

Human Tracking and Following in Dynamic Environment for Service Robots

Adiwahono, A. H.^{*}, Saputra, V. B.[†], Ng, K. P.[¶], Gao, W.^{**}, Ren, Q.^{||}, Tan, B. H.[§], Chang, T.W.[‡]

Institute for Infocomm Research

1 Fusionopolis Way, Singapore 138632

Email: (*adiwahonoah †saputrabv ¶ngkp **gaowc ||renqy §tanbh ‡changtw) @i2r.a-star.edu.sg

Abstract—Robot human interactions have significant applications in dealing with day-to-day robot service. This paper presents a novel approach for human detection, tracking and following. A flexible and robust detection method using a 2D laser scanner is proposed firstly. The detection method can be adapted to various leg characteristics. The detection result is then fed to a Kalman Filter based tracking algorithm. Owing to the inherent ability of the embedded Kalman Filter, the tracking algorithm is able to track specific persons and their motion. Then, a controller is designed to follow target persons while avoiding obstacles. Finally, a service robot, namely ISERA, is adopted in the experiments for verifying the proposed approach. The experimental results show the robot can following a target person both indoor and outdoor dynamic environment.

I. INTRODUCTION

Mobile service robots, such as robotic butlers, robotic waiters, and robotic pets, are designed to assist people in their daily lives. In these robots, the capability to autonomously detect, track and follow a person has been regarded as an important functionality of the robotic systems. However, until now, it still remains a challenge that endowing this capability to these robots when they operate in human environment.

The approaches of autonomous person following without the person having to wear any active beacon can be classified into two categories: vision based and laser based. Due to the low cost, RGB-D cameras are widely adopted in the vision based approaches as the work presented in [1, 2]. However, the minimum distance requirement and the narrow field view of the RGB-D sensors limit the applications for following persons.

The laser based approaches have an advantage in terms of large field of view and robustness on any lighting condition. In [3] the map information is used to help tracking both moving and stationary person. However, the usage of this approach will be limited on environments that have been mapped. On the other hand, leg identification has been used for people tracking in an unknown area (e.g. in [4] - [5]). More recently, [5] introduces a framework for outdoor person tracking and following for a smart wheel chair robot by extending a trained legs detector from [6]. In [7, 8], robotic behaviors in relation to the speed and distance to the person being followed have been analyzed. However, the correlation from detecting legs into tracking a person, and the robustness when multiple people is around the person being followed has not been explored in these studies.

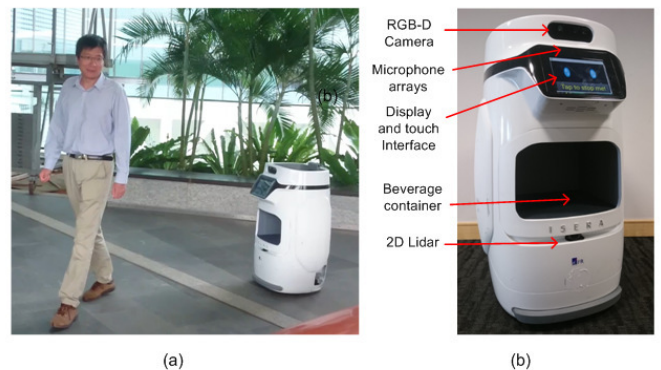


Fig. 1. (a) ISERA robot in action of tracking and following a target person (b) The sensor and interface of ISERA service robot

In this paper, we propose an approach to autonomously classify, track and follow persons in a dynamic environment for a service robot, namely ISERA (shown in Fig. 1). Using 2D Lidar and Odometry, our approach does not require prior map information, and benefit from wide field of view and consistent in any lighting condition. In our approach, a constraint based leg detection method is developed to regulate the strictness of the leg identifier as needed. Compared to previous works that solely track the legs, our proposed approach associate the data to form and track a virtual target person. In this way, brief false detections of a leg in crowded environment would not immediately made the person jump to another location, due to physical constraints in relation to the other leg. Two layers of Kalman filter is used to ensure the robustness of the person tracking despite close proximity to another person and temporary line of sight loss to the leg clusters occurred.

Another notable merit of the proposed approach is that a real time obstacle avoidance algorithm is embedded into the feed-back motion controller of the robot. Hence, experiments show ISERA can follow a target person while avoid obstacles autonomously in a cluttered environment.

The paper is organized as follows. Section II describes the mobile robot system. The people detection, tracking and following approach is detailed in Section III. The experiment and results are described in Section VI. Finally, a conclusion is given in Section V.

II. SYSTEM DESCRIPTION

The ISERA robot is designed as a human friendly service robot that could serve as a butler or courier in public area. ISERA is built on the top of a differential drive mobile platform with a maximum speed of $1m/s$. A PC is installed inside the robot for processing and a Hokuyo UTM-30-LX lidar is mounted about $30cm$ from the ground, which corresponds to sensor reading of legs below the knee, around the shank area of average adults. The lidar provide 2D laser scan, which is used to find groups of points that fits the criterions of clusters in the environment.

III. PERSON FOLLOWING

A. Leg Identification and Segmentation

A leg identification that build based on the principle of circle geometric fitting is proposed. To be classified as an arc or convex shape of a leg, the middle point of a cluster must be within a reasonable distance ratio (typically between 0.1-0.7) from the length between the cluster edges. The clusters that pass this filter will be processed further, otherwise it will be discarded. This filtering quickly discards any non-convex shape cluster with minimum computation cost.

Next, the center point of the cluster is approximated as the virtual point in the middle of the cluster leftmost and rightmost edges. For each cluster, the distance between the lidar and the center point of that cluster is denoted as z and will be used as a variable in the analysis. Another addition is the algorithm could also identify a joint dual leg, where the two legs of a person standing could be very close to each other and seen as one large cluster by the sensor. When a cluster has a local minimum that is farther then the center point of the cluster, it will be dissected at the minimum point into two parts. The part with more sampling points will be considered as a new single leg cluster and will be processed further.

To analyze the single leg clusters, three criterions are used as benchmark. First, the mean error of the inscribe angle variant (IAV) of the cluster [9]. The second is the error of the standard deviation of the cluster. The third is related to the error in the size of the cluster. The benchmark criterion is taken from numerous measurement of an adult humanoid model with typical cylindrical legs at varying distance z to the lidar. The data set collected are then interpolated into a function of distance z and kept as a benchmark values.

Instead of giving a crisp output, our algorithm calculates the confidence of a segmented cluster to be considered as a leg. Based on the normalized errors between the desired benchmark and the actual measurement, an equation to formulate the confidence of a cluster to be considered as a leg is derived as,

$$Leg_c = \max(1 - K_I I_N - K_S S_N - K_P P_N, 0). \quad (1)$$

with

$$\begin{aligned} I_N &= (|I_c - I_d(z)|)/I_d(z) \\ S_N &= (|S_c - S_d(z)|)/S_d(z) \\ P_N &= (|P_c - P_d(z)|)/P_d(z) \end{aligned}$$

Where Leg_c , which has a value between 0 and 1, is the confidence of cluster c to be considered a leg. I_c , S_c , and P_c are IAV mean, IAV standard deviation, and number of points that are calculated from sensor data, respectively. $I_d(z)$, $S_d(z)$, and $P_d(z)$ are the desired values of the IAV mean, IAV standard deviation, and number of points in the cluster for distance z , respectively. K_I , K_S , and K_P are weighting factor rule that tuned based on the preferred strictness of the leg segmentation.

Strict leg segmentation gives the least number of false positives, and could be used when the lidar identifies mostly bare legs. On the other hand, if the circumstance allows us to tolerate more false positives in the leg detection, loose segmentation could be used to detect people that are wearing loose pants. This strictness option would provide situational advantage compared to a rigid data trained learning based methods in other works. Finally, for clusters that have a confidence that exceed a cut off threshold, the center point and the confidence value are passed on to the leg and person tracking algorithm.

B. Leg and People Tracking

In order to follow certain person, multiple target tracking is developed to estimate positions of both detected legs and people. The estimation is implemented using Kalman Filters (KFs) [10] with two different motion model for legs and for people. In general, the state estimation of the legs and people can be modeled by a linear system, $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{w}_k$. The process noise, \mathbf{w}_k is modelled as Gaussian white noise with covariance matrix \mathbf{Q} . Fig. 2 shows the flow of the people following procedure. The input for the legs tracking are the leg detection results in each frame k . Furthermore, the input for the people tracking is the leg tracking states, treated as observations. In contrast with our approach, temporary people tracking are created in each update step in [5]. This approach could infer people directly from leg identification, but the movements of the people are not included in the consideration of the filter. In our method, the KF in the people tracking maintains the state of the tracked people at all times. Finally, the result of the people tracking is transmitted to the pursuit controller (Sec. III-C), so that the robot can follow the person.

1) *Legs Tracking*: The KF for the legs uses a constant velocity motion model with a pseudo velocity measurement during the KF update steps. At discrete time k , the KF maintains a set of leg tracks, ${}^L\mathbf{X}_k = \{{}^L\mathbf{x}_k^1, {}^L\mathbf{x}_k^2, \dots, {}^L\mathbf{x}_k^{L N_k}\}$, where $L N_k$ is the number of leg track at time k . Each leg track has a state estimate ${}^L\mathbf{x}_k^j = [x \ y \ \dot{x} \ \dot{y}]^T$ of a target position and

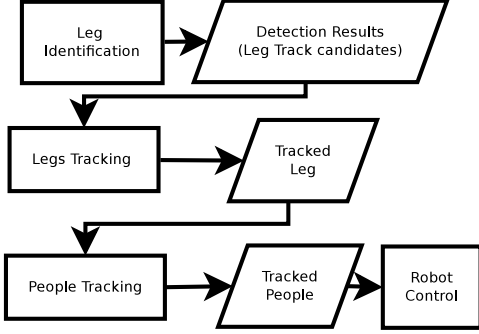


Fig. 2. Tracking process.

velocity in a 2D Cartesian coordinate. During the update step of the Kalman Filter, leg identification are processed using an observation model, ${}^L\mathbf{z}_k = {}^L\mathbf{H}^L\mathbf{x}_k + \mathbf{v}_k$. The observation includes position and velocity observations with white noise, \mathbf{v}_k governed by a covariance matrix, ${}^L\mathbf{R}$. The pseudo velocity measurement is determined from estimation of the difference from current state (after update step) and the previous state, normalized by the time step. The velocity measurement is used since we found that data from the laser perception is accurate and helps to improve legs tracking accuracy. The prediction and update cycle of the KF for every track can be summarized by the following equation. For clarity, the indices are omitted in this equation.

$$\tilde{\mathbf{x}} = \mathbf{A}\hat{\mathbf{x}} \quad (2a)$$

$$\tilde{\mathbf{P}} = \mathbf{A}\hat{\mathbf{P}}\mathbf{A}^T + \mathbf{Q} \quad (2b)$$

$$\mathbf{K} = \tilde{\mathbf{P}}\mathbf{H}^T(\mathbf{H}\tilde{\mathbf{P}}\mathbf{H} + \mathbf{R})^{-1} \quad (2c)$$

$$\hat{\mathbf{x}} = \tilde{\mathbf{x}} + \mathbf{K}(\mathbf{z} - \mathbf{H}\tilde{\mathbf{x}}) \quad (2d)$$

$$\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\tilde{\mathbf{P}}, \quad (2e)$$

where $\tilde{}$ and $\hat{}$ denote the predicted and filtered quantities, \mathbf{P} is the covariance matrix, \mathbf{H} is the observation matrix relating the state to the observation, \mathbf{I} is the identity matrix and \mathbf{K} is the Kalman gain. The update steps require associations between predicted state and current observation, which is solvable by employing Munkres algorithm [11]. The cost matrix for the assignment problem is computed using the Mahalanobis distance, so that the filter uncertainties are taken into account. This method is also known as Global Nearest Neighbor. However, a global gating threshold setting is employed so that association pairs that are too far away will be dropped. The data association has three interpretations: leg(s) that are uniquely associated with a track, the legs that cannot be associated with any of the existing track(s), and the track(s) that cannot be associated to any identified leg. Each track has also confidence level which is increased exponentially (with a constant α as a parameter) if there is evidence. If there is no evidence, the confidence will decrease and the track is removed if the confidence is below certain threshold. The outcomes of the interpretations are summarized in Table. I. The ${}^Lc_k^j$ and ${}^Ld_k^j$ are the confidence of the leg

TABLE I
POSSIBLE OUTCOMES FOR LEG TRACKING ASSOCIATION

Case	Outcome
Assigned track with leg observation	Update the corresponding KF state. Track confidence is updated, ${}^Lc_k^j = \alpha {}^Lc_{k-1}^j + (1 - \alpha) {}^Ld_k^j$
Unassociated track	Skip update, but propagate the track using the KF predict step. Track confidence is degraded, ${}^Lc_k^j = \alpha {}^Lc_{k-1}^j$
Unassociated leg observation	Initiate a new track with zero velocity and zero confidence.

track and the identified leg confidence associated with the leg track, with index j at time k , respectively.

2) *People Tracking*: In order to track people, additional KFs are also created similarly with the formulation elaborated in the leg tracking method. The differences lie on the motion model, observation model and how the data association is treated. People tracking maintains a set of people tracks, ${}^P\mathbf{X}_k = \{{}^P\mathbf{x}_k^1, {}^P\mathbf{x}_k^2, \dots, {}^P\mathbf{x}_k^{N_k}\}$, where PN_k is the number of people track at time k . Each leg track has a state estimate ${}^P\mathbf{x}_k^j = [x \ y \ \dot{x} \ \dot{y} \ \ddot{x} \ \ddot{y}]^T$ of a target position, velocity, and acceleration in a 2D Cartesian coordinate. The rationale behind the constant acceleration model for the people tracking is that humans have walking pattern that accelerates and decelerates periodically. The observation model takes the states of the legs tracks as measurements, ${}^P\mathbf{z}_k^m = {}^L\mathbf{x}_k^m$, where $m = 1 \dots {}^LN_k$. With this another layer of KF tracker, any spurious leg identification can be robustly filtered out.

The people tracking observation model, ${}^P\mathbf{z}_k = {}^P\mathbf{H}^P\mathbf{x} + \mathbf{v}_k$ takes only position as measurement variable. If both legs of a person can be detected, the position is computed as the center of the two legs, otherwise it is coincide as the position of the only leg detected. The people associations is solved using the Munkres algorithm, but the observation targets are firstly categorized into three groups: the legs that has unambiguous nearest leg, the legs that has uncertain leg pair (because there are a few candidates nearby), and the legs that are located far enough to other legs. These groups are subsequently referred as the two-legs group, the ambiguous one-legs group, and the certain one-legs group, respectively. The grouping is determined based on simple search with a maximum distance between two-legs pair threshold parameter. The data association is solved using similar cost matrix method, however we found that for people tracking, real distance metric is more robust. This is possibly caused by the fact that the uncertainties of the KFs are not as accurate compared to the leg tracking, because the observation are virtual measurement (the state of the leg tracks). There are four interpretations of the people tracking association: two-legs group that are uniquely associated with a people track, two-legs group that are unassociated, the people track(s) that are paired with only one in the one-legs group, the people track(s) that are unpaired with any observation. The summary

TABLE II
POSSIBLE OUTCOMES FOR PEOPLE TRACKING ASSOCIATION

Case	Outcome
Assigned people track with two-leg	Update the corresponding KF state. Track confidence is updated, ${}^P c_k^j = \beta {}^P c_{k-1}^j + (1 - \beta) {}^P d_k^j$
Unassociated people track	Skip update, but propagate the track using the KF predict step. Track confidence is degraded, ${}^P c_k^j = \beta {}^P c_{k-1}^j$
Associated people track with one-leg	If the one-leg belongs to certain one-legs group, update the corresponding KF state and its confidence using the properties of the one-leg. ,else if the one-leg is in the ambiguous group, the KF state is not updated, and the predict step is skipped for this track.
Unassociated two-leg	Initiate a new track with zero velocity, zero acceleration and zero confidence.

of the outcomes for these interpretations is shown in Table. II. People tracks carry additional confidence information derived from its observation from leg tracking data. The confidence is update in a similar manner as in the leg tracking, with a constant parameter β . Two-legs group confidence is computed by averaging its legs confidences, and a single confidence is derived when a people track is associated with the one-legs group. If associated with the uncertain one-legs group, people track confidence are not updated, i.e. maintain the same confidence value as it is deemed ambiguous. The track will be removed if the confidence is below certain threshold. The ${}^P c_k^j$ and ${}^P d_k^j$ are the confidence of the people track and the confidence of the associated observation, with index j at time k , respectively.

C. Pursuit Controller

The pursuit controller drives ISERA towards the tracked target person while avoiding obstacles. The driving control is based on the selection of discrete spatial zones with the lowest cost. Fig. 3 shows the variables used in the pursuit controller. The coordinate of the target person is described in polar coordinate with range R_P , and angle θ_P with respect to the robot heading. To determine the angle θ_{BP} where the robot should go to pursuit the target, the area surrounding the robot are first divided into m number of spatial ‘buckets’ zones. The costs of each bucket are subject to evaluation, determined by factors such as the angle of each bucket to the target and to obstacles.

A dynamic lookout distance is used to regulate the obstacle detection area where the lookout distance of each buckets R_{OD} , are kept to be less than the range to target. The area is shrunk and expanded proportional to the distance to the target R_P . In this way, the target person legs would not be regarded as an obstacle to be avoided. The delta angle of each bucket with respect to the target δ_{Bn} is calculated as,

$$\delta_{Bn} = |\theta_{Bn} - \theta_P|. \quad (3)$$

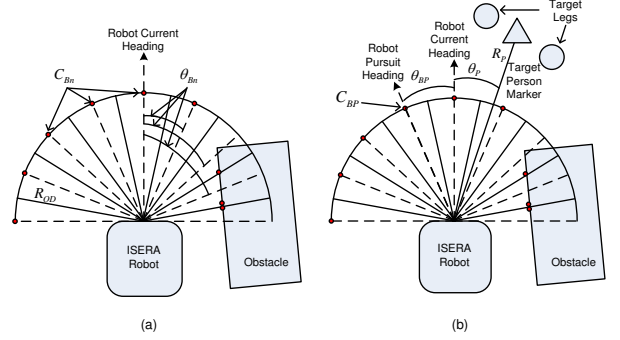


Fig. 3. Variables in the pursuit controller. (a) Illustration of the discrete spatial zone into m number of buckets (9 buckets for illustration purpose). Note that the nearest point to obstacle C_{Bn} in bucket $n = \{1, \dots, m\}$ may vary up depending on the environment (b) The parameters used in the pursuit controller

where θ_{Bn} is the angle of the center point in bucket n , and θ_P is the target angle. Then, the individual cost of each bucket is calculated as,

$$cost_{Bn} = \frac{\delta_{Bn} + 1}{C_{Bn}}. \quad (4)$$

where C_{Bn} is the distance to the nearest obstacle inside the buckets.

Naturally, when there is no obstacle inside the obstacle detection range, the nearest bucket in the direction of the target would have the lowest cost. However, to prevent the robot from going too near into an obstacle, each bucket cost is also influenced by the cost of its neighboring several buckets. By comparing the average cost of each buckets, the center point of the bucket with the lowest average cost is chosen to be the pursuit heading θ_{BP} .

The turn velocity ω_P is chosen to be proportional to pursuit heading angle θ_{BP} ,

$$\omega_P = K_\omega \times \theta_{BP}. \quad (5)$$

Where K_ω is a weighting constant.

The pursuit velocity v_P , is made to be proportional to obstacle clearance in the pursuit heading C_{BP} , the angle of the target person θ_P , and the angle of the pursuit heading θ_{BP} . For example, to pursuit with the highest speed, there should be no obstacle near the pursuit heading and the target person is in front of the robot. For other cases, the velocity will be reduced proportionally.

Another factor that has a significant impact on the robot behavior is the social zones of the person [8]. As a service robot, the robot is expected to behave in the manner that is socially acceptable. The closer the robot to the target person, the more gentle the motion should be. The control laws combining these considerations are,

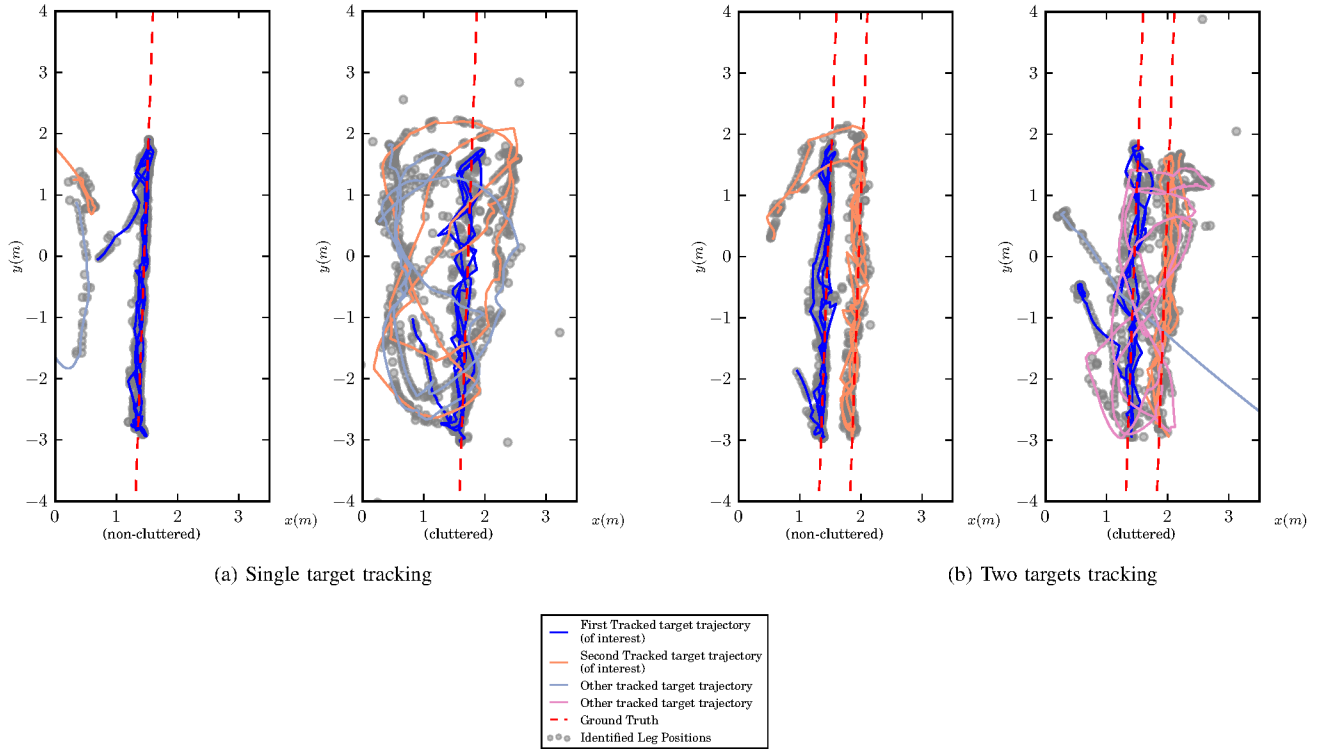


Fig. 4. Tracking results for stationary robot.

$$v_P = \frac{C_{BP}}{K_o} \times \frac{\theta_m - \theta_P}{\theta_m} \times \frac{\theta_m - \theta_{BP}}{\theta_m} \times D_P. \quad (6)$$

Where K_o is a weighting constant, and θ_m is an angle between the front and the edge of the perception zone. D_P is the parameter that changes the aggressiveness of the pursuit based on the social zone of the target person. In our ISERA, the typical values are,

$$D_P = \begin{cases} 1 & \text{if } R_P > 1.5m \\ 0.6 & \text{if } 1m < R_P < 1.5m \\ 0.2 & \text{if } 0.3m < R_P < 1m \\ 0 & \text{if } R_P < 0.3m \end{cases} \quad (7)$$

IV. EXPERIMENTATION AND RESULTS

The people detection, tracking, and pursuit algorithm has been tested both indoor and outdoor. The 2D Lidar has 180 degree towards the front of the robot, which gives 720 sampling points with 0.25 degree separating each point. The observation of clusters for leg detection is limited to be within 5 meters. Cluster size is limited to be within $0.07m - 0.2m$ and consist more than 4 sampling points. Clusters would be classified as legs if the confidence exceed a cut off threshold of 0.5. For the pursuit controller, the area in front of the robot are divided into 72 bucket zones.

A. People Detection and Tracking

A single target tracking tests are coordinated with a person walking repeatedly from one start point and an end point following a line marked on the floor, while ISERA is stationary. In a cluttered environment legs could be occluded by other people or obstacles. This condition is emulated by adding more people walking around while the primary target is still following the straight path. In addition, tracking of two targets of interest were conducted. The paths are two straight lines located with a roughly $65cm$ distance to each other.

The results of single and multiple target tracking are depicted in Fig.4a and Fig.4b, respectively. As can be seen, the tracking was executed robustly even in the middle of cluttered objects and some spurious leg measurements. It is observed that during occlusions, target can sometime deviate from the ground truth due to temporary uncertainties in the filter. By keeping the confidence level in each track, we can maintain which target to be preserved for the next filter cycle.

B. People Following

A series of experiments in a outdoor public area with obstacles to verify that ISERA can follow a person. The target person to be followed is selected by having the target person standing in front of ISERA. ISERA's pursuit controller maintains $30cm$ distance from the person. Some disturbance is emulated by having people walking around and in between ISERA and the target. The results shows ISERA consistently

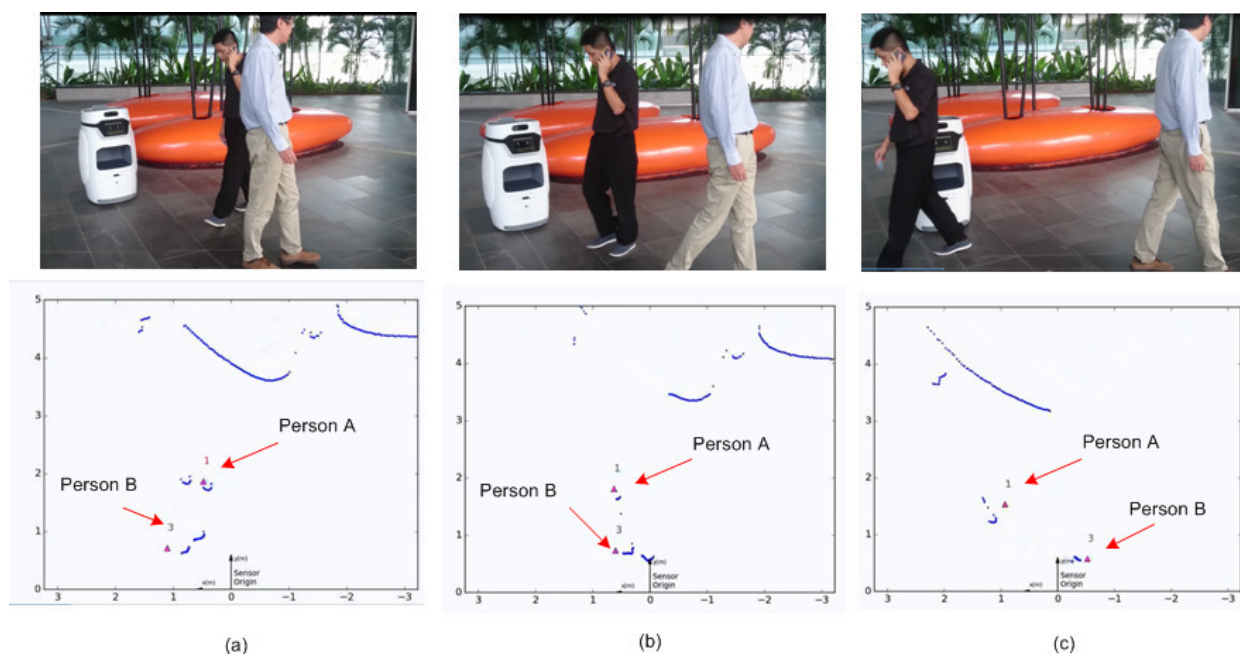


Fig. 5. ISERA tracks and follows person A (wearing blue shirt), while person B (wearing black shirt) is passing in between: (a) Person B is about to pass through between person A and ISERA. (b) For a few moments ISERA did not have line of sight of person A, and the tracker will start to decay. (c) ISERA regain line of sight and continued following person A

manage to track and follow the target person that walks below ISERA's maximum speed (limited to $1m/s$ in our hardware). When traveling between obstacles, ISERA also able to avoid going too close to the obstacle, and successfully follows the person.

V. CONCLUSION

In this paper, ISERA service robot algorithms to follow a target person have been presented. The person tracker associate the leg to a person and tracks based on the confidence level that is given by the leg identifier. The pursuit algorithm could regulate control between maintaining the distance to target person and to avoid obstacle. Future works should mitigate possible fail scenarios when the person walks too fast or being obstructed for a prolong period of time.

REFERENCES

- [1] M. Munaro and E. Menegatti, "Fast RGB-D people tracking for service robots," *Auton. Robots*, vol. 37, no. 3, pp. 227–242, 2014.
- [2] A. P. Gritti, O. Tarabini, J. Guzzi, G. A. D. Caro, V. Caglioti, L. M. Gambardella, and A. Giusti, "Kinect-based people detection and tracking from small-footprint ground robots," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, September 14-18, 2014*, 2014, pp. 4096–4103.
- [3] M. Montemerlo, S. Thrun, and W. Whittaker, "Conditional particle filters for simultaneous mobile robot localization and people-tracking," in *Proceedings of the 2002 International Conference on Robotics and Automation*, 2002, pp. 695–701.
- [4] E. A. Topp and H. I. Christensen, "Tracking for following and passing persons," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alberta, Canada, August 2-6, 2005*, 2005, pp. 2321–2327.
- [5] A. Leigh, J. Pineau, N. A. Olmedo, and H. Zhang, "Person tracking and following with 2d laser scanners," in *IEEE International Conference on Robotics and Automation, ICRA 2015, Seattle, WA, USA, 26-30 May, 2015*, 2015, pp. 726–733.
- [6] K. O. Arras, Ó. M. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data," in *2007 IEEE International Conference on Robotics and Automation, ICRA 2007, 10-14 April 2007, Roma, Italy*, 2007, pp. 3402–3407.
- [7] R. Gockley, J. Forlizzi, and R. G. Simmons, "Natural person-following behavior for social robots," in *Proceedings of the Second ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI 2007, Arlington, Virginia, USA, March 10-12, 2007*, 2007, pp. 17–24.
- [8] S. Hemachandra, T. Kollar, N. Roy, and S. J. Teller, "Following and interpreting narrated guided tours," in *IEEE International Conference on Robotics and Automation, ICRA 2011, Shanghai, China, 9-13 May 2011*, 2011, pp. 2574–2579.
- [9] J. Xavier, M. Pacheco, D. Castro, A. E. B. Ruano, and U. Nunes, "Fast line, arc/circle and leg detection from laser scan data in a player driver," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, ICRA 2005, April 18-22, 2005, Barcelona, Spain*, 2005, pp. 3930–3935. [Online]. Available: <http://dx.doi.org/10.1109/ROBOT.2005.1570721>
- [10] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [11] H. W. Kuhn and B. Yaw, "The hungarian method for the assignment problem," *Naval Res. Logist. Quart.*, pp. 83–97, 1955.