

Accurate Human Motion Capture in Large Areas by Combining IMU- and Laser-Based People Tracking

Jakob Ziegler Henrik Kretschmar Cyrill Stachniss Giorgio Grisetti Wolfram Burgard

Abstract—A large number of applications use motion capture systems to track the location and the body posture of people. For instance, the movie industry captures actors to animate virtual characters that perform stunts. Today’s tracking systems either operate with statically mounted cameras and thus can be used in confined areas only or rely on inertial sensors that allow for free and large-scale motion but suffer from drift in the pose estimate. This paper presents a novel tracking approach that aims to provide globally aligned full body posture estimates by combining a mobile robot and an inertial motion capture system. In our approach, a mobile robot equipped with a laser scanner is used to anchor the pose estimates of a person given a map of the environment. It uses a particle filter to globally localize a person wearing a motion capture suit and to robustly track the person’s position. To obtain a smooth and globally aligned trajectory of the person, we solve a least squares optimization problem formulated from the motion capture suite and tracking data. Our approach has been implemented on a real robot and exhaustively tested. As the experimental evaluation shows, our system is able to provide locally precise and globally aligned estimates of the person’s full body posture.

I. INTRODUCTION

The movie industry and creators of computer games are major markets for professional motion capture systems. Marker-based visual tracking approaches that use multiple statically mounted cameras, such as the Vicon mocap studio [18], belong to the most commonly used systems. These tracking systems are highly accurate but restrict the actor to a rather small area that can be captured by the cameras. To overcome this limitation, motion capture systems based on inertial measurement units (IMUs), such as the Xsens MVN motion capture suit [20], have been developed. These inertial motion capture suits attach multiple IMUs to a person and provide an estimate of the person’s full body posture. They can accurately track the person’s pose locally but suffer from global drift particularly when the person moves over large distances. As this drift accumulates over time due to the incremental nature of the IMU-based filter, interactions between the person and the environment itself are difficult to realize. Consequently, applications that require a locally accurate as well as a globally correctly aligned estimate of the person’s posture are still limited to rather small scenes.

The contribution of this paper is a novel approach that bridges this gap by combining locally accurate full body posture estimates with a precise global alignment of the

All authors are with the University of Freiburg, Department of Computer Science, Freiburg, Germany. G. Grisetti is also with Sapienza University of Rome, Department of Systems and Computer Science, Rome, Italy.

This work has been supported by the German Research Foundation (DFG) under contract number SFB/TR-8 and by Microsoft Research, Redmond. Their support is gratefully acknowledged.

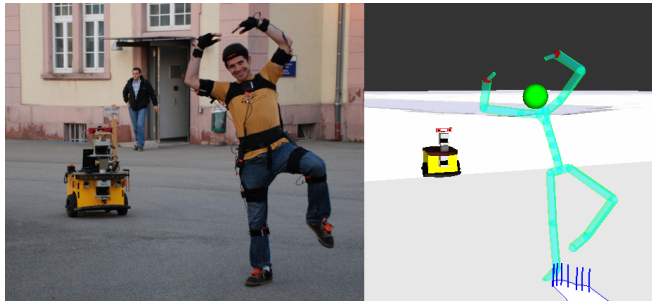


Fig. 1: Globally aligned full body posture estimate of a person obtained using an inertial motion capture system and a mobile robot equipped with a laser range finder.

tracked person with respect to the environment. Our approach combines the estimates of an inertial motion capture suit with particle filter-based people tracking techniques using a mobile robot equipped with a laser range finder (see Fig. 1). To obtain accurate trajectory estimates, our method computes the solution of a global least squares problem consisting of the estimates of both the inertial motion capture suit and the laser-based tracker. As a result, our approach is able to compute a globally accurate alignment of the smooth track of the person in the scene including the full body posture estimate of the person from the inertial motion capture suit.

II. RELATED WORK

Tracking people in the environment is a well studied problem in robotics and related disciplines. Approaches to people tracking in robotics can be classified according to the used sensor, e.g., laser scanners [1], [7], monocular cameras [3], stereo cameras [6], IMUs [16], or combinations of these sensors [4], [10], [12], [19], [21].

Key techniques found in most people tracking approaches are Kalman filters [5], [7], [8], particle filters [15], [19], and HMMs [2]. To tackle the data association problem, approaches rely on nearest neighbor techniques [15], the Hungarian method [3], multi-hypothesis data association trackers [1], probabilistic data association filters [12], or its joint variant called JPDAFs [2], [17].

Montemerlo *et al.* [15] present an approach to simultaneously localize a mobile robot and people in its vicinity. This method is related to our system in so far as we also estimate the trajectory of the robot and the person jointly. In contrast to Montemerlo *et al.*, we do not only track but also smooth the output to obtain a more accurate maximum likelihood solution. However, we do not consider that the

laser beams reflected by a person may negatively influence the pose estimate of the robot itself.

Grzonka *et al.* [9] present an approach which is able to reconstruct the trajectory of a person wearing an inertial motion capture suit along with localized events such as opening doors to estimate the path of the person as well as a topological-metric map.

The approach presented in this paper is based on a mobile robot that observes the tracked person with a laser-range finder and a mobile motion capture system based on IMUs, such as the Xsens MVN suit. Our method is able to provide globally consistent and locally accurate full-body motion estimates of a person, even over large distances. We use a combination of particle filter-based localization, people tracking, and global optimization to obtain the most likely motion path. In addition to that, our approach can deal with situations in which the person being tracked is temporarily outside the field of view of the robot.

III. APPROACH TO PARTICLE FILTER-BASED PERSON TRACKING USING AN INERTIAL MOTION CAPTURE SUIT

In our approach, we use a particle filter to track the pose x_t of a person. We estimate the posterior

$$p(x_t | z_{1:t}, u_{1:t}) \propto \int_{x'} p(x_t | x', u_t) p(x' | z_{1:t-1}, u_{1:t-1}) dx', \quad (1)$$

where u_t describes the movement of the person, and z_t corresponds to an observation of the person at time t . The particle filter estimates this posterior using the motion model $p(x_t | x_{t-1}, u_t)$ and the observation model $p(z_t | x_t)$. In our approach, the motion model is based on the estimate provided by the inertial motion capture suit, and the observation model is based on the laser measurements obtained by the robot.

A. Motion Model

An estimate of the motion of the person to be tracked is provided by the inertial motion capture suit. In our current implementation, the Xsens MVN suit estimates the motion of the person's full body posture using a set of IMUs. Consequently, the estimate of the inertial motion capture suit is affected by drift. Although the relative motion estimate is globally inaccurate, it is typically locally precise and therefore well-suited for the motion model.

The estimates provided by the inertial motion capture suit are with respect to its own reference frame. The motion $u_t = (T, \Delta\theta)^\top$ consists of a translation T and a change $\Delta\theta$ in orientation.

The particle filter algorithm samples a new pose for each particle according to the motion model based on u_t and its estimate of the current state. Assuming Gaussian noise in T and $\Delta\theta$, the sampled motion for a sample i is given by

$$\Delta\theta^i = \Delta\theta + \varepsilon_{\text{rot}}, \text{ where } \varepsilon_{\text{rot}} \sim \mathcal{N}(0, \alpha_1|T| + \alpha_2|\Delta\theta|) \quad (2)$$

$$T^i = T + \varepsilon_{\text{trans}}, \text{ where } \varepsilon_{\text{trans}} \sim \mathcal{N}(0, \alpha_3|T| + \alpha_4|\Delta\theta|). \quad (3)$$

The noise parameters $\alpha_1, \alpha_2, \alpha_3$, and α_4 determine the uncertainty with respect to translation and rotation. Finally, the new pose for sample i is computed as $x_t^i = x_{t-1}^i \oplus (T^i, \Delta\theta^i)^\top$.

B. Observation Model

The particle filter uses the observation model $p(z_t | x_t)$ to correct the prediction estimated by the motion model. In our approach, the observation model is based on a (localized) mobile robot equipped with a laser range finder.

To obtain $p(z_t | x_t)$, we compute the expected observation of the tracked person given its current estimated pose. We do so by constructing a 3D skeleton model of the person using the current estimate of the person's full body posture, which is provided by the inertial motion capture suit. The intersection of the skeleton and the plane of the laser range finder indicates where the legs of the person are expected to appear in the scan.

To evaluate how well the expected observation matches the actual observation, we detect potential legs in the current laser range scan. We consider the distances between the expected locations of the legs and the locations of the detected legs. Let $p(l_l) = p(l_r)$ be the probability of the detector identifying the legs of the person to track, which we assume to be independent from the state x_t and the measurement z_t . This probability can be easily learned from a labeled dataset. Since the data associations are unknown, we integrate/sum over all possible associations. Assuming Gaussian noise, we have

$$\begin{aligned} p(z_t | x_t) &\propto p(-l_l)p(-l_r) \\ &+ p(l_l)p(l_r) \sum_{l_l \neq l_r \in L} \exp(-\lambda(\|l_l^* - l_l\|^2 + \|l_r^* - l_r\|^2)) \\ &+ p(l_l)p(-l_r) \sum_{l_l \in L} \exp(-\lambda\|l_l^* - l_l\|^2) \\ &+ p(-l_l)p(l_r) \sum_{l_r \in L} \exp(-\lambda\|l_r^* - l_r\|^2), \end{aligned} \quad (4)$$

where l_l^* and l_r^* are the positions of the left and the right leg of the skeleton, according to the current particle, and L refers to the set of legs detected in the current range scan. Finally, λ is a scaling parameter that determines how peaked the distribution is.

Our approach applies a technique similar to that proposed by Kluge [11] to detect legs in the laser range scans. We group the laser endpoints into segments based on the Euclidean distance between the endpoints of the laser beams and subsequently examine the width and shape of each segment. If the width of a segment approximately corresponds to that of a human leg and the shape roughly resembles a half circle, the segment is classified as a leg candidate.

Note that for evaluating a particle it is more important to obtain few false-negatives (i.e., missing a leg) rather than a non-zero number of false positives (i.e., legs of other persons or similarly looking objects). This is similar to the observation likelihood in robot localization, which often is a multi-modal distribution over the environment. Hence, this relatively naive heuristic for leg extraction turned out to be sufficient in practice.

To update the weight w^i of each sample i we apply

$$w_t^i = \eta p(z_t | x_t^i) w_{t-1}^i, \quad (5)$$

where η is a normalization constant computed by summing up the individual likelihoods. Our approach carries out

the resampling step selectively by considering the effective number of samples, similar to Liu [14].

C. Globally Localizing a Person

A common approach to global localization is to uniformly sample poses from the entire state space, which is rather inefficient. An alternative method is to add samples to the particle filter such that the probability of adding a particular sample is proportional to its observation likelihood, similar to Lenser and Veloso [13]. In our application, however, such a technique is likely to make the particle filter diverge since the robot may observe a different person while the person that is supposed to be tracked is outside the robot’s field of view.

To mitigate this problem, we require that a potential pose be consistent with the measurements for some time before we incorporate it into the particle filter. For this purpose, we sample from the distribution

$$x_t \propto p(x_t \mid z_{t-k:t}, u_{t-k:t}), \quad (6)$$

taking into account the last k measurement updates and motion predictions. In this way, our algorithm prevents samples that correspond to a different person and that move differently from being added to the particle filter as the person to be tracked. To efficiently sample from (6), we run an additional particle filter at the corresponding locations and only incorporate the samples into the main particle filter whose motion and measurements are consistent.

D. Following a Person

Our approach is designed to track the person over longer periods of time and longer distances. Therefore, the robot needs to follow the person in order to obtain relevant observations, which are required to correct the pose estimate obtained from the inertial motion capture suit. To follow the tracked person, our current implementation uses a standard robot navigation approach to guide the robot to the vicinity of the person’s current location.

IV. OPTIMIZING THE TRAJECTORIES OF PERSONS BY MEANS OF LEAST SQUARES

When an animated character walks in a movie or in a virtual game according to the recorded trajectory, it is necessary that this trajectory accurately reflects the motion of the person. A tracking approach, especially based on particle filters, however, can lead to non-smooth tracks with small jumps in the trajectory. In an animation, this would look like the character being teleported between nearby locations because the track does not reflect the smooth motion of a person that the inertial motion capture suit itself provides.

To bridge this gap between local and global accuracy, we perform a post-processing step that refines the trajectory estimate provided by the particle filter. Our solution presented here bears resemblance with the graph-based SLAM problem, except that three different types of constraints are considered, namely the location of the robot, the person’s location

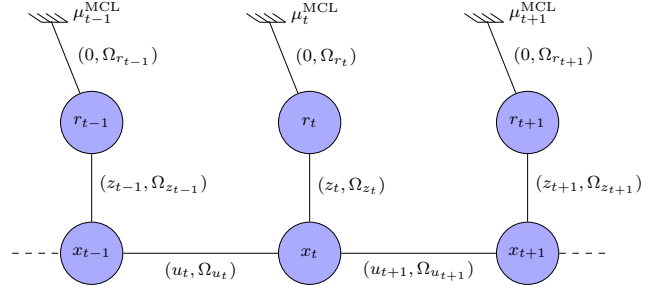


Fig. 2: The structure of the underlying optimization problem illustrated as a graph, similar to the pose graph in the SLAM problem. The nodes in the graph represent the poses r_t of the robot while observing the person and the tracker estimates of the person’s locations x_t . The edges model the laser-based observations z_t and the inertial motion capture suit data u_t . The poses r_t of the robot are anchored in the global reference frame according to the MCL estimate of the robot during data acquisition. Uncertainties are modeled by the corresponding information matrices Ω .

observation, and the inertial motion capture suit data. See Fig. 2 for an illustration.

As before, let x_t be the person’s pose at time t , let z_t be the observation of that person seen from the location r_t of the robot, and let u_t be the motion estimate of the inertial motion capture suit. Let Ω_{z_t} be the information matrix encoding the accuracy of the estimated poses of the particle-based tracker of the person (see Section III), let and Ω_{r_t} be the accuracy of the localization of the robot itself, which is, in contrast to Section III, considered explicitly here. Let Ω_{u_t} be the information matrix encoding the accuracy of the inertial motion capture suit.

The overall goal is to find the most likely configuration $x_{1:n}^*$ of the person’s trajectory of length n as

$$x_{1:n}^* = \operatorname{argmax}_{x_{1:n}} p(x_{1:n}, r_{1:n} \mid z_{1:n}, u_{1:n}). \quad (7)$$

Under the assumption of independent Gaussian noise, the maximum likelihood estimation is equivalent to a least squares error minimization given the error function

$$F = \sum_{t=1}^n \underbrace{(e_{z_t}^\top \Omega_{z_t} e_{z_t})}_{f_{z_t}} + \sum_{t=1}^n \underbrace{(e_{r_t}^\top \Omega_{r_t} e_{r_t})}_{f_{r_t}} + \sum_{t=2}^n \underbrace{(e_{u_t}^\top \Omega_{u_t} e_{u_t})}_{f_{u_t}}. \quad (8)$$

The individual error terms e_{r_t} for the robot’s own localization, e_{z_t} for the observed locations of the person, and e_{u_t} for the inertial motion capture suit estimate are defined as

$$e_{z_t} = (r_t \oplus z_t) \ominus x_t \quad (9)$$

$$e_{r_t} = r_t \ominus \mu_t^{\text{MCL}} \quad (10)$$

$$e_{u_t} = (x_{t-1} \oplus u_t) \ominus x_t. \quad (11)$$

Minimizing these error functions is not directly possible since they include non-linear terms. Therefore, we compute an approximation of these error functions by applying a Taylor expansion. For simplicity of notation, we introduce e_{k_t} as a generic error function that can be used for $k \in \{z, r, u\}$. Given this notation, the approximations of the individual error functions can be written as

$$e_{k_t}(\mathbf{x} + \Delta \mathbf{x}) \simeq e_{k_t}(\mathbf{x}) + \mathbf{J}_{k_t} \Delta \mathbf{x}, \quad (12)$$

where \mathbf{x} is a stacked vector of the person’s poses $x_{1:t}$ and the robot poses $r_{1:t}$. Furthermore, \mathbf{J}_{k_t} is the Jacobian of the corresponding error functions given in (9)-(11). Given this approximation of the error function, we can write

$$\begin{aligned} f_{k_t}(\mathbf{x} + \Delta\mathbf{x}) &\simeq \underbrace{e_{k_t}^\top \Omega_{k_t} e_{k_t}}_{c_{k_t}} + 2 \underbrace{e_{k_t}^\top \Omega_{k_t} \mathbf{J}_{k_t}}_{\mathbf{b}_{k_t}} \Delta\mathbf{x} + \Delta\mathbf{x}^\top \underbrace{\mathbf{J}_{k_t}^\top \Omega_{k_t} \mathbf{J}_{k_t}}_{\mathbf{H}_{k_t}} \Delta\mathbf{x} \\ &= c_{k_t} + 2\mathbf{b}_{k_t} \Delta\mathbf{x} + \Delta\mathbf{x}^\top \mathbf{H}_{k_t} \Delta\mathbf{x}, \end{aligned} \quad (13)$$

yielding the variables c_{k_t} , \mathbf{b}_{k_t} , and \mathbf{H}_{k_t} . Based on this equation, we can directly obtain an approximation of F that can be solved via the Gauss-Newton algorithm:

$$\begin{aligned} F &\simeq \sum_{t=1}^n \sum_{k \in \{z, r\}} c_{k_t} + 2\mathbf{b}_{k_t} \Delta\mathbf{x} + \Delta\mathbf{x}^\top \mathbf{H}_{k_t} \Delta\mathbf{x} \\ &\quad + \sum_{t=2}^n c_{u_t} + 2\mathbf{b}_{u_t} \Delta\mathbf{x} + \Delta\mathbf{x}^\top \mathbf{H}_{u_t} \Delta\mathbf{x} \end{aligned} \quad (14)$$

$$= c + 2\mathbf{b}^\top \Delta\mathbf{x} + \Delta\mathbf{x}^\top \mathbf{H} \Delta\mathbf{x}. \quad (15)$$

The quadratic form in (15) is obtained from (14) by setting $c = \sum c_{k_t}$, $\mathbf{b} = \sum \mathbf{b}_{k_t}$, and $\mathbf{H} = \sum \mathbf{H}_{k_t}$. To minimize (15), we derive it with respect to $\Delta\mathbf{x}$ and set it to zero. Thus, the minimum can be found by solving the linear system

$$\mathbf{H} \Delta\mathbf{x}^* = -\mathbf{b}. \quad (16)$$

Note that the matrix \mathbf{H} is sparse by construction, which allows us to efficiently solve (16) without computing \mathbf{H}^{-1} explicitly by applying a sparse Cholesky factorization.

The solution $\Delta\mathbf{x}^*$ to (16) is used to update the current estimate of the trajectory of the person and the robot by

$$\mathbf{x} \leftarrow \mathbf{x} + \Delta\mathbf{x}^*. \quad (17)$$

The Gauss-Newton algorithm iterates the procedure until convergence and yields a solution which represents the maximum likelihood trajectory of the person given all observations. Our experiments suggest that the resulting smooth trajectory estimate is well suited for rendering a virtual character locally and globally consistent in a virtual scene.

V. EXPERIMENTAL EVALUATION

To evaluate the presented approach, we carried out several experiments. We there used an Xsens MVN suit, which is an inertial mobile motion capture suit that consists of 17 IMUs and estimates the full body posture of the person that is wearing it. We furthermore used an ActivMedia PowerBot mobile robot equipped with a SICK LMS laser ranger finder.

A. Accuracy of the MVN Inertial Motion Capture Suit

The first experiment evaluates the global accuracy of the Xsens MVN suit without any external system, such as the mobile robot. A subject therefore walked around our campus multiple times in a loop passing predefined locations. Each loop had a length of 300 m-350 m. Figure 3 depicts a typical trajectory. The red solid line is the raw estimate of the

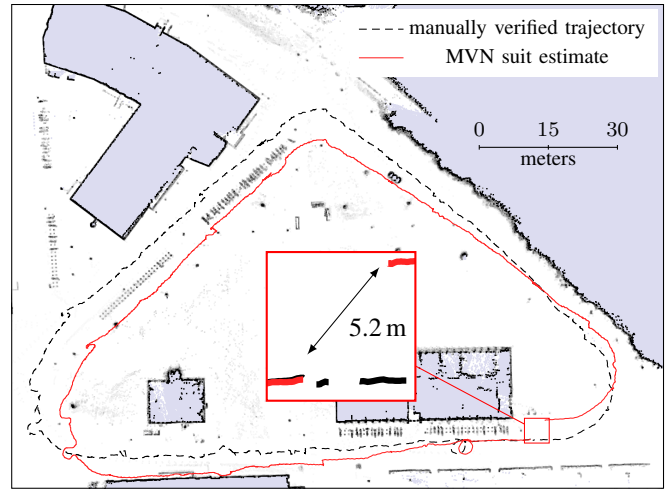


Fig. 3: The accuracy of the Xsens MVN suit in an experiment conducted on our campus. The manually verified trajectory of a person (obtained by inspecting every scan) is plotted as a dashed line. The raw trajectory estimate of the Xsens MVN suit is depicted in red. To measure the global drift, the start and the end locations of the trajectory were identical (see magnification).

Xsens MVN suit. The black dashed line is a ground truth trajectory that was recorded by an accurately localized mobile robot. This ground truth trajectory was obtained by manually verifying every single laser scan of the robot and providing the data association. After having walked a loop of 350 m, the subject returned to its starting point, which allowed us to measure an offset of 5.2 m. The run includes a number of rapid motions and small turns as can be seen by the trajectory.

In addition to that, we measured the tracking error when walking 10 times around the campus. Fig. 4 depicts the corresponding plot. Along this trajectory, we physically labeled 9 locations (1 and 10 are the same). We then evaluated the error for each round individually. The error-bars in Fig. 5 show the standard deviation around the mean error. The red and blue points correspond to the estimates of the Xsens MVN suit and our approach, respectively. We also carried out a paired sample t-test taking into account the errors at all checkpoints, which suggests that the results are statistically significant at a 99% confidence level.

B. Following a Person

The next experiment evaluates the capability of our approach to track and follow a person over long distances. We therefore used our approach to track a person that walked a distance of more than 1 km. In this experiment, the person wearing the Xsens MVN suit was additionally accompanied by three other persons walking closely to the tracked person, frequently blocking the field of view of the robot. Furthermore, there were several other people walking independently on our campus. During this experiment, the robot autonomously followed the tracked person. Fig. 6 shows the trajectory of the tracked person estimated by our approach (blue) and estimated by the Xsens MVN suit (red). In all of our experiments, our approach never lost the person that it followed, which suggests that our approach operates robustly.

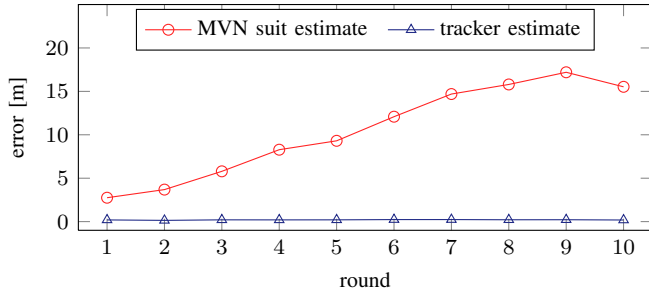


Fig. 4: A person walked on a 300 m loop ten times. After every round, we calculated the error of the pose estimate of the Xsens MVN suit with respect to a known position, which we had marked on the ground. In contrast to the Xsens MVN suit estimate that drifts, the error of our tracking system is significantly lower and stayed always below 20 cm during the entire experiment.

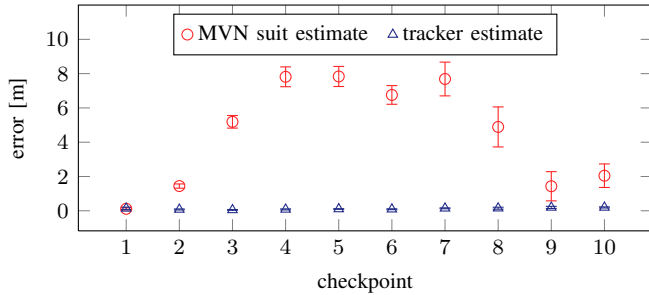


Fig. 5: The error of the estimate of the Xsens MVN suit (red) and of our tracker (blue) at predefined checkpoints along a 300 m loop (10 runs).

C. Optimization of the Person’s Trajectory

The particle filter-based tracking approach described in Section III is well suited for online person tracking. In certain situations, however, the estimate of a particle filter is non-smooth, which is suboptimal for our application. When rendering animated characters in virtual games or movies, we do not only require the person’s trajectory estimate to be globally consistent but also locally accurate and smooth. In this section, we evaluate our optimization procedure presented in Section IV.

To analyze the local smoothness of the trajectory obtained by our approach, we used the dataset from the previous experiment (see Fig. 6). In this experiment, we selected two relevant situations to analyze the smoothing. In the first situation, there was an error in the leg detection, which caused a local jump in the trajectory estimate of the person, as shown in Fig. 7. In animations, these jumps are undesirable. Fortunately, our smoothing approach is able to compensate for that by exploiting the high local accuracy of the estimate of the inertial motion capture suit. In other words, the determinant of Ω_{u_t} is bigger than the one of Ω_{z_t} .

The second situation illustrates that the optimization preserves local movements of the person, as seen in Fig. 8. The MVN suit estimate (red) and the estimate of our approach (blue) exhibit the same local behavior (apart from the global displacement), which suggests that local motions are appropriately conserved by the least squares optimization.

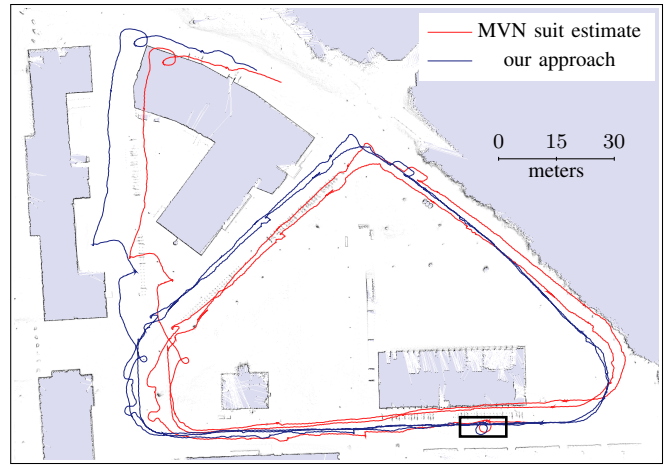


Fig. 6: Trajectory estimated while multiple people were walking close to the person to be tracked causing substantial occlusions. The black rectangle at the bottom corresponds to the magnified view shown in Fig. 8.

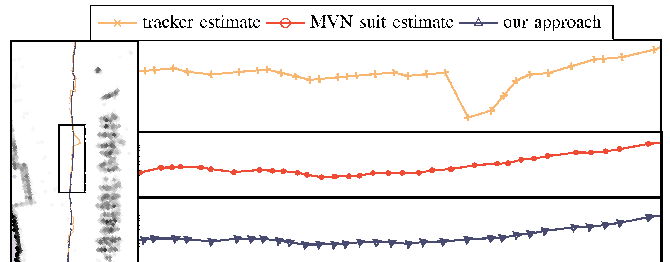


Fig. 7: Comparison of three trajectory estimates of a person, walking roughly on a straight line. The image on the left shows a part of the trajectory in the map whereas the right plots show the trajectories given the different estimation approaches. Top: particle filter-based tracker, Middle: Xsens MVN suit, Bottom: our approach using smoothing. The estimates of our approach preserve the local accuracy of the inertial motion capture suit estimate but are located correctly in the global reference frame.

Note that our smoothing approach can be implemented very efficiently. For instance, the dataset shown in Fig. 6, which comprises 5343 nodes and 7158 constraints, was optimized by our approach in around 45 ms per iteration. Given the good initial guess of the tracker, typically 2-3 iterations are sufficient for convergence.

D. Comparison to Other Motion Models

This final experiment illustrates that a standard motion model, such as a constant velocity motion model or a Brownian motion model, is not sufficient for robust person tracking in crowded scenes. This motivates the usage of the inertial motion capture suit (or something similar).

In this experiment, the robot aims to track one specific person of a walking group of three people. We initialized the tracker with the correct person and evaluated the tracking performance of three different motion models, i.e., the inertial motion capture suit motion model, a constant velocity motion model, and a Brownian motion model. As can be seen in Figure 9, the Brownian motion as well as the constant velocity motion model fail to keep track of the person. Both filters

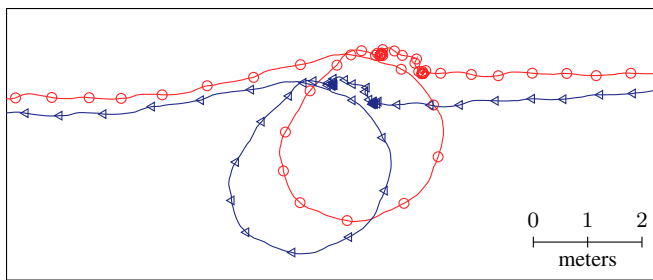


Fig. 8: The smoothed estimate (blue) preserves local movements in the trajectory (red) such as small loops (no differences in local shape).

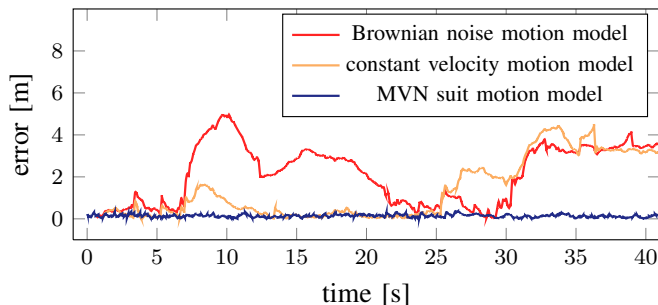


Fig. 9: Performance of the tracker using alternative motion models (Brownian motion and constant velocity) in the presence of a second person in the scene. As can be seen, the Brownian noise motion model and the constant velocity motion model fail to accurately track the person. In contrast to that, the error of the inertial motion capture suit motion model stays close to zero.

failed when other people came too close to the person to track. Once the filter diverged to a wrong person, it has no means to realize its error. In contrast, the motion model using the inertial motion capture suit data is able to correctly track the right person.

VI. CONCLUSION

In this paper, we addressed the problem of accurately and consistently tracking the full body posture of a person in large-scale environments over longer periods of time. Our approach relies on the locally accurate full body posture estimates of an inertial motion capture suit and tracks the position of the person in the environment using an autonomously navigating robot equipped with a laser range finder. To provide a smooth estimate of the person's trajectory, we formulate a joint least squares optimization problem that computes a maximum likelihood trajectory estimate given all measurements. We implemented our approach and exhaustively tested it in real world scenarios. As shown in our experiments, our algorithm yields smooth, consistent, and globally aligned estimates of the full body posture of the person.

Our method can be applied in areas which are too large for tracking systems based on static cameras and in which IMU-based systems suffer from substantial drift. It thus overcomes the key limitations of today's commercially available motion capture systems.

Note that our approach is not restricted to a single robot. Multiple robots or multiple static laser scanners can be used

with only marginal changes. In such a case, every robot introduces nodes for its own poses (r_t). Static sensors are a special case of mobile robots. Instead of multiple nodes representing the different robots' locations, a single node (r) is sufficient for each sensor. As a result, the optimization approach automatically calibrates the sensors in the global coordinate frame.

REFERENCES

- [1] K. Arras, S. Grzonka, M. Luber, and W. Burgard. Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 2008.
- [2] M. Bennewitz. *Mobile Robot Navigation in Dynamic Environments*. PhD thesis, University of Freiburg, Dept. of Computer Science, 2004.
- [3] M. Bennewitz, F. Faber, D. Joho, and S. Behnke. Fritz – A humanoid communication robot. In *Proc. of the IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, 2007.
- [4] Z. Byers, M. Dixon, K. Goodier, C.M. Grimm, and W.D. Smart. An autonomous robot photographer. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 2636–2641, 2003.
- [5] J. Cui, H. Zha, H. Zhao, and R. Shibasaki. Tracking multiple people using laser and vision. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 2116–2121, 2005.
- [6] A. Ess, K. Schindler, B. Leibe, and L. Van Gool. Moving obstacle detection in highly dynamic scenes. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2009.
- [7] A. Fod, A. Howard, and M.J. Mataric. Laser-based people tracking. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 3024–3029, 2002.
- [8] L.M. Fuentes and S.A. Velastin. People tracking in surveillance applications. *Image and Vision Computing*, pages 1165–1171, 2006.
- [9] S. Grzonka, F. Dijoux, A. Karwath, and W. Burgard. Mapping indoor environments based on human activity. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 476–481, 2010.
- [10] M. Kleinhagenbrock, S. Lang, J. Fritsch, F. Lomker, G.A. Fink, and G. Sagerer. Person tracking with a mobile robot based on multi-modal anchoring. In *Proc. of the IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, pages 423–429, 2002.
- [11] B. Kluge, C. Kohler, and E. Prassler. Fast and robust tracking of multiple moving objects with a laser range finder. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2001.
- [12] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia. People tracking and following with mobile robot using an omnidirectional camera and a laser. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 557–562, 2006.
- [13] S. Lenser and M. Veloso. Sensor resetting localization for poorly modelled mobile robots. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 1225–1232, 2000.
- [14] J.S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statist. Comput.*, 6:113–119, 1996.
- [15] M. Montemerlo, S. Thrun, and W. Whittaker. Conditional particle filters for simultaneous mobile robot localization and people-tracking. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 1259–1264, 2002.
- [16] L. Ojeda and J. Borenstein. Non-GPS navigation for emergency responders. *Sharing Solutions for Emergencies and Hazardous Environments*, 2006.
- [17] D. Schulz, D. Fox, and A.B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2001.
- [18] Vicon Motion Systems. <http://www.vicon.com>.
- [19] T. Wilhelm, H.J. Böhme, and H.M. Gross. Sensor fusion for vision and sonar based people tracking on a mobile service robot. In *Proc. of the Int. Workshop on Dynamic Perception*, pages 315–320, 2002.
- [20] Xsens. MVN - Inertial Motion Capture. <http://www.xsens.com>.
- [21] Z. Zivkovic and B. Krose. Part based people detection using 2D range data and images. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 214–219, 2007.