+ 호요성 · 광주과학기술원(GIST) 교수

TECH&TREND

3차원 영상 서비스와 기술 개발 동향(3)

1. 서론

2009년 영화 '아바타'가 흥행에 크게 성공하면서 3차원 영화에 대한 관심이 폭발적으로 증가하게 되었고, 이에 따른 파급 효과로 3차원 TV 수상기가 가정에까지 보급되었다 [1]. 3차원 영상 산업은 예전처럼 반짝인기에 그칠 것이라 비관적으로 예측했던 전문가들도 현재의 추세를 봤을 때 3차원 영상 산업의 열기는 당분간 계속 이어질 것으로 전망하고 있다.
시청자에게 3차원 효과를 주기 위한 기본적인 방법은 양안식(stereoscopic)카메라 혹은 컴퓨터 그래픽스 기술을 이용하여 얻은 좌우 영상을 시청자의 좌우 눈에 각각 분리해서 보여주는 것이다 [2]. 하지만, 이보다 앞선 기술로 3시점 혹은 그 이상의 넓은 시야각의 고해상도 3차원 영상을 시청자에게 제공하는 방법을 최근 MPEG 표준화 회의에서 논의하고 있고, 많은 연구 기관에서도 활발히 연구하고 있다 [3]–[5].

사용자가 원하는 임의 시점의 3차원 영상을 만들기 위해서는 다수의 카메라를 이용하여 획득한 넓은 시야각을 가지는 다시점 영상뿐만 아니라, 3차원 장면의 거리 정보를 표현하는 깊이 영상을 구해야 한다 [6][7]. 특히, 3차원 영상을 다양한 응용 분야에서 사용하기 위해서는 양안식 또는 다시점 영상의 깊이 정보를 획득하는 기술이 필수적이다. 3차원 정보를 획득하는 방법은 크게 능동적 깊이 센서 방식(active depth sensors)과 수동적 깊이 센서 방식(passive depth sensors)으로 나눌 수 있다 [8]. 능동적 깊이 센서 방식은 레이저 센서, 적외선 센서 등과 같은 센서 장치를 이용해서 3차원 공간상의 깊이 정보를 직접 획득하는 방식이다 [9]. 능동적 깊이 센서 방식은 깊이 정보를 실시간으로 획득할 수 있지만, 현재 하드웨어의 한계 탓에 해상도가 낮고, 센서의 비용이 상용화에는 무리가 있다 [10]. 반면에, 수동적 깊이 센서 방식은 2대 이상의 카메라로부터 얻은 영상의 상관관계를 이용해서 깊이 정보를 계산해서 획득하는 방식이다 [11]. 수동적 깊이 센서 방식은 깊이 정보를 획득하기 위해 시간이 오래 걸리고 상대적으로 정확도가 떨어진다는 단점이 있지만, 더욱 넓은 시야각을 제공할 수 있고 하드웨어 구성을 위한 비용이 낮다는 장점을 지니고 있다 [12].

다시점 3차원 영상의 가장 큰 장점은 시청자가 원하는 임의의 시점에서의 자유 시점 (free-viewpoint) 영상을 생성하여 제공할 수 있다는 것이다 [7]. 이상적으로 자유 시점 기능을 구현하기 위해서는 사용자가 원하는 위치의 모든 시점에 대한 정보를 가지고 있어야 한다. 하지만 카메라의 부피, 비용 등의 현실적인 문제가 존재하기 때문에 한정된 수의 카메라를 이용하여 모든 시점에서의 영상을 획득하는 데에는 한계가 있다. 이러한 문제점은 제한된 수의 다시점 영상과 깊이 영상을 이용해서 임의의 가상 시점의 영상을 합성하면 해결할 수 있다.

본 논문에서는 사용자의 조절에 의한 입체영상 재현이나 편안한 입체영상을 제공하기 위해 임의의 가상 시점의 영상을 생성하는 3차원 TV 기술에 필수 데이터인 3차원 깊이 정보를 획득하는 기술과 이를 이용한 가상시점 영상 생성 기술을 소개한다.

2. 3차원 깊이 정보 획득 기술

(1) 깊이와 변위의 상관관계

3차원 장면을 찍기 위해 두 대의 카메라가 있다고 가정하자. 또한, 이 두 대의 카메라는 정확하게 수평이 맞도록 정렬되어 있고, 두 카메라의 광축이 정확하게 평행을 이룬다고 가정한다. 이러한 카메라 구성에서 멀리 있는 물체는 좌영상과 우영상에서 시차가 작게 발생하지만, 가까이 있는 물체는 시차가 크게 나타난다. 이러한 기본적인 원리를 이용해서 기준 시점의 모든 화소가 참조 시점의 어느 위치에 존재하는지를 탐색하면 각 화소의 변위(disparity)를 얻을 수 있고, 이 변위 값을 이용하면 실제 깊이 정보를 계산할 수 있다. 그림 1은 변위 계산의 기본 원리를 도식화한 것이다.

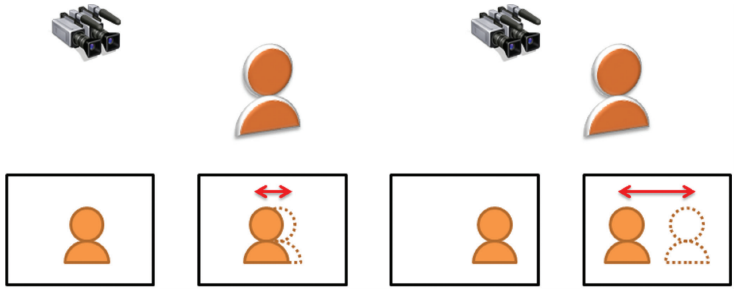


그림 1. 변위 계산의 기본 원리

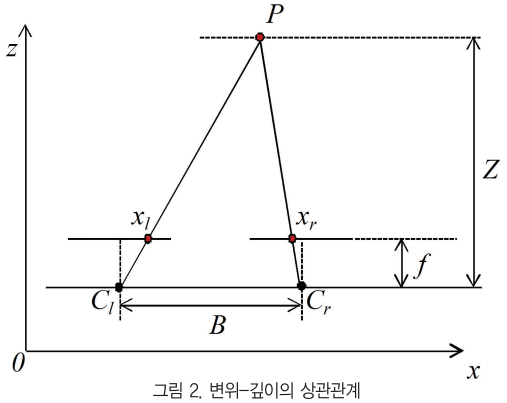


그림 2. 변위-깊이의 상관관계

그림 2는 변위와 깊이 정보의 상관관계를 보다 자세히 나타낸 것이다. 우선, 두 대의 카메라가 각각 C_l, C_r에 위치하고, 두 카메라의 광축이 z축 방향으로 평행하다고 가정한다. 이와 같은 환경에서 3차원 공간상에서의 한 점 P는 좌측 카메라의 영상 평면에서 (x_l,y)에 사상(mapping)되고, 우측 카메라의 영상 평면에서 (x_r,y)에 사상된다. 그렇다면, x_l과 x_r의 차이는 변위 d가 되고, 다음 비례식을 통해 실제 깊이 정보 Z를 쉽게 계산할 수 있게 된다.

$$Z = \frac{Bf}{x_l - x_r} = \frac{Bf}{d} \quad (1)$$

여기서 B는 두 카메라의 거리, f는 카메라의 초점 거리를 나타낸다.

(2) 스테레오 정합 기술

인간 시각 체계(human visual system)는 양쪽 눈으로부터 얻은 영상의 변위차로 깊이 정보를 추정하는 것으로 알려졌다. 스테레오 정합(stereo matching) 기술은 인간 시각 체계를 컴퓨터 시뮬레이션에 접목한 예로 컴퓨터 비전 분야에서 활발히 연구되고 있다 [13]. 일반적으로 스테레오 정합 기술은 다음과 같이 크게 네 단계로 분류된다.

- 정합 오차 계산 (matching cost computation)
- 오차 종합 (cost (support) aggregation)
- 변위 계산 / 최적화 (disparity computation / optimization)
- 변위 정제 (disparity refinement)

각 화소의 변위를 계산하기 위해서는 정합 함수를 이용하여 정합 오차를 계산하며, 정합 오차가 가장 작은 값을 가질 때의 변위를 취하도록 한다. 정합 함수는 두 가지 가정을 포함하고 있는데, 첫 번째는 두 시점에서 변위만큼 떨어진 화소는 서로 유사하다는 것이며, 두 번째는 주변의 화소들은 비슷한 변위를 가진다는 것이다. 이 두 가지 가정은 정합 함수에서 데이터 항(data term)과 평활화 항(smoothness term)으로 표현된다. 따라서 정합 함수는 다음과 같이 정의될 수 있다.

$$D(x,y) = \arg\min_{d}\{E(x,y,d)\} \tag{2}$$

$$E(x,y,d) = E_{data}(x,y,d) + E_{smooth}(x,y,d) \tag{3}$$

여기서 $D(x,y)$는 화소 (x,y)에서의 변위를 말하고, 정합 함수 $E(x,y,d)$는 데이터 항 $E_{data}(x,y,d)$와 평활화 항 $E_{smooth}(x,y,d)$로 구성된다. 데이터 항과 평활화 항은 다음과 같이 정의될 수 있다.

$$E_{data}(x,y,d) = U\{I_L(x,y), I_R(x-d,y)\} \tag{4}$$

$$E_{smooth}(x,y,d) = \sum_{(p,q)\in N} W(p,q) \tag{5}$$

여기서 $U\{\cdot\}$는 좌영상과 우영상의 화소값 차이를 나타내는데, 절대차(absolute difference) 혹은 제곱차(squared difference)를 가장 많이 사용한다. 또한, N은 주변 화소의 집합을 말하며 $W(p,q)$는 주변 화소들의 변위와 차이를 나타낸다.

스테레오 정합 방법은 변위 계산 방법에 따라 크게 지역적인 것과 전역적인 방법으로 나눌 수 있다. 일반적으로 지역적 방법은 계산 속도가 빠르지만, 화소별로 각각 대응점을 찾기 때문에 폐색(occlusion)영역, 반복되는 텍스쳐 혹은 텍스쳐가 없는 영역, Non-Lambertian 평면 등 정보가 모호한 영역에서는 성능이 저하되는 단점이 있다. 전역적 방법은 주변의 정보를 같이 활용하여 영상 전체의 오차를 최소화하기 때문에 보다 정확한 결과를 얻을 수 있지만 계산 시간은 더 많이 소요된다. 대표적인 전역적 방법으로는 마르코프 랜덤 필드(Markov Random Field, MRF)모형을 들 수 있는데 이 방법은 입력 영상을 이용하여 필드를 구성하고 최소 에너지를 구하여 변위를 계산한다 [14]. 따라서, 필드를 구성하는 방법과 최소 에너지 함수를 정의하는 것이 알고리즘의 성능을 좌우한다. 최근에는 전역적 방법이 지역적 방법에 비해 상대적으로 높은 성능을 보이면서 많은 연구들이 이 접근 방법을 따르고 있다 [15]–[18].

그림 3은 스테레오 정합 기술을 통해 획득한 변위 영상을 나타내고 있다. 그림 3에서 얻은 변위 영상은 전역적 방법의 하나인 신뢰 전파(belief propagation)기술을 이용해서 얻은 결과이다 [17]. 그림 3(c)와 그림 3(d)에서 알 수 있듯이, 스테레오 정합 기술을 이용하면 객체의 깊이 정보를 효과적으로 얻을 수 있다.

(a) 좌시점 색상 영상

(b) 우시점 색상 영상

(c) 좌시점 변위 영상

(d) 우시점 변위 영상

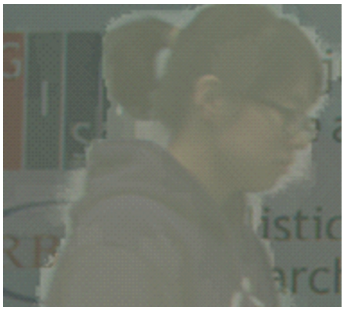
그림 3. 스테레오 정합 기술을 통해 획득한 변위 영상

(3) 깊이 정보 후처리 기술

스테레오 정합 기술을 이용하면 3차원 장면의 깊이 정보를 획득할 수 있지만, 그림 4에서 알 수 있듯이 색상 영상과 변위 영상의 객체 경계가 맞지 않는 문제점이 발생한다. 이러한 문제점은 전역적 방법을 사용했을 때 자주 발생하는 것으로 영상을 합성하는 과정에서 오차를 발생시킬 뿐만 아니라 다양한 3차원 응용 분야에 깊이 정보를 사용할 수 없게 한다.

(a) 색상 영상

(b) 변위 영상



(a) 오버레이 영상

그림 4. 색상 영상과 변위 영상의 경계 불일치

최근 이와 같은 문제점을 해결하기 위해서 획득한 변위 영상에 대해 후처리 필터링을 적용하는 방법들이 제안됐다[19][20]. 후처리 필터링은 기본적으로 양방향 필터(bilateral filter)로부터 출발한다. 기존의 양방향 필터는 한 장의 색상 영상에 적용되는 저역 필터로써, 인접 화소와의 거리차, 색상차 각각에 대한 두 개의 가우시안(Gaussian)함수를 사용한다. 반면에, 새롭게 제안된 결합형 양방향 필터(joint bilateral filter)에서는 두 개의 가우시안 함수에 추가로 인접 화소와의 변위차에 대한 항을 사용한다.

그림 5는 변위 영상에 후처리 필터링을 적용한 결과를 나타낸다. 그림 5(b)와 그림 5(d)에서 알 수 있듯이, 필터링을 적용하기 전 변위 영상에 비해 객체의 경계가 상당 부분 일치하는 것을 확인할 수 있다.

(a) 변위 영상 (b) 후처리된 변위 영상

(c) 필터링 전 오버레이 영상 (d) 필터링 후 오버레이 영상

그림 5. 변위 영상의 후처리 결과

3. 가상 시점 생성 기술

앞서 설명한 깊이 정보를 이용하여 사용자가 원하는 가상의 임의 시점에의 영상을 합성해야 한다. 가상 시점의 영상을 합성하는 기술은 그림 6과 같이 크게 3단계로 나눌 수 있다. 첫 번째 단계는 시점 이동 과정이다. 이 과정에서는 변위값을 이용해서 직접적으로 시점을 좌측 혹은 우측으로 이동한다. 시점 이동으로 인해 참조 시점에서 존재하지 않았던 영역은 빈 영역(hole)으로 나타나게 된다. 빈 영역은 좌, 우 참조 화면으로부터 가상시점에 이동된 두 영상을 하나로 합치는 영상 통합 과정을 통해 대부분 채워진다. 마지막 세 번째 단계는 영상 통합 과정 이후에도 남아 있는 빈 영역을 영상 보간법이나 인페인팅(inpainting)기술을 통해 채우는 과정이다.

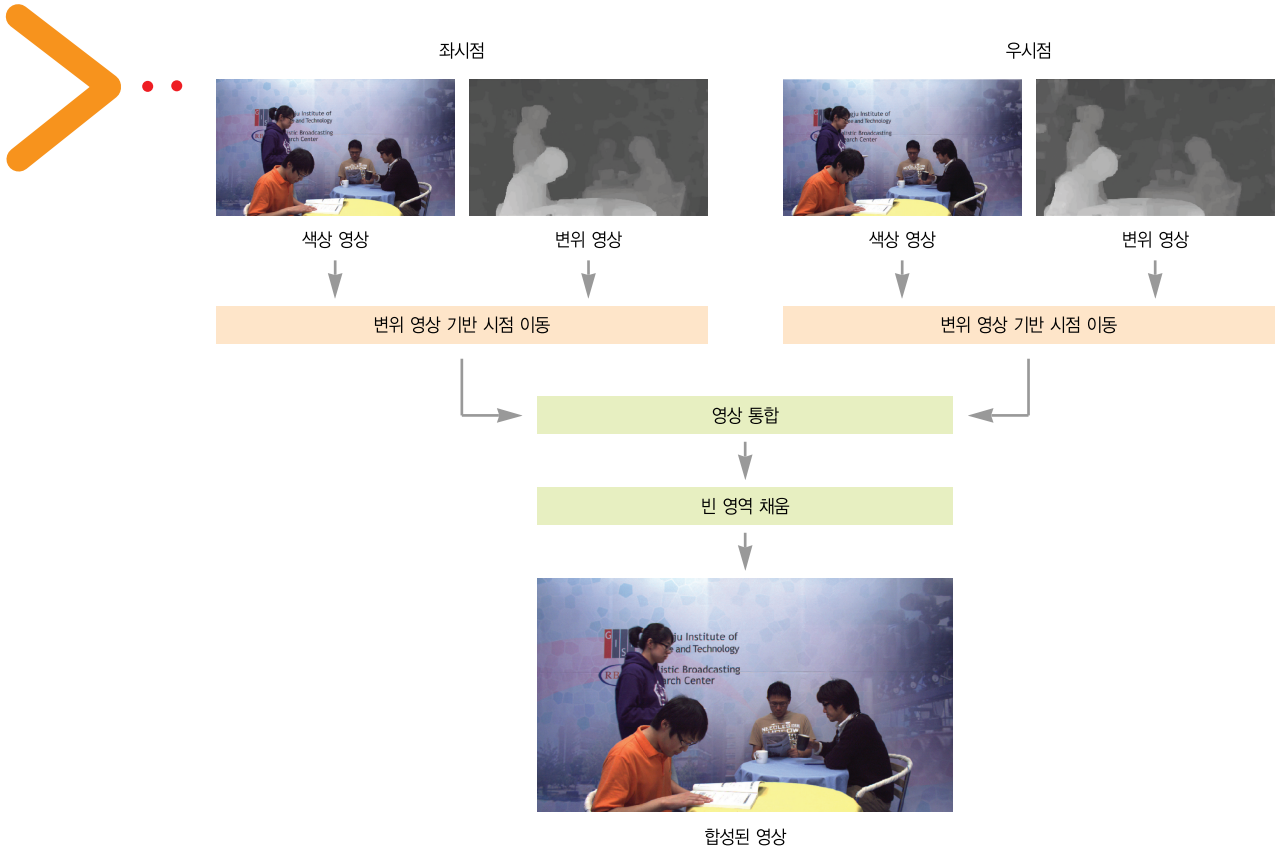


그림 6. 중간시점 영상합성의 전체 흐름도

(1) 변위 영상 기반 시점 이동

영상을 합성하기 위해서는 우선 좌영상과 우영상 각각을 가상 시점으로 이동한다. 시점을 이동하는 방법은 다음과 같은 식으로 표현될 수 있다.

$$I_v(x,y) = I_L\{x + \alpha D_L(x,y), y\} = I_R\{x - (1-\alpha)D_R(x,y), y\} \quad (6)$$

여기서 I_v, I_L, I_R 은 각각 가상 시점, 좌시점, 우시점을 나타내며, α는 좌시점과 우시점 사이에 존재하는 가상 시점의 위치를 나타낸다 (좌: α=0, 우: α=1).

변위 영상을 이용한 시점 이동은 기하학적인 차이로 인해 시점 이동 과정이 정확하게 1:1 사상(mapping)이 되지 않고, 또한, 원영상에서는 가려져 있던 부분이 가상 시점으로 시점 이동되면서 드러나는 영역. 즉, 비폐색(disocclusion)영역이 생기기 때문에 빈 영역이 발생하게 된다. 그림 7은 참조 시점에 따른 시점 합성 결과를 보여준다. 그림 7(a)와 같이, 좌영상만을 이용하여 시점 합성을 할 때에는 빈 영역이 객체의 오른쪽 영역에 존재한다. 반면, 그림 7(b)와 같이, 우영상만을 이용하면 빈 영역이 객체의 왼쪽 영역에 존재한다. 따라서 일반적으로는 좌우영상을 모두 이용하여 빈 영역을 채운다.

(a) 좌영상 → 가상 시점

(b) 우영상 → 가상 시점

그림 7. 참조 시점에 따른 시점 합성 결과

(2) 영상 통합 및 빈 영역 채움

영상 통합 과정은 크게 세 가지로 나눌 수 있다. 이 가운데 두 가지 방법은 좌영상(혹은 우영상) 우선 통합 방법으로 좌영상(우영상)을 먼저 가상 시점으로 시점 이동한 다음 우영상(좌영상)을 이용해서 빈 영역을 채우는 방법이다. 다른 한 가지 방법은 영상 혼합(view blending)방법으로 가상 시점의 위치에 따라 가상 시점으로부터 가까운 시점에 가중치를 높게 하는 가중합(weighting sum)으로 영상을 통합하는 방법이다. 하지만, 영상 혼합 방법은 다른 두 방법과는 다르게 객체의 경계가 두 개로 보이는 문제점을 유발하기 때문에 일반적으로는 좌영상(또는 우영상) 우선 통합 방법을 사용한다.

그림 8. 최종 합성 결과

영상 통합 과정을 통해 합성 영상 대부분의 영역이 채워지지만, 여전히 변위값의 오차 또는 비폐색 영역으로 인한 빈 영역이 남아 있게 된다. 이런 경우에는 빈 영역의 주변값을 고려하는 영상 보간법이나 인페인팅(inpainting)과 같은 방법을 통해서 빈 영역을 채우게 된다 [21][22]. 그림 8은 모든 과정을 마친 최종 합성 영상을 보여준다.

4. 결론

본 논문에서는 차세대 멀티미디어 콘텐츠의 하나인 3차원 영상을 만들기 위한 3차원 깊이 정보 획득 기술과 가상 시점 영상 생성 기술을 설명했다. 본 논문에서 소개된 기술들을 토대로 현재 3차원 콘텐츠를 획득하는 작업이 한창 진행 중이며, 급격히 늘어나고 있는 3차원 방송과 입체영상 산업의 발전 속도로 비추어 볼 때 상용화 단계가 그리 멀지 않았다고 생각된다. 머지않은 미래에는 다시점 영상과 깊이 영상을 이용하여 가상 시점 영상 생성 기술과 3차원 영상 재현 기술이 다양한 응용 분야에 적용될 것으로 예상하며, 앞으로 3차원 영상 산업에 엄청난 파급 효과를 가져다줄 것으로 기대된다.

참고문헌

[1] "제임스 캐머런 '상상력, 기술의 新 르네상스'," http://www.mt.co.kr/view/ mtview.php?type=1&no=2010051308483641428&outlink=1.

[2] 호요성, 김성열, "3DTV 3차원 입체영상 정보처리," 두양사, 2010.

[3] 호요성, 오관정, "다시점 비디오 기반의 3차원 실감방송 ," 방송과기술, 제173권, pp. 130-135, 2010.

[4] 호요성, 이상범, "3차원 비디오 부호화를 위한 MPEG 국제 표준화 작업," 방송과기술, 제175권, pp. 132-139, 2010.

[5] 호요성, 이상범, "3차원 비디오 부호화 기술의 국제 표준화 동향," TTA Journal, no. 123, pp. 77-83, 2009.

[6] ISO/IEC JTC1/SC29/WG11, "Applications and Requirements on FTV," N9466, 2007.

[7] ISO/IEC JTC1/SC29/WG11, "Vision on 3D Video," N10357, 2009.

[8] 호요성, 이은경, 강윤석, "다시점 깊이 카메라를 이용한 3차원 입체영상의 정보 획득 방법," 방송공학회지, 제15권, 제2호, pp. 88-100, 2010.

[9] 강윤석, 호요성, "다시점 카메라와 깊이 카메라를 이용한 3차원 장면의 깊이 정보 생성 방법," 전자공학회논문지 SP편, 제48권, 제3호, pp. 013-018, 2011.

[10] 이상범, 이천, 호요성, "3차원 영상 생성을 위한 깊이맵 추정 및 중간시점 영상합성 방법," 한국통신학회 논문지, 제34권, 제10호, pp. 1070-1075, 2009.

[11] In-Yong Shin and Yo-Sung Ho, "Disparity Estimation at Virtual Viewpointfor Real-time Intermediate View Generation," 3D Systems and Applications 2011 (3DSA 2011), paper S6-2, pp. 195-198, 2011.

[12] Woo-Seok Jang and Yo-Sung Ho, "Disparity Map Refinement using Occlusion Handling for 3D Scene Reconstruction," International Conference on Embedded Systems and Intelligent Technology (ICESIT), pp. 213-216, Feb. 2011.

[13] D. Sharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 131-140, 2001.

[14] P. Perez, "Markov Random Fields and Images," CWI Q., vol. 11, no. 4, pp. 413-437, 1998.

[15] M. Bleyer and M. Gelautz, "Graph-based Surface Reconstruction from Stereo Pairs using Image Segmentation," SPIE Electronic Imaging, vol. 5665, pp. 288-299, 2005.

[16] Y. Deng, Q. Yang, X. Lin, and X. Tang, "A Symmetric Patch-based Correspondence Model for Occlusion Handling," Int'l Conf. on Computer Vision, pp. 1316-1322, 2005.

[17] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Belief Propagation for Early Vision," Computer Vision and Pattern Recognition, pp. 261-268, 2004.

[18] J. Sun, Y. Li, S. Kang, and H. Shum, "Symmetric Stereo Matching for Occlusion Handling," Computer Vision and Pattern Recognition, pp. 399-406, 2005.

[19] Q. Yang, L. Wang, and N. Ahuja, "A Constant-space Belief Propagation Algorithm for Stereo Matching," IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1458-1465, 2010.

[20] P. Lai, D. Tian, and P. Lopez, "Depth Map Processing with Iterative Joint Multilateral Filtering," Picture Coding Symposium, pp. 9-12, 2010.

[21] K. Oh, S. Yea, and Y. Ho, "Hole Filling Method using Depth-based In-painting for View Synthesis in Free-viewpoint Television and 3-D Video," Picture Coding Symposium, pp. 39(1-4), 2009.