extract the foreground object and generate the depth map as of the color image. The experimental results show that the proposed method efficiently generates the depth map even for the object boundary and textureless regions.

I. 서 론

두 대 이상의 카메라를 사용하여 장면인 3차원 영상을 촬영하고 편집하여 좌안과 우안으로 각각 시차를 가지는 영상을 볼 때에, 시차의 정도에 따라서 시청자는 장면의 입체감을 느낄 수 있다. 양안식 영상보다 더 많은 시점을 가진 다시점 영상

(multi-view image)의 경우, 동일한 장면에 대하여 세 개 이상의 시점을 가지고 있으며, 다시점 디스플레이를 통해서 시청하였을 때에, 시청위치에 따라 시차를 가지는 두 영상을 보게 됨으로써 입체감을 느낄 수 있다^[1].

이와 같이 한 장면에 대하여 두 개 이상의 시점을 가지도록 하기 위해서는 여러 대의 카메라를 이용한

* 주저자 : 광주과학기술원 정보통신공학부, yunsuk@gist.ac.kr, 정회원

^o 교신저자 : 광주과학기술원 정보통신공학부, hoyo@gist.ac.kr, 종신회원

논문번호 : KICS2012-07-303, 접수일자 : 2012년 7월 4일, 최종논문접수일자 : 2012년 9월 5일

촬영이 필요하다. 하지만 한꺼번에 촬영하고 처리할 수 있는 카메라 개수는 제한되어 있고, 배치 간격을 조밀하게 하는 것도 한계가 있다. 따라서, 제한된 수의 카메라로 촬영된 영상으로부터 그 사이의 가상의 시점에 해당하는 영상을 생성하면 효과적인 입체 영상을 제작할 수 있다. 가상 시점 영상의 생성을 위해서는 해당하는 장면의 색상 영상에 대한 정확한 깊이 정보가 반드시 필요하다.

장면의 깊이 정보를 구하는 가장 일반적인 방법은 두 대의 카메라에서 촬영된 양안식 색상 영상만을 이용하는 스테레오 정합(stereo matching) 방법이다²⁾. 스테레오 정합은 색상 영상에서 각 화소에 해당하는 변위(disparity) 정보를 구하는 방법으로, 색상 영상만 가지고도 장면의 깊이를 구할 수 있는 장점이 있으나, 객체에 의해 가려져있는 영역, 텍스처가 없는 영역 등에서는 깊이 정보를 구하기가 매우 어렵다. 한편, 장면 내 물체들의 거리를 장비를 이용해 직접 측정할 수도 있다. Time-of-Flight (TOF) 기술을 이용하는 깊이 카메라가 대표적이다. 깊이 카메라는 적외선 혹은 광신호를 장면에 방사하고, 그 신호가 물체에 반사되어 돌아오는 위상차를 이용하여 거리를 측정하여 깊이 영상으로 출력한다. 이와 같은 방식은 실시간으로 장면의 깊이를 얻을 수 있지만, 출력영상의 낮은 해상도와 영상 잡음, 왜곡 등 해결해야 할 문제가 많다. 최근에는 융합 카메라 시스템이라는 이름으로 스테레오 정합에 깊이 카메라를 함께 이용하는 방법들이 많이 제안되었다³⁻⁶⁾. 하지만 이 방법들 역시 스테레오 정합이 여전히 추가 되는 방법으로, 결과를 얻기까지 많은 시간이 소요되며, 스테레오 정합의 단점들을 완벽하게 보완할 수 없다.

본 논문에서는 깊이 카메라를 이용하여 함께 촬영되는 색상 영상에서 전경 객체를 자동으로 분리하고, 알맞은 깊이 값을 찾아 깊이 맵을 생성하는 방법을 제안한다. 제안하는 방법은 스테레오 정합을 사용하지 않고, 깊이 카메라에서 촬영된 영상과 단일 시점의 색상 영상을 이용하여 전경 객체를 분리하고, 배경과 분리한 전경에 대해 각각 깊이 카메라 정보를 이용하여 깊이 값을 채운다.

II. TOF 깊이 카메라의 특성과 깊이 카메라를 이용한 고해상도 깊이 맵 제작 기술

TOF 깊이 카메라는 카메라의 센서로부터 발생된 광신호가 다시 카메라로 돌아오는 위상의 차이를 이용하여 장면의 거리정보를 획득한다. 본 논문의

실험에서 사용한 깊이 카메라 모델은 Mesa Imaging에서 개발한 SR4000이라는 모델로 그림 1과 같이 광신호 발생부분과 그 신호를 받아들이는 광필터로 이루어져 있다. 촬영할 수 있는 최소, 최대 거리는 각각 0.3m, 5.0m이며, 초당 최대 54프레임을 촬영할 수 있다. SR4000으로 촬영된 영상은 그림 1에 나타난 것과 같이, 장면의 거리정보를 나타내는 8bit의 깊이 영상(depth image)과 장면의 흑백 영상(intensity image) 영상 두 종류를 출력하며, 176x144의 해상도로 촬영된다.

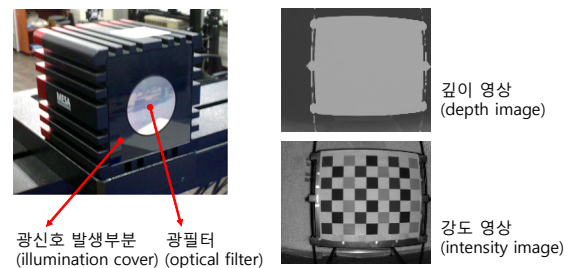


그림 1. 깊이 카메라와 촬영된 영상
Fig. 1. Depth camera and captured image

깊이 카메라는 일반적으로 해상도가 매우 낮기 때문에, 고해상도의 색상 영상과 함께 사용하여 고해상도의 깊이 맵을 획득하는데 사용된다. 고해상도의 정확한 깊이 맵을 얻기 위해서는 양안식 카메라와 한 대의 깊이 카메라를 사용하여 깊이 카메라 정보를 색상의 위치로 이동시킨 후 스테레오 정합의 초기 변위 값으로 설정하는 방법 [3, 4], 깊이 카메라 정보를 스테레오 정합의 에너지 함수 최적화를 위한 하나의 항으로 정의하여 사용하는 방법 [5] 등이 있으며, 고화질 색상 영상과 함께 사용하는 방법 [6], 그 외에 다양한 융합 형태로 두 종류의 카메라가 함께 사용되고 있다. 그러나 스테레오 정합의 복잡도가 여전히 높고, 객체의 정확한 경계 영역을 구하는 등에 있어서 한계를 가지고 있다.

III. 제안하는 깊이 맵 생성 방법

본 장에서는 제안하는 깊이 맵 생성 방법에 대해 설명한다. 그림 2는 제안하는 방법의 흐름도를 보여 준다. 장면을 촬영한 후 깊이 영상에 존재하는 왜곡을 보정하고, 각 카메라의 카메라 변수를 구한다. 그 후 깊이 카메라 영상 정보를 이용하여 깊이 불연속 맵을 생성하고, 이 정보를 이용하여 객체를 분리한 후, 깊이 카메라 영상의 각 화소 정보를 이용

하여 최종 깊이 맵을 제작한다.

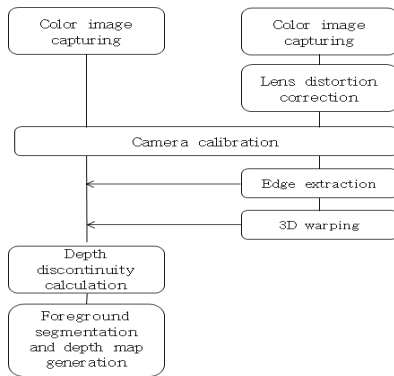


그림 2. 제안하는 깊이 맵 생성 방법
Fig. 2. Proposed depth map generation method

3.1. 영상 촬영 및 전처리

제안하는 방법의 실험을 위해서 한 대의 색상 카메라와 한 대의 깊이 카메라를 사용한다. 깊이 카메라는 색상 카메라의 아래쪽에 배치하였다. 그림 1(b)에서 보듯이, 깊이 카메라로 촬영된 영상에는 원형 렌즈 왜곡이 발생하기 때문에 촬영된 영상에 대하여 렌즈 왜곡 보정을 먼저 수행하고 [7] 각각의 카메라에 대해 카메라 변수를 계산한다. 카메라 변수는 카메라의 물리적 특성과 상대적인 방향 및 위치정보를 나타내는 변수들로써 카메라 보정(camera calibration) 과정을 통해 구할 수 있다^[8]. 초점 거리 등 카메라의 내부 특성을 나타내는 내부(intrinsic) 행렬과 3차원 공간에서의 방향을 나타내는 방향(orientation) 행렬, 그리고 위치를 나타내는 이동(translation) 벡터로 구성되며, 이 변수들을 이용하여 카메라 행렬(projection matrix)을 생성할 수 있다. 카메라 행렬은 깊이 영상의 3차원 투영(warping)시에 매우 중요한 역할을 한다.

3.2. 깊이 불연속 영역 계산 및 자동 객체 분리

장면의 깊이 맵을 계산할 때 가장 어려운 부분 중 하나는 깊이 불연속 구간에 대한 정확한 깊이 정보를 구하는 것이다. 일반적으로 전경 객체와 객체 간, 혹은 객체와 배경 간에는 깊이의 차이가 존재하기 때문에 객체에 의해 가려져 있다가 새롭게 드러나는 영역이 발생한다. 또한, 색상 영상에서의 객체 경계와 일치하지 않는 깊이 맵의 경우, 가상 시점을 생성했을 때 화질을 저하시키는 요인이 된다. 따라서 깊이 불연속 영역에서 정확한 깊이 정보를 구하는 것은 매우 중요하다.

제안하는 방법에서는 깊이 카메라 정보를 이용하여 색상 영상에서 배경과 객체를 미리 분리하고, 배경과 전경에 대하여 각각 깊이 정보를 구하고, 하나의 깊이

맵을 생성한다. 전경 객체의 자동 분리를 위해서는 먼저 깊이 불연속 영역을 계산해야 한다. 그림 3은 깊이 카메라로 촬영된 영상과 Canny 방법을 이용해 검출된 에지를 보여준다. 깊이 카메라 영상은 객체의 경계부분에서만 에지가 검출되기 때문에 경계 영역을 비교적 쉽게 찾아낼 수 있다.



그림 3. 깊이 카메라 영상과 에지 정보
Fig. 3. Depth camera image and edge information

깊이 카메라 영상에서 에지에 해당하는 영역들은 3차원 투영을 통해 색상 영상의 위치로 이동된다. 3차원 투영은 영상의 각 화소에 대하여 화소가 가지는 깊이 값과 카메라 변수를 이용하여 인접한 카메라의 위치로 이동하는 작업이다. 그림 4에 나타난 것과 같이, 깊이 카메라의 한 화소에 대하여 카메라 변수와 깊이 값을 이용하여 화소의 값을 3차원 공간으로 역투영시키고, 이 화소 값을 다시 3차원 공간으로부터 색상 카메라로 투영하게 된다.

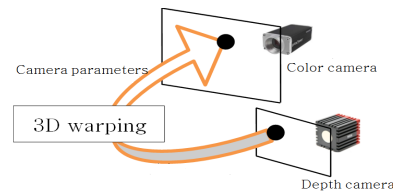


그림 4. 깊이 영상의 3차원 투영
Fig. 4. 3D warping of depth image

이 때, 이 화소들은 그림 3의 에지 영상에서처럼 연결된 선의 형태를 이루는 것이 아니라 몇 화소씩 오차를 두고 색상 영상의 위치에 투영되는데, 이는 카메라 변수에서 오는 오차, 깊이 값이 가지는 오차, 해상도 차이 등으로 인해 발생하는 것이다. 따라서 이러한 화소들을 바탕으로 색상 영상에 상응하는 깊이 불연속 영역을 계산하기 위해서는 화소들의 팽창 및 침식 연산을 이용한다. 해당 화소들에 대하여 영역이 만들어질 때까지 팽창시킨 후, 영역이 너무 넓어지는 것을 방지하기 위하여 침식 연산을 수행하면, 그림 5와 같은 결과를 얻게 된다. 그림 5(a)는 생성된 깊이 불연속 영역을 나타내며 그림 5(b)는 색상 영상과 대비하여 보았을 때, 깊이 불연속 영역이 올바르게 찾아진 것을 나타낸다.

이제 그림 5(a)와 같이 구한 깊이 불연속 영역을 이

용하여 전경의 객체를 분리해 낼 수 있다. Grab-cut 영역화 방법은 영상에서 분리해내고자 하는 객체에 대한 초기 정보를 사용자가 입력해주면 해당하는 범위 내에서 객체를 정확하게 분리해주는 알고리즘이다^[9]. 초기 정보는 여러 가지 형태로 입력될 수 있는데, 제안하는 방법에서는 사각형 형태로 입력하는 것은 전제로 하여, 그림 5(a)의 영상의 깊이 불연속 영역을 모두 포함하는 최소의 사각형을 그림 6(a)와 같이 찾고, 그 위치 정보를 Grab-cut 영역화 알고리즘의 초기 입력 값으로 사용한다. 그러면 그림 6(b)와 같이 전경 객체가 정확히 분리된 결과를 획득할 수 있다. 일부 영역에서 잡음이 발생할 수 있으나 중간 값 필터를 이용하여 간단히 제거할 수 있다.



그림 5. 깊이 불연속 영역 생성
Fig. 5. Depth discontinuity region generation

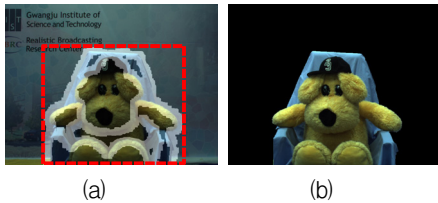


그림 6. Grab-cut 영역화를 이용한 전경 객체 분리
Fig. 6. Foreground object segmentation using Grab-cut

3.3 전경과 배경의 깊이 계산을 통한 깊이 맵 생성

그림 6과 같이 전경의 객체를 분리해 낸 후에는 깊이 카메라 정보를 이용해서 전경과 배경에 해당하는 깊이 값을 각각 계산해야 한다. 그림 7은 깊이 카메라에서 촬영된 깊이 영상이 색상 영상의 위치로 3차원 투영을 통해 이동된 것을 보여준다. 해상도 차이에 의해서 몇 화소씩 건너서 값이 존재한다. 따라서 값이 존재하지 않는 화소에 대하여 주변 값들을 이용해서 정확한 깊이 값을 찾아야 한다.

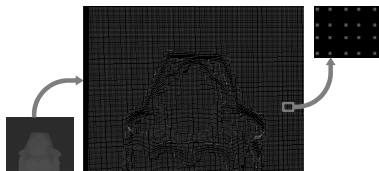


그림 7. 깊이 카메라 영상의 3차원 투영
Fig. 7. 3D warping of depth camera image

이제 그림 6(b)와 같이 전경과 배경 부분을 분리하고, 각각에 대하여 깊이 맵을 생성한다. 깊이 맵 생성에는 그림 7에서와 같이 3차원 투영된 깊이 카메라의 각 화소 값을 이용한다. 전경과 배경에 깊이 값이 존재하지 않는 화소에 대해서는 주변의 값들 $d(x+i, y+j)$ 을 이용하여 현재 화소의 값 $d(x, y)$ 를 계산하여 식 (1) 과 같이 채우게 되며, 전경은 배경의 화소를, 배경은 전경의 화소를 참조하지 않는다.

$$d(x, y) = \frac{1}{n} \sum_{i=-k}^k \sum_{j=-k}^k d(x+i, y+j) \quad (1)$$

여기서 n 은 값이 0이 아닌 화소 중 더해진 화소 수이다. 전경의 경우에 $d(x+i, y+j)$ 는 다음과 같이 정의된다. $c(x, y)$ 는 해당 위치에서의 색상 값이다. 변수 k 는 계산을 위한 블록의 크기를 나타낸다.

$$d(x+i, y+j) = \begin{cases} 0 & \text{if } |c(x, y) - c(x+i, y+j)| > th \\ d(x+i, y+j) & \text{otherwise} \end{cases} \quad (2)$$

즉, 참조하려는 화소의 색상 값이 비슷한 부분의 화소만을 참조하여 깊이 값이 다른 곳의 영향을 받는 것을 방지한다.

IV. 실험 결과 및 분석

제안한 방법의 실험을 위하여 그림 8과 같이 세 종류의 영상을 촬영하여 실험에 사용하였다. 색상 영상 및 깊이 카메라 영상의 해상도는 각각 1280x960과 176x144 이다. 실험에는 식 (1)의 $k=3$ 을 사용하였고, 해당 블록에 값이 없을 경우 k 를 증가시켰다. 그림 8(a)에 대한 중간 결과들은 그림 5와 그림 6에 각각 나타나있고, 그림 8(b)과 그림 8(c)에 대한 중간 결과인 깊이 불연속 영역과 분리된 전경의 결과가 그림 9(a)와 그림 9(b)에 각각 나타나 있다.

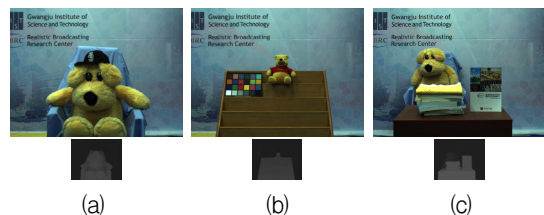


그림 8. 실험을 위해 촬영된 색상 및 깊이 카메라 영상
Fig. 8. Color and depth camera image for experiment

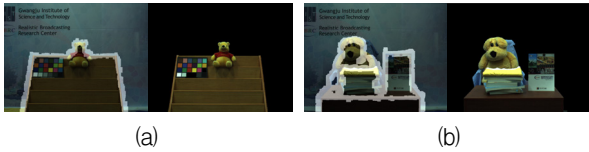


그림 9. 두 번째와 세 번째 영상에 대한 중간 결과
Fig. 9. Intermediate result for second and third images

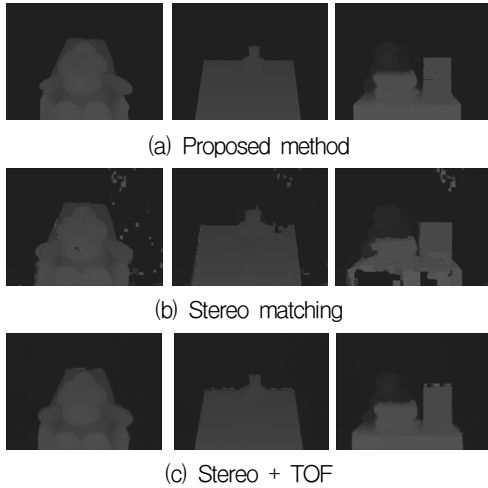


그림 10. 생성된 깊이 맵 결과
Fig. 10. Result of depth map generation

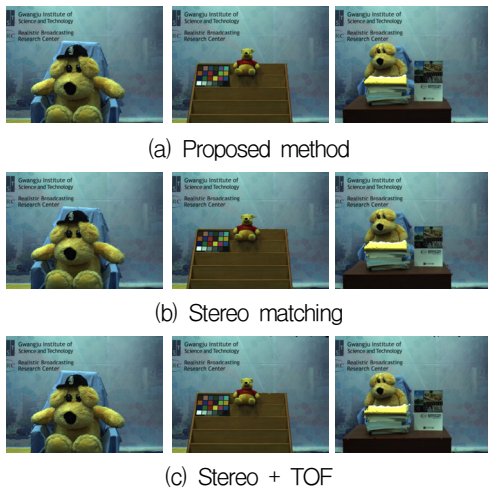


그림 11. 생성된 가상 시점 영상
Fig. 11. Virtual view image generation

이러한 과정을 거쳐 제안한 방법으로 구해진 깊이 맵이 그림 10(a)에 나타나 있다. 전경과 배경이 확실하게 분리되어 있으며, 전경 내 영역에서도 안정적인 깊이 정보를 생성함을 볼 수 있다. 그림 10(b)는 실험 영상과 우측 시점의 영상을 이용하여 신뢰 확산을 이용하는 스테레오 정합을 수행한 결과를 보여준다¹⁰⁾. 그림 10(c)는 역시 신뢰 확산을 이용한 스테레오 정합과 깊이 카메라의 깊이 정보를 함께 이용하여 구한 깊이 맵을 보여준다. 이 방법에서는 깊이 카메라의 정보

로 스테레오 정합을 위한 초기 변위 값을 설정하였다.

그림 10(b)와 그림 10(c)에 보인 것처럼, 기존의 방법들은 무늬가 없거나 약한 부분에서 잘못된 깊이 값을 가지며, 객체 경계부에서 안정적이지 못한 값을 가진다. 그림 10에 나타난 깊이 맵을 이용하여 우측 시점에 가상 시점 영상을 생성한 결과가 그림 11에 나타나 있다. 우측 시점 영상과의 비교를 위해 65mm 떨어진 곳에 있는 동일한 모델의 색상 카메라를 이용하여 그림 8에 나타난 세 종류의 영상에 대한 우측 시점의 영상을 촬영하였다. 그리고 좌영상과 함께 영상 정렬을 수행하여, 가상 시점의 화질 평가를 위한 원본 영상으로 사용하였다¹¹⁾. 생성된 가상 시점 영상에서 발생한 홀 영역은 인페인팅(impainting) 기법을 이용하여 채웠다¹²⁾. 표 1에 나타난 결과에서 볼 수 있듯이, 제안하는 방법은 다른 방법들에 비해 대체적으로 우수한 성능을 보이며 평균 수행시간도 매우 빠른 것을 볼 수 있다.

표 1. 가상 시점 합성 품질 및 수행 시간
Table. 1. Virtual view quality and running time

	PSNR (dB)		
	Proposed method	Stereo matching	Stereo + TOF
Image 1	33.47	34.18	31.55
Image 2	35.65	35.12	35.53
Image 3	32.69	31.43	32.39
Running time (초)	39.96	139.57	302.63

V. 결론

본 논문에서는 색상 카메라와 TOF 깊이 카메라를 이용하여 장면의 깊이 맵을 생성하는 방법을 소개하였다. 촬영된 영상에서의 잡음과 왜곡을 보정한 후, 깊이 영상은 색상 영상의 위치로 3차원 투영을 통해 이동되어 깊이 불연속 영역을 결정하고, 이를 통해 전경과 배경을 분리한 후, 깊이 카메라 정보를 이용하여 최종 깊이 맵을 생성하게 된다. 제안하는 방법은 복잡도가 높은 스테레오 정합에 비해 간결하고, 계산된 값이 아닌 직접 측정된 깊이 값을 기반으로 하여 무늬가 없는 영역과 같은 곳에서도 정확한 깊이 값을 빠른 시간에 생성하였다. 또한 전경과 배경이 분리되어 처리되므로 객체 경계부의 깊이 값이 깨끗하게 생성되었음을 볼 수 있다.

참 고 문 헌

[1] C. Fehn, R. Barre, and S. Pastoor, "Interactive 3-DTV- Concepts and Key Technologies," *Proc. of the IEEE*, vol. 94, no. 3, pp. 524-538, Mar. 2006.

[2] J. Sun, N.N. Zheng, and H.Y. Shum, "Stereo Matching Using Belief Propagation," *IEEE Trans. of Pattern Analysis and Machine Intelligence (PAMI)*, vol. 25, no. 5, pp. 787-800, Jul. 2003.

[3] S. A. Gudmundsson, H. Aanaes, and R. Larsen, "Fusion of Stereo Vision and Time-of-Flight Imaging for Improved 3D Estimation," *International Journal of Intelligent Sys. Technologies and Applications*, vol. 5, no. 3, pp. 425-433, Nov. 2008.

[4] B. Bartczak and R. Koch, "Dense Depth Maps from Low Resolution Time-of-Flight Depth and High Resolution Color Views," *Proc. of 5th International Symposium on Visual Computing*, pp. 1-12, Nov. 2009.

[5] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of Time-of-Flight Depth and Stereo for High Accuracy Depth Maps," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 231-236, June 2008.

[6] K. D. Kuhnert and M. Stommel, "Fusion of Stereo-Camera and PMD-Camera Data for Real-time Suited Precise 3D Environment Reconstruction," *Proc. of the IEEE/RSJ International Conf. on Intelligent Robots and Syst. (IROS)*, pp. 4780-4785, Mar. 2006.

[7] A. Wang, T. Qiu, and L. Shao, "A Simple Method of Radial Distortion Correction with Centre of Distortion Estimation," *Journal of Mathematical Imaging and Vision*, vol. 35, no. 3, pp. 165-172, Nov. 2009.

[8] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, Oct. 2000.

[9] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut - Interactive Foreground Extraction

using Iterated Graph Cuts," *ACM Trans. on Graphics*, vol. 23, no. 1, pp. 309-314, Jan. 2004.

[10] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Belief Propagation for Early Vision," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41-54, Oct. 2006.

[11] Y. S. Kang and Y. S. Ho, "Multi-view Image Enhancement by Compensating Geometric Error," *Telecomm. Review*, vol. 20, no. 3, pp. 507-518, 2010.

[12] A. Telea, "An Image Inpainting Technique based on The Fast Marching Method," *Journal Graphics Tools*, vol. 9, pp. 25-36, May 2004.

강 윤 석 (Yun-Suk Kang)



2007년 한국항공대학교 전자공학과 졸업(학사)
 2008년 광주과학기술원 정보통신공학부 졸업(석사)
 2008년~현재 광주과학기술원 정보통신공학부 박사과정
 <관심분야> 디지털 영상처리, 다시점 영상 획득 및 처리, 3차원 TV, 실감방송

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