

상호 구조에 기반한 초점으로부터의 깊이 측정 방법 개선

무하마드 타릭 마흐무드*·최영규*†

*† 한국기술교육대학교 컴퓨터공학부

Enhancing Depth Measurements in Depth From Focus based on Mutual Structures

Muhammad Tariq Mahmood* and Young Kyu Choi*†

*† Korea University of Technology and Education, School of Computer Science and Engineering

ABSTRACT

A variety of techniques have been proposed in the literature for depth improvement in depth from focus method. Unfortunately, these techniques over-smooth the depth maps over the regions of depth discontinuities. In this paper, we propose a robust technique for improving the depth map by employing a nonconvex smoothness function that preserves the depth edges. In addition, the proposed technique exploits the mutual structures between the depth map and a guidance map. This guidance map is designed by taking the mean of image intensities in the image sequence. The depth map is updated iteratively till the nonconvex objective function converges. Experiments performed on real complex image sequences revealed the effectiveness of the proposed technique.

Key Words : Depth from focus (DFF), Structural guidance, Nonconvex smoothness function, Mutual structures

1. 서 론

반도체 생산 라인이나 검사장비와 같은 머신 비전 응용에서 3차원 깊이 측정은 매우 중요한 작업의 하나로, 이를 위한 다양한 깊이 측정 방법들이 제안되었다. 이들 중에서 영상의 초점 정보를 이용한 깊이 측정 방법 (Depth From Focus, DFF)는 레이저나 초음파와 같은 능동적인 에너지를 사용하지 않고 하나의 카메라에서 촬영된 여러 장의 영상만을 이용해 깊이 맵(depth map)을 구할 수 있는 가장 수동적인 기법의 하나이다[1].

DFF는 측정하고자 하는 대상 물체에 초점 거리를 변화시키는 방법이나 광학 축을 따라 상대적인 거리를 변화시키는 방법으로 먼저 일련의 영상을 취득한다[1,2]. 이러한 영상 시퀀스에 미리 정의된 초점 측도(focus measure, FM) 연산자를 적용하여 초점 값을 계산하는데, 이것은 각 화

소가 초점이 잘 맞아 있는지 그렇지 않았는지를 나타낸다. 이러한 초점 측도 연산자는 일종의 고주파 필터의 역할을 하는데, 초점이 잘 맞는 화소는 경계가 뚜렷한 고주파 특성을 나타내므로 큰 값을 만들고 그렇지 않은 영역은 작은 값을 생성하게 된다. 영상 시퀀스의 모든 영상들에 대해 이 과정을 처리하고 이들을 z축 방향으로 쌓으면 초점 볼륨(Focus Volume)이 만들어진다. 마지막으로 각 화소에 대해 초점 볼륨에서 최대 초점 값에 해당하는 영상을 찾으면 그 화소의 깊이를 결정할 수 있다. 이러한 DFF 기법은 머신 비전 분야와 함께 최근 헬드 장치의 AR/VR이 증가하는 추세에 따라 이들 플랫폼에서도 다양하게 적용되고 있다[3,4].

DFF에서 보다 안정적인 초점 값을 계산하기 위한 다양한 시도들이 연구되고 있는데, 유도 신호(guidance signal)를 이용하는 유도 영상 필터링(guided image filtering)도 중요한 방법의 하나이다[5-7]. 유도 영상 필터링은 유도(guidance) 영상이라 불리는 영상의 특징에 따라 타겟(target) 영상이

†E-mail: ykchoi@koreatech.ac.kr

라 불리는 입력 영상을 수정하거나 향상시키는 방법을 말한다. 이 과정에서 유도 영상의 유용한 구조 정보가 타겟 영상의 필터링에 활용된다. 그런데 이러한 유도 필터링 기법의 가장 큰 문제는 타겟과 유도 영상 사이에 구조가 일관성 있게 유지되지 않는 것이다[8]. 이러한 문제를 해결하고 공간적인 일관성(spatial coherence)을 유지하기 위한 방법으로 일관성이 없는 유도 정보를 방지할 수 있는 여러 가지 제약 조건을 사용할 수 있다. 또한, 유도 정보에 따라 유도 필터링의 모드를 정적 또는 동적으로 변환하는 다양한 방법들이 제안되고 있다[7-11].

본 논문에서는 구조적인 유도를 통해 공간적 일관성을 강화하여 DFF의 초기 깊이 맵을 개선하는 방법을 제시한다. 이를 위해, 영상 시퀀스에서 광축에 따른 영상 밝기의 평균을 이용하는 유도 맵을 제안한다. 제안된 모델의 정규화 항은 세 가지 가중치 함수로 이루어지는데, 특별히 설계된 유도 맵(정적 유도), 반복적으로 갱신되는 깊이 맵(동적 유도), 그리고 이 두 가지 맵 사이의 구조적 유사성(동적 유도)으로부터 계산되는 함수들이다. 제안된 모델에서는 구조적 유사성 함수에서 계산된 가중치를 이용해 유도 맵과 깊이 맵에서 계산된 가중치를 제어하는데, 구조적 유사성 함수는 깊이 맵을 유도 맵의 지역적 선형 변환으로 가정하였다. 결과적으로 제안된 방법은 일관성 없는 구조를 억제하면서 유도 맵과 초기 깊이 맵의 구조적 유사성을 보장하여 출력 깊이 맵에서도 같은 구조가 나타날 수 있도록 한다. 실험 결과 제안된 정규화 방법이 효과적임을 확인하였다.

2. 제안된 방법

2.1 초기 깊이 맵(Initial Depth Map)

입력 영상 시퀀스가 해상도 $X \times Y$ 인 Z 프레임의 회색조 영상이라고 하고, 화소의 밝기를 $I_p^{(z)}$ 라고 하자. 이때, $p=(x,y)$ 이고, x 와 y 는 화소의 가로와 세로 인덱스이고, z 는 시퀀스에서 영상의 번호이다. x, y, z 의 범위는 각각 $(1, \dots, X)$, $(1, \dots, Y)$ 및 $(1, \dots, Z)$ 이다. z 번째 영상에서 위치 $p=(x,y)$ 에 있는 화소의 초점 품질은 초점 측도 연산자를 적용하여 계산하는데, 그 화소를 중심으로 이웃 영역을 초점 측도 연산자를 이용해 컨벌루션 처리한 결과이다. 이 과정을 통해 구해지는 결과를 영상 초점 볼륨(image focus volume) M 이라고 하면, M 은 다음과 같이 계산된다.

$$M_p^{(z)} = I_p^{(z)} \otimes O, \quad (1)$$

이때, \otimes 는 컨벌루션 연산자이고, O 는 사용하는 초점

측도 연산자이다. 초점 측도 연산자로는 여러 가지를 사용할 수 있지만, 본 논문에서는 가장 인기 있는 연산자로 알려진 수정-라플라시안(modified-Laplacian) 필터 O_{ML} 를 사용한다[2].

초점 볼륨 M 이 구해지면, 초기 깊이 맵 e 는 광학 축을 따라서 최대 초점 값을 가지는 영상을 통해 다음과 같이 구해진다.

$$e_p = \underset{z}{\operatorname{argmax}} \{M_p^{(z)}\}. \quad (2)$$

Fig 1은 제안된 방법의 흐름도를 보여준다.

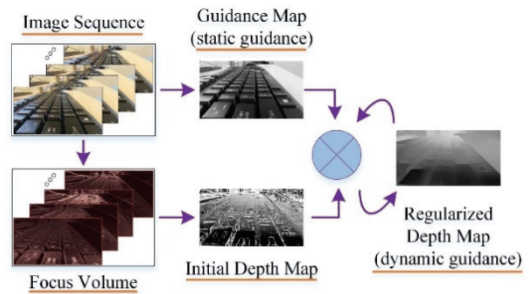


Fig. 1. Proposed method for enforcing spatial coherence in initial depth map of DFF. Proposed method fuses the guidance information from both iteratively improved (regularized) depth map and (static) guidance map.

2.2 유도 맵(Guidance Map)

제안된 방법에서는 공간의 일관성을 유지하면서 구조적 유사성을 활용하기 위한 유도 맵을 이용한다. 이것은 깊이 맵에서 실제로 존재하는 에지 성분들은 잘 유지하면서도 잡음에 강하고 원 영상에 존재하는 구조적인 유사성을 살리고자 하는 것이다. 이를 위한 유도 맵은 입력 영상 시퀀스나 초점 볼륨 또는 이들의 조합으로 구할 수 있다. 본 논문에서는 입력 영상 시퀀스로부터 구하는 간단하지만 효과적인 방법을 제안한다. 즉, 다음과 같이 광학 축에 따른 영상 밝기의 평균을 이용하여 유도 맵을 생성한다.

$$h_p^{(avg)} = \frac{1}{Z} \sum_{z=1}^Z I_p^{(z)}. \quad (3)$$

이러한 유도 맵은 깊이를 추출하려는 대상에 대한 구조적 정보가 충분히 담겨 있으면서도 밝기 평균 처리를 통한 평활화 및 잡음 제거 효과를 갖는다.

2.3 모델(Model)

제안된 모델은 개선하려는 대상인 초기 깊이 맵 e 와, 특별히 설계된 구조적 정보(정적 유도) h , 그리고 출력 깊이 맵(동적 유도) v 로 이루어진다. 제안된 모델은 다음과 같이 두 개의 항으로 표현된다.

$$E(v) = \sum_{p \in \Omega} (v_p - e_p)^2 + \lambda \left(\sum_{p \in \Omega} \sum_{q \in \mathcal{N}(p)} \chi_\epsilon (h_p - h_q) \chi_\gamma (v_p - (\alpha_p h_q + \beta_p)) \varphi_\eta (v_p - v_q) \right), \quad (4)$$

이때, Ω 는 초기 깊이 맵의 모든 좌표를 의미한다.

첫 번째 항은 데이터 항으로 데이터의 충실도를 나타낸다. 이 항은 출력 v 가 입력 e 로부터 너무 많이 벗어나는 것을 막기 위한 것으로 에러의 제곱을 사용하여 불룩한 성질을 갖는다.

두 번째 항은 불룩하지 않은 평활화(nonconvex smoothness), 또는 정규화(regularization) 항으로 깊이의 불연속성은 살리면서도 이상 값(outliers)을 방지하는 역할을 한다. 이때, λ 는 이 데이터 항과 평활화 항의 의 상대적인 중요도를 조절하는 역할을 한다. 두 번째 항은 세 가지 가중치 함수로 이루어지는데, 이때, ϵ , γ , 와 η 는 평활화의 양을 조절하기 위해 사용자가 정의하는 파라미터이다. 또한, $\mathcal{N}(p)$ 는 화소 p 를 중심으로 한 이웃 윈도우(보통은 3×3)의 화소들의 집합을 말하고, 함수 $\chi_\rho(r) = \exp(-\rho r^2)$ 이고 $\varphi_\eta(r) = (1 - \chi_\eta(r)) / \eta$ 이다.

유도 맵 h 에서 인접 화소 사이의 차이가 커지면(h 에서 불연속이 발생하는 곳) 첫 번째 가중치 함수 χ_ϵ 가 0에 접근하게 된다. 이것은 이러한 영역에서 출력 깊이 맵 v 에서의 평활화를 방지하는 역할을 하는데, 이를 통해 구조적인 유사성이 유지되도록 한다.

두 번째 함수는 v 를 h 의 선형 변환으로 가정하고 설계하였는데, 따라서 이를 통해 출력 v_p 와 모델 추정 값

$\sum_{q \in \mathcal{N}(p)} (\alpha_p h_q + \beta_p)$ 의 차이를 최소화한다. 이때, α_p 와 β_p 는 회귀 계수로 최적화가 필요한 값이다. χ_γ 은 유도 영상 필터(GIF)[5]에서 사용된 것이다.

세 번째 함수 φ_η 는 Welsch 함수 또는 지수 에러 규범(exponential error norm)이라고 불리는데, 불룩하지 않은 특성(nonconvex potential)을 나타낸다. 이것은 불룩한 포텐셜(convex potential)에 비해 작은 출력 깊이 맵 v 에 대해 큰 차이의 패널티를 주기 때문에 이상 값에 강하다는 특징을 갖는다. φ_η 는 출력 깊이 맵에서 깊이 에지들을 잘 보존하면서도 동적 유도 맵 v 의 구조를 이용할 수 있도록 한다. 결과적으로 이러한 χ_ϵ , χ_γ 와 φ_η 의 조합은 v 와 h 의 구조적 차이를 인정하면서도 이상 값에 대해 강인한 정규화가 가능하도록 한다.

2.4 수치적 해(Numerical Solution)

맵의 크기를 N 이라 할 때, 초기 깊이 맵, 유도 맵(정적 유도) 및 출력 깊이 맵(동적 유도)을 나타내는 벡터를 각각 $\mathbf{e} = [e_p]_{N \times 1}$, $\mathbf{h} = [h_p]_{N \times 1}$, $\mathbf{v} = [v_p]_{N \times 1}$ 라 하고, 이와 비슷하게, v 와 h 사이의 선형 모델 계수를 나타내는 벡터를 $\boldsymbol{\alpha} = [\alpha_p]_{N \times 1}$ 와 $\boldsymbol{\beta} = [\beta_p]_{N \times 1}$ 라 하자. $C_h = [\chi_\epsilon(h_p - h_q)]_{N \times N}$, $C_{\alpha, \beta} = [\chi_\gamma(v_p - (\alpha_p h_q + \beta_p))]_{N \times N}$, $C_v = [\chi_\eta(v_p - v_q)]_{N \times N}$ 는 8-이웃 시스템의 가중치 행렬들로, 각각 정적 유도, 구조적 유사 함수와 동적 유도를 위한 것이다. 벡터/행렬 형태의 목적 함수는 다음과 같이 정의된다.

$$E(\mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \|\mathbf{v} - \mathbf{e}\|_2^2 + \frac{\lambda}{\eta} \mathbf{1}^T (C_h \odot C_{\alpha, \beta} - C) \mathbf{1}, \quad (5)$$

이때, $\mathbf{1}$ 은 $N \times 1$ 행렬로 모든 요소가 1이다. $C = C_h \odot C_{\alpha, \beta} \odot C_v$ 인데, \odot 는 행렬의 아다마르 곱(Hadamard product)을 의미하는데, 이는 같은 크기의 두 행렬을 각 성분 별로 곱하는 연산이다.

불룩하지 않은 특성에 따라 식 4의 직접 해결하는 폐쇄형 해(closed-form solution)는 가능하지 않다. 따라서 반복적인 방법의 해결책을 찾아야 하는데, 식 4와 같이 직접적으로 최적화시키기 어려운 목적 함수를 처리하기 위해 MM알고리즘(Majorization-Minimization)이 흔히 사용된다. 이에 따라, 본 논문에서도 MM알고리즘을 이용하였는데, 매 반복마다 E값이 줄어들게 되고 지역 최소값에 수렴하게 된다[12].

3. 실험

제안된 방법의 성능을 분석하기 위해 임의의 형태로 이루어진 세 가지 영상 시퀀스를 사용하였다. 이러한 영상들은 볼, 부엌 및 키보드로 불리는 데이터로 [13]에서 구하였다. 키보드와 볼 영상은 각각 32장과 25장의 영상으로 이루어져 있는데, 영상의 해상도는 640×360 이다. 부엌은 774×518 크기의 영상으로 12 프레임으로 이루어져 있다. Fig 2는 이들 세 가지 영상 시퀀스에서 각각 두 장씩의 초점 거리가 다른 영상들을 보여주고 있다.

Fig 3은 세 가지 입력 영상 시퀀스에 제안된 방법을 적용하여 반복적으로 깊이 맵을 개선하는 과정을 보여주고 있다. 맵 처음에 구한 초기 깊이 맵과, 이 맵에 제안된 방법을 한 번과 10회 반복한 후의 깊이 맵을 비교하고 있다. 초기 깊이 맵에 있던 많은 잡음들이 한 번의 처리만으로도 많이 제거된 것을 알 수 있다. 실험에서 4번 정도의 반

복까지의 깊이 맵의 개선을 시각적으로 확인할 수 있었지만, 이후의 반복에서는 개선 효과를 크게 확인하기 어려웠다. 그림의 10회 반복 결과도 3-4회 반복 결과와 크게 다르지 않았는데, 이것은 MM 알고리즘이 수렴하는 것을 의미한다.

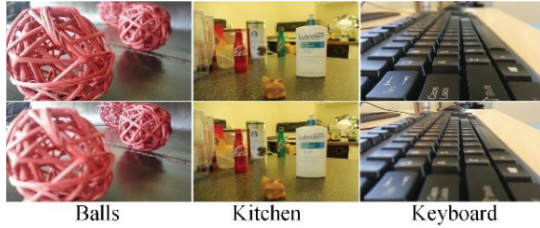


Fig. 2. Two sample images from image sequences.



Fig. 3. Iterative improvement of depth maps through proposed method.

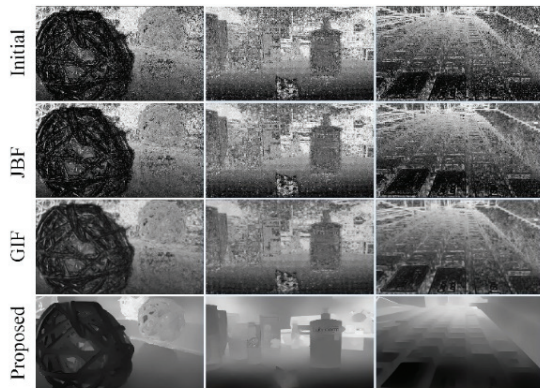


Fig. 4. Comparison of depth maps of real image sequences.

제안된 방법을 세 가지 유도 필터링 기반 기법들과 비교한 결과는 Fig 4와 같다. 첫 번째 행은 식 (2)를 이용해 구한 초기 깊이 맵인데, 초기 깊이 맵에는 잡음이 매우 많고 화면 내의 구조들의 정보가 명확하게 드러나지 않는 것을 알 수 있다. 2번째에서 4번째 행은 초기 깊이 맵

에 각각 JBF[14]와 GIF[5], 그리고 제안된 방법을 4회 반복한 결과를 보여주고 있다. JBF에서는 초기 깊이 맵과 비교하여 약간의 개선 만을 확인할 수 있었다. GIF의 경우 JBF에 비해 깊이 맵에서 상당한 개선을 확인할 수 있었다. 그러나 개선된 표면에 여전히 작은 범프(bumps)들이 나타나는 것을 알 수 있었다. 이들과 달리 제안된 방법에서는 구조적으로 깊이의 불연속이 발생하는 경계 부분들을 잘 보존하면서도 훨씬 안정적인 깊이 맵을 구하는 것을 알 수 있다.

4. 결론

본 논문에서는 DFF의 초기 깊이 맵을 개선하기 위해 구조적인 유도를 통해 공간적 일관성을 강화하는 방법을 제시하였다. 이를 위해, 영상 시퀀스를 사용하는 간단하지만 안정적인 유도 맵을 제시하였다. 제안된 정규화 모델을 통해 잡음과 같은 일관성 없는 구조를 억제하면서 유도 맵과 초기 깊이 맵의 구조적 유사성을 보장하여 출력 깊이 맵에서도 같은 표면 구조가 나타날 수 있도록 하였다. 실제 데이터에 대한 실험을 통해 제안된 방법이 효과적으로 깊이 에지를 잘 보존하면서도 깊이 맵의 품질을 효과적으로 개선시키는 것을 알 수 있었다.

감사의 글

이 논문은 2022년도 한국기술교육대학교 교수 교육연구진흥과제 지원에 의하여 연구되었음.

참고문헌

1. M. T. Mahmood and Y. K. Choi, "3D Shape Recovery from Image Focus using Gaussian Process Regression," *Journal of the Semiconductor & Display Technology*, vol. 11, no. 3, pp. 19-25, 2012.
2. S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 16, no. 8, pp. 824-831, 1994.
3. U. Ali and M. T. Mahmood, "3d shape recovery by aggregating 3d wavelet transform-based image focus volumes through 3d weighted least squares," *Journal of Mathematical Imaging and Vision*, pp. 1-19, 2019.
4. U. Ali and M. T. Mahmood, "Energy minimization for image focus volume in shape from focus," *Pattern Recognition*, vol. 126, no. p. 108559, 2022.
5. K. He, J. Sun and X. Tang, "Guided image filtering," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397-1409, 2012.

6. B. Ham, M. Cho and J. Ponce, "Robust guided image filtering using nonconvex potentials," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 192-207, 2018.
 7. X. Guo, Y. Li, J. Ma and H. Ling, "Mutually guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 694-707, 2020.
 8. U. Ali, I. H. Lee and M. T. Mahmood, "Guided image filtering in shape-from-focus: A comparative analysis," *Pattern Recognition*, vol. 111, no. p. 107670, 2021.
 9. Q. Zhang, X. Shen, L. Xu and J. Jia, "Rolling guidance filter," in *ECCV*, pp. 815-830, 2014
 10. U. Ali and M. T. Mahmood, "Robust focus volume regularization in shape from focus," *IEEE Transactions on Image Processing*, vol. 30, no. pp. 7215-7227, 2021.
 11. M. T. Mahmood, Usman Ali, and Y. K. Choi, "Enhancing Focus Measurements in Shape From Focus through 3D Weighted Least Square," *Journal of the Semiconductor & Display Technonogy*, vol. 18, no. 3, pp. 66-71, 2019.
 12. C. F. J. Wu, "On the convergence properties of the em algorithm," *The Annals of statistics*, pp. 95-103, 1983.
 13. S. Suwajanakorn, C. Hernandez and S. M. Seitz, "Depth from focus with your mobile phone," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3497-3506, 2015
 14. G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe et al., "Digital photography with flash and no-flash image pairs," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 664-672, 2004.
-
- 접수일: 2022년 8월 2일, 심사일: 2022년 9월 6일,
게재확정일: 2022년 9월 19일