Laser-based Person Tracking for Clinical Locomotion Analysis

Angus Leigh, Joelle Pineau
School of Computer Science, McGill University
{angus.leigh, jpineau}@cs.mcgill.ca

I. Introduction

Accurate knowledge of the positions of people over time provides rich, objective and quantitative data which can be highly useful in clinical research of human locomotion behaviour. Examples of recent studies which have relied on this information include analyses of the factors influencing gait speed of stroke subjects [1] and the effects of visual perceptions on the turning ability of stroke subjects [2]. One commonly used method for gathering this data in clinical scenarios is through motion capture systems, which require a set of infrared markers to be placed on each participant and a fixed array of infrared cameras, which must be recalibrated before each use.

Separate from this clinical research, significant developments have been made in person detection and tracking in the robotics realm, often with the aim of improving human-robot interaction or robot navigation in populated environments [3] [4]. This paper presents one such method of tracking people using planar laser range scanners, which was originally developed for use in robotics but we believe can also be useful in the field of clinical locomotion analysis.

The person tracking method presented here relies on planar laser range scanners, which are distance-measuring devices which use the time-of-flight principle to provide a fine array of distances in a wide field of view (typically 180° or more) on a plane with the laser. Since the presented approach uses a laser range scanner instead of a fixed motion-capture system, such as that used in [1] and [2], it benefits from greater portability, less cost and allowing its users to be tracked naturally as they appear without the need for added markers. Further, unlike pure vision systems, it can reliably provide accurate distances without the need for stereo correspondences.

II. Laser-based multi-person tracking

People are detected and tracked in laser scans using a combination of a machine learning classification technique designed to detect legs from laser scans, and a filtering method to track multiple individuals over time. The method itself is novel, and extends previous methods on tracking-by-detection such as [3] by incorporating a continuous detection confidence to improve data association and by using leg pairings to improve robustness in detection and tracking.

A. Physical laser setup

The laser is placed such that it is level with the ground and at a height between ankle and mid-thigh of most participants. In practice, we have experienced best results at mid-shin height but have found it to be relatively robust to other heights, as long as it lies in the participants' leg region. The laser is then connected to a computer via USB to process the range measurements.

Most of our experiments use either a Hokoyu UHG-08LX or URG-04LX-UG01 laser rangefinder, but any laser with a comparable angular resolution (1/3°) could be used as a drop-in replacement. Lasers of different angular resolution would work as well but would require retraining of the leg detection method presented in Sec. II-C to achieve best performance. Use of a higher resolution laser scanner would result in greater accuracy and longer-range tracking, while the opposite is true if a coarser resolution scanner were used.

B. Laser scan processing

Scan points returned from the laser are first clustered according to a distance threshold. Any points which are within a fixed threshold of each other are grouped together as a cluster. The threshold ($d < 13cm$) was chosen to be small enough to often separate a person's two legs into two distinct clusters, which is useful in later stages of the tracking, but rarely resulted in generating more than two clusters per person. Further, clusters containing less than three scan points are discarded, as they provide too little geometric information and may be the result of noise.

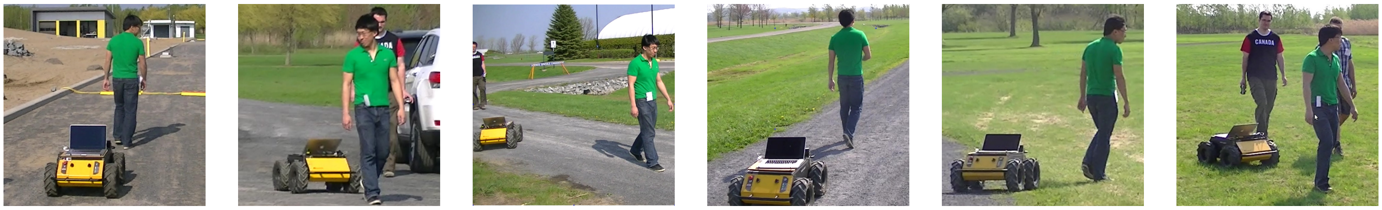
C. Leg-based person detection

Each cluster is then processed to extract a set of features describing its geometric shape. These features are listed in Table I and extend the feature set proposed in [3].

TABLE I. Features used for human detection from scan clusters.

Number of points	Width	Length
Standard deviation	Avg distance from median	Distance to adjacent cluster
Linearity	Circularity	Radius of best-fitting circle
Boundary length	Boundary regularity	Mean curvature
Mean angular difference	Inscribed angular var.	Distance from laser scanner
Occluded (boolean)		

Clusters are then automatically classified as human or non-human, based on these features. The classification is done by training a random forest classifier using a set of positive and negative examples [5]. Positive examples are obtained by setting up the laser scanner in a hallway or open area with significant pedestrian traffic; all clusters which lie in the open areas and meet the threshold in Sec. II-B are assumed to be the result of people and are used as positive training samples. Negative examples are obtained by moving the sensor around



(a) (b) (c) (d) (e) (f)

Fig. 1. The person tracking and following system implemented on a Clearpath Husky with a Hokuyo URG-04LX-UG01 laser sensor mounted at a height of approximately $40cm$. The robot autonomously followed the participant in an approximately $500m$ loop on gravel and grass at the Canadian Space Agency in Saint-Hubert. Two tracking failures occurred, one of which is shown in (c), which were caused by extensive sensor noise, as the laser scanner used is indoor-rated and highly susceptible to interference from the sun.

in an environment devoid of people; all clusters meeting the threshold in Sec. II-B are used as negative training samples.

One benefit of using a random forest classifier is that because it uses an ensemble of decision trees, a measure of confidence in the classification can be extracted by considering the number of individual decision trees predicting each category [1]. All clusters that have a classification confidence above a threshold value are considered as positive leg detections. Positively detected legs which are within a threshold Euclidean distance ($d < 0.6m$) of each other are then paired together, as it is assumed the two detections represent two individual legs belonging to the same person. This pairing is done in a greedy manner starting with the positively detected clusters that have the closest Euclidean distance to each other and stopping when the closest distance between any two clusters is greater than the threshold distance.

D. Kalman filter tracking

The position of each detected person is then tracked over time using a separate (non-extended) Kalman filter [7]. Each Kalman filter uses a state estimate containing the position and velocity of the person in 2D coordinates. The motion model is assumed to be constant velocity and only the position portion of the state is included in the observation matrix.

E. Global nearest neighbour data association

Since the system is designed to track multiple people, uncertainty arises pertaining to how detected people should be matched to tracked people. This can potentially become an intractable combinatorial optimization problem. To address this, we use a global nearest neighbour data association method which is solvable in polynomial time via the Munkres assignment algorithm [8]. The cost metric used in the assignment is a combination of Euclidean distance and detection confidence from the random forest and, more specifically, is

$$cost(det_i, track_j) = 1 - p_o(det_i, track_j)p_c(det_i) \quad (1)$$

where det_i is the i^{th} detected person, $track_j$ is the j^{th} tracked person, $p_o(det_i, track_j)$ is our estimated probability of observing a detection at the location of det_i from $track_j$ based on their relative Euclidean distance and $p_c(det_i)$ is the normalized confidence from the human classifier for det_i. The inclusion of the detection confidence in this equation discourages matchings between tracked people and non-human objects. Further, a maximum matching distance, $dist_{max}$, of $0.8m$ is imposed between detections and tracks. $p_o(det_i, track_j)$ is estimated as $(dist_{max} - dist(det_i, track_j))/dist_{max}$. Also note that detections can arise from either a pair of legs or a single leg, depending on whether or not the legs were matched using the method described in Sec. II-C. In the case of a pair of legs, the average of each leg's position and confidence is used in Eq. 1.

F. Track initiation and deletion

People tracks are initiated when a person detection arising from two simultaneously detected legs, which both have a detection confidence above an initiation threshold, is found but not matched to any existing person tracks. Tracks are deleted when they are not matched to a detection in the previous 3 seconds or have had low confidence values assigned to them from the automatic classifier in their previous updates.

G. Implementation

The tracking system is implemented in the Robot Operating System (ROS) [9] and is based on and reuses code from the leg_detector package[2] originally developed at Willow Garage by Caroline Pantofaru and extended by David Lu in [4]. It is provided open-source on the author's website and includes all data used to train the classifier as described in Sec. II-C.

III. System deployment

The laser-based multi-person tracking system has been previously deployed and evaluated onboard a robotic wheelchair, where the person detection and tracking modules are paired with a closed-loop control algorithm to allow the smart wheelchair to automatically follow a walking companion in a dynamic indoor public space, such as a university, mall or museum. The system has also been deployed on a Clearpath Husky in an outdoor environment. Pictures of this deployment are presented in Figure I. We have considered detection and tracking of up to 10 individuals in an unconstrained public setting. Formal data analysis from these experiments is underway and will provide benchmark information about the accuracy and reliability of the system.

[1]A function for extracting this confidence is included in the random forest classifier implementation for OpenCV [6].

[2]http://wiki.ros.org/leg_detector

Three sample videos have also been included with this paper to demonstrate the tracking capabilities of the deployed system for the target application of tracking individuals during rehabilitation therapy. Two videos show two participants being tracked as they walk arbitrarily around in open space in front of a stationary laser scanner and a third video shows tracking in a similar scenario except with a stationary obstacle placed in the centre of the scene.

IV. Discussion

We briefly present an algorithmic approach for multi-person detection and tracking in dynamic environments. The main motivation for developing this system is to permit assessment and evaluation of locomotion behaviour in individuals undergoing rehabilitation therapy in a variety of environments and conditions. The method we proposed is based on recent developments in the robotics literature. At the technical level, there are several aspects of the method that are currently under development to improve accuracy and reliability of the system. First, we are developing new methods for automatically estimating the motion and observation model of the Kalman filter directly from data. Second, we are working on incorporating information from other sensors (including camera) to improve disambiguation of individuals in cases of occlusion. At the clinical level, we are preparing to use the system to characterize the locomotion behavior of stroke patients both in a rehabilitation hospital and in natural living spaces. The latter deployment in natural living spaces is particularly promising, as this is the type of study which was not feasible with a fixed motion capture system. Thus, we hope to capture behaviours not observed in clinical settings.

Acknowledgment

The authors would like to thank Andrew Sutcliffe, Martin Gerdzhev, Anouk Lamontagne, Samir Sangani, Qiwen Zhang and Nicolas Olmedo for their helpful discussion and support.

References

[1] R. Bayat, H. Barbeau, and A. Lamontagne, “Speed and temporal-distance adaptations during treadmill and overground walking following stroke,” *Neurorehabilitation and neural repair*, vol. 19, no. 2, pp. 115–124, 2005.

[2] A. Lamontagne, J. Fung, B. McFadyen, J. Faubert, and C. Paquette, “Stroke affects locomotor steering responses to changing optic flow directions,” *Neurorehabilitation and neural repair*, vol. 24, no. 5, pp. 457–468, 2010.

[3] K. O. Arras, B. Lau, S. Grzonka, M. Luber, O. M. Mozos, D. Meyer-Delius, and W. Burgard, “Range-based people detection and tracking for socially enabled service robots,” in *Towards Service Robots for Everyday Environments*, pp. 235–280, Springer, 2012.

[4] D. V. Lu and W. D. Smart, “Towards more efficient navigation for robots and humans,” in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pp. 1707–1713, IEEE, 2013.

[5] L. Breiman, “Random forests,” *Machine learning*, 2001.

[6] G. Bradski, “The opencv library,” *Doctor Dobbs Journal*, vol. 25, no. 11, pp. 120–126, 2000.

[7] R. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, 1960.

[8] H. W. Kuhn, “The hungarian method for the assignment problem,” *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[9] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, “Ros: an open-source robot operating system,” in *ICRA workshop on open source software*, vol. 3, p. 5, 2009.