Pattern Recognition and Tracking Dynamic Objects with LIDAR

Levente TAMAS, Gheorghe LAZEA
Robotics Research Group, Technical Univeristy of Cluj-Napoca, Romania
{levente.tamas, gheorghe.lazea}@aut.utcluj.ro

Abstract

This paper presents a method to detect moving persons based on the information from a LIDAR type sensor mounted on a mobile robot. The detection of the objects is performed relative to the estimated robot position. The objects of interests are first described with Gaussian Mixture Model (GMM) and later on searched and classified with a Bayesian classifier. The estimated person positions are tracked via the Extended Kalman filter with two kind of motion models for humans.

1 Introduction

The perception capabilities of the mobile robots can be improved if multiple sensory information is fused in order to gain more relevant information as a result of the combination of several different sensors. This paper presents a multi-sensor architecture for processing the mobile robot's surrounding environment information for detecting moving person in order to avoid collision in an indoor environment. Examples of such moving object may be people or other mobile robots [1].
In the proposed architecture lidar based system is meant to detect and track the position of the predifined dynamic objects like humans. Similar strategies can be adopted in order to establish the coordinate correspondence between the lidar or the monocular vision camera to reduce the field in which the object detection is performed. The robot position estimation was based on traditional dead-reckoning sensors and the Kalman filtering algorithm [2].

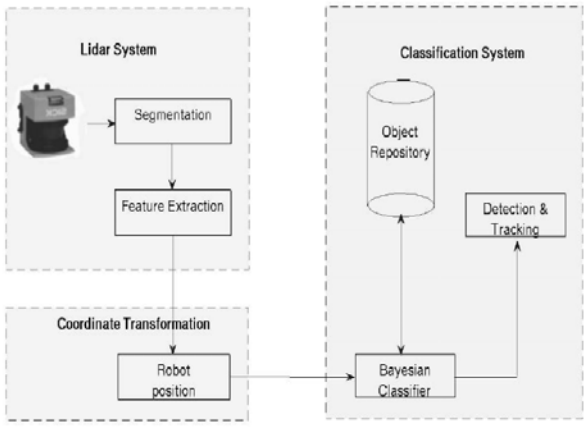

Figure 1: Architecture of the system

The architecture of the system is presented on Figure 1 is and composed from three subsystems: the robot position estimator, the lidar based classifier together with the coordinate transformation system the global classification subsystems. Based on the relative position information of the robot, the people relative to the robot are detected by the lidar based system. Also the this system can be used to measure the distance of the object relative to the robot with a good accuracy [3].
The people detection with lidar is based on the Gaussian Mixture Model representations of the predefined leg forms [4] and to extract the depth information too of the detected person.

2 Related Work

The robot localization problem represents a key aspect in making a robot a real autonomous one. The position of the robot has to be estimated accurately based on the information from the sensors about the surrounding world [2]. The human detection and tracking is an essential part of the human-robot interaction problem. This topic represents a major interest in the autonomous vehicle research domain [5].
The current trend in this field [6] is to fuse together relative and absolute measurements. The aim of this is to provide a better position estimation of the robot location based on the differing nature of the data from different kind of sensors. In order to fuse the information from different sources one of the widely applied methods is Kalman filtering [7].
In general detecting different objects on a moving platform using lidar and vision, or both sensors at the same time, for collision avoidance, mapping (or SLAM together) is well reported subject [8] [9].
Several research work have been performed using laser-scanners in object classification and moving object tracking including but not limited to localization and navigation[10] application or guarding systems [11]. For the object classification voting schemes, multi-hypotheses tracking [12] or even boosting approaches [13] were used. While the first two approaches lack the proper mathematical description framework they still offer reasonable performance.
The vision based systems are commonly used for object detection and classification with or without lidar [14]. In certain light/ambient conditions the performance of the vision system can degrade, and the range information .

ISBN 978-3-8007-3273-9

3 Lidar Based Classifier

In this section the lidar classifier is presented with the segmentation, feature extraction and classification components. Basically, the lidar measures bearing-range information about the surrounding objects with a relative good accuracy (in the performed experiments $1cm$ accuracy at a $10m$ range).

3.1 Scan Segmentation

The scan segmentation belongs to the primary modules of the lidar architecture among with the data acquisition and pre-filtering modules. The segmentation is the process of splitting a scan into several coherent clusters, i.e. point clouds. The choice of segmentation method is rather arbitrary and dependents on other design choices as the alignment and covariance estimation strategies [15]. The current strategy is the one based on the simple assumption of topological distances between segments adopted from [13].

The laser range scan information is a set of beams of the form $Z = \{b_1, ..., b_L\}$ Each element b_j of this set is a pair of (θ_j, ρ_j), where θ_j is the angle of the beam relative to the robot and ρ_j is the distance from the reflected surface.

The scan Z can be split into subsets according to the distance threshold computed for the segment. In case that the topological distance between two segments is greater than a preset threshold then a new segment is considered. Even if there are more sophisticated segmentation algorithms e.g. like the one presented by Premebida in [4], in the current problem setup we found appropriate this simpler approach.

The output of the splitting procedure is an angle ordered sequence $\mathscr{P} = \{S_1, ..., S_M\}$ of segments in such a way that $\bigcup S_i = Z$. The elements of each segment S contain pairs f Cartesian coordinates $\mathbf{x} = (x, y)$ which can be converted to polar coordinates with $x = \rho\cos(\theta)$ and $x = \rho\sin(\theta)$.

The adopted option is to consider the measurements in angle-order and segment point clusters at range discontinuities. By considering a full scan as an ordered sequence of $N_{\mathscr{S}}$ points in the form: $\mathscr{S} = \{(r_k, \theta_k) | k = 1, ..., N_{\mathscr{S}}\}$ in which case (r_k, θ_k) denotes the polar coordinates of the i^{th} scan. The point goups of a segment can be expressed as

$$S_j = \{((r_n, \theta_n))\}, \quad n \in [k_{begin}, k_{end}], \quad j = 1, ..., N_S \quad (1)$$

where N_S is the number of detected segments, k_{begin} and k_{end} are the first and last points in the scan of the current segment. The segment can also transformed in the Cartesian coordinates using a simple polar to Cartesian transformation.

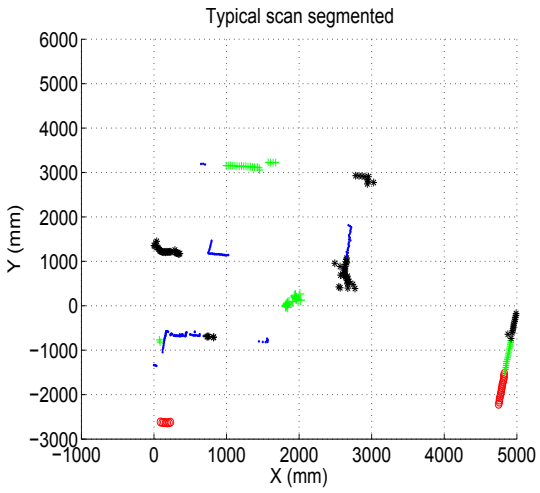

Figure 2: Segmentation output in a cluttered environment

A gating technique is applied in order to filter out the spurious data which can be summarized as follows: if the innovation ν_k is less than a γ_k gating threshold then a break point is observed. The innovation is defined as $\nu_k = \nu_k{}^T S_k^{-1} \nu_k$ where S_k is the associated measurement covariance matrix. The γ_k is chosen from the χ_1^2 test table. A typical scan segmentation can be seen in Figure 2. The segments are denoted with different colors, although in lack of easily distinguishable colors the same color appears as denoting another apart segment on the image. The size of the segments can be tuned via the distance threshold parameter between two consecutive segments.

3.2 Feature Extraction

This module extracts the relevant information from the segmented data and ensures robustness in the algorithm. The extracted information is used later on in the classifier module and can also be used for visualisation purposes too. The feature vector components may be chosen upon the required information [13]. The basic set of feature which was used in the experiments contained the following entries:

1. f1: object centroid;
2. f2: normalized Cartesian dimensions given by:

$$f2 = \sqrt{\Delta X^2 + \Delta Y^2} \quad (2)$$

3. f3: the standard deviation of the point from the centroid:

$$f3 = \sqrt{\frac{1}{n-1} \sum ||r_n - \bar{x}||} \quad (3)$$

These components are essential to the classifier.

ISBN 978-3-8007-3273-9

3.3 GMM Object Description

A Gaussian mixture model (GMM) is a weighted combination of Gaussian probability density functions (pdf). These densities are used to capture the particularities of an object. In a GMM model the probability distribution of a x random variable is defined as a sum of M weighted Gaussian probability density functions:

$$p(x|\Theta) = \sum_{m=1}^{M} \alpha_m p(x|\theta_m) \tag{4}$$

where $\theta_1, ..., \theta_M$ are the parameter of the Gaussian distributions and $\alpha_1, ..., \alpha_M$ is a weighted vector such that $\sum_{m=1}^{M} \alpha_m = 1$. A set of parameters for a mixture model is given by $\Theta = (\alpha; \theta_1, ..., \theta_M)$ where each parameter $\theta_m = (\mu_m, \Sigma_m)$ represents the mean and the covariance of the model with Gaussian pdf. The likelihood of a feature vector Ω is given by the linear combination of the Gaussian mixture probability density functions:

$$p(\Omega|q_i, \Theta^i) = \sum_{m=1}^{M} \alpha_m^i p(\Omega|\theta_m^i) \tag{5}$$

In this case each Gaussian density function for the two dimensional and gives as:

$$p(\Omega|q_i, \Theta^i) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma_m^i|}} e^{[-\frac{1}{2}(\Omega-\mu_m^i)^T (\Sigma_m^i)^{-1} (\Omega-\mu_m^i)]} \tag{6}$$

The Gaussian mixture parameters for each object of interest was determined using the expectation-maximization (EM) algorithm [16] . For each set of feature vectors $(\Omega^N = \Omega_1, ..., \Omega_N)$ the EM algorithm computes M Gaussian parameter vectors that maximizes the joint likelihood of the Gaussian density functions:

$$p(\Omega^N|q_i, \Theta^i) = \prod_{j=1}^{M} p(\Omega_j|q_i, \Theta_m^i) \tag{7}$$

Further more advanced optimization algorithms can be used for the parameter vector searching like *Figueiredo-Jain* method [17], but in this case the optimal parameter searching algorithm is not crucial, thus the faster EM is preferred.

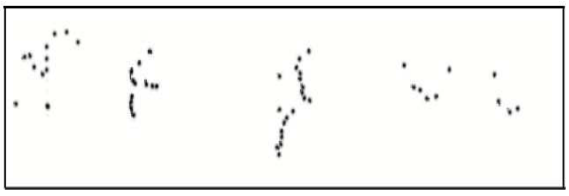

Figure 3: Possible leg forms in the laser scan

For a single leg pair the Gaussian models describing to the same leg forms are shown on Figure 3. As it can be seen on this figure, more Gaussian models are fitted to the leg pair. The number of the Gaussian model can be tuned with a parameter during the EM algorithm used for the fitting. In this case there it is not worth using more than 10 models for a leg pair, as the leg forms are rather simple ones, as this can be seen in Figure 4.

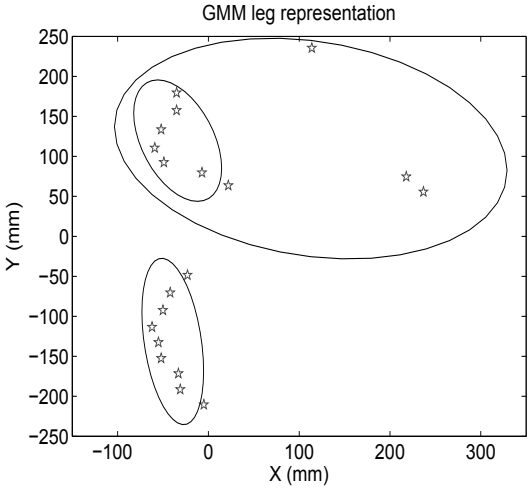

Figure 4: Gaussian models for a leg pair

Although not all the possible leg configurations are captured, with a proper classifier algorithm the distorted leg forms can also be detected in the segmented laser scan [18].

3.4 Bayesian Classifier

After a Gaussian mixture pdf for classified object is available, in order to classify which category (q_i) modelled by Θ^i fits the current observation feature-vector Ω_k a Bayesian decision framework based on the log-likelihood and on the log-prior probability is used.
Computing the log-likelihood has the advantage of reduced computational effort by avoiding the computation of the exponential in the pdf (6) and by turning the product (7) into sums . Furthermore, as the log-likelihood is a monotonically increasing function allows it can be used the former directly to classify the objects.
By considering the features equip probable, the logarithm of the posterior probability $log(P(\Theta^i|\Omega))$ for all categories is proportional to the sum of the log-likelihood of the logarithm of the prior probability:

$$log(P(\Theta^i|\Omega)) \approx log(p(\Omega|\Theta^i)) + log(P(\Theta^i)) \tag{8}$$

It is more convenient to use Bayes' law to estimate the posterior probability as it uses only the likelihoods and the prior probability. The former is computed at each scan, which will become in the next scan the last estimated posterior. Therefore the prior probability is updated dynamically as:

$$P(\Theta_k^i) = P(\Theta^i|\Omega_{k-1}) \tag{9}$$

ISBN 978-3-8007-3273-9

By knowing the initial prior probability for each class, the classification algorithm computes the maximum posterior probability for each segment. In order to decide which is the most likely class of object q_i for the segment S_j a decision rule of the following form was adopted:

$$S_j \in q_i \quad if \quad log(P(\Theta^i|\Omega_k)) = \max(log(P(\Theta^u|\Omega_k))) \quad (10)$$

where u spans from 1 to the number of classes.

4 Extended Kalman Filter for Tracking

A large number of mobile robots use position estimation based on the Kalman filters. Originally the theoretical backgrounds were formulated by Rudolf Kalman in 1960 and later there several extensions were developed [2]. The Kalman filter is an optimal recursive data processing algorithm for linear systems corrupted by noise.
The Extended Kalman filter (EKF) [19] uses a model to describe a discrete-time state transition. The filtering algorithm can be described in two steps: prediction and update. The *prediction step* is done at time instant $k-1$, before the information from the measurement is available and it is based on the previous state estimate $\mathbf{x}_{k-1}^{+}$.

$$\mathbf{x}_k^- = \mathbf{F}\mathbf{x}_{k-1}^+ + \mathbf{B}\mathbf{u}_k \quad (11)$$

$$\mathbf{P}_k^- = \mathbf{F}\mathbf{P}_{k-1}^+ + \mathbf{Q} \quad (12)$$

The *update step* is performed after the measurement from the time step k is available, and includes this information as a correction for the predicted state. This step can be summarised with the following equations:

$$\mathbf{x}_k^+ = \mathbf{x}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^-) \quad (13)$$

$$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^- \quad (14)$$

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R})^{-1} \quad (15)$$

The term with given by the difference between the estimate and the measured state is called the innovation or residual. The innovation sequence should be an uncorrelated, white sequence [6].

4.1 The Motion Models for Humans

Two motion models for were adopted for people tracking. For both models the measured state variables were the people's position in the Cartesian coordination (x_k, y_k).

4.1.1 Position-velocity-heading (PVH) Model

used to estimate the human motion the constant velocity model in our experiment was extended with the orientation ϕ_k and velocity v_k according to [10] as follows:

$$\begin{cases} x_k = x_{k-1} + \delta_k v_{k-1} cos\phi_{k-1} \\ y_k = y_{k-1} + \delta_k v_{k-1} sin\phi_{k-1} \\ \phi_k = \phi_{k-1} + n_{k-1}^{\phi} \\ v_k = v_{k-1} + n_{k-1}^{v} \end{cases} \quad (16)$$

with δ_k being the sampling time, n_{k-1}^{ϕ} and n_{k-1}^{v} the zero-mean Gaussian noises with $\sigma_\phi = \frac{\pi}{16}$ and $\sigma_v = 0.05ms^{-1}$.

4.1.2 Position-velocity-acceleration (PVA) Model

or referred as the $\alpha-\beta-\gamma$ filter [7] is the model of a particle in a Newtonian system represented in 2D a coordinate system. Along a single axes the motion equations are given as follows:

$$x_k = \begin{bmatrix} 1 & \delta_k & \delta_k^2/2 \\ 0 & 1 & \delta_k \\ 0 & 0 & 1 \end{bmatrix} x_{k-1} + \begin{bmatrix} \delta_k^2/2 \\ \delta_k \\ 1 \end{bmatrix} n_{k-1} \quad (17)$$

The same equations are valid for the y_k coordinates. When using this model special care must be taken for computing the model noise, which is a function of the sampling rate δ_k. Additional information on filter tuning can be found in [6].
In both cases the legs position are measured as bearing-range information with relative to the robot's position (x_k^R, y_k^R, ϕ_k^R) as follows:

$$\begin{cases} b_k = tan^{-1}\left(\frac{y_k - y_l}{x_k - x_l}\right) - \phi_k^R + n_k^b \\ r_k = \sqrt{(x_k - x_l)^2 + (y_k - y_l)^2} + n_k^r \end{cases} \quad (18)$$

where (x_l, y_l) are the offset of the laser device with respect to the robot. The noises n_k^b and n_k^r are device specific measurement Gaussian noises, considered for the experimental part $\sigma_b = \frac{\pi}{32}$ and $\sigma_r = 0.05m$.

4.2 Motion Model Comparison for the Tracking

To compare the two models, the EKF was used to estimate the position of the detected person. The same dataset was tested against the computational effort and the standard deviation of the innovation along the x and y axes. The results are summarised in Table 1.

Table 1: Comparison of the PVH and PVA models

Criteria	PVH Model	PVA Model
Runtime (s)	6.2	7.9
$X_{Std}(cm)$	171	112
$Y_{Std}(cm)$	78	23

As it can be seen in Table 1, the PVH model runs faster, but it gives larger standard deviation along the axis compared to PVA. This should be expected as in the case of PVA there are 6 states compared to PVH with only 4 state variables.

ISBN 978-3-8007-3273-9

5 Experimental Results

The purpose of this experimental part was to validate the presented classifier for the human leg form detection problem. The measured data contained random error points due to the imperfections of the laser scanner. These points have to be eliminated, as well as the effect of the varying leg pair forms.

For the detection of different leg pair forms, one option was to consider several leg pair configurations in the *GMM* training phase. This option has the folloing main advantage: the leg pairs with different distances between the legs (i.e. during walking) can be classified correctly. On the other hand, the main disadvantage of this type of classifier is that the recognized object space grows, thus the number of false positives also grows.

This is especially true in an office environment, where objects like chairs with thicker legs or trash bins can be easily classified as human leg pairs if the classifier is trained to recognize a wide range of leg pair configurations [20].

The experiment setup contained a P3 skid-steered mobile robot equipped with a LMS200 laser range finder and a camera connected to a PIV laptop.

Before performing the main experiments in the indoor some preliminary test were done. The position estimation of the robot was observed to be more reliable at low turning speed as the accelerations were not explicitly introduced in the process model. The camera was also tested in different light/background conditions and the best hit rate was achieved with the imagines containing high contrast parts especially about the people (like striped colorful clothes).

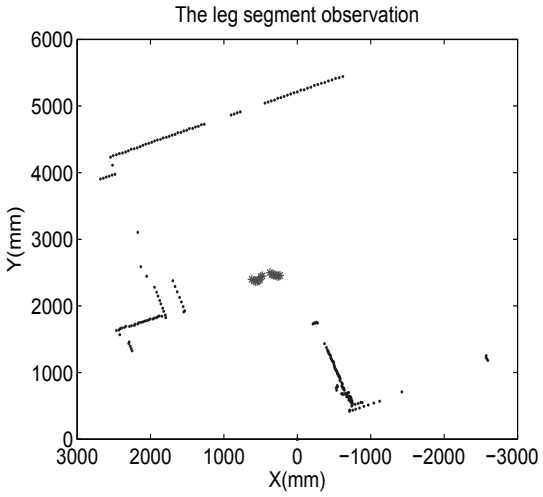

Figure 5: The detected laser leg forms

The lidar measurement were negatively influenced only by the surrounding objects in the laboratory which have very similar forms to the human legs. The a-prior map of the indoor environment was measured in a separate measurement process thereby decoupling the mapping and localization problem.

The observation of a leg form from the laser scanner is shown onn Figure 5. As it can be seen, the leg pair can be distinguished in case that is not too close to other objects.

A typical output of the Bayesian classifier is shown in Figure 6. The classification is done first by taking the points likely to be part of the *GMM* leg models considered earlier. Further on, a pdf value density quantile for each point is generated in order to evaluate the distribution density quantiles.

In order to decide whether a point is an outlier from the Bayesian model, a threshold pdf value is considered (around 0.9) and the decision is taken upon the previously computed density quantile for each point.

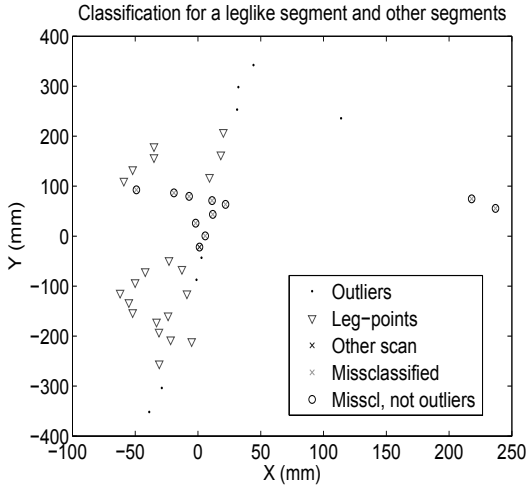

Figure 6: The classified leg points

A test bench on three different data sets was performed in an uncluttered environment and cluttered environement to test the human detection rate(DR) of the different classifiers. The results of the benchmark are presented in Table 2.

Table 2: Comparison of the different classifiers

Exp.	Avg. DR (%)	Avg FP (%)
Uncluttered	72	2
Cluttered	53	4

As expected best detection rate can be achieved in case of the uncluttered environment. Another remark regarding the detection rate is, although this is not so high as the ones reported in the literature review, the false positive rate is very low, i.e. the tuning of the classifier was performed in such a way to reduce as much as possible the false positive rate.

ISBN 978-3-8007-3273-9

6 Conclusions and Future Work

6.1 Conclusions

A multi-sensor object classifier was presented in this paper including a tracking of the detected objects. A cooperative technique was adopted to combine the information from the lidar. Details regarding the classification algorithms were presented for the both approaches. The computed coordinates of the moving human objects were used to track them in an indoor environment.

6.2 Future Work

We intended to test and validate against other classification algorithms the presented ones, and to extend the object detection of object tracking to multi person tracking and occlusion handling. Further on, the condition of the prior existent map can be relaxed, and the localization or the people detection can be performed simultaneously with the mapping task. This is referred as the *Simultaneous Localization and Mapping* and *Simultaneous Localization and People Tracking* [21] . Both of the problems require the idea of running the localization task in parallel with either the mapping or the detection task.
As an alternative information source, the stereo-vision camera system is proposed to be introduced.

References

[1] A. Mendes, L. C. Bento, and U. Nunes, "Multi-target detection and tracking with laser range finder," (Parma,), IEEE Intelligent Vehicles Symposium, 2004.

[2] J. Borenstein, H. R. Everett, L. Feng, and D. Wehe, "Mobile robot positioning sensors and techniques," *J. Robot. Syst*, vol. 14, pp. 231–249, Nov. 1997.

[3] A. J. Lipton, H. Fugiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," pp. 129–136, IEEE Image Understanding, 1998.

[4] C. Premebida and U. Nunes, "A multi-target tracking and gmm-classifier for intelligent vehicles," (Toronto,), 9th International IEEE Conference on Intelligent Transportation Systems, 2006.

[5] O. Arras and . Mozos, "People detection and tracking workshop," IEEE International Conference on Robotics and Automation, 2009.

[6] H. Durrant-Whyte, *Multi Sensor Data Fusion*. Australian Center for Field Robotics, 2006.

[7] Y. Bar-Shalom and X. Li, *Estimation and Tracking-Principles, Techniques and Software*. Norwood, MA: Artech House, 1993.

[8] J. Guivant, E. Nebot, and H. F. Durrant-Whyte, "Simultaneous localization and map building using natural features in outdoor environments," pp. 581–588, Intelligent Autonomous Systems VI, 2000.

[9] J. Vandorpe, H. V. Brussel, and H. Xu, "Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2d range finder," (Minneapolis,), p. 901ï¿½908, IEEE International Conference on Robotics and Automation, 1996.

[10] N. Bellotto and H. Hu, "Multisensor integration for human-robot interaction," *IEEE Journal of Intelligent Cybernetic Systems*, vol. 1, 2005.

[11] J. Neira, J. D. Tard, J. Horn, and G. Schmidt, "Fusing range and intensity images for mobile robot localization," pp. 76–84, IEEE Trans. Robotics and Automation, 1999.

[12] D. Streller and K. Dietmayer, "Object tracking and classification using a multiple hypothesis approach," (Parma,), IEEE Intelligent Vehicles Symposium, 2004.

[13] O. Mozos, K. Arras, and W. Burgard, "Using boosted features for detection of people in 2d range scans," in *In Proc. of the IEEE Intl. Conf. on Robotics and Automation*, 2007.

[14] M. Bertozzi and A. Broggi, "Pedestrian localization and tracking system with kalman filtering," IEEE Intelligent Vehicles Symposium 2004, 2004.

[15] G. A. Borges and M. J. Aldon, "Line extraction in 2d range images for mobile robotics," *Journal of Intelligent & Robotic Systems*, vol. 40, p. 267ï¿½297, 2004.

[16] P. Paalanen, J. K. Kamarainen, J. Ilonen, and H. Klviinen, "Feature representation and discrimination based on gaussian mixture model probability densities - practices and algorithms," tech. rep., 2005.

[17] M. Figueiredo and A. Jain, "Unsupervised learning on finite mixture models," in *IEEE transactions of pattern analysis and machine intelligence*, 2002.

[18] L. Tamas, G. Lazea, A. Majdik, M. Popa, and I. Szoke, "Laser and vision based people detection for mobile robots," in *In Proceedings of RAAD09*, (Brasov, Romania), 2009.

[19] P. Maybeck, *Stochastic Models, Estimaton and Control*, vol. 1. Academic Press, 1979.

[20] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia, "People tracking and following with mobile robot using an omnidirectional camera and a laser," in *In Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 557–562, 2006.

[21] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.

ISBN 978-3-8007-3273-9