

3D Periodic Human Motion Reconstruction from 2D Motion Sequences

Zonghua Zhang and Nikolaus F. Troje
BioMotionLab, Department of Psychology
Queen's University, Canada
zhang, troje@psyc.queensu.ca

Abstract

In this report, we present and evaluate a method of reconstructing three-dimensional (3D) periodic human motion from two-dimensional (2D) motion sequences. Based on a Fourier decomposition of a training set of 3D data, we construct a linear, morphable representation. Using this representation a low-dimensional linear model is learned by means of Principle Component Analysis (PCA). Two-dimensional test data are now projected onto this model and the resulting 3D reconstructions are evaluated. We present two different simulations. In the first experiment, we assume the 2D projection matrix to be known. In the second experiment, the horizontal viewpoint is unknown and is being recovered from the data.

1. Introduction

Human motion contains a wealth of information about the actions, intentions, emotions, and personality traits of a person and plays an important role in many application areas, such as surveillance, computer games, films, and biomechanics. The human visual system is so sensitive to human motion patterns that in a visual scene, one can quickly and efficiently recognize human behaviors and learn many aspects of human motion. What our visual system seems to have achieved effortlessly so far is still a challenging problem for artificial vision systems, although the computerized analysis of human motion is gaining more and more interest. Many researchers have conducted studies on human motion analysis, and general surveys can be found in recent review papers [1,2]. Most of the existing research [3~7] has been focused on the problem of tracking and recognizing human motion through motion sequences, while less attention has been paid to the problem of reconstructing 3D human motion from 2D motion sequences. 3D motion reconstruction is similar to 3D shape reconstruction from a sequence of images, but it

is widely believed that 3D motion reconstruction is much harder than the latter because it requires not only 3D postures derived from 2D motion sequences but also temporal correspondence between the model and 2D motion sequences. Therefore, 3D motion reconstruction from 2D motion sequences is still a challenging task.

The importance of 3D motion reconstruction stems from applications such as surveillance and monitoring, human body animation, and automatic recognition of human activities from video databases. Unlike 2D motion, which is highly view-dependent, 3D human motion can provide robust recognition and surveillance. However, multiple cameras are usually needed to get 3D motion information. For the purpose of realizing a virtual actor and multimedia application, even the most skilled modelers and animators are not able to accurately reproduce the respective motions from image sequences.

Human motion analysis is roughly divided into two categories: model-based methods and model-free methods [8]. In the model-based methods, a human model is used to represent the observed subjects, while in the model-free methods, the motion information is directly utilized in a sequence of images to recognize or analyze human motion. The use of a human model seems to be the preferred trend. Model-based methods require correspondence-based representations, which result in a separation of the overall information into range-specific information and domain-specific information [9,10]. In the case of biological motion data, range specific information refers to the state of the actor at a given time in terms of current position or of a number of feature points. Domain specific information, on the other hand, refers to when a given position occurs. In the model of Kakadiaris [11], motion estimation of human movement is obtained from multiple cameras. Bowden et al [12] mentioned the reconstruction of 3D posture from monocular image sequences. Just like reconstructing 3D face shape from a single image based on a morphable model [13], Bowden et al reconstructed the 3D

structure of a subject from a single view of its outline. In fact, only separate 3D poses can be reconstructed without using temporal correspondence. Recently, Troje [9] developed a framework that transformed biological motion data into a linear representation. Linearizing the data by using a PCA-based representation, he constructed a sex classifier with a reasonable classification performance. Giese et al [14] presented a learning-based approach for the representation of complex motion patterns based on linear combinations of prototypical motion sequences. Both authors [9,14] pointed out that spatio-temporal correspondences between image sequences are the key issue for the development of efficient models that perform well on recognition and classification tasks. The resulting models have been termed morphable models since linear combinations of existing data result in motions which represent members of the same motion class as the original motions.

Locomotion patterns such as walking and running are universal yet complex human activities. Locomotion involves a high level of interaction between the central nervous system and various muscles to allow an individual to keep the body upright and balanced while moving around in an orderly and stable manner. The solutions to the problem of generating a stable gait depend on masses and dimensions of the particular body and its parts and are therefore highly individualized. Gait patterns can in fact serve biometric purposes and have been used in the context of recognition and tracking tasks [3,4].

In this report, we will explore the problem of reconstructing 3D human walking from 2D motion sequences. Human walking is a special class of motion characterized not only by abundant similarities but also by variations along many different lines. Given a set of walking data represented in terms of a morphable model, the similarities are represented by the average motion pattern, while the variations are expressed in terms of the covariance matrix. Based on the set of walking data, principal component analysis (PCA) can be used to find a low dimensional, orthonormal basis system that would efficiently span a motion space in which individual walking patterns are approximated in terms of a linear model. We offer a representation that is able to capture the redundancy in gait patterns in an efficient way. The vision problem is not the purpose of this paper and here we make two assumptions that facilitate and focus the problem at this point. Firstly, we assume that the image processing problem of tracking feature points in the 2D image plane is solved. We represent human motion in terms of trajectories of

a series of markers located at the main joints of the body. Figure 1 illustrates the locations of these markers. Secondly, 2D motion sequences are assumed to be orthographic projections of 3D motion. Both assumptions are not critical to the general idea outlined in this paper. The tracking problem can be treated completely independently and in fact there are many laboratories working on this problem [5~7].

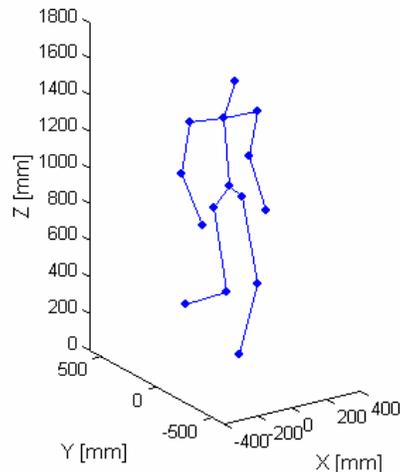


Figure 1. Human motion representation by joints

Establishing spatio-temporal correspondences is a prerequisite for constructing a linear morphable model. Here, our input data consist of the motion trajectories of discrete marker points so that spatial correspondence is solved in a canonical way and not an issue. Since we use low-dimensional Fourier expansions to approximate our data, establishing temporal correspondence is reduced to adjusting frequency and phase of the Fourier components.

We systematically develop a theory for recovering 3D walking data from 2D motion sequences. Section 2 briefly describes data acquisition with a marker-based motion capture system. The details of the linear model and the 3D motion reconstruction are given in Section 3. We reconstruct 3D walking data and evaluate the proposed method in Section 4. Finally, some conclusions and possible future extensions are given in Section 5.

2. Walking data acquisition

Eighty participants served as subjects to acquire walking data. A set of 41 retro-reflective markers was attached to their bodies. Participants were requested to

walk on a treadmill and 3D motion data were recorded using a motion capture system (Vicon; Oxford Metrics, Oxford, UK) equipped with 9 CCD high-speed cameras. The system tracked the 3D trajectories of the markers with a sub-millimeter spatial resolution and a sampling rate of 120 Hz. From the trajectories of these markers we computed the location of “virtual” markers positioned at the major joints according to a standard biomechanical model (BodyBuilder, Oxford Metrics). Each posture is described in terms of 15 markers positioned at the major joints of the body as illustrated in Fig. 1.

3. Motion reconstruction

Using the obtained walking data, we first construct a linear morphable model by PCA from Fourier representations of human examples, and then reconstruct the missing motion data from the linear model.

3.1 Linear model

Human walking can be regarded as a time series of postures $p_r(t) : t=1,2,\dots; T_r, r=1,2,\dots, R$, where R is the number of walkers and T_r is the number of sampled postures for walker r . Because each joint has three coordinates, the representation of a posture $p_r(t)$ including 15 joints is a 45-dimensional vector.

Human walking can be efficiently described in terms of low-order Fourier series [15]. A compact representation for a particular walker, which we call here Fourier representation, utilizes the average posture p_0 , the characteristic postures of the fundamental frequency (p_1, q_1) and second harmonic (p_2, q_2) of a discrete Fourier expansion, and the fundamental frequency (ω) itself to characterize a particular walker:

$$p(t) = p_0 + p_1 \sin(\omega t) + q_1 \cos(\omega t) + p_2 \sin(2\omega t) + q_2 \cos(2\omega t) \quad (1)$$

Since the average posture and each of the characteristic postures are 45-dimensional vectors and the frequency is a scalar, the dimensionality of a Fourier representation at this stage is $45 * 5 + 1 = 226$. The fundamental frequency ω is used to obtain 3D periodic human motion from Fourier representation.

For each specific walking pattern $p_r(t)$, using Eq.(1), we can get a Fourier representation w_r :

$$w_r = (p_{0,r}, p_{1,r}, q_{1,r}, p_{2,r}, q_{2,r}, \omega_r) \quad (2)$$

The advantage of Fourier representation is that it provides the possibility to successfully apply linear operations and to determine temporal correspondence. Temporal information is included in the frequency and phase. After computing the frequency and phase of a walking series by Fourier analysis, we can represent the walking series with zero phase and frequency-independent characteristic postures as shown in Eq. (1). Consequently, temporal correspondence is obtained by scaling and translating the time axis.

Every Fourier representation can be treated as a point in a 226-dimensional linear space. We applied PCA to all the Fourier representations in order to learn the main motion variation and to further reduce dimensionality. All the Fourier representations are concatenated into a matrix W with each column containing one walker w_r , concatenating $p_{0,r}, p_{1,r}, q_{1,r}, p_{2,r}, q_{2,r}$ and ω_r in turn. Computing PCA on the matrix W results in a decomposition of each parameter set w_r into an average value \bar{w} and a series of orthogonal eigen-walkers e_n :

$$w_r = \bar{w} + \sum_n k_{n,r} e_n \quad (3)$$

$\bar{w} = (1/R) \sum w_r$ denotes the average value of all the R columns and $k_{n,r}$ is the coefficient for walker r . A linear model is spanned by the first N eigen-walkers $e_1 - e_N$, which represent the principal variation. The exact dimensionality N of the model depends on the required accuracy of the approximation but will generally be much smaller than R . A Fourier representation of walker w_r can now be represented as a linear combination of eigen-walkers by using the obtained coefficients $k_{n,r}$.

3.2 Reconstruction

We denote a single 2D motion sequence as $\hat{p}(t) : t=1,2,\dots, T$, which is represented in terms of its discrete Fourier components:

$$\hat{w} = (\hat{p}_0, \hat{p}_1, \hat{q}_1, \hat{p}_2, \hat{q}_2, \omega) \quad (4)$$

The average posture \hat{p}_0 , and the characteristic postures $\hat{p}_1, \hat{q}_1, \hat{p}_2$ and \hat{q}_2 only contain 2D information. These postures and frequency are concatenated into a column vector \hat{w} with $30 * 5 + 1 = 151$ entries.

Reconstructing the full 3D motion means to find the right solution w of the equation:

$$\hat{w} = Cw \quad (5)$$

Here $C : \mathfrak{R}^{226} \mapsto \mathfrak{R}^{151}$ is the projection matrix and at this point we assume that it is known. In some cases with an unknown viewpoint, the projection matrix is a function of the view angle. In the following we only formulate the reconstruction with a known viewpoint and those cases with an unknown viewpoint can be formulated by the same procedure. Equation (5) is an under-determined equation system and can only be solved if additional constraints can be formulated. This is done in terms of the linear model outlined in the previous section. Substituting Eq. (3) into Eq. (5), we get:

$$\hat{w} = C \left(\bar{w} + \sum_{n=1}^N k_n e_n \right) \quad (6)$$

Denoting $\hat{\bar{w}} = C\bar{w}$ and $\hat{e}_n = Ce_n$, Eq. (6) can be rewritten as:

$$\hat{w} - \hat{\bar{w}} = \sum_{n=1}^N k_n \hat{e}_n \quad (7)$$

or, using matrix notation, as:

$$\hat{w} - \hat{\bar{w}} = \hat{E}K \quad (8)$$

This is an over-determined linear equation system. We approximate a solution according to a least-square criterion using the pseudo-inverse:

$$K = (\hat{E}^T \hat{E})^{-1} \hat{E}^T (\hat{w} - \hat{\bar{w}}) \quad (9)$$

After the coefficients K are estimated, the missing information $\tilde{w} = (\tilde{p}_0, \tilde{p}_1, \tilde{q}_1, \tilde{p}_2, \tilde{q}_2)$ of Fourier representation is synthesized in terms of the respective linear combination of eigen-walkers $e_1 - e_N$. The reconstructed Fourier representation w is the combination of the 2D Fourier representation \hat{w} and the reconstructed \tilde{w} :

$$w = [\hat{w}, \tilde{w}] = ([\hat{p}_0, \tilde{p}_0], [\hat{p}_1, \tilde{p}_1], [\hat{q}_1, \tilde{q}_1], [\hat{p}_2, \tilde{p}_2], [\hat{q}_2, \tilde{q}_2], \omega) \quad (10)$$

3D periodic human motion is obtained from the reconstructed Fourier representation w by using Eq. (1).

4. Experiments

Using walking data acquired with a marker-based motion capture system, we conducted two different experiments. In the first experiment we assume that the viewpoint from which the 2D test walker is seen is known. The 2D motion sequence was the orthographic projection of one 3D walker onto a vertical plane from

a known viewpoint. The missing dimensions were reconstructed and evaluated at one degree interval from the left (negative) to the frontal and finally to the right (positive) profile viewpoints. In the second experiment, we test whether we can retrieve 3D motion even without knowing the horizontal viewpoint from which a walker is seen. The 2D motion sequence was the orthographic projection of a 3D walker onto a vertical plane from an unknown viewpoint. In both experiments, we used a leave-one-out cross validation procedure: One walker is excluded from the data base when creating the linear model and is later used for testing. This procedure is repeated for every walker.

4.1. Linear model construction

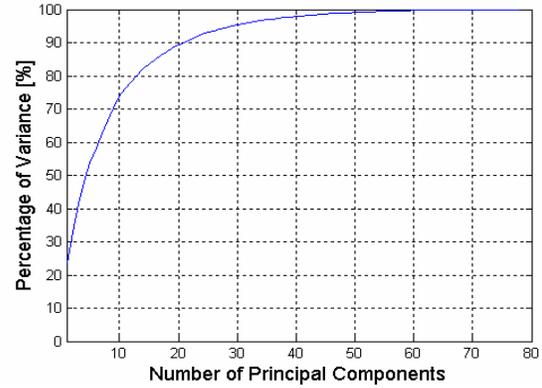


Figure 2. The cumulative variance covered by the principal components computed across 80 subjects

Fourier representations were applied separately to all walking data. On average, the fundamental frequency covered 91.9% of the total variance and the second harmonic accounted for another 6.0%, which means that the first two harmonics explain 97.9% of the overall postural variance of a walker. The sets of all the Fourier representations were now submitted to a PCA. The first 28 principal components accounted for 95% variance of all the Fourier representations, so they were chosen as eigen-walkers in the linear model (Fig. 2).

4.2. Motion reconstruction

Projecting the Fourier representation of a 2D motion sequence onto the linear model can lead to a 3D Fourier representation. 3D walking data were reconstructed at one degree intervals from the left profile-view to the right profile-view. Figure 3

illustrates the original and the reconstructed motion sequences for one subject from three viewing angles (0, 30 and 90 degrees), corresponding to the top, middle and bottom rows. In order to better visualize the reconstructed motion, we displayed the results from the directions orthogonal to missing directions. This figure shows five equal interval postures in one motion cycle. The dashed lines and the solid lines are the original and the reconstructed motion sequences, respectively. The results showed that the original and the reconstructed walking sequences were very similar in the corresponding postures.

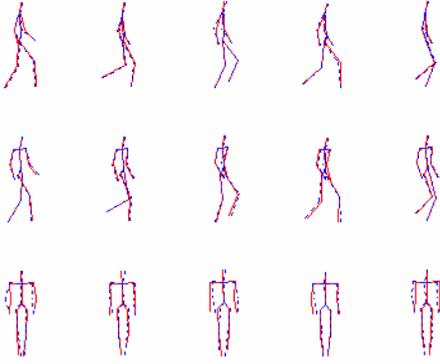


Figure 3. Original and reconstructed walking data from three different viewing angles.

In the experiments, the 2D motion sequence is the orthographic projection of a 3D walker onto a vertical plane, which means that the actual motion in the missing dimension is known beforehand. Therefore, the reconstruction method can be evaluated by comparing the original data with the reconstructed data. In order to give a quantified evaluation, an absolute mean square reconstruction error is defined for each joint:

$$E_{abs}(j) = \frac{1}{T} \sum_{t=1}^T (p_m^j(t) - \tilde{p}^j(t))^2 \quad (11)$$

where $p_m^j(\cdot)$ and $\tilde{p}^j(\cdot)$ are the original and reconstructed data for the j^{th} ($j=1,2,\dots,15$) joint in the missing dimension. From different viewpoints, human walking will have different variances, so a relative mean square reconstruction error is also defined for each joint:

$$E_{rel}(j) = \frac{1}{T\sigma_m^2} \sum_{t=1}^T (p_m^j(t) - \tilde{p}^j(t))^2 \quad (12)$$

where σ_m^2 is the average overall variance of the 15 joints in the missing dimension. The reconstruction error of one walker is the average value of the reconstruction errors for the 15 joints:

$$E_r = \frac{1}{15} \sum_{j=1}^{15} E(j) \quad (13)$$

In Eq. (13), E denotes the absolute reconstruction errors or the relative reconstruction errors of each joint.

The average reconstruction errors among all the walkers for a certain viewpoint were calculated and Fig. 4 shows these absolute average errors and relative average errors from different viewpoints (-90~90 degrees). As can be seen from Fig. 4(a), there are different absolute reconstruction errors in the negative and positive oblique viewpoints, which implies that human walking is slightly asymmetric. The very small asymmetry of variances in Fig. 4(b) also confirms this point. Gait symmetry and asymmetry is a popular topic of research [16] since human gait interchanges the left and right legs and the opposite side of arms with a phase shift of half a period. Both legs perform the same motion but out of phase from each other by half a period. In most cases, these motions are operated in space and time, satisfying the rules of spatial and temporal symmetry. Therefore, it is reasonable to assume that normal gait is symmetric. In fact, in able-bodied gait, there are differences in the right and left limbs, so gait asymmetry was also addressed in the literature [16]. Our experimental results provide evidence that the actions of four limbs are sometimes asymmetric during human walking. Human body expends a specific amount of energy for locomotion and it has been suggested that asymmetric walkers should expend more energy than symmetric walkers. Our reconstruction method can analyze amputee's gait to get symmetric artificial limbs to save energy.

Figure 4(c) illustrates that the frontal and the oblique viewpoints have small relative reconstruction errors, while the profile viewpoints have great relative reconstruction errors. The proposed method can retrieve most variances in the frontal and oblique viewpoints (near frontal) by the 2D projection of the linear model along these viewpoints. The great errors in the profile and oblique viewpoints (near profile) imply that fewer variances are covered by the 2D projection of the linear model along these viewpoints. The different relative reconstruction errors imply that human walking has more diverse actions in the frontal and oblique viewpoints (near frontal), while the profile and oblique viewpoints (near profile) contain more similar actions.

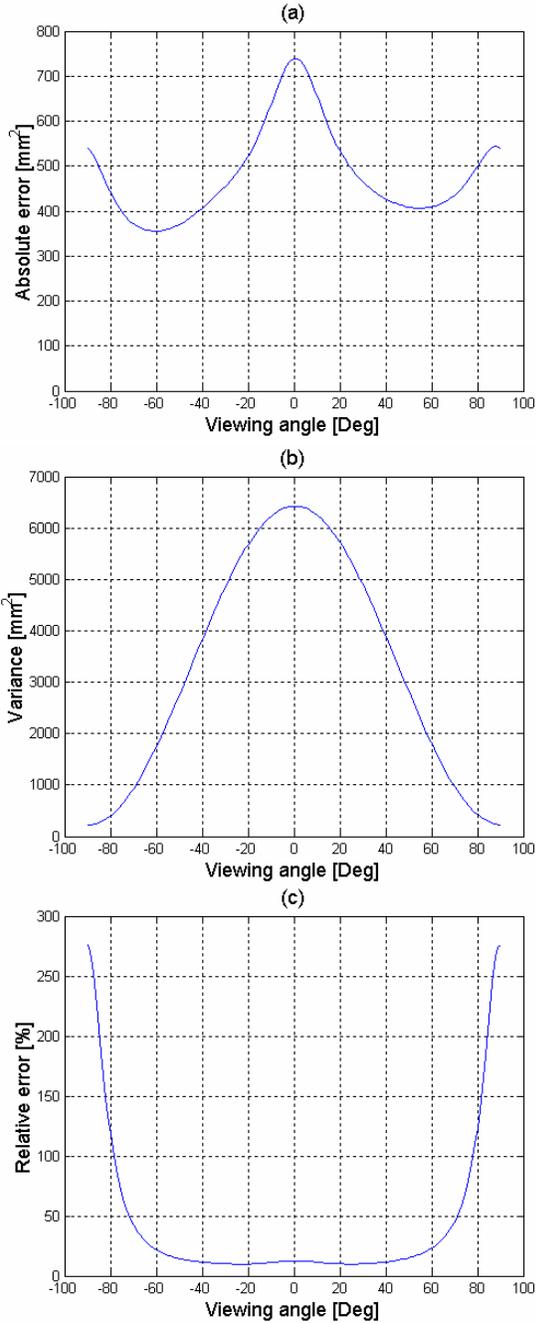


Figure 4. Illustration of the average reconstruction errors from different viewpoints of -90~90 degrees for 80 walkers. (a) Absolute reconstruction errors. (b) Variance. (c) Relative reconstruction errors.

4.3. Viewpoint reconstruction

Using 3D walking examples to generate 2D motion sequences and to build up the linear model, we can reconstruct and evaluate the viewpoint of a 2D walker. Assuming that a 2D walking sequence remains constant walking direction and the viewpoint has only rotation around the vertical axis (that is, horizontal viewpoint), we can estimate the view angle using the rotated average posture $\bar{w}(\alpha)$ and the rotated eigen-walkers $E(\alpha)$. Equation (8) can be represented as:

$$(\alpha_{opt}, K_{opt}) = \arg \min_{\alpha, K} \left\| \hat{w} - \hat{w}(\alpha) - \hat{E}(\alpha)K \right\| \quad (14)$$

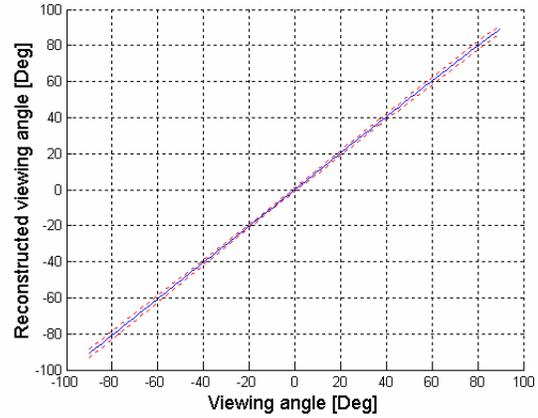


Figure 5. Illustration of the average reconstruction view angles from different viewpoints of -90~90 degrees for 80 walkers

The optimum solution of this minimization problem can be found by solving a nonlinear over-determined system and the view angle is α_{opt} . As such, the missing data is the linear combination of eigen-walkers by using the optimum coefficient K_{opt} . As in the first experiment, a leave-one-out procedure was applied to all the walkers and the average reconstruction angles from different viewpoints are illustrated in Fig. 5. The solid line is the average reconstructed angles and the dashed lines are the average standard deviations. The experimental results show that we can precisely obtain the viewpoints by our proposed linear model. It is possible to identify and recognize human gait from 2D image sequences of different viewpoints by using the obtained view angle and missing data. The poor reconstruction in profile views and the good reconstruction in frontal views fit with data on human performance in biological motion recognition tasks [9,17].

5. Conclusion and future work

We have investigated and evaluated the problem of reconstructing 3D periodic human motion from 2D motion sequences. A linear model, which is based on linear combination of eigen-walkers, was constructed from Fourier representations of the examples by using PCA. Fourier representation of a particular subject is a compact description of human walking, so not only is temporal correspondence determined by the frequency and phase, but also computational expenditure and storage space requirement are saved. Experiments and quantified evaluations were made on walking data. The results and quantified error analysis explained that the proposed method can reasonably reconstruct 3D human walking data from 2D motion sequences. We also verified that the linear model can estimate parameters like the view angle of walking direction, from 2D motion sequences with constant walking direction. It is possible to provide a view-independent gait identification and recognition method from monocular motion sequences. Another application of the proposed model is to use it as a prior for video-based tracking.

In our study, we assume that the positions of the joints are known. In practical applications, recognizing and tracking joints is a challenging task. Human motion is non-rigid and pivotal rotation about some joints, so the linear model is limited to capture global variations in human activities. In our future work, non-linear motion models will be incorporated into the reconstruction method and nonperiodic human motion reconstruction will also be studied by dynamic programming to determine temporal correspondence.

7. References

- [1] J. Aggarwal and Q. Cai. Human Motion Analysis: A Review. *Computer Vision and Image Understanding*, 73(3): 428-440, 1999.
- [2] L. Wang, W. Hu and T. Tan. Recent Developments in Human Motion Analysis. *Pattern Recognition*, 36(3): 585-601, 2003.
- [3] D. Cunado, M. Nixon and J. Carter. Automatic Extraction and Description of Human Gait Models for Recognition Purposes. *Computer Vision and Image Understanding*, 90(1): 1-41, 2003.
- [4] L. Wang, T. Tan, H. Ning and W. Hu. Silhouette Analysis-based Gait Recognition for Human Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12): 1505-1518, 2003.
- [5] A. Jepson, D. Fleet, and T. El-Maraghi. Robust Online Appearance Models for Visual Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10): 1296-1311, 2003.
- [6] H. Ning, T. Tan, L. Wang and W. Hu. Kinematics-Based Tracking of Human Walking in Monocular Video Sequences. *Image and Vision Computing*, 22(5): 429-441, 2004.
- [7] R. Urtasun and P. Fua. 3D Tracking for Gait Characterization and Recognition. In the 6th International Conference on Automatic Face and Gesture Recognition, 2004, South Korea
- [8] T. Moeslund and E. Granum. A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3): 231-268, 2001.
- [9] N. F. Troje. Decomposing Biological Motion: A Framework for Analysis and Synthesis of Human Gait Patterns. *Journal of Vision*, 2(5): 371-387, 2002.
- [10] J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer-Verlag, 1997.
- [11] I. Kakadiaris and D. Metaxas. Model-Based Estimation of 3D Human Motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12): 1453-1459, 2000.
- [12] R. Bowden, T. Mitchell and M. Sarhadi. Non-linear Statistical Models for the 3D Reconstruction of Human Pose and Motion from Monocular Image Sequences. *Image and Vision Computing*, 18(10): 729-737, 2000.
- [13] V. Blanz and T. Vetter. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9): 1063-1074, 2003.
- [14] M. Giese and T. Poggio. Morphable Models for the Analysis and Synthesis of Complex Motion Patterns. *International Journal of Computer Vision*, 38(1): 59-73, 2000.
- [15] M. Unuma, K. Anjyo and R. Takeuchi. Fourier Principles for Emotion-based Human Figure Animation. *Computer Graphics Proceedings of SIGGRAPH 95*, 91-96, 1995.
- [16] H. Sadeghi, P. Allard, F. Prince, H. Labelle. Symmetry and Limb Dominance in Able-Bodied Gait: a Review. *Gait and Posture*, 12(1): 34-45, 2000.
- [17] G. Mahter and L. Murdoch. Gender Discrimination in Biological Motion Displays Based on Dynamic Cues. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 258, 273-279, 1994.