Evaluation of a multiple target tracking approach for following and passing persons

Elin Anna Topp
Centre for Autonomous Systems
Royal Institute of Technology (KTH)
SE-10044 Stockholm, Sweden
topp@nada.kth.se

March 15, 2005

Abstract

This report describes a series of tests that were performed on a multiple target tracker. The tracker is to be used in the context of following a particular person through a populated indoor environment. Aim of the tests was to determine the usefulness of the tracking approach for this purpose and to define issues for future work and improvements. The setup of the tests is described together with the results and an analysis of critical situations.

1 Introduction

This report describes a couple of tests performed with a multiple moving target tracker on a mobile robot. Since the tracker is to be used in further work it had to be determined, in how far the implemented approach is sufficient for these purposes. The general idea for the tracker is to be one component in an interactively controlled mapping and navigation system. Tracking is needed for interaction purposes in context dependent ways as will be explained later. The context for this work is the Human Augmented Mapping framework that we propose within the European project COGNIRON[1]. Human Augmented Mapping is a term subsuming a couple of issues in service robotics, as human robot interaction, mapping, navigation, localisation, and environmental knowledge representation. A scenario that explains Human Augmented Mapping is a home tour scenario, where a user can take the robot on a tour around the house to explain locations and objects. This requires a couple of functionalities one of which is the here described method for tracking in the context of following and navigating in human populated environments.

1.1 The system setup

As a test bed the ActivMedia Performance PeopleBot is used. The robot is equipped with the built in set of sensors (sonar range finders, infrared sensors for collision avoidance with tables, bumpers and wheel encoders) and has additionally a camera and a SICK-LMS-200 laser range finder. The laser range finder is mounted at a height of about 33cm and is used for the tracking system. Odometry estimations from the wheel encoders are used for initial position estimation but other sensory data are currently not processed. To connect to sensors and the platform the open source software library “Player”[2] is used. For the reported tests the data transmission rate from the laser was set to its common maximum rate of 38400 baud. With a USB/Serial adaptor it is technically possible to reach transmission rates up to 500000 baud, but this is known to be an unstable setup[3]. In preliminary tests this communication instability could be experienced as well, so that for these tests the stable communication configuration was preferred. The data transmission rate results in a sampling rate of a about 4.7 Hz, which would mean that a person moving at a speed of $1\frac{m}{s}$ would translate about 21,3cm between two data samples. This is still not much compared to the size of a person’s projection to the ground plane. With the tests it should also be determined if this sampling rate is in fact sufficient in practical use.

[1] EU Integrated Project COGNIRON (”The Cognitive Robot Companion”), funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.
[2] http://playerstage.sourceforge.net
[3] http://www-robotics.cs.umass.edu/segway/sick.html

1.2 The tracking approach

The implemented tracking approach is based on the idea of probabilistic data association filters as proposed by Schulz et al. [3]. This method allows to track a number of targets and differentiate them. If not too many targets appear or disappear at the same time, it can also handle these situations according to the authors. With the current implementation the aim is to use the tracker for distinguishing a user from bystanders and detect and track a number of people for person passing, or "socially acceptable navigation". The detection of persons is based on two patterns that can be recognised: leg sized and body sized convex shapes in the laser data. Single legs are only accepted in a reasonable distance to an hypothesis already detected. The body size shapes are needed to represent skirts or other long pieces of clothing covering the legs of a person in the height of the laser range finder.

1.3 Outline

The report is divided into six sections, one of them being this introduction. Section 2 explains the critical issues for the tracking approach and the test setup, in section 3 the results from the different tests are presented. Section 4 explains the problematic situations that could occur during tests and section 5 refers to the effects of those situations on the functionality purposes of the tracking system. Section 6 summarises the report and gives some ideas on improvements and recommendations of use.

2 Test setup and background

The multiple target tracker will have to serve in at least two different contexts with slightly different foci. The first one is a "follow the user" situation, in which the tracker's results are needed to distinguish one person (the user) from other persons being around that might cross the user's way. The other situation is "human obstacle avoidance" in which the tracker is needed to detect and track all persons being around that could come into the robot's way to some goal. Both situations could of course as well occur at the same time in a following context with other persons being interpreted as human obstacles. Especially for the obstacle avoidance the tracker has to be stable in the sense that it must not lose any of the targets.

2.1 Situation and Scenario

As a real world scenario contains a lot of objects that appear from one or the other point of view as person-like in a laser range data set, the environment was simplified for the tests. With the help of large wood planks an area of about 4,5m x 4,0m was made into an "empty room" where the occurrence of persons could be controlled at any time.

2.2 Tests

The tests were run with three different base assumptions: "robot still", "robot moving independently" and "robot following". In the first two situation types a controlled number of persons were moving across the field of view, for the following test only one person (the one to be followed) was around.

3 Results

The results of the different tests are presented and explained in this section. Only the most relevant four types of tests are picked to illustrate the tracker behaviour. These test types represent the very basic case of a still robot with one person moving around, the case of a still robot with two moving persons in two representative constellations (occluding each other, and approaching and separating), the robot moving independently with up to three persons being around, and the robot following one user.

3.1 Robot still, one person

For this test the robot was not moving, the field of interest was set to an area of three meters of radial distance. One person walked into and through the field of interest for 9 times, following different trajectories and with different velocities. The test covered a time of about 140 seconds. Figures 1 and 2 show the x and y coordinates respective to the time steps and in three different pictures the trajectories of the person in the x/y-plane. The target was tracked correctly in all nine test cases, no loss occurred. The fact that

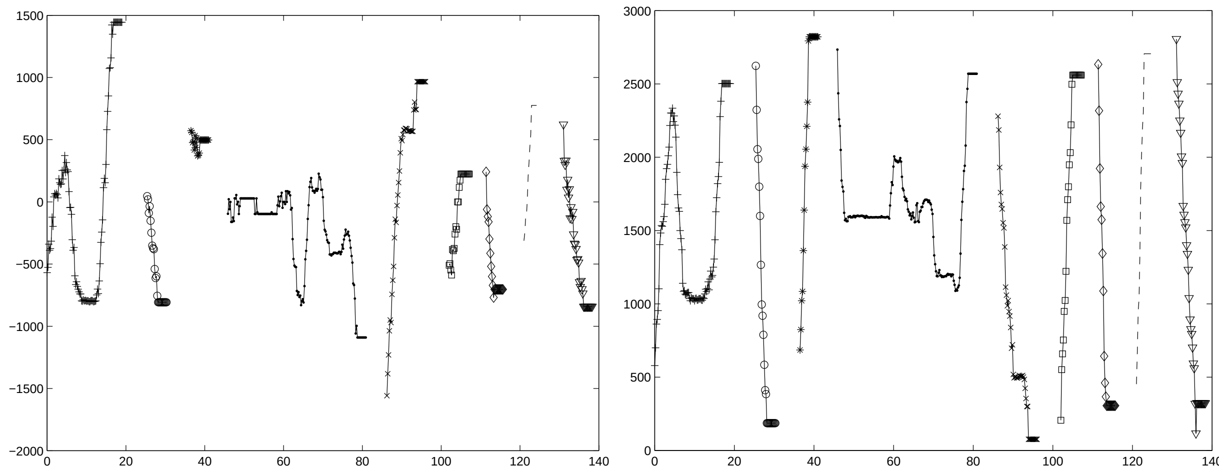

Figure 1: x (left) and y (right) coordinates of the person's trajectories over time

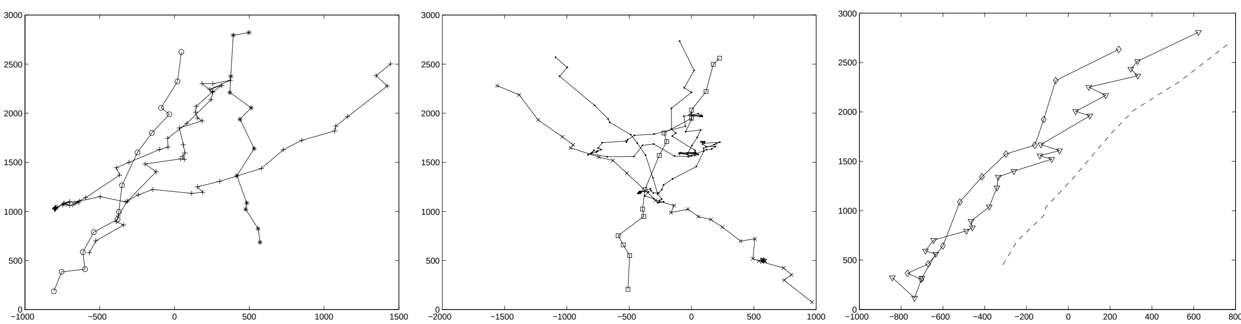

Figure 2: Person trajectories in the x/y-plane. The robot's position is fixed at (0,0). The trajectories are shown in three pictures for visibility

the trajectories in most of the cases are constant for the last few steps can be explained as follows: The tracking system considers only data and therefore features in a certain region. In this case the region is defined by a radial distance of three meters from the robot and the base line (y=0). Whenever a target is "lost" and was close (within 50cm) to the region's border in the last time step, a timer is started. This timer is used to decide if the target has left the area of interest, which happens after about three seconds of miss. "Long jumps" in the trajectory of the target can be explained as follows. Since the position of the target is represented as a single point, this point jumps from one leg to another (if there are two legs to be found, the closest one to the robot is used for position representation). If now the leg distance is rather long and the representation point switches, these jumps can occur. The same is the case when a trajectory point seems to "jump backward" according to the general direction of the target's movement.

Test summary With this very simple test scenario the general functions of the tracking system could be shown. The person was moving in different directions relative to the robot's position and went through the area of interest at different speeds up to 1m/s. The fact that all these motions were tracked correctly shows that the sample rate of about 5 Hz for the laser range data is still sufficient for the purpose, at least with the robot standing still.

3.2 Robot still, two persons

With this test the ability of the tracker to handle multiple targets is shown. Two persons enter the area of interest (the same as in the previously described scenario), walk past each other so that one is occluded from the robot's perspective. Figure 3 shows the x- and y-coordinates of the trajectories over time, in figure 4 the trajectories of the persons are shown. In this situation the occluded target is lost and replaced by a new one when visible again. This apparently erroneous behaviour of the tracker can be explained by looking

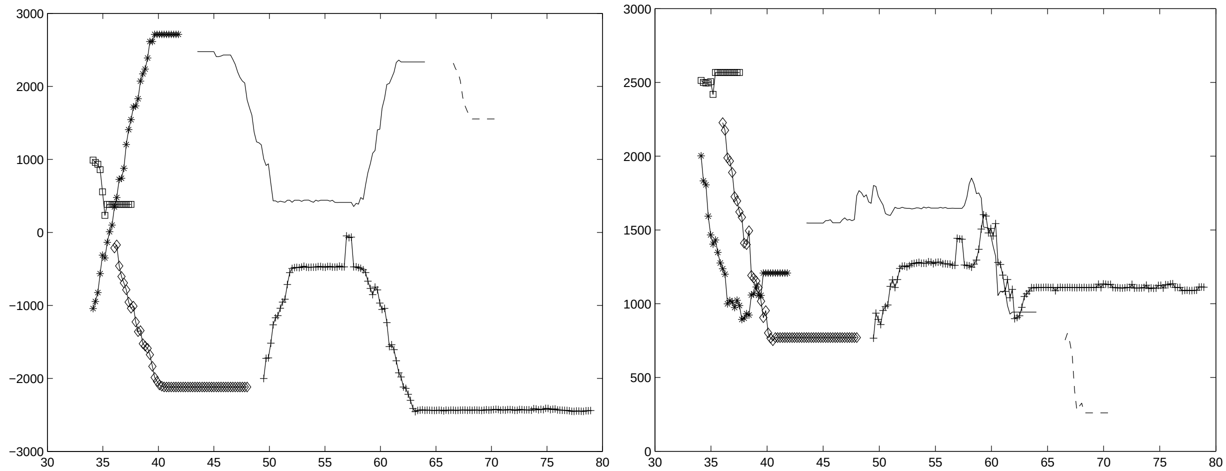

Figure 3: x (left) and y (right) coordinates of the persons' trajectories

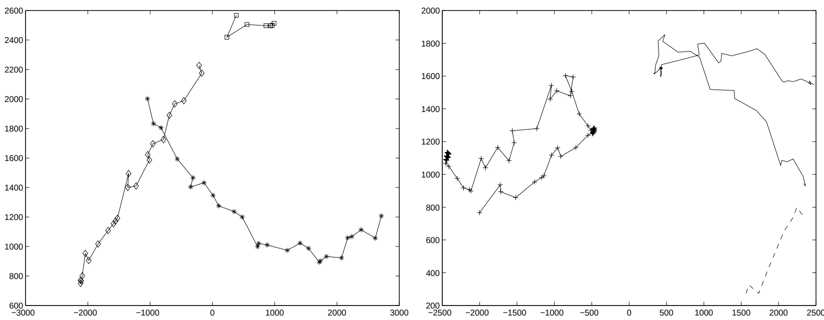

Figure 4: Person trajectories in the x/y-plane. The robot's position is fixed at (0,0). The trajectories are shown in two pictures for visibility. The two images refer to two different situations that occurred sequentially.

into the detecting and tracking process in a simulation. The figures 5 to 10 show the time steps in which the loss occurs. In the first picture both targets are fully visible, detected by one leg pattern each. The second picture shows the target number 1 occluding one of the legs of target number 0. Now target number 0 is detected by one leg pattern that refers to the other leg of the person which makes the position of the target jump in the y coordinate. This process causes an apparent change of velocity for the person. In the next step the target number 0 is fully occluded by target number one, which is not affected in any way. The particles for target number 0 spread according to the assumed velocity from the last step, which means into the wrong direction compared to the actual movement of the person. In the next frame the occluded target appears again, but the probabilistic data association ranks the probability for a loss higher than for a re-detection. A new target number 2 is created. Such a situation could be handled by an adjusted motion model, so that the particles do not take up short term changes immediately. Running tests with a changed model showed that this is possible.

In the following part of the test scenario the two persons approach each other, stop for while in a communication distance and depart in opposite directions. This situation is handled correctly by the tracking and association system. The last short trajectory reflects one of the persons walking out of the area of interest and then entering the area again.

In a test in simulation with the same data sets the speed parameter of the target velocity estimation was decreased whenever a target was not detected. This led to a less quickly drifting but still spreading particle cloud. Consequently the target was then tracked and re-detected correctly.

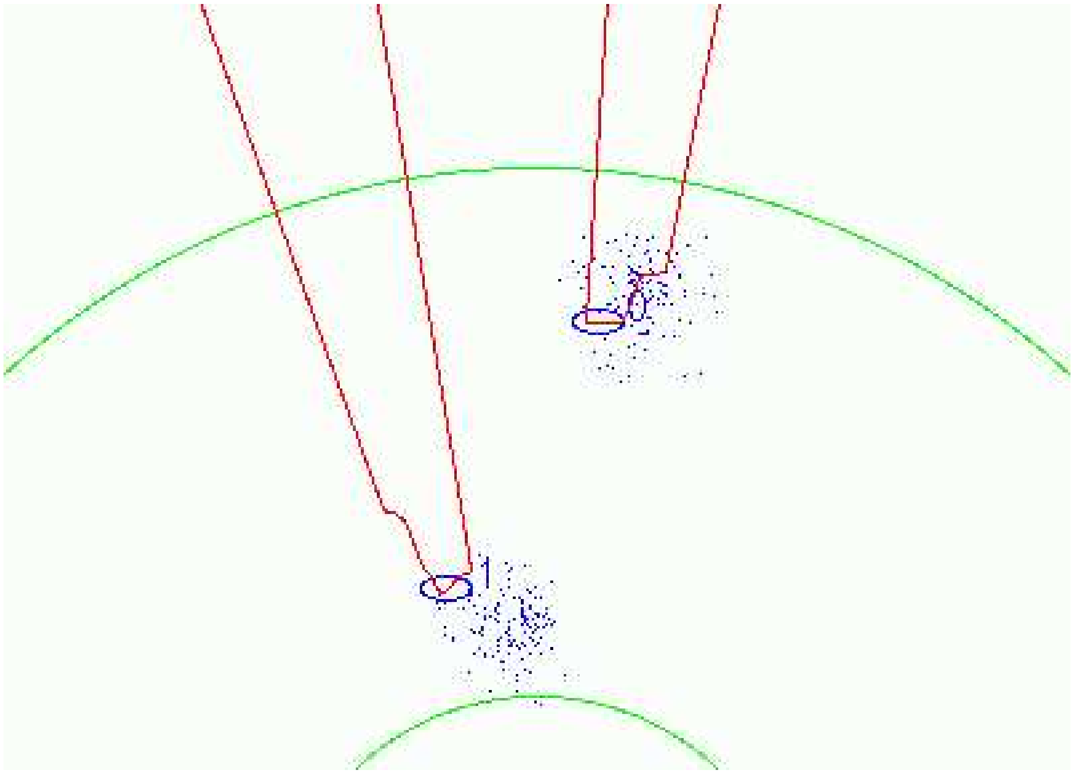

Figure 5: The two targets right before target number 1 occludes number 0

Test summary In the test with two persons passing each other one of the targets was lost and replaced by a new one. This happened due to a misinterpreted change of the target's position. Still the data association could handle the situation as specified and did not assign the continuously tracked target a wrong feature at any time. By modifying the motion model for the target with a decreasing speed parameter in case of detection misses this particular situation could be solved. This gives an idea of the dependability of the whole system on the motion model with which the particles predict the new position. Ideas for improvements are given in later parts of the report. The situation in which two persons approach each other, stop to interact and walk away, the tracking system can handle the association problem remarkably well.

3.3 Robot moving independently, three users

In this case the robot was moving straight forward until it reached some blocking obstacle (the artificial walls). There it turned randomly until it was no longer blocked (no readings in a certain distance) and continued moving straight forward. Three persons were entering the area of interest of which the distance parameter was in this case the maximum range of the laser range finder configuration (8m). The persons moved through the area on rather simple trajectories to make it possible to analyse them. Occlusions occurred in different constellations.

As can be seen in figures 11 and 12 all the occlusions were handled perfectly, none of the targets was lost. One disruption of a trajectory occurs for the sixth target (which becomes the seventh when re-detected). This severe loss is of the same type as the one in the previously described test situation: Due to an erroneously assumed direction of movement, the prediction for the positions of the target becomes incorrect. In this situation no feature is found due to misinterpretation of the data. Therefore the predicting particles spread further on their wrong path. After a while the person is detected again, but is now too far away for an assignment to the original target. This problem could be solved (as in the case described previously) by a simple modification of the motion model. More ideas on this will be given in later parts of the report.

Test summary In this situation the tracker handled all types of repeated occlusions of three persons moving around in the area of interest. One tracking loss of a target occurred, but independently from an

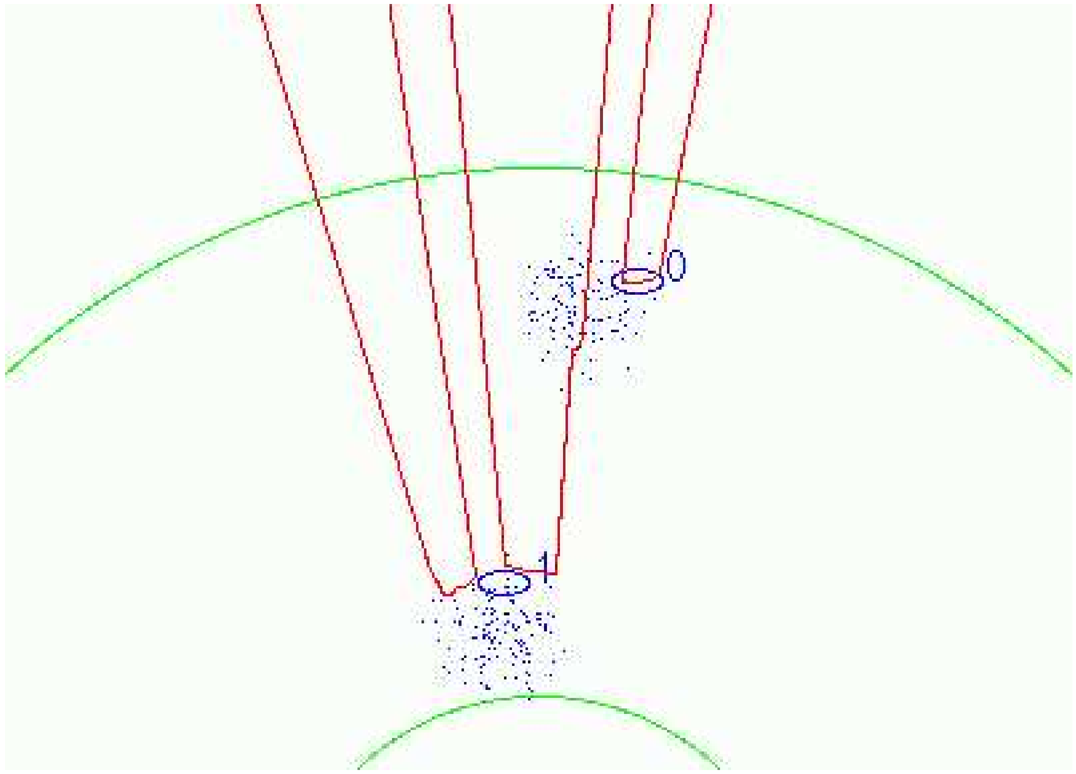

Figure 6: Only one leg of person (target) number 0 is visible, which makes the target position jump “backward”

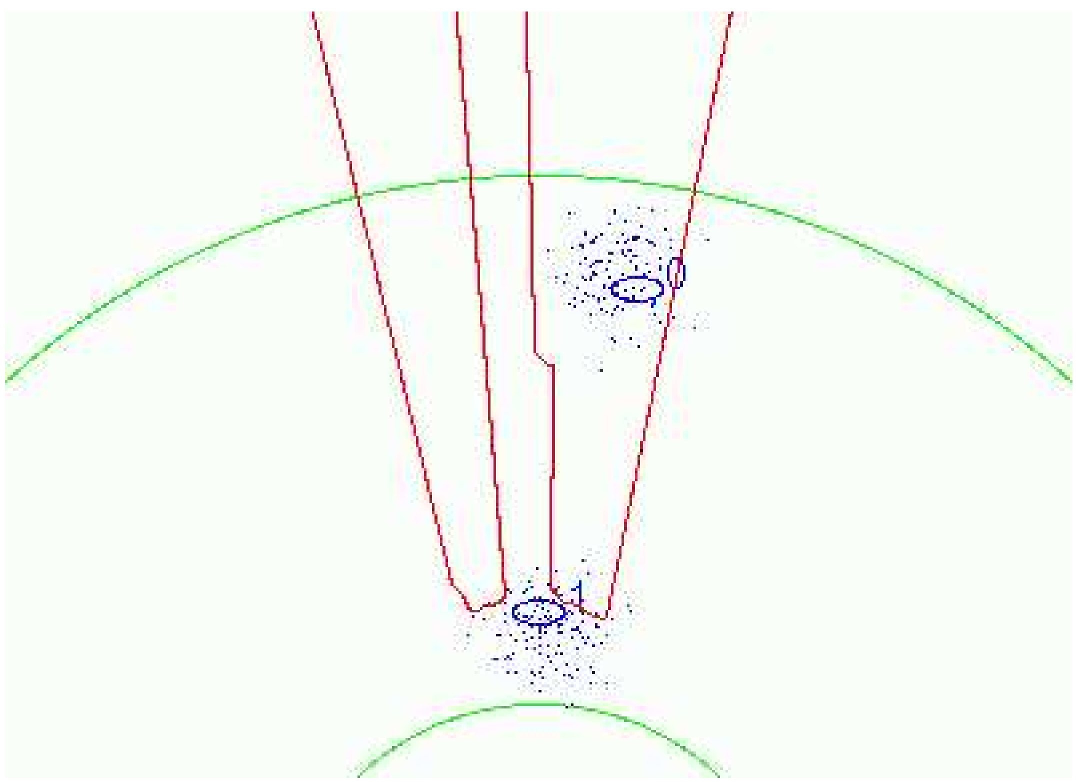

Figure 7: Target number 0 is completely occluded and its particles start spreading with the assumed velocity

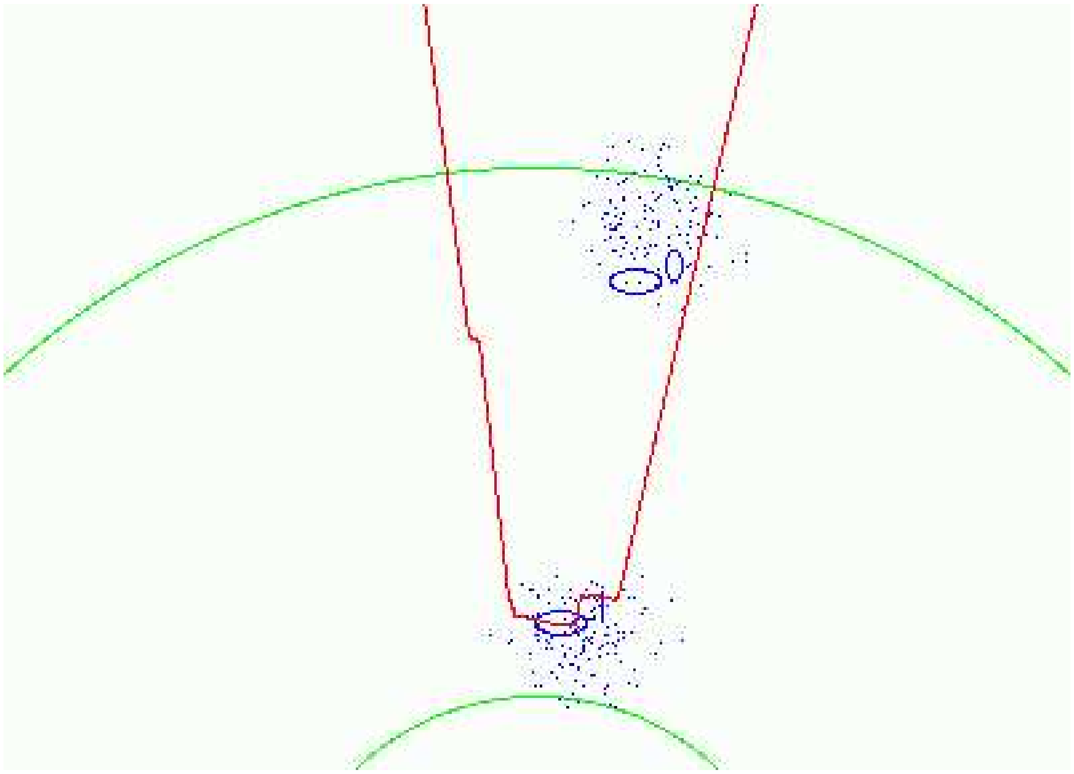

Figure 8: The target number 1 is still occluding the other target fully

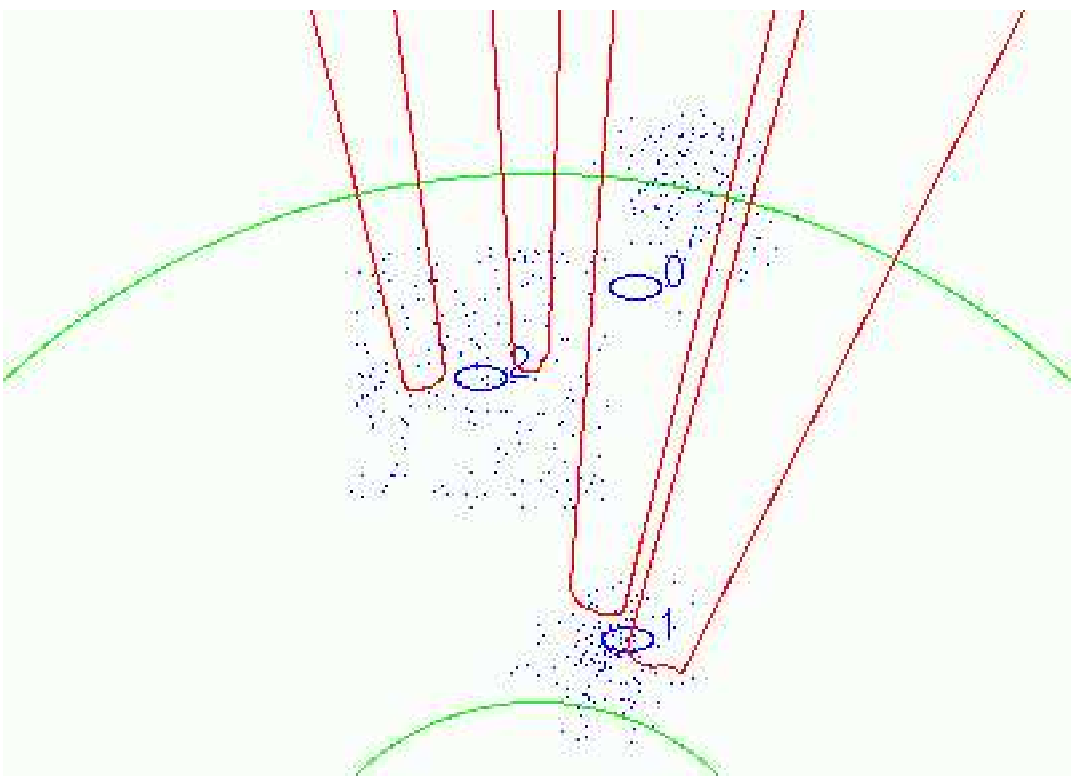

Figure 9: Due to the particles heading into the wrong direction, the feature appearing is assigned a new target number (2)

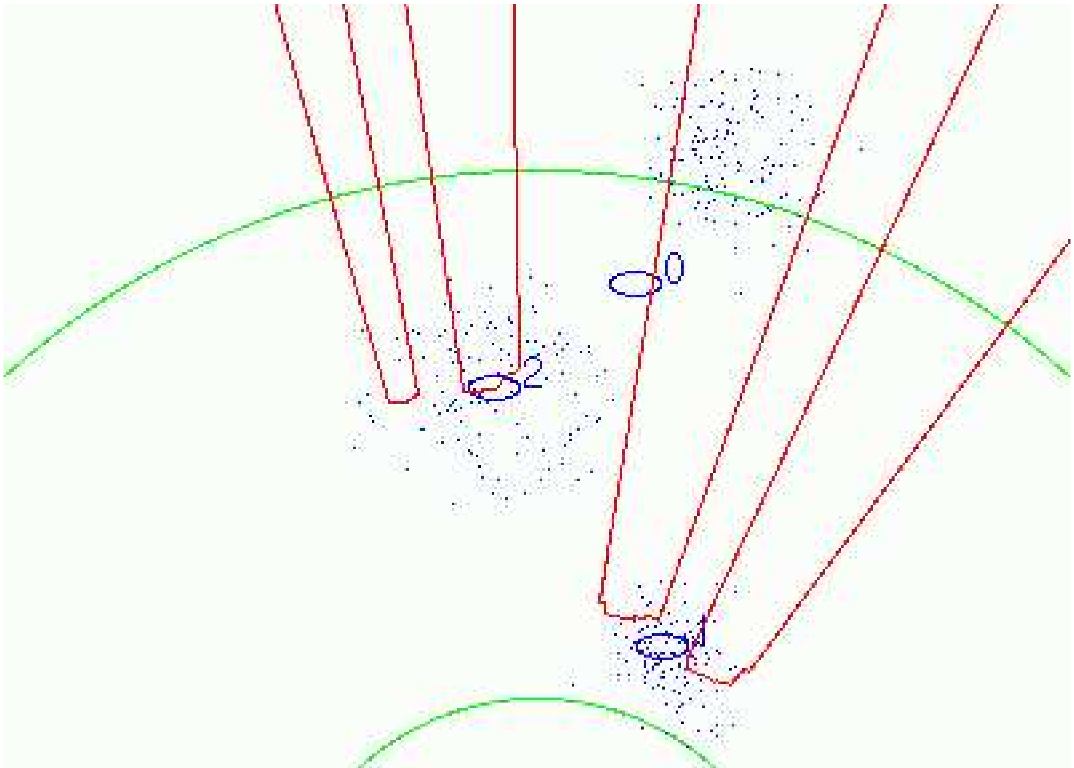

Figure 10: From now on the person who was initially assigned the target number 0 is now tracked as target number 2. Target number 0 is kept for a little while and then declared gone.

occlusion. This situation can be solved by adjusting the movement model.

3.4 Robot following, one user

In this test the robot followed one person for a time of about 3 minutes on a random trajectory. The person moved with different speeds and could be tracked all the time with one exception. Again, the situation evolved due to a combination of two problematic facts: After a quick step backward the person stopped and was not detected in the next few time steps. This was the case because the person was standing so close to the walls in a corner that the thresholds for the feature detection did not include this particular constellation. As the last detected movement was a very quick one, the tracker assumed the person continuing, which was not the case. A new target was created when the person was re-detected due to some movement. As in the other cases the problem could be solved with a differently parametrised movement model. As the trajectories got very long and confusing, it is not useful to present them in a figure.

4 Problematic situations

The situations in which the tracker can fail are described in the following.

4.1 Missing features

The detection of features that have to be matched with existing targets is up to now independent of previously collected knowledge. Neither is any kind of a priori knowledge or world knowledge applied. This results in situations, in which the constellation of person, robot and environment leads to failures of the detection. This happens rarely (once in a 3 minutes test), mostly when persons get very close to border lines of the field of view (walls, corners) or when the person's pattern in the data does not match any assumption for a feature (due to size or shape). Still it is hardly possible to measure such incidents in number per minute, as it cannot be predicted how often persons get into respective situations. A more sophisticated detection might help. A miss itself (if not occurring over a long time period) does not cause a target loss immediately.

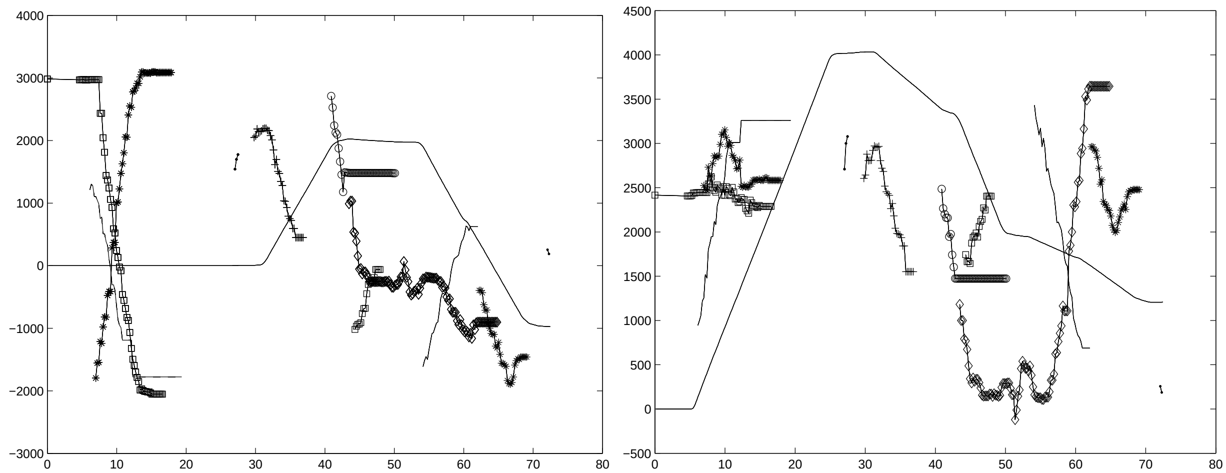

Figure 11: x (left) and y (right) coordinates of the persons' and the robot's trajectories over time

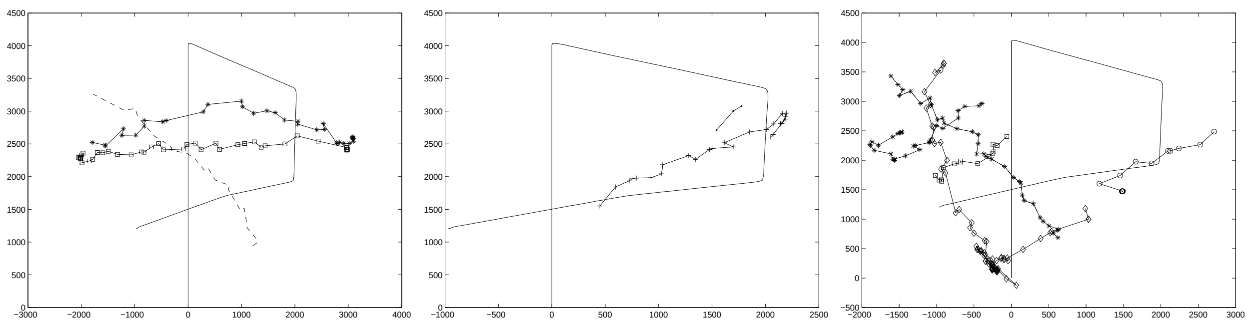

Figure 12: Person and robot trajectories in the x/y-plane. The robot followed a rather straight and simple path. The trajectories are shown in three pictures for visibility

This is the case only when additionally the prediction of the target's position is misleading. A long time period miss can occur if a person is completely static in a respective constellation, which is rather unlikely. Another possibility is of course, that the person went out of the field of view of the scanner.

4.2 Severely wrong motion assumption

Such wrong assumptions evolve when the position estimation for the person jumps within different parts of the feature, e.g. from the centre between both legs to one of the legs, that has a relative position "behind" the other one. In this case the target covers a rather large distance in short time, which leads to an erroneously high assumption for the speed. Additionally the person continues moving as before, but the predicted direction is wrong. If now the detection fails as well, the prediction diverges more and more from the actual position. Again, it is not possible to measure this in time, as it is dependent on situation and constellation.

4.3 Incorrect association of targets and features

Such a situation did not occur in any of the reported tests. In one preliminary test, when the area was not cleared of environmental objects (chairs, tables, bags, sitting persons and the like) a short term confusion (jumping associations in about two frames) happened, when a person passed a chair with a coat hanging on it very closely. The association recovered when the targets could be separated again. Nevertheless, this situation showed that there is a small likelihood of associating targets to features incorrectly. It is difficult to give numbers for this phenomenon, but it is definitely very unlikely, considering the normal distance people would have when passing each other. During the tests persons passed each other or came close together in at least 8 different constellations without any confusion.

5 Effects of tracker failures

The tracker has to serve different purposes, so the effects of situations that make the tracker produce erroneous results have to be analysed.

5.1 Following a person

In the scenario of person following the output of the tracker needs to be analysed to find the one and only person to follow. The purpose of the multiple target tracking is in this case to distinguish this user from other persons. As it is in the nature of the following process, the robot is moving such that the person is kept in focus at any time.

5.1.1 Losing track and replacing with new target

In such a situation two ways of recovering are possible. If it takes too long until the user can be detected again, the target is removed from the scene. This would result in a clear loss, which could only be solved by interactive means or higher level assumptions. If the tracker predicts the target in a different direction and therefore creates a new target when detecting the user again, the system would still keep the first user target as the one to follow. This target is in such a situation bound to disappear after a short time, as now with the new target already created, no feature is associated with the old one. Additionally its position is not updated while it cannot be detected, which would make the robot "follow" a static target. As soon as it has disappeared, the search for a new user target can be started automatically. The user would notice a hesitating behaviour for a moment and then the robot would follow again, when the system has come to the conclusion, that there is a new moving target in the right area. This type of situation is not critical for the following scenario. In the worst case the user manages to get out of the field of view of the robot, while the system tries to recover. This can be solved by interactive means with initiative from the robot. This worst case did not occur in the 3 minute test.

5.1.2 Confusions

A confusion of the user with any other target being around is critical for the following scenario. If the user gets confused with a static target the robot would get stuck, facing this static target, assuming that it still focuses the right target. As no error (loss of user) is reported, the system acts within specifications. In this case the user would have to solve the problem, by directing the focus of the system back to herself, either by means of dialogue or by walking between the assumed target and the robot, waiting until the erroneous "user" has been occluded long enough to be declared "disappeared". This situation did not occur, as the test area did not contain static person like objects when the following was tested. A confusion with another person would lead to the most critical behaviour: The robot would follow a person who is not expecting to be followed. In all tests and situations when the system was applied such a critical situation never occurred.

5.2 Human obstacle avoidance

When using the tracker output for human obstacle avoidance it has to be interpreted in a different way. Most importantly, all persons being around have to be detected and distinguished from person like static objects. This can be done by declaring a moving target a person. The interpretation is not done to pick one particular person, nor is it very important to differentiate between the detected persons. Therefore the effects of tracker failures are different for this type of scenario.

5.2.1 Losing one target and replacing it

In the case of a loss of any of the targets a critical situation can come up. Is the target replaced by a new one, the situation is still possible to be handled, as at least the actual person gets a target assigned and is reported to the system as a human to avoid. The old target is sitting for a short while as a virtual target in its last assigned position which might cause additional avoidance efforts. This is also the case with occluded targets. More critical is the case, that the target is lost completely due to miss of the target over long time, though it is still around. In such a case the system would still detect an obstacle but handle this in a classic way and not with the assumption of a person being the cause for the deviation. This is a clearly incorrect behaviour.

5.2.2 Confusing targets

This case is absolutely not critical for the obstacle avoidance in any case. If one person gets mixed with a static target, the former static target apparently starts moving and is therefore classified as person and can be avoided in appropriate manner. The former moving target, now assigned to the static object would still cause human avoidance behaviour for this object, but as it is static, the system would appear a little too polite, but still act within specification. In the case of two moving targets getting mixed up no problem for the human avoidance should occur.

5.3 Improvement suggestions

All critical situations depend on two basic parts of the tracking system: The detection of features and the motion model. As most of the reported target losses occurred under influence of an erroneous movement prediction this seems to be the more promising way to improve the tracker. The improvement of the feature extraction would involve a lot more knowledge handling in the basic parts of the system. This would slow down systems components that have to work in a time critical framework - if not in real time. Further “real world” tests would then have to show how likely a confusion based failure really is (with the previous tests it seems very unlikely).

6 Summary and recommendations

The multiple target tracker was tested in different test scenarios in an artificially emptied “room”. Under this circumstances it worked well and delivered good results with respect to both purposes, “following” and “human avoidance”. Nevertheless a few failures occurred, that showed that in particular situations and constellations problems can occur.

6.1 Application

In the current state the tracking system is applicable in a research context. Particular situations, that are more likely to cause erroneous behaviour than others are known and can be dealt with. In the current state it seems not appropriate to have “naive users” apply the system unsupervised-supervised.

6.2 Limits, not tested but known

One limit of the system is defined by the number of targets being around. As the data association is based on a statistical method that compares the likelihood for the different permutations of associations, calculations might take some time when the number of present targets is more than four. By now, the framing controller is dealing with situations, in which calculations overrun the time period for data collection, to avoid calculations on deprecated data. Still it is not recommended to have more than four possible targets around when working with the system or demonstrating it. This is another reason why “naive users” should not work with the system in the current state.

6.3 Improvement suggestions

Different kinds of improvement steps seem possible, but have to be thought of in a cost versus usefulness analysis.

6.3.1 Motion model

An easily to apply and probably effective improvement is to add a simple, but comparably more sophisticated movement prediction to the tracker. This does not seem a very time consuming step and would help to improve the handling of a miss. Such an improvement would still not help in situations where persons behave completely unpredictable, or in fact produce an unlikely trajectory. To deal with such situations would involve time consuming and deep studies of motion patterns, as started with in work done by Bennewitz et al. [1] or Bruce and Gordon [2]. Additionally a lot of world knowledge would have to be used, which would involve (the actually planned as “next step”) integration of map building into the system.

6.3.2 Feature extraction

For the feature extraction it seems at the current state not worth to spend a lot of time on integrating more or different mechanisms of distinguishing person like features from laser range data. As seen in a lot of test situations when taking a closer look into the data, persons tend to appear in a lot of different pattern types. The most likely ones are already covered, and the failures of the tracker seem to depend on the motion model rather than on a one frame miss of a feature. Still an improvement might be possible by using environmental knowledge that allows to adapt the criteria for accepting a feature as coming from a person (persons do rarely disappear in the middle of a room, but that involves conceptual knowledge about rooms and free space, doors as means of disappearing, and furniture as cause of clutter). Again, this knowledge would have to be derived from information drawn from maps.

6.3.3 Confusions

These occur rarely and if so, a strong motion model could even here help to recover from a confusing situation. The confusion problem itself is system immanent as the predictions and therefore the data association is based on a Monte Carlo method. This guarantees results, but those can be nonetheless (though unlikely) incorrect.

6.4 Improvement steps

To guarantee a reliable system for research purposes it seems sufficient to improve the position prediction for the targets integrating a velocity component into it to even out rough changes in the trajectory. All further improvements on the tracker would involve world and environmental knowledge which strongly suggests to deal with the representation of this knowledge first. Thus, after some efforts to improve this motion model, the representation of the environment is what will be dealt with in the near future.

References

[1] M. Bennewitz, W. Burgard, and S. Thrun, "Using EM to learn motion behaviors of persons with mobile robots", in Proc.of the Conference on Intelligent Robots and Systems (IROS), Lausanne, Switzerland, 2002.

[2] A. Bruce, and G. Gordon, "Better motion prediction for people tracking", in Proc. of the International Conference on Robotics and Automation (ICRA), New Orleans, USA, April 2004.

[3] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers, "Tracking multiple moving targets with a mobile robot using particle filters and statistical data association", in Proc. of IEEE International Conference on Robotics and Automation, Seoul, Korea, May 2001.