People Tracking

Constantin Schieber, 1228774, Technische Universitaet Wien

December 18, 2016

This paper summarizes and discusses current research in the field of people tracking. First, the established hardware, its limitations and possible combinations are presented. Second, the widely used Robotic Operating System (ROS) and its features of a distributed approach are introduced.
Then a detailed discussion on tracking of people with RGB-D hardware in a ground plane, indoor enviroment is done. The obtained results promise a good foundation for future work.

1 Introduction and Problem Statement

Detection and tracking of people is an important feature for many applications. Especially in the sphere of mobile robotics - where safe interactions with people are a basic requirement in any situation. But particularly in mobile applications there are several limiting constraints as e.g. computing power, field of view and time for decision making.
With the improvement of hardware in the relevant technology sector (e.g. RGB-D cameras, Laser, Thermal View) the number of papers focusing on this topic increased too.
Reliable tracking of multiple persons that are partially blocked is possible with RGB-D camera networks that are set up prior to usage in a room in combinations with solutions like OpenPTrack [1].
The same principles and hardware can be used in mobile applications as in [2]. A different approach is tracking by a combination of laser and thermal view as in [3].

This paper intents to highlight the currently used hardware, the Robot Operating System (ROS) as a framework for the hardware and the basic idea of currently used detection and tracking approaches.

1.1 Used Hardware

RGB Sensors won't be used on their own as they only deliver depth information when used in a stereo image approach and are usually combined with a depth sensor. This combination provides good enough performance to enable resource efficient and prompt computation of the environment.
Laser sensors provide accurate depth readings and a wide field of view but therefore depend on body features like legs for proper detection. They are well suited for close range following and tracking tasks as they provide accurate readings on close distances, in contrast to Depth Sensors [3].
Thermal sensors provide readings that can be interpreted accurately when the targets of interest have a distinct temperature from the rest of the environment [4].
Sonar sensors require an active counterpart on the person that is to be tracked. Therefore it is only suitable for tracking of single persons but works well in the outside.

The following Table 1 shows a set of the currently utilized hardware for detection and / or tracking of people.

Sensor-type	Depth Information	Works in sunlight	Requires Active Tag
RGB	Bad	Yes	No
Depth	Good	No	No
Laser	Good	Yes	No
Thermal	Bad	Yes	No
Sonar	Good	Yes	Yes

Table 1: Summary of common sensor functionality

1.2 Robotic Operating System

Most mobile robotic applications work with the same underlying software framework, the Robotic Operating System (ROS). This framework provides operating system like features, and can be used in a distributed heterogeneous group of nodes, enabling reliable communication between different sensors, actors and processing units.
The ROS Project is open source and provides a package system that makes sharing of highly specialized solutions between different working groups easy. Integration of these solutions is easy as only a communication between the new node and the existing nodes has to be established.
Figure 1 below shows the node based approach ROS takes. By providing communication between the different nodes it is even possible to stream data to an extern PC for further analysis and processing. All Nodes are registered at a master node and can then communicate directly with each other as needed [5]

2 Methodology

Data for this paper was obtained from the respective cited paper.
The cited papers themselves use a formal approach but omit predicted outcomes as testing in experiments is inevitable due to extern conditions that can't be modeled.
A typical experiment for tracking / detection will focus on following key points, as these indicate the quality of the solution very well: valid assignments, ID switches (ID), misses ($Miss$) and false positives (FP).

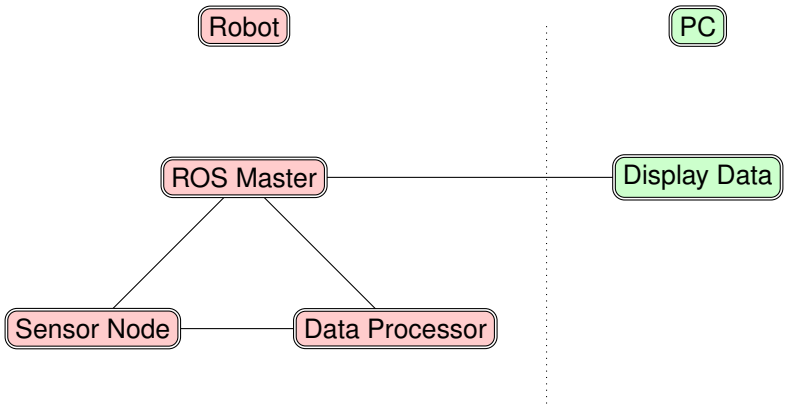

Figure 1: Typical distribution of task on nodes in ROS

These values can be aggregated together to evaluate the *multi-object accuracy tracking (MOTA)* score, as described in [3] by the formula:

$$MOTA = 1 - \frac{\sum_k (ID_k + Miss_k + FP_k)}{\sum_k g_k}$$

g_k represents the ground truth annotations and k is the time. This weighs all errors with the same weight, which may shed a wrong light on the application and its intended use case scenario.
To quantify how precise a target is tracked if it is properly tracked over a given time frame the *muli-object tracking precision (MOTP)* is introduced [3]. c_k is the number of matchings between estimated people positions and ground truth values. It is put into proportion with d_k^i, which describes the distance between the *i*th match as follows:

$$MOTP = \frac{\sum_{i,k} d_k^i}{\sum_k c_k}$$

3 Major Findings of the Papers

3.1 Tracking with RGB-D Data

The paper [2] discusses tracking of people with the help of RGB-D Data in a mobile platform.
Figure 2 shows the steps that are needed to be performed for yielding a close to real time performance (approx. 26 Images per Second).

3.1.1 Voxel Grid Filter and Ground Plane Removal

To reduce the amount of data that needs to be processed a filter is applied on the point cloud that the sensor delivers. This is done by dividing the frame into voxels (volumetric pixel) of 0.06m. All points in a voxel are then approximated with coordinates of their centroid.
Since it can be assumed that people stand on the ground the ground plane can also be removed from the already shrinked point cloud.

3.1.2 3D Clustering

After removing the ground plane, distinct people should no longer be connected to each other and the 3D points could be clustered on that basis. To overcome problems like splitting one person into multiple

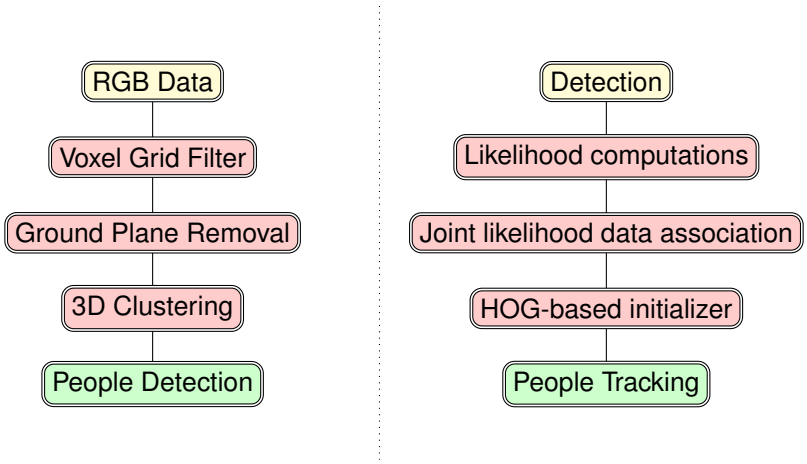

Figure 2: Process of detecting and tracking a person (from [2])

clusters if it is partially occluded or merging multiple persons into the same cluster due to being to close to each other another feature of the human body is considered: The head. As the head should be the highest part of any human and also is the one to be least likely occluded it can be used as a metric for splitting clusters correctly. Local maxima in the clusters are determined and, assuming that 0.3m is the intimate distance between two individual, all points in that distance are associated with the cluster of the local maxima. If clusters have too few points or don't match a certain height they are discarded.
A HOG (Histogram of Oriented Gradients, Matlab Toolbox) Detector is then applied on the parts of the RGB Image where Subclusters were detected to further reduce the chance for mismatches.

3.1.3 Tracking

Tracking is composed of three components: Motion of the tracked object, color appearance and people detection confidence.
A maximization of a joint likelihood is performed between these three components in order to solve the data association problem.

Color Appearance
The challenge in using the color appearance of an object is to select relevant colors of this object that make it distinct to other tracked objects in the same environment. The solution reflects this and uses an online classifier based on Adaboost for weighing the color appearance. A color histogram is computed from the RGB Image of the current detection (that already has an associated track).
Weak classifiers in the histogram are pooled by randomized parallelepides, the sum of histogram elements within this parallelepide is the feature value.
To improve the training process not only random negative examples are selected but also histograms of detections that are not associated with the current track. This improves the distinction between different tracks further by increasing the difference in confidence between the currently observed and other tracks.

Detection and Motion
The velocity of the tracked object is determined by an Unscented Kalman filter to predict a future track. A constant velocity model was chosen, as it yields good results for full occlusion scenarios.

The Mahalanobis distance (The distance between two points in a multidimensional vector space) between the current detected position and the and the predicted state of the track is calculated, based on the velocity of the object.

3.2 Experiments

The paper chose three different test scenarios to validate its setup and evaluated them with the *MOTA* and *MOTP* metrics explained in the methodology section. The videos that were tested were manually annotated with ground truth values.
The results are good, with *MOTA* values of over 80% and *MOTP* values over 90% for every scenario. The three test stages were designed as follows:

- No obstacles, linear trajectories of human objects
- No obstacles, more difficult trajectories and interaction between human objects
- Obstacles, more difficult trajectories and interaction between human objects

A second test was performed on a publicly available RGB-D Dataset that was recorded with 3 Kinect Sensors. This dataset is fully annotated and allows for performance comparison between different approaches that use RGB-D sensor data as input. The performance is again good, but *MOTA* values were down to roughly 70%, compared to 78% from the algorithm the paper chose to benchmark against [6]. The number of ID Switches and False Positives were lower than the ones in the other approach.

4 Critical Reflection

The paper shows an robust detection and tracking algorithm that performs well in the concluded experiments.
Nothing is said on the outdoor performance though, which could suffer due to the use of RGB-D sensors as main source for depth perception. Stereo Cameras replace the RGB-D sensors in such a scenario but may result in excessive use of computing power. The approach is also designed for working well in even terrain by removing the ground plane from the voxel grid. This will be another point of failure for a good outdoor performance.
This is also indicated in the second experiment where the *MOTA* value mainly decreases due to the presence of many people on stairs in the dataset.
More experiments in this direction would be desirable.

5 Conclusion

The paper shows a very fast and resource efficient approach that is suitable for a wide variety of scenarios (static and mobile, different hardware). Due to the assumpation that people move on an even pane and the subclustering approach enables the algorithm to work well with people close to each other or people that are close to the background.
Due to the use of ROS multiple sensors sources can be utilized and processed in an multi-threaded way. New test scenarios emerge for the future, e.g. using a network of robots for imaging and computing the enviroment.

References

[1] M. Munaro, A. Horn, R. Illum, J. Burke, and R. B. Rusu, "Openptrack: People tracking for heterogeneous networks of color-depth cameras," in *IAS-13 Workshop Proceedings: 1st Intl. Workshop on 3D Robot Perception with Point Cloud Library*. Citeseer, 2014, pp. 235–247.

[2] M. Munaro, F. Basso, and E. Menegatti, "Tracking people within groups with rgb-d data," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 2101–2107.

[3] A. Leigh, J. Pineau, N. Olmedo, and H. Zhang, "Person tracking and following with 2d laser scanners," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 726–733.

[4] I. Ćirić, Ž. Ćojbašić, V. Nikolić, and D. Antic, "Computationally intelligent system for thermal vision people detection and tracking in robotic applications," in *Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), 2013 11th International Conference on*, vol. 2. IEEE, 2013, pp. 587–590.

[5] "Ros Documentation official wiki," http://wiki.ros.org/, accessed: 2016-12-15.

[6] M. Luber, L. Spinello, and K. O. Arras, "People tracking in rgb-d data with on-line boosted target models," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 3844–3849.