

반자동 실내 3D 공간 구축을 이용한 사람 크기 예측

길종인 김만배

강원대학교 컴퓨터정보통신공학과

{jigil, manbae}@kangwon.ac.kr

Human Size Estimation via Semi-automatic Indoor 3D Space Layout

Jong-in Gil Manbae Kim

Dept. of Computer and Communications Eng., Kangwon National University

요약

사람 검출 시스템은 카메라의 위치 및 각도 등에 큰 영향을 받는다. 이로 인해 획득한 2D 영상에서 사람은 위치에 따라 각기 다른 크기를 갖는 형태로 나타난다. 이러한 요인들은 사람 검출 시스템의 실시간 구현을 어렵게 만드는 요인이 된다. 본 논문에서는 실내 공간의 구조를 깊이맵으로 구성하여, 이로부터 3D 공간을 구성한다. 3D 공간에서는 어느 위치에서든지 사람의 크기가 일관되므로 이를 2D 영상으로 투영하게 되면 2D 영상의 좌표에 따른 정확한 사람의 크기를 추정할 수 있다. 실험 결과로부터 제안 방법의 타당성을 입증하였다.

1. 서론

실내환경의 사람 검출 방법은 많은 연구가 이루어지고 있고, 나날이 발전하고 있다. 주로 기계학습을 통해 분류기를 생성하고, 이 분류기가 사람인지 아닌지를 판단하는 방식으로 처리하고 있다. 그러나 이러한 알고리즘은 수많은 연산을 수행해야 하기 때문에, 실시간 처리에 많은 어려움을 가지고 있다. 그러므로, 알고리즘의 복잡도를 낮추기 위한 연구가 꾸준히 지속되고 있으나, 복잡도가 낮아지면 성능 또한 낮아지기 때문에 많은 어려움이 존재한다. 기존의 분류기를 통한 기계학습 기법은 주로 적응적 윈도우로 영상의 각 탐색영역을 모두 살펴가며 사람을 탐지하는 방식을 사용하였다. 그러나 이러한 방법은 이미지 피라미드 영상해상도 조절 및 탐색윈도우의 크기 조절 등 많은 불필요한 연산을 수행하게 되므로 효율적이지 못하다. 영상내의 위치에 따라 사람은 제각기 다른 크기를 갖게 된다.

본 논문에서는 수작업으로 깊이맵 및 3D 공간을 구성한 후에, 2D 영상을 3D 공간에 역투영함으로써 2D 영상에 존재하는 사람의 크기를 측정하고자 한다. 비록 2D 영상에서는 좌표마다 사람의 크기가 달라지지만 3D 공간에서는 어느 위치에서는 실제 객체의 크기는 변하지 않는다는 현상을 활용한다. 이를 다시 2D 영상으로 투영한다면 정확한 사람의 크기를 예측할 수 있다.

2. 깊이맵 생성

2D 영상 좌표에서 사람의 크기를 예측하기 위해서는 먼저 3D 공간에 대한 깊이맵을 필요로 한다. 이를 위해 먼저 깊이맵을 생성하는 과정이 필요하다. TOF 카메라와 같은 깊이 카메라를 이용하여 깊이맵을 획득할 수도 있지만, 깊이 카메라를 이용하게 되면 내부에 존재하는 책상 및 의자와 같은 여러 가지 물체들로 인해 깊이맵으로부터 장소의

구조를 파악하기 어려워진다. (그림 1) 본 논문에서 제안하는 사람의 크기 추정문제를 위해서는 이러한 불필요한 객체는 모두 제거하고, 내부 실내의 공간정보가 획득해야 한다.

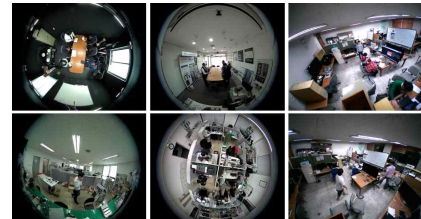


그림 1. 카메라 위치에 따른 다양한 2D영상

깊이맵을 생성하기 위해 본 논문에서는 반자동 방법인 폴리곤 분할 기법을 이용하여 깊이맵을 생성하였다. 폴리곤 분할 기법은 Delaunay 삼각화[1]와 Gouraud 셰이딩[2]의 두 단계로 이루어져 있다. 실내환경은 다수의 벽과 천장 및 바닥이 모여서 공간을 구성한다. 이러한 평면은 폴리곤의 형태로 나타낼 수 있다. 따라서 여러 개의 폴리곤을 생성하여, 각 폴리곤의 깊이를 설정하면 전체 구조의 깊이를 설정할 수 있다. 먼저 카메라로부터 입력받은 영상으로부터 구조를 파악한 후, 천장, 바닥 및 벽의 위치를 파악하여 이를 폴리곤으로 생성함으로써 결정할 수 있다. 각 폴리곤을 구성하는 정점(vertex)들에 깊이를 할당하면 Gouraud 셰이딩을 이용함으로써 폴리곤의 내부의 모든 픽셀에 대해 깊이를 할당할 수 있다. 그림 2는 입력 영상으로부터 폴리곤을 구성하고 각 폴리곤을 구성하고 있는 정점을 노란색으로 표시하였고, 이로부터 Delaunay 삼각화 및 Gouraud 셰이딩을 수행하여 획득한 깊이맵을 보여준다.



그림 2. 입력 영상으로부터 생성된 깊이맵

3. 사람 크기 추정

먼저 깊이맵으로부터 Ground Plane(GP)을 추출한다. 기존에 GP를 검출하는 몇몇 알고리즘이 연구되고 있으나 [3, 4], 본 논문에서는 GP는 이미 검출되었다고 가정한다. GP 검출 이외에 카메라 캘리브레이션도 미리 수행되어야 한다. 카메라 캘리브레이션으로부터 내부 및 외부 파라미터가 얻어졌다면 월드 좌표계가 구성되었다고 할 수 있다. 이때, 월드 좌표계는 외부 파라미터를 추정하는 단계에서 이용자의 편의대로 설정이 가능하다. 다음 그림 3은 깊이맵과 깊이맵으로부터 추출한 GP, 그리고 구성된 월드 좌표계의 예를 보여준다.

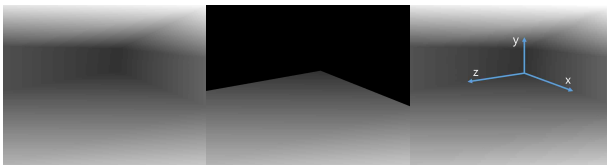


그림 3. Ground plane과 월드 좌표계.

이제 좌표계의 원점과 x축, y축 그리고 z축의 방향을 알게 되었으므로, 3D 공간상에서 자유자재로 객체를 위치시키는 것이 가능해졌다. 이때, 앞서 언급하였듯이 사람은 항상 발을 ground에 지지한채로 서있다. 앉아있는것과 누워있는 등의 자세는 고려하지 않고 오로지 서있는 상태만을 고려한다. 따라서, ground plane으로부터 3D 공간상에 사람을 위치시키는 것이 가능하다. 이를 위해서 먼저 2D 영상의 GP에 해당하는 픽셀의 좌표로부터 3D 공간상의 좌표를 계산해야 한다. 3D 공간상의 점들을 다음 식 (1)의 과정을 통해 모델링한 변환관계를 이용함으로써 3D의 한 점 (X, Y, Z)를 2D 영상의 점 (u,v)에 투영시킬 수 있다.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R| - Rt] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

여기서 R은 rotation matrix, t는 translation vector이다. 그리고 내부 파라미터 K와 외부 파라미터 R, t를 식 (2)와 같이 하나로 결합하여 투영 행렬(Projection Matrix) M으로 표현한다면 식 (3)과 같이 변경할 수 있다.

$$M = K[R| - Rt] = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

위와는 반대로 (u,v)를 3D (X,Y,Z)의 한점으로 변환하기 위해서는 역투영(Inverse projection)을 수행함으로써 월드 좌표계의 좌표를 찾

을 수 있다.

만일 GP의 어떠한 한 픽셀이 3D 공간상의 한 점으로 매핑되었다면, 해당 점을 ground point로 설정하고, 이를 기준으로 하여 cube를 생성한다. 이때, cube는 Δx, Δy, Δz로 크기를 설정할 수 있으며, 물론 Δy는 사람의 키에 해당한다고 말할 수 있다. Δx, Δy, Δz는 cm 단위로 써, 사람의 크기에 적절하게 미리 설정한다. 이렇게 생성된 cube를 사람으로 가정한다. 본 논문에서는 Δx=30, Δy=180, Δz=30으로 설정하였다. 하나의 ground point로부터 cube를 생성함으로써 cube의 정점에 해당하는 8개의 점 P_i를 추출할 수 있다. 또한 생성된 cube의 중심점도 추출할 수 있다.

4. 실험 결과

사람의 크기는 세로 너비와 가로 너비가 다르기 때문에 Euclidean 거리를 측정하여 값을 할당하였다. 이러한 사람의 크기는 ground plane으로부터 계산하였으므로, ground plane으로부터 멀리 위치한 천장과 같은 곳에서는 사람의 크기가 정의되지 않음을 알 수 있다. 즉 사람의 크기가 정의되지 않는곳에는 사람이 존재할 수 없음을 의미한다. 그림 3에서는 제안 방법을 이용한 실험 결과를 보여주고 있다.



그림 3. 실험 결과. 노란색으로 경계 상자 설정

5. 결론

본 논문에서는 2D 영상의 Ground Plane을 3D 공간에 역투영 한 후, 이로부터 사람을 가정하는 cube를 생성하고 이를 다시 2D 영상으로 투영함으로써 2D 영상의 위치에 따른 사람의 크기를 측정하는 방법을 제안하였다. 실험 결과를 통하여 본 논문에서 제안하는 방법의 타당함을 입증하였다.

참고 문헌

[1] Cignono, P., Montani, C. and Scopigno, R., "DeWall: A fast divide and conquer Delaunay triangulation algorithm in Ed," Computer-Aided Design, 30(5), 33-341, 1998.
 [2] Gouraud, H., "Continuous shading of curved surfaces," In Seminal graphics. ACM. pp. 87-93, 1998, July.
 [3] S, Choi., J, Park., J, Byun. and W, Yu., "Robust ground plane detection from 3D point clouds," 14th Intl. Conference on Control, Automation and System, pp. 1076-1081, 2014, October.
 [4] Kircali, D., Tek, F. B. and Iyidir, I. K., "Ground Plane Detection using Kinect sensor," 21st Signal Processing and Communications Applications Conference, pp. 1-4, 2013, April.