G Taylor and L Kleeman, "A Multiple Hypothesis Walking Person Tracker
with Switched Dynamic Model", Proceedings of the Australasian
Conference on Robotics and Automation, Dec 2004, Canberra, Australia

A Multiple Hypothesis Walking Person Tracker with Switched Dynamic Model

Geoffrey Taylor and Lindsay Kleeman
ARC Centre for Perceptive and Intelligent Machines in Complex Environments
Monash University, Australia
{Geoffrey.Taylor; Lindsay Kleeman}@eng.monash.edu.au

Abstract

This paper presents the first comprehensive model for a walking person in range data from a scanner mounted at leg height. Our model helps to distinguish people from other types of moving targets and provides good tracking robustness. The central assumption of the walking model is that at least one leg always remains stationary. The current implementation tracks a single person in a multiple hypothesis framework. We extend the multiple hypothesis framework to allow for both association uncertainty and a switched dynamic model depending on the currently moving leg. Furthermore, an occlusion model and non-stationary dynamic state transition probabilities are used in the evaluation of hypotheses to further improve tracking robustness. Experimental results demonstrate the robustness and efficiency of the proposed framework.

Figure 1: Range scanner mounted on a mobile robot.

1 Introduction

This paper addresses the problem of tracking a walking person using measurements from a laser range sensor mounted on the mobile robot shown in Figure 1. While the robot currently only acts as a stationary platform, the potential applications of people tracking in mobile robotics are numerous. For example, a service robot could follow a person or verify that the robot is being followed. For mapping and localization, people tracking enables the robot to identify persistent distractions that should be discarded during map construction, and predict the motion of obstacles for more effective path planning and obstacle avoidance. Apart from robotics, range-based people tracking could play an important role in surveillance applications.

Various approaches to range-based people tracking have already been demonstrated. In most cases the state of the person is described by a simple centre of mass, tracked in successive scans using nearest neighbour matching [Prassler *et al.*, 1999] or feature-based matching of the range profile [Fod *et al.*, 2002]. Motion models are limited to Brownian [Brooks and Williams, 2003; Montemerlo *et al.*, 2002] and constant velocity [Cielniak and Duckett, 2003] approximations. State estimation is usually performed with a Kalman filter, but more sophisticated methods have also been adopted to deal with uncertain measurement associations, such as particle filter [Montemerlo *et al.*, 2002] and joint probabilistic data association [Schulz *et al.*, 2001] based people trackers.

Almost none of the above methods attempt to model the appearance or motion of a walking person in range data, and in reality can only be considered *moving object* trackers. The exception is the method presented by Brooks and Williams [2003], which simply requires two closely spaced blobs to validate a pair of legs. The absence of explicit models is usually justified by noting that human behaviour is complex and unpredictable. Conversely, we propose that walking is highly regular, and exploit this notion to improve tracking robustness and distinguish people from other moving objects. In this paper, we develop for the first time a complete kinematic and dynamic model for a walking person in range data from a scanner mounted at leg height.

A central feature of our tracker is a switched dynamic

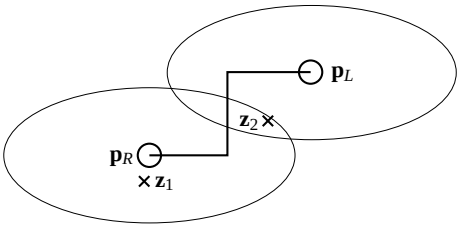


Figure 2: Association of measurements using validation gate.

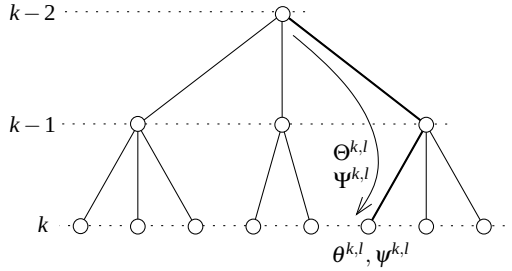


Figure 3: Three generations of a hypothesis tree.

model for a left or right moving leg (while the other remains stationary). The tracker must continuously choose the most likely dynamic model to explain the current observations, which is achieved by extending the multiple hypothesis tracking (MHT) framework [Reid, 1979] to cope with uncertain dynamic state transitions in addition to uncertain measurement associations. Occlusions and correlations between dynamic states are explicitly modelled in the MHT to further improve tracking robustness. In related work, particle filtering has been used to solve dynamic state transitions and measurement associations for people tracking in vision [Isard and Blake, 1998]. However, the required number of particles makes the approach less computationally tractable than MHT.

The following section outlines the framework of our multiple hypothesis people tracker. Section 3 then details the kinematic and dynamic models of a walking person in range data, while Section 4 describes the generation, evaluation and elimination of hypotheses used to estimate the state of the person in each measurement cycle. Finally, the tracker is applied to real laser scans and the results are presented in Section 5.

2 Overview of Tracking Framework

Our SICK LMS laser range scanner is mounted at a height of about 20 cm on a stationary robot, as shown in Figure 1. The sensor scans a 180 degree slice at a rate of 40 Hz, with an angular resolution of 0.5 degrees and a range resolution of 1 cm. For the kth update cycle, the measurement set consists of 2D candidate leg positions $z^k = \{\mathbf{z}_i^k\}$, $i \leq i \leq n$, measured in the plane of the laser. To generate z^k, the scan is segmented at range discontinuities and segments are classified as leg candidates if the distance between end-points is roughly equal to a typical leg diameter. Measurement $\mathbf{z}_i^k$ is calculated as the centroid of samples in the ith candidate leg segment. This measurement process introduces a bias since the scanner samples only one side of any leg, but the effect is ignored in practice.

To deal with the uncertainty of associations between measurements and the legs of the walking person, multiple hypothesis tracking generates and evaluates all possible associations in a Bayesian probability framework. Measurements are considered to arise from one of three sources: the left leg L, right leg R or noise (false alarm) F. Figure 2 illustrates the association problem for two measurements, $\mathbf{z}_1$ and $\mathbf{z}_2$, with the legs at predicted positions $\mathbf{p}_L$ and $\mathbf{p}_R$. Only those measurements within the error ellipse of each leg (arising from uncertainty in the state) can form an association; here $\mathbf{z}_2$ can be associated with either leg. Rather than choosing a particular association, MHT generates hypotheses for all possibilities, including those in which $\mathbf{z}_1$ and $\mathbf{z}_2$ are false alarms. Each hypothesis is then allowed to evolve independently and give rise to new hypotheses in subsequent measurement cycles. Figure 3 illustrates three generations of a so-called *hypothesis tree*. The *association event* $\theta^{k,l}$ at leaf l denotes a particular set of associations between z^k and measurement sources. In people tracking, uncertainty also arises in the dynamic state: whether the left/right leg is moving, or the person reversed direction (see Section 3). Thus, in addition to the measurement associations the tracker must hypothesize about the dynamic state, as indicated by the *dynamic state event* $\psi^{k,l}$ in Figure 3.

A particular history of association and dynamic state hypotheses leading up to leaf l in the hypothesis tree (for example, the bold line in Figure 3) is described by the *cumulative hypotheses* $\Theta^{k,l}$ and $\Psi^{k,l}$. Furthermore, the history of measurements over all generations is denoted by the cumulative measurement set Z^k. Each path through the hypothesis tree also has an associated extended Kalman filter (EKF) to estimate the state $\mathbf{x}^{k,l}$ and state covariance $\mathrm{P}^{k,l}$ resulting from the history of measurements and dynamic states. The basic task in multiple hypothesis tracking is then to evaluate the probability of each cumulative hypothesis conditioned on the cumulative measurements, written as $P(\Theta^{k,l}, \Psi^{k,l} | Z^k)$. The current state is then chosen as the one corresponding to the most likely cumulative hypothesis. Generation and evaluation of new hypotheses in each measurement cycle is detailed in Section 4. MHT provides good tracking robustness by allowing the system to recover from association/dynamic state errors by choosing an alternative history in each cycle, rather than the single-tracked history of alternative techniques.

3 Walking Person Model

Our walking model is based on the simple physical assumptions that the legs extend in equal and opposite directions from the torso and at least one leg always remains stationary. Figure 4 gives a top view of the walking model. As described earlier, only the location of left and right legs ($\mathbf{p}_L$ and $\mathbf{p}_R$ in Figure 4) are observable from the calf-height slice provided

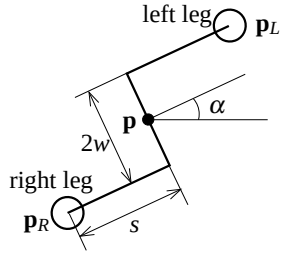


Figure 4: Walking model parameters.

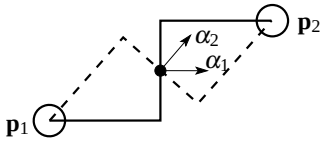


Figure 5: Alternative solutions for walking model orientation.

by the range sensor. The parameters of the model are:

- $\mathbf{p}$=$(x,y)^\top$ = torso centre of mass
- α = walking direction relative to x-axis of sensor
- s = projected leg length, positive for right leg forward
- $2w$ = width of hips, assumed fixed and known

Given measurements $\mathbf{p}_L = (x_L, y_L)^\top$ and $\mathbf{p}_R = (x_R, y_R)^\top$ for both legs, the unknown parameters of the model are given by

$$\mathbf{p} = \frac{1}{2}(\mathbf{p}_L + \mathbf{p}_R) \quad (1)$$

$$\alpha = \text{atan}\left(\frac{y_L - y_R}{x_L - x_R}\right) \pm \text{asin}\left(\frac{w}{\sqrt{|\mathbf{p}_L - \mathbf{p}_R|}}\right) \quad (2)$$

$$s = \pm\frac{1}{2}\sqrt{|\mathbf{p}_L - \mathbf{p}_R|^2 - 4w^2} \quad (3)$$

As indicated by equation (2), the leg measurements give rise to two possible orientations. The physical interpretation of these solutions are shown by the solid and dashed lines in Figure 3 (note also that the interpretation of $\mathbf{p}_1$ and $\mathbf{p}_2$ as the left and right legs switch between solutions). During walking, people often transition between these configurations when turning around to walk backwards (particularly when a person is being followed by a robot). Thus, both configurations must be considered when generating hypotheses about a tracked target (see Section 4.1).

The inverse of the above equations is the measurement model used by the EKF to predict the position of the legs from the estimated model parameters. The measurement model is

$$\mathbf{p}_{L,R} = \mathbf{p} + \mathrm{R}(\alpha) \cdot (\pm s, \mp w)^\top \quad (4)$$

where the positive sign is taken for R and the negative for L, and $\mathrm{R}(\theta)$ is a 2×2 rotation matrix for angle θ. If $\mathbf{z}_L$ and $\mathbf{z}_R$ are the measured location of both legs, the observation passed to the EKF is the combined measurement $\mathbf{y} = (\mathbf{z}_L^\top, \mathbf{z}_R^\top)^\top$.

To track a walking person, the parameters in Figure 4 are augmented with the linear velocity v and angular velocity ω, and the complete state vector is therefore $\mathbf{x} = (x, y, \alpha, s, v, \omega)^\top$. The evolution of the state from time step $k-1$ to k is described by one of two possible dynamic models: the left leg swings while the right leg remains stationary, or the right leg swings while the left leg remains stationary (a stationary person can be described by either model with zero velocity). In either case, the torso is assumed to undergo constant velocity motion. The position $\mathbf{p}^k$ at time step k therefore evolves from the previous state $\mathbf{x}^{k-1}$ according to

$$\mathbf{p}^k = \mathbf{p}_{L,R}^{k-1} + \mathrm{R}(\omega^{k-1}\Delta_t)[\mathbf{p}^{k-1} + v\Delta_t \mathrm{R}(\alpha^{k-1})\hat{\mathbf{x}} - \mathbf{p}_{L,R}^{k-1}] \quad (5)$$

where Δ_t is the sample period, $\hat{\mathbf{x}}$ is a unit vector in the x-direction and $\mathbf{p}_{L,R}$ is the location of the stationary leg, given by equation (4). Equation (5) describes a linear displacement by $v\Delta_t$ in the direction α, followed by a rotation about the stationary leg by angle $\omega\Delta_t$. To preserve the location of the stationary leg, the projected leg length must increase or decrease by $v\Delta_t$ according to

$$s^k = s^{k-1} \pm v^{k-1}\Delta_t \quad (6)$$

where the positive sign describes a stationary left leg, and negative describes a stationary right leg. The remaining parameters evolve according to a constant velocity model.

4 Multiple Hypothesis People Tracking

4.1 Hypothesis Generation

The purpose of hypothesis generation is to enumerate all likely associations of new measurements to the predicted state for all dynamic models. This section describes hypothesis generation for a single parent node of the hypothesis tree, and the process is repeated for all parent nodes.

As discussed above, a walking person switches between two possible dynamic models depending on the moving leg, and between the two configurations shown in Figure 3 when reversing direction. Thus, given the estimated state at the previous time step, four possible predicted states must be considered. Let the predicted states be described by state vector ${}^j\mathbf{x}^k$ and covariance matrix ${}^j\mathrm{P}^k$, where $1 \le j \le 4$. The state and covariance corresponding to the left and right moving legs ($j = 1, 2$) are predicted using the EKF with the dynamic models in equations (5)-(6). Then, the alternative configurations in Figure 3 ($j = 3, 4$) are generated by recalculating the orientation, leg length and velocity as

$${}^{j+2}\alpha^k = {}^j\alpha^k + \pi - 2\tan(w/{}^j s^k) \quad (7)$$

$${}^{j+2}s^k = -({}^j s^k) \quad (8)$$

$${}^{j+2}v^k = -({}^j v^k) \quad (9)$$

The next step is to find all associations between predicted leg positions, ${}^j\mathbf{p}_L^k$ and ${}^j\mathbf{p}_R^k$ (from equation (4)), and measurements in z^k. Using the validation gate procedure illustrated

in Figure 2, measurement $\mathbf{z}_i^k$ is associated with the leg at ${}^j\mathbf{p}_\nu^k$ ($\nu = L, R$) when the Mahalanobis distance satisfies

$$(\mathbf{z}_i^k - {}^j\mathbf{p}_\nu^k)^\top ({}^j\mathrm{S}_\nu^k)^{-1} (\mathbf{z}_i^k - {}^j\mathbf{p}_\nu^k) < d_{th}^2 \tag{10}$$

for threshold d_{th}^2. The 4×4 innovation covariance matrix ${}^j\mathrm{S}^k$ of measurement vector $\mathbf{y}^k = (\mathbf{z}_L^\top, \mathbf{z}_R^\top)^\top$ is calculated from the predicted state covariance ${}^j\mathrm{P}^k$, measurement model Jacobian M and measurement error covariance R as

$${}^j\mathrm{S}^k = \mathrm{M}\,{}^j\mathrm{P}^k\mathrm{M}^\top + \mathrm{R} = \begin{pmatrix} {}^j\mathrm{S}_L^k & {}^j\mathrm{S}_{LR}^k \\ {}^j\mathrm{S}_{LR}^k & {}^j\mathrm{S}_R^k \end{pmatrix} \tag{11}$$

To evaluate equation (10), the 2×2 submatrix ${}^j\mathrm{S}_L^k$ or ${}^j\mathrm{S}_R^k$ is selected from ${}^j\mathrm{S}^k$ depending on the leg associated with $\mathbf{z}_i^k$.

Measurements that do not fall within any validation gate are simply discarded, since hypotheses associated with these measurements would have negligible probability. The result is a reduced set of $m \leq n$ *validated* measurements in z^k. A *hypothesis matrix* Ω^k is then constructed in which row i indicates the possible source of $\mathbf{z}_i^k$ in the reduced set. For example, the hypothesis matrix for the scenario in Figure 2 is

$$\Omega^k = \begin{array}{c} \begin{matrix} F & L & R \end{matrix} \\ \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array} \tag{12}$$

Finally, the association hypotheses $\theta^{k,l}$ are generated from Ω^k. To make physical sense, hypothesis $\theta^{k,l}$ associates each measurement with only one source (F, L or R) and allows each source (except F) to produce only one measurement. Hypotheses are therefore generated by selecting one valid source from each row of Ω^k such that each leg appears no more than once. For example, Figure 2 gives rise to five association hypotheses: (F,F), (F,L), (F,R), (R,F) and (R,L). The final set of hypotheses $(\theta^{k,l}, \psi^{k,l})$ is obtained by repeating the process for each possible dynamic model.

4.2 Probability Evaluation

Let $P(\Theta^{k,l}, \Psi^{k,l}|Z^k)$ represent the probability of the cumulative hypothesis $(\Theta^{k,l}, \Psi^{k,l})$ for leaf l in the hypothesis tree, conditional on the measurement history Z^k. The probability calculation presented below applies to all leaves, and index l is henceforth dropped for convenience. Our evaluation of $P(\Theta^k, \Psi^k|Z^k)$ follows the general framework outlined in [Bar-Shalom and Fortmann, 1988]. Using Bayes rule, the conditional probability of the hypothesis can be factored as

$$P(\Theta^k, \Psi^k|Z^k) = \frac{1}{c} p(z^k|\Theta^k, \Psi^k, Z^{k-1}) P(\theta^k|\Theta^{k-1}, Z^{k-1}) \cdot P(\psi^k|\Psi^{k-1}, Z^{k-1}) P(\Theta^{k-1}, \Psi^{k-1}|Z^{k-1}) \tag{13}$$

where the normalization constant c is summed over all leaves. The terms on the right hand side of equation (13) represent the measurement likelihood, conditional probability of the current measurement association, conditional probability of the current dynamic model and the prior probability of the parent hypothesis. Each factor will now be evaluated in turn.

Measurements associated with the left or right leg in θ^k are assumed to be independent with normally distributed error (as required by the EKF). False alarms in θ^k are assumed to be independent and uniformly distributed across the N range samples with a probability of N^{-1} (an approximation since false alarms are actually often persistent). The measurement likelihood can thus be written as the product

$$p(z^k|\Theta^k, \Psi^k, Z^{k-1}) = N^{-\phi} |2\pi \mathrm{S}^k|^{-1} e^{-\frac{1}{2}(\mathbf{y}^k - \hat{\mathbf{y}}^k)^\top (\mathrm{S}^k)^{-1} (\mathbf{y}^k - \hat{\mathbf{y}}^k)} \tag{14}$$

where ϕ is the number of false alarms, $\hat{\mathbf{y}}^k$ is the predicted measurement vector from equation (4), and S^k is the innovation covariance. When measurements are assigned to both legs in θ^k, equation (14) is evaluated using the measurement vector $\mathbf{y} = (\mathbf{z}_L^\top, \mathbf{z}_R^\top)^\top$ and the 4×4 innovation covariance in equation (11). When θ^k involves only one leg, the measurement vector is $\mathbf{y}^k = \mathbf{z}_\nu^k$ and the innovation covariance is the submatrix S_ν^k in equation (11), where $\nu = L$ or R.

To calculate the conditional probability of the association event, $P(\theta^k|\Theta^{k-1}, Z^{k-1})$, the detection indicators δ_L and δ_R are introduced, where $\delta_i = 1$ if the associated leg is detected in θ^k and $\delta_i = 0$ otherwise. Then, dropping the irrelevant conditionings, the probability of event θ^k can be written as

$$P(\theta^k|\Theta^{k-1}, Z^{k-1}) = P(\theta^k|\delta_L, \delta_R, \phi) P(\delta_L, \delta_R, \phi) \tag{15}$$

In evaluating $P(\theta^k|\delta_L, \delta_R, \phi)$, all association events that give rise to detection indicators δ_L and δ_R with ϕ false alarms are assumed to be equally likely. The number of such events is the number of permutations of $(m - \phi)$ detected legs out of m measurements (permutations since the legs have identity), and the probability of each event is

$$P(\theta^k|\delta_L, \delta_R, \phi) = \left[\frac{m!}{(m - (m - \phi))!} \right]^{-1} = \frac{\phi!}{m!} \tag{16}$$

False alarms are assumed to arise independently of the legs. Thus the second term in equation (15) can be factorized as $P(\delta_L, \delta_R, \phi) = P(\delta_L, \delta_R) P(\phi)$. We follow the convention of modelling false alarms as a Poisson process with fixed density λ over all N laser samples (ie. an average of λN false measurements are expected per scan). Thus, the probability of ϕ false alarms is $P(\phi) = [(\lambda N)^\phi / \phi!] e^{-\lambda N}$.

In conventional MHT, the detection of established targets (events δ_L and δ_R) is assumed to be independent with fixed probability, P_D. This is a good approximation for sparse targets, but breaks down for a walking person as one leg occludes the other with almost every step. However, occlusions can be predicted and should be included in the evaluation of the hypothesis to improve tracking robustness against false measurements. For example, consider the case of $\mathbf{p}_L$ occluded by $\mathbf{p}_R$ in Figure 2: a hypothesis that does *not* assign $\mathbf{z}_2$ to $\mathbf{p}_L$ should have a higher probability than one that does.

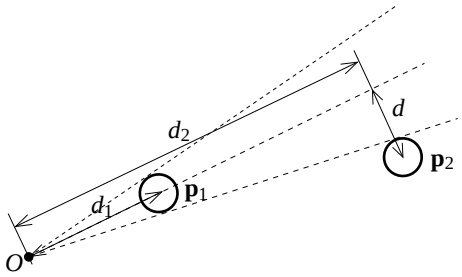


Figure 6: Occlusion model parameters.

Figure 6 illustrates the model used for predicting occlusions, where $\mathbf{p}_1$ and $\mathbf{p}_2$ are the leg measurements predicted by equation (4) and O is the world origin (location of the range scanner). Assuming the legs have fixed radius r, the closer leg casts the occluding shadow indicated by the dotted lines. Let $d_1 = |\mathbf{p}_1|$ and $d_2 = (\mathbf{p}_1^\top \mathbf{p}_2)\mathbf{p}_1/|\mathbf{p}_1|$ represent the distance from the laser to the front and rear leg along the central axis of the occlusion region. The distance of the rear leg from the axis is $d = (|\mathbf{p}_2|^2 - d_2^2)^{1/2}$, and occlusions occur when d falls below a threshold D_{th} given by

$$D_{th}(\mathbf{p}_1, \mathbf{p}_2) = r(d_1 + d_2)/d_1 \tag{17}$$

Figure 6 shows the rear leg just on the verge of this threshold. The probability $P_D(\mathbf{p}_1, \mathbf{p}_2)$ of detecting the rear leg is assumed to decrease gradually as the occlusion becomes more severe, as approximated by the piecewise linear model

$$P_D(\mathbf{p}_1, \mathbf{p}_2) = \begin{cases} P_O + (P_V - P_O)D/D_{th} & \text{if } d < D_{th} \\ P_V & \text{if } d \geq D_{th} \end{cases} \tag{18}$$

where P_V is the detection probability when $\mathbf{p}_2$ is outside the occlusion shadow and P_O is the detection probability when $\mathbf{p}_2$ lies on the shadow axis (typically $P_V = 0.9$ and $P_O = 0.2$). The detection probability for the front leg at $\mathbf{p}_1$ is simply $P_D = P_V$, and the probability of the combined event $\delta = (\delta_L, \delta_R)$ is the product of the detection probability for both legs.

Finally, combining equation (16) with the Poisson distribution for the false alarms and the detection probability above, the conditional probability of the association event in equation (15) can be written as

$$P(\theta^k|\delta,\phi)P(\delta,\phi) = \frac{(\lambda N)^\phi}{m!} e^{-\lambda N} \prod_{i=L,R} (P_D)^{\delta_i}(1-P_D)^{1-\delta_i} \tag{19}$$

For tracking filters with switched dynamic models, the conditional probability $P(\psi^k|\Psi^{k-1}, Z^{k-1})$ is usually assumed to be Markovian [Isard and Blake, 1998]. In this case, the conditional probability reduces to $P(\psi^k|\psi^{k-1})$, which can be implemented as a two dimensional transition probability table. This assumption breaks down in the case of walking, which exhibits periodic state transitions and short-term correlations between sequential dynamic states. We exploit these properties to improve tracking robustness (provided the walking pattern is not too irregular) by modelling the dynamic state transitions as a non-stationary process. Let t be the time since the transition to the current dynamic model ψ^k. When t is small the probability of finding the same dynamic state is high (short-term correlation), but decays as t increases until a transition to another state is likely. We model this effect as an exponentially decaying/increasing conditional probability

$$P(\psi^k|\Psi^{k-1}) = \begin{cases} 0.25 + (P_I - 0.25)e^{-t/\tau} & \text{if } \psi^k = \psi^{k-1} \\ 0.25 - \frac{1}{3}(P_I - 0.25)e^{-t/\tau} & \text{if } \psi^k \neq \psi^{k-1} \end{cases} \tag{20}$$

where τ is a time constant modelling the gait period, and P_I determines the initial probability after a transition.

The final term in equation (13) is the prior probability $P(\Theta^{k-1}, \Psi^{k-1}|Z^{k-1})$, which is the probability of the parent hypothesis. Finally, the probability of the new hypothesis $P(\Theta^k, \Psi^k|Z^k)$ is calculated as the product of the prior with the conditional probabilities in equations (14), (19) and (20).

4.3 Hypothesis Reduction

In any MHT filter, persistent pruning of the hypothesis tree is necessary to satisfy real-time constraints. However, care must also be taken to ensure that sufficient variation remains to adequately approximate the distribution of possible states. Our tracking filter employs a combination of common hypothesis reduction techniques to achieve these goals, and result in about 10-50 hypotheses maintained per generation.

N-scan-back pruning [Kurien, 1990] dramatically culls hypotheses by discarding entire branches attached to a single node at the $(k-N)$th generation of the hypothesis tree. The probability of each branch is evaluated by tallying the probabilities of descended leaves, and only the most likely branch is retained. Using this method, the hypothesis tree always provides N generations of variation and contains only a single history above the $(k-N)$th node. Clearly, N is carefully chosen to provide sufficient variation while keeping the hypothesis tree manageable, and $N = 4$ was empirically found to provide good results for our walking person tracker.

Due to the high number of hypotheses descending from each node, a second culling stage is applied to remove low probability leaves. A low probability test is usually implemented as a fixed threshold [Cox and Leonard, 1994], but this approach was found to be too indiscriminant. Instead, we cull hypotheses that have a probability less than a fixed fraction (say 0.05) of the best hypothesis, which retains the most likely hypotheses regardless of the absolute probability.

Finally, the hypothesis tree is likely to contain several hypotheses with similar states, particularly when the person is stationary. Hypotheses are therefore removed if a more likely hypothesis with the same dynamic model is found within a validation gate defined by the state covariance. Let $\mathbf{x}_l$ and P_l represent the state and covariance of the lower probability hypothesis and $\mathbf{x}_h$ represent the state of the higher probability hypothesis. Then, the lower probability hypothesis is

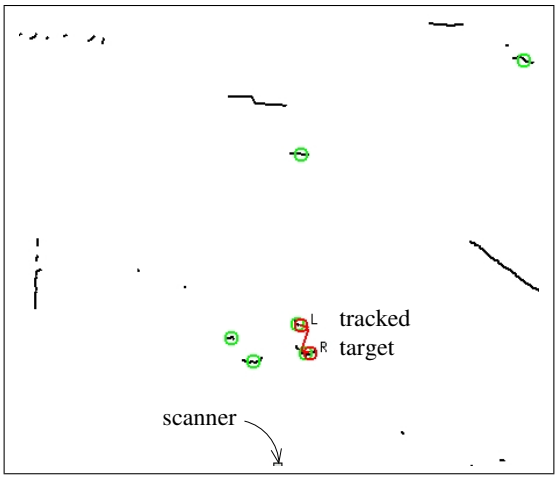


Figure 7: Typical tracking result showing raw range data (black), leg measurements (green) and estimated state (red).

removed if the Mahalanobis distance between states satisfies

$$(\mathbf{x}_l - \mathbf{x}_h)^\top \mathrm{P}_l^{-1} (\mathbf{x}_l - \mathbf{x}_h) < d_{th}^2 \qquad (21)$$

where $d_{th}^2 = 1$ is used in practice. The probability of the eliminated hypothesis is added to the surviving hypothesis.

4.4 Track Initiation and Termination

The current implementation tracks a single person at a time, and new targets are initiated and terminated automatically. When no target currently exists, a new hypothesis tree is initiated when two leg candidates are separated by less than a fixed threshold. The initial state is calculated using equations (1)-(3), along with zero velocity and a large state covariance. Either solution of equations (2)-(3) may be chosen for the new target, since the other configuration will be considered in the next generation. An existing track is terminated if no valid measurements are found for a threshold number of generations, or the probability of the best hypothesis is sufficiently small (ie. the target is unlikely to be a walking person).

5 Results

The multiple hypothesis walking person tracker presented in this paper has been applied to a number of experimental sequences. A sequence of 1000 scans involving two people standing still, walking, crossing paths, meeting, changing direction and moving in and out of the field of view is shown in Video 1. In this video sequence, raw range points are shown in black (depth discontinuities removed), leg candidates are shown in green and the most likely hypothesis is shown in red or blue to indicate transitions between dynamic states. Dynamic model switches are observed to occur frequently as the tracker switches between alternative histories, particularly when the tracked person is stationary. Importantly, one or the other person is successfully tracked during the entire sequence, with the MHT only losing track when a person leaves the field of view. Figure 7 gives the tracking result for a typical scan in this sequence, while the people briefly meet.

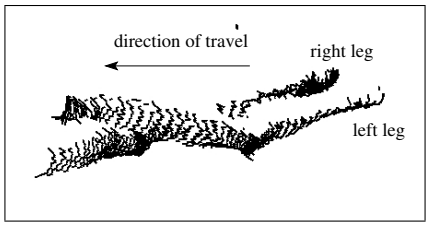


Figure 8: Range profiles of legs for turning event.

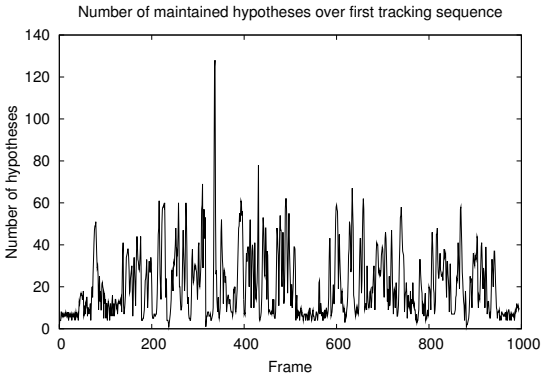


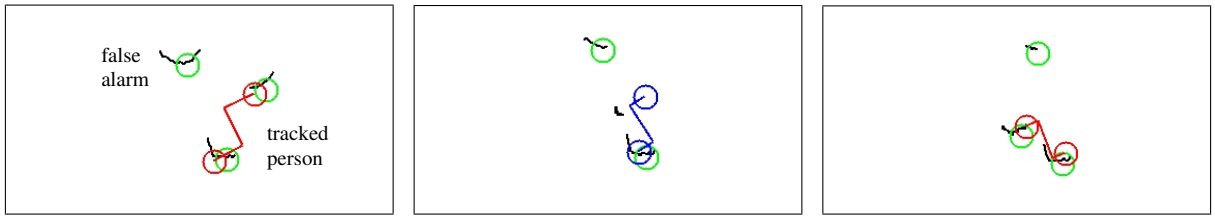
Figure 10: Number of maintained leaves after each scan.

Two challenging sequences in this experiment are highlighted in Figures 8 and 9. The sequence of laser profiles in Figure 8 show the path of the legs as the person switches between the walking configurations in Figure 3 (around scan number 650). In this case, the left leg crosses over the right leg as the person turns around to walk backwards. The state of the walking person is successfully maintain during this manoeuvre as can be seen in Video 1. Figure 9 shows representative scans before, during and after occlusion of the right leg by the left. Again, the state of the target is successfully maintained despite the presence of a false measurement (the leg of the second person) in the same neighbourhood.

The number of leaves at each new generation of the hypothesis tree (after applying the pruning in Section 4.3) is shown in Figure 10, which verifies that the tracker operates reliably with about 10-50 hypotheses per generation. The required processing time in this case is 5-25 ms per scan (depending on the number of hypotheses) on a 3 Ghz Pentium 4 desktop PC, which is well within the constraints of real-time tracking.

6 Summary and Future Work

This paper has addressed the problem of tracking a person using range measurements at leg height. Unlike previous range-based people trackers, our framework explicitly models walking kinematics and dynamics to evaluate the probability that the target is truly a walking person. A switched



(a) Before occlusion event. (b) During occlusion event. (c) After occlusion event.

Figure 9: Successful recovery of tracking after occlusion in the presence of a false alarm.

dynamic model allows for the possibility of a moving left or right leg (with the other stationary), and for changing configuration when turning on the spot. Our tracker employs a multiple hypothesis framework with several novel features: both measurement associations and dynamic model transitions are included in the hypothesis tree, dynamic model transitions are treated as a non-stationary process, and the effect of occlusion is modelled in the evaluation of association probabilities. Experimental results highlight the usefulness of this additional modelling, and demonstrate the robustness and efficiency of our tracker in real scenes. It would be interesting in future work to compare the efficiency of our approach with alternative multiple hypothesis frameworks such as particle filtering.

While our method efficiently tracks a single person, tracking multiple people introduces several new problems. In particular, multiple people cannot be tracked independently as interactions arise when one person occludes another or validation gates overlap. These interactions dramatically increase the number of hypotheses and additional pruning techniques are likely to be required, such as splitting and merging independent hypothesis trees for different groups of targets.

The ultimate extension of this work will be to identify tracked individuals, possibly by applying a Bayesian learning model to the observed dynamic state sequence. Real-time learning would allow individuals to be robustly re-acquired after extended occlusions, or identified after grouping together and subsequently dispersing. Such a system would have widespread applications in service robotics and surveillance, and is the subject of continuing research.

Acknowledgment

This work was supported by the ARC funded Centre for Perceptive and Intelligent Machines in Complex Environments.

References

[Bar-Shalom and Fortmann, 1988] Y. Bar-Shalom and Thomas E. Fortmann. *Tracking and Data Association*. Academic Press, San Deigo, 1988.

[Brooks and Williams, 2003] A. Brooks and S. Williams. Tracking people with networks of hetrogeneous sensors. In *Proc. 2003 Australasian Conf. on Robotics and Automation*, pages 1–7, 2003.

[Cielniak and Duckett, 2003] G. Cielniak and T. Duckett. Person identification by mobile robots in indoor environments. In *1st Int. Workshop on Robotic Sensing*, pages 1–5, 2003.

[Cox and Leonard, 1994] I. J. Cox and J. J. Leonard. Modeling a dynamic environment using a bayesean multiple hypothesis approach. *Artificial Intelligence*, 66:311–344, 1994.

[Fod *et al.*, 2002] A. Fod, A. Howard, and M. J. Mataric. A laser-based people tracker. In *Proc. 2002 IEEE Int. Conf. on Robotics Automation*, pages 3024–3029, 2002.

[Isard and Blake, 1998] M. Isard and A. Blake. A mixed-state CONDENSATION tracker with automatic model-switching. In *Sixth International Conference on Computer Vision*, pages 107–112, 1998.

[Kurien, 1990] T. Kurien. Issues in the design of practical multitarget tracking algorithms. In Y. Bar-Shalom, editor, *Multitarget-Multisensor Tracking: Advanced Applications*, pages 43–83. Artech House, 1990.

[Montemerlo *et al.*, 2002] M. Montemerlo, S. Thrun, and W. Whittaker. Conditional particle filters with simultaneous mobile robot localization and people-tracking. In *Proc. 2002 IEEE Int. Conf. on Robotics Automation*, pages 695–701, 2002.

[Prassler *et al.*, 1999] E. Prassler, J. Scholz, and A. Elfes. Tracking people in a railway station during rush-hour. In *Proc. 1st Int. Conf. on Computer Vision Systems*, pages 162–179, 1999.

[Reid, 1979] D. B. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, AC-24(6):843–854, 1979.

[Schulz *et al.*, 2001] D. Schulz, W. Burgand, D. Fox, and A. B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In *Proc. 2001 IEEE Int. Conf. on Robotics Automation*, pages 1665–1670, 2001.