Perceiving People from a Low-lying Viewpoint

Armando Pesenti Gritti[*] Oscar Tarabini[*] Alessandro Giusti[†] Jerome Guzzi[†]
Gianni A. Di Caro[†] Vincenzo Caglioti[*] Luca M. Gambardella[†]
[*] DEIB, Politecnico di Milano, Italy [†] IDSIA, USI/SUPSI, Lugano, Switzerland

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding; I.4.8 [**Image Processing And Computer Vision**]: Scene Analysis

Keywords

Robot perception, Machine learning, Person tracking

1. VIDEO DESCRIPTION

Reliable perception and tracking of nearby humans in unstructured environments is one of the main issues to be solved in order to enable effective human-robot interaction and sharing of common spaces [5].

RGB-D sensors, such as the Microsoft Kinect or Asus Xtion, are an ideal tool for solving this problem indoors, and many human-tracking algorithms have been developed in the last few years. These algorithms are based on a number of assumptions: the subject is expected to lie at a given minimum distance from the sensor, such that at least part of its upper body is visible – most importantly the head, which represents an easy-to-detect seed for human detection and segmentation [6]; in some cases, the human is expected to face the sensor, and the sensor is assumed to be still [2]; the viewpoint is always assumed to lie in an elevated position, at least at the level of the subjects' chest or eyes.

Small-footprint ground robots must by necessity be short, with sensors lying close to the ground. Detecting humans from such a viewpoint is an hard and mostly unexplored problem: occlusions are frequent and severe; unless the subject is far from the robot, only legs are visible, exhibiting a complex, highly-variable appearance with irregular motion and no obvious markers for detection; when the sensor is mounted on a mobile platform, background-subtraction techniques can not be used unless very reliable ego-motion estimates are available. In this work we demostrate human detection and tracking under these challenging conditions, by means of a low-lying, forward-pointing RGB-D sensor, considering the TurtleBot [3] as an example platform.

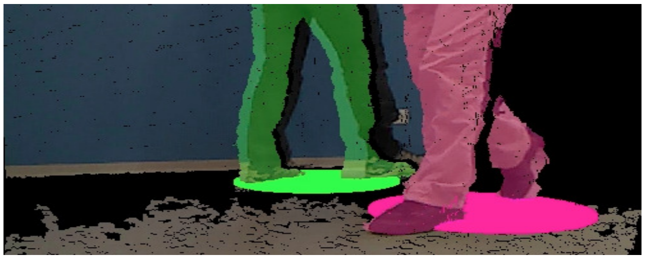

Figure 1: Tracked position (colored circle) and segmented silhouette of two people in a video frame.

The video illustrates the different steps of our algorithm using real-world datasets, and showcases results in various complex, realistic scenarios. Initially, the point cloud is downsampled [1] and the floor is detected as the most prominent horizontal plane; then, we search for clusters of points emerging from the floor, which comply with reasonable foot dimensions, and we expand them to knee level to get a set of candidate legs. Because in realistic environments most of such candidates are not actual legs, we apply an SVM classifier based on Histogram of Oriented Gradients features, which returns the probability that a candidate is a leg. Such classifier was previously learned from a large, manually-labeled training dataset comprising more than 2200 frames shot in 8 different, cluttered real-world environments. In such dataset, 6839 candidates were present of which 2413 were legs. Finally, individual legs and then people are tracked using a Kalman filter [4], accounting for the fact that a person has two legs, but often one or both may be occluded or fused to a single detection. The resulting output is the position and the velocity of each visible person.

The system currently runs at 10 FPS under MATLAB; we are currently working towards a realtime, reuseable implementation as a ROS node.

References

[1] The point cloud library. `http://pointclouds.org/`.
[2] Primesense nite middleware. `http://bit.ly/IAIHRb`.
[3] The turtlebot robot development kit. `http://bit.ly/lqRgND`.
[4] Y. Bar-Shalom et al. The probabilistic data association filter. *IEEE Control Systems*, 29(6):82–100, 2009.
[5] J. Guzzi et al. Bioinspired obstacle avoidance algorithms for robot swarms. In *Proc. of BIONETICS*, 2012.
[6] L. Xia, C.-C. Chen, and J. Aggarwal. Human detection using depth information by kinect. In *Proc. of CVPR Workshop on Human Activity Understanding from 3D Data (HAU3D)*, pages 15–22, 2011.

Human-Robot Interaction 2014 Bielefeld University