3D Human Animation from 2D Monocular Data Based on Motion Trend Prediction

Li Zhang
Department of Computing
Curtin University of Technology
Perth, WA 6845
GPO Box U1987 Perth
zhangl@cs.curtin.edu.au

Ling Li
Department of Computing
Curtin University of Technology
Perth, WA 6845
GPO Box U1987 Perth
ling@cs.curtin.edu.au

ABSTRACT

A model-based method is proposed in this paper for 3-dimensional human motion recovery, taking un-calibrated monocular data as input. The proposed method is able to generate smooth human motions that resemble the original motion from the same viewpoint the sequence was taken, and look continuous from any other viewpoint. The core of the proposed system is the motion trend prediction for reconstruction. To focus the research effort on motion reconstruction, "synthesized" input is first employed to ensure that the reconstruction algorithm is developed and evaluated accurately. Experiment results on real video data indicate that the proposed method is able to recover human motion from un-calibrated 2D monocular images with very high accuracy.

Keywords: human animation, 3D motion reconstruction, motion trend prediction

1. INTRODUCTION

Animation is the production of consecutive images, which, when displayed, conveys a feeling of motion [HOB99]. In the past decade, with the rapid development of computer technology, computer animation has become very popular in many applications. In computer animation, the representation of human body and its motion receives great attention, since human animation are widely employed in many areas, such as games, movies, surveillance, scientific visualization, etc. As monocular images and video sequences are easily available, many great efforts have been made to reconstruct 3D human motion from monocular images. However, such attempt remains very much under-developed due to many technical difficulties.

[Tay00] and [RR03] suggested adjusting the posture of a human model according to camera calibration information and biomechanical constraints applied on the model. Orthographic projection is used in their approaches, which is greatly different from the perspective projection used in any real camera. [LZP99] and [PCS02] made use of the motion library. The former took motion attributes achieved through reconstruction as guidance for estimation of unknown human motion, while the latter use motion library to resolve the depth ambiguity in recovering 3D configuration from 2D image features. In both attempts, a large motion library needs to be maintained and upgraded continuously. In [CB04] the concept of prioritized constraints is introduced. Based on it the proposed method can get quite good results, which is only suitable for adding variations to motions known before the reconstruction. [DTJR01] introduced an interactive system which combines biomechanical constraints on 3D motion with user interferences to reconstruct sequences in 3D; similarly three possibilities for solving inverse kinematics problem during human animation are discussed when interactive direct manipulation is applied [Fěd03]. There are also attempts in automatically generating accurate inverse dynamics solutions to simulate and deform human motion [TSF04] and [KM04], however such efforts have been concentrated on hand posture recovery only. [ZL04] proposed a Criterion Function (CF) to represent the residuals between the image feature and the projected features from the reconstructed 3D model in a Global Adjustment (GA) system. In their

SHORT COMMUNICATIONS proceedings,
ISBN 80-86943-05-4
WSCG'2006, January 30-February 3, 2006
Plzen, Czech Republic.

method the accuracy and the consistency of the recovered postures are only guaranteed from the same viewing direction as the original.

Most existing methods introduce simplifications on human motion or require assistance such as user interference or motion library. This paper aims to propose a novel model-based human motion reconstruction method from un-calibrated 2D monocular data without user interferences and the human motion is truly unrestricted.

The key component of our proposed system is a Motion Trend Prediction (MTP) technique which aims to achieve the posture reconstruction at every frame based on information from previous frames (except the 1st frame). Unlike the proposed method for estimating 3D motion of an articulated object in [HNHQ94], we do not divide our skeletal model into many small components, recover their motions separately and integrate them together to compose the motion of the skeleton itself. In that way although the estimation of each small component might be simple and feasible, the recovered motion of the entire skeleton cannot be ensured to be favorable. In our algorithm, the whole skeletal model is always treated as a single object to ensure the consistency between motions of different body parts and their smoothness. Some systems utilized silhouettes for estimation of the posture with human body depth and collision constraints [MG01], where only arms were studied. Later such efforts were extended [FRDC04]. To better produce silhouettes from 2D monocular view and to resolve the 3D pose prediction ambiguities Principal Component Analysis (PCA) and Radial Basis Function (RBF) are both employed. Besides them, motion library and key frame technique are also needed for accurately rendering 3D animal gaits. Simplicity and computational efficiency is sacrificed in the method to achieve accurate motion simulation. Compared to them our proposed MTP technique is fairly simpler yet more accurate. As the MTP technique is a per-frame approach, we introduce a filtering process to smooth the results. The filtering algorithm can be repeated if necessary in order to re-enforce important constraints. Therefore the accuracy of the MTP can be strictly guaranteed. Further more, unlike the Kalman or the particle filtering prediction tracking algorithm [WB95], calculations of partial derivatives or an even more complicated infrastructure are not necessary in our algorithm. Through a very small set of simple Deviation Function (DF) equations, 3D human postures in a motion sequence can be tracked and recovered. The settings of the weighting parameter (WP), which plays an important role in determining the MTP's accuracy, can be easily found through a low-pass filtering process as well. The MTP technique proposed here is very simple, but it can efficiently generate smooth human animation. It demonstrates great advantages over most of the methods proposed before, when only 2D monocular data is available.

The rest of this paper is organized as follows: 3D skeleton model used in motion reconstruction is introduced in Section 2; in Section 3 assumptions on the input data are explained; MTP and the detailed reconstruction procedure are described in Section 4; experimental results are presented and analyzed in Section 5; and a brief conclusion is presented in Section 6.

2. MODEL

To reconstruct 3D motion a camera model and a human model are set up. The camera model is located at a fixed position in virtual space with pre-defined focal length, and it does not require knowledge of the actual cameras from which the video sequence is taken. This project focuses on recovering the whole body motion from monocular data; hence an articulated 3D skeletal model as shown in Fig.1 is sufficient for our purpose. The joint *pelvis* is set as the root in the skeletal tree structure, and the 5 leaf joints are *left wrist, right wrist, left ankle, right ankle* and *head*.

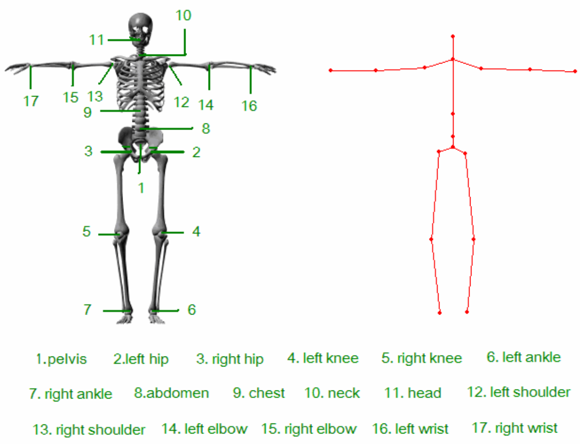

Figure 1: 3D skeletal model, its 2D stick figure and the 17 joints

At the current moment, our attention is focused on the motion reconstruction for segments including pelvis, waist, chest, neck, upper arms, forearms, thighs, and crura. The tips of hands and feet, and the top of the head are not considered as individual joints in our work. Their motions will be determined by the motions of forearm, crus and neck segments respectively. Besides, the leaf joints are assumed to have no degrees of freedom (DOF) since they are located at the end of skeletal branches. Movements of these leaf joints are resulted from the rotations of the connecting parent joints. For reconstruction purpose, we assign each intermediate joint (joints that are not leaf joints mentioned above) 1 to 3 DOF(s) in the world coordinate system (WCS), and each intermediate joint is associated with a local coordinate system (LCS) as shown in Fig.2. However there are 6 DOFs (3 for translations and the other 3 for rotations) at *pelvis*, since the translation of the whole body is represented by the movement of *pelvis*. Therefore our human model actually has 17 joints, 12 segments and 37 DOFs.

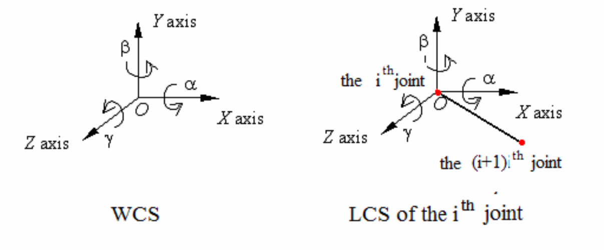

Figure 2: WCS and LCS of the i^{th} joint

The rotational angles of each intermediate joint about the 3 axes in its associated LCS are governed by biomechanical constraints [WBF87, RGBC96, PW99]. The joint constraints in [ZL04] are used here with some modifications as listed in Table 1.

	x-axis	*y*-axis	*z*-axis
Chest	-25 ° ~ 30 °	-30 ° ~ 30 °	-15 ° ~ 15 °
Left Shoulder	-90 ° ~ 90 °	-135 ° ~ 45 °	-135 ° ~ 90 °
Left Elbow	0 ° ~ 0 °	-135 ° ~ 0 °	0 ° ~ 0 °
Right Shoulder	-90 ° ~ 90 °	-45 ° ~ 135 °	-90 ° ~ 135 °
Right Elbow	0 ° ~ 0 °	0 ° ~ 135 °	0 ° ~ 0 °
Left Knee	0 ° ~ 135 °	0 ° ~ 0 °	0 ° ~ 0 °
Right Knee	0 ° ~ 135 °	0 ° ~ 0 °	0 ° ~ 0 °

Table 1: Modified angular constraints based on biomechanical constraints

Through proper translations of the *pelvis* joint and rotations of the segments about each joint in the skeletal model, any human posture could be obtained from the initial pose in Fig.1.

3. PREPARATION OF DATA

In order to generate reliable motion reconstruction, 2D feature information, such as the joints' positions on the 2D image plane, extracted from the source images or video must be highly accurate. Any error in feature extraction could lead to incorrect 2D configuration of human body geometry and human posture. There have been a large number of approaches to feature extraction in the image processing and computer vision area [AT04, BK03, GODC05, LF04, LZZP04, SJ04, TSF04]. However up to date, no technique is able to produce 2D image feature extraction that is sufficiently accurate. Extraction error remains an unavoidable issue. As a matter of fact, such inaccuracy is one of the main factors preventing the progress of reconstruction technology. Besides, the lack of depth information in monocular image source makes it extremely difficult to evaluate the performance of any 3D reconstruction algorithm.

The extraction of 2D feature information and the 3D motion reconstruction are actually two independent modules. They should be addressed separately and in parallel for the ultimate development of the complete 3D motion reconstruction system. In this project, the research is focused on the 3D motion reconstruction. To ensure the proper development and evaluation of the reconstruction algorithms, computer synthesized source videos is first used. The popular BVH (Bio-vision hierarchical data) motion files are employed in any computer animation software to generate various animation series on a fixed human model with known geometry. For the time being, the motion is generated for a skeletal model to show the motion more clearly. The 3D animation is projected on the image plane to produce the 2D joint features in every frame, which is the only input to our motion reconstruction system. The 3D motion data and the camera settings are not utilized in any way during the reconstruction process. Such information can later be used for comparison purpose after the motion is reconstructed in 3D to ensure proper evaluation of the reconstruction algorithms developed.

The reconstruction algorithms developed can be used on any kind of actual monocular video to produce reconstructed human motion truthful to the extracted 2D feature information. The "synthesization" part of this project is merely to enable accurate evaluation and assessment of any reconstruction algorithms. The algorithms are later tested on real video to demonstrate their applicability.

Manual adjustment for the 1st frame is required in our system, since information of the 1st frame is the starting point for the MTP. The human posture and body location recovered at the 1st frame have to be very accurate. The skeletal model will first be resized to fit the human geometry shown in the 2D monocular data. Human posture in the 1st frame will then be calculated and recovered using MTP as discussed in the next section. Minor manual rotations about certain joint might be necessary to ensure that the recovered posture resembles the original one in all details. Such manual adjustment is only allowed in the first frame. The reconstruction process of all other frames is completely automatic and user interference is not allowed and totally unnecessary.

4. MOTION TREND PREDICTION

Human motion reconstruction is actually a process to reconstruct human posture at every frame. The goal is to recover human postures which resemble the original postures as much as possible. Our algorithm is based on two observations about human motion. 1. Despite the complexity of human motion, there are actually only two types of movements involved: the translation of the whole body and the rotations of the body segments about the joints. The former put the human body in a certain location, while the latter generates a particular body posture. Hence, the 3D movement of any joint can be treated as the composition of the following transformations: translation of the whole body, and rotations of all ancestor joints of this particular joint. 2. Most of the human motion is assumed generally smooth (although it's not so true in real human motion), which means that the 3D human postures in neighboring frames are similar.

Based on the above observations, a Motion Trend Predication technique is proposed. As discussed in Section 1, the technique has many advantages over the existing methods. It is developed for two different types of human motion: human motion with or without the body relocation.

4.1. Motion Reconstruction without Body Relocation

In this case, there is no body relocation in the input monocular human motion, which means that the position of the joint *pelvis* is fixed. Therefore the movement of each joint can be considered as a composition of the rotations about all its ancestor joints. According to kinematics principles [IC05], the adjusting order applied in posture reconstruction should be from the root to the leaf joints. The *pelvis* is taken as the origin of the WCS. By minimizing the residuals between the skeleton's projected figure and the 2D image feature, all joints can be rotated to certain 3D positions, which ensure the recovered posture resembles the original one from the same viewpoint. However recovering posture at each frame individually may violate the inter-frame consistency; to ensure natural looking motions could be produced effectively with the MTP, two important 3D predictors are introduced. When a segment connecting the i^{th} joint and one of its direct descendants is being rotated about the i^{th} joint, coordinates of that descendant should be utilized to evaluate if the segment has been rotated to the correct location. Since the 3D coordinates of any joint in neighboring frames should be similar to ensure smooth motion, the 3D coordinates of every joint in the WCS could be used for motion prediction. Hence the first 3D predictor is defined to represent the distance between a certain joint's 3D positions at neighboring frames as follows:

$$predictor_{3D\_pos\_i} = \sqrt{(x^K - x^{K-1})^2 + (y^K - y^{K-1})^2 + (z^K - z^{K-1})^2}$$
$$(K \geq 2) \qquad (1)$$

where (x^K, y^K, z^K) and (x^{K-1}, y^{K-1}, z^{K-1}) stand for the WCS coordinates of the i^{th} joint's direct descendant(s) at the K^{th} and $(K\text{-}1)^{th}$ frame respectively. Another 3D predictor is derived from the *Z* component of each joint. This predictor is used to stress the movements caused by the absolute depth changes in WCS:

$$predictor_{z\_component\_i} = abs\left(z^K - z^{K-1}\right) \qquad (2)$$

During the reconstruction process, once the reconstruction of a frame is finished, the posture configuration of the skeletal model obtained for this frame will be used as the reference for predicting the human posture in the next frame. In other words, the 3D coordinates of every joint and its Z component in WCS at the $(K\text{-}1)^{th}$ frame will be utilized when recovering the 3D posture at the K^{th} frame.

Therefore a DF is developed as shown in Eq.(3), which is the parametric representation of the MTP technique for adjusting the skeletal model. It looks similar to the popularly-used energy function defined in [WBF87]. However in our DF, 3D predictors are introduced, which are based on the 3D position and Z component of every joint. The new items greatly enhance the accuracy of the reconstruction in the Z direction, which is the main concern in any 3D reconstruction. Meanwhile the way the DF is formulated in our MTP makes the solution process much simpler than all existing methods.

$$\begin{aligned} DF_i = {} & weighting\_parameter_{orientation\_i} \times deviation_{orientation\_i} + \\ & weighting\_parameter_{position\_i} \times deviation_{position\_i} + \\ & weighting\_parameter_{length\_i} \times deviation_{length\_i} + \\ & weighting\_parameter_{3D\_pos\_i} \times predictor_{3D\_pos\_i} + \\ & weighting\_parameter_{z\_component\_i} \times predictor_{z\_component\_i} \end{aligned} \qquad (3)$$

The above DF is in its basic form. When dealing with limbs, possible ambiguities could be resulted from occlusion due to the high flexibility of limbs. A method is proposed to better handle such situations. Usually postures of limbs differ slightly between neighboring frames, thus it is first assumed that poses of the forearms or shins are exactly the same when rotating the upper arms or thighs. This way the ambiguities could be reduced greatly. Before upper arms or thighs are being rotated at the K^{th} frame, configurations of the forearms or shins obtained at the $(K\text{-}1)^{th}$ frame are applied temporarily. The position residuals of the *wrists* or *ankles* between the projection and image features are then included into the DF. Hence the DF for rotating upper arms or thighs is evolved as follows:

$$\begin{aligned} DF_i^{limb} = {} & DF_i + \\ & weighting\_parameter_{leaf\_joint\_i} \times deviation_{leaf\_joint\_i} \end{aligned} \qquad (4)$$

$$deviation_{leaf\_joint\_i\_} = D_2\left(P_{j\_i}, P_{j\_p}\right) \quad (5)$$

Here the j^{th} joint (leaf joint – *wrist* or *ankle*) is an indirect descendant of the i^{th} joint (*shoulder* or *hip*); P_{j_i} is the image feature of the j^{th} joint, and P_{j_p} is its corresponding projection feature, as illustrated in Fig.3.

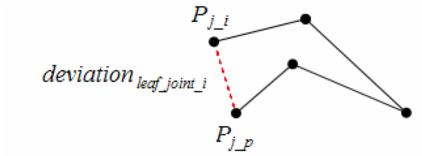

Figure 3: Deviation in the position of indirect child joint (leaf joint)

When the joint *pelvis* is fixed at a certain position in WCS, the 3D coordinates of every joint in WCS are used to define a 3D predictor no matter how the *pelvis* is rotated. Each joint of the skeleton model is rotated one by one from the root to the leaf joints according to the DFs defined above. For instance, when rotating joint *pelvis*, the residuals between the image and projection features of its three direct child joints (*abdomen*, *left* and *right hips*) must be optimized to nearly zero. The 3D positions and Z components of these three joints in the previous frame are taken into consideration as well. This way the 3D position of *pelvis* can be recovered.

Once the posture configuration resulting in the minimum DF values for every joint is obtained, all joint are believed to have been moved to their approximately correct 3D positions. As continuity between recovered postures at each frame is strictly guaranteed by the two 3D predictors in the DF, the whole reconstructed human motion will be smooth, and looks natural from any viewpoint in 3D space.

4.2. General Motion Reconstruction

Next we attempt to reconstruct unrestricted human motions. Two types of movements have to be taken into consideration: the relocation of the human body which is represented by the translations of the *pelvis* joint, and the rotations of every intermediate joint.

Translations parallel to the image plane is easy to implement, since in this case the distance between the human object and the camera is fixed. The DF for recovering such translations is defined as:

$$DF_{translation\_1} = deviation_{position\_pelvis} \quad (6)$$

$$deviation_{position\_pelvis} = D_2\left(P_{pelvis\_i}, P_{pelvis\_p}\right) \quad (7)$$

where P_{pelvis_i} and P_{pelvis_p} are image and projection features of *pelvis* respectively.

However, usually translations perpendicular to the image plane are also present in the movement of the *pelvis*. Such translation is much more difficult to handle since it is not easy to detect when and how such translation exactly happens. It is a common knowledge that the distance of an object with the projection plane determines its projected sizes on the image plane. Such understanding is used to determine the translation of the *pelvis* perpendicular to the image plane. As 3D human postures in consecutive frames should appear similar, we can first assume the human posture already recovered for the $(K\text{-}1)^{th}$ frame and the posture to be recovered in the K^{th} frame to be "the same". Before translating the *pelvis* in the K^{th} frame, the 3D posture configuration obtained at the $(K\text{-}1)^{th}$ frame can be applied to the skeletal model temporarily. If the skeleton model is translated to its correct location at the K^{th} frame, the sum of the projected segment lengths should be equivalent to that of the image features. The following DF is then defined to perform the body translation perpendicular to the image plane:

$$DF_{translation\_2} = deviation_{total\_length} \quad (8)$$

$$deviation_{total\_length} = abs\left(\sum_{i=1}^{16}\left\|\overrightarrow{S_{i\_i}}\right\| - \sum_{i=1}^{16}\left\|\overrightarrow{S_{i\_p}}\right\|\right) \quad (9)$$

where S_{i_i} and S_{i_p} represent lengths of the image feature and projection feature of the i^{th} segment respectively.

If more than 3 frames prior to the current frame have been recovered, the average perpendicular translation factor of the *pelvis* at those frames is used as an additional deviation factor for further refinement.

After the translation of the joint *pelvis*, the 3D skeletal model is considered as positioned to the right location in 3D space. Rotations of every joint will then follow using the methods described in Section 4.1. As the *pelvis* position changes in most of the frames, to recover the human posture through the same DFs as discussed in Section 4.1, a new reference coordinate system (RCS) is introduced. It is defined after the translation of *pelvis* in every frame, with its origin located at *pelvis* and its three axes parallel to those of WCS. Such RCS is used as the substitute of the "WCS" for calculating the 3D predictors in MTP. Based on the RCS, rotations of all intermediate joints can be derived.

5. EXPERIMENTS AND STATISTICS

Varies experiments on both the computer synthesized animation and real video are conducted to test the accuracy of the MTP technique. Monocular sequences with resolution of 640x480 and frame rate of 20fps are used.

5.1. Results from "Synthesized" Data

In this section, the MTP technique is evaluated on computer synthesized monocular video data. During the process a parameter search routine is performed to find the best WP settings to be used in the DFs.

Firstly, two sequences (stretching and sneaking) are reconstructed, both of which mainly concern the motions nearly parallel to image plane. Part of the sneaking action is illustrated in Fig.4.

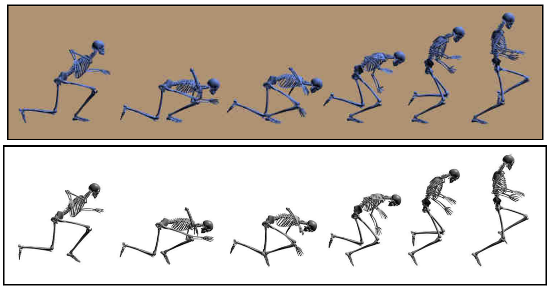

Figure 4: Top: Input motion (Sneaking); Bottom: Reconstruction from the same view direction;

The statistics data obtained during reconstruction from above sequences is shown in Fig.5, where the total 2D residuals of all the 17 joints between image and projection features are presented in form of stacked line. The horizontal axis represents the 17 joints in numerical order as defined in Section 2. It can be seen from Fig.5 that for both sequences, the maximum total residual is below 3 pixels, which is highly satisfactory given the image resolution of 640*480.

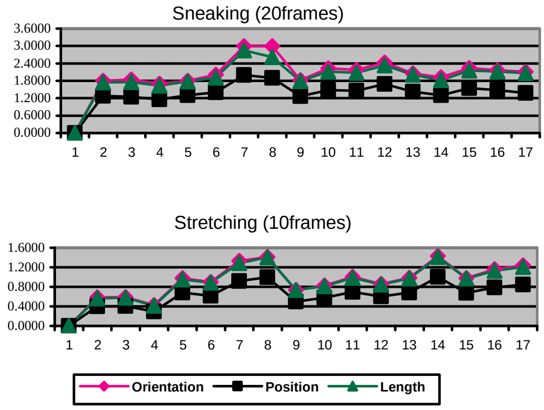

Figure 5: Total 2D deviation value at all frames

Next our efforts are put into animating a human kicking action (22frames), where the joint *pelvis* is moving all the time and the motion is more flexible in 3D space. As shown in Fig.6, the reconstructed results closely resemble the original motion from the same viewing angle.

Fig.7 illustrates the total value of DFs of all joints at each frame for the kicking sequence. Such value reaches its peak at the 11th frame, which is 1.660629 in pixels. Again the result indicates that the projection of the recovered 3D posture at each frame is very close to the original, given the image resolution of 640*480.

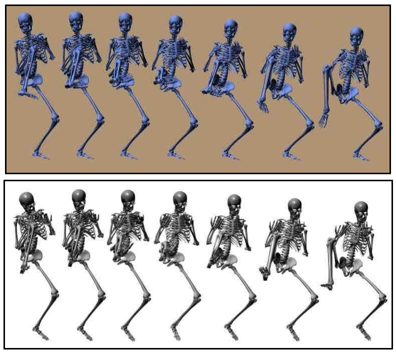

Figure 6: Top: Input motion (Kicking); Bottom: Reconstruction from the same view direction.

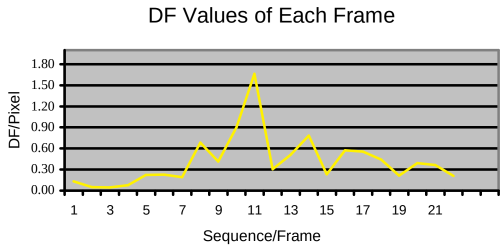

Figure 7: Total DF value at all frames (Kicking)

Since the sequence is computer synthesized, the actual 3D position of every joint is available. Thus the comparison between the original and the reconstructed postures could be done in 3D. In Fig.8 the original and reconstructed motions are still very similar when viewed from another angle.

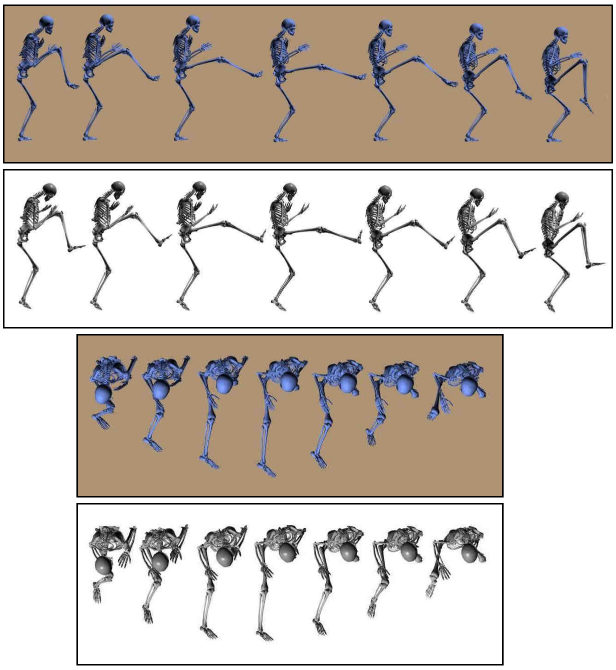

Figure 8: Input and reconstructed motion from other viewing directions

5.2. Results from Real Video Data

Next the technique is tested on real monocular video sequences. One motion sequence composed of walking and squatting (49frames), and the reconstructed human animation based on it are presented in Fig.9. To ensure accurate image feature extraction from real video sequence and to reduce unnecessary noises, color labels are stuck to the human object at joint positions. Image processing techniques such as those mentioned in [BK03, JB04, MRC05] are used to extract the 2D joint features from each frame of the video sequences; however we can only guarantee the noises will be minimized as possible as we can, which will affect the final 3D motion reconstruction.

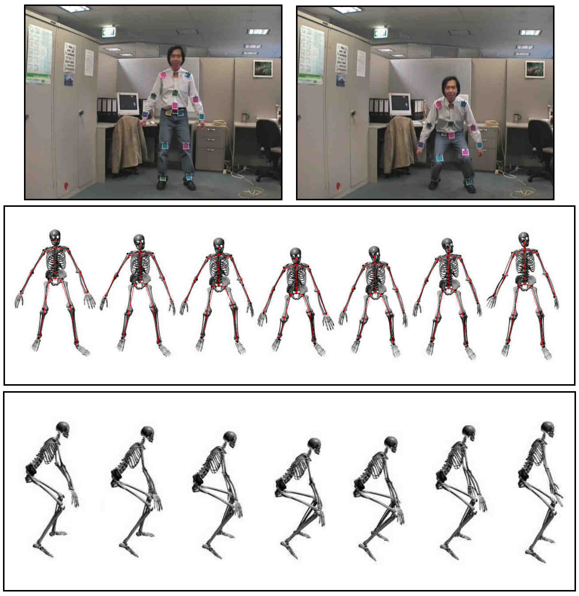

Figure 9: Top: Two frames from the input video Middle: Reconstructed motion - front view Bottom: Reconstructed motion - side view (Walking and Squatting)

Fig.10 shows the total values of DFs of all joints at each frame of this sequence. Compared with Fig.7, the total values of DFs for real video reconstruction is much higher than those from "synthesized" data. However the maximum DF sum value is only 6.28112 at the 18th frame in a resolution of 640*480. The reconstruction from real video still can be considered as highly accurate.

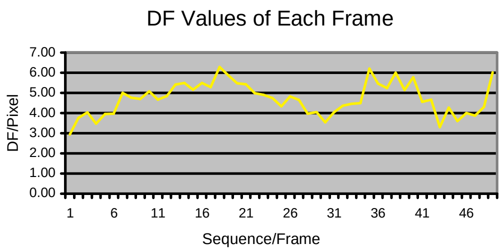

Figure 10: Total DF value at all frames (Walking and squatting)

6. CONCLUSION

A model-based technique is proposed in this paper for motion reconstruction from un-calibrated monocular video sequences containing unrestricted human motion. MTP technique is first developed to reconstruct human motion with no body relocation. The technique is then extended to derive a new RCS at each frame, and hence enable reconstruction of any unrestricted human motion.

The main advantage of our approach is that through it a truly wide range of monocular sequences could be reconstructed, and there is no requirement for camera calibration. From experimental results presented in the paper, the reconstruction results are highly satisfactory as long as the 2D image features are reasonably accurate.

As a future work we are planning to introduce control on the leaf joint into MTP. Plausible motions about these parts are expected to be simulated.

REFERENCES

[AT04] Agarwal A., Triggs B.: Learning to track 3D human motion from silhouettes. In *Proceedings of the 21st International Conference on Machine Learning* (July 2004), pp. 9-16.

[BK03] Barrón C., and Kakadiaris I. A.: A convex penalty method for optical human motion tracking. *First ACM SIGMM international workshop on Video surveillance*, IWVS'03, pp. 1-10.

[CB04] Callennec B. L., Boulic R.: Interactive motion deformation with prioritized constraints. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aug. 2004), pp. 163-171.

[DTJR01] David E. DiFranco, Tat-Jen Cham, James M. Rehg.: Reconstruction of 3D figure motion from 2D correspondences. In *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2001.

[Fěd03] Fdor M.: Application of inverse kinenatics for skeleton manipulation in real-time. In *Proceedings of the 19th spring conference on Computer graphics* (Apr. 2003), pp. 203-212.

[FRDC04] Favreau L., Reveret L., Depraz C., Cani, M. P.: Animal gaits from video. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aug. 2004), pp. 277-286.

[GODC05] Gibson D. J., Oziem D. J., Daltion C. J., Campbell N. W.: Capture and synthesis of insect motion. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aug. 2005), pp. 39-48.

[HNHQ94] Holt R. J., Netravali A. N., Huang T. S., Qian R. J.: Determining Articulated Motion from Perspective Views: A Decomposition Approach. *Pattern Recognition, Vol. 30* (1997), pp. 1435-1449.

[HOB99] Hodgins J. K., O'Brien J. F., Bodenheimer R.

E.: *Computer Animation*. In *Wiley Encyclopedia of Electrical and Electronics Engineering*, John G. Webster, ed., v. 3, 686-690, 1999.

[IC05] Ieronutti L., Chittart L.: A virtual human architecture that integrates kinematic, physical and behavioral aspects to control h-anim characters. In *Proceedings of the tenth international conference on 3D Web technology* (Mar. 2005), pp. 39-48.

[JB04] Juan C.D, Bodenheimer B.: Cartoon texture. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aug. 2004), pp. 267-276.

[KM04] Kurihara T., Miyata N.: Modeling deformable human hands from medical images. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aug. 2004), pp. 355-363.

[LZP99] Liu Xiaoming, Zhuang Yueting, Pan Yunhe.: Video based Human animation technique. In *Proceeding of the 7th ACM International Conference on Multimedia* (Oct. 1999), pp. 353-362.

[LZZP04] Li Chuanjun, Zhai Peng, Zheng S. Q., Prabhakaran B.: Segmentation and recognition of multi-attribute motion sequences. In *Proceeding of the 12th ACM Annual International Conference on Multimedia* (Oct. 2004), pp. 836-843.

[MG01] Moeslund T. B., Granum E.: A Survey of Computer Vision-based Human motion capture. *Computer Vision and Image Understanding, Vol. 81, Issue 3* (Mar. 2001), 231-268.

[MRC05] Müller M., Röder T., Clausen M.: Efficient content-based retrieval of motion capture data. *ACM Transactions on Graphics 24*, 3 (July 2005), 677-685.

[PCS02] Park M. J., Choi M. G., Shin S. Y.: Human Motion Reconstruction from inter-frame feature correspondences of a single video stream using a notion library. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation* (July 2002), pp. 113-120.

[PW99] Popovic Z., Witkin A.: Physically based motion transformation. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (July 1999), pp. 11-20.

[RGBC96] Rose C., Guenter B., Bodenheimer B., Cohen M. F.: Efficient generation of motion transitions using spacetime constraints. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (Aug. 1996), pp. 147-154.

[RR03] Remondina F., Roditakis A.: Human figure reconstruction and modeling from single Image or monocular video sequence. In *Proceeding of the 4th International Conference on 3D Digital Imaging and Modeling* (Oct. 2003), pp. 116-123.

[SJ04] Sminchisescu C., Jepson A.: Generative modeling for continuous non-linearly embedded visual inference. In *Proceedings of the 21st International Conference on Machine Learning* (July 2004), pp. 9-16.

[Tay00] Taylor C. J.: Reconstruction of articulated objects from point correspondences in a single image. *Computer Vision and Image Understanding, Vol. 80, No. 3* (Dec. 2000), 349-363.

[TSF04] Tsang W., Singh K., Fiume E.: Helping hand: an anatomically accurate inverse dynamics solution for unconstrained hand motion. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aug. 2004), pp. 319-328.

[WB95] Welch G., Bishop G..: An introduction for kalman filter. Technical report TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, 1995.

[WBF87] WITKIN A., BARR A., FLEISCHER K.: Energy constraints on parameterized models. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques* (Aug. 1987), pp. 225-232.

[ZL04] Zhao Jianhui, Li Ling.: Human motion reconstruction from monocular images using genetic algorithms. *Computer Animation and Virtual World, Vol. 15, No. 3-4* (July 2004), 407-414.