Lidar and Vision Based People Detection and Tracking

L. Tamas, **M. Popa**, **Gh. Lazea**, **I. Szoke**, **A. Majdik**

Robotics Research Group, Technical University of Cluj-Napoca, Romania
http://rrg.utcluj.ro

Abstract: This paper presents a multi-sensor architecture to detect moving persons based on the information aquired from a lidar and vision systems. The detection of the objects are performed relative to the estimated robot position. For the lidar the Gaussian Mixture Model (GMM) classifier and for the vision the AdaBoost classifier is used from which the outputs are combined with the Bayesian rule. The estimated person positions are tracked via the Extended Kalman filter. The main aim of the paper was to reduce the false positives in the detection process with the use of a sequentially combined classifiers.

Keywords: Detection and tracking systems, visual pattern recognition, laser range finders.

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 persons in order to avoid collision in an indoor environment. Examples of such moving object may be people or other mobile robots (Mendes et al., 2004).

In the proposed architecture two different object classifiers, based on different sensors, are combined in order to gain a higher level of inference and meaningful information to achieve robustness in the classification process. A cooperative strategy was adopted in order to establish the coordinate correspondence between the lidar and the monocular vision camera to reduce the field in which the object detection is performed. The robot position estimation was based on dead-reckoning sensors and the Kalman filtering algorithm (Borenstein et al., 1997).

The architecture of the system is composed from four subsystems: the robot position estimator, the lidar based classifier, the vision based classifier and the coordinate transformation system for 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. Further on, coordinates are transformed and used in the in the field of view of the camera to refine the people detection (Neira et al., 1999). Also the lidar can be used to measure the distance of the object relative to the robot with a good accuracy (Lipton et al., 1998).

The people detection with lidar is based on the Gaussian Mixture Model representations of the leg forms proposed by (Premebida and Nunes, 2006), and the detected person positions are used to restrict the field of view (FOV) of the camera and to extract the depth information of the detected person. The *AdaBoost* classifier (Monteiro et al., 2006) is based on *Haar-Like* features. The two classifiers are used sequentially in order to reduce the false positives in the detection. The general overview of the system is presented on the Figure 1.

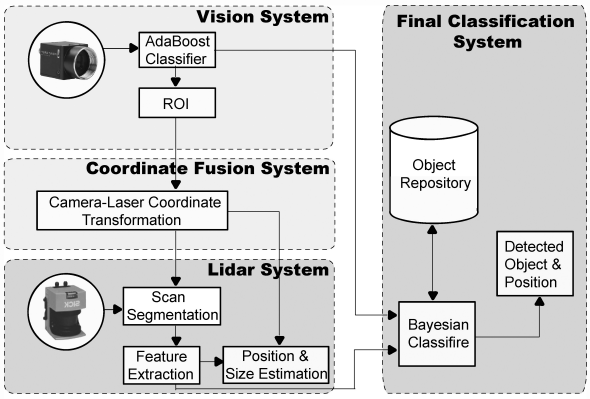

Fig. 1. Classification system overview

2. RELATED WORK

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, see (Arras and Mozos, 2009).

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 is well reported subject (Guivant et al., 2000) (Vandorpe et al., 1996).

Several research work have been performed using laser-scanners in object classification and moving object tracking including but not limited to localization and navigation proposed in (Bellotto and Hu, 2005) application or guarding systems shown in (Neira et al., 1999). For the object classification voting schemes, multi-hypotheses tracking presented in (Streller and Dietmayer, 2004) or even boosting approaches (Mozos et al., 2007) 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 (Bertozzi and Broggi, 2004). In certain light/ambient conditions the perfor-

mance of the vision system can degrade and the range information even if it is available is not appropriate. In such cases the use of additional sensors like the laser range finders is highly motivated.

Papageorgiu (Oren et al., 1997) introduced a trainable object detection architecture based on wavelet templates that defines the shape of an object by considering a subset of the wavelet coefficients of the image. Based on this type of architecture in (Viola and Jones, 2001) is proposed an algorithm for detection of people, which can recognise a person through a possible constellation of his body parts. These body parts are described by Haar-like features (Freund and Schapire, 1995), so images are processed at high rates. Moreover, an object classifier with good performances is obtained using Adaboost algorithm (Monteiro et al., 2006). Some extensions of the algorithm were developed in (Zivkovic and Krose, 2007) and in (Schulz et al., 2007). Zivkovic uses the algorithm in order to learn not just people shapes, but appearance of upper and lower human body. Schulz adds a new final stage to the Viola's algorithm which consists of a neural network which is trained to identify the false alarms. The objec t detection with the monocular vision presented in this paper is based on the approach proposed by Viola.

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 prefiltering 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 (Borges and Aldon, 2004). The current strategy is the one based on the simple assumption of Eucledean distances between segments adopted from (Mozos et al., 2007).

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 laser beam relative to the robot and ρ_j is the distance from the reflecting 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, than a new segment is considered. Even if there are more sophisticated segmentation algorithms e.g. like the one presented in (Premebida and Nunes, 2006), in the current problem setup we found appropriate this 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 Cartesian coordinates $\mathbf{f} = (x, y)$ which can be converted to polar coordinates with $x = \rho\cos(\theta)$ and $y = \rho\sin(\theta)$.

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 gating threshold γ_k then a break point is observed.

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 (Mozos et al., 2007). The basic set of feature which was used in the experiments contained the following e_1, e_2 and e_3 entries:

(1) e_1: object centroid;
(2) e_2: normalized Eucledian distances given by:

$$f2 = \sqrt{\Delta X^2 + \Delta Y^2} \tag{1}$$

(3) e_3: the standard deviation of the point from the r centroid computed for n points:

$$f3 = \sqrt{\frac{1}{n-1}\sum ||r_n - \bar{x}||} \tag{2}$$

These components are essential to the classifier.

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:

$$p(x|\Theta) = \sum_{m=1}^{M} \alpha_m p(x|\theta_m) \tag{3}$$

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:

$$p(\Omega|q_i, \Theta^i) = \sum_{m=1}^{M} \alpha_m^i p(\Omega|\theta_m^i) \tag{4}$$

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

$$p(\Omega|q_i, \Theta^i) = \frac{1}{\sqrt{(2\pi)^2|\Sigma_m^i|}} \exp[-\frac{1}{2}(\Omega - \mu_m^i)^T (\Sigma_m^i)^{-1} (\Omega - \mu_m^i)] \tag{5}$$

The Gaussian mixture parameters for each object of interest was determined using the expectation-maximization (EM) algorithm. 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:

$$p(\Omega^N|q_i, \Theta^i) = \prod_{j=1}^{M} p(\Omega_j|q_i, \Theta_m^i) \tag{6}$$

3.4 Bayesian Classifier

After a Gaussian mixture pdf for classified object is available a Ω_k feature-vector is considered in order to classify which category (q_i) fits the current observation. Based on a Bayesian decision framework the log-likelihood of the fitness is computed.

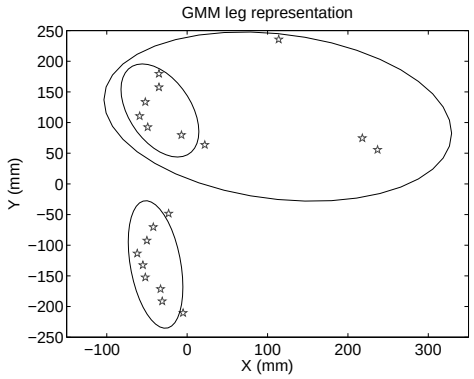

Fig. 2. Possible leg forms in the laser scan

Computing the log-likelihood has the advantage of reduced computational effort by avoiding the computation of the exponential in the pdf (5) and by turning the product (6) 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 equipprobable, 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:

$$log(P(\Theta^i|\Omega)) \approx log(p(\Omega|\Theta^i)) + log(P(\Theta^i)) \tag{7}$$

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 pdf is computed at each scan, which will become in the next scan the last estimated posterior. Therefore the prior probability is updated dynamically as:

$$P(\Theta^i_k) = P(\Theta^i|\Omega_{k-1}) \tag{8}$$

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:

$$S_j \in q_i \quad if \quad log(P(\Theta^i|\Omega_k)) = \max(log(P(\Theta^u|\Omega_k))) \tag{9}$$

where u spans from 1 to the number of classes.

3.5 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 on several extensions were developed (Borenstein et al., 1997). The Kalman filter is an optimal recursive data processing algorithm for linear systems corrupted by noise.

The Extended Kalman filter (EKF) (Maybeck, 1979) 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 $\mathbf{x}^+_{k-1}$. 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.

3.6 The Motion Models for Humans

Two motion models were adopted for people tracking. For both models the measured state variables were the positions in the Cartesian coordinates (x_k, y_k).

Position-velocity-heading (PVH) Model – used to estimate the human motion with constant velocity model. In our experiments this model was extended with the orientation ϕ_k and velocity v_k according to (Bellotto and Hu, 2005) as follows:

$$\begin{cases} x_k = x_{k-1} + \delta_k v_{k-1} cos\phi_{k-1} \\ y_k = y_{k-1} + \delta_k v_{k-1} sin\phi_{k-1} \\ \phi_k = \phi_{k-1} + n^{\phi}_{k-1} \\ v_k = v_{k-1} + n^{v}_{k-1} \end{cases} \tag{10}$$

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

Position-velocity-acceleration (PVA) Model – or referred as the $\alpha-\beta-\gamma$ filter (Bar-Shalom and Li, 1993) is the model of a Newtonian system represented in 2D coordinate system. Along a single axes the motion equations are given as follows:

$$x_k = \begin{bmatrix} 1 & \delta_k & \delta_k^2/2 \\ 0 & 1 & \delta_k \\ 0 & 0 & 1 \end{bmatrix} x_{k-1} + \begin{bmatrix} \delta_k^2/2 \\ \delta_k \\ 1 \end{bmatrix} n_{k-1} \tag{11}$$

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 (Durrant-Whyte, 2006).

In both cases the legs position are measured as bearing-range information with relative to the robot's position (x^R_k, y^R_k, ϕ^R_k) as follows:

$$\begin{cases} b_k = tan^{-1}\left(\dfrac{y_k - y_l}{x_k - x_l}\right) - \phi^R_k + n^b_k \\ r_k = \sqrt{(x_k - x_l)^2 + (y_k - y_l)^2} + n^r_k \end{cases} \tag{12}$$

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

3.7 Motion Model Comparison for Tracking

To compare the two models, the EKF was used to estimate the position of the detected person with respect to the robot position. The same dataset was tested checking the computational effort and the standard deviation of the innovation along the x and y axes during the experiments. 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.

4. VISION BASED CLASSIFIER

The vision-based system is used in this paper to estimate the positions of the people. A special attention is paid to the reduction of computing time with respect for a good detection rate, therefore the detection algorithm can work in a dynamic environment in real time. The detection procedure uses the gradient based segmentation algorithm for the reduction of the interest regions in images and also the AdaBoost classification algorithm (Freund and Schapire, 1995).

4.1 Gradient Based Segmentation

The gradient segmentation used in this paper is based on clustering horizontal gradients (Jaap, 2006). The structures of people appearances in images denote that these gradients can be useful for the reduction of searching area for people detection. When applying the horizontal gradients on an image, vertical edges are highlighted while the horizontal ones are masked, so that vertical structures can be clustered. The regions of interest from the images are considered those regions which accommodate the vertical structures with respect to a ratio between the region's height and width.

4.2 The Object Classifier

In (Viola and Jones, 2004) is proposed a multilevel classification procedure, using Haar features and Adaboost algorithm. Haar-like features are preferred to other features based on pixel values because they are computed in constant time, speeding up the detection process. Moreover, they codify domain datasets, which are difficult to extract from a finite input with other methods, so the classification becomes easier. The Adaboost algorithm is used to train a classifier $f(P)$ in order to split a dataset D into homogeneous partitions. The classifier is computed as a linear combination of some less discriminating classifiers, named weak classifiers $h^t(P)$, weighted according to their classification error. The algorithm has T iterations. A new weak classifier is determined at each iteration, being described by a feature's type and size, parity and a threshold value which splits the data set into partitions as homogeneous as possible (Wiegersma, 2006).

$$h^t_{a,p,\mu}(x) = \begin{cases} +1, if\ p \cdot a(x) < p \cdot \mu \\ 0,\ otherwise \end{cases} \tag{13}$$

where: p takes values in the $\{-1,1\}$ set and $\mu \in X, X = \{a(x)|x \in D\}$

The accuracy of the classifier proposed by Viola depends on the training datasets completeness and on the features used for classification. It is not a trivial task to endow the classifier with complete datasets for training. The dataset of positive images (those which contain people) has to include examples with front view and side view of different human body shapes and different body positions. A complete dataset of negative examples is much harder, almost impossible, to provide because the high number of object structures which could be found in the images and which could have some people-like features. Hence, there will be always a number of false positive detections. The attempt to reduce this number modifying the classical algorithm parameters may lead to the rise of the number of missclassified examples. The method for people detection presented in (Popa et al., 2009) should reduce the number of false positive detections, by implementing a new weight assignment mechanism for the features of the classifier, with a slight modification of the algorithm training phase.

The Adaboost algorithm is provided with one positive and multiple negative image sets for training, and for each image, it computes a weak classifier. After the training phase, for each feature type there are multiple weak classifiers with different threshold values, chosen in such a way that they constitute some feature poles which minimize the classification errors. Having this information, a finite interval defined by two thresholds can be found for any new feature value computed in the classification phase. The weight assignment mechanism used in this work computes the weights α_i of the weak classifier according to the classification errors at the interval limits and with interval's length, see Figure 3.

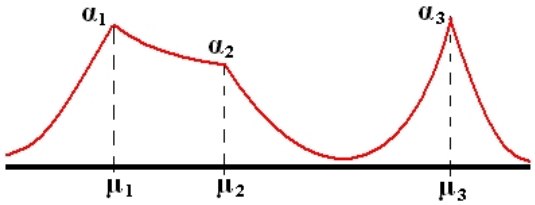

Fig. 3. Weight assignment function

This mechanism makes the classification less dependent of the negative image training dataset and raises the importance of providing a complete positive one, which reduces the dependence of the classifier performance by the environment. By training the classifier with outdoor image datasets and testing it indoor, it can be seen that the false positive alarm rate is maintained low. To compensate the variations in illumination which are more frequent indoor, a parameter c_i is computed according to the image histogram. Its role is to relax the classification process in case of low level of illumination or low contrast. So, the classification decision is taken in this work with the following formula:

$$f(P) = \begin{cases} 1, \text{ if } \sum_{a\in A} \alpha_a h_a(x) \geq \frac{1}{2} \sum_{a\in A} \alpha_a c_i \\ 0, \text{otherwise} \end{cases} \tag{14}$$

where the classification weights α_i are computed with the mechanism described above and c_i is a image histogram parameter, and h_a represents the week classifier function.

5. COMBINED CLASSIFIERS

The basic idea of the combined object classifiers is presented in Figure 1. The information from the laser and camera is used as input to two different kinds of classifiers in order to enhance the robustness of the moving person detection. First of all the information is structured in segments and then it is introduced to the classifier (Wang et al., 2003).

5.1 Calibration

The lidar and camera calibration is important to perform in order to transform the point coordinates from the camera $\{C\}$ frame to the laser reference $\{L\}$. For avoiding overwhelming computations, the camera and the laser are aligned "ideally"

in the same plane parallel to the robot displacement plane. 2 In order to obtain the transformation matrix between the two coordinate systems, a special measurement set was considered, and the geometric transformation matrix was obtained by least squares error minimization technique.

The camera intrinsic and extrinsic parameters were approximated based on the online camera calibration toolbox (Caltech, 2005), and in this way it can be easily achieved the transformation between the image frame and the camera $\{C\}$ frame.

5.2 Classifier Combination

After the region of interest(ROI) is detected by the laser ranger based on the Bayesina classifier applied to the trained GMM forms, this information is transmitted to the camera module. This module transforms the coordinates from the laser device to the camera frame (if they at least partially overlap) and searches only in the common ROI for possible targets.

Even though with the lidar it is possible to obtain only a single horizontal information about the object, this is with rather superior accuracy compared to the position information from the camera. With the combination of the two sensors and by assuming that the vehicle moves on a flat surface the bottom limit of the detected object can be easily found out, while the height it can estimated from the size on the image.

As it was observed in the experimental part, the camera captures the entire human body from distances larger than $3m$. On the other hand, the laser detects with a better detection rate the leg pairs which are closer than $5m$ as the number of points representing the leg pair is greater. Based on this observation, for distances less than $3m$ only the laser leg classifier is used, while for distances larger than $3m$ the combined laser&camera classifier is employed with the ROI transmitted from laser to camera. Also at this phase, the tracking information from the laser is used to give ROI hints to the camera in case that the laser temporarily lost a target. For distances above $6m$ only the camera classifier is used.

5.3 Experimental Results

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). For regions with similar background to the clothes of the people, the only information which could be used was the one from the laser ranger classifier.

The lidar measurement were negatively influenced only by the surrounding objects in the laboratory which have very similar forms to the human legs.

The people detection was performed using data sequences in different positions and lighting conditions. The detection was done by applying the gradient segmentation to reduce the interest area from images and then labelling regions from this area, which contain people, with a strong classifier obtained from AdaBoost algorithm.

The strong classifier is trained on a 6000 images set, 3000 of them as positive examples (containing people), with the feature resolution choosen to be 30x30 pixels. A typical measurement output from the camera is shown in Figure 4. The high contrast background enables the detection of the full body in the test case.

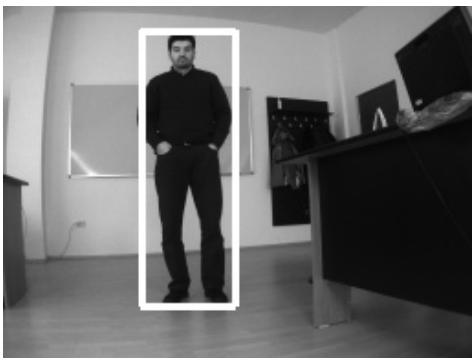

Fig. 4. The selected region classified as people

The same observation from the laser scanner is shown in Figure 5. Although there are similar regions to human legs on the laser scan which may give false alarms, the narrowed ROI by the camera enables to classify the correct segment as a human leg.

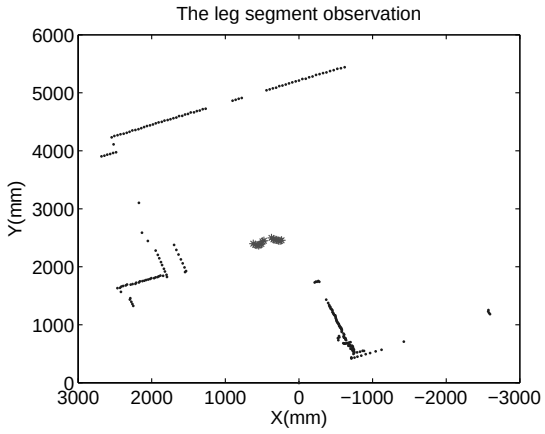

Fig. 5. The selected segment classified as people

In most of the vision based classifier achieved better results than the GMM classifier in terms of the hit rate, but this situation changes drastically in case of reduced camera visibility cases (extreme light conditions). Also for side view of the persons, the vision system performs better, as for the lidar a single leg looks similar to many common objects (like refuse bean) in an indoor environment.

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

As expected best detection rate can be achieved with the combined classifier. 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 both classifiers were performed in such a way to reduce as much as

Table 2. Comparison of the different classifiers

Classifier	Avg Detection Rate (%)	Avg False Positive (%)
Laser	52	2
Camera	55	0.8
Laser&Camera	68	0.5

possible the false positive rate. The runtime for the combined classifier is also better than the runtime of the camera, as the search for the ROI is reduced by the information from the laser classifier.

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 and the visual systems. 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 mapping problem based on landmark detection and classification is intended to be included. As an alternative information source, the stereo-vision camera system is proposed to be introduced.

7. ACKNOWLEDGEMENTS

This project was partially supported by the POSTDRU (Project of Doctoral Studies for Development in Advanced Technologies) at the Technical University of Cluj-Napoca and partially by CNCSIS TD-213 project.

REFERENCES

Arras, O. and Mozos, . (2009). People detection and tracking workshop. IEEE International Conference on Robotics and Automation.

Bar-Shalom, Y. and Li, X. (1993). *Estimation and Tracking-Principles, Techniques and Software*. Artech House, Norwood, MA.

Bellotto, N. and Hu, H. (2005). Multisensor integration for human-robot interaction. *IEEE Journal of Intelligent Cybernetic Systems*, 1.

Bertozzi, M. and Broggi, A. (2004). Pedestrian localization and tracking system with kalman filtering. IEEE Intelligent Vehicles Symposium 2004.

Borenstein, J., Everett, H.R., Feng, L., and Wehe, D. (1997). Mobile robot positioning sensors and techniques. *J. Robot. Syst*, 14, 231–249.

Borges, G.A. and Aldon, M.J. (2004). Line extraction in 2d range images for mobile robotics. *Journal of Intelligent & Robotic Systems*, 40, 267297.

Caltech (2005). Camera calibration toolbox for matlab. URL `www.vision.caltech.edu/bouguetj/calib`.

Durrant-Whyte, H. (2006). *Multi Sensor Data Fusion*. Australian Center for Field Robotics.

Freund, Y. and Schapire, R.E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 23–37.

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

Jaap, W.A. (2006). *Real-time pedestrian detection in FIR and grayscale images*, 39–45. Ruhr University, Bochum.

Lipton, A.J., Fugiyoshi, H., and Patil, R.S. (1998). Moving target classification and tracking from real-time video. 129–136. IEEE Image Understanding.

Maybeck, P. (1979). *Stochastic Models, Estimaton and Control*, volume 1. Academic Press.

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

Monteiro, G., Peixoto, P., and Nunes, U. (2006). Vision-based pedestrian detection using haar-like features. Proc. 6th National Festival of Robotics, Scientific Meeting (ROBOTICA), Guimaraes,.

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

Neira, J., Tard, J.D., Horn, J., and Schmidt, G. (1999). Fusing range and intensity images for mobile robot localization. 76–84. IEEE Trans. Robotics and Automation.

Oren, M., Papageorgiou, C., and Sinha, P. (1997). Pedestrian detection using wavelet templates. Proc. IEEE CVPR 1997.

Popa, M., Lazea, G., Majdik, A., Tamas, L., and Szoke, I. (2009). An effective method for people detection in grayscale image sequences. In *IEEE International Conference on Intelligent Computer Communication and Processing*, 181 – 185.

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

Schulz, W., Enzweiler, M., and Ehlgen, T. (2007). Pedestrian recognition from a moving catadioptric camera. *Lecture Notes in Computer Science*, 456–465.

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

Vandorpe, J., Brussel, H.V., and Xu, H. (1996). Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2d range finder. 901908. IEEE International Conference on Robotics and Automation, Minneapolis,.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *IEEE Conf. Computer Vision and Pattern Recognition*, 511–518.

Viola, P. and Jones, M. (2004). Robust real time face detection. 137–154. International Journal of Computer Vision.

Wang, C., Thorpe, C., and Suppe, A. (2003). Ladar-based detection and tracking of moving objects from a ground vehicle at high speeds. In *In Proceedings of the IEEE Intelligent Vehicles Symposium*.

Wiegersma, A.J. (2006). *Real-time pedestrian detection in FIR and grayscale images*. Master's thesis, Bochum University.

Zivkovic, Z. and Krose, B. (2007). Part based people detection using 2d range data and images. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 214–219.