

단안단서를 이용한 자연영상의 상대적 깊이지도 생성

한중원^o 조진수 이일병
연세대학교 컴퓨터과학과
{zeross^o, hamster, yblee}@csai.yonsei.ac.kr

Relative Depth-Map Generation of Natural Scenes using Monocular Cues

Jongwon Han^o, Jinsu Jo, Yillbyoung Lee
Dept. of Computer Science, Yonsei University

요 약

사람은 일반적으로 깊이를 지각하는데 두 눈으로 들어오는 영상의 시차(binocular disparity)를 이용하며 6~15m 정도의 범위 내에서는 매우 뛰어난 깊이 판별 능력을 보인다. 그러나 사람은 하나의 눈만으로도 깊이를 지각하는데 별 어려움을 느끼지 못한다. 이것은 공간의 깊이 지각 단서로 양안단서만이 아니라 다양한 단안단서(monocular cue)들이 함께 사용되기 때문이다. 본 논문에서는 사람이 공간 깊이정보 파악에 사용하는 것으로 알려진 여러 단안 단서들 중 영상의 채도(saturation) 정보와 디포커스(defocus) 정보, 기하학적 깊이(geometric depth) 정보에 기반을 둔 단안 영상에서의 상대적 깊이지도의 생성방법을 제안한다.

1. 서 론

2차원 영상으로부터의 3차원 공간 구조 복원은 컴퓨터 비전의 기본적인 문제들 중 하나이다. 이와 관련된 연구들은 주로 양안부동(binocular disparity)에 근거한 스테레오 비전 기술이며, 단일영상에서 추출한 단안단서에 대한 연구는 결 기울기(texture gradient)[1]나 포커스(focus)[2]와 디포커스(defocus)[3] 정보에 의한 깊이 지각연구 등이 있는데, 사람이 활용하는 단안단서들은 여러 종류가 밝혀져 있으나 그 단서들이 어떻게 경쟁 혹은 융합되는지에 대한 연구는 아직 초기 단계이다.

회화나 사진 등과 같은 2차원 영상에는 물체와 물체, 혹은 물체와 배경간의 거리에 대한 직접적인 정보가 없다. 그러나 인간은 하나의 2차원 영상만으로도 영상 내에 적절한 공간 단서들이 주어지면 그로부터 3차원적 구조를 복원하여 물체들이나 배경간의 상대적 위치 관계를 비교적 정확하게 판단할 수 있다.

한 눈으로 얻을 수 있는 이런 적절한 단서들을 단안단서(monocular cues)라 부르며, 단안단서의 종류에는 결 기울기(texture gradient), 중첩(occlusion), 상대적 크기 변화(relative size), 채도(saturation), 디포커스(defocus), 선형 원근감(linear perspective) 등이 있다. 이런 단서들을 이용할 때 하나의 단안 단서만으로는 정확한 깊이 지각이 어렵기 때문에 두 가지 이상의 단서를 종합하여 깊이 판단에 활용하는 것이 유용하다. 본 논문에서는 위의 단안단서들 중에서 채도 정보와 디포커스 정보에 초점을 맞추어서 깊이지도의 생성방법을 제안한다.

채도의 차이에 의한 깊이정보의 복원은 대기원근법(aerial perspective)에 기반을 둔다. 대기원근법은 레오나르도 다빈치에 의해 명명된 회화기법으로 눈과 물체 사이의 공기층과 빛의 작용에 의해 발생하는 색체의 변화로 거리감을 표현하는 기법이다. 이런 현상은 대기 중에서의 빛의 산란이나 먼지, 수증기 등에 의해 나타나며, 그 결과는 영상에서 대상 물체의 채도 감소로 나타난다. 따라서 대기원근법의 역으로 2차원 영상에서 채도변화에 따른 깊이지각이 가능하다.

채도정보에 의한 깊이지도를 얻기 위해 입력된 RGB영상을 HSV space로 변환하여 saturation image를 구하고, 자연영상의 노이즈를 줄이기 위해 가우시안 필터링을 수행한 결과에 K-means clustering하여 다단계의 saturation depth map을 구성한다.

그러나 HSV color model에서 서로 다른 색상(hue)에서 같은 채도 값을 가질 수 있기 때문에 채도깊이는 전체 이미지에 대한 깊이를 나타낸다고 볼 수 없다. 따라서 유사색상영역의 클러스터에 대한 채도깊이를 구하고 각 색상영역의 기하학적 깊이에 따라 채도깊이를 재구성할 필요가 있다.

이를 위해 hue image를 구하고, K-means clustering으로 색상영역에 대한 segmentation을 수행하여 다단계의 hue range map을 구한다. hue range map은 색상영역간의 기하학적 깊이를 구하여 ordering하고, 각 색상영역의 order value를 해당되는 픽셀의 채도깊이에 가중치로 적용하여 saturation depth map을 재구성한다.

그림 1에서 보는 바와 같이 채도정보만으로 구성된 깊이지도 영상(b)은 유사색상영역에서 거리에 따른 채도차가 반영된 깊이지도를 구성하지만, 연결된 다른 색상영역의 깊이를 반영하지 못하고 있다. 그러나 (c)의 색상영

2 채도 정보에 의한 깊이지도 생성

역의 기하학적 깊이순서를 가중치로 적용한 (d)영상에서는 (b)에서의 문제가 개선된 것을 확인할 수 있다.

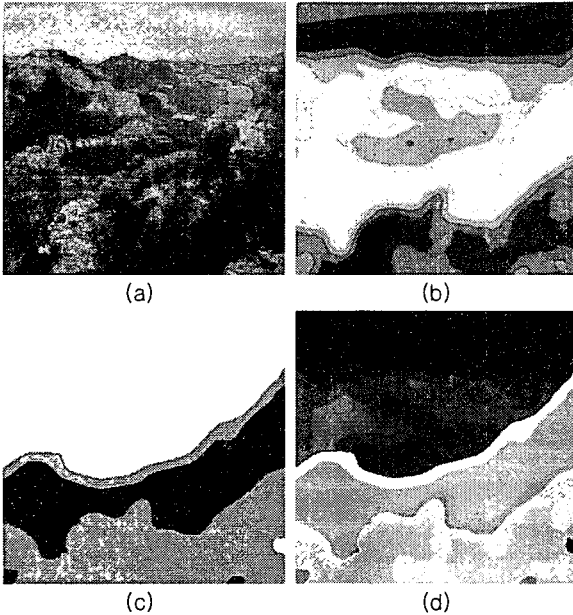


그림1. (a) 원본영상, (b) 채도(saturation) 깊이지도 (K=6), (c) 색상(hue) 영역 segmentation (K=3), (d) (c) 의 기하학적 순서를 (b)에 가중치로 적용한 결과

3. 디포커스 정보에 의한 깊이지도 생성

특정 물체에 초점을 맞추어 단안 카메라로 촬영했을 때 결과영상에서 피사체와 피사체 주변의 피사계심도(depth of field)에 포함되는 영역은 선명하게 나타나지만 배경을 포함한 그 외의 영역은 흐릿하게 나타난다. 이러한 특성을 이용하여 초점이 맞은 영역과 맞지 않은 영역간의 상대적 깊이를 얻을 수 있다.

RGB영상의 HSV space 변환을 통해 얻은 intensity value 영상으로부터 각 픽셀의 4방향 주변 픽셀과의 intensity value 차를 구하고 평균을 계산하여 spatial frequency 이미지를 구한다. 노이즈를 줄이기 위해 가우시안 필터링을 수행한 결과 영상에 대해 K-means clustering하여 다단계의 깊이지도를 구성한다.

그림2에서 보는 바와 같이 초점이 맞은 피사체를 포함한 피사계심도 영역은 초점이 맞지 않은 배경 영역에 비해 더 가까운 영역으로 나타난 것을 확인할 수 있다.

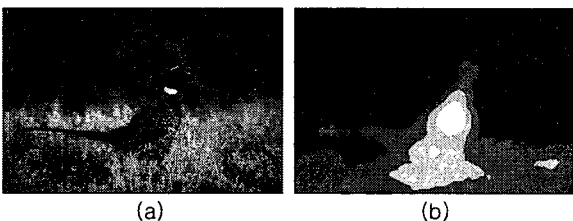


그림2. (a) 원본영상, (b) defocus 깊이지도 (K=4)

4. 기하학적 깊이

자연영상에서 물체는 영상의 하단경계에서 horizon boundary까지 근거리에서 원거리로 배치된다. 그리고 horizon boundary에서 영상의 상단경계까지의 범위는 하늘영역으로 판단할 수 있다. 이 기하학적 깊이정보는 앞서의 채도 깊이지도와 디포커스 깊이지도의 결과를 보완하는 수단으로 이용될 수 있다.

일반적으로 단안영상에서 하늘 영역은 낮은 공간주파수를 가지기 때문에 상단 경계에 접한 low frequency 영역은 하늘로 판단할 수 있으며, spatial frequency가 증가하기 시작하는 지점은 horizon boundary가 된다.

horizon boundary를 구하기 위해 앞서 구한 디포커스 깊이지도에서 low frequency 영역을 pixel by pixel로 영상의 상단경계로부터 구한다. 각 column마다 spatial frequency가 threshold를 넘어서는 경계의 y축 평균 좌표로 horizon boundary를 측정한다.

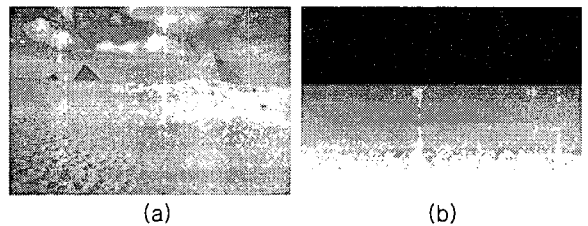


그림3. (a) 원본영상, (b) 기하학적 깊이

5. 결론 및 향후연구

실험결과 각각의 단안단서에 의한 깊이지도로부터 실제 사람이 단안영상으로부터 얻을 수 있는 깊이 지각과 유사한 깊이 정보를 얻을 수 있음을 확인하였다.

단, 채도 깊이지도는 색상이 풍부한 영상일수록 부정확한 결과가 도출되었는데, 이는 자연영상의 특성상 색상(hue) 정보가 일정치 않기 때문에 색상에 따른 채도변화의 정확한 판단이 어려운 것으로 판단할 수 있다. 향후 연구에서는 채도의 감소와 밝기, 색상변화 등에 대한 좀 더 구체적인 관계 연구가 필요할 것으로 보인다.

디포커스 정보에 의한 깊이지도는 카메라의 조리개 값에 따라 정확성에서 차이를 보였다. 이는 조리개 값에 따라 피사계심도(depth of field)의 범위가 변하기 때문이다. 이에 대한 연구와 더불어 고충빌딩과 같이 강한 에지가 많은 영상의 spatial frequency 분포에 대한 연구도 추가적으로 필요할 것이다.

단안단서들은 입력영상에 따라 각각 중요도가 다르게 나타날 수 있기 때문에 어느 한 가지의 단안단서만을 사용해서는 안 되며, 두 가지 이상의 단안단서를 복합적으로 적용해야 한다. 단안영상의 특성에 따라 각 단서의 가중치를 다르게 부여해야 보다 정확한 상대적 깊이지도 결과영상을 얻을 수 있을 것이다.

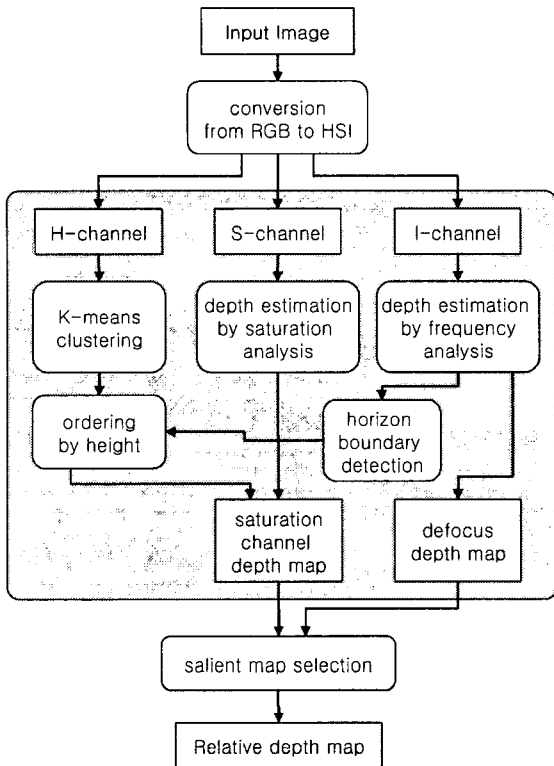


그림4. 단안단서에 의한 깊이지도 생성 시스템

위의 그림4는 단안단서에 의한 깊이지도 생성 시스템을 나타낸다. 입력영상은 RGB channel에서 HSI channel로 변환된다. intensity channel은 frequency analysis를 거쳐서 defocus depth map을 구하게 된다.

영상의 hue channel을 K-means clustering에 의해 segmentation하고, intensity channel의 frequency analysis 결과로부터 얻은 horizon boundary를 참고하여 영상에서의 height를 기준으로 한 hue channel segment간의 order를 얻는다.

saturation channel에 대해 saturation analysis를 수행하고, 그 결과에 앞에서 구한 hue channel segment order를 반영하여 saturation depth map을 구한다.

생성된 두 가지의 depth map중 적절한 depth map을 선택하여 최종 relative depth map을 구한다.

참고문헌

- [1] J. J. Gibson, "Perception of the Visual World", Houghton Mifflin, Boston, 1950.
- [2] T. Darrell and K. Wohn, "Pyramid based depth from focus", Proc. IEEE CVPR, pp.504 - 509, June 1988.
- [3] A. Pentland, "A new sense of depth of field",

Proc. IEEE Trans. PAMI, pp.523-531, 1987.

[4] D. Marr, "Vision: a computational investigation into the human representation and processing of visual information", W.H. Freeman, 1982.

[5] E. Bruce Goldstein, "Sensation and Perception", Wadsworth Publishing Company, 2003.

[6] J. E. Cutting, "How the eye measures reality and virtual reality", Behavior Research Methods, Instruments & Computers, 29(1), 27-36, 1997.