3D Reconstruction of Human Motion from Monocular Image Sequences

Bastian Wandt, Hanno Ackermann and Bodo Rosenhahn

Abstract—This article tackles the problem of estimating non-rigid human 3D shape and motion from image sequences taken by uncalibrated cameras. Similar to other state-of-the-art solutions we factorize 2D observations in camera parameters, base poses and mixing coefficients. Existing methods require sufficient camera motion during the sequence to achieve a correct 3D reconstruction. To obtain convincing 3D reconstructions from arbitrary camera motion, our method is based on a-priorly trained base poses. We show that strong periodic assumptions on the coefficients can be used to define an efficient and accurate algorithm for estimating periodic motion such as walking patterns. For the extension to non-periodic motion we propose a novel regularization term based on temporal bone length constancy. In contrast to other works, the proposed method does not use a predefined skeleton or anthropometric constraints and can handle arbitrary camera motion.
We achieve convincing 3D reconstructions, even under the influence of noise and occlusions. Multiple experiments based on a 3D error metric demonstrate the stability of the proposed method. Compared to other state-of-the-art methods our algorithm shows a significant improvement.

Index Terms—human motion, structure and motion, factorization, 3D reconstruction.

1 Introduction

The recovery of 3D human poses in monocular image sequences is an inherently ill-posed problem, since the observed projection on a 2D image can be explained by multiple 3D poses and camera positions. Nevertheless experience allows a human observer to estimate the pose of a human body, even with a single eye. The purpose of this paper is to achieve a correct 3D reconstruction of human motion from monocular image sequences as shown in Fig. 1.

Recent works considering the non-rigid structure from motion problem (e.g. [2], [3], [4]) work well as long as there is a camera rotation around the observed object. Due to ambiguity in camera placement and 3D shape deformation they fail in realistic scenes such as a fixed camera filming a person walking by as shown in Fig. 2. Several single image pose recovery approaches (e.g. [5], [6], [7], [8]) use strong constraints on the observed shape to overcome this problem. These methods achieve acceptable results but are too restrictive for general 3D reconstructions as they limit the solution to a predefined skeleton. Obviously, applying these single image approaches for image sequences results in an unstable 3D motion reconstruction.

In this article, we use a trilinear factorization approach similar to [3], [8], [9] and [6]. We assume that a set of feature points on the skeleton of the person is tracked throughout the sequence. Our goal is to decompose it into three factors for camera motion, base poses and mixing coefficients. Different to [9] and [3], we keep the second factor fixed which corresponds to 3D structure, similar to [8] and [6]. Furthermore, we propose to regularize the third factor, commonly interpreted as the mixing coefficients: Firstly, we impose a prior well suited for periodic motion. Secondly, constraints on the limb lengths are applied. As opposed to [8] and [6] where lengths or relations of particular limbs need to be *a-priorly* known, we *constrain* the limbs lengths to be invariant.

We demonstrate that our algorithm works on motion capture data (CMU MoCap [10], HumanEva [11]) as well as on challenging real world data as for example the KTH Football Dataset [1] shown in Figure 1. Additionally we are analyzing the influence of the number of base poses and the regularization factor on the reconstruction result. Furthermore we demonstrate that our algorithm is robust to noise and also able to handle occlusions and reconstruct the occluded body parts correctly. We show that it can also be used for motion classification tasks.

Our method allows to correctly reconstruct 3D human motion from feature tracks in monocular image sequences with arbitrary camera motion. It does not use a predefined skeleton or anthropometric constraints. Additionally it can handle occlusions and noisy data. Summarizing, the contributions of this article are as follows:

- A periodic model for the mixing coefficients for periodic and quasi-periodic motions such as walking is introduced.
- We propose a novel regularization term for non-periodic motions.

2 Related Work

The factorization of a set of 2D points tracked over a sequence of images was proposed by Tomasi and Kanade [12]. It rests upon the idea that the input data is decomposed into two sets of variables, one of which is associated with the motion parameters, the other with the coordinates of the rigid 3D structure. This algorithm was generalized to deforming shapes by Bregler et al. [13] by expressing the observed

• *B. Wandt, H. Ackermann and B. Rosenhahn are with the Institut für Informationsverarbeitung, Leibniz University Hannover.*
E-mail: wandt@tnt.uni-hannover.de

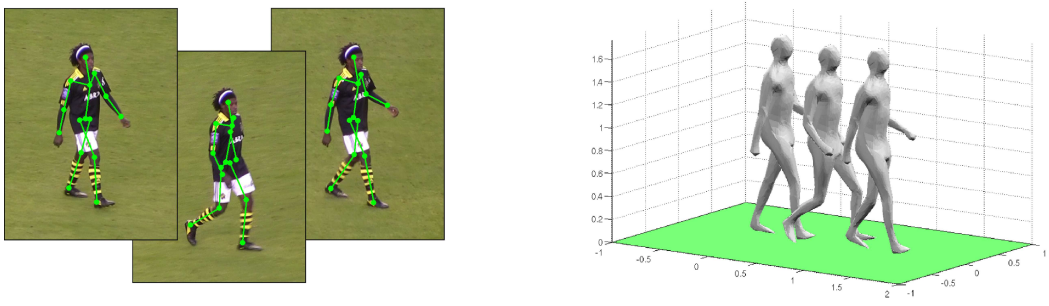


Fig. 1. Real world scenario of KTH database [1]. Left: frames 115, 136 and 143 of Sequence 1 from Football Dataset II. Right: 3D reconstruction using our proposed method

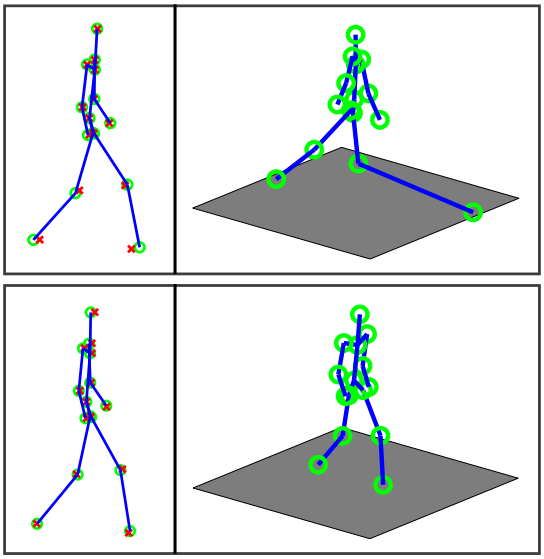


Fig. 2. 3D reconstruction (green circles, blue lines) and ground truth data (red crosses). Top: Using approach of Gotardo and Martinez [4]. Most non-rigid structure from motion approaches with no rotation and unknown base poses fail, although they produce a small reprojection error (left). From other perspectives (right) a wrong reconstruction can be observed. Bottom: Our approach. Correct reconstruction in all views.

shape in any particular image as a linear combination of multiple rigid basis shapes. Xiao et al. [14] showed that this decomposition is non-unique. They extended a well-known problem of rigid 3D reconstruction, namely the problem of self-calibration, to the non-rigid case. Akther et al. [15] showed that the solution proposed in [14] still is ambiguous. Torresani et al. [16], [17], [18] independently proposed to avoid the troublesome step of non-rigid self-calibration by imposing a Gaussian prior on the linear mixing coefficients. Akhter et al. [15] built on this idea and fixed the linear coefficents in advance by selecting them from a cosine function. This approach both adds a strong prior that the non-rigidity can be explained by periodic base function, and it also determines in advance the frequencies that the observations need satisfy. Gotardo and Martinez later extended this approach by assuming smoothly moving cameras [2], [3], [4] and 3D points. Akhter et al. [19] use a bilinear spatiotemporal basis to apply it to graphics tasks including labeling, gap-filling, de-noising, and motion touch-up. Zhu et al. [20] proposed to use a small number of keyframes to avoid ambiguities between point and camera motion. Li introduced an approach based on L_1-minimization [21] where the number of mixing coefficients that are non-zero is minimized.

Anthropometric priors have been used before in multi camera human motion capture applications. Theobalt et al. [22], [23] use an initialization step to estimate a template of the skeletal structure through a silhouette-based fitting process. Li et al. [24] use bone length constraints from a template to reconstruct missing markers in motion capture data. A relaxation of fixed bone lengths assumptions is proposed by Kovar et al. [25]. They allow a slight change of the template model to avoid the footskate effect resulting from noisy motion capture data. Hasler et al. [26] enforce mesh constancy for 3D body shape estimation and thus implicitly enforce bone length constancy.

Several works have been proposed regarding the 3D reconstruction of human poses given single images only [5], [6], [7], [8]. State-of-the-art methods such as the work of Ramakrishna et al. [6] represent a 3D pose by a linear combination of a set of base poses that are learned from motion databases. They are minimizing the reprojection error using the sum of squared limb lengths as constraint. This is a very weak constraint considering all the possible but incorrect poses which satisfy this constraint. Wang et al. [8] extended that model. Different to [6] they enforce the proportions of eight selected limbs to be constant. However, limb proportions differ from one person to another.

While all these approaches assume multi camera setups or use fixed bone lengths priors, to the best of our knowledge this is the first work using a temporal bone lengths constancy prior for monocular 3D reconstruction of human motion.

3 Our Approach

Our approach consists of three main steps (see Figure 3). First we assume, that every 2D motion sequence can be factorized in a camera model and a series of 3D poses (section 3.1), like in standard structure from motion approaches. The 3D poses are composed of a linear combination of base poses, that are retrieved by a PCA on different motion databases (section 4.1). To model periodic motion (eg.

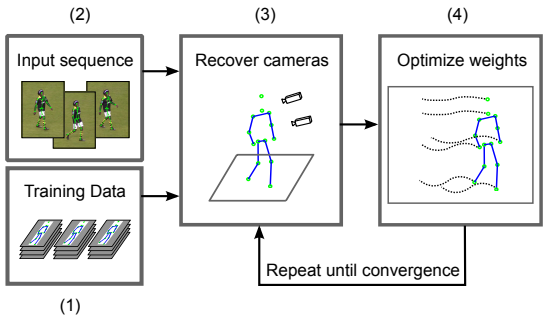


Fig. 3. Our method. (1) First we are learning 3D base poses from training data. (2) Input sequence. (3) Cameras are recovered from estimated 3D poses and 2D poses. (4) Weights for base poses are calculated by minimizing the reprojection error. Steps (3) and (4) are alternated until the algorithm converges

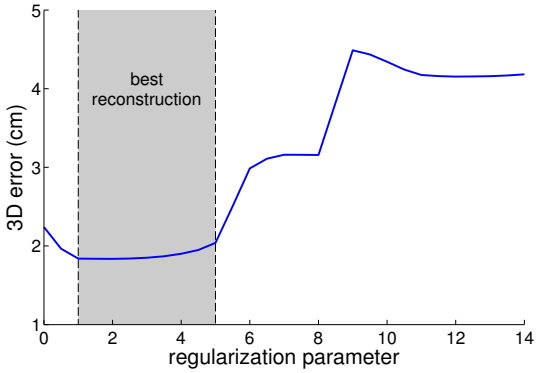


Fig. 4. Influence of the camera path regularization on the reconstruction result. A low value for the regularization parameter γ avoids flips while a high value enforces a static camera. The best results are obtained for values between 1 and 5.

walking and running), we show that it is possible, to assume a periodic weight for the base poses to significantly reduce the number of variables, that have to be calculated (section 3.3). Our algorithm (section 3.5) is alternatingly recovering the camera matrices (section 3.2) and the 3D poses. Our extension to non-periodic motion calculates the weights for the base poses for each frame. We handle the large number of variables by using a regularization term enforcing bone length constancy over time. This leads to a highly realistic 3D reconstruction of different types of non-periodic motion (section 3.4).

3.1 Factorization model

A single 3-dimensional pose $\boldsymbol{P} \in \mathbb{R}^{4\times a}$ with a joints in homogeneous coordinates can be written as a linear combination of k previously learned base poses $\boldsymbol{Q}_l \in \mathbb{R}^{4\times a}$

$$\boldsymbol{P} = \boldsymbol{Q}_0 + \sum_{l=1}^{k} \boldsymbol{\theta}_l \boldsymbol{Q}_l, \tag{1}$$

where $\boldsymbol{Q}_0$ is the mean pose of all poses used for training and $\boldsymbol{\theta}_l \in \mathbb{R}^{4\times 4}$ is the weight matrix for the base pose $\boldsymbol{Q}_l$. With ϑ_l as the scalar weight for the l-th base pose each $\boldsymbol{\theta}_l$ has the form

$$\boldsymbol{\theta}_l = \begin{pmatrix} \vartheta_l \boldsymbol{I}_3 & \\ & 0 \end{pmatrix}, \tag{2}$$

where $\boldsymbol{I}_3$ is the 3×3 identity matrix. Note that only the coordinates in the mean pose $\boldsymbol{Q}_0$ are describing a point in homogeneous coordinates, while $\boldsymbol{Q}_{1,\dots,k}$ are directions that define *deformations*. By stacking poses we can write a 3D sequence as $\boldsymbol{W} \in \mathbb{R}^{4f\times a}$ of f images, with $\boldsymbol{P}_{1,\dots,f}$ as the poses in frames $1, \dots, f$

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{P}_1 \\ \vdots \\ \boldsymbol{P}_f \end{pmatrix}. \tag{3}$$

With Eq. (1) we can do a factorization

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{Q}_0 + \sum_{l=1}^{k} \boldsymbol{\theta}_{l,1} \boldsymbol{Q}_l \\ \vdots \\ \boldsymbol{Q}_0 + \sum_{l=1}^{k} \boldsymbol{\theta}_{l,f} \boldsymbol{Q}_l \end{pmatrix} = \boldsymbol{\Theta} \begin{pmatrix} \boldsymbol{Q}_0 \\ \boldsymbol{Q}_1 \\ \vdots \\ \boldsymbol{Q}_k \end{pmatrix} = \boldsymbol{\Theta}\boldsymbol{Q}, \tag{4}$$

where $\boldsymbol{\Theta} \in \mathbb{R}^{4f\times 4k}$ contains the weight matrices $\boldsymbol{\theta}_l$.

The projection of a 3D pose $\boldsymbol{P}_i$ in the i-th frame to a 2D pose $\boldsymbol{P}_{i,2D} \in \mathbb{R}^{2\times a}$ is done by the camera matrix $\boldsymbol{M}_i \in \mathbb{R}^{2\times 4}$

$$\boldsymbol{P}_{i,2D} = \boldsymbol{M}_i \boldsymbol{P}_i. \tag{5}$$

To project the whole 3D sequence described by the matrix $\boldsymbol{W}$, the camera matrix $\boldsymbol{M} \in \mathbb{R}^{2f\times 4f}$ is used. Let $\boldsymbol{M}$ be a sparse block diagonal matrix

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{M}_1 & & \\ & \ddots & \\ & & \boldsymbol{M}_f \end{pmatrix}. \tag{6}$$

The factorization of a 2D sequence given by the matrix $\boldsymbol{W}_{2D} \in \mathbb{R}^{2f\times a}$ can now be written as

$$\boldsymbol{W}_{2D} = \boldsymbol{M}\boldsymbol{\Theta}\boldsymbol{Q}. \tag{7}$$

When dealing with missing feature points (for example caused by partly occluded body parts) the equations corresponding to these feature points can be excluded from the optimization. This is further explained and evaluated in section 4.7. This model is very similar to the models proposed by [13], [9] and [3]. While they are fixing $\boldsymbol{\Theta}$ and optimize for $\boldsymbol{M}$ and $\boldsymbol{Q}$, our approach is using a previously learned $\boldsymbol{Q}$ and optimize for the weights $\boldsymbol{\Theta}$ like [6] and [8] did for single images.

3.2 Camera Parameter Estimation

To reconstruct the camera parameters we are assuming a weak perspective camera. The pose in the i-th frame $\boldsymbol{w}_{2D}^i$ can be factorized with the above notation as

$$\boldsymbol{w}_{2D}^i = \boldsymbol{M}_i \boldsymbol{\Theta}_i \boldsymbol{Q}, \tag{8}$$

where $\boldsymbol{\Theta}_i \in \mathbb{R}^{4\times 4k}$ denotes the weight matrix for this frame. For the estimation of the camera parameters we assume the 3D pose described by $\boldsymbol{\Theta}_i\boldsymbol{Q}$ to be known. The solution for the camera matrices for each frame can be obtained by least squares minimization of the reprojection error

$$\min_{\boldsymbol{M}_K} \left|\left| \boldsymbol{w}_{2D}^i - \boldsymbol{M}_i \boldsymbol{\Theta}_i \boldsymbol{Q} \right|\right|_F. \tag{9}$$

In our model each $\boldsymbol{M}_i$ describes a weak perspective camera. Therefore we give the optimization algorithm used to solve

Eq. (8) correct starting values for $\boldsymbol{M}_i$ which satisfy the constraints for a weak perspective camera. We rewrite Eq. (8) with $(\boldsymbol{\Theta}_i\boldsymbol{Q})^+$ as the right-inverse of $\boldsymbol{\Theta}_i\boldsymbol{Q}$

$$\boldsymbol{M}_i = \boldsymbol{w}_{2D}^i(\boldsymbol{\Theta}_i\boldsymbol{Q})^+. \tag{10}$$

The scale parameter s of the weak perspective camera can be determined by

$$s = \frac{1}{2}\sqrt{||\boldsymbol{M}_{i,1}||^2 + ||\boldsymbol{M}_{i,2}||^2}, \tag{11}$$

with $\boldsymbol{M}_{i,1}$ as the first row and $\boldsymbol{M}_{i,2}$ as the second row of $\boldsymbol{M}_i$. We receive an unscaled camera matrix by dividing $\boldsymbol{M}_i$ by s. Next we orthonormalize the first 2×3 block of the unscaled matrix with the help of a singular value decomposition, where all singular values are set to 1. Recombining the orthonormalized block with the scale s and the last column of the unscaled camera matrix gives a good estimation for the starting values.

If we reconstruct the cameras for each frame separately the camera orientations can *flip*. I.e. the camera matrix of the flipped camera not only describes a weak perspective projection but also a reflection at the origin of the coordinate system. As this effect rarely occurs it can be easily avoided by penalizing rapid changes in the camera path. Therefore we propose a regularization term that calculates the difference between the current camera matrix $\boldsymbol{M}_i$ and the previous camera matrix $\boldsymbol{M}_{i-1}$

$$r_{K,i} = \gamma\,||\boldsymbol{M}_i - \boldsymbol{M}_{i-1}||_F\,, \tag{12}$$

with γ as regularization parameter. Eq. (12) is equivalent to forward differences. While central differences are also possible, we will show in Section 3.5 that using forward differences leads to a much faster optimization.

The whole minimization problem can now be written as

$$\min_{\boldsymbol{M}_i} \left|\left|\boldsymbol{w}_{2D}^i - \boldsymbol{M}_i\boldsymbol{\Theta}_i\boldsymbol{Q}\right|\right|_F + r_{K,i}. \tag{13}$$

While the regularization term also allows smoothing of the camera path, its sole purpose is to avoid camera flips. The regularization is not necessary in most cases as the flips only occur very rarely. Setting the parameter γ to a high value would result in a static camera. Therefore we set γ to a very low value where it avoids flips and only slightly effects the camera path as shown in a small experiment in Fig. 4. Although the reconstruction error without the camera regularization ($\gamma = 0$) seems low there are three flips in the camera path causing wrong 3D reconstructions. In contrast to Zhu et al. [20] who solved the problem by using keyframes, we do not assume any prior camera positions or poses.

Considering the entries in

$$\boldsymbol{M}_i = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \end{pmatrix} \tag{14}$$

we can enforce a weak perspective camera by the constraints

$$m_{11}^2 + m_{12}^2 + m_{13}^2 - (m_{21}^2 + m_{22}^2 + m_{23}^2) = 0 \tag{15}$$

and

$$m_{11}m_{21} + m_{12}m_{22} + m_{13}m_{23} = 0. \tag{16}$$

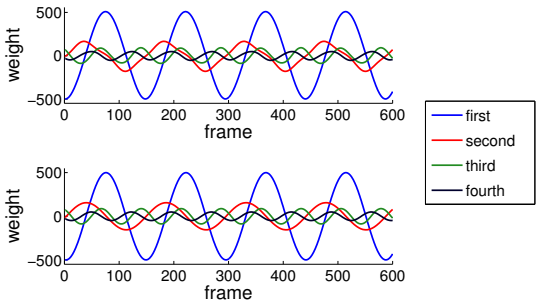


Fig. 5. Comparison of ground truth coefficients of the first four base poses (top) with fitted periodic function (bottom) using the data set of N. Troje [27].

3.3 Periodic Motion

With the camera matrix $\boldsymbol{M}$ calculated as described in Section 3.2 the weights $\boldsymbol{\Theta}$ for the base poses can be reconstructed. Trying to optimize the reprojection error for all variables in $\boldsymbol{\Theta}$ fails, as there are to many degrees of freedom. For periodic motion the number of unknowns can be reduced by using a sine function to model the temporal behavior of the weights in $\boldsymbol{\Theta}$.

Figure 5 shows the weights of the first four base poses of a gait sequence and the corresponding fitted sine functions. For this specific sequence the mean absolute error of the periodic reconstruction to the ground truth data is $2.05mm$. It verifies the results obtained by N. Troje in [27], [28]. They used the same periodic assumption to describe human gait patterns and did an extensive research on a large set of persons. These observations can be made with running motions as well. So the periodic assumption appears to be appropriate for periodic motion.

As shown in Section 3.1 the number of unknowns in $\boldsymbol{\Theta}$ equals fk. By modelling the temporal behavior of ϑ as

$$\vartheta(t) = \alpha\sin(\omega t + \varphi) \tag{17}$$

the number of unknowns can be decreased to $3k$. Note that the number of variables does not depend on the number of frames anymore yet only on the number of base poses. We can thus minimize the 2D reprojection error

$$\min_{\alpha,\omega,\varphi} ||\boldsymbol{W}_{2D} - \boldsymbol{M}\boldsymbol{\Theta}\boldsymbol{Q}||_F\,. \tag{18}$$

Note, that the objective function in Eq. (18) is nonlinear and nonconvex.

The use of sine functions to approximate human motion was firstly proposed by Troje et al. [27], [28]. We use a similar representation in Eq. (17) which can be motivated from [15], since a sine function can be represented by a linear combination of DCT bases. Modeling a structure from motion problem in *trajectory space* using DCT bases, requires a manually set or estimated number of DCT bases which mostly results in too many degrees of freedom. In Fig. 2 we show that 3D reconstructions of approaches derived from [15] (e.g. Gotardo and Martinez [4]) fail when there is no sufficient camera motion in the sequence (i.e. low reconstructibility as defined by Park et al. [9]). Combining the use of a single sine function as weight as proposed by

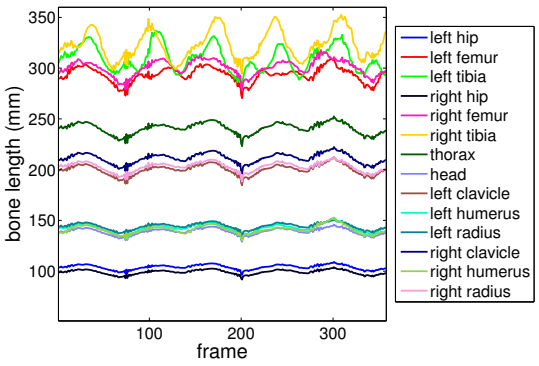


Fig. 6. Temporal behavior of bone lengths obtained by unconstrained optimization. The maximal variation is about 40mm. Computed on CMU MoCap (subject7/walk1).

N. Troje in [27], [28] with trained base poses results in a low number of variables and plausible 3D reconstructions.

3.4 Non-Periodic Motion

To model non-periodic motion, periodic functions for the weights of the base poses are not applicable anymore. Trying to optimize all weights at once without constraints gives good results for the 2D reprojection, but does not ensure a realistic 3D reconstruction. Figure 6 shows the temporal behavior of the bone lengths using the unconstrained optimization. There are variations in lengths up to $40mm$. This is caused by a slightly wrong initial camera position, which the optimizer later tries to compensate by weighting base poses wrongly. It results in a 3D reconstruction where unrealistic bone length changes occur. To compensate this we propose a regularization term, which holds the bone lengths constant over time. Different to [6] and [8] we are not using bone length constraints. Such a constraint would restrict the model to a particular person.

The length of a bone is defined by the euclidean distance between the 3D joint coordinates of that bone. These can be directly obtained from the 3D reconstruction described by $\boldsymbol{\Theta Q}$. We denote the length of bone s as

$$b_s = ||\boldsymbol{j}_{s,2} - \boldsymbol{j}_{s,1}||_2 \,, \tag{19}$$

where $\boldsymbol{j}_{s,1}$ and $\boldsymbol{j}_{s,2}$ are the coordinates of the endpoints of that bone. We want to hold the bone lengths nearly constant over time to ensure a realistic reconstructed skeleton, but do not want to be too restrictive to the optimizer. In other words the bone lengths should not change much. In the optimal case they are not changing at all. We are using the variance of the length changes over time of each bone as a measure. To build the regularization term r_B, we sum the variances $\mathrm{Var}(\bullet)$ of all bone lengths over time

$$r_B = \beta \sum_i \mathrm{Var}(b_i), \tag{20}$$

with β as the regularization parameter. This regularizer holds the bone length constant but is not fixing it to a specific value. Note, that the same variance for a short bone allows larger relative changes in length than for longer bones. Using the relative variance, i.e. normalizing $\mathrm{Var}(b_i)$ by the mean of the bone length avoids this effect. However, as experimentally shown in Fig. 11 there is no significant difference in using the variance or the relative variance. Due to this finding and to keep computational effort as low as possible, all experiments are using Eq. (20) as regularizer.

The optimization problem can be written as

$$\min_{\boldsymbol{\Theta}} ||\boldsymbol{W}_{2D} - \boldsymbol{M\Theta Q}||_F + r_B. \tag{21}$$

For the minimization of Eq. (18), the parameters α, ω and φ of the functions defined by Eq. (17) are estimated. Here, for minimizing the nonlinear and nonconvex objective function in Eq. (21) we can estimate the coefficients $\boldsymbol{\Theta}$ of the linear combination $\boldsymbol{\Theta Q}$ subject to the constraints defined by Eq. (20) since $\boldsymbol{Q}$ defines the prior knowledge on the possible deformations of human shapes.

The number of variables equals fk, i.e. it linearly depends on the number of frames f. Using a skeleton with 15 joints gives the same number of 2D/3D point correspondences per frame. By keeping the number of used base poses k low there are more equations than unknowns.

3.5 Algorithm

To estimate the $f+1$ sets of variables $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_f$ and $\boldsymbol{\Theta}$ we alternatingly optimize for each of the sets while keeping the others fixed. The optimization of each camera matrix $\boldsymbol{M}_j$, $j = 2, \ldots, f$, requires the regularization terms $r_{K,j}$ and $r_{K,j+1}$. If we use central differences in Eq. (12), we need to optimize all the sets $\boldsymbol{M}_j$, $j = 1, \ldots, f$, simultaneously. Using the proposed forward differences allows to sequentially estimate them, i.e. given $\boldsymbol{M}_1$ we estimate $\boldsymbol{M}_2$, then $\boldsymbol{M}_3$ etc. The precision of the estimated solution is hardly affected while the computation time in our experiments reduces by the factor 5. Shape parameters are estimated by minimizing Eq. (18) in the case of periodic motion, and Eq. (21) in the case of non-periodic motion, respectively. These constrained nonlinear and nonconvex problems are optimized using a second-order gradient descent algorithm.

In the first iteration we use the mean pose as initialization. This means setting all values in $\boldsymbol{\Theta}$ to zero except the ones weighting the mean pose $\boldsymbol{Q}_0$. With that the initial cameras are estimated framewise as described in Section 3.2. The optimization for the weights of the base poses follows. This step is depending on whether we are using the periodic (Section 3.3) or the non-periodic model (Section 3.4). The last two steps are repeated until the reprojection error is not changing anymore.

Alternating the estimation of the parameter sets can be seen as a variant of a block-coordinate descent by formulating one objective function for all parameters:

$$f(\boldsymbol{M}_1, \ldots, \boldsymbol{M}_f, \boldsymbol{\Theta}) = f(\boldsymbol{\Theta}) + \sum_{i=2}^{f} g_i(\boldsymbol{M}_i), \tag{22}$$

where

$$f(\boldsymbol{\Theta}) = \quad ||\boldsymbol{W}_{2D} - \boldsymbol{M\Theta Q}||_F + r_B \tag{23}$$

$$g_i(\boldsymbol{M}_i) = \quad r_{K,i}. \tag{24}$$

The objective function for the periodic reconstruction can be formulated in the same way. Convergence of coordinate gradient descent is guaranteed if the joint objective

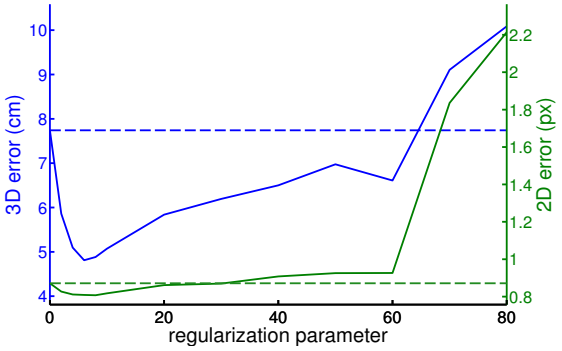


Fig. 7. 2D reprojection error and 3D reconstruction error with different regularization parameter β. While the 2D error is not changing much or getting worse, the 3D error gets significantly better at most parameter values. Computed on CMU MoCap (subject35/walk1). Qualitatively there is no difference between different motion categories.

function is strongly-convex [29]. More recently, results on convergence were established if at least one of the terms is convex (see, e.g. [30]). Since neither of the terms in Eq. (22) is convex, convergence cannot be guaranteed. However, we will experimentally show that the proposed algorithm converges to a reasonable local minimum in Section 4.5.

Algorithm 1 Recover camera and shape

$\boldsymbol{Q} \leftarrow$ base shapes
while no convergence **do**
  for $t = 1 \rightarrow f$ **do**
    calculate starting values for $\boldsymbol{M}_t$
    optimize $||\boldsymbol{w}_{2D}^i - \boldsymbol{M}_t \boldsymbol{\Theta}_i \boldsymbol{Q}||_F + r_{K,i}$
    insert $\boldsymbol{M}_t$ in $\boldsymbol{M}$
  end for
  optimize $\|\boldsymbol{W} - \boldsymbol{M}\boldsymbol{\Theta}\boldsymbol{Q}\|_F + r_B$
end while

4 Experimental Results

To evaluate our method, we were using three different databases: CMU MoCap [10], HumanEva [11] and KTH Football [1]. We trained base poses (see Section 4.1) of different motion categories, for example walking, jogging, running and jumping to demonstrate the generality of our method.

Instead of the reprojection error we define a 3D error e as evaluation criterion

$$e = \frac{1}{f} ||\boldsymbol{W}_{in} - \boldsymbol{W}_{rec}||_F \,, \tag{25}$$

with $\boldsymbol{W}_{in}$ as the ground truth 3D data and $\boldsymbol{W}_{rec}$ as the reconstruction. To compare sequences of different lengths, we are dividing the error by the number of frames f. As shown in Section 2, the reprojection error is a bad criterion for judging a 3D reconstruction. Therefore it is important to use the 3D error instead of the reprojection error when evaluating 3D reconstructions. For example with our bone length regularizer we achieve a worse reprojection error but a significantly better 3D reconstruction (see Figure 7). While the reprojection error remains nearly constant for values of the regularization parameters up to 60, the 3D error is getting better. Only for very high values both errors are getting worse. This is further evaluated in Section 4.4.

4.1 Learning base poses

For learning the base poses we were using different databases: the well-known CMU Motion Capture Database [10], the HumanEva dataset [11] and as a real world example the KTH Football Dataset II [1]. These three databases are using slightly different joint annotations, so it is important to learn the base poses for each database separately.

We are learning the base poses by stacking pose vectors of all frames and executing a PCA on this matrix. For each of the used motion categories a linear combination of the first ten eigenvectors obtained by the PCA is enough to cover more than 99% of the variance in the dataset. It is also possible to learn base poses for multiple motions at once. If doing so, the number of base poses should be increased to be able to fully cover all possible motions. The influence of the used number of base poses on the reconstruction result is evaluated later in Section 4.3.

4.2 Periodic Motion

As shown in Section 3.3, the number of unknowns can be reduced when using periodic base functions. This results in a much faster solving of the optimization problem. Figure 21 shows some frames of a reconstruction of a gait sequence by just using four base poses. Even with only 12 unknowns to optimize the reconstruction is close to the real 3D data. Note that the number of variables does not depend on the number of frames. That means that the computational effort does not increase much if longer sequences are used as long as the motion does not change. The reconstruction of the shown sequence of 450 frames took about 15 seconds, which is about two magnitudes faster than the non-periodic reconstruction on the same sequence. For periodic motion this method is a fast and efficient way for the 3D reconstruction. Comprehensive results of the periodic reconstruction on different periodic motions can be seen in Section 4.6.

If bone length constancy is used to additionally regularize the reconstructions we observed no improvement. The reason is that the periodic assumption is such a strong prior that an additional regularization term has no effect. Setting the weight of the bone length regularizer too high results in a local minimum where the skeleton is not moving at all and stays in the mean pose.

4.3 Number of base poses

One of the main questions is how many base poses should be used to achieve a good reconstruction. More base poses can model more deformation but using too many can cause unnatural deformation.

It is important to notice that all motions used for training lie in the space spanned by the base poses. However, not every linear combination of the base poses defines a correct human pose. In fact, every base pose allows for some non-human deformations. Thus the more base poses are used for the reconstruction, the more distorted the reconstruction gets. As shown in Figure 8 using 4 to 10 base poses results in the best reconstructions for periodic and non-periodic

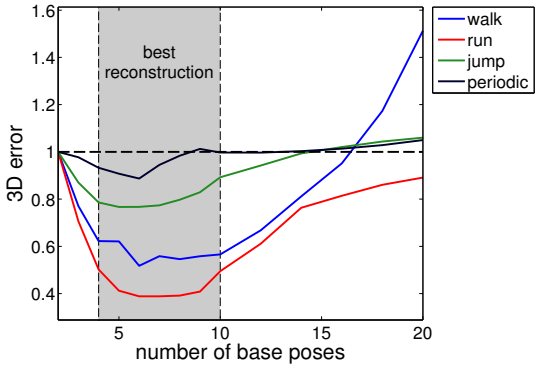


Fig. 8. Influence on the number of used base poses on the 3D error using the non-periodic reconstruction (labels: walk, run, jump) and the periodic reconstruction (label: periodic). The number of used base poses is crucial for a good 3D reconstruction. Using more than 10 base poses for each motion category worsens the reconstruction error. For better visibility, the errors are normalized on the 3D error when using 2 base poses. The periodic reconstruction is done on the same walking sequence as the non-periodic reconstruction. Computed on CMU MoCap (subject35/walk2/run1, subject13/jump1).

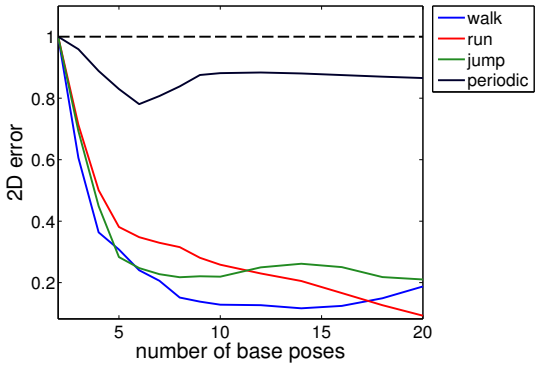


Fig. 9. Influence on the number of used base poses on the 2D error using the non-periodic reconstruction (labels: walk, run, jump) and the periodic reconstruction (label: periodic). The 2D error decreases when more base poses are used. For better visibility, the errors are normalized on the 2D error when using 2 base poses. The periodic reconstruction is done on the same walking sequence as the non-periodic reconstruction. Computed on CMU MoCap (subject35/walk2/run1, subject13/jump1).

motions. On the test data sets six base poses appear to be the optimum with respect to the 3D error. If too many base poses are used the reconstruction deteriorates, whereas the reprojection error reduces. Comparing Figure 8 to Figure 9 shows the correlation between the 2D error and 3D error for the same sequences.

4.4 Influence of regularization

Figure 10 shows the influence of the regularizer on the 3D reconstruction for the motion categories walk, run and jump. For better comparability the error is normalized for each motion class on the error value without regularization. Even a small value for the parameter causes a significant improvement of the 3D reconstruction. In a wide range of parameter settings the reconstruction is much better with the regularizer than without it. The selection of values for the regularization factor is crucial. If the value is too high, the reconstruction is getting worse. Using a too strong factor

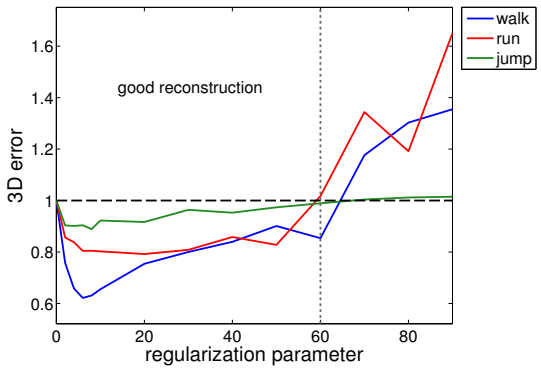


Fig. 10. Influence of the regularization parameter β on the normalized 3D error. In a wide range, the reconstruction improves (left of dotted line) if the regularizer is used as compared to optimization without it ($\beta = 0$). Computed on CMU MoCap (subject7/walk1/run1, subject13/jump2).

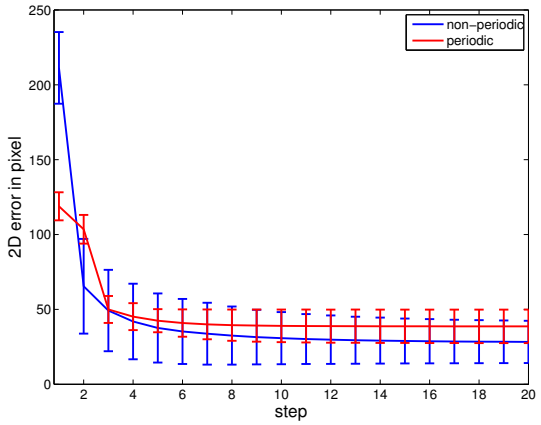


Fig. 12. Mean 2D error and standard deviation for periodic and non-periodic reconstruction of the CMU data set (subjects 7,9,13,16,35). Evaluated on 57 different sequences including the motion categories walk, run and jump. Odd steps refer to camera estimation while even steps refer to pose estimation.

causes the reconstruction to not move at all over time. This is an expectable behavior in the sense of constant bone lengths, but unwanted for a realistic 3D reconstruction.

A comparison of the temporal behavior of the bone lengths of the same sequence with different values for the regularization factor is shown in Figure 11. The bone lengths of the periodic reconstruction (first image) are fluctuating heavily. The second image shows the best non-periodic reconstruction in terms of the 3D error. The fluctuation is less than the one of the periodic reconstruction. The maximal difference in bone length is about 8mm. Considering possible noisy measurements, this should be an acceptable value. On the third image the bone lengths are not changing much, but the 3D error is larger than in the second.

4.5 Convergence and stability

As stated in Section 3.5 the alternatingly optimized objective functions are nonlinear and nonconvex for the periodic and non-periodic case, respectively. Thus we cannot prove convergence of the proposed algorithm. Instead we demonstrate it experimentally. Figure 12 shows the mean and standard deviation of the 2D error during the first 10 iterations

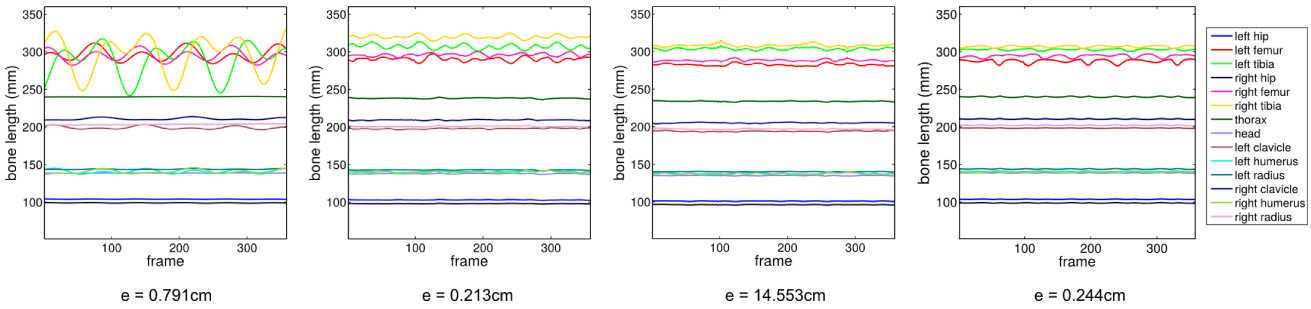


Fig. 11. Comparison of the temporal behavior of the bone lengths with different regularization factors. First: periodic reconstruction with 3D error of $0.791cm$. Second: Non-periodic reconstruction with best 3D error of $0.213cm$. Third: Non-periodic reconstruction with very high regularization factor. Bone lengths are nearly constant over time but the 3D error of $14.553cm$ is larger. Fourth: Using relative variance for regularization. Computed on CMU MoCap (subject35/walk1).

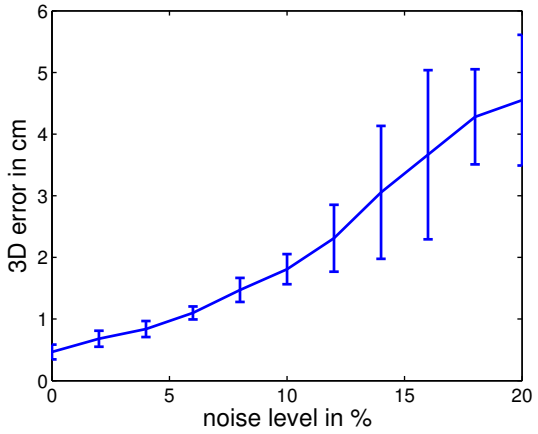


Fig. 13. Mean 3D error and standard deviation of 57 different sequences of the CMU data set (subjects 7,9,13,16,35) obtained by optimization with noisy starting values. The noise level is given in percent of body size of the respective subject.

of 5 different subjects of the CMU MoCap database. An odd step refers to camera estimation while an even step refers to pose estimation. All experiments done during the evaluation (including those in Figure 12) are converging to a plausible local minimum and the value of the 2D error decreases in every step.

As all nonconvex optimization algorithms the proposed algorithm is sensitive to initialization. When initialized with bad starting values it converges to a bad local minimum. As described in Section 3.5, initialization is done by the mean pose of the corresponding motion category which is an appropriate assumption. However, it is reasonable to evaluate the stability of the algorithm with bad or noisy initializations. Figure 13 shows the mean and standard deviation of the 3D error with gaussian noise added onto the starting values. Up to a noise level of 10% the reconstructions still look plausible and close to the reconstructions without noise. Above 10% the 3D reconstructions degenerate to unrealistic poses.

4.6 Different Motion classes

We trained our algorithm on multiple motion classes including periodic (walking, running, jogging) and non-periodic motions (jump up/forward). Different data sets are used (CMU MoCap [10], HumanEva [11], KTH Football [1]). The ground truth for the CMU Mocap and the HumanEva data sets are generated from marker based motion capture data of humans performing different actions. The KTH Football data set contains video sequences with manually labeled joints. The 3D reconstruction which we use as ground truth data was computed using a multi camera system. Overall this data set is more noisy than the other two data sets and offers a real world scenario. Table 1 shows the 3D reconstruction error of our different methods on some of the used motion sequences compared to the results of Gotardo and Martinez [4] and Bregler et al. [13]. It is noticeable that the reconstruction results of the jumping sequences are worse compared to the other sequences. The reason is that the variance between jumping motions of different persons is much larger than between walking motions. So a new (not trained) jumping motion is unsufficiently explained by the base poses, while every new walking pattern is very similar to those in the training data. Nevertheless the reconstructions appear realistic (cf. Figure 23). All results except the row labeled "np all" are obtained by training on the specific motion categories. When training all motions at once (here we are using walk, run, jog, jump up, jump forward) to get more general base poses, the results are getting worse but stay realistic and are still superior to [13] and [4]. The results of [13] and [4] are obtained with the source code provided by the authors.

TABLE 1
Average 3D reconstruction error in cm on the CMU dataset (walk, run, jump), HumanEva walking dataset (HE) and KTH Football dataset. First row: reconstruction with periodic constraints. Second row: non periodic reconstruction without bone length regularizer. Third row: Best reconstruction result achieved with bone length regularizer. Fourth row: best result when using all motions for training simultaneously. Fifth and Sixth row: comparison to other approaches.

Method	walk	run	jump	HE	KTH
periodic	0.784	0.968	-	1.200	0.357
np ($\beta = 0$)	0.295	0.661	1.226	0.564	0.292
best	**0.183**	**0.523**	**1.090**	**0.423**	**0.187**
np all	0.334	2.805	1.313	-	-
[13]	4.557	10.821	8.531	17.824	4.427
[4]	16.359	11.395	17.139	5.714	14.673

Our 3D reconstructions are highly realistic, which was

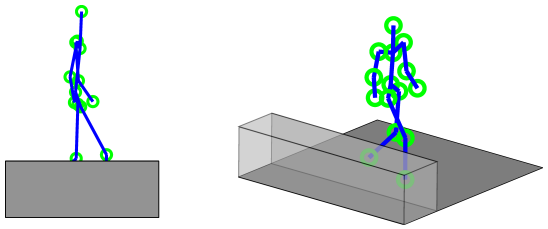


Fig. 14. Left: Observation data of a person walking behind a box. The legs are partly occluded. Right: 3D reconstruction of occluded body parts using our method.

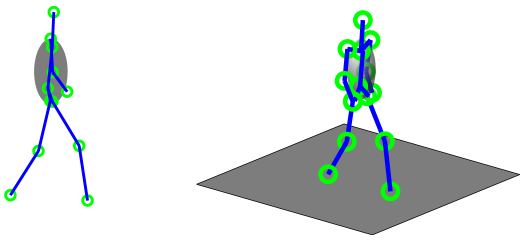


Fig. 15. Left: During the whole sequence the left hand is occluded by the body. Right: 3D reconstruction of occluded body parts using our method.

shown by surveying the 3D error. Figures 21, 22, 23, 24 show reconstructed motions taken from the CMU MoCap database. Figure 21 uses the periodic reconstruction with only 4 base poses. Figure 22, 23 and 24 are using the non-periodic approach.

4.7 Occlusions

In realistic scenes, body parts can be occluded. This happens for example if parts of the observed person are behind an object, for instance as shown in Figure 14. Another common case is self-occlusion where one body part occludes another body part. The integration of occlusions in our algorithm is simple. Since we are using the frobenius norm of the reprojection error it is possible to set occluded values to zero in the observation matrix $\boldsymbol{W}_{2D}$ and the reprojection $\boldsymbol{M\Theta Q}$ while using the same objective function (Eq. (18) or Eq. (21)) as in the non occluded case. This equals to canceling the corresponding equations in the objective function.

Figure 14 shows a person walking (CMU MoCap, subject7/walk2) behind an artificial box so that the legs cannot be seen in the input data. Our algorithm is able to reconstruct a realistic leg motion that is very close to the original motion. Figure 15 shows the problem of self occlusion. In the whole sequence, the back arm (shoulder, elbow and hand) is fully occluded, i.e. 20% of the input data is unknown. On the right of Figure 15 the back arm is correctly reconstructed by our method.

For further evaluation of the occlusion handling we randomly delete data points in the input data. Figure 16 shows the maximal 3D error of the periodic and non-periodic reconstruction. While the non-periodic reconstruction produces a high maximal 3D error for occlusions higher than 3%, the periodic reconstruction benefits from the smoothness constraint it puts on the reconstruction and remains stable for occlusions up to 20%.

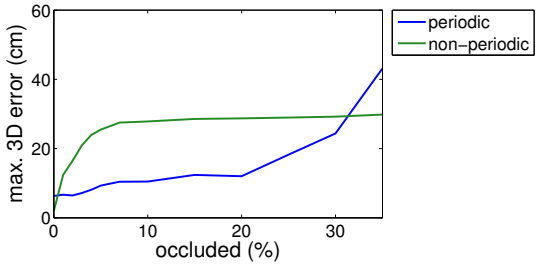


Fig. 16. Comparison of the maximal 3D error of periodic and non-periodic reconstruction with randomly occluded data points. The periodic reconstruction appears to be more stable as it puts a smoothness constraint on the reconstruction. Computed on CMU MoCap (subject35/walk1)

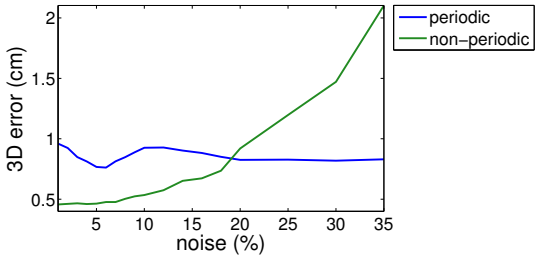


Fig. 17. Influence of additional noise on periodic and non-periodic reconstruction. While the 3D error of the non-periodic reconstruction raises, the error for the periodic reconstruction remains nearly constant.

4.8 Noise stability

To evaluate the stability of our method we put additional noise on the 2D input data. Figure 17 shows the 3D reconstruction error with respect to the noise level for the periodic and non-periodic reconstruction. In this case 5% noise means 5% of the maximal range of motion of the most moving 2D point. With a very high noise level the reconstruction is still good. Apparently the periodic reconstruction appears to be more stable than the non-periodic reconstruction, because it puts a strong smoothness constraint on the weights $\boldsymbol{\Theta}$ of the base poses. The result is still a smooth motion as shown in Figure 18. While the non-periodic reconstruction (center) is getting unstable the periodic reconstruction (right) still achieves a realistic output compared to the ground truth data (left).

4.9 Classification

We also used our proposed method for classification of a mixed motion. In this example we reconstruct the outdoor sequence from [31] of a person running and jumping over an obstacle (cf. Figure 20). For the classification the reconstruction is done for 10 frames wide sections over the whole sequence. Figure 19 shows the corresponding 2D error when using the periodic reconstruction with base poses trained from the CMU running sequences (35/17-26). The 2D error increases for non-trained motions, as these can not be reconstructed with the used base poses. In this example the jump over the obstacle around frame 30 can be clearly seen. By setting a threshold for the 2D error a classification in running and non-running motion is possible. For the jumping part

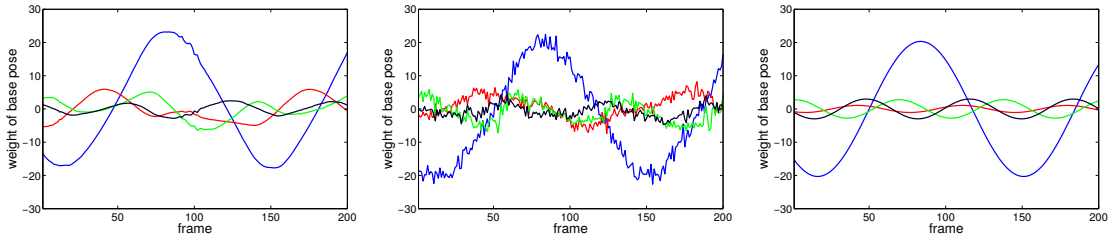


Fig. 18. Comparison of the weights for the base poses. Left: Ground truth weights. Center: Non-periodic reconstruction with 20% noise (3D error: $1.09cm$). Right: Periodic reconstruction with 20% noise (3D error: $1.02cm$). Computed on the first 200 frames of CMU MoCap (subject7/walk1).

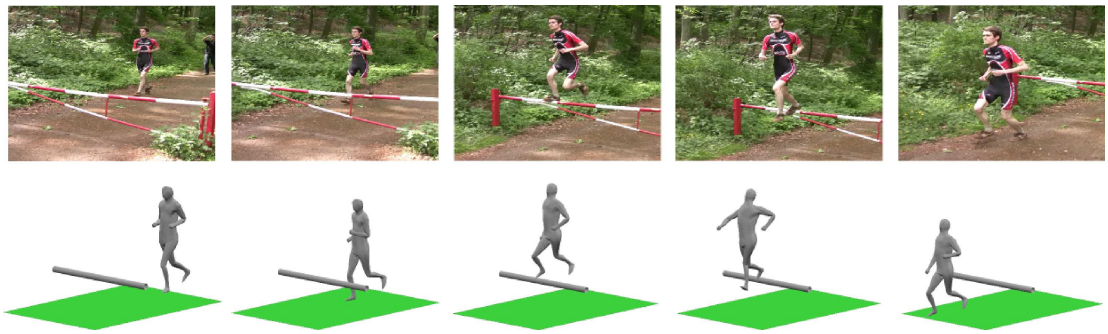


Fig. 20. Combined running and jumping sequence from [31]. The first two frames are reconstructed using the periodic reconstruction, the others are using the non-periodic reconstruction. Although the base poses are trained on another data set that does not contain this specific motion, the reconstruction is not perfect but realistic.

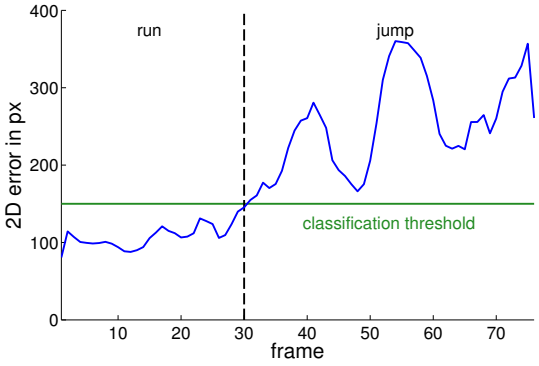


Fig. 19. 2D error using the periodic reconstruction. The base poses are trained from the CMU running sequences (35/17-26). Reconstructing poses belonging to the jumping motion results in a large 2D error.

of the sequence we may therefore switch from the periodic reconstruction (cf. Section 3.3) to the less constrained non-periodic algorithm (cf. Section 3.4). Since there is no similar jumping motion in the other data sets (only jumping with legs closed or on one leg), we use base poses trained on the motions walk, run and jump simultaneously as mentioned in Section 4.6. Although the example sequence is manually labeled and the base poses are trained on another data set our method achieves realistic results as shown in Figure 20.

5 Conclusion

We presented a new method for the 3D reconstruction of human motion from monocular image sequences. Using periodic functions to model the weights of the base poses turned out to be very effective and stable on periodic motion. Reconstruction of non-periodic motion was successfully done with our new regularization term. In contrast to state of the art methods for estimation of nonrigid shapes from monocular image sequences (e.g. [4], [13]) the proposed regularizations enable us to reconstruct plausible human motion even under low reconstructibility. We showed the generality of our approach on multiple common datasets with different motion types. It even performs well under occlusions, noise and on the real world data of the KTH dataset as well as on our outdoor obstacle jump sequence.

6 Acknowledgements

The work has been partially supported by the ERC-Starting Grant (Dynamic MinVIP) and the DFG-project RO 2497/11-1. The authors gratefully acknowledge the support.

References

[1] V. Kazemi, M. Burenius, H. Azizpour, and J. Sullivan, "Multi-view body part recognition with random forests," in *British Machine Vision Conference (BMVC)*, 2013.

[2] P. Gotardo and A. Martinez, "Kernel non-rigid structure from motion," in *International Conference on Computer Vision (ICCV)*. IEEE, 2011.

[3] P. Gotardo and A. Martinez, "Non-rigid structure from motion with complementary rank-3 spaces," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[4] O. Hamsici, P. Gotardo, and A. Martinez, "Learning spatially-smooth mappings in non-rigid structure from motion," in *European Conference on Computer Vision (ECCV)*, 2011.

[5] Y.-L. Chen and J. Chai, "3d reconstruction of human motion and skeleton from uncalibrated monocular video," in *Asian Conference on Computer Vision (ACCV)*, ser. Lecture Notes in Computer Science, H. Zha, R. I. Taniguchi, and S. J. Maybank, Eds., vol. 5994. Springer, 2009, pp. 71–82.

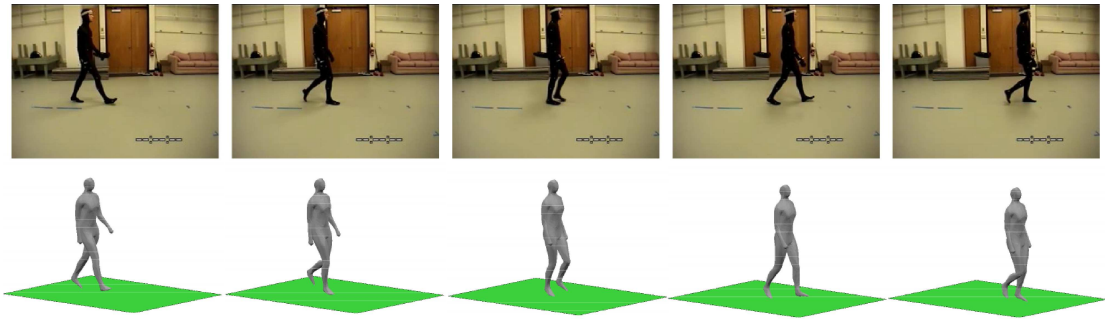


Fig. 21. Walking sequence 35/02 of the CMU MoCap data set reconstructed with the periodic reconstruction using only 4 base poses

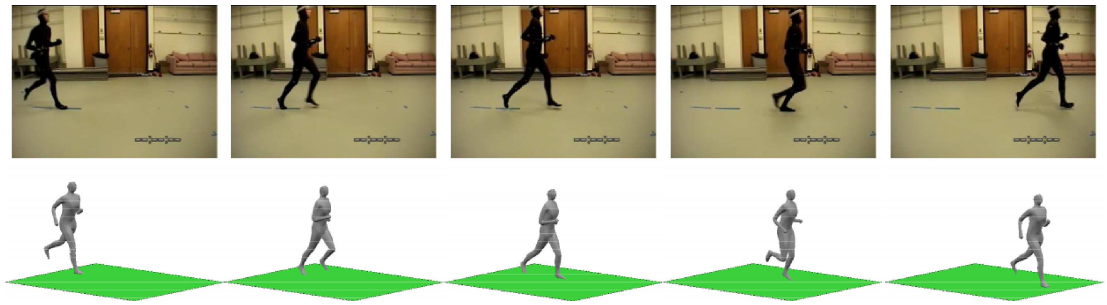


Fig. 22. Running sequence 35/17 of the CMU MoCap data set reconstructed with the non-periodic reconstruction

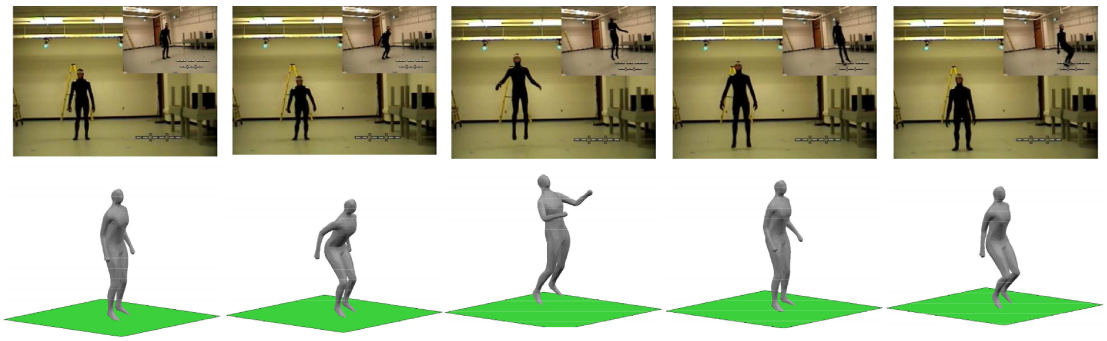


Fig. 23. Jumping sequence 13/11 of the CMU MoCap data set with the non-periodic reconstruction

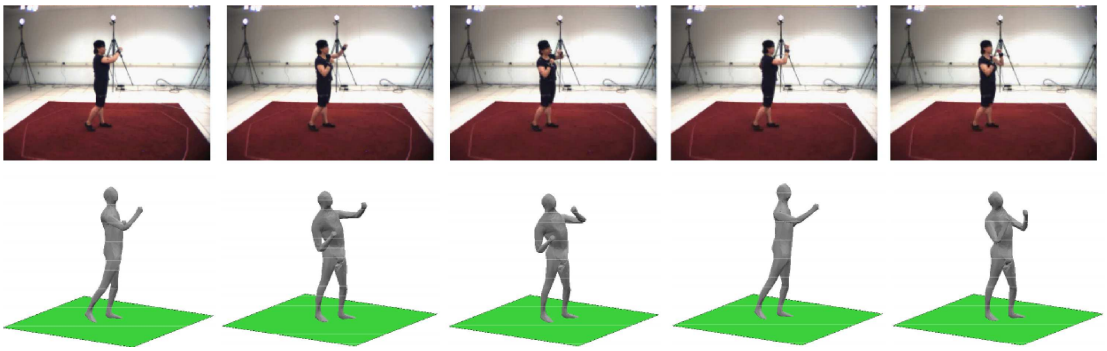


Fig. 24. Boxing sequence of the HumanEva data set with the non-periodic reconstruction

[6] V. Ramakrishna, T. Kanade, and Y. A. Sheikh, "Reconstructing 3d human pose from 2d image landmarks," in *European Conference on Computer Vision (ECCV)*, October 2012.

[7] E. Simo-Serra, A. Ramisa, G. Aleny, C. Torras, and F. Moreno-Noguer, "Single image 3d human pose estimation from noisy observations." in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2673–2680.

[8] C. Wang, Y. Wang, Z. Lin, A. Yuille, and W. Gao, "Robust estimation of 3d human poses from a single image," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[9] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh, "3d reconstruction of a moving point from a series of 2d projections," *European Conference on Computer Vision (ECCV)*, September 2010.

[10] CMU. (2014) Human motion capture database. [Online]. Available: http://mocap.cs.cmu.edu/

[11] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *International Journal of Computer Vision*, vol. 87, no. 1-2, pp. 4–27, 2010.

[12] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, pp. 137–154, 1992.

[13] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3d shape from image streams," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000, pp. 690–696.

[14] J. Xiao, J. Chai, and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," in *European Conference on Computer Vision (ECCV)*, May 2004.

[15] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Trajectory space: A dual representation for nonrigid structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1442–1456, 7 2011.

[16] L. Torresani, A. Hertzmann, and C. Bregler, "Learning non-rigid 3d shape from 2d motion," in *Neural Information Processing Systems (NIPS)*, S. Thrun, L. K. Saul, and B. Schölkopf, Eds. MIT Press, 2003.

[17] L. Torresani, A. Hertzmann, and C. Bregler., "Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors," *IEEE Transactions Pattern Analysis and Machine Intelligence*, 2008.

[18] L. Torresani, D. B. Yang, E. J. Alexander, and C. Bregler, "Tracking and modeling non-rigid objects with rank constraints," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. 493–500.

[19] I. Akhter, T. Simon, S. Khan, I. Matthews, and Y. Sheikh, "Bilinear spatiotemporal basis models," *ACM Transactions on Graphics*, vol. 31, no. 2, pp. 17:1–17:12, Apr. 2012.

[20] Y. Zhu, M. Cox, and S. Lucey, "3d motion reconstruction for real-world camera motion." in *CVPR*. IEEE Computer Society, 2011, pp. 1–8.

[21] Y. Dai and H. Li, "A simple prior-free method for non-rigid structure-from-motion factorization," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, ser. CVPR '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 2018–2025.

[22] C. Theobalt, J. Carranza, M. Magnor, and H.-P. Seidel, "Enhancing silhouette-based human motion capture with 3d motion fields," in *11th Pacific Conference on Computer Graphics and Applications (PG-03)*, J. Rokne, R. Klein, and W. Wang, Eds., IEEE. Canmore, Canada: IEEE, October 2003, pp. 185–193.

[23] J. Carranza, C. Theobalt, M. A. Magnor, and H.-P. Seidel, "Free-viewpoint video of human actors," in *ACM SIGGRAPH 2003 Papers*, ser. SIGGRAPH '03. ACM, 2003, pp. 569–577.

[24] L. Li, J. McCann, N. S. Pollard, and C. Faloutsos, "Bolero: A principled technique for including bone length constraints in motion capture occlusion filling." in *Symposium on Computer Animation*, Z. Popovic and M. A. Otaduy, Eds. Eurographics Association, 2010, pp. 179–188.

[25] L. Kovar, J. Schreiner, and M. Gleicher, "Footskate cleanup for motion capture editing," in *Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ser. SCA '02. ACM, 2002, pp. 97–104.

[26] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H.-P. Seidel, "Multilinear pose and body shape estimation of dressed subjects from image sets." in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2010, pp. 1823–1830.

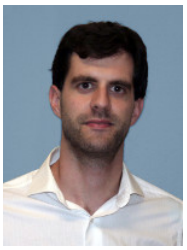
[27] N. F. Troje, "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns," *Journal of Vision*, vol. 2, no. 5, pp. 371–387, 2002.

[28] N. F. Troje, "The little difference: Fourier based synthesis of gender-specific biological motion," *AKA Press*, pp. 115–120, 2002.

[29] Z. Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *Journal of Optimization Theory and Applications*, vol. 72, no. 1, pp. 7–35, Jan. 1992.

[30] P. Tseng and S. Yun, "A coordinate gradient descent method for nonsmooth separable minimization," *Mathematical Programming*, vol. 117, no. 1-2, pp. 387–423, 2009.

[31] N. Hasler, B. Rosenhahn, T. Thormählen, M. Wand, and H.-P. Seidel, "Markerless motion capture with unsynchronized moving cameras," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.



Bastian Wandt studied Mechatronics at the Leibniz University Hannover. During his bachelor studies he focused on robotics and automation. He received his B. Sc. in 2012 with a thesis on path planning of autonomous mobile robots. His master thesis dealt with 3D reconstruction and animation of human motion. Since 2014 he is working towards his Dr.-Ing. at the Institut für Informationsverarbeitung (TNT) in Hannover. His main research interests are markerless motion capture and 3D reconstruction.

Hanno Ackermann studied Computer Engineering at the University of Mannheim. He received his masters degree (Dipl.-Inf.) in 2003. From 10/2004 until 3/2008 he did his Phd at the University of Okayama, Japan. From 5/2008 until 9/2008 he worked as PostDoc at the Max-Planck-Institute for Computer Science in Saarbruecken, Germany. Since 10/2008 he is a member of the group of Prof. Rosenhahn at Leibniz University Hannover. He is currently funded by a scholarship of the German Research Foundation (DFG).

Bodo Rosenhahn studied Computer Science (minor subject Medicine) at the University of Kiel. He received the Dipl.-Inf. and Dr.-Ing. from the University of Kiel in 1999 and 2003, respectively. From 10/2003 till 10/2005, he worked as PostDoc at the University of Auckland (New Zealand), funded with a scholarship from the German Research Foundation (DFG). In 11/2005-08/2008 he worked as senior researcher at the Max-Planck Insitute for Computer Science in Saarbruecken. Since 09/2008 he is Full Professor at the Leibniz-University of Hannover, heading a group on automated image interpretation.