Tracking of People in Paper Mill Warehouse Using Laser Range Sensor

Hassan Nemati
Intelligent Systems Lab
Halmstad University
Halmstad, Sweden
hassan.nemati@hh.se

Björn Åstrand
Intelligent Systems Lab
Halmstad University
Halmstad, Sweden
bjorn.astrand@hh.se

Abstract — In this paper a laser scanner based approach for simultaneous detection and tracking of people in an indoor environment is presented. The operation of an autonomous truck, for transporting paper reels in a dynamic environment shared with humans, is considered as the application setting for this work. Here, a human leg detection procedure and an Extended Kalman Filter (EKF) based tracking method are employed for real-time performance. Several experiments with different data sets collected from an autonomous forklift truck in a paper mill warehouse have been performed in an offline situation. The results show how the system is able to detect and track multiple moving people.

Keywords — human detection; tracking; extended kalman filter; autonomous vehicles; load handling; intelligent systems

I. INTRODUCTION

Detection and tracking of people is one of the most challenging issues in terms of robotics, intelligent vehicles, surveillance and safe automation. For this purpose, two important tasks should be performed by the autonomous vehicles. First, they should navigate and localize their position, and second, analyse the dynamic environment in their vicinity. The latter concern is very essential when the vehicle needs to work with other autonomous vehicles in an environment shared with humans.

Various techniques and algorithms have been proposed in literature for movement and human detection. The detection of moving objects, by using a 2D laser scanner, can be done simply by comparing two consecutive laser scans [1] and [2] or by employing learning techniques to create classifier for people detection [3].

The number of lasers varies in different methods. In [4] three laser in different layers are employed to detect legs, upper body, and the head of human. This method uses a supervised learning algorithm to train one classifier for each layer. The classifiers are combined in a probabilistic way for human detection. In another approach [5] a double layered laser range finder is exerted to detect chest and legs.

According to the information of these two scanners a person model will be created. Then, based on this model, the human position in an indoor environment can be estimated.

Moreover, there are some other techniques where a combination of visual and laser based people detection is employed [6-9]. In [10], how efficiently this type of fusion could improve the performance of human detection is examined.

After movement detection, more advanced algorithms are needed for tracking. The human tracking is addressed in [1], [2], and [11] using Kalman filter (KF). This method is valid even if the moving objects are occluded but considers tracking as a linear problem. In [12] and [13] another tracking method, based on extended Kalman filter (EKF) is presented to include the nonlinear characteristics of the motion. Some other techniques, such as in [14] have introduced methods which are based on joint probabilistic data association filters (JPDAF) to take into account any non-Gaussian noise in the system.

There are many research groups who are working on autonomous trucks and forklifts [15-17]. In 2006, the Real Systems Group (RTS) of Hannover University presented an unmanned forklift truck, based on two laser scanners, which was able to perform localization and picking-up pallets [18]. After that in 2008, the RTS developed their material handling autonomous forklift and called it as “FMX autonom” [19]. The truck was able to detect and load pallets in a warehouse and bring them to a predefined destination by using a 3D laser scanner. Furthermore, a self-localization system was introduced based on ceiling structures and 3D analysis. The experiment was held in a static environment, where moving objects were not considered.

In [20], an Automatic Guided Vehicle (AGV) system is presented. The AGV is an unmanned vehicle which can transport loads in dynamic environments. This application consists of picking up loads from particular spots and unloading them at some predefined places. Moreover, collision avoidance is discussed in two situations: a) with other AGVs, and b) with static obstacles. However, the human movement is not mentioned or discussed.

The researchers in Computer Science and Artificial Intelligence Laboratory (CSAIL) in Massachusetts Institute of Technology (MIT) with Laboratory for Information and Decision Systems have created a truck which is able to interact with humans in a shared environment [21]. The truck can transport pallet loads under voice commands in a minimally-prepared outdoor environment. In their experiment, they have used 15 laser sensors and 4 cameras around the vehicle at different levels. Accordingly, employing this number of sensors increases the cost of the vehicle and needs more repair and maintenance.

In this paper, we present a human detection and tracking system in highly dynamic environments. The implemented method is based on only two laser scanners mounted on a forklift truck. The aim is to present a real time, low cost and efficient people detection and tracking system for autonomous trucks. These trucks are used for loading and
unloading materials in industrial environments shared with humans. By acquiring perception of human movement, the truck can perform more optimal path planning and consequently operate safely in highly dynamic environments. Moreover, it can drive continuously to reach its destination instead of stopping regularly to avoid hitting human workers.

The remainder of this paper is structured as follows. Section II explains the system description and working environment. In section III, the first two steps of laser based systems: segmentation and feature extraction steps are explained. Section IV is devoted to human movement detection, while section V describes the EKF based tracking. In section VI, the experiments and results are presented and finally we draw conclusion in section VII.

II. SYSTEM DESCRIPTION

The truck, in the experiments, is used in paper production facilities and warehousing terminal. In the warehouse, paper reels have a cylindrical shape and are transported by the manually driven trucks.

The autonomous vehicle is a Linde H50D diesel forklift truck which is fitted with a clamp for gripping paper reels. Three SICK 2D laser scanners are mounted on the truck. In front of the truck, one S300 laser sensor is employed for localization and human detection, and one S300 for paper reels detection. At the backside of the truck, another S300 is mounted for tracking human activities.

To control the truck, an AGV system is employed. The control system provides localization and path planning from initial position to destination by using reflective markers installed in the environment. The working environment and the forklift truck are shown in Fig. 1 and Fig. 2 respectively.

![Figure 1. Paper mill warehouse](image1)

![Figure 2. The Linde H50D diesel forklift truck](image2)

A general DTMO (detection and tracking moving objects) system architecture [22] for human detection and tracking procedure is shown in Fig. 3. In this system, the first step is segmentation which is detecting distinct objects at each laser scan. These distinct objects are classified as geometrical features e.g. lines, circles (paper reels), ellipse, and blobs in the feature extraction step. Each object could be grouped into two models: Background model and Foreground model. The background model consists of information that belongs at the background of the field of view and from that the position of paper reels can be obtained. The foreground model estimates the position of objects which are in front of the background. The foreground model might include people, other vehicles, and static objects with small sizes.

At each scan, the foreground objects of the current state are compared with their previous state. If there is a difference between their positions, it is concluded that the corresponding object is a moving object; otherwise, it is a static object. The detected moving objects are sent to tracking step. In the tracking step, it should be possible to predict a position ahead and be able to alleviate the consequence of occlusion.

![Figure 3. DTMO System architecture](image3)

III. SEGMENTATION AND FEATURE EXTRACTION

Segmentation is the first step in most laser based algorithms which is classifying the scan data into a set of connected objects [1], [2], [23]. The segmentation algorithm is based on an assumption that there is a gap between each distinct object. One of the disadvantages of such an approach is that the tracking becomes difficult, for example in case of a group of people moving together. In this case, the algorithm defines the group as one moving object. Other methods such as background subtraction and clustering could be used to solve such a problem [24]. However, the group of people and the number of people in a group is not distinguishable easily in our experimented situations. The reason comes from the fact that, relative position of people to the forklift might vary between 0 to 20 meters. Therefore, in some cases, only a single point (or two points) declares a human leg on the field. In such situations, detecting a group of people is not possible.

After the segmentation step, the connected points must be determined as features and objects. For most indoor applications, finding primitive features like lines, circles,
arcs, ellipses and corners are precious and easy to extract. Accordingly, the connected points could be classified into four categories: lines, circles, ellipse, and blobs. Blobs are defined as segments with length smaller than 50cm and can be easily extracted. The term 'length' in this regard is the Euclidian distance between the first and the last point in a segment.

IV. HUMAN MOVEMENT DETECTION

The result of feature extraction algorithm could be classified as background model and foreground model. All the detected lines, circles, and ellipses are considered as static objects and grouped as background model. The foreground model contains static objects with length smaller than 50cm, and moving people in the field. Since the diameter of human’s leg is always smaller than 50cm, this algorithm guarantee to find the legs.

Movement detection algorithm is employed to distinguish the difference between human movement and static objects. In this algorithm, the blobs in two consecutive scans are stored into two matrices and compared i.e. 1) newly scanned blobs from the current scan, and 2) old blobs from the previous scan. If there is a distinct distance between these two generations of blobs, the corresponding blob is classified as a humans’ leg; otherwise, it is a static object. In fact, this algorithm could only detect human movement in the field.

Because the laser is mounted on a moving vehicle, the laser and world coordinate systems are not the same. Thus, each candidate scan point \( p_i = \left( \theta_i, r_i \right), i \in [1,2,...,N] \) must be converted from laser to world coordinate system by the following equations:

\[
\begin{align*}
x_i(t) &= X_{Track}(t) + r_i \cos(\alpha_{Track}(t) + \theta_i(t)) \\
y_i(t) &= Y_{Track}(t) + r_i \sin(\alpha_{Track}(t) + \theta_i(t))
\end{align*}
\]

Where \( (X_{Track}(t), Y_{Track}(t)) \) represent the current position of the vehicle in the world coordinate system at time \( t \). After this conversion, the two consecutive set of blobs could be compared together in the same coordinate system.

The next step in the human detection algorithm is finding the position of a person according to his legs coordinate. In this case, each detected single blob (human’s leg) corresponds to a person. If two legs are very close to each other (less than 40 cm), they are considered to belong to the same person. Then the position estimation of the human body is calculated based on the mean value of center of gravity of each leg.

In literature, finding position of a person based on his legs coordinate and walking model is called people modeling. This kind of modeling depends on several variables such as motion velocity and the length of the person’s leg. In [25] a simplified walking model is proposed which can estimate the position of a person. In this method, the distance between each leg is considered as \( 2L \sin \theta \), where \( L \) is the length of human’s leg and \( \theta \) is angle between leg and body. Moreover, the velocity is assumed constant. However, in our algorithm, the human position is estimated by averaging the position of two legs. This method can be applied in any condition without making any assumption about the length of a person’s leg or the moving velocity.

In the next step, these detected human legs are sent to the tracking procedure.

V. EKF BASED TRACKING

After finding people, the truck needs to be able to track them in order to perform path planning and safe navigation. The main purpose of tracking is to assign each individual measurement to the detected objects. The general way to do this assignment is to use Kalman filter, and in nonlinear situations, Extended Kalman filter [26-28]. In this section, extended Kalman filter based tracking method is investigated.

In the EKF [26-28], the system is considered as a nonlinear model with Gaussian noise. Thus, the system is estimated by a differentiable function as:

\[
\begin{align*}
\hat{X}_k &= f(\hat{X}_{k-1}, u_{k-1}) + w_k \\
Z_k &= h(X_k) + v_k
\end{align*}
\]

Where the non-linear function \( f \) relates two consecutive states \( (k \text{ and } k-1) \) and it contains the control input \( u_{k-1} \). The nonlinear function \( h \) relates the state \( X_k \) to the observation \( Z_k \). The \( v_k \) and \( w_k \) are the observation (measurement) and the system (process) noise respectively. They are assumed to be uncorrelated with each other, white, and with Gaussian distribution.

At the prediction phase, a posteriori estimation of the system and noise covariance will be determined by the following equations:

\[
\begin{align*}
\hat{X}_k &= f(\hat{X}_{k-1}, u_k) + w_k \\
\hat{P}_{kk-1} &= F_k\hat{P}_{k-1,k-1}F_k^T + B_k\Sigma_kB_k^T
\end{align*}
\]

In these equations \( P_{kk-1} \) is the prior estimate covariance matrix at time \( k \) given its value at time \( k-1 \) and is calculated by the law of error propagation (in the matrix form).

The matrix \( \hat{X}_k \) could be written as:

\[
\hat{X}_k = f(\hat{X}_{k-1}, u_k) = \begin{bmatrix} x_{k-1} + d_k \cos(\theta_k + \alpha_k) \\
y_{k-1} + d_k \sin(\theta_k + \alpha_k) \\
\theta_k + \alpha_k \end{bmatrix}
\]

In this equation, \( d_k \) is the relative movement of the object at time \( k \) compared to time \( k-1 \), \( \theta_k \) is the current direction of object, and \( \alpha_k \) is the difference angle between two consecutive movements (at times \( k \) and \( k-1 \)). Matrixes \( F_k \) and \( B_k \) in (2) are the Jacobian matrices of partial derivatives of the function \( f \) with respect to state variable \( X \).
\((x, y \text{ and } \theta)\), and input variables \(u(t)(V, a)\), which consists of velocity and movement angle, respectively:

\[
F_k = \begin{bmatrix}
1 & 0 & -d_x \sin(\theta_t + \alpha_k) \\
0 & 1 & d_y \cos(\theta_t + \alpha_k) \\
0 & 0 & 1
\end{bmatrix}
\]

\[(4)\]

\[
B_k = \begin{bmatrix}
\cos(\theta_t + \alpha_k) & -d_x \sin(\theta_t + \alpha_k) \\
\sin(\theta_t + \alpha_k) & d_y \cos(\theta_t + \alpha_k) \\
0 & 1
\end{bmatrix}
\]

\[(5)\]

Furthermore, the \(\Sigma_u\) in Eq.2 is the system noise covariance matrix at time \(k\).

The correction step contains the following calculations:

- Innovation covariance: \(S_k = HP_{k-1}H^T + R\)
- Kalman gain: \(K_k = P_{k-1}H^T S_k^{-1}\)
- Updated state estimate: \(\hat{X}_k = \hat{X}_{k-1} + K_k(Z_k - \hat{X}_{k-1})\)
- Updated or posteriori estimate covariance:

\[
P_{k} = P_{k-1} - K_kH_{k}\]

\[(9)\]

where

\[
Z_k = [x \ y] \quad R_k = \begin{bmatrix}
\sigma_x^2 & 0 \\
0 & \sigma_y^2
\end{bmatrix} \quad H_k = \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

Consider that, there is no information about the human direction from the observation \(Z_k\). In fact, beacons just give us information about position with a fixed covariance \(R_k\) and not about orientation. Therefore, in the above equations, \(R_k\) and \(H_k\) only include \(x\) and \(y\) coordinates.

At each time step, an EKF is assigned to the moving objects to estimate their next positions. In the next scan, newly identified targets are matched with these estimated positions. The data association between multi-target tracking is based on nearest-neighbor approach. This approach is performed by defining a fix and certain threshold.

The state estimation of moving objects contains an area of uncertainty based on the estimated covariance matrix. The size of this area indicates how much the procedure is sure about the position estimation. This uncertainty area could be used to estimate the position of a person when he is hidden by an obstacle.

VI. EXPERIMENTS AND RESULTS

Several tests have been carried out with the Linde H 50D diesel forklift truck in the paper mill warehouse. The collected scanned data are explored for human detection and tracking in an offline situation. Although, the processing time of the system is short enough to be used in the real time procedures. As mentioned earlier, the truck was fitted with two 2D laser scanners for human detection at ground level: one in the front of the truck, and another in the backside of the vehicle. In these experiments, the unmanned truck was driven while people were moving in the environment. Fig. 4 and Fig. 5 show our first attempt in two different scans, where the truck was driving on the field and two people were following it. In Fig. 4 (a), the relative coordinate system of the truck and the corresponding scan points for the rear and front laser scanners are depicted. Moreover, the estimated position of paper reels and their corresponding center which are achieved in the feature extraction step are shown. The Fig. 4 (b) shows the human position during this experiment in the world coordinate system; while Fig. 4 (c), shows the path taken by the truck. Accordingly, the system is able to detect and track people in a dynamic environment.

In this experiment, the total distance traversed by the truck is about 168 meters. During this path, two people are walking around the truck. Along this route, one person is correctly detected and tracked in about 155 meters while the other one is tracked for about 140 meters.

Note that, the rear laser scanner is placed at position \([-2.2 0.0 3.14]\) while the front laser scanner is mounted at position \([0.46 0.0 0.0147]\). Therefore, there is an uncovered gap between these two laser sensors. Indeed, if a person walks along with the truck between these areas he will not be visible for the truck. This situation is marked in Fig. 5 (b) with a ‘T’.

The feature extraction algorithm is able to correctly detect paper reels and blobs in the environment within 20
mometer around the truck. Moreover, the human detection algorithm can detect human movements between candidate blobs. However, in some cases, it falsely detects static blob as a human and a human as a static object (false acceptance and true rejection respectively). Furthermore, in the tracking step, the system is able to correctly track the pass of human movement in its surrounding.

To evaluate the False Acceptance (FA) in the detecting algorithm, the truck was driven in the environment while there were no human activities around it. The result is demonstrated on Fig. 6 and shows the amount of falsely accepted static objects as human movement on the environment by blue dots. In this experiment, 28 objects are falsely accepted as human. Accordingly, these FAs are shown as single points which declare that only in a single scan (maximum two) they are detected as human movement. Thus, by filtering these kinds of detected movements, the FA blobs could be removed from the list of humans to be tracked.

![Figure 6](image)

Figure 6. (a) Relative coordinate system of the truck. (b) Falsely accepted points as human movement during the experiment shown in Fig. 5. (c) Path taken by truck in world coordinate system.

Fig. 7 and Fig. 8 show two different experiment scenarios. In Fig. 7 the truck is stand still but Fig. 8 depicts the situation where the truck is moving forward and backward for a few times while people are walking around it. In these experiments only the rear laser scanner is examined and used for people tracking. Indeed, the system can correctly detect and track multiple people walking around the vehicle.

VII. CONCLUSION

Interaction between autonomous vehicles and humans is one of the long-standing challenges in robotics and intelligent vehicles. In this regard, a safe operation of the vehicle in human existing workplace is essential.

In this paper a system for detection and tracking of multiple people in an industrial environment is employed. The vehicle in the experiments is an autonomous forklift truck which is used for loading and unloading materials in a paper mill warehouse. Two 2D laser scanners are mounted in the front and in the backside of the truck and used for human detection and tracking.

The first step in the system is segmentation and feature extraction where the paper reels and blobs (as candidate features of human’s leg) can be detected. In the next step, the problem of finding and detecting moving humans is solved by comparing features and especially blobs in two consecutive scans. In the experimented results, the detection phase had some falsely accepted features classified as human’s leg. This situation could be relieved by filtering the single scan movement. Although, employing two laser scanners for each side of the truck (front and rear) or using 3D laser scanner to detect different level of the human body are more suitable methods to resolve the problem of human detection. This could significantly reduce FA and TR in the human detection step.

The last section in the system is tracking multiple moving people in the truck’s surrounding. This could be achieved by using tracking methods such as EKF. The EKF based tracking approach calculates the maximum likelihood estimation to perform ideal prediction of the state. In this case, the posteriori probability of the true state of the moving object at the current scan is calculated based on the previous state of the object and the current observation from laser sensor.
Several experiments have been performed in an offline situation and the results show how efficiently the system can detect and track multiple moving people. Consequently, the truck can perform optimal path planning, cost saving, and safe operation.

ACKNOWLEDGMENT

This work has being carried out with data from the MALTA (Multiple Autonomous Forklifts for Loading and Transportation Applications) research project supported by the Swedish KK foundation. The authors would like to acknowledge Henrik Andreasson at Örebro University for his support with data collection.

REFERENCES


