2013 IEEE RO-MAN: The 22nd IEEE International Symposium on
Robot and Human Interactive Communication
Gyeongju, Korea, August 26-29, 2013

Multiple People Tracking from 2D Depth Data by Deterministic Spatiotemporal Data Association

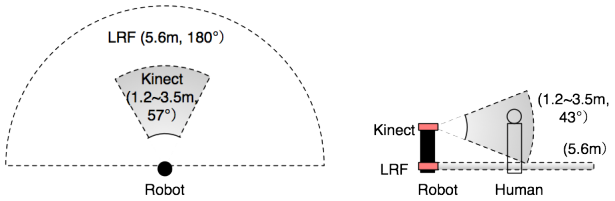
Seongyong Koo and Dong-Soo Kwon

Abstract— **This paper proposes a deterministic approach to track people in a populated environment from 2D depth data by a laser range finder attached on a mobile robot. This work aims to improve robustness of multiple people tracking in the presence of change of the number of people, missing data, and long-term occlusions by using spatiotemporal data association. The temporal data association method is based on the multi-frame tracking (MFT) and the improved MFT (IMFT) is proposed for enhancing computational efficiency in the long-term occlusions. A spatial data association algorithm used a matching algorithm from the leg history data for detecting a human subject from leg tracks. The proposed methodology has been assessed in the three walking patterns of two people and compared with MFT and MHT methods.**

I. INTRODUCTION

Tracking multiple people in populated environments is an important technique for service robots in terms of allowing long-term interaction with humans. For example, a service robot should be able to track and follow a partner in an interaction [1], and select a person who has the greatest interest in the robot among people [2], [3]. In addition, the history of behaviors of each person can be used to infer semantic information of a human such as intentions or goals [4], [5], [6].

Many critical issues for robust tracking have been addressed in previous works [7]. They are caused mainly by occlusion, objects moving in/out of the scene, and densely populated environments. In order to overcome these issues in hardware-wise terms, previous works have employed multiple sensors attached on a robot [3], [8], environment sensors [9], and cooperation of multiple robots [10]. In recent years, the advent of low-cost RGB-D cameras such as Microsoft Kinect and ASUS Xtion allow to detect environment as 3D point clouds in real-time. Despite rich information captured by the sensor with high frequency, the relatively narrow and limited view area needs to be compensated by other sensors for service robots operating in wide working area; in turn, it is required to use 2D Laser Range Finder as illustrated in Fig. 1. When there are multiple people walking around a robot, a LRF at the bottom side of the robot, equipped for safety, detects ankles of humans and tracks human candidates by the two-leg positions. A RGB-D camera at the top side of the robot can meanwhile find a human and his/her gestures to interact with among the candidates and track behaviors in detail during the interaction.



(a) Top view (b) Side view

Fig. 1. Range of LRF and Kinect sensors attached on a robot

In order for a robot to track multiple people robustly by means of a LRF (URG-04LX, Hokuyo) attached on the robot, this paper proposes a novel two-leg based deterministic tracking method to improve the performance in three generally occurring cases in populated environments: random movements, long-term occlusions, and a densely distributed area. In the remainder of this paper, related works involving model-based probabilistic approaches are described in chapter II, and the concept of deterministic data association method (MFT) is delineated in chapter III. The temporal and spatial data association methods for tracking people based on the two-leg data are presented in chapter IV. Experiments involving the three aforementioned cases are presented and the results are compared with other methods in chapter V. Finally, a conclusion to this work is given in chapter VI.

II. RELATED WORKS

In order to facilitate robust tracking of humans from detected human legs by a laser scanner, previous works have attempted to resolve several issues arising in real situations. The first issue is involved with detecting humans. [8], [10] used two-leg pattern models while [11], [12] detect one-leg clusters and find a human with the spatial relation of two legs. The measured leg or human positions, however, could suffer from false or missing measurements due to noise and errors in the previous step. This issue has been tackled by several KF-based filtering methods with prediction models or particle filtering without a model. The prediction model for estimating human position is assumed by a constant velocity model in many cases [6], [8], [10], [11], [15], [17], while [9], [12], [16] applied a human walking model by the dynamics of two legs. Employing certain assumptions of moving patterns according to the space, the model of

S. Koo is with Mechanical Engineering Department and HRI Research Center, KAIST, Deajeon, Republic of Korea koosy@robot.kaist.ac.kr

D. Kwon is with Faculty of Mechanical Engineering Department and HRI Research Center, KAIST, Deajeon, Republic of Korea kwonds@kaist.ac.kr

978-1-4799-0509-6/13/$31.00 ©2013 IEEE

the environment can be applied to the prediction of human positions [17].

The temporal data association between frames including multiple objects encompasses issues of variable number of tracks, initialization and termination of tracks, and false matching. A basic and widely used method for the association is the Nearest Neighbor (NN), wherein a new object is matched to the closest object in the previous frame within a particular region of interest. Many studies have used this approach with specific filtering methods and prediction models [6], [8], [9], [17]. However, if many objects are distributed densely in the scene and the movements do not correspond with the prediction models, the data association methods should consider the possibility of false matching. Joint Probabilistic Data Association Filter (JPDAF) and Multi-Hypothesis Tracking (MHT) methods are probabilistic approaches for temporal matching of objects according to frames. They calculate the probability of each track for all possible matches with a probability density function around new points. Although JPDAF assumes a fixed number of tracks [18], MHT extends to a variable number of tracks and hence has been used in many applications of multiple people tracking [11], [13], [15], [16]. On the other hand, several deterministic approaches have been proposed for searching optimal matches to minimize the correspondence cost, which is formulated as a combinatorial optimization problem. [19] proposed the Two Frame Matching (TFM) algorithm using graph theory. The Greedy Optimal Assignment (GOA) algorithm [20] enhanced the performance of finding optimal associations to allow occlusion and detection errors in a constant number of points. Shafique and Shah [21] proposed the Multi-Frame Tracking (MFT) algorithm to improve tracking performance by considering point information in multi-frames for a variable number of points.

Although the probabilistic approach has been applied successfully in many tracking applications, it suffers from some intractable problems. First, it requires assumptions of a probability density function for all points, which do not always hold, and the tracking performance is sensitive to a number of parameters of the model. In addition, the algorithm is computationally demanding, since the complexity grows exponentially with the number of points. These drawbacks require variable assumptions and models to make the algorithm efficient in specific applications, and the parameters should be tuned by experiments for the given situation. [13] invested several models about occlusion, deletion, new tracks, and false matching to find the best model of MHT in pedestrian tracking.

In this paper, for people tracking in general situations, i.e., densely populated, random walking, variable number of humans, and allowing occlusion and detection errors, the models of human movements and detected points need to be eliminated as much as possible, parameters for the algorithm should be minimized, and the developed method should be able to operate in real-time. The following sections explain the concept and processes of temporal data association using Improved MFT (IMFT) and spatial data association using the history information of each leg track.

III. BACKGROUND : MUMTI-FRAME TRACKING

MFT is an efficient and robust data association method based on the noniterative greedy algorithm [21]. The maximum matching algorithm among multi-frame data allows corrections of existing correspondences, which compensates occlusion and detection errors of points. Fig. 2 summarizes the processes of MFT. In the first two frames, there are three points each in the first frame($v_{11}v_{12}, v_{13}$) and in the second frame(v_{21}, v_{22}, v_{23}), respectively. Extension edges can then be connected to all points between two frames, as seen in Fig. 2(a). Based on the weight values at each edge, the maximum

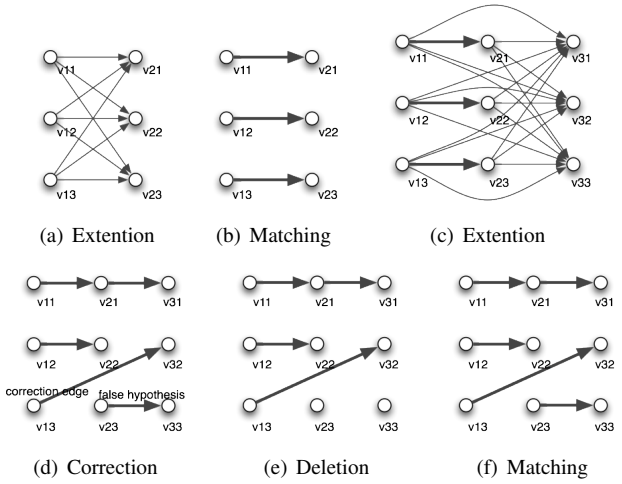
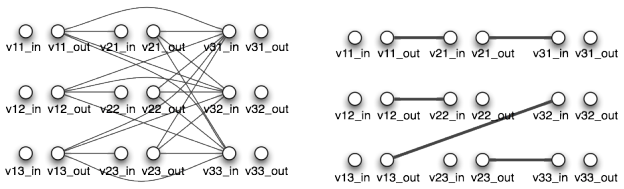


Fig. 2. Processes of Multi-Frame Tracking algorithm

matching algorithm is performed to find optimal correspondences, as indicated by the bold arrows in Fig. 2(b). When points in a new frame come into the graph, the extension edges are generated from all points in the existing k-frames, as seen in Fig. 2(c). These extended edges could share points with existing correspondences, which results in correction edges and false hypotheses after maximum matching. Fig. 2(d) shows the correction edge($v_{13} - v_{32}$) of the previous correspondence($v_{13} - v_{23}$), and the false hypothesis($v_{23} - v_{33}$) caused by the false correspondence($v_{13} - v_{23}$). Because the false hypothesis is meaningless with the correction edge, it is removed in the correspondences, as in Fig. 2(e). After the deletion step, the remaining unconnected points perform maximum matching between adjacent frames, as presented in Fig. 2(f).

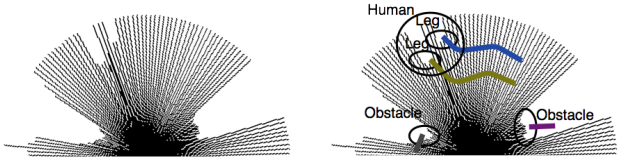
The purpose of the matching step is to find a set of edges among extended edges to maximize the summation of cost values. In addition, the selected edges should not share start or end nodes with other edges. Fig. 3(a) is a bipartite graph of the digraph in Fig. 2(c). The maximum weighted matching algorithm can find correspondences where the sum of the values of the edges have a maximum value without sharing common points, as seen in Fig. 3(b).

The weight value of each edge represents the degree of the association between two points. It can be defined as a real



(a) Extention (b) Matching

Fig. 3. Maximum weighted matching in a bipartite graph of multi-frame leg points



(a) 2D depth data (D^t) (b) Results of people tracking

Fig. 4. An example of people tracking from laser scanned data

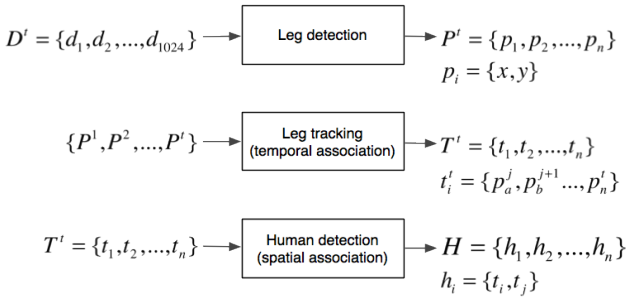


Fig. 5. Processes and annotations of people tracking

value between 0 and 1 by any criteria pertaining to the relations. This allows the adoption of prediction models, filtering methods, and even probabilistic assumptions. For example, (1) is a gain function to represent the convex combination of magnitude and orientation differences between estimated and measured vectors [21]. x_1 is a point of a existing track, and x_2 is a new point at the current frame. $\hat{x_1}$ represents the estimated position of a track, x_1, at the new frame. S_x and S_y are the x, y sizes of the scene.

$$gain(x_1, x_2) = \alpha \left[\frac{1}{2} + \frac{\overrightarrow{x_1\hat{x_1}} \cdot \overrightarrow{x_1x_2}}{2\left\|\overrightarrow{x_1\hat{x_1}}\right\| \left\|\overrightarrow{x_1x_2}\right\|} \right] + (1-\alpha) \left[1 - \frac{\left\|\overrightarrow{x_1\hat{x_1}} - \overrightarrow{x_1x_2}\right\|}{\sqrt{S_x^2 + S_y^2}} \right], \alpha \in [0,1] \quad (1)$$

The size of multi-frames, k, plays a role as a sliding window in which all point histories of all tracks are extended to the points in a new frame; this means corrections of existing correspondences are possible in the window. If k is larger than the occlusion time, the disappeared or mismatched tracks can be recovered to the detected points after the occlusion. However, because a longer window necessitates a larger search space, the determined value of k should be appropriate for real-time calculation.

IV. TWO-LEG BASED PEOPLE TRACKING

A. Processes and Anotation

The goal of people tracking is to find position histories of tracks of people from 2D depth data at each time frame measured by LRF. In this research, we assumed a human on the basis of two legs that are associated spatially for minimizing detection error and acquiring the human orientation as well as the positions of the two legs. Fig. 4 shows an example of people tracking results. If there are leg tracks that are not associated with others, they are regarded as not being human subjects. The details of the procedures and annotations of people tracking are depicted in Fig. 5.

The purpose of leg detection is to find points of leg candidates (P) at each time step by checking the size of the cluster, which is separated by the discontinuity condition. The leg size is assumed to be between 7 cm and 30 cm, and the discontinuity condition is (2). n is the size of normalization of the distance in the cluster, which is set as 5 in this research. th_c determines the size of depth in the cluster, and is 10 cm in this research. This method is simple and robust for clustering laser scanned data, and many works have accordingly used variations of this method [6], [9], [11], [12], [16].

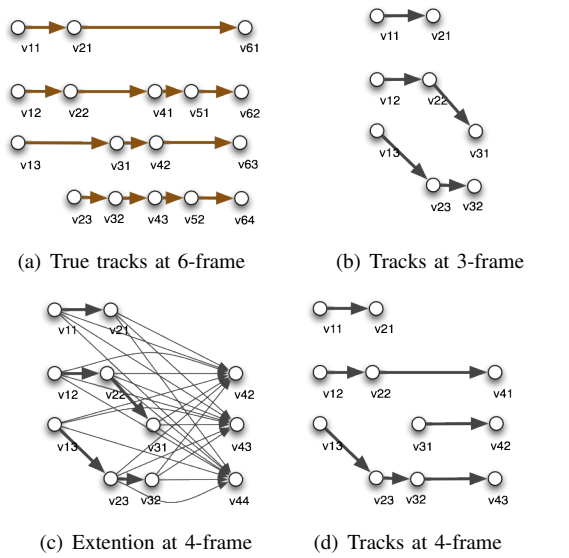
$$\sqrt{(\frac{1}{n}\sum_{j=1}^{n} d_{i-j})^2 + (d_{i+1})^2} > th_c \quad (2)$$

The leg tracking is a process of temporal data association for each point of leg candidates. The human detection process entails establishing pairs of legs, which have close spatial associations, based on the history information. The human data including two individual leg histories can be used to calculate the human position and orientation.

B. Temporal Data Association : Improved MFT

The temporal data association is based on MFT with a modification of a sliding window in order to overcome characteristics of leg data. The measured human legs are frequently missed by occlusions due to another leg or obstacles, and by detection errors or environment noise. Occlusions normally cause longer missing time than that induced by detection errors. Fig. 6(a) shows an example of true tracks from frames 1 to 6. A relatively long edge of $v_{21} - v_{62}$ occurs due to the occlusion, while short edges ($v_{22} - v_{41}, v_{13} - v_{31}, v_{42} - v_{63}$) are caused by detection errors. The short edges can be generated by correction edges. For example, the wrong edge of $v_{22} - v_{31}$ in Fig. 6(b) is modified to the correction edge of $v_{22} - v_{41}$ by the extension edges in Fig. 6(c). The initial wrong edge of $v_{13} - v_{23}$ can be corrected by backtracking in the reverse direction at the first k-frame [21]. The long edge of $v_{21} - v_{61}$ can be generated at frame 6 by extension of the terminal nodes of tracks, v_{21}, and new points. This means the sliding window k should be large enough to cover all the terminal nodes during the occlusion time. However, it not only covers terminal nodes, but also

connected nodes to make extensions to the new points, which almost do not need to be corrected.

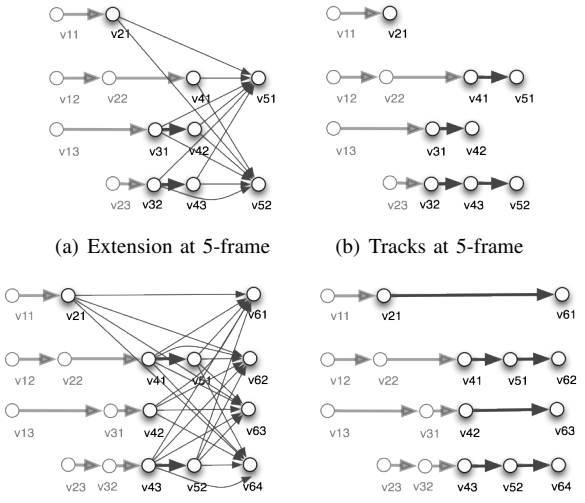


(a) True tracks at 6-frame (b) Tracks at 3-frame

(c) Extention at 4-frame (d) Tracks at 4-frame

Fig. 6. Multi-frame leg tracks with long-term and short-term occlusions

In order to tackle the long-term occlusion problem in human leg tracking for real-time calculation, we introduced two kinds of sliding windows, a short-term window, s-frame, and a long-term window, l-frame. All the existing nodes in s-frame can be extended while only terminal nodes in l-frame can be extended to the new points. Only a few nodes are terminal nodes in the long-term and the wrong edges are corrected in the short-term; these two kinds of windows make the algorithm more efficient and allow robust tracking in a multi-frame sequence without increasing computational complexity. Fig. 7 shows the effects of the improved MFT with 5 of l and 3 of s.



(a) Extension at 5-frame (b) Tracks at 5-frame

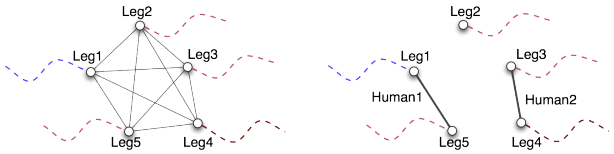
(c) Extention at 6-frame (d) Tracks at 6-frame

Fig. 7. An example of people tracking by improved MFT

In Fig. 7(a), the nodes out of s-window, except the terminal nodes (v_{21}), which are gray points, do not generate extension nodes to the points of frame 5. The terminal node (v_{21}) in the l-window can be extended to the new frame, as shown in Fig. 7(c). The number of edges in the extension graphs of Fig. 7 is substantially smaller than that of the extended graph of MFT with the same l window size.

C. Spatial Data Association : Maximum Matching based on the Leg Histories

Each leg track contains trajectory information. At each time frame, all leg tracks develop a graph including nodes of tracks and edges of weight values between two tracks, as seen in Fig. 8(a). After maximum weighted matching is performed, pairs of two legs are associated to generate the maximum value of total cost, as shown in Fig. 8(b).



(a) Extension graph of leg tracks (b) Maximum matching

Fig. 8. Maximum weighted matching for spatial data association

Each leg track has a different start, end times, and duration, as presented in Fig. 9.

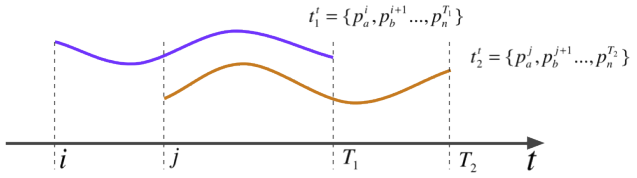


Fig. 9. Two leg tracks having different durations

In order to calculate the weight value between two leg tracks, the weight value is defined by (3).

$$gain(t_1^t, t_2^t) = \frac{\sum_{k=max(i,j)}^{min(T_1,T_2)} \gamma^{max(T_1,T_2)-k}\left(1 - \frac{\left\|p_1^k - p_2^k\right\|}{\sqrt{S_x^2 + S_y^2}}\right)}{max(T_1, T_2)} \quad (3)$$

The weight is the average of the degree of closeness between two tracks existing at the same time frames. γ is a forgetting factor that allows recent data to have greater effects on the weight value than old data. This function reflects the close spatial relation and historical contemporary of two legs of a human.

V. EXPERIMENTS AND RESULTS

The proposed people tracking method was tested in three experimental cases that occur generally and cause tracking errors frequently. The tracking performance and computational time in these three cases were compared with those of a deterministic method, MFT, and a probabilistic method, MHT.

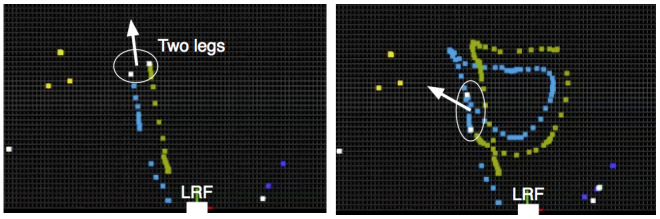
A. Three Experimental Cases

1) A person randomly walking: This case tests tracking of a human whose movements cannot be predicted. A human is walking around the space randomly, as in Fig. 10(a), and the legs occlude each other during natural walking.

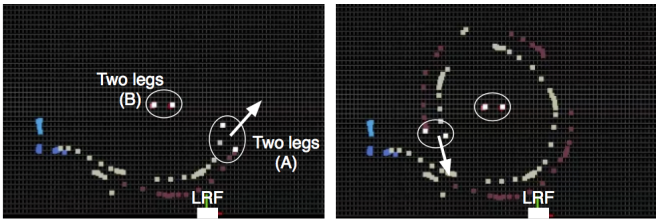
2) A person circularly walking around another person: When many people are walking in a space their legs occlude each other, which causes long-term missing points. This case was tested by a person walking around a standing person, as in Fig. 10(b).

3) Two people walking cross each other: Legs of people in a densely populated environment could be mismatched relative to the true tracks. The two people cross paths very closely two times, as in Fig. 10(c).

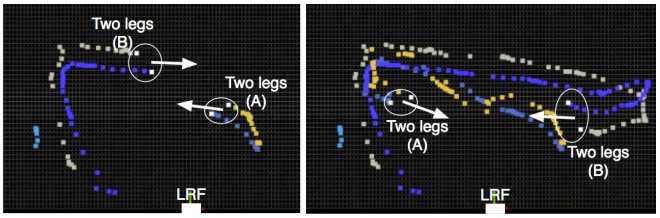
The leg data of the three cases were measured by the leg detection process illustrated in Fig. 5 at every 100 ms for 20 seconds. IMFT, MFT, and MHT were performed as temporal data association methods for tracking human legs and the proposed spatial data association method was applied to the three methods.



(a) Case 1: A person randomly walking



(b) Case 2: A person circularly walking around another person



(c) Case 3: Two people walking cross each other

Fig. 10. Detected Leg points and tracks in the three cases

B. Results

In order to assess and compare performances of different methods for multiple object tracking, CLEAR MOT metrics were proposed by [22]. There are two metrics, multiple object tracking precision (MOTP), which represents the ability to estimate precise object positions, and multiple object tracking accuracy (MOTA), which is defined to account for all object tracking errors over all frames. In this research, MOTP is ignored, because the ability of position prediction is not our concern and it needs the truth position of humans, which we cannot ascertain without other localization devices. MOTA is defined as (4) by three parameters at each time, missing tracks(m_t), false positives(fp_t), and mismatches(mme_t) to the ground truth(g_t).

$$1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \tag{4}$$

TABLE I

RESULTS OF LEG TRACKING EVALUATION BY THE THREE METHODS FOR THE THREE CASES

Model	Case	MT	FP	MM	MOTA	FPS
IMFT	Random	37	0	0	0.815	16.7
	Circular	27	2	6	0.815	28.5
	Cross	3	1	5	0.955	20
MFT	Random	38	0	0	0.81	18.1
	Circular	33	12	11	0.72	28.5
	Cross	8	0	6	0.93	25
MHT	Random	40	0	0	0.8	12.1
	Circular	35	16	5	0.72	15.2
	Cross	10	2	8	0.9	17.3

Although sliding windows of larger size can improve the tracking performances of IMFT and MFT, they were defined as the largest values to real-time operation, over 10 fps. For this test and comparison, the window of MFT, k, was 20, and the two windows of IMFT, l, s, were selected as 50 and 12, respectively. The gain function of (1) for the weights was used with a constant velocity assumption. Table I shows the results of leg tracking evaluation by the three aforementioned methods for the three cases. In all the three cases, IMFT performs better than the other two methods. Especially for the case 2, circularly walking, the errors of IMFT is significantly less than the others, because it can stand the long occlusion by the long window. The two deterministic methods performs better in terms of the computational load together with the tracking performance.

VI. CONCLUSIONS AND FURTHER WORKS

In this paper we presented a deterministic people tracking methodology based on an Improved MFT (IMFT) as a temporal data association technique and matching based on the leg history data as a spatial data association technique. Three difficult cases for tracking that frequently occur in populated environments have been proposed and experimented to assess the performance of the proposed tracking method in comparison with MFT and MHT. The results showed that the IMFT performs better than the other methods and also provides higher efficiency. This stems from the multi-frame association allowing corrections and extensions in the long-term data missing cases by a sliding window that is relatively larger than the occlusion time.

However, the proposed method was evaluated only in three representative cases involving two people without any other objects. Although the deterministic method has fewer

parameters than probabilistic methods, the number of points that can be tracked in real-time should be verified. In addition, the 2D depth data for detecting human legs may not be sufficient to track people in some cases such as subjects standing side by side with duplicated legs or where the lower body is occluded by objects. This deterministic approach for tracking can be extended with other sensor data.

ACKNOWLEDGMENTS

This work was supported by the Industrial Strategic Technology Development Program(10044009, Development of a self-improving bidirectional sustainable HRI technology) funded by the Ministry of Knowledge Economy(MKE, Korea).

REFERENCES

[1] R. Gockley, J. Forlizzi, and R. Simmons, Natural Person-Following Behavior for Social Robots, in Proc. of the ACM/IEEE Int. Conf. on Human-robot interaction (HRI), 2007, pp .17-24.

[2] S. Satake, T. Kanda, D.F. Glas, M. Imai, H. Ishiguro, and N. Hagita, How to Approach Humans? Strategies for Social Robots to Initiate Interaction, in Proc. of the 4th ACM/IEEE Int. Conf. on Human robot interaction (HRI), 2009. pp. 109-116.

[3] 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, vol. 43, no. 2-3, pp. 133-147, 2003.

[4] K.A. Tahboub, Intelligent Human-Machine Interaction Based on Dynamic Bayesian Networks Probabilistic Intention Recognition, Journal of Intelligent and Robotic Systems, vol. 45, no. 1, pp. 31-52, 2006.

[5] R. Kelley, A. Tavakkoli, C. King, M. Nicolescu, M. Nicolescu, and G. Bebis, Understanding human intentions via hidden markov models in autonomous mobile robots, in Proc. of the 3rd ACM/IEEE Int. Conf. on Human robot interaction (HRI), 2008, pp. 367-374.

[6] S. Koo and D.S. Kwon, Recognizing human intentional actions from the relative movements between human and robot, in Proc. of the 18th IEEE Int. Symp. on Robot and Human Interactive Communication, (RO-MAN), 2009, pp. 939-944.

[7] A. Yilmaz, O. Javed, and M. Shah, Object tracking: A survey, ACM Computing Surveys (CSUR), vol. 38, no. 4, pp. 13-48, 2006.

[8] N. Bellotto, H. Hu, Multisensor-based human detection and tracking for mobile service robots, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 39, no. 1, pp. 167-181, 2009.

[9] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, Multi-modal tracking of people using laser scanners and video camera, Image and vision Computing, vol. 26, no. 2, pp. 240-252, 2008.

[10] C.T. Chou, J.Y. Li, M.F. Chang, and L.C. Fu, Multi-robot cooperation based human tracking system using Laser Range Finder, in Proc. of the 2011 IEEE Int. Conf. on Robotics and Automation (ICRA), 2011, pp. 532-537.

[11] 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 Proc. of the 2008 IEEE Int. Conf. on Robotics and Automation (ICRA), 2008, pp. 1710-1715.

[12] J.H. Lee, K. Abe, T. Tsubouchi, R. Ichinose, Y. Hosoda, and K. Ohba, Collision-free navigation based on people tracking algorithm with biped walking model, in Proc. of the 2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2008, pp. 2983-2989.

[13] M. Luber, G.D. Tipaldi, and K.O. Arras, Better Models For People Tracking, in Proc. of the 2011 IEEE Int. Conf. on Robotics and Automation (ICRA), 2011, pp. 854-859.

[14] N. Bellotto and H. Hu, Computationally efficient solutions for tracking people with a mobile robot: an experimental evaluation of Bayesian filters, Autonomous Robots, vol. 28, no. 4, pp. 425-438, 2010.

[15] B. Lau, K.O. Arras, and W. Burgard, Multi-model hypothesis group tracking and group size estimation, International Journal of Social Robotics, vol. 2, no. 1, 2010, pp. 19-30.

[16] G. Taylor and L. Kleeman, A multiple hypothesis walking person tracker with switched dynamic model, in Proc. of the Australasian Conference on Robotics and Automation, 2004.

[17] K.S. Tseng and A.C.W.Tang, Goal-oriented and map-based people tracking using virtual force field, in Proc. of the 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2010, pp. 3410-3415.

[18] D. Schulz, W. Burgard, D.Fox, and A.B. Cremers, People tracking with mobile robots using sample-based joint probabilistic data association filters, The International Journal of Robotics Research, vol. 22, no. 2, pp. 99-116, 2003.

[19] B. Kluge, C. Kohler, and E. Prassler, Fast and robust tracking of multiple moving objects with a laser range finder, in Proc. of the 2001 IEEE Int. Conf. on Robotics and Automation (ICRA), 2001, pp. 1683-1688.

[20] C.J. Veenman, M.J.T. Reinders, and E. Backer, Resolving Motion Correspondence for Densely Moving Points, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 1, pp. 54-72, 2001.

[21] K. Shafique and M. Shah, A Noniterative Greedy Algorithm for Multiframe Point Correspondence, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 1, pp. 51-65, 2005.

[22] K. Bernardin and R. StiefelhagenK, Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics, Journal on Image and Video Processing, 2008.