Low level detection and tracking for robust following of a single person in cluttered environment

Olivier Aycard

Low level detection and tracking for robust following of a single person in cluttered environment

Olivier Aycard
National Polytechnic Institute of Grenoble (Grenoble INP - UGA) - GIPSA Lab, France
olivier.aycard@grenoble-inp.fr

***Abstract*—To deploy mobile robots in spaces that they will share with humans, mobile robots should have social navigation methods. One important feature to design such methods is the ability to follow a moving person. In this paper, we present how we detect and track a moving person that our mobile robot is following. A low level detection of the moving person followed, to detect the person independently of their position and orientation with respect to the mobile robot, is combined with a sliding window approach to track the moving person. Some experimental results show the robustness of the method on real scenarios.**

***Index Terms*—Human Detection and Tracking; Sensor Fusion; Robot Companions**

I. Introduction

Many companies (such as Amazon, Bossanova, Hease) and researchers in the robotics and IoT fields are increasingly interested in deploying mobile robots in spaces that they will share with humans. Among the few current examples of such robots, there are accounts of people interpreting robot behavior as a cue of interaction and communication processes.

Observing human behavior when interacting with robots already suggests that the dimensions linked to navigation (such as distance, orientation or kinematics) cannot be modeled with pre-determined and fixed constants, even when their values originate from ecological experiments. In previous work [1], we designed social navigation methods in an iterative manner, with an emphasis on adaptivity of the method, evolvability of the system, ecological experimentation, and thorough analysis of the experimental data and observed HRI both from the human and robot's perspectives.

One important point in this previous work is that the robot should be able to detect and track the followed person. In this paper, we describe our approach to detect and track the followed person.

To perceive its environment, a mobile robot is equipped with sensors (2D laser sensors most of the time). Using the information provided by its sensors, the capability to autonomously detect, track and follow a moving person has been identified as an important functionality for many robot systems.

*This work is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02).

Leg detection has been a popular method since, for most robots, the laser scanners are placed at ankle level for navigation. One of the main drawbacks of this approach is that legs are not very discriminant when perceived with a laser scanner [2]. To overcome this limitation, [3] employ a learning method to determine which properties and in what amounts to consider to improve detection. However, detection of multiple people in cluttered environments is still difficult especially considering occlusions when people walk side by side. Additional features, provided at a different height where occlusion stops, can overcome this problem. For instance, [4] defined a setup with two laser scanners one at the legs level and the other one at the chest level to improve the perception of a person. The main drawback of this approach is that the way they fused the information provided by the two lasers doesn't consider occlusions of legs, and the chest could only be recognized when the person is facing the robot. But one limitation still present in those systems [3], [5] is occlusion of the body feature for an extended time, for example if the person stopped behind the dust bin or is seated behind a desk. An other important problem in cluttered environment where many similar features appear like legs or chests is that it is very difficult to identify which detected features belong to which person.

To improve detections, tracking techniques are used to identify which detected features belong to which person essentially when we have occlusions and where are in cluttered environment with many similar features, like legs of people in a public event. This problem is known as the data association problem. The most powerful and robust technique to perform data association and multi objects tracking is known as Multi Hypothesis Tracking [6] (MHT), which can deal with temporal occlusions and manage several hypothesis of data association at each time. But one limitation still present in those systems [3], [5] is occlusion of the tracked body feature for an extended time, for example if the person stopped behind the dust bin or is seated behind a desk. MHT based systems will delete of the occluded track if it is missing for more than some maximum time. Moreover most of the work on the following of a single person, detect and track all the persons present in the envi-

ronment which is useless and imposes to implement complex tracking techniques especially in cluttered environment [7]. On the contrary, we decide to detect and track only the followed person as [2]. The main advantage of our method regarding [2] is that we manage several hypothesis at each time and work with a long time horizon to avoid confusions.

In this paper, we present a robust method to follow a single person, detecting and tracking this person. The interests of our approach are:

- To avoid occlusions, we use two lasers to perceive the two legs and the chest of the person. In this way, we extend the work of [4] to take into account the occlusions of one or several parts (legs or chests) of the followed person. To achieve this, we design a model-based tracking that takes advantage of the fusion of the two lasers data in a less restrictive way than [4];
- To avoid to implement a complex MHT, we design a low level detection which enables us to track only the moving person followed and the objects in its close surrounding;
- an adapted tracking method based on our prevous work in tracking [8] that (i) is real time and robust, (ii) takes into account several hypothesis for the moving person followed, (iii) estimates the whole trajectory of the moving person followed than only 2 consecutive detections in most of the classical tracker.

This paper is organized as follows. In next section, we describe our experimental platform. Section III details the architecture we designed and implemented to follow a moving person. In section IV, we present some experimental results. Finally, we give some conclusions and perspectives in section V.

II. Experimental platform

Fig. 1. RobAIR. The white circle at the bottom shows the position of the bottom laser scanner and the blue circle the position of the top one.

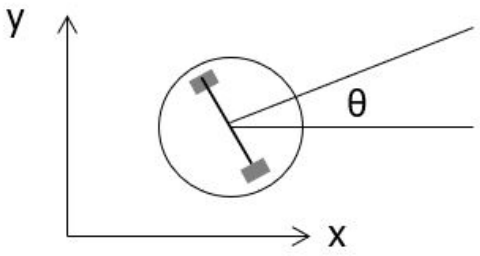

Fig. 2. Circular base of RobAIR showing the 2 driven wheels and its position (x, y, θ) in its environment.

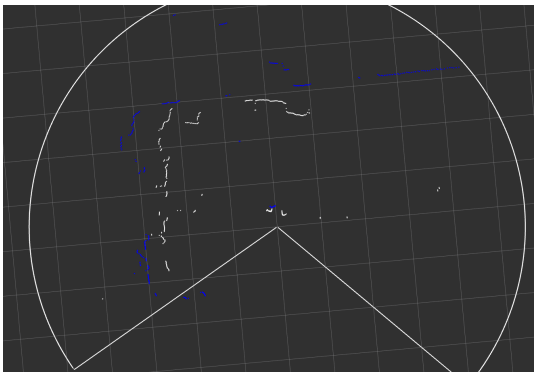

Fig. 3. Range, field of view of the bottom laser in white color and of the top laser in blue color. The background is composed of a grid where each square has a size of 1m x 1m.

Our mobile robot named RobAIR is a differential drive mobile robot (see figure 1). This is a home made robot designed and built in our fablab. It is made of flower pots: a large pot for the bottom and a smaller one for the top. Its head is equipped with LED strips. Its base is circular with 2 drive wheels and 2 idler wheels for stability (see figure 2). To control it, we can send translation and rotation speeds to its base. Each wheel is equipped with encoders and we have designed an odometer to estimate its motion. It is equipped with two Hokuyo laser scanners to perceive its environment at a frequency of 10 Hz: one located at the leg level and the other one is located at the chest level. The range of the laser is of 5.5 meters and its field of view is of 240 degrees with an angular resolution of 1/3 degree (see figure 3). So a scan of the laser is composed of 724 values corresponding to the 724 hits of the laser.

Its position is defined by its pose (x, y) in a plane and its orientation θ (see figure 2).

III. Architecture of moving person followed

In this section, we detail our architecture to follow a moving person (see figure 4). This architecture has two modes: (i) first of all, we detect a moving person that we will follow and (ii) we track this moving person.

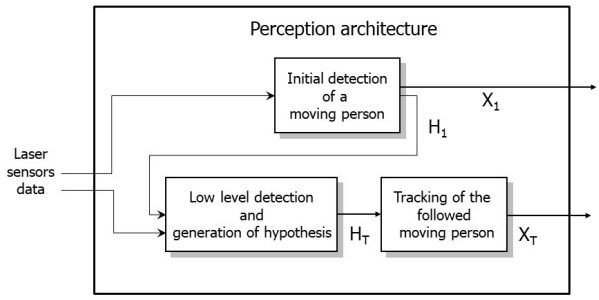

Fig. 4. Our architecture to follow a moving person

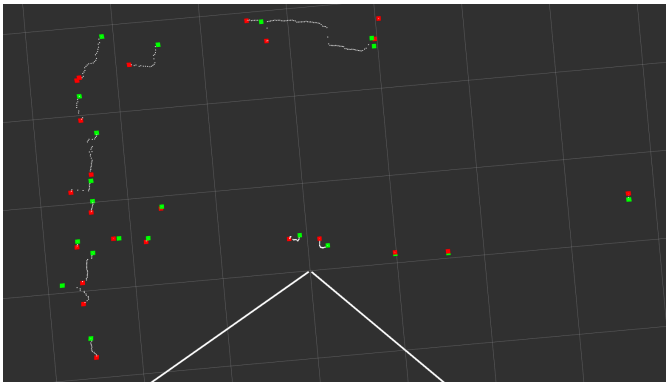

Fig. 5. Illustration of the clustering of the bottom laser. A green hit illustrates a start of a cluster and a red one the end.

A. Initial detection of a moving person

In this section, we detail how we initialize the following of a moving person. The main idea is to follow the first moving person in the environment of the robot. First of all, we detect the persons present in the environment. Secondly, we focus our attention on one moving person and finally we label the hits of the two lasers for future detections. These 3 steps are detailed in the next subsections.

1) Detection of a person with data fusion of the two lasers

First of all, for each laser, we cluster the data to form objects. The clustering is performed by comparing two consecutive hits of the laser and if they are closer than a given threshold, they belong to the same cluster, otherwise we create a new cluster (figure 5). In a second step, we search for persons by performing a fusion of the clusters of each laser. When a person is facing the laser scanners, the person is perceived by the two lasers as illustrated in figure 6. We start by searching for a chest. On the contrary to [4], a chest is not only an ellipse because most of the time the two arms are perceived as well and belong to the same cluster than the chest. So we define a chest as a box model with a length of maximum 90 centimeters and a width of maximum 20 centimeters. When we have found a chest, we search for the two corresponding legs. A leg is modeled by a circle of a diameter smaller than 25 centimeters. So in the data of the bottom laser, we search for two legs and we associate them with each chest. To check that the chest and the two legs belong to the same person, we check that the center of the chest is close to the center of the two legs and that the two legs are separated by less than 70 centimeters (ie, maximum possible distance between two legs of a walking person) to ensure that the two legs belong to the same person.

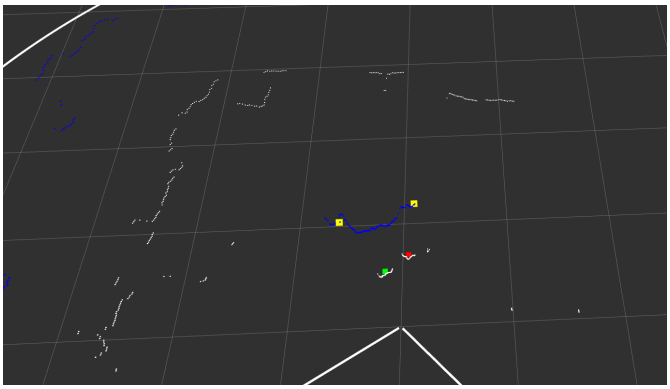

Fig. 6. Illustration of a person perceived by the 2 laser scanners. The red dot shows the right leg, the green dot shows the left leg and the two yellow dots show the beginning and the end of the chest.

2) Detection of a moving person

To focus the attention of the robot on one person and to avoid all false positives, we decided to follow the first moving person present in the environment of the robot. To take into account only moving persons, we use a background subtraction technique to detect clusters that are composed of hits of the laser that are considered as moving.

These two first phases eliminate all the false positives but sometimes due to the fact that we consider a perfect model of a person, we miss a moving person that is not facing the robot and is not perfectly perceived by the 2 lasers.

3) Classification of laser hits for low level detection

We consider that the chest or the legs of the moving person will be perceived most of the time, although the patterns could change while the person and robot are moving. To account for this change, we perform a classification of the hits of both lasers. The goal of this classification is to perform a scan matching between two consecutive scans of the laser to find the legs and the chest of the moving person in the current scan of the laser.

To achieve this task, we classify as a chest all the hits of the top laser corresponding to the chest of the moving person and as a leg all the hits of the bottom laser corresponding to the two legs of the moving person. Figure 7 illustrates the process of classification. The bold white and blue hits will be used to perform a low level detection in the next scan of both lasers.

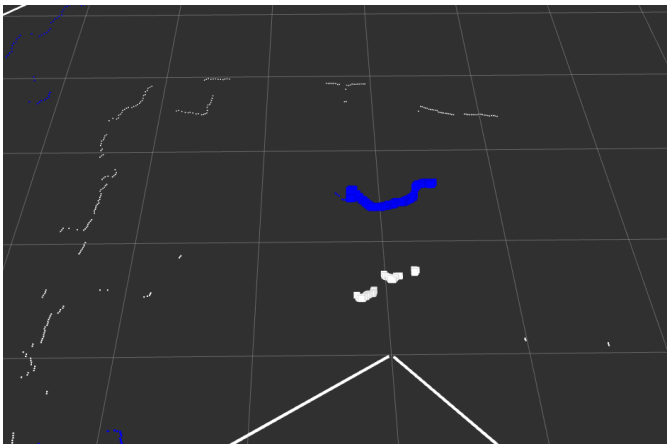

Fig. 7. Classification of hits of the laser corresponding to the moving person of figure 6. Bold white dots correspond to hits that belong to one leg and bold blue dots correspond to hits that belong to the chest.

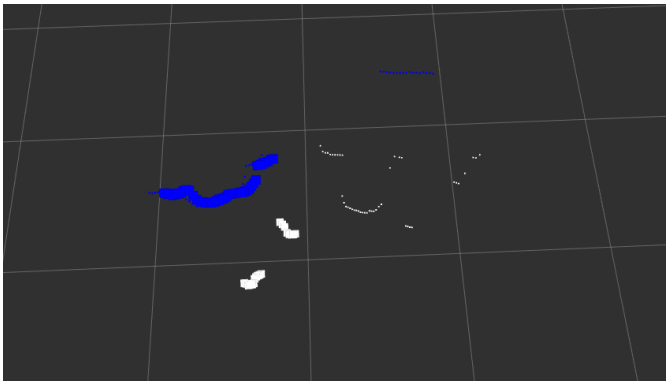

Fig. 8. Illustration of low level detection for both lasers. Bold white and blue dots correspond to hits that are detected and the other dots correspond to hits that are not detected.

B. Low level detection of the followed moving person and generation of hypothesis for tracking

In this section, we detail how we detect the moving person that the robot is following in order to track this person. First, we perform a low level detection to detect the moving person independently of its position and orientation with respect to the mobile robot. Secondly, taking into account that sometimes objects present in the environment could be close to the moving person and could be confused with some parts of the moving person (for instance a dust bin could be confused with a leg), we generate a set of hypothesis for the moving person that will be used for the tracking in the next phase.

1) Low level detection of the followed moving person

The goal of this part is to find the two legs and the chest in the current scan of the two lasers. Using this approach, we avoid detecting and tracking all the persons present in the environment and moreover we can detect legs and chest independently of their pattern, which could change with distance and orientation. This is a kind of scan matching [9] where for each hit of each laser in the current observation, we search for the closest hit in the previous observation of the same laser. Figure 8 illustrates the low level detection process for both lasers. We see that the hits corresponding to the two legs and to the chest are clearly detected and will be used, in the next step, to generate some hypothesis for tracking. All the hits of both lasers that are not detected will not be used to generate hypothesis.

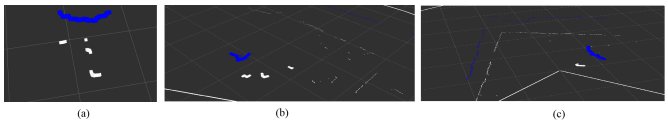

Fig. 9. Examples of generation of hypothesis for tracking

2) Generation of hypothesis for tracking

Now that we have identified the hits of each laser that could correspond to the followed moving person, we will cluster these hits to form objects and according to our model of a person, we will generate hypothesis to perform tracking. To take into account that some parts of the followed moving person could be occluded or not detected, we consider that a person could be composed by any combination of a chest and two legs. Figure 9 shows 3 examples of detection and the hypothesis generated. In figure 9(a), a chest and two legs have been detected and a dust bin as well located on the right of the two legs. It is not possible to discriminate between the legs and the dust bin so the dust bin will be used as a leg to generate hypothesis. In this case, according to our person model, we will generate two hypothesis: the first one with the chest and the two legs and the second one with the chest, the right leg and the dust bin. In figure 9(b), a chest and two legs have been detected and a dust bin as well located on the right of the two legs. It is still not possible to discriminate between the legs and the dust bin so the dust bin will be used as a leg to generate hypothesis. But, in this case, according to our person model, we will generate two hypothesis: the first one with the chest and the two legs and the second one with the dust bin alone considered as a leg. In figure 9(c), a chest and only one leg have been detected because the second leg is occluded. In this case, one hypothesis will be generated with the chest and the leg.

At each time t, the output of the detection level is a set of detections (called hypothesis in the rest of this article) H_t composed by $\{H_t^1,...,H_t^{N_t}\}$ where N_t is the number of detections at time t and d_t^i is the position of the i^{th} moving person detected at time t. This set of hypothesis H_t will be the input of the tracking level that is described in the next subsection.

C. Tracking of the followed moving person

Let H be the sequence of detections or hypothesis within the time interval $[1,T]$ where T corresponds to the current time and let X be the sequence of real position of the followed moving person within the time interval $[1,T]$.

By definitions:

$$X = X_{1:T} = \{X_1, X_2, ..., X_T\} \quad (1)$$
$$H = H_{1:T} = \{H_1, H_2, ..., H_T\} \quad (2)$$

The tracking problem can be treated as a process of taking inputs from measurements H to estimate X. In the probabilistic form, the tracking problem involves estimating the probability distribution $P(X|H)$.

To improve the robustness of the process, we consider tracking as finding the whole trajectory of the moving person given the whole sequence of hypothesis. More precisely among the set of possible whole trajectories of the moving person X and the set of sequence of hypothesis H, we would like to find the couple (X, H) that maximizes $P(X|H)$.

$$\underbrace{\max_{X,H} P(X|H)}_{\text{posterior at } T} = \max_{X_{1:T},H_{1:T}} P(X_{1:T}|H_{1:T})$$
$$\propto \max_{X_T,H_T} P(H_T|X_T)\, P(X_T|X_{T-1}) \underbrace{\max_{X_{1:T-1},H_{1:T-1}} P(X_{1:T-1}|H_{1:T-1})}_{\text{posterior at } T-1}$$
$$\underbrace{\qquad\qquad}_{\text{prediction}}$$
$$\underbrace{\qquad\qquad\qquad\qquad}_{\text{estimation}} \quad (3)$$

This equation can be interpreted as transformations over distributions of probability. Using the state transition function $P(X_T|X_{T-1})$ and the previously estimated probability $P(X_{1:T-1}|H_{1:T-1})$, we obtain a distribution $P(X_{1:T}|H_{1:T-1})$ which is commonly called the prediction step. Then introducing the new hypothesis H_T, we estimate the distribution, with the likelihood $P(H_T|X_T)$ (although called the sensor model), to obtain the desired result $P(X_{1:T}|H_{1:T})$. The Viterbi algorithm [10] is a recursive algorithm that provides a solution to this kind of discrete linear optimization problem. It is used for finding the most likely sequence of hidden states (ie, the sequence of tracks X corresponding to the sequence of positions of the followed moving persons) that results in a sequence of observed events (ie, the sequence of hypothesis or detections H).

To implement this algorithm, we need to define the state transition function $P(X_T|X_{T-1})$ and the sensor model $P(H_T|X_T)$. $P(X_T|X_{T-1})$ is a linear prediction based on previous estimation of the successive positions of the moving person. $P(H_T|X_T)$ is based on a product of three independent gaussian distributions: one gaussian for each leg and one gaussian for the chest. Each gaussian is centered on the corresponding part (ie, left leg, right leg and chest) of X_T with a small variance. The intuitive idea is that if H_T is closed to the predicted position X_T the likelihood that H_T corresponds X_T is high otherwise, it is low.

IV. Experimental results

The experiment described here is extracted from a video available at: https://youtu.be/6hhKI53g2b8[1]. We extract and interpret some interesting parts of this video. In bold blue and white color, we see the hits of the laser that are the results of the low level detection. Moreover, the chest is shown by the two bold yellow dots, the red dot represents the right leg and the left leg is represented by the green dot. The goal of the figures presented in this section is to illustrate the robustness of our approach.

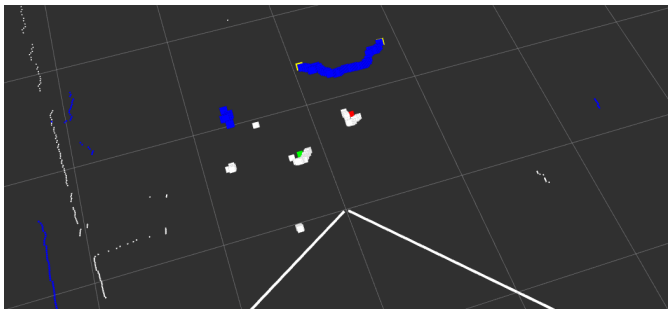

Fig. 10. Example of low level detection + tracking.

Figure 10 shows an example of the low level detection process + tracking. We have some small objects that have appeared close to the moving person, so they are labeled as possibly belonging to the moving person. But our tracker is able to estimate the position of the different features composing the followed moving person.

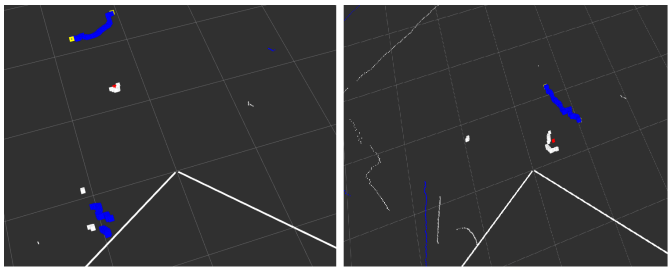

Fig. 11. Example of bad perception of legs.

Figure 11 shows two examples where the legs are difficult to perceive and detect. In the left figure, the left leg is occluded by the right leg. In the right figure, the two legs are very close each one from the other. So it is impossible to distinguish them but our low level detection still detect them and we are still able to track them.

Figure 12 illustrates what happen when the moving person is close to a pole. In the top left figure, we see that as the

[1]Some more videos of our "follow me" behavior could be found at: https://youtube.com/playlist?list=PL8ZyzBKlMS50B_jNKEYciU0tBKUHbbPXu&si=h83wfCtvDlX240qd.

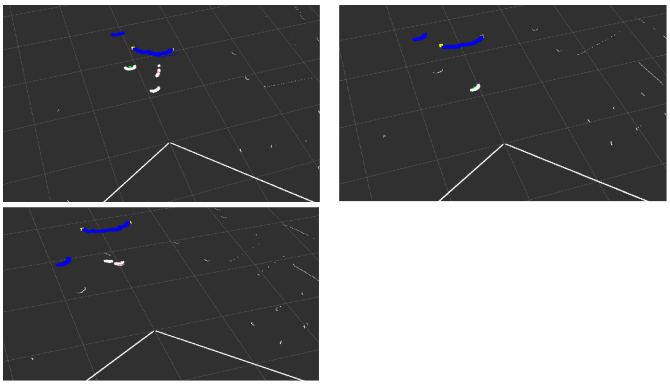

Fig. 12. Example of following when the moving person is close to a pole.

pole appears for the first time close to the moving person, its hits are labeled as possibly belonging to the moving person. Moreover, as the right leg is partially occluded by the left leg, the association of the two legs is wrong: the right leg is confused with the left one (red dot) and the left one is confused with the bottom of the pole (green dot). In the top right figure, due to the fact that we take into account the whole trajectory, the misassociation has been corrected: the left leg is now completly occluded by the right leg (green dot) and the bottom part of the pole is not associated with any legs because it is too far from the right leg and the chest. Finally, in the bottom right figure, we see that now the two legs are detected and well associated.

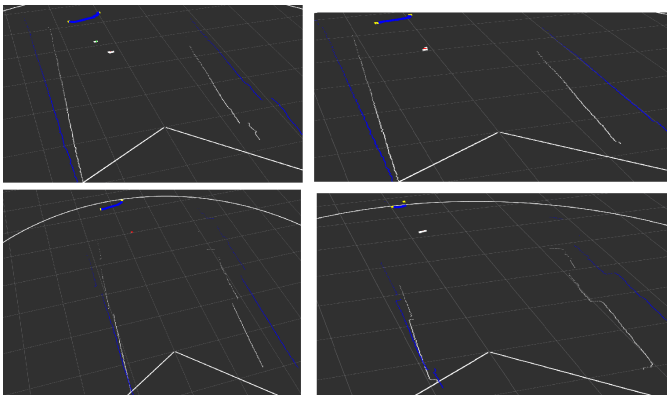

Fig. 13. Example of following when the moving person is far from the mobile robot.

An illustration of the following of a moving person that is far from the mobile robot is illustrated in figure 13. It is important to note that the moving person is located at more than 4 meters from the mobile robot and it is difficult to perceive him because we are close to the maximum range of the laser. In the top left figure, the two legs and the chest are still perceived by the two legs are very small. In the top right, we dont perceive the left leg but due to our low level detection and tracker, the moving person is still tracked. In the bottom left, the perception is worse because we dont perceive any legs. Finally, in the bottom right, we perceive the right leg again and the perception of the chest is very bad because it is very far from the mobile robot. This 4 illustrations show that our low level detection still detect the moving person even if it is closed to the maximum range of the laser and combined with our tracker, the moving person is not lost.

V. Conclusion and perspectives

In this paper, we presented a method to perform a robust detection and tracking of a moving person in cluttered environment. Our method is based on (i) a sensor setup that enables us to perceive a moving person at the legs level and at the chest level to improve the robustness of the perception, (ii) a low level detection to detect legs and chest independently on their position and orientation relative to the robot, (iii) a model based tracking that fusion features provided by both lasers to generate hypothesis for the position of the moving person and (iv) and a real time estimation of the whole trajectory of the moving person followed than only 2 consecutive detections in most of the classical tracker.

Our approach has been intensively tested in indoor environments and in some public events.

The next step is to augment the laser information by some features provided by a camera to be able to distinguish two different persons that are very close in a cluttered environment and could be confused using only information provided by two lasers.

References

[1] P. Scales, O. Aycard, and V. Auberge, "Studying navigation as a form of interaction: a design approach for social robot navigation methods," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[2] A. Cosgun, D. A. Florencio, and H. I. Christensen, "Autonomous person following for telepresence robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[3] O. Arras and W. Burgard, "Range-based people detection and tracking for socially enabled service robots," *Towards Service Robots forEveryday Environments*, pp. 235–280, 2012.

[4] S. Y. Alexander Carballo, Akihisa Ohya, "Fusion of double layered multiple laser range finders for people detection from a mobile robot," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008, pp. 677–682.

[5] M. Mucientes and W. Burgard, "Multiple hypothesis tracking of clusters of people," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 692–697.

[6] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, 1979.

[7] A. Leigh, J. Pineau, N. Olmedo, and Z. Hong, "Towards more efficient navigation for robots and humans," in *IEEE International Conference on Intelligent Robots and System (IROS)*, 2013.

[8] T.-D. Vu and O. Aycard, "Lased-based detection and tracking moving object using data-driven markov chain monte carlo," in *IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, May 2009.

[9] J. Prassler and P. Fiorini, "Navigating a robotic wheelchair ina railway station during rush hour," *Int. Journal on Robotics Research*, vol. 18, no. 7, pp. 760–772, 1999.

[10] D. Forney, "The viterbi algorithm," *Proceedings of The IEEE*, vol. 61, no. 3, 1973.