Proceedings of the IEEE ICRA 2009
Workshop on People Detection and Tracking
Kobe, Japan, May 2009

Multi-model Hypothesis Group Tracking and Group Size Estimation

Boris Lau Kai O. Arras Wolfram Burgard

***Abstract*— People in densely populated environments typically form groups that split and merge. In this paper we track groups of people so as to reflect this formation process and gain efficiency in situations where maintaining the state of individual people would be intractable. We pose the group tracking problem as a recursive multi-hypothesis model selection problem in which we hypothesize over both, the partitioning of tracks into groups (models) and the association of observations to tracks (assignments). Model hypotheses that include split, merge, and continuation events are first generated in a data-driven manner and then validated by means of the assignment probabilities conditioned on the respective model. Observations are found by clustering points from a laser range finder given a background model and associated to existing group tracks using the minimum average Hausdorff distance. We further propose a method to estimate the number of people in groups based on the number of human-sized clusters. Experiments with a stationary and a moving platform show that, in populated environments, tracking groups is clearly more efficient than tracking people separately. The results also show a high accuracy in the estimation of group sizes. Our system runs in real-time on a typical desktop computer.**

I. Introduction

The ability of robots to keep track of people in their surrounding is fundamental for a wide range of applications including personal and service robots, intelligent cars, or surveillance. People are social beings and as such they form groups, interact with each other, merge to larger groups or separate from groups. Tracking individual people during these formation processes can be hard due to the high chance of occlusion and the large extent of data association ambiguity. This causes the space of possible associations to become huge and the number of assignment histories to quickly become intractable. Further, for many applications, knowledge about groups can be sufficient as the task does not require to know the state of every person. In such situations, tracking groups that consist of multiple people is more efficient and furthermore contains semantic information about activities of the people.

This paper focuses on group tracking in populated environments with the goal to track a large number of people in real-time. The approach attempts to maintain the state of groups of people over time, considering possible splits and merges as illustrated in Fig. 1. For our experiments we use a mobile robot equipped with a laser range finder, but our method should be applicable to data from other sensors as well.

All authors are with the University of Freiburg, Germany, Department of Computer Science {lau,arras,burgard}@informatik.uni-freiburg.de.

This work was partly funded by the European Commusion under contract number FP6-IST-045388

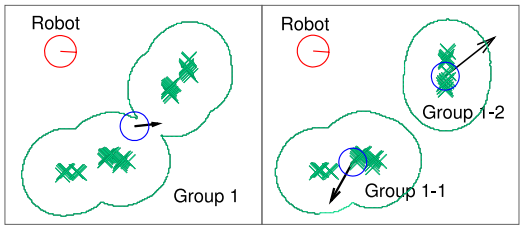

Fig. 1. Tracking groups of people with a mobile robot. Groups are shown by their position (blue), velocity (black), the associated laser points (green) and a contour for visualization. In the two frames, a group of four people splits up into two groups with two people each.

In most related work on laser-based people tracking, tracks correspond to individual people [1], [2], [3], [4], [5]. In Taylor *et al.* [6] and Arras *et al.* [7], tracks represent the state of legs which are fused to people tracks in a later stage. Khan *et al.* [8] proposed an MCMC-based tracker that is able to deal with non-unique assignments, i.e., measurements that originate from multiple tracks, and multiple measurements that originate from the same track. Actual tracking of groups using laser range data was, to our knowledge, first addressed by Mucientes *et al.* [9]. Most research in group tracking was carried out in the vision community [10], [11], [12]. Gennari *et al.* [11] and Bose *et al.* [12] both address the problem of target fragmentation (splits) and grouping (merges). They do not integrate data association decisions over time – a key property of the Multi-Hypothesis Tracking (MHT) approach, initially presented by Reid [13] and later extended by Cox *et al.* [14]. The approach belongs to the most general data association techniques as it produces joint compatible assignments, integrates them over time, and is able to deal with track creation, confirmation, occlusion, and deletion.

The works closest to this paper are Mucientes *et al.* [9] and Joo *et al.* [15]. Both address the problem of group tracking using an MHT approach. Mucientes *et al.* employ two separate MHTs, one for the regular association problem between observations and tracks and a second stage MHT that hypothesizes over group merges. However, people tracks are not replaced by group tracks, hence there is no gain in efficiency. The main benefit of that approach is the semantical extra information about formation of groups.

Joo *et al.* [15] present a visual group tracker using a single MHT to create hypotheses of group splits and merges and observation-to-track assignments. They develop an interesting variant of Murty's algorithm [16] that generates the k-best *non-unique* assignments which enables them to make multiple assignments between observations and tracks, thereby describing target splits and merges. However, the method only produces an approximation of the optimal k-

best solutions since the posterior hypothesis probabilities depend on the number of splits, which, at the time when the k-best assignments are being generated, is unknown. In our approach, the split, merge and continuation events are given by the model *before* computing the assignment probabilities, and therefore, our k-best solutions are optimal.

In this paper we propose a tracking system for groups of people using an extended Multi-Hypothesis Tracking (MHT) approach to hypothesize over both, the group formation process (models) and the association of observations to tracks (assignments). Each model, defined to be a particular partitioning of tracks into groups, creates a new tree branch with its own assignment problem. As a further contribution we propose a group representation that includes the shape of the group and we show how this representation is updated in each step of the tracking cycle. This extends previous approaches where groups are assumed to have Gaussian shapes only [11], [9]. We also present an estimation method to determine the number of people in groups which extends the approach presented by the same authors in [17]. Finally, we use the psychologically motivated *proxemics* theory introduced by Hall [18] for the definition of a group. The theory relates social relation and body spacing during social interaction.

It is structured as follows: the following section describes the extraction of groups of people from laser range data. Section III introduces the definition of groups. Section V briefly describes the cycle of our Kalman filter-based tracker. Section VI explains the data-driven generation of models and how their probabilities are computed. Whereas Section VII presents the multi-model MHT formulation and derives expressions for the hypothesis probabilities, Section VIII describes the experimental results.

II. Group Detection in Range Data

Detecting people in range data has been approached with motion and shape features [1], [2], [3], [4], [5], [9] as well as with a learned classifier using boosted features [19]. However, these recognition systems were designed (or trained) to extract single people. In the case of densely populated environments, groups of people typically produce large blobs in which individuals are hard to recognize. We therefore pursue the approach of background subtraction and clustering. Given a previously learned model (a map of the environment for mobile platforms), the background is subtracted from the scans and the remaining points are passed to the clustering algorithm. This approach is also able to detect standing people as opposed to [9] which relies on motion features.

Concretely, a laser scanner generates measurements $\mathbf{z}_i = (\phi_i, \rho_i)^T,\ \ i \in \{1, \dots N_z\}$, with ϕ_i being the bearing and ρ_i the range value. The measurements $\mathbf{z}_i$ are transformed into Cartesian coordinates and grouped using *single linkage clustering* [20] with a distance threshold d_P. The outcome is a set of clusters $\mathcal{Z}_i$ making up the current observation $Z(k) = \{\mathcal{Z}_i \,|\, i = 1, \dots, N_\mathcal{Z}\}$. Each cluster $\mathcal{Z}_i$ is a complete set of measurements $\mathbf{z}_i$ that fulfills the cluster condition, i.e., two clusters are joined if the distance between their closest points is smaller than d_P. A similar concept, using a connected components formulation, has been used by Gennari and Hager [11]. The clusters then contain range readings that can correspond to single legs, individual people, or groups of people, depending on the cluster distance d_P.

III. Group Definition

This section defines the concept of a group and derives probabilities of group-to-observation and group-to-group assignments.

What makes a collection of people a *group* is a highly complex question in general which involves difficult-to-measure social relations among subjects. A concept related to this question is the proxemics theory introduced by Hall [18] who found from a series of psychological experiments that social relations among people are reliably correlated with physical distance during interaction. This finding allows us to infer group affiliations by means of body spacing information available in the range data. The distance d_P thereby becomes a threshold with a meaning in the context of group formation.

A. Representation of Groups

Concretely, we represent a group as a tuple $G = \langle \mathbf{x}, C, \mathcal{P} \rangle$ with $\mathbf{x}$ as the track state, C the state covariance matrix and $\mathcal{P}$ the set of contour points that belong to G. The track state is composed of the position $(x,\, y)$ and the velocities $(\dot{x},\, \dot{y})$ to form the state vector $\mathbf{x} = (x,\, y,\, \dot{x},\, \dot{y})^T$ of the group.

The points $\mathbf{x}_{\mathcal{P}_i} \in \mathcal{P}$ are an approximation of the group's current shape or spatial extension. Shape information will be used for data association under the assumption of *instantaneous rigidity*. That is, a group is assumed to be a rigid object over the duration of a time step Δt, and consequently, all points in $\mathcal{P}$ move coherently with the estimated group state $\mathbf{x}$. The points $\mathbf{x}_{\mathcal{P}_i}$ are represented relative to the state $\mathbf{x}$.

B. Group-to-Observation Assignment Probability

For data association we need to calculate the probability that an observed cluster $\mathcal{Z}_i$ belongs to a predicted group $G_j = \langle\, \mathbf{x}_j(k+1|k),\, C_j(k+1|k),\, \mathcal{P}_j \,\rangle$. A distance function $d(\mathcal{Z}_i, G_j)$ is sought that, unlike the Mahalanobis distance used by Mucientes *et al.* [9], accounts for the shape of the observation cluster $\mathcal{Z}_i$ and the group's contour $\mathcal{P}_j$, rather than just for their centroids. To this end, we use a variant of the Hausdorff distance. As the regular Hausdorff distance is the *longest* distance between points on two contours, it tends to be sensitive to large variations in depth that can occur in range data. This motivates the use of the minimum average Hausdorff distance [21] that computes the minimum of the averaged distances between contour points,

$$d_{\text{HD}}(\mathcal{Z}_i, G_j) = \min\left\{d(\mathcal{Z}_i, \mathcal{P}_j), d(\mathcal{P}_j, \mathcal{Z}_i)\right\} \qquad (1)$$

where $d(\mathcal{Z}_i, \mathcal{P}_j)$ is the directed average Hausdorff distance. Since we deal with uncertain entities, $d(\mathcal{Z}_i, \mathcal{P}_j)$ is calculated using the squared Mahalanobis distance $d^2 = \nu^T\, S^{-1}\, \nu$,

$$d(\mathcal{Z}_i, \mathcal{P}_j) = \frac{1}{|\mathcal{Z}_i|} \sum_{\mathbf{z}_i \in\, \mathcal{Z}_i} \min_{\mathbf{x}_{\mathcal{P}_j} \in\, \mathcal{P}_j} \left\{d^2(\nu_{ij},\, S_{ij})\right\}, \qquad (2)$$

with ν_{ij}, S_{ij} being the innovation and innovation covariance between a point $\mathbf{z}_i \in \mathcal{Z}_i$ and contour point $\mathbf{x}_{\mathcal{P}_j}$ of the predicted set $\mathcal{P}_j$ transformed into the sensor frame,

$$\nu_{ij} = \mathbf{z}_i - (H\mathbf{x}_j(k+1|k) + \mathbf{x}_{\mathcal{P}_j}) \quad (3)$$

$$S_{ij} = H\,C_j(k+1|k)\,H^T + R_i \quad (4)$$

where $H = \left(\begin{smallmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{smallmatrix}\right)$ is the measurement Jacobian and R_j the 2×2 observation covariance whose entries reflect the noise in the measurement process of the range finder.

The probability that cluster $\mathcal{Z}_i$ originates from G_j is finally

$$\mathcal{N}_i := \mathcal{N}(d^2_{\text{HD}}(\mathcal{Z}_i, G_j), S_{ij}) \quad (5)$$

where $\mathcal{N}(\mu, \Sigma)$ denotes the normal distribution.

C. Group-to-Group Assignment Probability

To determine the probability that two groups G_i and G_j merge, we compute the distance between their closest contour points in a Mahalanobis sense. In doing so, we have to account for the clustering distance d_P that states identity of G_i, G_j as soon as their contours come closer than d_P. Let $\Delta\mathbf{x}_{\mathcal{P}_{ij}} = \mathbf{x}_{\mathcal{P}_i} - \mathbf{x}_{\mathcal{P}_j}$ be the vector difference of two contour points of G_i and G_j respectively, we then subtract d_P from $\Delta\mathbf{x}_{\mathcal{P}_{ij}}$ unless $\Delta\mathbf{x}_{\mathcal{P}_{ij}} \leq d_P$ for which $\Delta\mathbf{x}_{\mathcal{P}_{ij}} = 0$. Concretely, the modified difference becomes $\Delta\mathbf{x}'_{\mathcal{P}_{ij}} = \max(0,\ \Delta\mathbf{x}_{\mathcal{P}_{ij}} - d_P\,\mathbf{u}_{\mathcal{P}_{ij}})$ where $\mathbf{u}_{\mathcal{P}_{ij}} = \Delta\mathbf{x}_{\mathcal{P}_{ij}} / |\Delta\mathbf{x}_{\mathcal{P}_{ij}}|$.

In order to obtain a similarity measure that accounts for nearness of group contours *and* similar velocity, we augment $\Delta\mathbf{x}'_{\mathcal{P}_{ij}}$ by the difference in the velocity components, $\Delta\mathbf{x}^*_{\mathcal{P}_{ij}} = (\Delta\mathbf{x}'^T_{\mathcal{P}_{ij}},\ \dot{x}_i - \dot{x}_j\,,\ \dot{y}_i - \dot{y}_j)^T$. Statistical compatibility of two groups G_i and G_j can now be determined with the (four-dimensional) minimum Mahalanobis distance

$$d^2_{\text{min}}(G_i, G_j) = \min_{\mathbf{x}_{\mathcal{P}_i} \in \mathcal{P}_i,\, \mathbf{x}_{\mathcal{P}_j} \in \mathcal{P}_j} \left\{ d^2(\Delta\mathbf{x}^*_{\mathcal{P}_{ij}},\, C_i + C_j) \right\}.$$

The probability that two groups actually belong together, is finally given by $\mathcal{N}_{ij} := \mathcal{N}\left(d^2_{\text{min}}(G_i, G_j),\, C_i + C_j\right)$.

IV. Estimating the Number of People in Groups

As described above, our group tracking approach considers the joint state of groups rather than the states of the individuals that form the groups. However, knowing the number of people in a group is interesting information, e.g., for interaction, data association or motion planning. We therefore augment the state vector of group tracks by a fifth state variable, n_s, the group size. A group state is then the vector $\mathbf{x} = (x,\, y,\, \dot{x},\, \dot{y},\, n_s)^T$.

As an observation of the group size, we take the number of human-sized clusters in the set of contour points $\mathcal{P}$ of a group track G. Reapplying single-linkage clustering with a cluster distance of $d_P = 0.3\,m$ yields groups of points that are likely to correspond to human individuals.

For state prediction and in case of a track confirmation event, we assume, analogous to the constant velocity motion model, constant group size. Noise in the motion model accounts for people joining or leaving the group without being noticed. If two tracks are merged, the resulting size estimate is the sum of the sizes of the joining groups. The variances simply sum up, as we assume independent size estimates across groups. If two tracks are split, we split the group size in half and increase the variance to account for uneven splits.

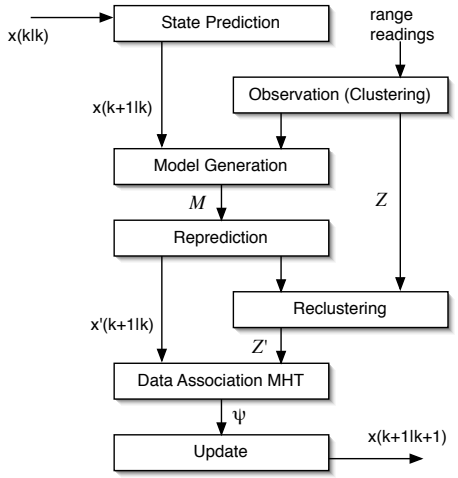

Fig. 2. Flow diagram of the tracking system. See explanations in section V.

V. Tracking Cycle

This section describes the steps in the cycle of our Kalman filter-based group tracker. An overview is given by the flow diagram in Fig. 2. The structure differs from a regular tracker in the additional steps *model generation*, *track reprediction* and *reclustering*.

- *State prediction:* The state prediction of a group track based on the previous posterior estimates $\mathbf{x}(k|k)$, $C(k|k)$ is given by $\mathbf{x}(k+1|k) = A\ \mathbf{x}(k|k)$ and $C(k+1|k) = A\,C(k|k)\,A^T + Q$, where A is the state transition matrix for a constant velocity motion model and Q the 4×4 process noise covariance matrix whose entries reflect the acceleration capabilities of a typical human. The set of contour points $\mathcal{P}$ is now relative to $\mathbf{x}(k+1|k)$.
- *Observation:* As described in section II, this step involves grouping the laser range data into clusters $\mathcal{Z}$.
- *Model Generation:* Models are generated based on the predicted group tracks and the clusters $\mathcal{Z}$, see section VI.
- *Reprediction:* Based on the model hypotheses that postulate a split, merge or continuation event for each track, groups are repredicted so as to reflect the respective model:

  If a model hypothesis contains a split of a group, two new groups are created by duplicating its predicted state. The same applies for the set $\mathcal{P}$.

  If a model hypothesis contains a merge of two groups G_i, G_j, the repredicted group state $\mathbf{x}_{ij}$, C_{ij} is computed as the multivariate weighted average (omitting $(k+1|k)$),

$$C_{ij}^{-1} = C_i^{-1} + C_j^{-1}$$

$$\mathbf{x}_{ij} = C_{ij}\,(C_i^{-1}\mathbf{x}_i + C_j^{-1}\mathbf{x}_j)\,. \quad (6)$$

  The set of contour points of the merged group is the union of the two former point sets, $\mathcal{P}_{ij} = \mathcal{P}_i \cup \mathcal{P}_j$.

- *Reclustering:* Reclustering an observed cluster $\mathcal{Z}_i$ is necessary when it has been produced by more than one group track, that is, it is in the gate of more than one track. If the model hypothesis postulates a merge for the involved tracks, nothing needs to be done. Otherwise, $\mathcal{Z}_i$ needs to be reclustered, which is done using a nearest-neighbor rule: those points $\mathbf{z}_i \in \mathcal{Z}_i$ that share the same nearest neighbor track are combined in a new cluster. This step follows from the uniqueness assumption – common in target tracking – which says that a target can only produce a single observation.
- *Data Association MHT:* This step involves the generation, probability calculation, and pruning of data association hypotheses that assign repredicted group tracks to reclustered observations. See section VII.
- *Update:* A group track G_j that has been assigned to a cluster $\mathcal{Z}_i$ is updated with a standard linear Kalman filter using the centroid position $\bar{\mathbf{z}}_{\mathcal{Z}_i}$ of $\mathcal{Z}_i$. The contour points in $\mathcal{P}_j$ are replaced by the points in $\mathcal{Z}_i$, transformed into the reference frame of the posterior state $\mathbf{x}(k+1|k+1)$. Thereby, $\mathcal{P}_j$ contains always the group's most actual shape approximation.

VI. Model Generation and Model Probability

A model is defined to be a partitioning of tracks into groups. It assumes a particular state of the group formation process. New models, whose generation is described in this section, hypothesize about the evolution of that state.

The space of possible model transitions is large since each group track can split into an unknown number of new tracks, or merge with an unknown number of other tracks. We therefore bound the possible number of model transitions by the assumption that merge and split are binary operators. We further impose the gating condition for observations and tracks using the minimum average Hausdorff distance, thereby implementing a data-driven aspect into the model generation step. Concretely, we assume:

- A track G_i can split at most into two tracks in one frame provided two compatible observations with G_i.
- At most two group tracks G_i, G_j can merge into one track at the same time but only if there is an observation which is statistically compatible with G_i and G_j.
- A group track can only split into tracks that are both matched in that very time step. Splits into occluded or obsolete tracks are not allowed.
- A group track can not be involved in a split and a merge action at the same time.

Gating and statistical compatibility are both determined on a significance level α. The limitation to binary operators is justified by the realistic assumption that we observe the world much faster than the rate with which it evolves. Even if, for instance, a group splits into three subgroups at once, the tracker requires only two cycles to reflect this change.

A new model now defines for each group track if it is continued, split or if it merges with another group track. The probability of a model is calculated using constant prior probabilities for continuations and splits, p_C and p_S respectively, and the probability for a merge between two tracks G_i and G_j as $p_G \cdot \mathcal{N}_{ij}$. The latter term consists of a constant prior probability p_G and the group-to-group assignment probability $\mathcal{N}_{ij}$ defined in section III-C. Let N_C and N_S be the number of continued tracks and the number of split tracks in model M respectively, then the probability of M conditioned on the parent hypothesis Ω^{k-1} is

$$P(M|\Omega^{k-1}) = p_C^{N_C} \cdot p_S^{N_S} \prod_{G_i,G_j \in \Omega^{k-1}} (p_G \cdot \mathcal{N}_{ij})^{\delta_{ij}} \quad (7)$$

with δ_{ij} being 1 if G_i, G_j merge and 0 otherwise.

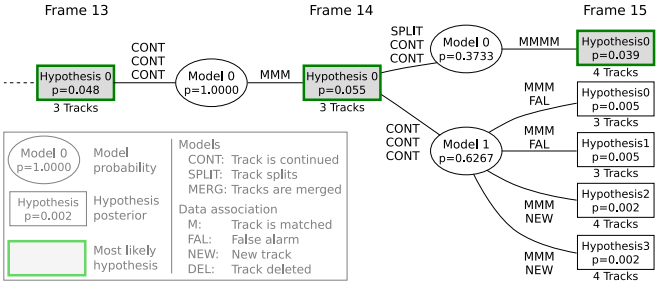

Fig. 3. The multi-model MHT. For each parent hypothesis, model hypotheses (ellipses) branch out and create their own assignment problems. In our application, models define which tracks of the parent hypothesis are continued, split or merge. The tree shows frames 13 to 15 of figure 4. The split of group 1 between frames 14 and 15 is the most probable hypothesis following model branch 0. See the legend for details.

VII. Multi-Model MHT

In this section we describe our extension of the original MHT by Reid [13] to a multi-model tracking approach that hypothesizes over both, data associations and models.

Let Ω_i^k be the i-th hypothesis at time k and $\Omega_{p(i)}^{k-1}$ its parent. Let further $\psi_i(k)$ denote a set of assignments which associates predicted tracks in $\Omega_{p(i)}^{k-1}$ to observations in $Z(k)$. As there are many possible assignment sets given $\Omega_{p(i)}^{k-1}$ and $Z(k)$, there are many children that can branch off a parent hypothesis, each with a different $\psi(k)$. This makes up an exponentially growing hypothesis tree.

The multi-model MHT introduces an intermediate tree level for each time step, on which models spring off from parent hypotheses (Fig. 3). In each model branch, the tracks of the parent hypothesis are first repredicted to implement that particular model and then assigned to the (reclustered) observations. Possible assignments for observations are *matches* with existing tracks, *false alarms* or *new tracks*. Using the generalized formulation of Arras *et al.* [7] to deal with more than two track interpretation labels, tracks are interpreted as *matched*, *obsolete* or *occluded*.

A. Probability Calculations

The probability of a hypothesis in the multi-model MHT is calculated as follows. According to the Markov assumption, the probability of a child hypothesis Ω_i^k given the observations from all time steps up to k, denoted by Z^k, is the joint probability of the assignment set $\psi_i(k)$, the model M

and the parent hypothesis $\Omega_{p(i)}^{k-1}$, conditioned on the current observation $Z(k)$. Using Bayes rule, this can be expressed as the product of the data likelihood with the joint probability of assignment set, model and parent hypothesis,

$$\begin{aligned} P(\Omega_i^k|Z^k) &= P(\psi, M, \Omega_{p(i)}^{k-1}|Z(k)) \quad (8) \\ &= \eta \cdot P(Z(k)|\psi, M, \Omega_{p(i)}^{k-1}) \cdot P(\psi, M, \Omega_{p(i)}^{k-1}). \end{aligned}$$

By using conditional probabilities, the third term on the right hand side can be factorized into the probabilities of the assignment set, the model and the parent hypothesis,

$$P(\psi, M, \Omega_{p(i)}^{k-1}) = P(\psi|M, \Omega_{p(i)}^{k-1}) \cdot P(M|\Omega_{p(i)}^{k-1}) \cdot P(\Omega_{p(i)}^{k-1}).$$

The last term is known from the previous iteration while the second term was derived in section VI.

The first term is the probability of the assignment set ψ. The set ψ contains the assignments of observed clusters $\mathcal{Z}_i$ and group tracks G_j either to each other or to one of their possible labels listed above. Assuming independence between observations and tracks, the probability of the assignment set is the product of the individual assignment probabilities. They are: p_M for matched tracks, p_F for false alarms, p_N for new tracks, p_O for tracks found to be occluded and p_T for obsolete tracks scheduled for termination. If the number of new tracks and false alarms follow a Poisson distribution (as assumed by Reid [13]), the probabilities p_F and p_N have a sound physical interpretation as $p_F = \lambda_F V$ and $p_N = \lambda_N V$ where λ_F and λ_N are the average rates of events per volume multiplied by the observation volume V (the sensor's field of view). The probability for an assignment ψ, given a model M and a parent hypothesis Ω^{k-1} is then computed by

$$P(\psi|M, \Omega^{k-1}) = p_M^{N_M}\ p_O^{N_O}\ p_T^{N_T}\ \lambda_F^{N_F} \lambda_N^{N_N}\ V^{N_F+N_N}, \quad (9)$$

where the Ns are the number of assignments in ψ to the respective labels.

Thanks to the independence assumption, also the data likelihood $P(Z(k)|\,\psi, M, \Omega_{p(i)}^{k-1})$ is computed by the product of the individual likelihoods of each observation cluster $\mathcal{Z}_i$ in $Z(k)$. If ψ assigns an observation $\mathcal{Z}_i$ to an existing track, we assume the likelihood of $\mathcal{Z}_i$ to follow a normal distribution, given by Eq. 5. Observations that are interpreted as false alarms and new tracks are assumed to be uniformly distributed over the observation volume V, yielding a likelihood of $1/V$. The data likelihood then becomes

$$P(Z(k)|\psi, M, \Omega^{k-1}) = \left(\tfrac{1}{V}\right)^{N_N+N_F} \prod_{i=1}^{N_{\mathcal{Z}}} \mathcal{N}_i^{\delta_i}, \quad (10)$$

where δ_i is 1 if $\mathcal{Z}_i$ has been assigned to an existing track, and 0 otherwise.

Substitution of Eqs. (7), (9), and (10) into Eq. (8) leads, like in the original MHT approach, to a compact expression, independent on the observation volume V.

Finally, normalization is performed yielding a true probability distribution over the child hypotheses of the current time step. This distribution is used to determine the current best hypothesis and to guide the pruning strategies.

TABLE I

SUMMARY OF THE DATA USED IN THE TWO EXPERIMENTS.

	Experiment 1	Experiment 2
Number of frames	578	991
Avg. / max people	6.25 / 13	8.99 / 20
Avg. / max groups	2.60 / 4	4.16 / 8
Number of splits / merges	5 / 10	48 / 44
Number of new tracks / deletions	19 / 15	34 / 39

B. Pruning

Pruning is essential in implementations of the MHT algorithm, as otherwise the number of hypotheses grows boundless. The following strategies are employed:

K-best branching: instead of creating all children of a parent hypothesis, the algorithm proposed by Murty [16] generates only the K most probably hypotheses in polynomial time. We use the multi-parent variant of Murty's algorithm, mentioned in [22], that generates the global K best hypotheses for all parents.

Ratio pruning: a lower limit on the ratio of the current and the best hypothesis is defined. Unlikely hypotheses with respect to the best one, being below this threshold, are deleted. Ratio pruning overrides K-best branching in the sense that if the lower limit is reached earlier, less than K hypotheses are generated.

N-scan back: the N-scan-back algorithm considers an ancestor hypothesis at time $k-N$ and looks ahead in time onto all children at the current time k (the leaf nodes). It keeps only the subtree at $k-N$ with the highest sum of leaf node probabilites, all other branches at $k-N$ are discarded.

VIII. EXPERIMENTS

To analyze the performance of our system, we collected two data sets in a large entrance hall of a university building. We used a Pioneer II robot equipped with a SICK laser scanner mounted at 30 cm above floor, scanning at 10 fps. In two unscripted experiments (experiment 1 with a stationary robot, experiment 2 with a moving robot), up to 20 people are in the sensor's field of view. They form a large variety of groups during social interaction, move around, stand together and jointly enter and leave the hall (see Tab. I).

To obtain ground truth information, we labeled each single range reading. Beams that belong to a person receive a person-specific label, other beams are labeled as non-person. These labels are kept consistent over the entire duration of the data sets. People that socially interact with each other (derived by observation) are said to belong into a group with a group-specific label. Summed over all frames, the ground truth contains 5629 labeled groups and 12524 labeled people.

The ground truth data is used for performance evaluation and to learn the parameter probabilities of our tracker. The values, determined by counting, are $p_M = 0.79$, $p_O = 0.19$, $p_T = 0.02$, $p_F = 0.06$, $p_N = 0.02$ for the data association probabilities, and $p_C = 0.63$, $p_S = 0.16$, $p_G = 0.21$ for the group formation probabilities. When evaluating the performance of the tracker, we separated the data into a training set and a validation set to avoid overfitting.

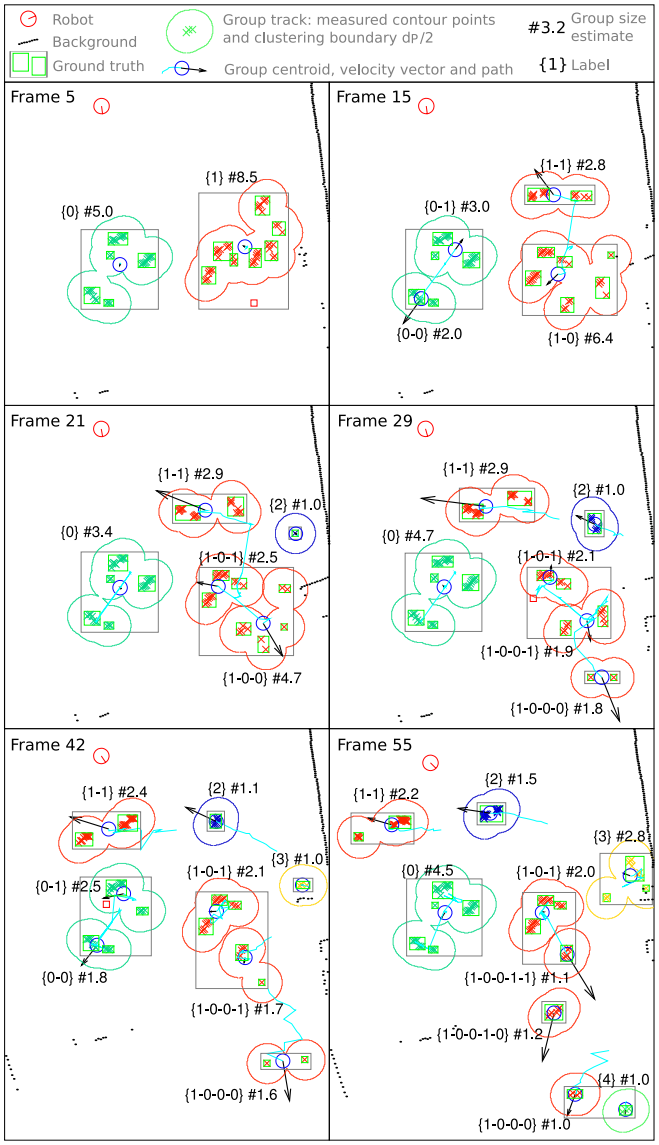

Fig. 4. Tracking results from experiment 2. In frame 5, two groups are present. In frame 15, the tracker has correctly split group 1 into 1-0 and 1-1 (see Fig. 3). Between frames 15 and 29, group 1-0 has split up into groups 1-0-0 and 1-0-1, and split up again. New groups, labeled 2 and 3, enter the field of view in frames 21 and 42 respectively.

Six frames of the current best hypothesis from experiment 2 are shown in Fig. 4, the corresponding hypothesis tree is shown in Fig. 3. The sequence exemplifies movement and formation of several groups.

A. Clustering Error

Given the ground truth information on a per-beam basis we can compute the clustering error of the tracker. This is done by counting how often a track's set of points $\mathcal{P}$ contains too many or wrong points (undersegmentation) and how often $\mathcal{P}$ is missing points (oversegmentation) compared to the ground truth. Two examples for oversegmentation errors can be seen in Fig. 4, where group 0 and group 1-0 are temporarily oversegmented. However, from the history of group splits and merges stored in the group labels, the correct group relations can be determined in such cases.

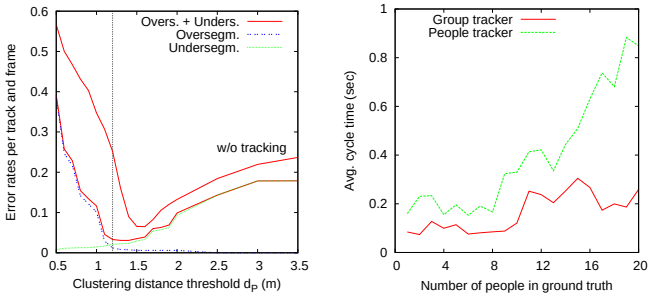

Fig. 5. Left: clustering error of the group tracker compared to a memory-less single linkage clustering (without tracking). The smallest error is achieved for a cluster distance of 1.3 m which is very close to the border of personal and social space according to the proxemics theory, marked at 1.2 m by the vertical line. Right: average cycle time for the group tracker versus a tracker for individual people plotted against the ground truth number of people.

For experiment 1, the resulting percentages of incorrectly clustered tracks for the cases undersegmentation, oversegmentation and the sum of both are shown in Fig. 5 (left), plotted against the clustering distance d_P. The figure also shows the error of a single-linkage clustering of the range data as described in section II. This implements a memory-less group clustering approach against which we compare the clustering performance of our group tracker.

The minimum clustering error of 3.1% is achieved by the tracker at $d_P = 1.3\,m$. The minimum error for the memory-less clustering is 7.0%, more than twice as high. In the more complex experiment 2, the minimum clustering error of the tracker rises to 9.6% while the error of the memory-less clustering reaches 20.2%. The result shows that the group tracking problem is a *recursive* clustering problem that requires integration of information over time. This occurs when two groups approach each other and pass from opposite directions. The memory-less approach would merge them immediately while the tracking approach, accounting for the velocity information, correctly keeps the groups apart.

In the light of the proxemics theory the result of a minimal clustering error at 1.3 m is noteworthy. The theory predicts that when people interact with friends, they maintain a range of distances between 45 to 120 cm called personal space. When engaged in interaction with strangers, this distance is larger. As our data contains students who tend to know each other well, the result appears consistent with Hall's findings.

B. Tracking Efficiency

When tracking groups of people rather than individuals, the assignment problems in the data association stage are of course smaller. On the other hand, the introduction of an additional tree level on which different models hypothesize over different group formation processes comes with additional computational costs. We therefore compare our system with a person-only tracker which is implemented by inhibiting all split and merge operations and reducing the cluster distance d_P to the very value that yields the lowest error for clustering single people given the ground truth. For

experiment 2, the resulting average cycle times versus the ground truth number of people is shown in Fig. 5 (right). The plots are averaged over different k from the range of 2 to 200 at a scan-back depth of $N = 30$.

With an increasing number of people, the cycle time for the people tracker grows much faster than the cycle time of the group tracker. Interestingly, even for small numbers of people the group tracker is faster than the people tracker. This is due to occasional oversegmentation of people into individual legs tracks. Also, as mutual occlusion of people in densely populated environments occurs often, the people tracker has a lot more occluded tracks to maintain than the group tracker, as occlusion of entire groups is rare. Also, the additional complexity of multiple models in the group tracker virtually disappears when the tracks are isolated due to the data-driven model generation.

This result clearly shows that the claim of higher efficiency holds for this group tracking approach. With an average cycle time of around 100 ms for up to 10 people on a Pentium IV at 3.2 GHz, the algorithm runs in real-time even with a non-optimized implementation.

C. Group Size Estimation

To evaluate the accuracy of our group size estimation approach, we define the error as the absolute difference between the estimated number of people in a group and the true value according to the labeled ground truth.

In experiment 1, we find that the average error is 0.23 people with a standard deviation of 0.30. In the more complex experiment 2, the average error is 0.33 people with a standard deviation of 0.49. If the estimated group sizes are rounded to integers, the tracker determined the correct value in 88.9% of all cases in experiment 1 and in 84.3% for experiment 2.

If only deviations of more than one person are considered an error, the system was correct in 99.5% of all cases in experiment 1 and 97.5% in experiment 2.

IX. Conclusion

In this paper, we presented a multi-model hypothesis tracking approach to track groups of people. We extended the original MHT approach to incorporate model hypotheses that describe track interaction events that go beyond what data association can express. In our application, models encode the formation of groups during split, merge, and continuation events. We further introduced a representation of groups that includes their shape, and employed the minimum average Hausdorff distance to account for the shape information when calculating association probabilities.

The proposed tracker has been implemented and tested using a mobile robot equipped with a laser range finder. It is able to robustly track groups of people as they undergo complex formation processes. Given ground truth data with over 12,000 labeled occurrences of people and groups, the experiments showed that the tracker could reproduce such processes with a low clustering error and very accurate estimates of the number of people in groups.

Further experiments carried out from a stationary and a moving platform in populated environments with up to 20 people demonstrated that tracking groups of people is clearly more efficient than tracking individual people. They also showed that our system performs significantly better than a memory-less single-frame clustering which underlines the recursive character of this model selection problem.

References

[1] B. Kluge, C. Köhler, and E. Prassler, "Fast and robust tracking of multiple moving objects with a laser range finder," in *Proceedings of the IEEE Int. Conf. on Robotics and Automation*, 2001.

[2] A. Fod, A. Howard, and M. J. Mataric, "Laser-based people tracking," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Washington DC, May 2002, pp. 3024–3029.

[3] D. Schulz, W. Burgard, D. Fox, and A. Cremers, "People tracking with a mobile robot using sample-based joint probabilistic data association filters," *Intl. J. of Robotics Research (IJRR)*, vol. 22, no. 2, 2003.

[4] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Tracking multiple people using laser and vision," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Alberta, Canada, 2005.

[5] W. Zajdel, Z. Zivkovic, and B. Kröse, "Keeping track of humans: Have I seen this person before?" in *IEEE International Conference on Robotics and Automation*, Barcelona, Spain, 2005.

[6] 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*, Canberra, Australia, 2004.

[7] 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 *IEEE International Conference on Robotics and Automation (ICRA)*, Pasadena, CA, USA, May 2008.

[8] Z. Khan, T. Balch, and F. Dellaert, "MCMC data association and sparse factorization updating for real time multitarget tracking with merged and multiple measurements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, December 2006.

[9] M. Mucientes and W. Burgard, "Multiple hypothesis tracking of clusters of people," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2006, pp. 692–697.

[10] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Computer Vision and Image Understanding*, vol. 80, no. 1, pp. 42–56, October 2000.

[11] G. Gennari and G. D. Hager, "Probabilistic data association medhods in visual tracking of groups," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[12] B. Bose, X. Wang, and E. Grimson, "Multi-class object tracking algorithm that handles fragmentation and grouping," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.

[13] D. B. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. on Automatic Control*, vol. AC-24, no. 6, pp. 843–854, 1979.

[14] I. Cox and S. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 138–150, February 1996.

[15] S.-W. Joo and R. Chellappa, "A multiple-hypothesis approach for multiobject visual tracking," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2849–2854, November 2007.

[16] K. Murty, "An algorithm for ranking all the assignments in order of increasing cost," *Operations Research*, vol. 16, pp. 682–687, 1968.

[17] B. Lau, K. O. Arras, and W. Burgard, "Tracking groups of people with a multi-model hypothesis tracker," in *International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, May 2009.

[18] E. Hall, *Handbook of Proxemics Research*. Society for the Anthropology of Visual Communications, 1974.

[19] K. O. Arras, Óscar Martínez Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA'07)*, Rome, Italy, 2007.

[20] J. Hartigan, *Clustering Algorithms*. John Wiley & Sons, 1975.

[21] M. P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Intl. Conference on Pattern Recognition*, vol. 1, Jerusalem, Israel, 1994, pp. A:566–568.

[22] I. Cox and M. Miller, "On finding ranked assignments with application to multi-target tracking and motion correspondence," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 486–489, 1995.