

Discriminative Human Full-Body Pose Estimation from Wearable Inertial Sensor Data

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. In this paper, a method is presented that allows reconstructing the full-body pose of a person in real-time, based on the limited input from a few wearable inertial sensors. Our method uses Gaussian Process Regression to learn the person-specific functional relationship between the sensor measurements and full-body pose. We generate training data by recording sample movements for different activities simultaneously using inertial sensors and an optical motion capture system. Since our approach is discriminative, pose prediction from sensor data is efficient and does not require manual initialization or iterative optimization in pose space. We also propose a SVM-based scheme to classify the activities based on inertial sensor data. An evaluation is performed on a dataset of movements, such as walking or golfing, performed by different actors. Our method is capable of reconstructing the full-body pose from as little as four inertial sensors with an average angular error of 4-6 degrees per joint, as shown in our experiments.

Keywords: Human pose estimation, Gaussian process regression, wearable inertial sensors, ambulatory motion analysis.

1 Introduction

Human motion analysis plays an important role in various fields of application, such as industrial ergonomic studies or sports engineering, but especially in the medical domain. Precise and quantitative representations of human motion can help physicians in diagnosis, treatment planning and progress evaluation. For instance, orthopedists use motion analysis to assess the status of diseases affecting the skeletal system [1]. Neurologists analyze the movements of epilepsy patients to identify the cortical locations that cause epileptic seizures [2].

Typical systems for acquisition of full-body motion work with optical tracking systems based on a stationary multiple camera setup. The person is often required to wear cumbersome markers or suits and to stay within a dedicated tracking volume. However, medical motion analysis can be most beneficial when patient movements are monitored over extended periods of time and during activities of daily life. In such cases, a pose estimation system using wearable body sensors as an input is of great advantage. Patients equipped with a few wireless

and light-weight sensors on their extremities could move naturally and follow their everyday routine, while their movements are recorded. In contrast to existing ambulatory motion analysis systems that record only specific signals, such as heart rate or step count [1,3,4], a mobile pose estimation system could provide physicians with a reconstruction of full-body movements.

However, inferring full-body pose from inertial sensor data, such as acceleration or orientation of individual body segments, is a challenging problem. The sensor measurements typically exhibit high drift rates and provide only a limited number of constraints for analytical calculation of human pose, e.g. using inverse kinematics [5,6,7]. Instead, we propose to use prior knowledge about feasible human poses obtained from sample motions in a learning-based approach. In a training phase, the person-specific functional relationship between the low-dimensional sensor data and the full-body pose, acquired with a motion capture system, is learned. After training, our method allows reconstructing the pose of a person in real-time, given only the data of a few inertial orientation sensors attached to the limbs (Figure 1). The focus of our method is not a generalization across different actors, but the ability to model the particular motions of individual persons for evaluation, as in the case of medical motion analysis.

The problem of estimating human pose from low-dimensional input data can be approached using *generative* or *discriminative* learning techniques. In the former case, one searches for the joint angle configuration of a simplified skeleton model that best fits the measured input data [8,9,10,11]. This typically requires using costly iterative optimization algorithms. On the other hand, *discriminative* methods learn a direct functional relationship from input data to joint angles, allowing efficient pose prediction. Gaussian Process Regression (GPR) is a discriminative learning technique that has been applied to predict full-body pose from image-based input data, such as human silhouettes [12,13,14,15]. We propose to employ GPR for learning the direct mapping from inertial sensor data to full joint angle configurations. To the best of our knowledge, a discriminative learning-based pose estimation approach using observations from wearable sensors has not been proposed before.

In our method, person-specific GPR models are learned for each activity of interest, such as walking or particular disease-related movements. The person is required to perform sample movements using a motion capture system to generate training data. After training, the predictive means of the GPR models provide pose estimates from wearable sensor data. Additionally, in order to determine which of the learned activities a person is performing, we propose to use a multi-class Support Vector Machine (SVM) classifier, trained on the wearable sensor dataset with the known activities as class labels.

The main contribution of this paper is to demonstrate that full human body pose can be estimated based on just a few wearable inertial sensors using a discriminative learning approach. This opens the way for creating a wearable pose estimation system suitable for long-term evaluation of full-body motion during daily life. The quantitative evaluation of our method shows that the full-body pose can be estimated for the activities we considered (Section 3.1)

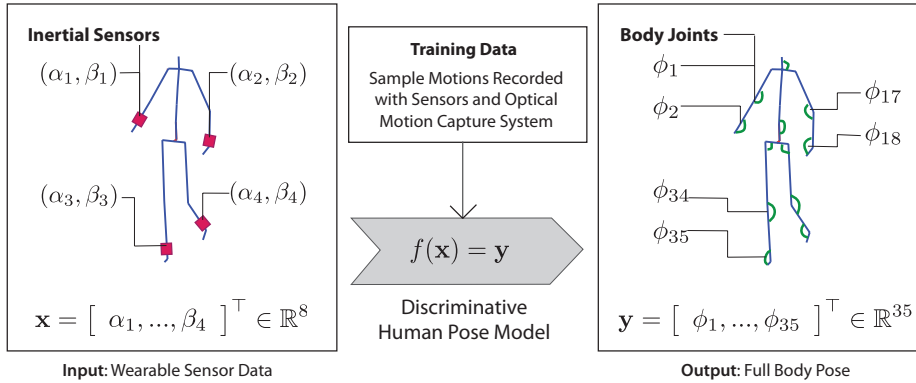


Fig. 1. Schematic of the proposed pose estimation approach. An actor is equipped with a few inertial sensors measuring orientation angles. Full body pose is predicted using a Gaussian Process Regression model that has been trained on the functional relationship between the sensor data \mathbf{x} and the joint angle configuration \mathbf{y} .

with average errors of 4-6 degrees per joint, using four, three and even two inertial orientation sensors. This reconstruction quality is comparable to other state-of-the-art methods that use higher-dimensional input for discriminative pose estimation [12,13,16,17].

1.1 Related Work

There is a large variety of learning-based approaches for human pose estimation in the literature [12,13,14,15,16,17]. Many recently published methods are *generative* and focus on learning low-dimensional representations of human poses using so-called latent variable models [8,9,10,11]. After projecting a training dataset of poses into a low-dimensional (latent) subspace, these methods predict new poses by optimizing for the latent positions that best account for new observations. For instance, Urtasun et al. [11] use Gaussian Process Latent Variable Models (GPLVM) to learn a 3D latent space from motion capture data. Human poses are then tracked in video sequences by optimization in each frame.

Disadvantages of such *generative* pose estimation methods are their sensitivity to initialization and the high computational cost of pose inference [13]. Several authors therefore propose to directly learn a mapping from simple, low-dimensional observations to the corresponding full body poses by non-linear regression [12,17,15]. Such *discriminative* approaches have the advantage that poses can be predicted efficiently, without the need for iterative optimization.

Different types of input data have been proposed for discriminative pose estimation, such as silhouettes in 2D images [12], SIFT-features [15] or 3D voxel-representations of humans [17]. Fossati et al. [13] use trajectories of body parts in videos as an input. In contrast, our method estimates human pose from wearable inertial sensor measurements. Learning methods that have been described

in the context of discriminative human pose estimation include Relevance Vector Machines [12,17], Support Vector Machines [18] and Gaussian Process Regression [13,14,15]. We use the latter technique because of its compact mathematical formulation and its computational efficiency.

Similar to our goal of reducing the dimension of the input data, Chai and Hodgins [19] reconstruct full-body pose from a low number of visual markers in an optical setup. However, the person is still constrained to stay within the camera volume. Wearable sensors allow us to acquire information on human pose without being bound to a specific location.

The goal of estimating human pose based on wearable sensors has been targeted by several authors [6,7,20], however, mostly relying on physical models, inverse kinematics and stochastic filtering. For instance, Zhou et al. [7] use accelerometers to compute the pose of one arm based on kinematic constraints. Roetenberg et al. [6] fuse inertial and ultrasonic sensor streams using a Kalman filter, allowing to track position and orientation of the four limbs and the torso. A Kalman-based approach is also described by Zhu et al. [20], where magnetic sensors are used for tracking of a single arm. Our method removes the need for complex, physical motion models, since learned prior knowledge about human movements is used to cope with limitations of low-dimensional sensor data.

Paper Organization. In section 2, the methods we apply for pose estimation are presented. Training of human motion models is explained in section 2.3. Section 2.4 describes how our method recognizes activities from inertial sensor data. Section 2.5 explains how pose is estimated, given the learned pose models and an activity classification. A quantitative evaluation of the proposed method is given in section 3.

2 Human Pose Estimation Approach

In this paper, we propose a method to reconstruct full-body pose of an actor, given only the readings from wearable orientation sensors attached to the extremities of the actor. We propose to learn person-specific motion models for each considered activity from corresponding sets of sensor data and full-body joint angles, obtained offline by motion capture. In the testing phase, only the sensor data is used for pose estimation, allowing the person to move freely and follow their daily routine.

2.1 Method Outline

Let $\mathbf{x}_i \in \mathbb{R}^d$ be a vector containing the concatenated measurements of all inertial sensors for a single pose. Also, let $\mathbf{y}_i \in \mathbb{R}^D$ ($d \ll D$) denote a vector of joint angles that fully determines a pose in terms of the articulated body model (Figure 2(a)). The sensor measurements for a specific activity, consisting of a sequence of M poses, are arranged in a matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M]^\top$. The corresponding joint angle configurations are stored in a matrix $\mathbf{Y} = [\mathbf{y}_1 - \bar{\mathbf{y}}_n, \dots, \mathbf{y}_M - \bar{\mathbf{y}}_n]^\top$. Here, $\bar{\mathbf{y}}_n$ is

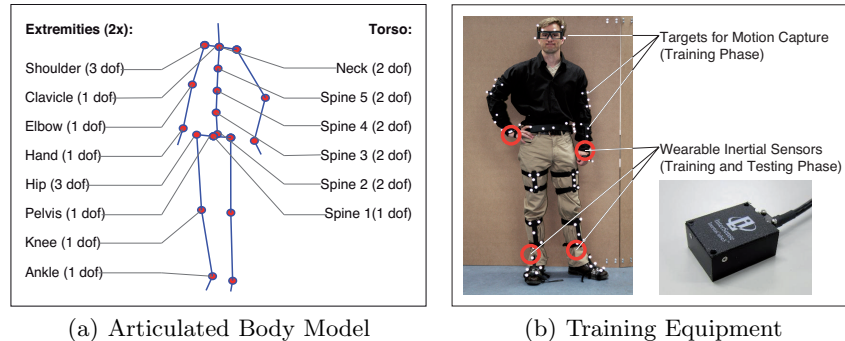


Fig. 2. a) Articulated body model consisting of five kinematic chains (arms, legs and torso). Each chain contains joints with one, two or three degrees of freedom (dof), totaling 35 dof. b) Actor equipped with reflective markers for optical motion capture and wireless inertial sensors. Markers are only worn during training.

the mean joint configuration for activity $n \in \{1, \dots, N\}$ and N is the number of considered activities.

We train a GPR model $\mathcal{G}_n = \{\mathbf{X}, \mathbf{Y}, \theta\}$ for each activity, modeling the non-linear functional relationship $f_n(\mathbf{x}_i) = \mathbf{y}_i$ between the sensor readings \mathbf{x}_i and the joint angles \mathbf{y}_i . This implies optimizing the so-called hyperparameters θ of the GPR model. In order to compute the pose estimate \mathbf{y}_* from new, unseen sensor data \mathbf{x}_* , we first need to determine the activity n_* performed by the actor. For this purpose, we train a multi-class SVM to classify the activities directly from the sensor measurements. The full-body pose estimate is then computed from the predictive mean of the GPR model corresponding to activity n_* .

2.2 Gaussian Process Regression

This section briefly introduces the general principles of GPR. Regression methods can be used to recover an unknown functional relationship $f(\mathbf{x}) = z$, given noisy observations of the output variable $z_i \in \mathbb{R}$ at certain values of the input variable $\mathbf{x}_i \in \mathbb{R}^d$. Knowledge about this relationship allows predicting z_* for a new value \mathbf{x}_* . For the following paragraphs, let $\{\mathbf{x}_i, z_i\}_{i=1 \dots M}$ be a set of training data pairs and let $\mathbf{z} = [z_1, \dots, z_M]$ denote the observed function values.

By definition, a Gaussian process is a collection of random variables, any finite number of which have a joint (multivariate) Gaussian distribution [21]. One can thus model a function f with a Gaussian process, by assuming that the function values \mathbf{z} are a sample from the multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{K})$. Here, \mathbf{K} is a covariance matrix with entries given by a covariance function k , such that $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. The type of covariance function that is used, along with the hyperparameters, determines properties of the sought functional relationship $f(\mathbf{x}) = z$. In this paper, we follow the typical choice of the squared exponential covariance function [22,23]:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \alpha^2 \exp\left(-\frac{1}{2\beta^2}(\mathbf{x}_i - \mathbf{x}_j)^2\right) + \gamma^2 \delta_{ij}, \quad (1)$$

where δ_{ij} equals 1 if $i = j$ and 0 otherwise. Here, the hyperparameters are $\theta = (\alpha, \beta, \gamma)$ [21]. γ represents the assumed noise level in the training data and β represents the length-scale. Suitable values for θ are obtained by optimization over a training dataset (section 2.3).

Adding a new pair of input and output values $\{\mathbf{x}_*, z_*\}$ to the initial distribution of function values \mathbf{z} , results again in a multivariate Gaussian distribution, since the function values are assumed to follow a Gaussian process:

$$\begin{bmatrix} \mathbf{z} \\ z_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{k}(\mathbf{x}_*) \\ \mathbf{k}(\mathbf{x}_*)^\top & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix}\right), \quad (2)$$

where $\mathbf{k}(\mathbf{x}_*)_i = k(\mathbf{x}_i, \mathbf{x}_*)$. The distribution in Equation 2 can be used to state the conditional probability $p(z_*|\mathbf{z})$ of the new output value z_* given the training data [21]. This probability, in turn, follows the Gaussian distribution $\mathcal{N}(\mu(\mathbf{x}_*), \sigma^2(\mathbf{x}_*))$, with mean and variance given by

$$\mu(\mathbf{x}_*) = \mathbf{k}(\mathbf{x}_*)^\top \mathbf{K}^{-1} \mathbf{z}, \quad (3)$$

$$\sigma^2(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_*)^\top \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}_*). \quad (4)$$

When only a new value \mathbf{x}_* is available and $z_* = f(\mathbf{x}_*)$ is unknown, the GPR estimate for z_* given the training data is the predictive mean (Equation 3). The predictive variance (Equation 4) can be seen as an assessment of prediction uncertainty. Note that prediction and uncertainty can be computed in closed form. While in the derivation above the functional relationship is assumed to be real-valued, the extension to the case of vector-valued functions is simply achieved by concatenating the one-dimensional regression result for each dimension of output space.

2.3 Learning Human Pose Models

We now describe our use of GPR to learn the functional relationship between inertial sensor data and full joint angle configurations. Given a matrix of sensor measurements \mathbf{X}_n for activity n and a matrix of full joint angle configurations \mathbf{Y}_n , we train a GPR model $\mathcal{G}_n = \{\mathbf{X}_n, \mathbf{Y}_n, \theta\}$ by minimizing the negative log marginal data likelihood with respect to the hyperparameters θ :

$$-\log p(\mathbf{Y}_n|\mathbf{X}_n, \theta) = \frac{1}{2} \text{tr}(\mathbf{Y}_n^\top \mathbf{K}^{-1} \mathbf{Y}_n) + \frac{1}{2} \log |\mathbf{K}| + c, \quad (5)$$

where c is a constant and the matrix \mathbf{K} is a function of the hyperparameters (see Equation 1). During minimization, the first term fits the model to the training data, whereas the second term prevents overfitting by penalizing complex models [22]. Optimization can be done offline, for instance using standard gradient-based approaches. We use the conjugate gradient method [24].

2.4 Classification of Sensor Measurements

Since we train individual GPR models \mathcal{G}_n for each activity performed by an actor, new wearable sensor data needs to be classified to one of the activities before pose estimation to choose the appropriate model. To this end, we propose to use a multi-class SVM classifier based on the standard one-against-all approach and using an RBF kernel [25]. The training data for the classifier is generated by concatenating the sensor measurement matrices \mathbf{X}_i over all activities in the training dataset into a matrix \mathcal{X} and supplying the known class labels for each sensor measurement in a vector \mathcal{Y} . We denote the SVM classification for a given input measurement \mathbf{x}_* as n_* .

2.5 Pose Estimation

Once a new, unseen sensor reading \mathbf{x}_* has been classified as belonging to activity n_* , we can recover the full joint angle configuration \mathbf{y}_* using the corresponding GPR model \mathcal{G}_{n_*} . To this end, we use the predictive mean of the model, in analogy to Equation 3 (extended to the vector-valued case), which is essentially a weighted combination of the joint angle configurations in the training dataset:

$$\tilde{\mathbf{y}}_{n_*}(\mathbf{x}_*) = \bar{\mathbf{y}}_{n_*} + \mathbf{Y}_{n_*}^\top \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}_*). \quad (6)$$

The notation $\tilde{\mathbf{y}}_{n_*}$ implies that this is only an estimate of the true, unknown joint configuration \mathbf{y}_* corresponding to \mathbf{x}_* , as given by the model for activity n_* . The equation can be efficiently evaluated online for each incoming sensor reading, since \mathbf{K}^{-1} can be precomputed.

In order to increase robustness with respect to scattered misclassifications, we extend this simple frame-by-frame classification scheme as follows. Let \mathbf{x}_* be the j -th sensor measurement in the sequence. All preceding frames in a sliding window of size w are classified using the SVM classifier. The pose estimate for frame j is then a linear combination of the predictions from all models, weighted by their relative contribution within the current window:

$$\tilde{\mathbf{y}}(\mathbf{x}_*) = \sum_{n=1}^N \frac{\lambda_n}{w} \cdot \tilde{\mathbf{y}}_n(\mathbf{x}_*), \quad (7)$$

where λ_n is the number of frames in the current window classified as activity n . The more frames in a window are classified as a certain activity, the higher the influence of the corresponding GPR model will be. When all frames are classified identically, only one GPR model is used for pose estimation. An inherent advantage of this weighting approach is that transitions between different activities are smooth in the pose reconstruction. When the activity performed by the actor changes from one to another, the influence of both corresponding GPR models is shifted gradually (for an evaluation, see section 3.4).

3 Experimental Results

We performed several types of experiments in order to quantitatively assess our pose estimation method. We first present separate studies for the activity classification and pose estimation parts (sections 3.2 and 3.3). An evaluation of the combined approach is given thereafter (section 3.4). Finally, we illustrate the influence of the number of inertial sensors and that of the training dataset size on pose estimation results (section 3.5).

3.1 Data Acquisition

For training and evaluation purposes, we use a database consisting of motion sequences that were recorded simultaneously with the wearable sensors and a motion capture system. We considered the following generic activities: *clapping hands*, *golfing*, *jumping on one foot*, *scratching head*, *walking* and *waving hand*. Each type of motion was recorded with three different actors and six repetitions per actor. Individual sequences are between 200 and 900 frames in length.

In our experiments, we used wireless InertiaCube sensors that output orientation values (yaw, pitch and roll), filtered from an internal accelerometer, a gyroscope and a compass (www.intersense.com). We discard the yaw values, since they give the absolute orientation of the actor with respect to the earth's magnetic field, which is unnecessary for learning general properties of movements. Unless stated otherwise, we used four orientation sensors in our experiments, such that the input space is eight-dimensional ($d = 8$). Figure 2(b) shows an actor equipped with inertial sensors and markers required by the infrared tracking system (www.ar-tracking.de).

3.2 Pose Estimation Accuracy

In a first set of experiments, we evaluated the pose estimation approach assuming we know which activity is being performed in the testing data. We measured the error between the joint angles estimated with our approach and the ground truth obtained from the optical motion capture system. The acquired database of motions was used for cross-validation experiments, i.e. repeated measurements were performed for every activity and actor, each time using one of the sequences in the dataset for testing and the remaining ones for training. The error for activity n is computed per frame as

$$e_{\text{ang}} = \frac{1}{D} \sum_{i=1}^D |[\mathbf{y}]_i - [\tilde{\mathbf{y}}_n]_i|, \quad (8)$$

where D is the number of joints, $\tilde{\mathbf{y}}_n$ is the vector of joint angles estimated using Equation 6 and \mathbf{y} is the corresponding ground truth joint configuration. The mean error over all frames in the testing sequence and the standard deviation are listed in Figure 3(a). Typical reconstruction errors lie between 4 and 7 degrees

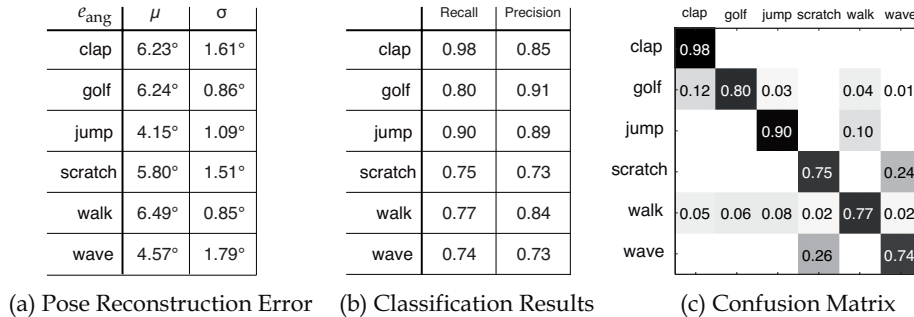


Fig. 3. Experimental results for pose reconstruction and pose classification, computed in a six-fold cross-validation study for 3 actors. a) Average angular error per joint per frame. b) Recall and precision for SVM-based pose classification for each activity class. c) Confusion matrix with average classification rates per combination of ground truth (rows) and prediction (columns).

per joint per frame, with standard deviations between 0.8 and 2.0 degrees. These values are comparable to the results of other state-of-the-art discriminative pose estimation methods that use higher-dimensional input data, e.g. obtained from restrictive camera setups [12,13,16,17,15]. Figure 4 visualizes the pose reconstruction result for two sample sequences, showing the predicted angles for the main joints, the corresponding ground truth and the angular error throughout the sequences.

3.3 Activity Classification Reliability

In a second series of experiments we separately evaluated the SVM-based activity classification approach. As in section 3.2, we performed a set of six-fold cross-validation studies for each activity in the training dataset, repeated for each of the three actors. Figure 3(b) indicates recall¹ and precision² for each activity class and Figure 3(c) shows the confusion matrix. Entry (i, j) of the matrix gives the relative number of frames belonging to class i that were classified as j .

The highest correct classification rates are achieved for *clapping* (98%) and *jumping* (90%), since these activities consist of repetitions of a few similar and characteristic poses. Around 20% of frames belonging to *walking* are classified as other activities. This can be explained by the multitude of poses taken during walking, some of which are similar to poses of other activities, e.g. jumping. The most significant misclassifications occur between *scratching head* and *waving*. The precision for these activities drops below 75% and the confusion matrix indicates relatively high numbers in the corresponding entries. However, the quality of the pose estimates for frames that are misclassified

¹ Recall: relative number of poses belonging to class n that are classified as such.

² Precision: relative number of poses classified as n that really belong to this class.

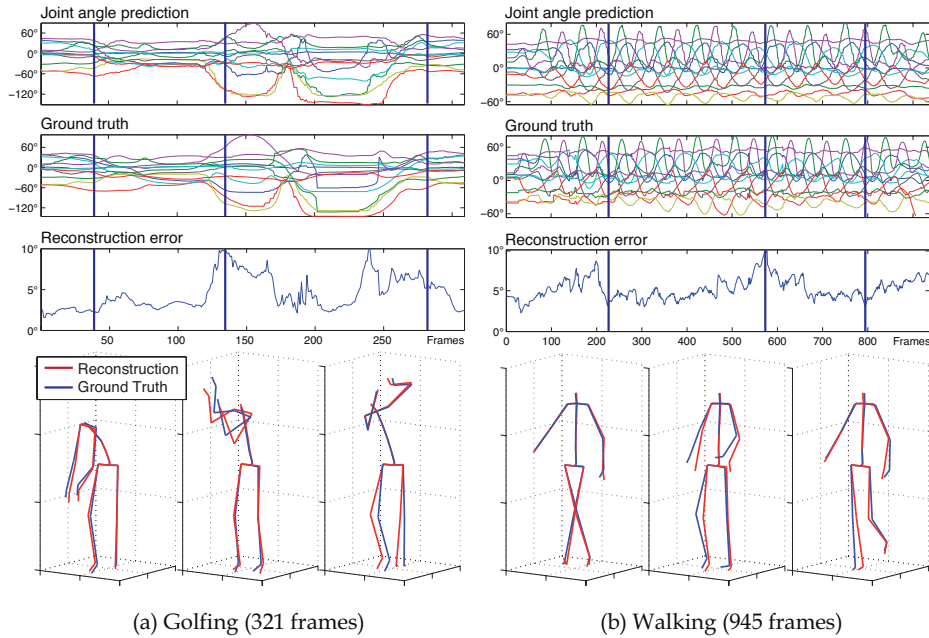


Fig. 4. Illustration of pose estimation results for two sample sequences. The top graph shows the reconstructed angles for the main body joints over the length of the sequence. The second graph is the corresponding ground truth. Reconstruction error is plotted in the third graph. Skeletons are displayed for the frames indicated with black vertical bars (red: reconstructed pose, blue: ground truth pose).

between similar activities is not severely affected, as shown in the following section.

3.4 Combined Activity Classification and Pose Estimation

In order to demonstrate the combined performance of the activity classification and pose estimation parts, we created sequences consisting of all the activities considered in this paper performed in a row. For each of the three actors, a six-fold cross-validation study was done. As described in section 2.5, pose estimates were computed by combining the pose predictions of multiple GPR models based on the classification of frames in a sliding window ($w = 15$).

Figure 5(a)-(c) shows the results for one of the sequences. The error does not significantly increase in areas where the pose was misclassified. Averaged over all experiments in this study, over 78% of frames are correctly classified. The angular error is 5.6° per joint on average. This value is equivalent to what can be achieved in the more controlled setting with prior knowledge of the performed activity (section 3.2). The SVM-based activity classification scheme therefore significantly increases the flexibility of our pose estimation approach without having a negative effect on pose reconstruction accuracy.

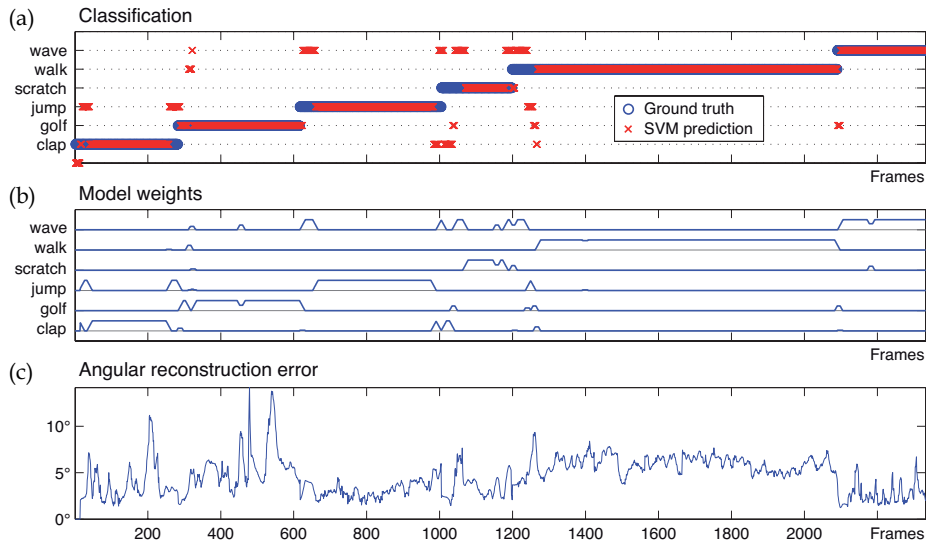


Fig. 5. Activity classification and pose estimation over a sequence of six consecutive movements. a) Predicted and ground truth activity classification over the length of the sequence (2380 frames). b) Weights between 0 and 1 for each of the learned GPR models derived from the classification in a sliding window. c) Pose reconstruction error.

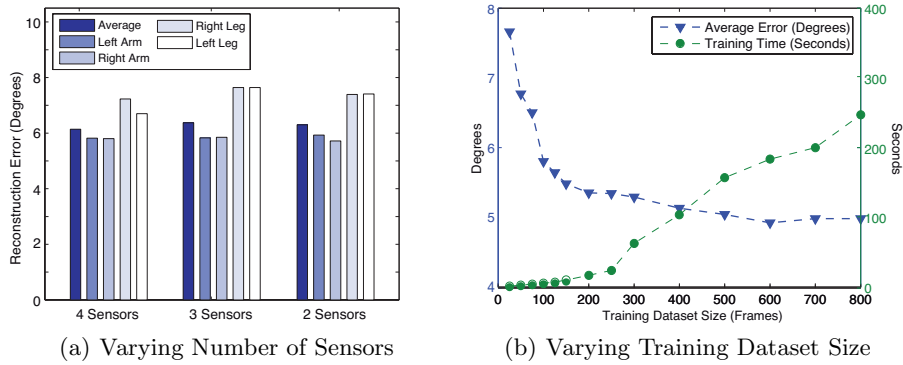


Fig. 6. a) Pose estimation error for varying numbers of inertial sensors. Average errors over all joints and errors for selected body segments are reported. Walking sequences were used in a six-fold cross-validation experiment. b) Influence of training dataset size on reconstruction error and training duration. Five walking sequences of one actor were used for training, subsampled to match the designated frame count.

3.5 Number of Sensors and Training Dataset Size

Although the inertial sensors do not impose a burden on a wearer, we evaluated the ability of our method to estimate full-body pose from less than four inertial sensors. For this experiment, we only used the *walking* sequences in our dataset, since all four extremities are moved equivalently in this activity. We selected the following sensor configurations: four sensors (left wrist, right wrist, left shinbone, right shinbone), three sensors (left wrist, right wrist, left shinbone) and two sensors (left wrist, right shinbone). Figure 6(a) shows the average angular error values achieved in our cross-validation study. The error for the legs is higher than that of the arms, since apparently the variation between individual walking sequences is higher for the legs. Most remarkably, the pose estimation error does not increase for any of the configurations with less than four inertial sensors, even when only two sensors are used. The experiments show that, for activities as regular as *walking*, our method is capable of predicting 35-dimensional full-body poses from input data of 6 or even 4 dimensions (3 or 2 sensors).

The size of the training dataset influences the processing time required to train a GPR model and the achievable reconstruction quality. Training duration increases with the number of frames, but can improve the ability of the GPR model to generalize to different motion styles. In order to evaluate this behavior, GPR models for the walking motion were learned with a varying training dataset size, ranging from 25 to 800 frames. Figure 6(b) gives the cross-validation results. While it takes considerably more time to train a GPR model with a dataset size of 800 frames, the reconstruction quality does not improve above a number of approximately 500 frames. In fact, when using up to 250 frames for training, significantly shorter training times (< 40 seconds in our implementation) can be achieved by the price of just a slight increase in average angular error ($\sim 5.5^\circ$).

4 Conclusion

We have presented a discriminative pose estimation method that allows reconstructing the full-body pose of a person, given only the limited observations of a few wearable inertial sensors attached to the person's limbs. Our method makes use of Gaussian Process Regression for learning the non-linear mapping between sensor measurements and full joint angle configurations. The system is trained on the specific movements of individual persons for each activity of interest. Since our method only requires an installed motion capture system for training, it can be applied in many settings, especially in complex, occlusive environments and outdoors. It is particularly suited for applications where movement patterns of individual persons are in focus, such as in ambulatory medical motion studies. Our experiments show that a low number of 2-4 inertial sensors are sufficient for reconstructing full-body pose for movements the system has been trained on. We plan to evaluate the feasibility of our approach in the medical scenario of motion analysis for epilepsy patients. In addition, the method will be extended in the direction of reconstructing more arbitrary movements, based on a training dataset of general human poses.

References

1. Aminian, K., Andres, E.D., Rezakhanlou, K., Fritsch, C., Robert, P.: Motion analysis in clinical practice using ambulatory accelerometry. In: Magnenat-Thalmann, N., Thalmann, D. (eds.) CAPTECH 1998. LNCS (LNAI), vol. 1537, pp. 1–11. Springer, Heidelberg (1998)
2. Cunha, J., Li, Z., Fernandes, J., Feddersen, B., Noachtar, S.: Movement quantification during epileptic seizures: a new technical contribution to the evaluation of the seizure semiology. In: International Conference of the IEEE EMBS (2003)
3. Jovanov, E., Milenkovic, A., Otto, C., Groen, P.D.: A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of NeuroEngineering and Rehabilitation* 2(1), 6 (2005)
4. Najafi, B., Aminian, K., Paraschiv-Ionescu, A.: Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *IEEE Transactions on Biomedical Engineering* (January 2003)
5. Luinge, H., Veltink, P., Baten, C.: Ambulatory measurement of arm orientation. *Journal of Biomechanics* 40(1), 78–85 (2007)
6. Roetenberg, D., Slycke, P., Veltink, P.: Ambulatory position and orientation tracking fusing magnetic and inertial sensing. *IEEE Transactions on Biomedical Engineering* 54(4), 883–890 (2007)
7. Zhou, H., Stone, T., Hu, H., Harris, N.: Use of multiple wearable inertial sensors in upper limb motion tracking. *Medical Engineering and Physics* 30, 123–133 (2008)
8. Grochow, K., Martin, S., Hertzmann, A., Popović, Z.: Style-based inverse kinematics. *ACM Transactions on Graphics* 23(3), 522–531 (2004)
9. Gupta, A., Chen, T., Chen, F., Kimber, D., Davis, L.: Context and observation driven latent variable model for human pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
10. Shon, A., Grochow, K., Hertzmann, A., Rao, R.: Learning shared latent structure for image synthesis and robotic imitation. *Neural Information Processing Systems (NIPS)* 18, 1233 (2006)
11. Urtasun, R., Fleet, D., Hertzmann, A., Fua, P.: Priors for people tracking from small training sets. In: IEEE International Conference on Computer Vision, ICCV (2005)
12. Agarwal, A., Triggs, B., Montbonnot, F.: Recovering 3d human pose from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 28(1), 44–58 (2006)
13. Fossati, A., Salzmann, M., Fua, P.: Observable subspaces for 3d human motion recovery. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), March 2009, pp. 1–8 (2009)
14. Urtasun, R., Darrell, T.: Sparse probabilistic regression for activity-independent human pose inference. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (January 2008)
15. Zhao, X., Ning, H., Liu, Y., Huang, T.: Discriminative estimation of 3d human pose using gaussian processes. In: International Conference on Pattern Recognition (ICPR), pp. 1–4 (2008)
16. Sminchisescu, C., Kanaujia, A., Li, Z., Metaxas, D.: Discriminative density propagation for 3d human motion estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2005)
17. Sun, Y., Bray, M., Thayananthan, A., Yuan, B., Torr, P.: Regression-based human motion capture from voxel data. In: British Machine Vision Conference, BMVC (2006)

18. Okada, R., Soatto, S.: Relevant feature selection for human pose estimation and localization in cluttered images. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 434–445. Springer, Heidelberg (2008)
19. Chai, J., Hodgins, J.: Performance animation from low-dimensional control signals. In: International Conference on Computer Graphics and Interactive Techniques, pp. 686–696 (2005)
20. Zhu, R., Zhou, Z.: A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 12(2), 295–302 (2004)
21. Rasmussen, C., Williams, C.: *Gaussian processes for machine learning* (2006)
22. Ek, C., Torr, P., Lawrence, N.: Gaussian process latent variable models for human pose estimation. *Machine Learning for Multimodal Interaction*, 132–143 (2008)
23. Lawrence, N.: Gaussian process latent variable models for visualisation of high dimensional data. *Neural Information Processing Systems, NIPS* (2004)
24. Lawrence, N.: *Gaussian process library* (2009), <http://www.cs.man.ac.uk/neill/gplvm/>
25. Abe, S.: *Support Vector Machines for Pattern Classification*. Springer, Heidelberg (2005)