Multiple-Activity Human Body Tracking in Unconstrained Environments

Loren Arthur Schwarz, Diana Mateus, and Nassir Navab

Computer Aided Medical Procedures, Technische Universität München, Germany,
{schwarz,mateus,navab}@cs.tum.edu, http://campar.cs.tum.edu/

Abstract. We propose a method for human full-body pose tracking from measurements of wearable inertial sensors. Since the data provided by such sensors is sparse, noisy and often ambiguous, we use a compound prior model of feasible human poses to constrain the tracking problem. Our model consists of several low-dimensional, activity-specific motion models and an efficient, sampling-based activity switching mechanism. We restrict the search space for pose tracking by means of manifold learning. Together with the portability of wearable sensors, our method allows us to track human full-body motion in unconstrained environments. In fact, we are able to simultaneously classify the activity a person is performing and estimate the full-body pose. Experiments on movement sequences containing different activities show that our method can seamlessly detect activity switches and precisely reconstruct full-body pose from the data of only six wearable inertial sensors.

Keywords: Human pose tracking, manifold learning, wearable sensors

1 Introduction

Approaches for human full-body pose tracking have mostly been studied in the field of computer vision, where observations are typically image features, such as human silhouettes [1,2,3,4]. Vision-based methods depend on illumination, viewpoint and line of sight between the tracked person and one or more cameras. In applications where long-term tracking is addressed or when everyday-life activities need to be studied, such constraints are not practicable. Typical cases are motion analysis for ergonomic studies of factory workers or for medical diagnosis of diseases involving motion-disorders, e.g. Multiple Sclerosis [5]. These applications require the recovery of full-body motion for a set of activities of interest, while subjects move freely. We take an alternative to using vision-based observations for full-body pose estimation and rely on measurements from wearable inertial sensors. Our proposed method allows us to capture full-body motion data in situations where visual tracking systems cannot be used.

Tracking human motion using inertial sensors is challenging, since the measurements provided by such sensors are sparse, noisy and often ambiguous. Prior models of human motion are therefore a prerequisite for achieving satisfactory tracking results. We rely on identifying the low-dimensional *manifold of feasible*

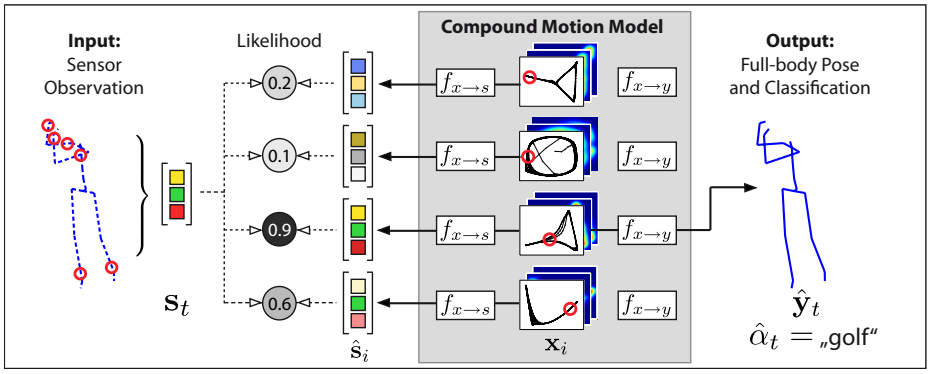

Fig. 1. The proposed compound motion model is comprised of several activity-specific models. Each of these consists of a manifold embedding of feasible poses, pose likelihood priors and learned mappings to sensor space ($f_{x\to s}$) and full-body pose space ($f_{x\to y}$). We use a particle filter in embedding space to track multiple pose hypotheses $\mathbf{x}_i$ and select the hypothesis that best matches the true sensor observation $\mathbf{s}_t$.

human poses inside the high-dimensional space of pose parameters [2]. In particular, we use Laplacian Eigenmaps [6], a manifold learning technique, to create a prior motion model from full-body pose training data. Manifold learning methods are known to produce meaningful embeddings that efficiently parameterize human poses for single activities, e.g. walking [7,2]. Unfortunately, the generalization to multiple activities, as required by the above-mentioned applications, is not straightforward [8]. In fact, a global embedding including all activities will be dominated by inter-activity differences and characteristics of individual activities will be represented inadequately. We propose to address the multiple-activity tracking problem by means of a *compound motion model* comprised of several activity-specific models and an efficient *activity switching mechanism*.

The activity-specific motion models consist of separate low-dimensional manifold embeddings generated from full-body pose training data. Together, the embeddings provide a compact representation of likely poses for multiple activities and allow us to significantly restrict search space during tracking. Additionally, we learn kernel regression mappings for each activity, which relate the low-dimensional embeddings both to observation space and to full-body space (Figure 1). We formulate the tracking problem within the Bayesian framework and use a particle filter for efficient inference. This way, we are able to track multiple pose hypotheses and select the one that best explains the sensor observations. Since a pose hypothesis in our case consists of both, a pose in embedding space and an activity index identifying the most likely motion model, we can simultaneously estimate full-body pose and classify performed activities.

Our tracking method is at the same time *general*, in that motions corresponding to multiple activities can be tracked, and *specific*, since our compound model provides specialized motion models for each activity of interest. The cost of a one-time training phase is compensated by the ability to faithfully track full-body pose from simple and limited wearable sensor observations.

1.1 Related Work

Generative full-body pose tracking methods are based on modelling the mapping from poses to observations and searching for the most likely pose, given new observations. The major difficulty is the high dimensionality of full-body pose space. Several authors have addressed this issue by sampling pose space using a particle filter [3,9]. Computation cost can also be reduced by restricting search space using learned low-dimensional human motion models [1,7,13]. For instance, Gaussian Process Latent Variable Models (GP-LVM) can provide a compact representation of human motion from training data [1,10].

With a similar purpose, our method uses a spectral embedding technique [6] for obtaining prior motion models. Spectral embeddings are low-dimensional and, as opposed to GP-LVMs, reflect local structural properties of the high-dimensional training data. Approaches using spectral embedding methods for human tracking commonly rely on a single motion model [11,7] and the generalization to various activities is not straightforward. We propose to employ a compound model built from separate, activity-specific manifold embeddings, making it possible to track various types of motions.

Mechanisms for using multiple specialized motion models for tracking can be found in several domains. The classical particle filter algorithm was extended in [12] to handle multiple, discrete dynamics models. An efficient approach for full-body tracking using multiple low-dimensional motion models is proposed in [13]. The authors demonstrate a switching mechanism for the two actions of running and walking. Unfortunately, the computational cost of their method grows significantly with the number of considered activities. In contrast, our method can be trained on a potentially arbitrary number of activities.

Most of the methods estimate pose from visual cues. Since our final goal is long-term motion analysis, where visual features are hard to obtain, we focus on mobile inertial sensors. However, sensor data is typically less informative and suffers from issues, such as drift. Existing approaches for full-body tracking using inertial sensors [14,15] recover the pose directly from the measurements. To the authors' knowledge, learning prior constraints for tracking full-body pose from sensor data is new. In [16], accelerometer measurements are compared to a database of poses and motion sequences matching the measurements are replayed in an approximation of the true motion. The method proposed in [15] is able to track full-body pose using ultrasonic sensors and accelerometers. However, their approach is computationally expensive. Our method uses a low-dimensional, efficient parameterization of human poses for reducing search space.

2 Full-body Tracking Method

We address the problem of human full-body tracking from measurements of inertial orientation sensors. Given a set of M activities of interest, we start by building a compound motion model from training data containing both full-body poses $\mathbf{y} \in \mathbb{R}^{d_y}$ and sensor readings $\mathbf{s} \in \mathbb{R}^{d_s}$. Then, during testing, we estimate the full-body pose $\hat{\mathbf{y}}_t$ at each time step t only from sensor observations $\mathbf{s}_t$.

Our compound motion model contains multiple activity-specific motion models, each of which consists of (1) a low-dimensional manifold embedding of the full-body pose training data, (2) predictive mappings from the embedding to full-body pose space and to observation space, (3) a pose likelihood prior in embedding space and (4) an activity switching prior. We formulate the tracking problem in the Bayesian framework and estimate the system state at each time step t. The system state is given by an activity index $\alpha \in \{1, \ldots, M\}$ and a pose $\mathbf{x} \in \mathbb{R}^{d_x}$ in low-dimensional embedding space ($d_x \ll d_y$). Applying a particle filter allows us to seamlessly evaluate multiple pose hypotheses and to select the most appropriate motion model for each new sensor observation. The learning tasks are described in section 2.1, the tracking approach in sections 2.2 and 2.3.

2.1 Learning Multiple Low-dimensional Motion Models

In a training phase, we learn activity-specific motion models from full-body pose data and corresponding sensor measurements. Each motion model consists of the following components (see Figure 2 for an illustration):

Manifold Embedding. Let the set of N_α full-body training poses for activity α be denoted by $\mathbf{Y}^\alpha = [\mathbf{y}_1^\alpha \ldots \mathbf{y}_{N_\alpha}^\alpha]$. We obtain a corresponding set of dimensionality-reduced points $\mathbf{X}^\alpha = [\mathbf{x}_1^\alpha \ldots \mathbf{x}_{N_\alpha}^\alpha]$ by applying Laplacian Eigenmaps, a spectral embedding technique [6]. The low-dimensional points $\mathbf{x}_i^\alpha$ efficiently represent the *manifold of feasible poses* for activity α. Using this representation for tracking, we are able to restrict search space to likely poses, instead of exhaustively searching the high-dimensional full-body pose space.

Predictive Mappings. In order to relate poses in low-dimensional embedding space to sensor measurements and to full-body poses, we learn predictive mappings from training data. We follow the approach in [11,7] and use nonlinear kernel regression, with the difference that we learn separate mappings from each of the activity-specific manifold embeddings. The mapping $f_{x \to y}^\alpha(\mathbf{x})$ for *prediction of full-body poses* is learned from corresponding training pairs of embedding points $\mathbf{x}_i^\alpha$ and full-body poses $\mathbf{y}_i^\alpha$ for an activity α. Similarly, the mapping $f_{x \to s}^\alpha(\mathbf{x})$ is learned from training pairs of embedding points and sensor measurements $\mathbf{s}_i^\alpha$ and allows *predicting sensor values*.

Pose Likelihood Prior. Using the training data for each activity, we can derive the likelihood for arbitrary poses in low-dimensional embedding space. Intuitively, poses $\mathbf{x}$ that are close to the embedding points $\mathbf{x}_i^\alpha$ learned from training data should have the highest likelihood. The pose likelihood prior for activity α is obtained using a kernel density estimate [11,7] as $p_{\text{pose}}^\alpha(\mathbf{x}) = \frac{1}{N_\alpha} \sum_{i=1}^{N_\alpha} k(\mathbf{x}, \mathbf{x}_i^\alpha)$, where $k(\cdot, \cdot)$ is a Gaussian kernel function.

Activity Switching Prior. We also define a prior distribution $p_{\text{switch}}^\alpha(\mathbf{x})$ for every motion model that describes how likely a switch of activity is, given a pose $\mathbf{x}$ in embedding space. To ensure generality, we allow activity switching from any pose with constant minimum probability p_k. However, we let the probability of switching increase for poses that typically occur between subsequent activities. In our experiments, the upright standing pose was used as an intermediate pose

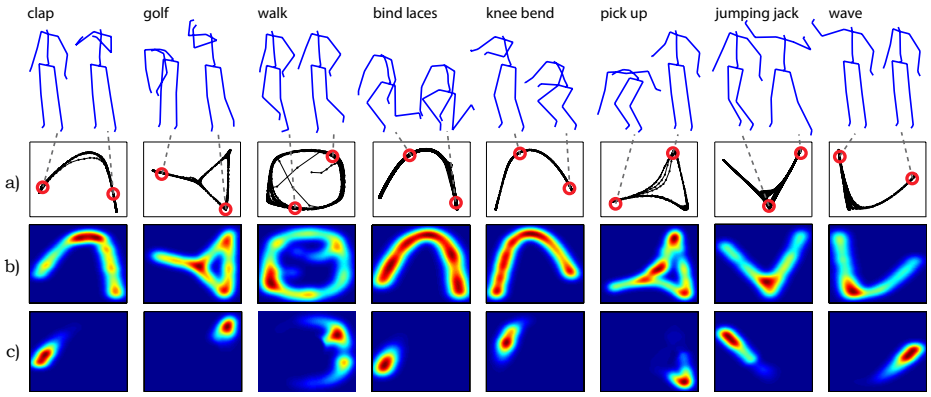

Fig. 2. Learned motion models for 8 different activities. a) Two-dimensional manifold embeddings obtained using Laplacian Eigenmaps on full-body pose training data. Each point on the manifolds corresponds to a valid full-body pose, a few examples are shown above. b) Static pose priors in latent space. c) Activity switching priors in latent space.

that encourages activity switching. We model the switching prior with a normal distribution $p^{\alpha}_{\text{switch}}(\mathbf{x}) = \mathcal{N}(f^{\alpha}_{x\to y}(\mathbf{x}); \mathbf{y}_0, \mathbf{\Sigma}^{\alpha}_y) + p_k$, where $\mathbf{y}_0$ represents the intermediate pose in full-body space, $f^{\alpha}_{x\to y}(\mathbf{x})$ is a predicted full-body pose and $\mathbf{\Sigma}^{\alpha}_y$ is the diagonal covariance matrix of the training data $\mathbf{Y}^{\alpha}$.

2.2 Bayesian Tracking Using Multiple Motion Models

The testing phase of our method consists of tracking pose in low-dimensional embedding space. In a standard Bayesian tracking formulation, we wish to find the optimum of the posterior $p(\mathbf{x}_t|\mathbf{s}_t) = p(\mathbf{s}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{s}_{t-1})$, with $p(\mathbf{x}_t|\mathbf{s}_{t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{s}_{t-1})d\mathbf{x}_{t-1}$. In other words, we seek the most likely pose $\mathbf{x}_t$ in embedding space at time t, given the observations up to $\mathbf{s}_t$. The *dynamics model* $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ determines how pose estimates are updated from one time step to the next and the *observation model* $p(\mathbf{s}_t|\mathbf{x}_t)$ links poses in embedding space to observations. Since we are using multiple motion models for tracking, we need to include the discrete activity index $\alpha_t \in \{1, \ldots, M\}$, leading to the posterior

$$\underbrace{p(\mathbf{x}_t, \alpha_t|\mathbf{s}_t)}_{\text{posterior}} = \underbrace{p(\mathbf{s}_t|\mathbf{x}_t, \alpha_t)}_{\text{observation model}} \underbrace{p(\mathbf{x}_t, \alpha_t|\mathbf{s}_{t-1})}_{\text{prior}}. \tag{1}$$

Following [12], we also augment the dynamics model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ with an activity index, yielding the factored model

$$\underbrace{p(\mathbf{x}_t, \alpha_t|\mathbf{x}_{t-1}, \alpha_{t-1})}_{\text{dynamics model}} = \underbrace{p(\mathbf{x}_t|\mathbf{x}_{t-1}, \alpha_t, \alpha_{t-1})}_{\text{pose dynamics}} \underbrace{p(\alpha_t|\mathbf{x}_{t-1}, \alpha_{t-1})}_{\text{activity dynamics}}. \tag{2}$$

The pose dynamics model governs the evolution of poses in embedding space and the activity dynamics model describes the activity switching process.

Pose Dynamics Model. We define the new pose dynamics model as follows:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \alpha_t, \alpha_{t-1}) = \begin{cases} p(\mathbf{x}_t|\mathbf{x}_{t-1}) & \text{if } \alpha_t = \alpha_{t-1}, \\ p^{\alpha_t}_{\text{pose}}(\mathbf{x}_t) & \text{else.} \end{cases} \quad (3)$$

When there is no switch of activity ($\alpha_t = \alpha_{t-1}$), dynamics are governed by a random walk, modeled as a normal distribution centered at the previous pose in embedding space, $p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_{t-1}, \mathbf{\Sigma}^{\alpha_t}_x)$. Here, $\mathbf{\Sigma}^{\alpha_t}_x$ is the diagonal covariance matrix of the low-dimensional training data $\mathbf{X}^{\alpha_t}$. In the case of activity switching ($\alpha_t \neq \alpha_{t-1}$), the dynamics model follows the pose likelihood prior $p^{\alpha_t}_{\text{pose}}(\mathbf{x})$ of activity α_t (section 2.1). In other words, the most likely poses after switching to activity α_t are those learned from the training data.

Activity Dynamics Model. We assume that all sequences of consecutive activities are equally likely, i.e. $p(\alpha_t = j|\mathbf{x}_{t-1}, \alpha_{t-1} = i)$ is equal for all activity indices $j \neq i$. The probability of switching from a given activity α_{t-1} to any other activity then only depends on the previous pose $\mathbf{x}_{t-1}$ in embedding space. Thus, we state our activity dynamics model using the activity switching prior defined in section 2.1 as $p(\alpha_t|\mathbf{x}_{t-1}, \alpha_{t-1}) = p^{\alpha_{t-1}}_{\text{switch}}(\mathbf{x}_{t-1})$.

Observation Model. Our observation model $p(\mathbf{s}_t|\mathbf{x}_t, \alpha_t)$ relates observations to the learned embedding space. We define it as a product of three terms:

$$p(\mathbf{s}_t|\mathbf{x}_t, \alpha_t) = \underbrace{\mathcal{N}(\mathbf{s}_t; f^{\alpha_t}_{x\to s}(\mathbf{x}_t), \mathbf{\Sigma}^{\alpha_t}_s)}_{\text{prediction term}} \underbrace{\mathcal{N}(\mathbf{y_{t-1}}; f^{\alpha_t}_{x\to y}(\mathbf{x}_t), \mathbf{\Sigma}^{\alpha_t}_y)}_{\text{full pose smoothness term}} \underbrace{p^{\alpha_t}_{\text{pose}}(\mathbf{x}_t)}_{\text{prior}}. \quad (4)$$

The prediction term uses the learned mapping $f^{\alpha}_{x\to s}(\mathbf{x})$ to predict sensor observations from a pose $\mathbf{x}_t$. The likelihood of $\mathbf{x}_t$ based on this term is maximal if the prediction perfectly matches the true observation $\mathbf{s}_t$. In order to reduce the influence of outlier observations, the smoothness term penalizes embedding locations if their predicted full-body pose differs strongly form the previous pose $\mathbf{y}_{t-1}$. The pose likelihood prior encourages poses that are likely with respect to the training data. $\mathbf{\Sigma}^{\alpha_t}_s$ and $\mathbf{\Sigma}^{\alpha_t}_y$ are the diagonal covariance matrices of the training observations $\mathbf{S}^{\alpha_t}$ and full-body poses $\mathbf{Y}^{\alpha_t}$ belonging to activity α_t.

2.3 Particle Filtering and Full-body Pose Inference

We employ a particle filter [17,12] to sample the posterior density in Eq. 1. The particle filter, adapted to use our compound motion model, allows simultaneously evaluating pose hypotheses of different motion models and selecting the most appropriate model. Particle filtering is computationally efficient in our setting, since it is applied in the low-dimensional space of manifold embeddings.

We initialize n particles $(\mathbf{x}^i_0, \alpha^i_0)$, $i \in \{1 \dots n\}$, with locations across all manifold embeddings of our compound model. At each time step t, we first resample the particles according to their weights w^i_{t-1}. Each particle is then updated by sampling from the dynamics model $p(\mathbf{x}_t, \alpha_t|\mathbf{x}_{t-1} = \mathbf{x}^i_{t-1}, \alpha_{t-1} = \alpha^i_{t-1})$. This implies switching the i-th particle to a randomly chosen other activity with probability $p^{\alpha'}_{\text{switch}}(\mathbf{x}^i_{t-1})$, where $\alpha' = \alpha^i_{t-1}$. The weights are re-computed using

the observation model, $w_t^i = p(\mathbf{s}_t|\mathbf{x}_t = \mathbf{x}_t^i, \alpha_t = \alpha_t^i)$. We then determine the estimated activity $\hat{\alpha}_t$ as the most frequent activity among the highest-weight particles. The pose estimate $\hat{\mathbf{x}}_t$ in low-dimensional space is computed as a convex combination of the positions of the highest-weight particles with activity $\hat{\alpha}_t$. The full-body pose at time t is finally obtained as $\hat{\mathbf{y}}_t = f_{x \to y}^{\hat{\alpha}_t}(\hat{\mathbf{x}}_t)$.

3 Experiments and Results

We acquired a synchronized dataset of full-body poses $\mathbf{Y}^\alpha$ and sensor values $\mathbf{S}^\alpha$, $\alpha \in \{1 \dots M\}$, using a motion capture system and six wearable inertial *orientation* sensors. A full-body pose is given by a vector of $d_y = 35$ dimensions representing the joint angles of our skeleton body model. An observation vector has $d_s = 12$ dimensions, representing pitch and roll for each of the sensors. The yaw values were omitted for independence of magnetic north. We placed sensors on the wrists, upper arms and shinbones of each person. In the training phase, we learned a manifold embedding $\mathbf{X}^\alpha$ of $d_x = 2$ dimensions for each activity.

We considered $M = 10$ activities: clapping, golfing, hurrah (arms up), jumping jack, knee bends, binding laces, picking something up, scratching head, walking and waving. Each of the movements was recorded 6 times with 9 actors. Every movement recording has a length of ~ 600 frames. The testing data consists of 5 sequences per actor containing all activities (~ 2000 frames each). See Figure 3 for an illustration. For tracking, only the inertial sensor values were used, the motion capture data served as ground truth. All experiments were performed in a cross-validation scheme, i.e. each testing sequence was generated from one of the recordings per activity and actor, using the remaining five for training.

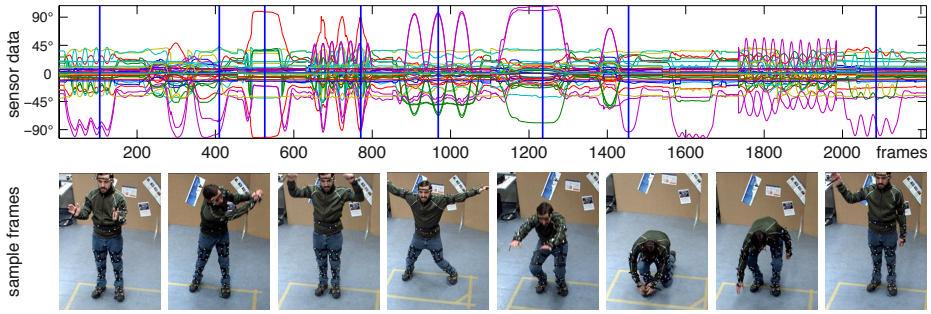

Fig. 3. *Top:* Inertial sensor data for a sequence of 10 activities in a row. *Bottom:* Pictures of the person at the time instants marked above with vertical lines. The person is equipped with motion capture markers and six wearable inertial sensors.

Full-body Pose Tracking. Noting that the appropriate number of particles grows *linearly* with the number of considered activities, we used $n = 400$ particles. Figure 4 illustrates how particles sample the activity-specific manifold embeddings (only two are shown for clarity) for a testing sequence switching from

waving to *golfing*. Initially, the person is waving and most particles are concentrated around a pose on the *waving* embedding. A small number of particles also samples all other manifolds. As the person leans forward for golfing, particles quickly accumulate on the *golfing* manifold, since the sensor predictions of these particles increasingly match the real observations. Subsequent resampling steps cause the majority of particles to follow.

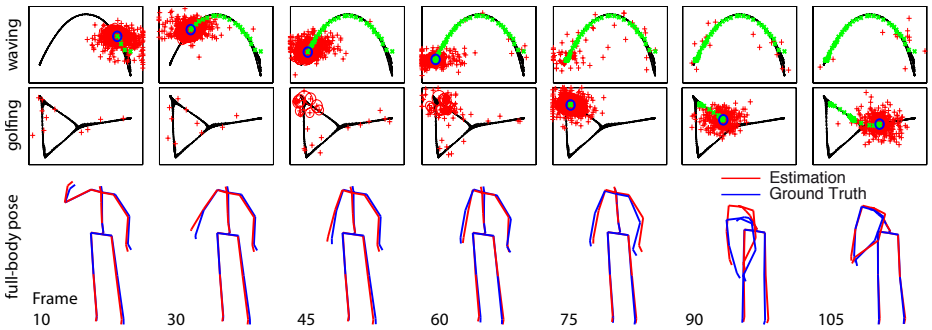

Fig. 4. The particle filter-based activity switching mechanism on a sample sequence. Two of the ten activity manifold embeddings (*waving* and *golfing*) are displayed for several frames. Particles are shown as red crosses. The particles used for predicting full-body pose are circled in dark color. Green crosses indicate the trace of previous frames. Shown below are the corresponding predicted and ground-truth body poses.

Activity Classification. The number of particles per activity manifold is an indicator of activity class membership. Figure 5.(a) shows classification results for the testing sequence in Figure 3. The particle count over time for four of the manifold embeddings is displayed, along with predicted and true activity classifications. Misclassifications mainly occur at the beginning and end of activities. In fact, these frames can be classified as any activity, since the person is standing idle. The confusion matrix in Figure 5.(b) gives the classification rates for all activities over all testing sequences. On average, we achieved a correct classification rate of 89% for all non-idle frames. The matrix is mostly diagonal, significant confusion only occurs between *waving* and *scratching head*, which both consist of raising the right arm close to the head. Misclassification in this case therefore does not necessarily affect the precision of full-body pose estimation.

Pose Estimation Accuracy. We measured how precisely the poses estimated by our method match the ground truth using two metrics. The angular error e_{ang} gives the deviation from the ground truth in terms of joint angles. The distance error e_{dist} is the difference in 3D space between predicted joint locations and the ground truth. Averaged over all frames of the testing sequences, we achieved $\bar{e}_{\text{ang}} = 6.23°$ per joint and $\bar{e}_{\text{dist}} = 45.2$mm. As shown in Table 1, the deviation from the ground truth only increases for fast movements with a large variability, such as *jumping jack* or *walking*. Our results are comparable to other state-of-the-art methods that use visual observations [4,15].

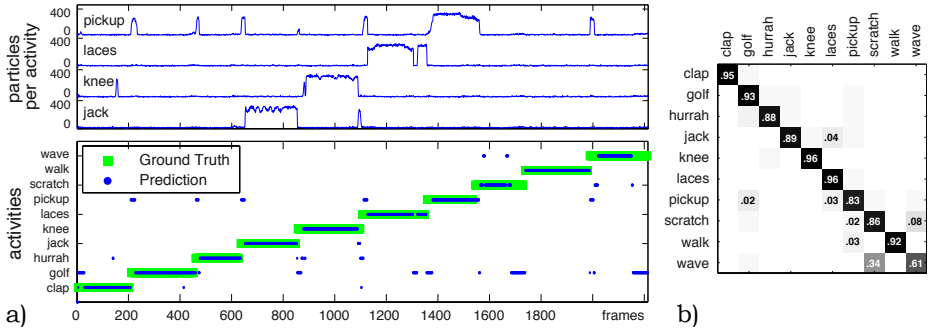

Fig. 5. (a) Activity classification results for the sequence shown in Fig. 3. *Top:* Number of particles per frame sampling four of the activity manifolds. *Bottom:* Ground truth classification and predicted activities for each frame of the sequence. (b) Confusion matrix computed from the classification results for all testing sequences.

	clap	golf	hurrah	jack	knee	laces	pickup	scratch	walk	wave
e_{ang}	4.95	6.10	6.79	8.80	4.87	5.90	5.90	4.43	9.65	4.93
e_{dist}	37.8	51.2	40.4	58.1	45.6	60.5	51.4	27.7	50.7	28.5

Table 1. Pose estimation accuracy for all considered activities. Deviations from ground truth poses are provided as joint angles (e_{ang} in degrees per joint) and as distances (e_{dist} in millimeters per joint), averaged over all experiments.

4 Discussion and Conclusion

The learned compound model of feasible poses provides a reliable framework for multiple-activity tracking from limited, low-dimensional observations. Apart from wearable sensor data, other observations can also be used, such as sparse visual features. A requirement of our method is that the motion model is initially trained on a set of activities. While the model allows stylistic variation between instances of the same motion, completely unseen movements will not be reconstructed precisely. However, our method will still provide a pose estimate that matches the new observations as close as possible. The multiple-hypothesis tracker will furthermore quickly recover the correct pose, as soon as a known movement is performed. We also do not require pose initialization, since when tracking begins, particles are distributed to sample all learned feasible poses.

Since inertial sensors only measure relative movement, we do not track the global position of a person. However, integrating global tracking can be easily achieved using conventional positioning systems. We particularly target scenarios where the focus lies on the movement itself, not the person's location. An application of interest to us is medical motion analysis for Multiple Sclerosis patients [5]. Currently, physicians evaluate the disease state by analyzing patient motion in a short protocol including movements such as walking and jumping. Training our method on the movements of the protocol would allow acquiring motion data over longer periods of time in the patient's everyday environment.

We currently investigate how to extend the method for being able to detect anomalies (i.e. unknown activities) in such scenarios.

In conclusion, we have presented a method for tracking human full-body pose given only limited observations from wearable inertial sensors. For dealing with the sparse and often ambiguous sensor data, we learn a compound model of human motion from full-body pose training data. The method is efficient, since we track poses in a low-dimensional space of manifold embeddings and use sparse non-linear regression to relate the embedding space to observations and to full-body poses. Our experiments showed that we can reliably recognize motions of multiple activities and precisely track human full-body pose.

References

1. Urtasun, R., Fleet, D., Fua, P.: 3d people tracking with gaussian process dynamical models. CVPR, June (2006)
2. Elgammal, A., Lee, C.: The role of manifold learning in human motion analysis. Human Motion Understanding, Modeling, Capture and Animation (2008) 1–29
3. Bandouch, J., Engstler, F., Beetz, M.: Accurate human motion capture using an ergonomics-based anthropometric human model. AMDO (2008)
4. Agarwal, A., Triggs, B., Montbonnot, F.: Recovering 3d human pose from monocular images. PAMI **28**(1) (2006) 44–58
5. Weikert, M., Motl, R.W., Suh, Y., McAuley, E., Wynn, D.: Accelerometry in persons with multiple sclerosis: Measurement of physical activity or walking mobility? Journal of the neurological sciences **290**(1) (2010) 6 – 11
6. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation **15**(6) (2003) 1373 – 1396
7. Lu, Z., Carreira-Perpinan, M., Sminchisescu, C.: People tracking with the laplacian eigenmaps latent variable model. NIPS (Jan 2007)
8. Datta, A., Sheikh, Y., Kanade, T.: Modeling the product manifold of posture and motion. THEMIS Workshop (2009)
9. Deutscher, J., Reid, I.: Articulated body motion capture by stochastic search. IJCV **61**(2) (2005) 185 – 205
10. Wang, J., Fleet, D., Hertzmann, A.: Gaussian process dynamical models for human motion. PAMI (2008) 283–298
11. Kanaujia, A., Sminchisescu, C., Metaxas, D.: Spectral latent variable models for perceptual inference. ICCV (2007) 1–8
12. Isard, M., Blake, A.: A mixed-state condensation tracker with automatic model-switching. ICCV (1998) 107–112
13. Jaeggli, T., Koller-Meier, E., van Gool, L.: Learning generative models for multi-activity body pose estimation. IJCV **83** (2009) 121 – 134
14. Roetenberg, D., Slycke, P., Veltink, P.: Ambulatory position and orientation tracking fusing magnetic and inertial sensing. IEEE Transactions on Biomedical Engineering **54**(4) (2007) 883–890
15. Vlasic, D., Adelsberger, R., Vannucci, G., Barnwell, J.: Practical motion capture in everyday surroundings. ACM TOG (2007)
16. Slyper, R., Hodgins, J.: Action capture with accelerometers. ACM SIGGRAPH Symposium on Computer Animation (2008) 193–199
17. Isard, M., Blake, A.: Condensation—conditional density propagation for visual tracking. IJCV (1998)