

도메인 변환을 이용한 키넥트 깊이 정보 품질 향상 기법

김영중 최성환 손광훈

연세대학교

khsohn@yonsei.ac.kr

Kinect Depth Map Refinement Based on Domain Transform

Kim, Youngjung Choi, Sunghwan Sohn, Kwanghoon

Yonsei University

요약

최근 많은 영상처리 연구자들 사이에서 마이크로소프트사의 실시간 깊이센서 ‘키넥트’가 상당한 관심을 받고 있다. ‘키넥트’는 실시간으로 깊이정보를 제공함과 동시에 별도의 센서를 부착하지 않고도 컴퓨터와의 인터랙션할 수 있는 가능성을 제공한다. 하지만 ‘키넥트’의 깊이영상은 홀 영역, 부정확한 경계, 낮은 해상도등의 많은 문제점을 지니고 있다. 이러한 부정확한 깊이 정보는 3차원 렌더링, 가상시점 영상 합성, 모션 인식 등에서 성능 저하를 야기한다. 따라서 본 논문에서는 깊이 정보 품질 향상기법에 관하여 깊이영상 신뢰도를 이용한 도메인 변환기반 해상도 상향 알고리즘을 제안한다. 정확하고 빠르게 홀 영역 정보를 추정하기 위해 도메인 변환 기반의 경계 보존 필터링이 사용된다. 또한 다양한 깊이 영상의 노이즈를 효율적으로 제거하기 위해 깊이 영상의 신뢰도를 이용한다. 실험결과를 통하여 제안하는 방법이 효율적으로 홀 영역을 채우고, 부정확한 경계를 제거하여 깊이 영상의 품질을 향상시키는 것을 확인할 수 있다.

1. 서론

최근 학계와 산업을 중심으로 높은 품질의 깊이 정보 획득을 위한 영상처리 기술연구가 활발히 이뤄지고 있는 가운데, 많은 영상 처리 연구자들 사이에서 마이크로소프트사의 실시간 깊이센서 ‘키넥트’가 상당한 관심을 받고 있다. ‘키넥트’는 적외선 프로젝터와 카메라 사이의 NCC(Normalized Cross Correlation matching)을 이용하여 실시간 깊이영상을 제공한다. 따라서 ‘키넥트’로부터 얻어지는 깊이정보는 크게 3가지의 한계점을 지니고 있다. 첫째 적외선을 이용하기 때문에 물체의 표면에 매우 민감하다. 즉 유리나 모니터와 같은 표면에선 정확한 깊이정보를 얻을 수 없다. 둘째, 스테레오 정합 방식이기 때문에 부정확한 경계와 깊이 정보 팽창현상이 발생한다. 셋째, 깊이영상은 컬러영상에 비해 낮은 해상도를 가지고 있다. 위와 같은 문제점을 해결하기 위해 본 논문에서는 깊이영상 신뢰도를 이용한 도메인 변환(이하 DT)[1] 기반 해상도 상향 알고리즘을 제안한다.

초기의 깊이 영상 품질 향상 기법은 Bilateral filtering[2] 기반의 기법이 많이 연구 되었다. 특히 컬러영상을 이용하여 깊이 영상을 상향 변환 하는 JBU(Joint Bilateral Up-sampling)[3] 기법이 가장 널리 연구되었다. 하지만 JBU의 경우 깊이영상의 경계와 컬러영상의 경계가 일치하지 않는 영역에서 깊이정보가 흐려지는 단점이 있다. 또한 높은 복잡도로 인하여 실시간 동작이 어렵다. 이를 해결하기 위해 Chan은 JBU를 확장하여 multi-lateral up-sampling 기법인 NAFDU(Noise Aware Filter for Depth Upsampling)을 제안하였다[4]. Garcia는 깊이 정보의 기울기 값을 이용하여 영상을 상향 변환하는 PWAS(Pixel Weighted Average Strategy)를 제안하였다[5]. 하지만 두 방식 모두

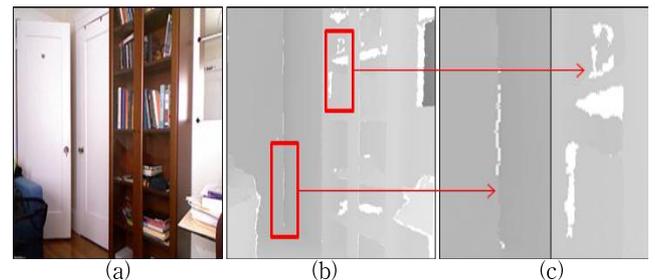


그림 1. 부정확한 키넥트 깊이정보: a) 컬러영상, b) 깊이영상, 홀 영역(흰색), c) 홀 영역 확대

깊이 정보 경계의 흐려짐을 줄여줄 뿐 완전히 제거하지는 못하였다. 따라서 본 논문에서는 신뢰도를 측정하여 컬러와 깊이 영상의 경계불일치 문제를 완화 하는 깊이 영상 상향 알고리즘을 제안한다. DT 필터링을 이용하여 영상 상향 시 깊이정보의 경계를 보존하도록 한다. 이때 깊이영상의 신뢰도를 측정하여 부정확한 깊이 정보의 영향을 최소화한다. 신뢰도는 마스크 내부의 부정확한 픽셀 정보의 양과 깊이 정보의 기울기에 따라 결정된다. 즉, 주변에 부정확한 픽셀 정보가 많을수록, 깊이 정보의 기울기가 클수록 신뢰도는 낮아진다. 실험결과를 통하여 제안하는 방법이 효율적으로 홀 영역을 채우고, 부정확한 경계를 제거하여 깊이 영상의 품질을 향상시키는 것을 확인할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서 DT를 이용한 경계 보존 필터링을 소개하고 3장에서는 신뢰도를 이용한 영상 상향 변환 기법을 제안한다. 제안한 방식으로 상향 변환한 깊이 영상의 결과를 4장에서 분석하고 5장에서 본 논문의 결론을 맺는다.

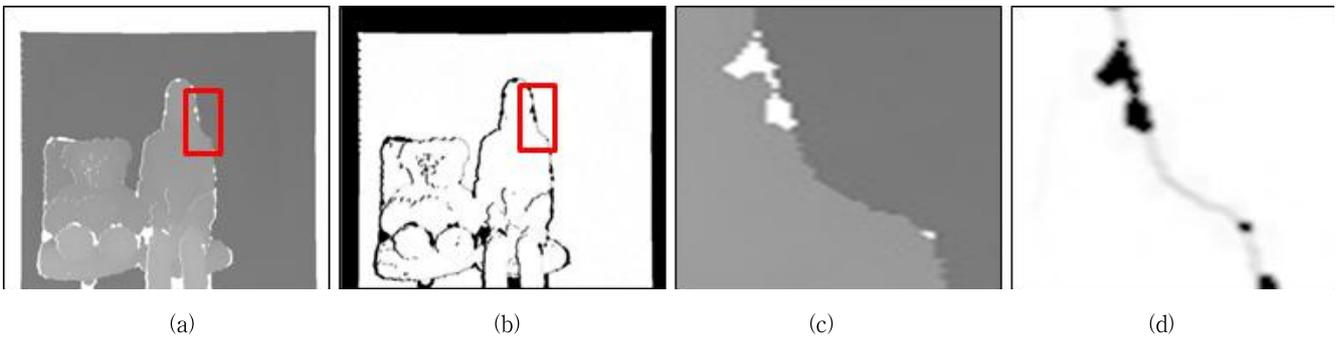


그림 2. 깊이 영상의 신뢰도 측정 예: a) 원본 깊이 영상 b) '키넥트' 깊이 영상의 신뢰도 c) a)의 확대 영상, d) b)의 확대 영상

2. 도메인 변환

경계 보존 필터링을 수행하기 위하여 본 논문에서는 DT을 이용한다. DT는 n차원의 신호를 geodesic 거리가 보존되도록 $l(l < n)$ 차원으로 변환하여 실시간 경계 보존 필터링을 수행하는 방법이다. 1차원 DT는 다음과 같이 주어진다.

$$ct(u) = \int_0^u 1 + \sum_{k=1}^c |I'_k(x)| dx \quad (1)$$

식(1)에서 I_k 는 신호의 k번째 채널, c는 신호 채널의 수, I'_k 는 k번째 채널 신호의 기울기 값을 각각 나타낸다. 이로부터 변환된 신호는 기울기 값이 큰 정보들은 서로 멀리 위치하게 되고, 기울기 값이 작은 정보들은 서로 가까운 곳에 위치하게 된다. 이 신호를 고정된 너비의 l차원 가우시안 필터링을 한 후 원래의 n차원 도메인으로 변환하면 기울기가 큰 영역이 보존되는 경계보존 필터링을 수행할 수 있다. 영상의 빠른 가우시안 필터링을 수행하기 위해 RF(Recursive filtering)를 이용한다.

3. 깊이영상 상향 변환 기법

'키넥트'로부터 얻어지는 깊이 영상은 적외선 프로젝터와 카메라 사이의 시차로 인해 그림 1과 같이 홀 영역, 부정확한 경계등의 문제점을 지니고 있다. 이를 해결하기 위해 본 논문에서는 부정확한 깊이 정보를 홀 영역 주변의 깊이 정보와 팽창된 깊이 정보로 분류한다. 분류한 깊이 정보를 이용하여 신뢰도를 계산한 뒤 DT을 사용하여 최종 깊이 영상을 획득한다.

3.1 신뢰도 측정

정확한 깊이 정보로부터 홀 영역의 깊이 정보를 추론하기 위하여 본 논문에서는 깊이 정보의 신뢰도를 계산한다. '키넥트'로부터 얻어진 깊이 정보의 신뢰도는 깊이 정보의 기울기와 마스크 내 홀 영역의 크기로부터 구해진다. 깊이 정보의 기울기 값을 통해 팽창된 영역의 신뢰도를 낮출 수 있다[5]. 깊이 정보의 기울기는 다음과 같은 식을 통해 얻어진다.

$$C_f(p) = \frac{1}{|N_p^{w_f}|} \sum_{q \in N_p^{w_f}} \exp\left(\frac{|D_x(p)| + |D_y(p)|}{2\sigma^2}\right) \quad (2)$$

식(2)에서 $N_p^{w_f}$ 는 p점을 중심으로 크기가 $w_f = 9 \times 9$ 인 윈도우, $|N_p^{w_f}|$ 는 윈도우 내의 픽셀의 수, $\sigma = 0.3$ 는 가우시안 함수의 표준편차, D_x 와 D_y 는 각각 원본 깊이 영상의 x축, y축 방향 기울기를 나타낸다. 기울기의 값이 작을수록 $C_f(p)$ 는 1에 가까워지며, 기울기의 값이 클수록 0에 수렴한다.

'키넥트'로부터 얻어진 깊이정보의 홀 영역을 통해 오차 마스크 Ω 를 얻을 수 있다. 다시 말해, p지점이 홀 이면 $\Omega(p) = 0$, 아닌 경우 1을 나타낸다[6]. 오차 마스크를 기반으로 홀 영역 주변의 신뢰도를 정의하면 다음과 같다.

$$C_h(p) = \frac{1}{|N_p^{w_h}|} \sum_{q \in N_p^{w_h}} \Omega(q) \quad (3)$$

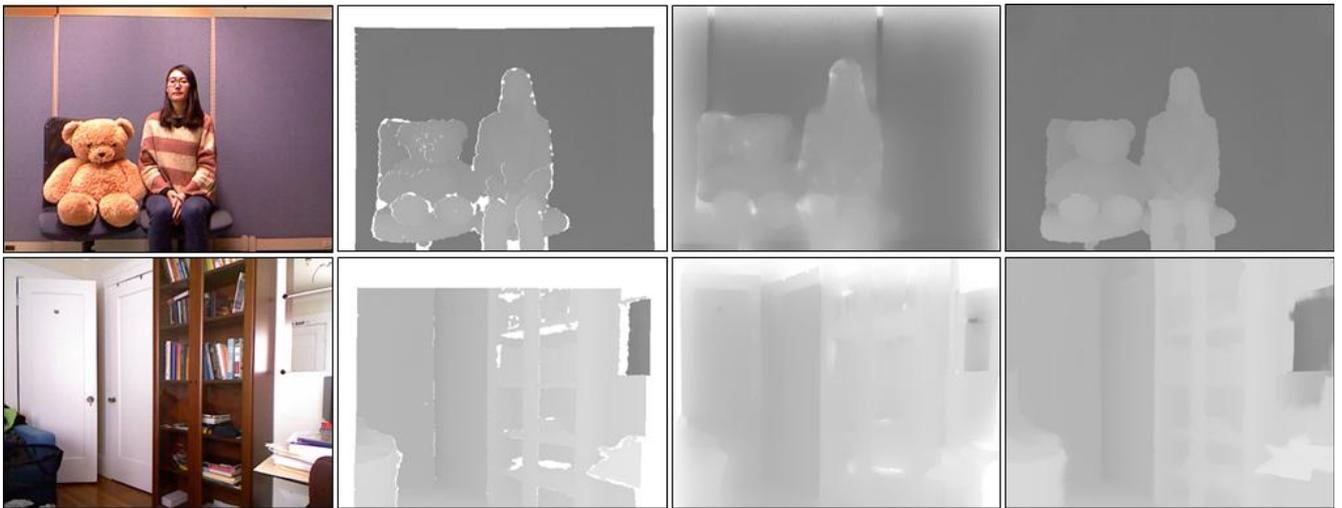
식(3)에서 $N_p^{w_h}$ 는 p점을 중심으로 크기가 $w_h = 9 \times 9$ 인 윈도우, $|N_p^{w_h}|$ 는 윈도우 내의 픽셀의 수를 의미한다, 따라서 제안하는 최종 신뢰도 함수는 다음과 같다.

$$C(p) = C_f(p) C_h(p) \quad (4)$$

그림 2에서처럼 $C(p)$ 를 통해 각각의 깊이 영상의 픽셀은 기울기가 작고 주변의 홀인 픽셀이 적을수록 높은 신뢰도를 가지며, 반대의 경우 낮은 신뢰도를 가진다. 홀 영역 주변의 깊이 정보는 부정확하기 때문에 그림 2-(d)에서 매우 낮은 신뢰도를 갖는 것을 확인할 수 있다. 또한 팽창된 깊이 영상의 신뢰도를 줄이기 위하여 깊이 영상의 기울기 값을 이용하였기 때문에 그림 2-(d)의 어께 영역은 홀이 없음에도, 깊이 영상의 경계 영역임으로 신뢰도가 낮게 측정되는 것을 확인할 수 있다.

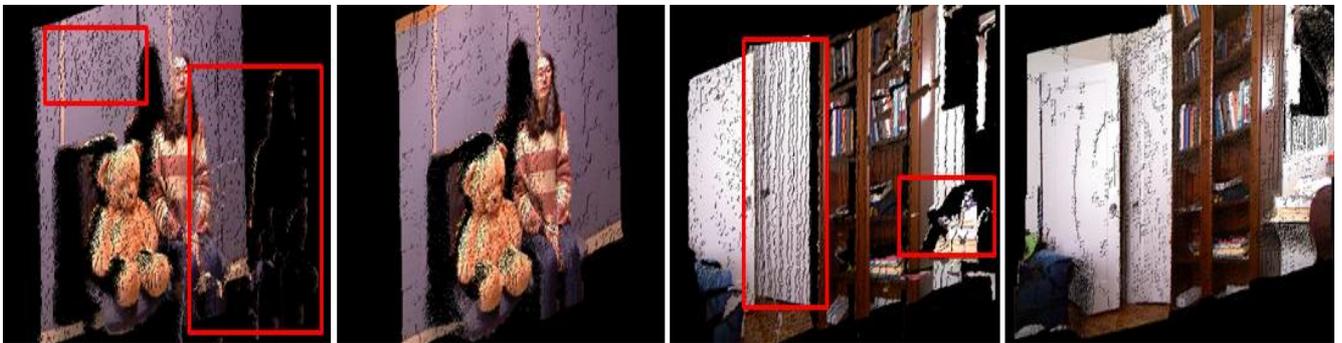
3.2 깊이 영상 상향 변환

최종 깊이 영상은 도메인 변환과 신뢰도 영상을 이용하여 얻을 수 있다.



(a) (b) (c) (d)

그림 3. 실험 결과: a) 컬러 영상, b) 원본 깊이 영상, c) 신뢰도를 사용하지 않은 DT, d) 제안한 방법



(a) (b) (c) (d)

그림 4. 3차원 렌더링 결과: a) 원본 깊이 영상(획득 영상), b) 제안한 방법(획득 영상), c) 원본 깊이 영상(‘키넥트’ Data set), d) 제안한 방법(‘키넥트’ Data set)

$$D' = \frac{DT(D \cdot C \cdot \Omega)}{DT(C \cdot \Omega)} \quad (5)$$

식(5)에서 D 와 D' 는 각각 원본 깊이 영상과 보정된 깊이 영상을 나타내며, $DT(\cdot)$ 는 도메인 변환을 이용한 경계 보존 필터링을 의미한다. 신뢰도가 곱해진 깊이 영상을 필터링 한 후, 신뢰도를 필터링한 영상으로 정규화 하여 최종 깊이 영상을 계산한다. 정규화 영상으로 나눠주는 이유는, 홀 픽셀 값이 유효한 값으로 필터링 되는 것을 막기 위한 것이다. 즉 윈도우내의 홀을 제외한 필터 계수의 합이 1이 되도록 한다.

4. 실험 결과 및 분석

4.1 실험 조건

본 논문에서는 마이크로소프트사의 ‘키넥트’를 이용하여 직접 획득한 영상과 ‘키넥트’ 영상 Data set[7]을 사용하여 실험을 진행하였다 실험

환경은 Windows 7, Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz, 4.00GB RAM, 64비트 운영체제이다. 깊이 영상과 칼라 영상의 캘리브레이션(calibration)을 수행하기 위해 ‘키넥트’ SDK가 제공하는 함수를 이용하였다. 실험 결과의 성능 평가를 위해 OpenGL을 이용한 3차원 렌더링을 이용하였다. 제안하는 신뢰도 측정 방식의 성능을 확인하기 위해 신뢰도를 사용하지 않은 DT 방식이 비교에 이용되었다. 모든 실험에는 $w_h, w_f = 9 \times 9, \sigma = 0.3$ 으로 동일한 파라미터가 쓰였다.

4.2 실험 결과

그림 3에서 제안한 방식이 신뢰도를 사용하지 않은 방식에 비해 시각적으로 의미 있는 결과를 보여주는 것을 확인할 수 있다. 그림 3-(c)의 결과를 통하여 큰 홀에서 깊이 정보가 번지는, 블러링 문제가 발생하는 것을 볼 수 있다. 그림 3-(d)에서 제안한 방식은 깊이 정보의 신뢰도를 바탕으로 블러링 및 텍스처 침투 문제를 해결하였고 큰 홀 영역에서 우수한 결과를 보여주는 것을 확인할 수 있다. 또한 그림 4를 통해서 제안하는 방식이 우수한 3차원 렌더링 결과를 나타내는 것을 확인할

수 있다. 그림 4-(a),(c)의 결과에서 부드럽지 않은 깊이 영상에 의한 지터링 효과와 홀 영역에 의한 부정확한 렌더링 결과를 볼 수 있다. 그림 4-(b),(d)에서 제안하는 방식은 경계 보존 필터링을 바탕으로 지터링 효과와 부정확한 렌더링 영역을 제거하였다.

5. 결론

본 논문에서는 DT를 이용한 깊이 영상 품질 향상의 정확성을 높이기 위하여 신뢰도를 고려한 깊이 영상 상향 변환 기법 과정을 제안하였다. 제안한 방식은 부정확한 깊이 정보를 홀 영역 주변의 깊이 정보와 팽창된 깊이 정보로 분류한 뒤 신뢰도를 측정한다. 이를 DT기반의 깊이 영상 상향 변환 기법에 적용함으로써 깊이 영상 상향 변환 시 경계 값이 블러링되는 문제와 텍스처 침투 현상을 해결하여 결과의 정확성을 높였다. 실험 결과에서 제안한 방식이 신뢰도를 고려하지 않는 방식과 비교하여 우수한 결과를 보임을 확인하였다.

참고문헌

- [1] E. Gastal, M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1-11, 2011.
- [2] C. Tomasi, R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vis.*, 1998.
- [3] J. Kopf, M. Cohen, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 96-99, 2007.
- [4] D. Chan, H. Buisman, and S. Thrun, "A noise-aware filter for real-time depth upsampling," in *Proc. Workshop Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications.*, 2008.
- [5] F. Garcia, B. Mirbach, and B. Ottersten, "Pixel weighted average strategy for depth sensor data fusion," in *Proc. Int. Conf. Comput. Vis.*, 2010.
- [6] S. Choi, B. Ham, and K. Sohn, "Hybrid approach for accurate depth acquisition with structured light and stereo camera," in *Proc. IEEE Conf. BMSB.*, 2012
- [7] <http://kinectdata.com>