

3D Human Motion Capture from Monocular Image Sequences

Bastian Wandt, Hanno Ackermann, Bodo Rosenhahn
Leibniz University Hannover
Hannover, Germany
www.tnt.uni-hannover.de

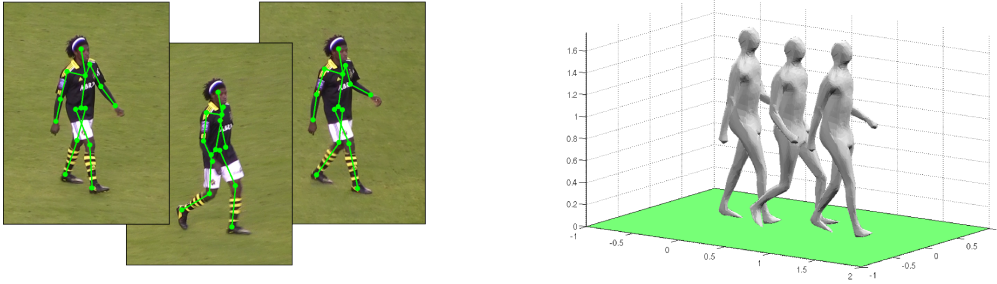


Figure 1: Real world scenario of KTH database [9]. Left: frames 115, 136 and 143 of Sequence 1 from Football Dataset II. Right: 3D reconstruction using our proposed method

Abstract

This paper 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 our 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.

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.

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. [6, 7, 8]) 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. [3, 11, 13, 20]) 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 paper, we use a trilinear factorization approach similar to [7, 10, 11, 20]. 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 [7] and [10], we keep the second factor fixed which corresponds to 3D structure, similar to [11] and [20]. 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 [11] and [20] 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 [4], HumanEva [12]) as well as on challenging real world data as for example the KTH Football Dataset [9] shown in Figure 1. Additionally we are analyzing the influence of the number of base poses and the regularization factor on the reconstruction result.

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. Summarizing, the contributions of this paper are as follows:

- A periodic model for the mixing coefficients for periodic and quasi-periodic motions such as walking
- 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 [14]. 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. [2] by expressing the observed shape in any particular image as a linear combination of multiple rigid basis shapes. Xiao et al. [21] 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. Akhter et al. [1] showed that the solution proposed in [21] still is ambiguous. Torresani et al. [15, 16, 17] 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. [1] built on this idea and fixed the linear coefficients 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 to satisfy. Gotardo and Martinez later extended this approach by assuming smoothly moving cameras [6, 7, 8] and 3D points. Li introduced an approach

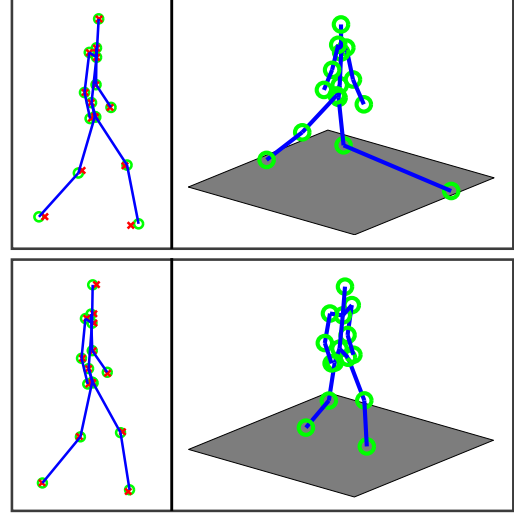


Figure 2: 3D reconstruction (green circles, blue lines) and ground truth data (red crosses). Top: Using approach of Gotardo and Martinez [8]. Most non-rigid structure from motion approaches with no rotation and unknown base poses fail, although they produce a small reprojection error (left). From other perspectives (right) a wrong reconstruction can be observed. Bottom: Our approach. Correct reconstruction in all views.

based on L_1 -minimization [5] where the number of mixing coefficients that are non-zero is minimized.

Several works have been proposed regarding the 3D reconstruction of human poses given single images only [3, 11, 13, 20]. State-of-the-art methods such as the work of Ramakrishna et al. [11] 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. [20] extended that model. Different to [11] they enforce the proportions of eight selected limbs to be constant. However, limb proportions differ from one person to another.

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. walking and running), we show that it is possible, to assume a periodic weight for the base poses to significantly

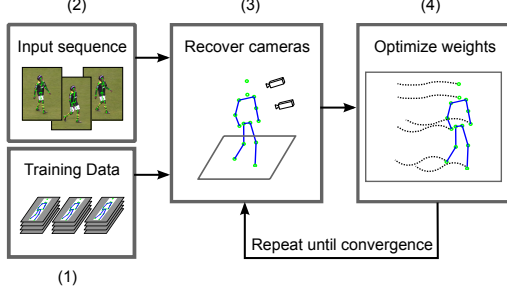


Figure 3: Our method. (1) First we are learning 3D base poses from training data. (2) Input sequence. (3) Cameras are recovered from estimated 3D poses and 2D poses. (4) Weights for base poses are calculated by minimizing the reprojection error. Steps (3) and (4) are alternated until the algorithm converges

reduce the number of variables, that have to be calculated (section 3.3). Our algorithm (section 3.5) is alternately 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 $P \in \mathbb{R}^{4 \times a}$ with a joints in homogeneous coordinates can be written as a linear combination of k previously learned base poses $Q_l \in \mathbb{R}^{4 \times a}$

$$P = Q_0 + \sum_{l=1}^k \theta_l Q_l, \quad (1)$$

where Q_0 is the mean pose of all poses used for training and $\theta_l \in \mathbb{R}^{4 \times 4}$ is the weight matrix for the base pose Q_l . With ϑ_l as the scalar weight for the l -th base pose each θ_l has the form

$$\theta_l = \begin{pmatrix} \vartheta_l I_3 & \\ & 0 \end{pmatrix}, \quad (2)$$

where I_3 is the 3×3 identity matrix. Note that only the coordinates in the mean pose Q_0 are describing a point in homogeneous coordinates, while $Q_{1,\dots,k}$ are directions that define *deformations*. By stacking poses we can write a 3D sequence as $W \in \mathbb{R}^{4f \times a}$ of f images, with $P_{1,\dots,f}$ as the poses in frames $1, \dots, f$

$$W = \begin{pmatrix} P_1 \\ \vdots \\ P_f \end{pmatrix}. \quad (3)$$

With Eq. (1) we can do a factorization

$$W = \begin{pmatrix} Q_0 + \sum_{l=1}^k \theta_{l,1} Q_l \\ \vdots \\ Q_0 + \sum_{l=1}^k \theta_{l,f} Q_l \end{pmatrix} = \Theta \begin{pmatrix} Q_0 \\ Q_1 \\ \vdots \\ Q_k \end{pmatrix} = \Theta Q, \quad (4)$$

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

The projection of a 3D pose P_i in the i -th frame to a 2D pose $P_{i,2D} \in \mathbb{R}^{2 \times a}$ is done by the camera matrix $M_i \in \mathbb{R}^{2 \times 4}$

$$P_{i,2D} = M_i P_i. \quad (5)$$

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

$$M = \begin{pmatrix} M_1 & & \\ & \ddots & \\ & & M_f \end{pmatrix}. \quad (6)$$

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

$$W_{2D} = M \Theta Q. \quad (7)$$

This model is very similar to the models proposed by [2], [10] and [7]. While they are fixing Θ and optimize for M and Q , our approach is using a previously learned Q and optimize for the weights Θ like [11] and [20] 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 w_{2D}^i can be factorized with the above notation as

$$w_{2D}^i = M_i \Theta_i Q, \quad (8)$$

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

$$\min_{M_i} \|w_{2D}^i - M_i \Theta_i Q\|_F. \quad (9)$$

Considering the entries in

$$M_i = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \end{pmatrix} \quad (10)$$

we can enforce a weak perspective camera by the constraints

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

and

$$m_{11}m_{21} + m_{12}m_{22} + m_{13}m_{23} = 0. \quad (12)$$

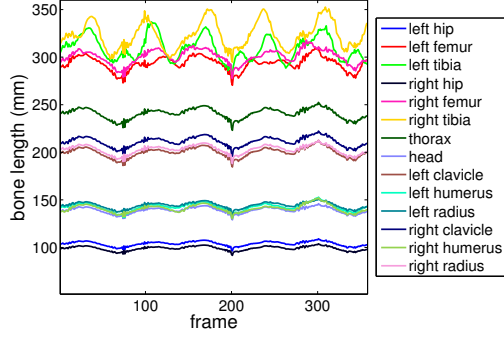


Figure 4: Temporal behavior of bone lengths obtained by unconstrained optimization. The maximal variation is about 40mm.

3.3. Periodic Motion

With the camera matrix M calculated as described in Section 3.2 the weights Θ for the base poses can be reconstructed. Trying to optimize the reprojection error for all variables in Θ fails, as there are too 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 Θ . This assumption has been shown to be appropriate for periodic motion (for instance gait sequences [18, 19]).

As shown in Section 3.1 the number of unknowns in Θ equals fk . By modeling the temporal behavior of ϑ as

$$\vartheta(t) = \alpha \sin(\omega t + \varphi) \quad (13)$$

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

$$\min_{\alpha, \omega, \varphi} \|W_{2D} - M\Theta Q\|_F. \quad (14)$$

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 4 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 wrong. 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 [11] and [20] 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 ΘQ . We denote the length of bone s as

$$b_s = \|\mathbf{j}_{s,2} - \mathbf{j}_{s,1}\|_2, \quad (15)$$

where $\mathbf{j}_{s,1}$ and $\mathbf{j}_{s,2}$ 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 $Var(\bullet)$ of all bone lengths over time

$$r_B = \beta \sum_i Var(b_i), \quad (16)$$

with β as the regularization parameter. This regularizer holds the bone length constant but is not fixing it to a specific value. So the optimization problem can be written as

$$\min_{\Theta} \|W_{2D} - M\Theta Q\|_F + r_B. \quad (17)$$

3.5. Algorithm

To estimate the two factors M and Θ we alternately optimize for each factor while keeping the other fixed. In the first iteration we use the mean pose as initialization. This means setting all values in Θ to zero except the ones weighting the mean pose 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.

Algorithm 1 Recover camera and shape

```

 $Q \leftarrow$  base shapes
while no convergence do
  for  $i = 1 \rightarrow f$  do
    calculate starting values for  $M_i$ 
    optimize  $\|w_{2D}^i - M_i \Theta_i Q\|_F$ 
    insert  $M_i$  in  $M$ 
  end for
  optimize  $\|W - M\Theta Q\|_F + r_B$ 
end while

```

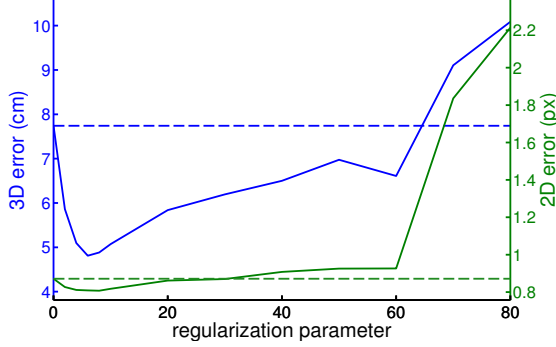


Figure 5: 2D reprojection error and 3D reconstruction error with different regularization parameter β . While the 2D error is not changing much or getting worse, the 3D error gets significantly better at most parameter values.

4. Experimental Results

To evaluate our method, we were using three different databases: CMU MoCap [4], HumanEva [12] and KTH Football [9]. We trained our algorithm on 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

$$e = \frac{1}{f} \|W_{in} - W_{rec}\|_F, \quad (18)$$

with W_{in} as the real 3D data and 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. With our bone length regularizer we achieve a worse reprojection error but a significantly better 3D reconstruction (see Figure 5). 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 [4], the HumanEva dataset [12] and as a real world example the KTH Football Dataset II [9]. 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

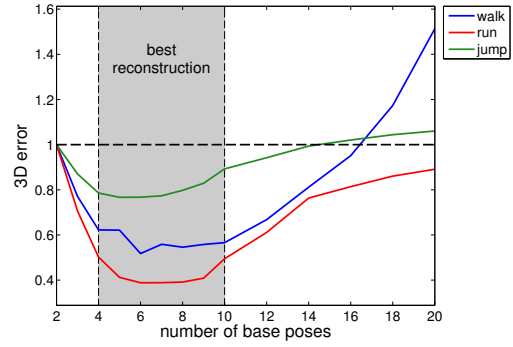


Figure 6: The number of used base poses is crucial for a good 3D reconstruction. Using more than 10 base poses for each motion category worsens the reconstruction error. For better visibility, the errors are normalized on the 3D error when using 2 base poses.

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 9 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 very 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 very fast and efficient way for the 3D reconstruction.

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 results in more degrees of freedom for the optimization, which can cause unnatural deformation. For simple periodic motion, such as walking, four base poses appear to be enough and more just worsen the reconstruction (see Fig-

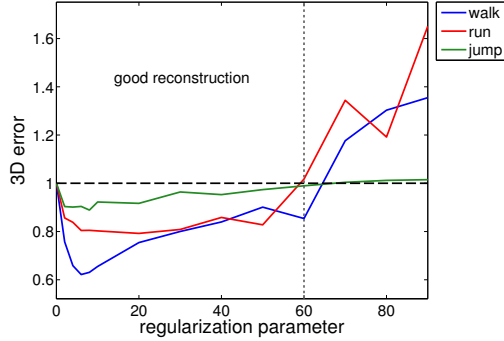


Figure 7: Influence of the regularization parameter β on the normalized 3D error. In a wide range, the reconstruction improves (left of dotted line) if the regularizer is used as compared to optimization without it ($\beta = 0$)

ure 6). The range from 4 to 10 base poses gives the best reconstruction results. On the test data sets six base poses appear to be the optimum. Again, it is very important to look at the 3D error, because the reprojection error reduces if the more base poses are used, but the reconstruction worsens as shown in Figure 6.

4.4. Influence of regularization

Figure 7 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 causes the reconstruction to not move at all over time. This is an expectable behavior in 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 8. The bone lengths of the periodic reconstruction (left) are fluctuating heavily. The middle 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 right the bone lengths are not changing much, but the 3D error is bigger than in the middle.

4.5. Different Motion classes

We trained our algorithm on multiple motion classes including periodic (walking, running, jogging) and non-periodic motions (jump up/forward). Also different data sets were used (CMU MoCap [4], HumanEva [12], KTH Football [9]). Table 1 shows the 3D reconstruction error of our different methods on some of the used motion sequences compared to the results of Bregler [2] and Gotardo and Martinez [8]. The used error metric is explained in Section 4. It is noticeable that the reconstruction results of the jumping sequences are worse compared to the other sequences. The reason is that the difference between two jumping persons is much bigger than between two walking persons. So this difference is not covered that much in the training set. Nevertheless the reconstructions appear to be very realistic (see Figure 11). 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 still stay realistic.

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	-	-
[2]	4.557	10.821	8.531	17.824	4.427
[8]	16.359	11.395	17.139	5.714	14.673

Table 1: Average 3D reconstruction error in *cm* on the CMU dataset (walk, run, jump), HumanEva walking dataset (HE) and KTH Football dataset.

5. Conclusion

We presented a new method for the 3D reconstruction of human motion. 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. We showed the generality of our approach on multiple common datasets with different motion types. It even performs well on the noisy real world data of the KTH dataset. Our 3D reconstructions are highly realistic, which was shown by surveying the 3D error. In Figures 9, 10, 11 and 12 are some of the reconstructed motions of the CMU MoCap database and the HumanEva dataset. Figure 9 uses the periodic reconstruction with only 4 base poses. Figure 10, 11 and 12 are using the non-periodic approach.

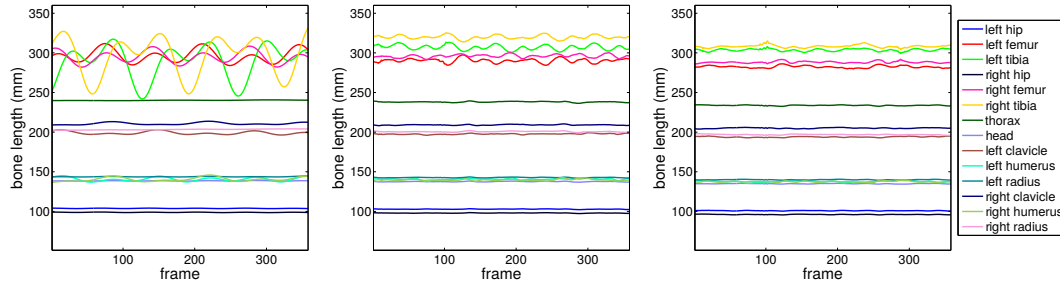


Figure 8: Comparison of the temporal behavior of the bone lengths with different regularization factors. Left: periodic reconstruction. Middle: Non-periodic reconstruction with best 3D error. Right: Non-periodic reconstruction with very high regularization factor; bone lengths are nearly constant over time but the 3D error is larger

References

- [1] 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*, 33(7):1442–1456, 7 2011. 2
- [2] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 690–696, 2000. 2, 3, 6
- [3] Y.-L. Chen and J. Chai. 3d reconstruction of human motion and skeleton from uncalibrated monocular video. In H. Zha, R. I. Taniguchi, and S. J. Maybank, editors, *Asian Conference on Computer Vision (ACCV)*, volume 5994 of *Lecture Notes in Computer Science*, pages 71–82. Springer, 2009. 1, 2
- [4] CMU. Human motion capture database, 2014. 2, 5, 6
- [5] 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)*, CVPR ’12, pages 2018–2025, Washington, DC, USA, 2012. IEEE Computer Society. 2
- [6] P. Gotardo and A. Martinez. Kernel non-rigid structure from motion. In *International Conference on Computer Vision (ICCV)*. IEEE, 2011. 1, 2
- [7] 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. 1, 2, 3
- [8] 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. 1, 2, 6
- [9] V. Kazemi, M. Burenius, H. Azizpour, and J. Sullivan. Multi-view body part recognition with random forests. In *British Machine Vision Conference (BMVC)*, 2013. 1, 2, 5, 6
- [10] 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. 1, 2, 3
- [11] 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. 1, 2, 3, 4
- [12] 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*, 87(1-2):4–27, 2010. 2, 5, 6
- [13] 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)*, pages 2673–2680. IEEE, 2012. 1, 2
- [14] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9:137–154, 1992. 2
- [15] L. Torresani, A. Hertzmann, and C. Bregler. Learning non-rigid 3d shape from 2d motion. In S. Thrun, L. K. Saul, and B. Schölkopf, editors, *Neural Information Processing Systems (NIPS)*. MIT Press, 2003. 2
- [16] 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. 2
- [17] 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)*, pages 493–500, 2001. 2
- [18] N. F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5):371–387, 2002. 4
- [19] N. F. Troje. The little difference: Fourier based synthesis of gender-specific biological motion. *AKA Press*, pages 115–120, 2002. 4
- [20] 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. 1, 2, 3, 4
- [21] 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. 2

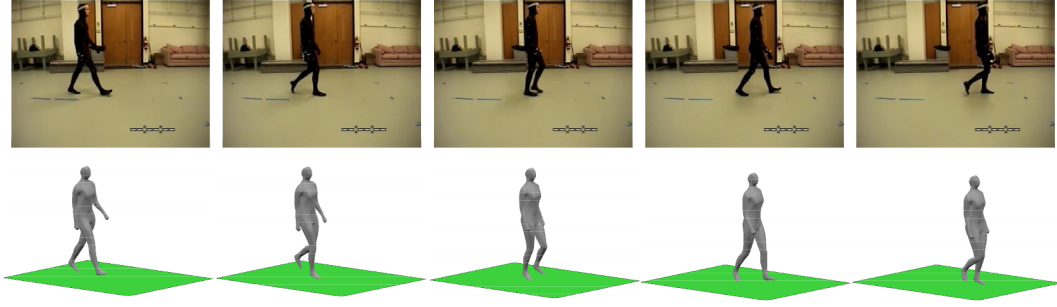


Figure 9: Walking sequence 35/02 of the CMU MoCap data set reconstructed with the periodic reconstruction using only 4 base poses

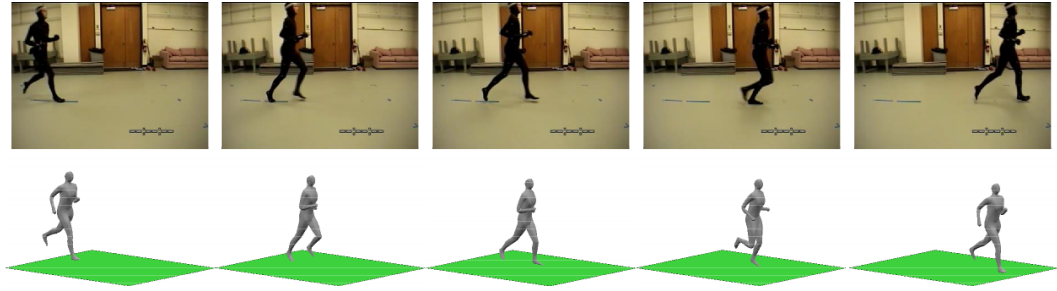


Figure 10: Running sequence 35/17 of the CMU MoCap data set reconstructed with the non-periodic reconstruction

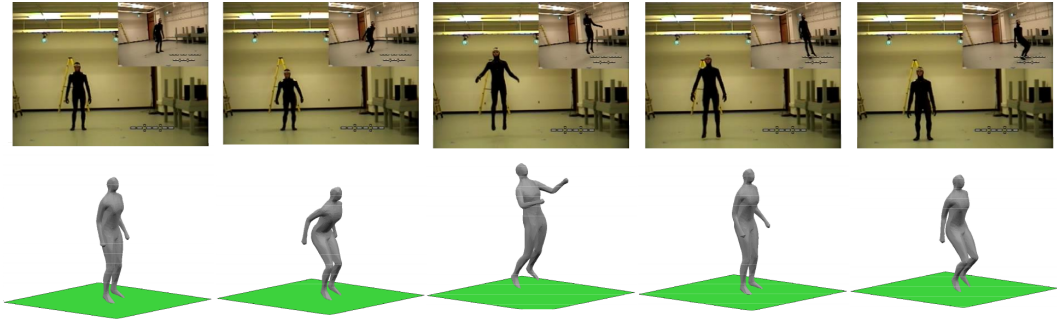


Figure 11: Jumping sequence 13/11 of the CMU MoCap data set with the non-periodic reconstruction

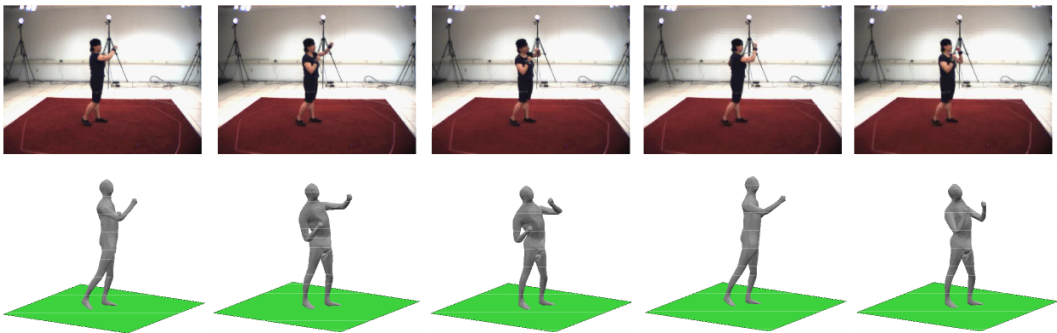


Figure 12: Boxing sequence of the HumanEva data set with the non-periodic reconstruction