RGBD 카메라를 이용한 실내에서의 물체 검출 알고리즘

허선, 이상화, *김명식, *한승범, 조남익
서울대학교, *삼성전자
hsfra111@ispl.snu.ac.kr, lsh529@snu.ac.kr, *ms.1.kim@sanmsung.com,
*sb85.han@samsung.com, nicho@snu.ac.kr

Indoor object detection method using a RGBD image

Seon Heo, Sang Hwa Lee, *Myung Sik Kim, *Seung Beom Han, Nam Ik Cho
Seoul National University, *Samsung Electronics Co., Ltd.

요 약

본 논문에서는 실내에서 RGBD 영상을 이용하여 물체를 검출하는 방법을 제안한다. 특정 물체가 아닌 일반적인 여러 가지 물체에 대한 특징을 규정하기 어려우므로 본 논문에서는 영상 정보에 의존하기 보다 물체와 픽셀의 기하학적 구조에 기반하여 물체를 검출한다. 우선 컬러 정보를 이용하여 대략적인 영상 영역분할을 하고 이를 같은 레이블로 분류하여 물체와 배경의 후보를 얻는다. 대체로 실내 환경에서 바닥은 평면이라 가정할 수 있으므로 바닥의 평면 모델을 만들어서 물체 후보에서 이를 제외시킨다. 또한, 물체에 대한 간단한 가정을 통해 바닥 이외의 배경 역시 물체와 구분하여서 물체 후보들을 가려낸다. 최종적으로 3 차원 공간에서 가까이 위치하는 레이블을 하나로 통합하는 과정을 통해 최종적인 물체 영역을 검출하고 이를 bounding box 로 표시한다. 직접 촬영한 몇몇 실내 RGBD 영상에서 실험한 결과, 제안하는 방법이 기존 방법들에 비해 물체 검출 성능이 좋은 것을 확인하였다.

1. 서론

물체 검출은 컴퓨터 비전의 주요 분야인 검출 알고리즘을 상위 레벨에서 진행하는 것으로서, 하위 레벨 검출 방법인 특징점 검출보다 여러 가지 이유로 많은 어려움이 있다. 우선 물체마다 모양과 크기, 내부 텍스처가 다르고, 여기에 물체와 카메라 사이의 각도와 거리에 따라 같은 물체도 다르게 보여지기 때문에 일반적인 물체의 특징을 특정하기가 어렵다. 따라서 대부분의 물체 검출 방법은 물체의 종류를 한정하고, 제한된 크기와 시점에서 본 것을 가정을 하고 진행하게 된다. 대표적인 예가 P. Viola 와 M. Jones [1]가 제안한 boost 방법으로 영상에서 얼굴 검출을 빠르고 정확하게 해내었다. 하지만 이렇게 물체의 종류가 한정되어 있고, 학습에 기반한 물체 검출 방법은 일반적인 물체 검출에 적용하기에 한계가 명확하다. 물체의 종류는 무한히 많고, 이들을 모두 학습할 수 없기 때문이다. 다른 물체 검출 방법으로는 salient region 을 검출하는 방법 [2], [3]이 있다. 이 방법은 영상에서 salient region 이 물체라고 가정하고 이들을 찾는 방법이다. 특정 물체에 대하여 학습할 필요 없이 일반적인 물체를 찾아낼 수 있지만 영상에서 물체가 한 개가 아니라 다수라면 물체의 경계를 알 수 없고 몇 개의 물체가 있는지 모른다는 단점이 있다. 다른 한편으로, 물체 검출을 정확히 하기보다 물체에 대한 많은 “ proposal” 을 결과로 내주는 방법 [4]이 있다. 이 방법은 특정 물체에 대하여 학습이 되어 있을 때, 이를 검출하기 위해 영상을 전역 탐색 해야 하는 어려움이 있기 때문에 이를 보완하고자 탐색 범위를 proposal 로만 줄이는 것이 목적이다. 결과로 proposal 의 신뢰도가 대부분 같이 나오므로 이를 이용하여 일정 문턱치 이상의 신뢰도를 갖는 proposal 을 물체 검출의 결과로 생각하고 사용할 수 있다. 하지만 이러한 접근의 방법들은 proposal 의 정확도보다는 물체에 대한 recall 을 중요시하기 때문에 무수히 많은 false positive 가 나올 수 있는 단점이 있다.

본 논문에서는 특징 추출보다는 물체와 배경의 기하학적 구조에 기반한 일반적인 물체 검출 방법을 제안한다. 일반적인 물체에 대한 특징을 쉽게 정할 수 없기 때문에, 배경 모델을 만들고 이 모델에 맞지 않는 부분을 물체로 인식하게 된다. 이를 위하여 본 논문에서는 최근 많이 연구되고 있는 depth sensor 를 사용하여 depth 정보를 추가로 얻고, 이 정보를 쉽게 얻을 수 있는 환경인 실내 환경에서의 영상으로 실험 환경을 제한한다. 실내 환경에서는 바닥이 평면이라 가정할 수 있기 때문에 배경에 대한 모델도 간단한 평면 모델로 추정할 수 있다.

2. 제안하는 방법

모든 물체를 학습할 수 없기 때문에 일반적인 물체 검출에는 학습 기반의 방법을 적용하기 어렵다. 본 논문에서는 일반적인 물체 검출을 위하여 크게 3 단계를 거치는데, 대략적인 영상 영역 분할(rough segmentation), 배경 추출, 물체 영역 다듬기(refinement) 단계가 그것이다. 본 논문에서 자주 사용되는 기호를 정리하면, 깊이 영상을 D , i 번째

픽셀을 $p_i=(x_i,y_i)$ 라 표기하기로 한다.

같은 물체를 나타내는 픽셀들은 대체로 서로 컬러가 유사하고 공간상에서 비슷한 영역에 존재한다. 따라서 비슷한 성질을 가지는 픽셀을 그룹화하여 물체의 후보들을 만드는 rough segmentation 과정을 우선 실행한다. 영상이 몇 개의 레이블로 구분될 지 모르는 상황에서 빠르게 레이블링을 하기 위해서 잘 알려진 mean shift clustering 방법 [5]을 사용한다. Depth 카메라를 사용하기 때문에 p_i 에 대응하는 3 차원 상의 점 q_i 를 다음과 같이 알 수 있다.

$$q_i=(X_i,Y_i,Z_i)=\left(\frac{D(p_i)x_i}{f},\frac{D(p_i)y_i}{f},D(p_i)\right)$$

여기서 f 는 카메라의 초점거리이다. 하지만 depth 카메라의 한계로 인하여 depth 정보를 모르는 픽셀이 존재하고, 알고리즘의 빠른 수렴을 위하여 mean shift clustering 을 실시하는 피쳐 공간은 컬러와 이미지 좌표만 사용하는 5 차원 공간이다. Mean shift clustering 을 통하여 만들어진 label set

$$L=\{l_1,\cdots,l_n\}$$

이 기본적으로 물체의 후보가 되는데, l_i 에 속하는 모든 픽셀의 집합을 M_i 라 하면, 안정성을 위하여 임의의 i 에 대하여 M_i 의 원소의 수가 일정 수보다 적으면 l_i 를 L 에서 제외시키므로,

$$|M_i|>N \quad for \quad \forall i$$

를 만족한다. 여기서 N 은 미리 정해진 상수 값이다.

그 다음, 영상에서 배경을 분리해내어 물체의 후보를 줄이게 되는데, 우선 배경 중 바닥 영역을 추출한다. 대부분의 물체가 바닥에 놓여 있으므로 물체 검출을 위하여 바닥의 추출은 중요한 과정이기 때문이다. 앞서 구한 레이블을 분석하여 바닥 영역의 후보 레이블 집합 $L_f\subset L$ 를 구하는데, 만약 $l_i\in L_f$ 라면 다음의 2 가지 조건을 만족해야 한다.

$$\begin{aligned} y&=h && for \quad \exists p\in M_i \\ |D(q_u)-D(q_v)|&>T_1 && for \quad \forall p_u,p_v\in M_i \end{aligned}$$

여기서 h 는 영상의 세로 길이이고, T_1 은 정해진 threshold 이다. 그리고 다음과 같이 픽셀들의 집합 $M_f=\{M_i|\forall i,\ l_i\in L_f\}$ 가 정의된다.

이렇게 L_f,M_f 를 선정하고 난 후, 바닥 모델을 만든다. 실내에서 바닥은 대부분 평면이므로 평면 모델

$$\vec{n}\cdot q=d,\quad \|\vec{n}\|=1$$

으로 바닥을 모델링 할 수 있다. 여기서 $\vec{n}$ 은 평면의 법선 벡터이고, d 는 상수이다. M_f 에 속하는 임의의 세 픽셀을 뽑고 이들의 이미지 좌표와 depth 값을 이용하여 이 점들의 3 차원 위치를 얻는다. 이렇게 공간에 투영된 세 점을 이용하면 평면이 유일하게 결정되는데, 이 평면이 바닥 모델의 후보가 된다. 최종 평면 후보의 선정은 RANSAC 방법을 이용하여 얻는다. M_f 에 속하는 픽셀 p 의 3 차원 투영 점 q 가

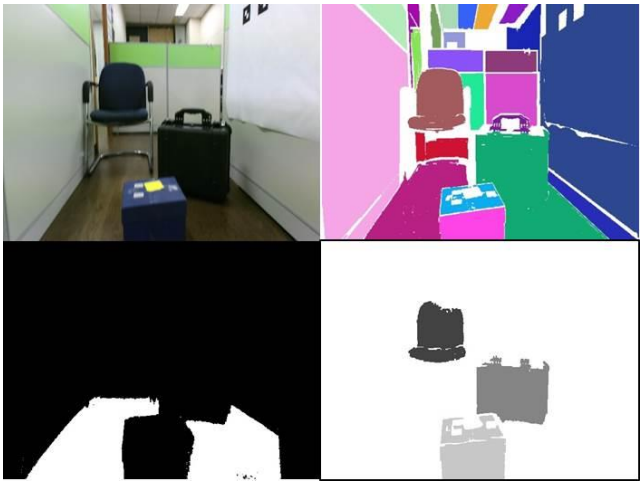

그림 1. 왼쪽 위부터 시계 방향으로 입력 RGB 영상, rough segmentation 결과, 최종 물체 검출, 바닥 영역 검출의 예시다.

만들어진 평면 후보 $(\vec{n},d)$ 와 일정 거리 내에 있으면, 즉,

$$dist(q,(\vec{n},d))=\frac{|\vec{n}\cdot q-d|}{\|\vec{n}\|}=|\vec{n}\cdot q-d|<T_2$$

이면, 그 p 는 $(\vec{n},d)$ 의 inlier 가 되고, 가장 inlier 가 많은 평면이 최종 평면 후보가 된다. 최종 평면 후보의 inlier 들의 집합을 C 라고 하면, 다음과 같이 행렬

$$A=\begin{pmatrix}\vdots & \vdots\\ q_i^T & -1\\ \vdots & \vdots\end{pmatrix}\quad for\quad \forall p_i\in C$$

를 정의할 수 있고, 다음 식

$$\begin{aligned}&\arg\min_{\vec{n_f},d_f}\left\|A\begin{pmatrix}\vec{n_f}\\ d_f\end{pmatrix}\right\|\\ &s.t.\quad \|\vec{n_f}\|=1\end{aligned}$$

을 풀면, 최종 평면 $(\vec{n_f},d_f)$ 를 얻을 수 있다. 위 식의 해는 A 를 singular value decomposition 한 후, 가장 작은 singular value 에 대응하는 right singular vector 를 통해 얻을 수 있다. 영상의 모든 픽셀을 바닥 모델 $(\vec{n_f},d_f)$ 에 부합하는지 시험하여, 부합하면 그 픽셀은 바닥 영역으로 구분하게 된다.

바닥 이외의 배경은 물체에 대한 간단한 가정을 통해 이에 맞지 않은 영역들로 찾는데, 물체는 영상의 가운데 부분에 있고 카메라와 가까운 거리에 있으며 물체는 적당한 크기여서 같은 물체 내 depth 변화가 작다고 가정한다. 바닥에 속한 픽셀을 모두 제외하고 connected component 를 이용하여 새롭게 label set L^{new} 과 픽셀 set M^{new} 을 만들고, 이 가정에 맞지 않는 label 은 배경으로 간주하여 물체 후보에서 제외한다.

남아 있는 label 은 모두 물체 후보가 되고, 같은 물체가 여러 개의 label 로 나뉘어졌을 경우, 이를 하나의 label 로

그림 2. 위에서부터 차례대로 [4], [3], 제안하는 방법의 물체 검출 결과이다.

통합하는 과정을 거친다. 2 개 label 사이의 거리를

$$dist(l_i, l_j) = \min_{q_u \in M_i^{new}, q_v \in M_j^{new}} \|q_u - q_v\| \quad for \quad \forall l_i, l_j \in L^{new}$$

으로 정의하고, 거리가 가까운 label 은 같은 물체로 간주하여 하나로 통합한다. 마지막으로 각 label 에 대한 bounding box 를 구하여 이를 물체 검출의 결과로 출력한다.

3. 실험 결과

제안하는 방법을 직접 촬영한 몇 개의 RGBD 영상에 대해 실험하여 물체를 검출하였다. 영상은 VGA 해상도($w = 640,$ $h = 480$)를 가지고, 실내에서 촬영되었으며 depth sensor 로는 Kinect version 2 를 이용하였다. 모든 실험에는 같은 파라미터를 사용하였으며 그 값은 $N = \frac{wh}{300}, T_1 = 100, T_2 = 0.5$ 이다. 제안하는 방법은 물체가 많은 복잡한 영상에서도 물체 검출을 잘 하고, VGA 해상도 영상에 대해 일반적인 PC 에서 초당 1~2 프레임 정도의 속도로 동작한다. 또한, 비교를 위하여 기존의 방법인 물체 proposal 을 찾아주는 방법 [4]과 salient 영역을 찾아주는 방법 [3]을 같은 실험 영상에 대하여 실험하였다. [4]의 결과는 원래 물체 proposal 을 내주는 것이기 때문에 false alarm 이 많이 나오며, [3]의 결과를 보면 salient 영역에 물체가 대체로 존재하지만 다수의 물체를 구분하기에는 어려움이 있다. 실험 결과 예시는 그림 2 에서 보여주고 있다.

4. 결론

본 논문에서는 실내 환경에서 RGBD 영상을 이용하여 일반적인 물체에 대한 검출 방법을 제안하였다. 일반적인 물체를 표현하는 특징을 알기 어렵기 때문에 물체와 배경에 대한 기하학적 구조를 파악하여 물체를 검출하는 방법을 제안하였는데, 먼저 컬러와 영상 좌표만을 이용하여 영상을 대략적으로 구분하는 영상 영역 분할(rough segmentation)을 하고, 바닥에 대한 평면 모델을 추측하여 바닥 영역을 구하였다. 그 다음, 물체에 대한 간단한 가정을 통하여 나머지 배경과 물체를 분리하고, 마지막으로 3 차원 공간의 거리를 이용하여 물체 검출 영역을 다듬는(refinement) 과정을 수행하였다. 제안하는 방법을 이용하면 실내 환경에서의 일반적인 물체 검출에 있어서 기존의 방법들보다 좋은 성능을 보이는 것을 실험을 통해 확인하였다. 또한, 이 방법은 로봇 청소기의 주행경로 설정 등 다른 알고리즘의 전처리 과정으로 사용될 수 있을 것이다.

감사의 글

본 연구는 삼성전자 (DMC 연구소 Frontier Research Lab) 의 지원을 받아서 수행하였음.

참고문헌

[1] P. Viola and M. Jones, “ Rapid object detection using a boosted cascade of simple features,” In CVPR, IEEE, 2001.

[2] T. Liu, Z. Yuan, J. Sun, J. Wang and N. Zheng, “ Learning to detect a salient object,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 33, no. 2, pp. 353-367, 2011.

[3] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng and S. Li, "Salient object detection: A discriminative regional feature integration approach," in CVPR, IEEE, 2013.

[4] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in ECCV, Springer, 2014.

[5] D. Comaniciu and P. Meer, “ Mean shift: A robust approach toward feature space analysis,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603-619, 2002.