Torso Detection and Tracking using a 2D Laser Range Finder

Zulkarnain Zainudin
Faculty of Electronics and Computer Engineering
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, MALAYSIA
Email: zul@utem.edu.my

Sarath Kodagoda and Gamini Dissanayake
ARC Centre of Excellence for Autonomous Systems
Faculty of Engineering and Information Technology
University of Technology, Sydney
PO Box 123 Broadway, Ultimo, NSW 2007, Australia
Email: {sakoda,gdissa}@eng.uts.edu.au

Abstract

Detecting and tracking people in populated environments has various applications including, robotics, healthcare, automotive, security and defence. In this paper, we present an algorithm for people detection and tracking based on a two dimensional laser rage finder (LRF). The LRF was mounted on a mobile robotic platform to scan a torso section of a person. The tracker is designed to discard spurious targets based on the log likelihood ratio and can effectively handle short term occlusions. Long term occlusions are considered as new tracks. Performance of the algorithm is analysed based on experiments, which shows appealing results.

1 Introduction

Human robot interaction has become an emerging area of research in the past years. Robots are emerging as helpers, carers, security officers and entertainers in day today life. People detection and tracking can play an important role in such situations. Various researchers have developed algorithms utilizing diverse sensor modalities, however with different levels of success. Cameras and laser range finders (LRFs) are commonly exploited in those applications. In this paper, we propose to use a single LRF to detect and track people using a mobile platform.

There are several techniques that have been proposed in the literature for the detection of people with laser range finders, such as motion-based, feature-based and heuristic approach [Zhang and Kodagoda, 2005] [Zivkovic and Krose, 2007a] [Mozos et al., 2009c] [Arras et al., 2007b]. In general, detection based on motion can have limitations due to stationary people. Feature-based people detection reported in the literature use single-layered, double-layered or triple-layered approaches, which may detect legs, upper body and head [Carballo, 2009b] [Zivkovic and Krose, 2007a] [Mozos et al., 2009c] [Arras et al., 2007b]. Leg detection is an appealing approach, however it may lead to complex algorithms due to leg movements and affected by the attire. Triple layered approaches have height limitations and need three LRFs. It is our belief that, a single LRF could still be exploited as a cost effective solution to detect and track people. In this work, we propose to detect and track the torso of people. It has a general cross section of an ellipse and do not change much with the attire. It also does not have complex dynamics as legs and can be classified into standard torso categories [Harrison and Robinette, 2002] [Pheasant, 2003] which means a template matching (as in computer vision) approach can be used.

In general, people detection problem is handled by utilizing classification algorithms. Thus, the selection of a classifier is an important part in the detection process. Commonly used classifiers in people detection based on LRF are AdaBoost and Support Vector Machines. Arras et al [Arras, 2007b] used the AdaBoost algorithm with 14 features that were based on the characteristics of a laser segment. Further, Spinello et al [Spinello and Siegwart, 2008b] applied Support Vector Machines classifier on 2D laser data and vision data. In this paper, we compared the performance of different number of classifiers for a given set of features in order to choose the best classifier.

This paper is organized as follows. In section 2, we present the details of feature extraction and classifier selection with regard to people detection. Section 3 presents people tracking algorithm. Section 4 presents the experimental results using a mobile robotics platform. Section 5 concludes the paper with direction of future research.

2 People Detection

Our strategy to people detection based on laser data comprises of extracting significant features followed by a classification process.

2.1 Features

The first processing step of laser range/bearing data based people detection is data segmentation. This is based on detecting range discontinuities in the laser scan.

The laser range finder provides range and bearing, $\{r_i, \theta_i\}$ to objects in its field of view, where, suffix i refers to a specific range/bearing data with $i = 1......n$. By using a model based technique, which is realized using the Extended Kalman Filter (EKF) [Kodagoda et al., 2002], it is possible to partition the data into segments, $S = \{s_1, s_2, ..., s_M\}$ as shown in Figure 1. M is the number of segment in a particular laser range/bearing data. In the Figure 1, symbol 'o' refers to discontinuity points, which define start and end points of segments.

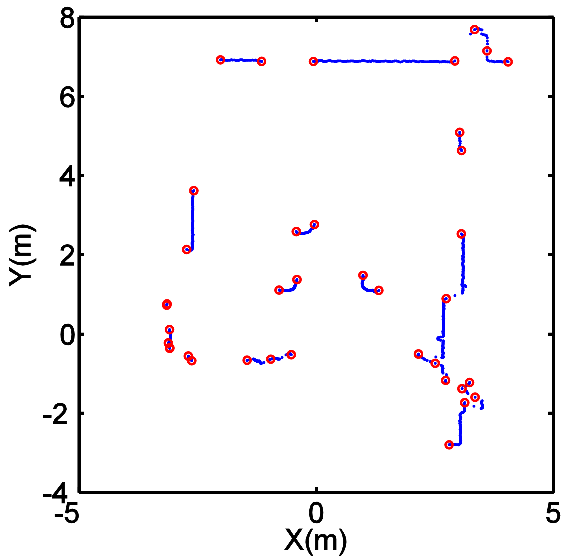

Figure 1: Segmented laser data

Once the data segmentation is performed, next step is to extract meaningful features to be used in people detection.

Feature 1: Length of a segment, which is given by, $\sqrt{(x_n - x_1)^2 + (y_n - y_1)^2}$. (1)

Feature 2: Ratio of major to minor axis of the ellipse. The laser range finder is mounted in such a way that it scans torso part of an average person. Cross section of torso of a human can generally be approximated by an ellipse. Therefore, an ellipse fitting algorithm [Fitzgibbon et al., 1999a] is implemented on segmented laser range data.

The Cartesian coordinates of each element in an i^{th} segmented laser data, $s_i = \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{in}\}$ can be transformed into a matrix, $\mathbf{D}_i = [x_{i1}, x_{i2}, ..., x_{in}]^T$. Then the solution for fitting of ellipses is a general conic equation [Fitzgibbon et al., 1999a]:

$$F(\mathbf{a}, \mathbf{x}) = \mathbf{a}.\mathbf{x} = ax^2 + bxy + cy^2 + dx + cy + f = 0 \quad (2)$$

$$\mathbf{Sa} = \lambda \mathbf{Ca} \quad (3)$$

$$\mathbf{a}^T \mathbf{Ca} = 1 \quad (4)$$

where, $\mathbf{a}$ =[a b c d e f]^T, $\mathbf{x}$ =[x^2 xy y^2 x y 1]^T, $\mathbf{S} = \mathbf{D}_i \mathbf{D}_i^T$, and C is $\begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. Here, the maximum and minimum values of λ, $\lambda_{\max}$ and $\lambda_{\min}$ define the length of major and minor axes respectively. The ellipses fitted for the segmented data in Figure 1 are shown in Figure 2. The features that we consider include length of major and minor axes, and the ratio of major and minor axes.

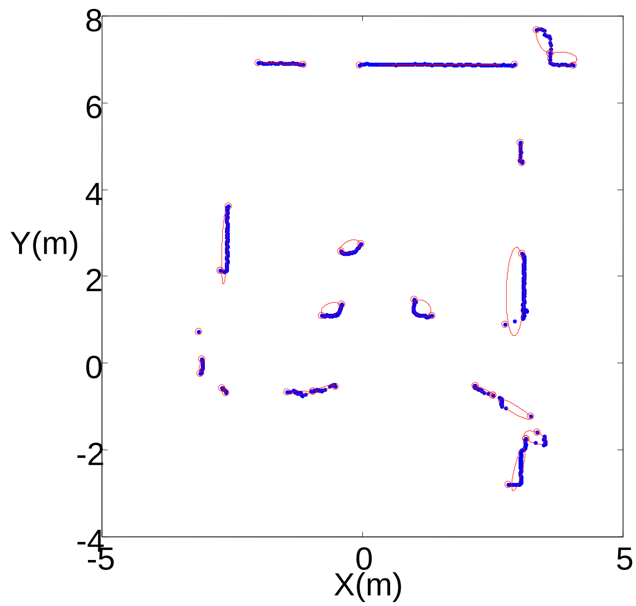

Figure 2: Segmented data with ellipse fitting

Feature 3: Mean curvature characteristic of segment, S_i. Given three sequential Cartesian coordinates, $\mathbf{x_1}$, $\mathbf{x_C}$ and $\mathbf{x_n}$, let A denote the area of the triangle enclosed by $\mathbf{x_1 x_C x_n}$ and d_1, d_C, d_n denotes the distance of three legs of the triangle. Then, an approximation of discrete curvature of the boundary at x_C is given by [Arras et al., 2007b],

$$k = \frac{4A}{d_1 d_C d_n} \quad (5)$$

Feature 4: The fourth feature is the ratio of the distance between laser source to the centre of segmentation over number of points, which is given by

$$\frac{\sqrt{x_C{}^2 + y_C{}^2}}{n} \quad (6)$$

where, x_c, y_c and n are centre point of x and y; and number of points respectively.

2.2 Classification

Once the features have been extracted, a classification routine was implemented. In order to compare the performance of different classifiers, we have used Weka [Weka 3.6-2, 2010], a popular open source machine learning software. The data was captured using a Hokuyo laser range finder while people are freely wandering in an office like environment. The laser range finder was mounted to scan torso of a person. Totally there were 500 number of laser range scans. Out of them, 200 scans were used for the training and another 200 scans were used for testing.

Table 1 shows the results of different classifiers with few people wandering in an environment. When there is only one person in the vicinity of the laser range finder, all classifiers performed well. However, with more people the classifiers tend to have poorer performances.

This could be mainly due to differences in sizes, costumes of people and artifacts due to occlusions. Out of the given classifiers, it could be seen that the RBFSVM (Radial Basis Function Support Vector Machines) performed better and therefore, it was chosen as the classifier to be used in this study.

TABLE 1. COMPARISON WITH OTHER CLASSIFIERS

No. of people	***Classifier***	***Training Data Accuracy (%)***	***Testing Data Accuracy (%)***
One	RBFSVM	98.5965	96.8643
	AdaBoostM1	96.5789	92.5884
	Simple Logistic	97.9825	90.9350
	MultiBoostAB	94.1228	62.5998
	BayesNet	98.5965	95.6770
	Complement Bayes Net	94.1228	63.2269
	Naïve Bayes	98.5088	98.0616
	Naïve Bayes Simple	98.5088	98.0616
	Naïve Bayes Updateable	98.5088	98.0616
More than one	RBFSVM	96.7196	95.2197
	AdaBoostM1	96.8254	64.3022
	Simple Logistic	96.9312	61.3724
	MultiBoostAB	93.6508	69.5451
	BayesNet	97.4603	64.8419
	Complement Bayes Net	87.9365	57.1318
	Naïve Bayes	96.5079	75.6361
	Naïve Bayes Simple	96.4021	70.7016
	Naïve Bayes Updateable	96.5079	68.6047

Given a training data set $T = \{(\mathbf{F}_i, l_i) \mid l_i \in (-1,1)\}$, where $i = 1, 2, ..., n$ and SVM requires the following optimization [Hsu et al., 2009e]

$$\frac{1}{2}(\mathbf{w}^T \mathbf{w}) + C\sum_{i-1}^{n} \xi_i \quad (7)$$

subject to $l_i(\mathbf{w}^T \phi(\mathrm{F}_i) + b) \geq 1 - \xi_i$ where $\xi_i \geq 0$. **F** and l are the features and the label of data set. Training vectors $\mathbf{F}_i$ are mapped into a higher dimensional space into by function ϕ. C is the penalty parameter of the error term. For radial basis function SVM, the kernel function is

$$K(\mathbf{F}_i, \mathbf{F}_j) = \exp(-\gamma \| \mathbf{F}_i - \mathbf{F}_j \|^2), \gamma > 0, \quad (8)$$

where γ is kernel parameter.

3 People Tracking

Once people were detected based on the laser data, it was temporally tracked based on an Interactive Multiple Model (IMM) tracker [Kodagoda et al., 2007c]. A constant velocity and constant turn rate models have been used to model the human motion. Due to the large scatter present in the environment due to various furniture, glass walls and various metallic parts, there were obvious false detections. The tracking problem was complex and nontrivial to handle due to disappearing, reappearing and maneuvering of target in clutter. This problem was handled by an IMMPDAF filter with track confirmation and deletion.

Using the Markov relationship, the probability of existence of a true person $P_T(k+1|k)$ before the receiving data in scan $k+1$ is given by [Blackman and Popoli, 1999b],

$$P_T(k+1|k) = P_{22}P_T(k|k) + P_{12}[1 - P_T(k|k)] \quad (9)$$

where P_{22} is the transition probability from an observable to observable state, and P_{12} is the transition probability from an unobservable to observable state. Then, probability update of person existence is

$$P_T(k+1|k+1) = \frac{1-\delta_{k+1}}{1-\delta_{k+1}P_T(k+1|k)} P_T(k+1|k) \quad (10)$$

where

$$\delta_{k+1} = \begin{cases} P_D P_G & N_{k+1} = 0 \\ P_D P_G \left[1 - \bar{V} \sum_{i=1}^{N_{k+1}} \frac{1}{P_G (2\pi)^{M/2} \sqrt{|S(k+1)|}} e^{-d_i^2/2} \right], & \text{others} \end{cases}$$

and $\bar{V} = V_{G_{k+1}} / (N_{k+1} - P_D P_G P_T(k+1|k))$, P_D is the probability of detection, P_G is the gate probability, V_G is the gate volume, N_{k+1} is the number of measurements inside the validation gate, S is the innovation covariance, and d_i^2 is the normalized innovation squared of the ith measurement.

The log-likelihood ratio (LLR) is defined as,

$$LLR_{k+1} = \ln\left(\frac{P_T}{1-P_T}\right). \quad (11)$$

Once the LLR is obtained, confirmation and termination of track thresholds are determined as

$$\mathrm{LLR}_{k+1} \geq \ln\left(\frac{1-\beta_T}{\alpha_T}\right), \text{ declare track confirmation}$$

$$\ln\left(\frac{\beta_T}{1-\alpha_T}\right) < \mathrm{LLR}_{k+1} < \ln\left(\frac{1-\beta_T}{\alpha_T}\right), \text{ continue test}$$

$$\mathrm{LLR}_{k+1} \leq \ln\left(\frac{\beta_T}{1-\alpha_T}\right), \text{ delete track}$$

where α_T and β_T are the probability of false-track confirmation and the probability of true-track termination, respectively.

4 Experimental Results

Figure 3: Robot used for the experiments

4.1 Experimental Setup

The robot used in our experiments is a Segway equipped with sensors and computers, which is shown in Figure 3. It has an onboard computer, AMD Athlon II X2 255/ Dual Core/ 3.1 GHz with 4GB DDR3 running on Linux Ubuntu 9.10 operating system. Robot system uses HOKUYO UTM-30LX laser range finder that has 30 meters of detection range, 0.25° angular resolution, angular field of view of 270° and 25millisecond sampling period. A robot based on Segway is used to monitor the environment, while in motion. The experiments were carried out in a common area in our university.

4.2 Classifier Section

The LRF data consists of various furniture, structures, people and their poses. As given in Section 2.1, the LRF data was first segmented filtered and ellipses were fitted for feature extraction. Ellipses fitted on torso of a person with different poses are shown in Figure 4. Although there are slight changes due to the position of hands (this could also happen due to different types of clothing), the ellipses were fitted reasonably well.

The features described in section 2 were estimated and used in the Weka [Weka 3.6-2, 2010] with several numbers of classifiers as shown in Table 1. The data was analyzed categorizing the scenarios into three cases based on the number of people present in the environment (and hence possible occlusions).

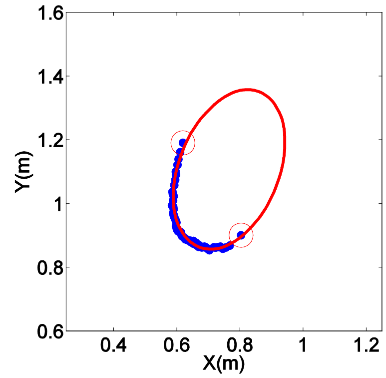

Figure 4(a): Ellipse fitting: a person with hands up.

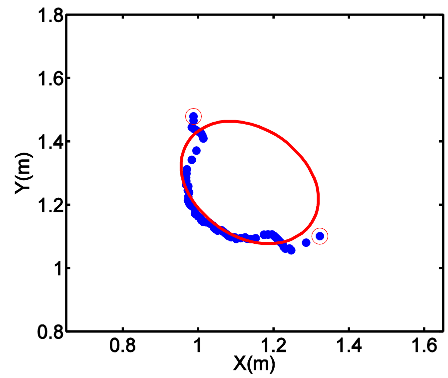

Figure 4(b): Ellipse fitting: a person with hands down.

In general, it could be seen that the classifier performance degrades with increased number of people due to rise in occlusions. Although in simple scenarios, classifier such as BayesNet performs well, it is susceptible to errors with increased complexity. On the other hand classifiers such as, radial basis function SVM leads to better classification accuracies in both simple and complex scenarios.

4.3 People Detection

Experiments have been conducted to assess the performance of the people detection algorithm. In order to have a better understanding of the errors and their causes, we have chosen specific scenarios such as, people looking at the sensor with hands up or down and people looking sideways with hands up or down. In all cases the laser range finder was mounted at torso height. Table 2 summarizes the accuracies. Detection accuracy with side facing people is generally higher than that of straight facing people because of the complexity of the curve on the straight facing scans.

On the other hand, not surprisingly people with hands up pose have higher accuracies than that of hands down (normal pose). It could also be noted that the false positives are always smaller than false negatives. Therefore, the algorithm provides more candidates, which can be further filtered to improve the detection accuracy. Total computing time is not more than 0.06s in all scenarios.

Although, in general the ellipse fitting algorithm worked well, it had some problems with segmented data relevant to occluded scenarios. This is explained in Figure 5. In the figure, ellipse fitting was done reasonably well from (a) to (d). The problem started at (e), where one ellipse has undergone a significant change to its size and shape. In (f), one ellipse has completely disappeared due to an occlusion and started to re-appear in (g).

TABLE 2. CONFUSION MATRIX FOR STRAIGHT AND SIDE FACING

STRAIGHT FACING (HANDS UP)		
True Label	***Person***	***Others***
Person	86.41%	0.77%
Others	13.59%	99.23%
Total Accuracy		96.54 %
Computing Time		0.03 s
STRAIGHT FACING (HANDS DOWN)		
True Label	***Person***	***Others***
Person	83.33%	3.00%
Others	16.67%	97.00%
Total Accuracy		94.77 %
Computing Time		0.06s
SIDE FACING (HANDS UP)		
True Label	***Person***	***Others***
Person	91.97%	2.14%
Others	8.03%	97.86%
Total Accuracy		97.22 %
Computing Time		0.06 s
SIDE FACING (HANDS DOWN)		
True Label	***Person***	***Others***
Person	86.75%	3.56%
Others	13.25%	96.44%
Total Accuracy		92.84 %
Computing Time		0.02 s

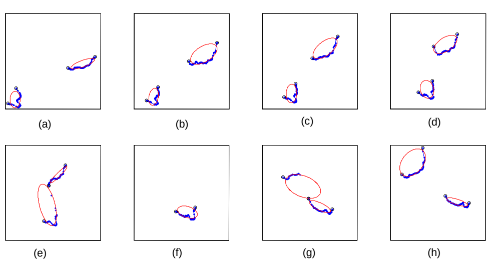

Figure 5: Occlusion of two people.

4.4 People Tracking

People detection part can be integrated in the IMM based temporal tracking algorithm discussed in the Section 3. Figure 6 shows tracking of two people (T1 and T2) using a stationary observer. The motions of T1 and T2 caused an occlusion, where T1 disappears from observations. However, due to the predictions of the IMM tracker, the lost track could be re-associated for further tracking.

Figure 7 shows results of tracking of two people with a dynamic observer. The motions of T1 and T2 again causing the occlusion where T2 disappears from the observation and the scenario is quite similar to the stationary observer. If T2 has the possibility of being terminated if disappears for a long time and it will re-appear as a new target. The process of determining the tracks is based on log-likelihood ratio (LLR) which is shown in Figure 7. A new track is confirmed, if the LLR is higher than an upper threshold and a track is deleted, if it falls down bellow a lower threshold (as defined in section 3).

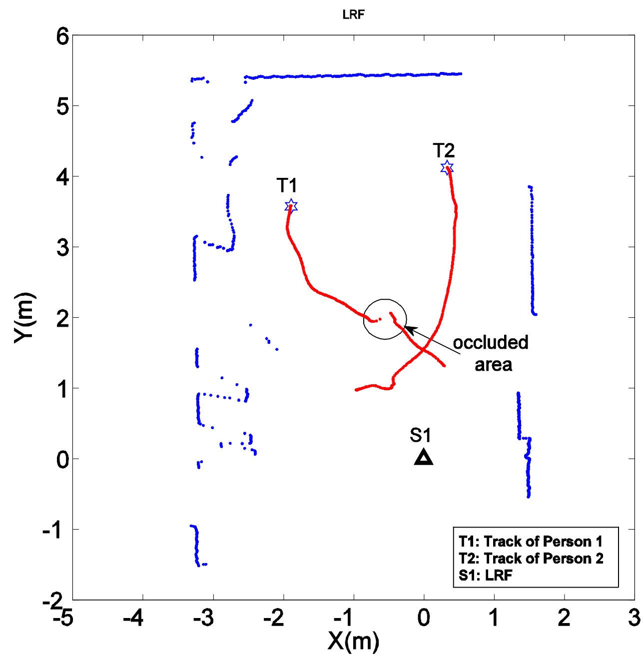

Figure 6: People tracking results: with a stationary observer. T1 and T2 denote the tracked two people.

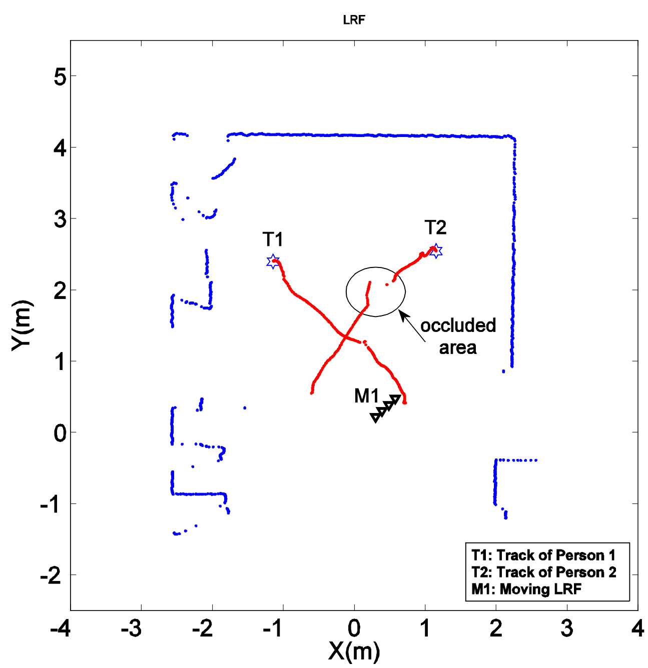

Figure 7: People tracking results: with a dynamic observer. T1 and T2 denote the tracked two people. Triangles denote the observer position.

The tracks that are occluded for a long time have the possibility of being deleted and re-appear as new tracks. It is reasonable as far as the application does not require the identification and tracking of a particular individual.

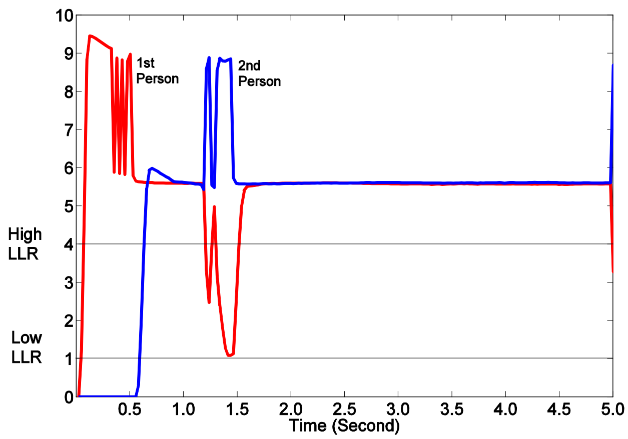

Figure 8: The log-likelihood ratio (LLR) of two targets tracking

5 Conclusion and Further Work

The problem of multiple people detection using features extracted on torso of a person using a single layer LRF was presented in this paper. It is shown that the detection of a person is possible by supervised learning classifier such as SVM. Experimental results were presented with a stationary and dynamic LRF which was observing a common area. Objects were tracked by an IMM tracker. Various tracks and spurious data were effectively handled by a log likelihood ratio based decision making. It showed appealing results. We are currently working on ways to address the tracking problem in more complicated environments.

6 Acknowledgement

This work is supported by the ARC Centre of Excellence programme and funded by the Australian Research Council (ARC), the New South Wales State Government and Ministry of Higher Education (MOHE), Malaysia.

References

[Carballo, 2009b] A. Carballo, A. Ohya, and S. Yuta. Multiple People Detection from a Mobile Robot using Double Layered Laser Range Finders (Published Conference Proceeding Style). In *Proc. IEEE ICRA 2009 Workshop on People Detection and Tracking*, Kobe, 2009.

[Zhang and Kodagoda, 2005] Z. Zhang, K.R.S. Kodagoda. Multi-sensor Approach for People Detection. In *Proc. IEEE 2005 Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing*, pages 355-360, 2005.

[Zivkovic and Krose, 2007a] Z. Zivkovic and B. Krose. Part based people detection using 2D range data and images. In *Proc. IEEE 2007 Int. Conf. on Intelligent Robots and System*, pages 214-219, San Diego, CA, USA, 2007.

[Mozos et al., 2009c] O.M. Mozos, R. Kurazume and T. Hasegawa. Multi-Layer People using 2D Range Data. In *Proc. IEEE ICRA 2009 Workshop on People Detection and Tracking*, Kobe, 2009.

[Arras et al., 2007b] K.O. Arras, O.M. Mozos and W. Burgard. Using Boosted Features for the Detection of People in 2D Range Data. *IEEE 2007 Int. Conf. on Robotics and Automation*, pages 3402-3407, Rome, Italy, 2007.

[Spinello and Siegwart, 2008b] L. Spinello, R. Siegwart. Human Detection using Multimodal and Multidimensional Features. *IEEE 2008 Int. Conf. on Robotics and Automation*, pages 3264-3269, Pasadena, USA, 2008.

[Hata and Wolf, 2009d] A.Y. Hata, D.F. Wolf. *Terrain Mapping and Classification using Support Vector Machines*. IEEE Latin American Robotics Symposium, pages 1-6, Valparaiso, Chile, 2009.

[Kodagoda et al., 2002] K.R.S. Kodagoda, W.S. Wijesoma and A.P. Balasuriya. Road Curb and Intersection Detection using a 2D LMS. In *Proc. 2002 IEEE RSJ Int. Conf. on Intelligent Robots and Systems*, pages 19-24, Lausanne, Switzerland, 2002.

[Fitzgibbon et al., 1999a] A. Fitzgibbon, M. Pilu, and R.B. Fisher. Direct Least Square Fitting of Ellipses. *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 476-480, vol. 21, May 1999.

[Hsu et al., 2009e] C.W. Hsu, C.C. Chang, C.J. Line. A Practical Guide to Support Vector Classification. [Online]. 2009, May 19Available: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[Kodagoda et al., 2007c] K.R.S. Kodagoda, S.S. Ge, W.S. Wijesoma, A.P. Balasuriya. IMMPDAF Approach for road-Boundary Tracking. *IEEE Trans. Vehicular Technology*, pages 478-486, vol. 56, 2008.

[Weka 3.6-2, 2010] Weka 3.6-2. Data Mining Software in Java. The University of Waikato, 2010.

[Harrison and Robinette, 2002] C.S. Harrison, K.M. Robinette. CAESAR: Summary Statistics for the Adult Population (Ages 18-65) of United States of America. United States Air Force Research Laboratory, Ohio, USA, June 2002.

[Pheasant, 2003] Stephen Pheasant. Bodyspace: Anthropometry, Ergonomics and the Design of Works. Taylor & Francis, 2003.

[Blackman and Popoli, 1999b] Samuel Blackman, Robert Popoli, Design and Analysis of Modern Tracking Systems, 1999.