Pose Estimation and Adaptive Robot Behaviour for Human-Robot Interaction

Mikael Svenstrup, Søren Tranberg, Hans Jørgen Andersen and Thomas Bak

Abstract—This paper introduces a new method to determine a person’s pose based on laser range measurements. Such estimates are typically a prerequisite for any human-aware robot navigation, which is the basis for effective and time-extended interaction between a mobile robot and a human. The robot uses observed information from a laser range finder to detect persons and their position relative to the robot. This information together with the motion of the robot itself is fed through a Kalman filter, which utilizes a model of the human kinematic movement to produce an estimate of the person’s pose. The resulting pose estimates are used to identify humans who wish to be approached and interacted with. The behaviour of the robot is based on adaptive potential functions adjusted accordingly such that the persons social spaces are respected. The method is tested in experiments that demonstrate the potential of the combined pose estimation and adaptive behaviour approach.

I. INTRODUCTION

Mobile robots are moving from factory floors out into less controlled human environments such as private homes or institutions. The success of this shift relies on the robots ability to be responsive to and interact with people in a natural and intuitive manner and accordingly human-robot interaction is a novel and growing research field [5], [12].

To allow close, more effective and time-extended relationships it is first necessary to determine the persons willingness to engage in interaction, followed by a coordination in time and space that respects the persons interest and privacy. Several authors [2], [3], [6] have investigated the willingness of people to engage in interaction with robots that exhibit different expressions or follow different spatial behavior schemes. In [13] models are reviewed that describe social engagement based the spatial relationships between a robot and a person with emphasis on the movement of the actors. Human-aware detection, tracking and navigation were discussed in [15], [10].

As a step in the direction of human aware robot behaviour, we present a novel method for inferring a human’s pose from 2D laser range measurements. Here we define pose as the position, and orientation of the body. Compared to vision based pose estimation, such as [7], or 3D range scans [14], using 2D laser range scanners provide extra long range and lower computational complexity. The extra range enables the robot to detect the movement of persons moving at a higher speed. The approach takes advantage of the inherent mobility and typical sensors of a mobile robotic platform and does not require any determination of the persons facial expressions or other gestures, and hence the person does not have to be facing the robot. The method relies on an algorithm for detecting legs of persons. The algorithm has been tested in a public transit space [16].

While the focus of this paper is on the pose estimation, the use of the estimates in determining persons willingness to engage in interaction is analyzed from the recorded kinematic state of the person. By looking at knowledge from previous encounters, the robot behavior is adjusted as described in [1]. When the persons willingness has been determined it is used as a basis for human-aware navigation using adaptive potential functions centered around the person, inspired by [15]. The navigation is adapted such that it respects the persons social spaces as discussed in [6].

The pose estimation is validated through a number of experiments. In addition, experiments indicate the effectiveness of the combined algorithm for human-aware navigation.

II. MATERIALS AND METHODS

The basis for any interaction is the ability of the robot to detect if a person is present, and if this is the case to estimate the kinematic state of that person. The algorithm for detecting legs of persons and converting these to person position estimates is described in [8], and [19] been implemented and adapted to keep track of individual persons. The implementation has been proved robust for person speeds up to $2 \frac{m}{s}$ in a real world public space setting described in [16]. The position estimates from this algorithm form the basis for deriving a pose estimate of a person.

A. Person Pose Estimation

The setup, where a person is moving around, while the robot is following, can be seen in Fig. 1. The basic idea for estimating the pose of a person ($\theta$), is to take the position estimates from the laser range finder algorithm and combine them with robot odometry information to obtain a pose estimate. Because we have both position and velocity measurements to estimate a pose, it can not be calculated directly, and a Kalman filter is therefore used to filter the measurements. The filter produces a velocity estimate of the person relative to the robot. After this a post filter is added to obtain the person pose from the velocity estimates.
A standard discrete state space model formulation for the system is used:

\[
\begin{align*}
\dot{x}(k+1) &= \Phi \ddot{x}(k) + \Gamma \ddot{u}(k) \quad (1) \\
\dot{y}(k) &= H \ddot{x}(k), \quad (2)
\end{align*}
\]

where the state is comprised of the person position and velocity and the robot velocity

\[
\ddot{x} = \begin{bmatrix} p_{x,pers} \\ p_{y,pers} \\ v_{x,pers} \\ v_{y,pers} \\ v_{x,rob} \\ v_{y,rob} \end{bmatrix}. \quad (3)
\]

Here \( p \) is positions and \( v \) is velocities, all given in the robot coordinate frame. The position of the person relative to the robot depends both on the person velocity and the robot velocity. In this stage we omit the rotation of the robot:

\[
\begin{align*}
p_{x,pers}(k+1) &= p_{x,pers}(k) + T(v_{x,pers}(k) - v_{x,rob}(k)) \quad (4) \\
p_{y,pers}(k+1) &= p_{y,pers}(k) + T(v_{y,pers}(k) - v_{y,rob}(k)) \quad (5)
\end{align*}
\]

where \( T \) is the sampling time. This yields the following state transition matrix:

\[
\Phi = \begin{bmatrix} 1 & 0 & T & 0 & -T & 0 \\ 0 & 1 & 0 & T & 0 & -T \\ 0 & 0 & 1 & 0 & T & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \quad (6)
\]

The measurements used are the person position, and the odometry velocity data from the robot:

\[
H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \quad (7)
\]

To overcome the nonlinear effect the robot rotation has on the state, a measurement driven Kalman filter, which is used in [11], [17]. The idea is to use sensor readings to drive the process model as an input - in this case odometry data from the robot. Using polar coordinates, the position vector of the person relative to the robot can be written:

\[
\vec{\rho}_{pers} = \begin{bmatrix} d \cos(\phi(t)) \\ d \sin(\phi(t)) \end{bmatrix} \quad (8)
\]

where \( d \) is the distance to the person, and \( \phi(t) \) is the angle to the person. The change, i.e. the derivative becomes

\[
\dot{\rho}_{pers} = \begin{bmatrix} -d \sin(\phi(t)) \\ d \cos(\phi(t)) \end{bmatrix} \quad , (9)
\]

where \( \dot{\phi}(t) \) has the opposite sign of the rotation of the robot itself \( \psi \), so the derivative becomes

\[
\dot{\rho}_{pers} = \begin{bmatrix} -d \sin(\phi(t)) \\ d \cos(\phi(t)) \end{bmatrix} (-\psi) = \begin{bmatrix} p_{y,pers} \\ -p_{x,pers} \end{bmatrix} \tilde{\psi}. \quad (10)
\]

Using an Euler integration, we can substitute \( \Gamma \ddot{u} \) in the model by

\[
\Gamma \ddot{u} = \begin{bmatrix} p_{y,pers}(k) \\ -p_{x,pers}(k) \end{bmatrix} T \tilde{\psi}, \quad (11)
\]

where \( \tilde{\psi} \) is the estimated robot rotation from the odometry data.

The velocity vector is not necessarily equal to the pose of the person. Consider a situation where the person is standing almost still in front of the robot, but is moving slightly backwards. This means that the velocity vector suddenly is in the opposite direction, but the actual pose is the same. Therefore the velocity estimate is filtered through a first order autoregressive filter. The filter is made with adaptive coefficients relative to the velocity. So when the person is moving fast, we rely very much on the direction of the velocity, but if the person is moving slow, we do not change the pose estimate very much and rely on the previous estimate. The autoregressive filter

\[
\theta(k+1) = \beta \theta(k) + (1 - \beta) \arctan \left( \frac{v_{y,pers}}{v_{x,pers}} \right), \quad (12)
\]

where \( \beta \) is chosen relative to the absolute velocity \( v \) as:

\[
\beta = \begin{cases} 0.9 & \text{if } v < 0.1 m/s \\ 1.04 - 1.4v & \text{if } 0.1 m/s \geq v \leq 0.6 m/s \\ 0.2 & \text{else} \end{cases} \quad (13)
\]

**B. Evaluating a Person’s Willingness to Interact**

Although the robot is not perceived as a human being when encountering people, the hypothesis is that human behavioral reactions are the same as in human to human encounters. If a person is interested, he or she will undoubtedly approach the robot in a straightforward manner. On the contrary, if in no interest the person will carefully avoid the path of the robot.

However, there may be many trajectories where the interest of the person will be difficult to determine, i.e. many...
valid trajectories are possible and trajectories variate for each robot-person encounter. In previous work [1] an adaptive person evaluator based on a Case Based Reasoning (CBR) system has been used. The CBR system is basically a database system which holds a number cases describing each encounter. The specification of a case is a question of determining a representative set of features connected to the event of a human robot encounter which can serve as input. In this case we use the features velocity, position and pose illustrated in Fig. 1. The output is the Person Interest indicator, \( P.I. \) which store the probability or indication of the detected person’s interest to interact. The value 0 indicates no interest whereas 1 indicate interest.

The starting point of the CBR system is an empty database holding no a priori correspondence between trajectories and interest. When interacting, interest is confirmed by handover on an object to the robot. Lack of interest is triggered if no interaction has occurred after a fixed period of time. By adding cases, the system gradually learns how to decode trajectories into a interest level.

C. Human-aware Navigation

The robot behavior is inspired by the spatial relation between humans (proxemics) as described in [9]. Hall divides the zone around a person into to four categories, 1) the public zone > 3.6m, 2) the social zone > 1.2m, the personal zone > 0.45m, and the intimate zone < 0.45m. Social spaces between robots and humans were studied in [18] supporting the use of Hall’s proxemics distances and the human robot interaction is therefore designed to be able to experiment with different distances.

For modeling the robots navigation system a person centered potential field is introduced. The potential field has high values where the robot is not allowed to go, and low values where the robot should be. All navigation is done relative to the person, and hence no global positioning is needed in the proposed model. The method is described in [1], but is slightly changed in this implementation. The potential field is designed by the weighted sum of four Gaussian distributions of which one is negated. The covariances of the distributions are used to adapt the potential field according to \( P.I. \).

The four Gaussian distributions are illustrated in Fig. 2 and has the following functions:

**Attractor**, this is a negated distribution used to attract the robot to the person. Its variances \( \sigma_x^2 \) and \( \sigma_y^2 \) are both set to 7.5 and its covariance \( \sigma_{xy} \) is set to 0.

**Rear**, this distribution ensures that the robot does not approach a person from behind. Its variances \( \sigma_x^2 \) and \( \sigma_y^2 \) are respectively set to 2 and 1 and its covariance \( \sigma_{xy} \) to 0. It is only applied when the robot is behind the person.

**Parallel**, this distribution is initially placed with its major axis parallel to the \( x_p \)-axis in the persons coordinate frame. Its variances and covariance are adapted according to the person interested in interaction.

**Perpendicular**, this distribution is initially placed with its major axis perpendicular to the parallel distribution.

Its variances and covariance are adapted according to the person interested in interaction. This distribution as well as the parallel is only applied when the robot is in front of the person.

The attractor and rear distribution are both kept constant for all instances of the person interest indication \( P.I. \). But the parallel and perpendicular distributions are interactively scaled and rotated according to the person’s changing \( P.I. \) during interaction. This means that the robot continuously adapt its behaviour to the current \( P.I. \) value. The potential functions are scaled and adjusted according to Hall’s proximity distances, and the preferred robot to person encounter reported in [6].

The resulting potential field contour can be seen in Fig. 3 for three specific values of \( P.I. \). In the extreme case with \( P.I. = 0 \) the potential field will like look Fig. 3(a) where the robot will move to the dark blue area, i.e. the lowest potential app. 2 meters in front of the person. The other end of the scale for \( P.I. = 1 \) is illustrated in Fig. 3(c), where the person is interested in interaction and as result the potential function is adapted so the robot is allowed to enter the space right in front of him or her. In between, Fig. 3(b), is the default configuration of \( P.I. = 0.5 \), in which the robot is forced to encounter the person in approximately 45°, while keeping just outside the personal zone.

Instead of just moving towards the lowest point at a fixed speed, then the gradient of the potential field is derived. This allows the robot to move fast when the potential field is steep, for example if the robot has to move fast away from a person if getting in the way. On the other hand, the robot has slow comfortable movements when it is close to where it is supposed to be, i.e. near a minimum of the field.

III. Experimental Setup

In order to validate all parts of the system, the above algorithms have been tested in three steps by integrating one feature at the time. First the basic person detection and pose estimation algorithm is tested. Then this is combined with the human aware navigation. Finally, also the person interest estimation has been included. Each step has been validated...
in real world experiment. All experiments have been done with only one human in the area, since the purpose is to demonstrate the proof of concept of the methods. However, the methods should be valid with more persons around the robot.

A. Test Equipment and Implementation

The basis for evaluation of the proposed methods was a FESTO Robotino platform on which a head, capable of showing simple facial expressions, is mounted (see Fig. 4). On the platform, the robot control software framework Player/Stage [4] has been implemented, which also enables simulation before real world tests. The robot is equipped with a URG-04LX line scan laser range finder and a contact to press if you are interested in interaction. The case database has been implemented using MySQL.

A 3D motion tracking system from Vicon (typically used for indoor UAV applications) has been used to validate the functionality of the algorithm through laboratory experiments.

B. Pose Estimation

The pose estimation algorithm was validated through laboratory experiments. The pose algorithm was tested isolated from the system, i.e. all navigation and learning algorithms were disabled. During the experiment, the robot was placed on a fixed position in the lab, while a test person entered the robot field of view and wandered around following different patterns at velocities around $0.5 - 1.5 \frac{m}{s}$. While the robot estimated the pose of the test person, the stationary tracking system was concurrently reading the movements.

C. Human Aware Navigation

In this step, the human aware navigation algorithm was added to the pose estimation system and tested in the robot laboratory. In order to isolate the navigation algorithm, the level of $PI$ was set to a fixed value through each experiment and was completed for $PI = \{0, 0.5, 1\}$. As in the former experiment, the stationary tracking system was set to read the position of the test person while he would approach the robot following different patterns. Since the navigation algorithm was enabled in this experiment, the movement of the robot was also tracked.

D. Integration Test

In this step, the complete system was tested, i.e. the CBR system was added. The test took place in a foyer at the University campus with an open area of 7 times 10 meters. This allowed for easily repeated tests with no interference from other objects than the test person. The contact on the robot was used to get feedback from the test persons whether they were interested in interaction or not. The test persons were asked to approach or pass by the robot in different ways. In approximately half the cases, the test person would end the trajectory by pressing the switch indicating interest in interaction. The test started with an empty database, so that the robot had no experience to start from. During the experiments the $PI$ values, the position, pose estimates were logged in the database.

IV. RESULTS AND DISCUSSION

A. Pose Estimation

Fig. 5 shows the different trajectories performed. The black dots are the position estimates from the robot, and the lines extending from the dots are the corresponding pose estimates. Some underlying coloured dots can be seen in Fig. 5(a) and in the left line in Fig. 5(b). These are the correct position estimates as measured by the motion tracking system. The correct underlying positions are omitted in the other plots, since they clutter the image unnecessarily. In Fig. 5(a), a u-turn movement is performed where the person
stands still for a few seconds close to the robot in the lower left corner. This demonstrates that even though the person stands still (and might even move slightly backwards), the pose estimate keeps being correct towards the robot. When the robot recognizes a person, it assumes that the pose is close to 0, which explains the blue lines always pointing towards the robot in the beginning of a trajectory. However, as soon as the person start to walk, the pose estimate turns and follows the motion. Note that it is not the Kalman filter which estimates the velocity wrong, when the lines do not follow the trajectory exactly. But it is the autoregressive filter that does not allow the pose estimate to change too quickly. The figures show that the pose estimator works satisfactory and can be used to estimate pose in real world. Although this laboratory test confirm that the pose algorithm works as expected, the test is limited to only a few test runs and the algorithm is not really fine tuned yet.

B. Human Aware Navigation

In this experiment, the robot is set to move according to prespecified person indication (PI) values. The person moves from the bottom left corner (in Fig. 6(a)-6(c)) towards the robot. The elapsed time is shown by the trajectory colour changing from blue to red. When $PI = 0$, it can be seen that the robot tries to get away from the person. When it reaches a comfortable distance, it settles around that position relative to the person. When the person is partially interested, the robot avoids the person and tries to stay at an angle of approximately 45° degrees. Finally, when the person is interested, the robot approaches the person from the front until the border of the intimate zone is reached at around 45 cm. The major colour change at the person trajectory indicates that the person has been standing still. As soon as the person starts to move away, the robot finds out that it is behind and too close to the person, so it starts to move away.

The shaky sinusoidal movement of the person trajectory is due to the tracking of the person. It is caused by the fact that the central part of the body moves like this when a human walks. The experiment proves that the potential field enables the robot to keep at the correct position relative to the person, and that the pose estimator also works when the robot is moving.

C. Integration Test

The output of the integration test was a trained CBR database. The database can be seen in Fig. 7. The dots show recorded test person positions, which was rounded to a grid size of $40 \times 40$ cm. Note that the positions are not global coordinates, but relative to the robot while it is moving. The extending lines, show the direction of the pose estimates, and the colour the corresponding $PI$ value, where red indicates
an interested person, and blue indicates a person which is not interested. The database shows that persons right in front of the robot with a pose close to $0^\circ$ is typically interested in interaction, whereas a pose pointing away from the robot indicates that the person is not interested. The results reflect the fact that the potential field makes the robot move so only persons who are estimated as interested are allowed to come close to the robot.

Clearly, the experimental work is still in its initial stage and is not exhaustive. However, the tests demonstrates the potential of the methods and of combining the pose estimation algorithm with the proposed method for human aware navigation.

V. CONCLUSIONS

This paper describes a new method for estimating the pose of a person in an interaction scenario with a mobile robot. The algorithm only relies on laser range finder data, which makes it applicable for moving persons at larger distances than normal vision techniques allow. A Kalman filter is used to filter the measured positions of persons within view and outputs a pose estimate.

The position and pose estimates are used in a Case Based Reasoning system to estimate the person’s interest in interaction, and the spatial behavior strategies of the robot are adapted accordingly using adaptive potential functions. The human robot interaction methodology described in this paper is supported by laboratory and real world tests which demonstrate the applicability of the pose estimator and the spatial behaviour of the robot.

The real world tests demonstrate the potential of the integrated system, which can be used for robots moving in human environments. Generally, the conducted experiments on the robots cognitive functionality show that the method of CBR implemented can advantageously be applied to a robot, which needs to evaluate the behavior of a person.

An interesting aspect for future work would be to combine this pose estimation technique with a vision based technique.

This could give a more accurate pose estimate for close interaction.

REFERENCES


