Leg detection and tracking for a mobile robot and based on a laser device, supervised learning and particle filtering

Eugenio Aguirre *, Miguel Garcia-Silvente, and Javier Plata

Department of Computer Science and A.I., CITIC-UGR.
E.T.S. Ingenierías en Informática y en Telecomunicaciones
University of Granada. 18071 - Granada, Spain
{eaguirre,M.Garcia-Silvente,jplata}@decsai.ugr.es
http://decsai.ugr.es

Abstract. People detection and tracking is an essential skill to obtain social and interactive robots. Computer vision has been widely used to solve this task but images are affected by noise and illumination changes. Laser range finder is robust against illumination changes so that it can bring useful information to carry out the detection and tracking. In fact, multisensor approaches are showing the best results. In this work, we present a new method to detect and track people using a laser range finder. Patterns of leg are learnt from 2d laser data using supervised learning. Unlike others leg detection approaches, people can be still or moving at the surroundings of the robot. The method of leg detection is used as observation model in a particle filter to track the motion of a person. Experiments in a real indoor environment have been carried out to validate the proposal.

Keywords: Leg detection and tracking, laser range finder, supervised learning, particle filter, mobile robots

1 Introduction

A social robot or a service robot should have located the people at its surroundings and to be aware of their presence. Therefore, people detection and tracking is an essential skill in order to achieve a social and interactive robot. This is a very challenging task as people is freely moving by the environment. Many researches have focused their efforts in this field in last years using different kind of sensors. Vision systems have been widely used to deal with this problem [10]. In our research group this problem has been attacked through the use of stereo vision [14], [15], [16]. Stereo vision systems and other kinds of devices as the Kinect [13], allow the acquisition of various kinds of information such as color and depth. However the depth information is not always reliable and

* This work have been partially supported by Andalusian Regional Government project P09-TIC-04813 and the Spanish Government project TIN2012-38969.

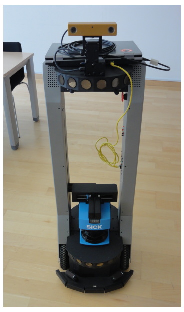

Fig. 1. Peoplebot robot equipped with laser range finder SICK LMS200.

the images are affected by noise and illumination changes. Other kind of sensor used in people detection and tracking is the laser range finder (LRF) [2], [8], [17]. Compared with vision, the use of a LRF is advantageous since it is robust against illumination changes in the environment. Nevertheless, the LRF has also some limitations because we obtain only distance information in one 2d-plane. The best solutions are obtained fusing the information from different sensor systems. There are several authors who combine LRF and vision [1], [9]. Thus, the trend in this field is to use multisensor approaches [4] because integrating several sources of information the obtained system is more robust and efficient.

In this paper we focus our attention in the development of a system to detect and track people using a LRF. In future works we will integrate the information from the LRF system with previous stereo vision systems for people detection and tracking [16]. In the current work, first a new method to detect legs using LRF and supervised learning is proposed. This method analyses certain geometric features present in the laser data and then a classifier is trained using Support Vector Machines (SVM) [6]. The classifier will be able to identify patterns of legs in the surroundings of the robot both for static or moving people. After that, a particle filter is designed to track the motion of people. The classifier is used as an observation model to detect the legs of the people. In the current work the robot is immobile and observing the motion of people at its surroundings.

The rest of this paper is organized as follows. Section 2 describes the hardware of our system and presents the human leg detection algorithm. Section 3 shows the tracking algorithm based on particle filter. In Section 4, experimental verifications are described. Finally some concluding remarks and future works are commented in Section 5.

2 Leg Detection

Our hardware system comprises a PeopleBot mobile robot equipped with a laser range finder SICK LMS200. It scans 180° with a 1° resolution at 75 Hz. The maximum range of distance in the current operation mode is 8 meters. It is mounted at a height of 30 cm from the ground. Fig. 1 shows this robot.

In the detection stage, our aim is to detect legs when people are moving and static. It is a challenging objective because legs patterns are different in both situations. In our approach we separate the laser measurements into clusters and geometrical properties of these clusters are studied. Such as their width or depth, which let us to distinguish between legs and other objects such as waste baskets or table legs. In order to detect if a cluster is a leg or not a SVM classifier is used. It has been trained with a large number of positive and negative samples. Bellow we will explain deeply all the steps performed.

Our system is designed to be used in Human Robot Interaction (HRI) applications, so we are going to consider that the person who desire to interact with the robot is near. Thus we have defined a maximum interaction distance and all range measurements further than this distance are ignored. In the experimentation, we have estimated this distance as 3m.

The robot mostly operates in an indoor polygonal environment. Therefore, the following step is to remove the information associated to long lines, that correspond to walls. For this operation a Line Hough Transform [11] has been implemented. The Hough transform is a feature extraction technique used in computer vision. The purpose of the technique is to find lines by a voting procedure.

The next step is to cluster the remaining points by the application of the clustering threshold distance $D_{threshold}$ and the clustering minimum number of points C_{min}. We consider as a cluster a group with at least C_{min} points where the distance between two consecutive points is lower than $D_{threshold}$. As geometrical properties we have computed the same three attributes used in [8]. They are: the contour of the neighbor points in a cluster from P_1 to P_m, the width defined as the distance from P_1 to P_m and the depth as the maximum distance between a point P_i and the line $\overline{P_1 P_m}$. Figure 2 shows these attributes.

A large data set has been acquired for training the SVM. This dataset contains positives samples, which was registered with people walking and standing in front of the robot and negative samples corresponding to table legs, waste baskets, boxes and an extinguisher. Notice that some of this objects could have geometrical properties similar to an human leg, and they will increase the pre-

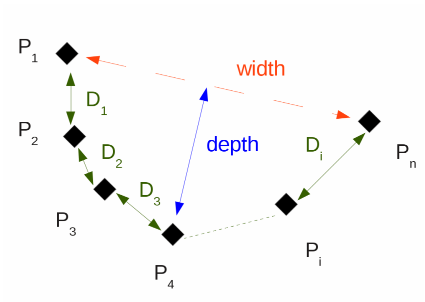

Fig. 2. Illustration of the defined geometrical parameters in a cluster.

cision and robustness of the classifier. It is important to remark that positive samples were acquired while people were moving, so our system will be able to detect a person who is walking, not only people standing. We use a balanced dataset containing 1156 both positive and negative samples.

In order to apply the SVM classifier, the LibSVM was used [6]. Different kernels have been considered and the best precision is obtained with radial basis function (RBF). A wide grid-search using cross-validation has been performed in order to find the optimal value for these parameters [7] obtaining a precision about 91.5%.

Table 1 corresponds to contingency table for a 10-fold cross validation using the parameter values described before. Results are very promising since the rates of true positives and true negatives are very high.

Table 1. Contingency table

		Observation class	
		Positive	Negative
Predicted class	Positive	93.67 %	12.35 %
	Negative	6.33 %	87.65 %

3 Tracking

Particle filter is well known for its many applications in tracking. Target tracking problem is expressed by recursive Bayesian estimation. Essentially, two steps are given in each iteration: prediction and estimation. Both steps take into account the information of an observation model. Equations of particle filter are well known [3].

In this paper, a particle filter is used to estimate the position of the detected people. In the current work only one person will be tracked. The state in time t is defined as a pair of coordinates (x, y). The coordinates correspond to the middle point of the straight line which links the two points representing both detected legs of the person. That is, the people position S_t is represented by a pair $S_t = [x_t, y_t]$. The prediction is carried out by the model of the state transition. We do not consider in the model the motion of the people. Thus, a model of the movement of the target is intentionally not defined. The state transition is defined as $S_t = S_{t-1} + R_{t-1}$ where S_{t-1} is the previous state vector and R_{t-1} is the process noise. The noise is modeled using a Gaussian with average μ_R and standard deviation σ_R. Experimental data have been taken into account to establish the values of μ_R and σ_R in order to model the conditions of the real world.

Condensation algorithm [12] is used to generate a weighted set of particles $(s_i(t), \Pi_i(t))$ where $s_i(t)$ represents a possible position of the people, and $\Pi_i(t)$ is a factor called *importance weight* representing an estimation of the observation.

At the beginning the algorithm is provided by an initial sample $(s_i(0), \Pi_i(0))$ of N equally weighted particles. At each iteration, the algorithm uses the sample set $(s_i(t-1), \Pi_i(t-1))$ to create a new one. A resample mechanism is used to solve the divergence problem by eliminating particles having low importance weights. Afterwards, the model of state transition is used to predict the motion of the person obtaining the prediction of the state S'_t. The weight $\Pi_i(t)$ of each particle is computed based on the new observation $Z(t)$. Then the weights are normalized so that $\sum_{i=1}^{N} \Pi_i(t) = 1$.

The observation model is needed to carry out the update. As model of observation, the leg detection method explained in Section 2 is used. Thus, each laser reading set is analyzed and the positions of possible legs are obtained. To calculate the probability for each particle the nearest possible legs are taken into consideration. The probability of each particle is calculated as:

$$\Pi_i(t) = \frac{1}{\sigma_z \sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{dist-\mu_z}{\sigma_z}\right)^2} \tag{1}$$

being *dist* the euclidean distance between $s_i(t)$ and the position of the nearest legs. The parameters μ_z and σ_z correspond to the average and standard deviation of a normal distribution, respectively. $\mu_z = 0$ and σ_z has been experimentally tuned. The final person position corresponds to the mean of the state $\mathcal{E}[S(t)]$, calculated as $\mathcal{E}[S(t)] = \sum_{i=1}^{N} \Pi_i(t) s_i(t)$.

4 Experimentation

Different experiments have been carried out in order to validate the proposal. The goal is to test the accuracy of the tracking algorithm in a real environment having several people. A set of seven paths on the floor were defined with different trajectories and including easy paths like a straight line as well as more complicated paths such as circles or curves in which motion direction continuously changes. Fig. 3 shows the acquisition scenario and the seven paths defined.

Fig. 3. Acquisition scenario for tracking system test.

Then, different samples of people following these trajectories have been acquired. Four different people have participated in the experiments. It is important to acquire data from different people since each one has his or her own gait. Every person has walked on each trajectory six times with different speed walking. In total 168 trajectories have been analyzed.

In order to evaluate the tracking performance of each trajectory, we have calculated the distance of the predicted position by the tracker to the real trajectory, which has been manually computed. The MOPT measurement [5] has been used for this purpose. The algorithm has been evaluated with different numbers of particles: 50, 100, 150, 200 and 250. Table 2 shows the results obtained for each trajectory in mm, where column T indicates the trajectory. For each trajectory and each number of particles, mean error (MOTP) and standard deviation in mm are indicated. As it can be observed the error obtained is very low and it decreases when the number of particles increases. Fig. 4 shows as the tracking error decreases when number of particles increases. For a number of particles higher than 200 the error decreases in a lower rate. Because of that, we have used 200 particles. Using this number of particles, the average tracking error for all the trajectories is computed. This value is 24.11 mm and its standard deviation is 4.16 mm. The average run time obtained is very low, 2.4 ms, so we can claim that the system can operate in real time.

Table 2. Results in mm for particle filter with 50, 100, 150, 200 and 250 particles

	Number of particles									
	50		100		150		200		250	
T	MOTP	Std	MOTP	Std	MOTP	Std	MOTP	Std	MOTP	Std
1	30.78	17.31	28.92	16.70	27.41	16.20	26.71	16.62	26.66	16.47
2	34.39	42.08	31.30	40.28	30.47	40.23	30.44	40.53	30.24	41.30
3	24.12	17.24	21.87	15.71	20.53	15.33	20.31	15.92	19.21	15.46
4	28.77	18.68	27.13	17.84	25.44	17.67	25.05	18.22	24.65	18.17
5	28.40	17.25	26.23	16.85	25.42	17.16	25.49	18.10	24.64	17.37
6	26.78	18.33	23.77	17.72	23.74	17.41	22.90	17.18	22.83	17.52
7	21.47	16.90	19.39	17.20	18.58	16.93	17.90	17.07	17.19	16.76

5 Conclusions and Future Work

The contributions of this paper can be summarized in two. The first one is the method to detect the human legs. It does not assume any specific shape of legs. Also, it is not needed that the human is still and looking towards the robot. We use supervised learning to learn the different patterns of legs which can appear in the robot surroundings corresponding to moving or still people. These patterns are learned from a large number of sample data using SVM. The second contribution is the robust tracking of leg positions using the method for leg detection as an observation model within a particle filter. Experiments

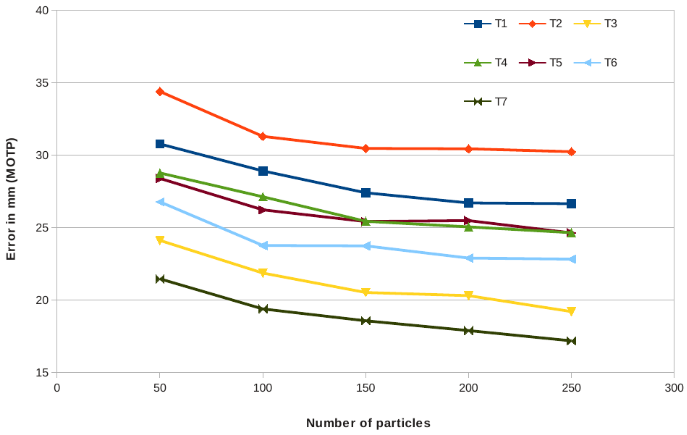

Fig. 4. MOTP error in mm vs number of particles for each trajectory.

on leg detection and on people tracking have been carried out to validate the proposal. The experimental results show that human legs are properly detected in the most of the cases. Also, the experiments on the tracking show that our LRF based tracking system is working obtaining a low error tracking the human position.

As future work, the LRF based tracking system will be integrated with other previous stereo vision based systems in order to obtain a multisensor tracking system. The multisensor tracking system will be used to follow a person in an indoor environment.

References

1. A. Ansuategui, A. Ibarguren, J.M. Martínez-Otzeta, C. Tubío, and E. Lazkano. Particle filtering for people following behavior using laser scans and stereo vision. *International Journal on Artificial Intelligence Tools*, 20(02):313–326, 2011.
2. K. O. Arras, S. Grzonka, M. Luber, and W. Burgard. Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities. In *In Proc. IEEE International Conference on Robotics and Automation (ICRA*, pages 1710–1715, 2008.
3. M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.
4. N. Bellotto and H. Hu. Multisensor-based human detection and tracking for mobile service robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(1):167–181, 2009.
5. K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: The clear mot metrics. *Eurasip Journal on Image and Video Processing*, 2008, 2008.

6. Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.
7. Chih-Chung Chang Chih-Wei Hsu and Chih-Jen Lin. *A Practical Guide to Support Vector Classification.* National Taiwan University, 2010.
8. W. Chung, H. Kim, Y. Yoo, C.-B. Moon, and J. Park. The detection and following of human legs through inductive approaches for a mobile robot with a single laser range finder. *IEEE Transactions on Industrial Electronics*, 59(8):3156–3166, 2012.
9. J. Fritsch, M. Kleinehagenbrock, S. Lang, T. Plötz, G. A. Fink, and G. Sagerer. Multi-modal anchoring for human-robot interaction. *Robotics and Autonomous Systems*, 43(2-3):133–147, 2003.
10. D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding: CVIU*, 73(1):82–98, 1999.
11. V. Hough and C. Paul. Method and means for recognizing complex patterns, 1962. U.S. Patent 3069654.
12. M. Isard and A. Blake. Condensation-conditional density propagation for visual trackings. *International Journal of Computer Vision*, 29:5–28, 1998.
13. Microsoft. Kinect official webpage, 2010.
14. R. Muñoz-Salinas, E. Aguirre, and M. García-Silvente. People detection and tracking using stereo vision and color. *Image and Vision Computing*, 25:995–1007, 2007.
15. R. Muñoz-Salinas, E. Aguirre, M. García-Silvente, and R. Paúl. A new person tracking method for human-robot interaction intended for mobile devices. *MICAI 2007: Advances in Artificial Intelligence Lecture Notes in Computer Science*, 4827:747–757, 2007.
16. R. Paúl, E. Aguirre, M. García-Silvente, and R. Muñoz-Salinas. A new fuzzy based algorithm for solving stereo vagueness in detecting and tracking people. *International Journal of Approximate Reasoning*, 53:693–708, 2012.
17. K. Schenk, M. Eisenbach, A. Kolarow, and H. Gross. Comparison of laser-based person tracking at feet and upper-body height. In Joscha Bach and Stefan Edelkamp, editors, *KI 2011: Advances in Artificial Intelligence*, volume 7006 of *Lecture Notes in Computer Science*, pages 277–288. Springer Berlin Heidelberg, 2011.