Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Human skeleton proportions from monocular data

PENG En, LI Ling
(*Department of Computing, Curtin University of Technology, Perth, Australia*)
E-mail: perryp@cs.curtin.edu.au; ling@cs.curtin.edu.au

Received Apr. 11, 2006; revision accepted Apr. 24, 2006

Abstract: This paper introduces a novel method for estimating the skeleton proportions of a human figure from monocular data. The proposed system will first automatically extract the key frames and recover the perspective camera model from the 2D data. The human skeleton proportions are then estimated from the key frames using the recovered camera model without posture reconstruction. The proposed method is tested to be simple, fast and produce satisfactory results for the input data. The human model with estimated proportions can be used in future research involving human body modeling or human motion reconstruction.

Key words: Modeling, Human figure, Monocular
doi:10.1631/jzus.2006.A1266 **Document code:** A **CLC number:** TP39

INTRODUCTION

The problem of creating a virtual human has received a lot of attention in the last few decades, due to the increasing popularity of applications involving human figures, such as movies and computer games. There are many approaches to create a virtual human body. Traditional ways for creating a virtual human figure include employing 3D body scanners or using 3D modeling techniques based on understanding of human anatomy. Detailed 3D human body model can be acquired easily using such methods. However, it is expensive and clumsy to use the scanner, and more seriously, the person to be modeled might not be available for scanning. Meanwhile, the human models based on human anatomy generally fall short in representing personalized individuals. Another way to recover the human body is by using images. Since images are the popular media that record the human figure, recovering human model from images can avoid many limitations in traditional ways.

The research to model human body using images can be divided into two groups: (1) using multi-view images; and (2) using single-view images. Multi-view images are employed by many researchers. Some researchers (Hilton and Gentils, 1998; Lee *et al*., 2000; Villa-Uriol *et al*., 2003) recover the human body from multi-view images of a static human figure, while others reconstruct the virtual human from multi-view images of a dynamic person performing specified motions (Cheung *et al*., 2003; Cohen and Lee, 2002; D'Apuzzo *et al*., 2000; Kakadiaris and Metaxas, 1995; Starck and Hilton, 2003). Methods using multiple cameras suffer the same drawback: the target person has to pose for the cameras at a specific location, normally in a fully-equipped laboratory or studio. On the contrary, single-view images are conveniently available in various formats (e.g., photos, DVD, VHS video tape) to general public. Hence, human body reconstruction from single-view images is a very attractive idea.

The approaches to recover the human figure from single-view images can be separated into two groups depending on the camera model adopted: (1) using affine camera model; and (2) using perspective camera model. Affine camera model is only an approximation of the real camera model. Scaled-orthographic camera model is an important instance of the affine camera model and is popularly used by many researchers (Barrón and Kakadiaris, 2003; Remondino and Roditakis, 2003). However, such camera model can only handle images with very little perspective effects. To handle images with any perspective effects, the perspective camera model is required since it represents a real camera.

There are limited research efforts (Peng and Li, 2005; Zhao *et al*., 2005) working on human model reconstruction based on perspective camera models. Some researchers (Zhao *et al*., 2005) restrict all body segments of the human figure as almost parallel to the image plane in order to acquire accurate human skeleton proportions. Some researchers (Peng and Li, 2005) require estimating the virtual scale parameters for each frame. The perspective camera models used in these researches are either pre-defined (Zhao *et al*., 2005) or manually defined (Peng and Li, 2005).

To automatically recover a perspective camera model, camera calibration is a popular technique which addresses the issue of reconstructing the camera parameters using 2D image points and the corresponding known 3D object points (Bacakoglu and Kamel, 1997; Ji and Zhang, 2001; Memon and Khan, 2001; Batista *et al.*, 1998). However if only the 2D information is available as in monocular images or videos, such camera calibration will not be applicable.

The human body is an articulated 3D object. Skeleton proportions dominate the appearance of a human figure, regardless of the height of the human figure. Skeleton proportions may be different between and among different populations (e.g., men vs women, adults vs children). Estimating the skeleton proportions is the crucial step before reconstructing the 3D shape of a human body, and it is critical for any model-based human motion reconstruction and tracking.

Therefore, it is clearly desirable to develop new algorithms to automatically recover a perspective camera model based on 2D features only and then recovering the skeleton proportion of a human figure using the recovered camera model.

This paper proposes a novel system to estimate the skeleton proportions from single-view image sequence using an automatically recovered perspective camera model. In the proposed system, the camera is assumed to be fixed and almost parallel to the floor during capturing and there should be at least one foot of the human figure touching the floor at any moment during capturing.

The rest of this paper is organized as follows: The next section provides an overview of the proposed algorithm; Sections "KEY FRAMES EXTRACTION", "CAMERA MODEL RECOVERY", and "SKELETON PROPORTIONS ESTIMATION" discuss the major parts of the proposed system; Section "RESULTS" show the results with discussion; and the last section concludes the paper.

OVERVIEW

Fifteen feature points are assumed known from any input source, as shown in Fig.1.

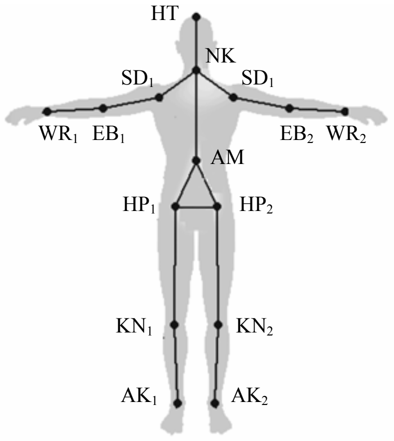


Fig.1 Feature points of human figure

The proposed system is shown in the chart in Fig.2. There are three important modules in this system: (1) key frames extraction; (2) camera model recovery; and (3) skeleton proportions estimation.

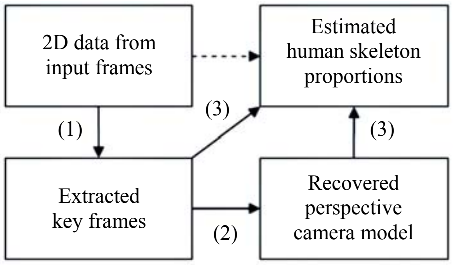


Fig.2 Human proportions estimation system
(1) Key frames extraction; (2) Camera model recovery; (3) Skeleton proportions estimation

The key frames extraction module identifies the frames with both feet touching the floor from all input frames. It is assumed that in the input image sequence at least one foot touches the ground at any moment. Hence there will be two possibilities for each frame: (1) one foot is the stance foot; (2) both feet are stance feet. The proposed algorithm can determine the probable case for each frame from noisy input data. The module is further discussed in a later section.

The camera model recovery module will reconstruct the perspective camera model. Traditional camera calibration method requires correspondence between both the 2D and 3D information and hence could not be adopted in this project. The proposed method will be based only on 2D data, which are the projections of 15 feature points. The module will also be further discussed later.

The skeleton proportion estimation module will reconstruct the human skeleton proportions from the 2D data in the key frames using the recovered perspective camera model. A virtual human figure will be reconstructed to match the projections from all input frames under the virtual environment. The skeleton proportions of this virtual human figure are the desired human skeleton proportions, since scaling does not change the proportions. Again this module will be further discussed in a later section.

KEY FRAMES EXTRACTION

The input 2D data contain a moving human figure, assumingly with at least one foot placed on the floor at any moment. Each input frame displays either both feet touching the floor or only one foot touching the floor. The key frames extraction module first determines which foot touches the ground for each frame, then automatically choose the frames with both feet touching the floor.

Two steps are required for the determination of the stance foot. First it is determined whether a foot in the frame is a possible static foot. Next it is determined whether the static foot is a stance foot.

Possible static foot

Since the camera is fixed, if a foot remains static for a period, its projection should also stay at the same position. Hence, if the input data are perfect, when the projection of a foot remains at the same 2D position within a few neighboring frames, this foot may remain at the same 3D position during this time, it may also move along the straight line passing through the camera center and the foot itself. In this system, the foot from any of these two possibilities is considered as a possible static foot.

Due to errors in feature extraction, such as noises, the extracted feature point of a static foot may not accurately represent its actual projection. It is possible that both feet “appear” to be dynamic in all frames, judging only from their projections. If the noises in the input data are not significant, an error threshold δ can be utilized. When the projections of the same feature point within k frames fall inside a circle with radius δ, this feature point is considered possibly static within these k frames. In this way, the static foot can be determined for the input frames. There are four possibilities for each frame: “N”—none of the feet is static; “1”—foot 1 is possibly static; “2”—foot 2 is possibly static; “B”—both feet are possibly static.

Stance foot

To determine whether a possible static foot is a stance foot, the sequence of foot status is investigated.

If a frame has the status “1” or “2”, the corresponding foot must be a stance foot. But when a frame has the status “N” or “B”, it needs to be further analyzed.

Status “N” indicates that neither foot is static according to the extracted projections. It happens mainly when the input error exceeds the given threshold. Such situations can be handled as follows: for the case “11...1NN...N11…1” and case “22...2NN…N22…2”, if the number of “N” frame is very small, the projection in the frame with “N” can be adjusted according to the projections around this frame, in which way the status “N” can be eliminated. Similarly, cases “11…1NN...N22…2” and “22…2NN...N11…1” can be adjusted and become “11…1BB…B22…2” and “22…2BB…B11…1”. However, when the number of “N” frames is large, this system will not be able to be handled because it indicates the input data are very unstable.

Status “B” indicates that both feet are static. However one of the feet may be static in the air or moving along the straight line passing through the camera center and the foot itself. Such foot cannot be considered as a stance foot. In the case “11...1BB…B11…1” or “22…2BB…B22…2”, one foot remains static during the whole period while the other foot keeps still only for a short period of time. So, the other foot is unlikely a stance foot and such cases can be revised to “11…1” and “22…2”.

Only in cases “11…1BB…B22…2” or “22…2BB…B11…1” can the status “B” be considered as both feet touching the floor. Such frames are the desired key frames.

CAMERA MODEL RECOVERY

The camera model recovery module uses a novel method to obtain a pin-hole camera model from 2D data. To recover a pin-hole camera model, the position, orientation, film size and focal length of the camera need to be recovered. In this project, the camera coordinate system is used, which means that the camera center is the origin of the coordinate system, the orientation of the camera defines the z-axis. The film size is set as the conventional 35 mm (36 mm×24 mm). Therefore, the focal length of the camera is the only factor needed to be recovered.

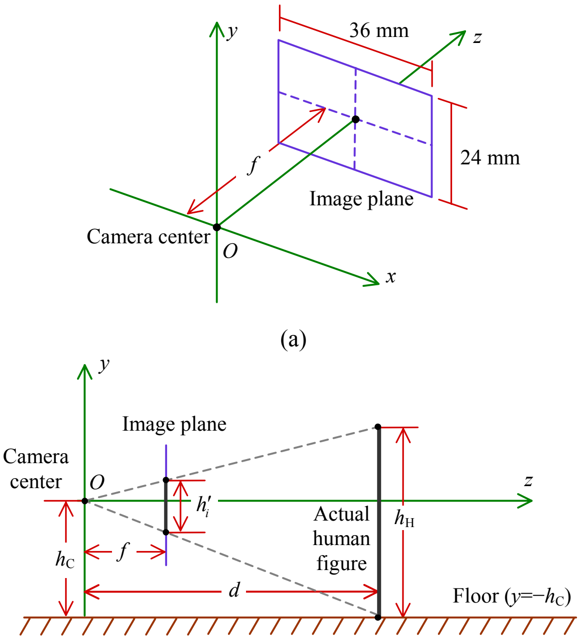
As shown in Fig.3a, the camera with focal length f is located at the center of the world coordinate system, and the image has the size of 36 mm×24 mm. Since the orientation of the camera is assumed parallel to the floor, the floor can be represented as $y=-h_C$, where h_C is the height of the camera from the floor. Assume that the human figure with height h_H is upstanding at the moments M_i and M_j, and the distance between the camera and the human figure in z direction is d_i at the moment M_i, and d_j at the moment M_j. Fig.3b shows the scene at moment M_i, where the projection of the human figure has the height of h_i' in the image plane. The following relationships can be determined for the moment M_i and M_j:

$$\text{Moment } M_i\text{: } h_i'/h_H = f/d_i, \tag{1}$$

$$\text{Moment } M_j\text{: } h_j'/h_H = f/d_j. \tag{2}$$

With Eq.(1) and Eq.(2), the following relationship between the focal length f and the other factors can be established:

$$f = \frac{|d_i - d_j|}{h_H} \cdot \frac{h_i' h_j'}{|h_i' - h_j'|}. \tag{3}$$



(a)

(b)

Fig.3 Camera model used in this project. (a) Camera model; (b) Side view of the scene

As h_i' and h_j' can be obtained directly from the images, if the ratio $|d_i-d_j|/h_H$ is also known, the focal length f of the camera is possible to be calculated directly from 2D data using Eq.(3).

Human walking is a common human motion that exists in many sources. During human walking, one foot lands before the other takes off. Therefore, there is at least one foot that touches the floor at any moment. Each time one foot going forward will make a step, and two successive steps make a stride, as shown in Fig.4. The step length is the distance from one foot strike to the next (left to right or right to left) while the stride length is double the step length. The normal stride length is roughly the same as the human height (stature) (Kirtley, 1998). Therefore the normal step length can be considered as half of the human height. Each time the human figure makes a normal foot step, the figure moves about half of its height.

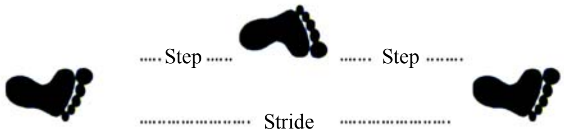


Fig.4 Foot step and foot stride (Kirtley, 1998)

If the moments M_i and M_j represent two successive steps, the distance between the human figures in the two moments is about $h_H/2$, which leads to $|d_i-d_j|\le h_H/2$. Thus, Eq.(3) can be simplified to:

$$f \le \frac{h_i' h_j'}{2|h_i' - h_j'|}. \tag{4}$$

In this project, the stance foot for each frame is determined after applying the algorithms in Section "KEY FRAMES EXTRACTION". Key frames with both feet on the floor should have been extracted. Hence, it is not difficult to automatically obtain a pair of key frames representing two successive steps. Actually, a number of pairs can be obtained if the video contains a long human walking sequence. Let the number of such pairs of key frames be n, according to Eq.(4), each pair of such key frames will yield:

$$f \le f_k \ (k=1, 2, \ldots, n). \tag{5}$$

Thus, the focal length f of the camera model can be approximated as:

$$f \approx \min(f_1, f_1, \ldots, f_n). \tag{6}$$

SKELETON PROPORTIONS ESTIMATION

The basic idea of the skeleton proportion estimation is to find a virtual human model which can reproduce the projection data in the virtual environment. The skeleton proportion can be calculated from this virtual human.

It is well-known that any projection point on the viewing plane can be back-projected to an infinite number of possible 3D positions. Fortunately, the stance foot is determined in each frame. Thus, given the virtual ground and the perspective camera model, the 3D position of the stance foot can be calculated. The virtual human figure on the virtual ground will be modeled from the stance foot. The estimation process is shown in Fig.5. During the estimation process, there are three basic actions "PICK", "TEST" and "CALC".

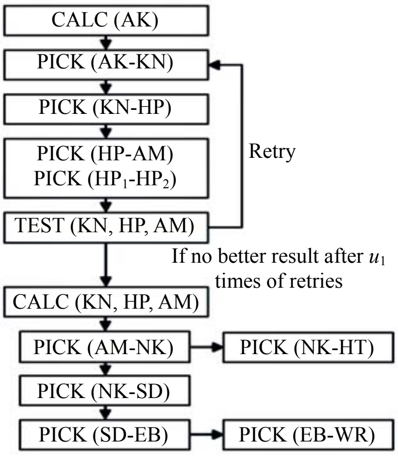


Fig.5 Skeleton proportions estimation system

"PICK"

The purpose of this action is to assign a length for a specified skeleton segment. This length should be within the range of all possible skeleton lengths. Hence the lower boundary and the upper boundary of the range should be determined first.

The lower boundary of the range is defined as follows: if the skeleton segment has length below this lower boundary, it is not possible to have its projection matching all key frames under the estimated camera model from the given 3D position. Instead, if a skeleton is longer than the lower boundary of the range, it will forever be able to produce the given projection.

The upper boundary of the range is introduced for the purpose of minimizing the searching range. It is defined to allow a skeleton segment to give the projection from any possible 3D position.

For the ith input key frame, the lower boundary and upper boundary of the possible skeleton length can be calculated as $L_{\min}(i)$ and $L_{\max}(i)$, indicating that the skeleton length l should satisfy $L_{\min}(i)\le l\le L_{\max}(i)$ in the ith frame. Therefore, if all key frames are considered, the global upper and lower boundaries can be acquired as:

$$L_{\min}=\max(L_{\min}(i), i=1, 2, \ldots, n), \tag{7}$$

$$L_{\max}=\max(L_{\max}(i), i=1, 2, \ldots, n). \tag{8}$$

Any length within this range, $L_{\min}\le l\le L_{\max}$, is possible for the skeleton segment.

"TEST"

The purpose of this action is to test whether the estimated skeleton could produce reasonable postures in frames.

Starting from joint AK_i (i=1, 2 as shown in Fig.1), there are up to 4 possible 3D positions for the corresponding joint HP_i. Subsequently, there are up to 16 possible combinations of positions for the joint HP_1 and HP_2. Each of such combinations forms a unique triangle HP_1-AM-HP_2 from the projection. The sides of these triangles are compared to the estimated length of HP_i-AM and HP_1-HP_2. If none of these triangles match the estimated triangle in side lengths, more iteration should be conducted, until the accumulated difference is under a given threshold or the maximum number of iterations is reached.

"CALC"

The purpose of this action is to calculate a unique 3D position of the feature points in the virtual scene for each key frame. For example, CALC(AK_i) will calculate the positions of the joint AK_i with the help of the virtual camera and virtual ground. CALC (KN_i, HP_i, AM) will calculate the positions of joint KN_i, HP_i, and AM that can provide the best result in the action TEST (KN_i, HP_i, AM).

RESULTS

The proposed system is tested on the 2D data extracted from two real video sequences. Both video sequences record the walking motion of a human figure. The video camera is fixed and mounted on a tripod, with its orientation parallel to the floor.

The first sequence—Video 1 contains 113 frames. It is captured by the video camera with focal length 29.05 mm (in 35 mm format). The second sequence—Video 2 contains 141 frames and is captured by the video camera with the focal length 400 mm (in 35 mm format).

Figs.6a and 6b show an example frame of Video 1 and Video 2, while Figs.6c and 6d illustrate the 2D data extracted from these two example frames. The proposed system will work on the extracted 2D data.

The 2D data from each input video sequence are processed following the flow chart shown in Fig.2. Three major steps are involved: (1) extract the key frames from the 2D data; (2) recover the camera model from the key frames; (3) estimate the human skeleton proportions from the key frames using the recovered camera model. The algorithm used in each step is tested and the results are presented and discussed in the following subsections.

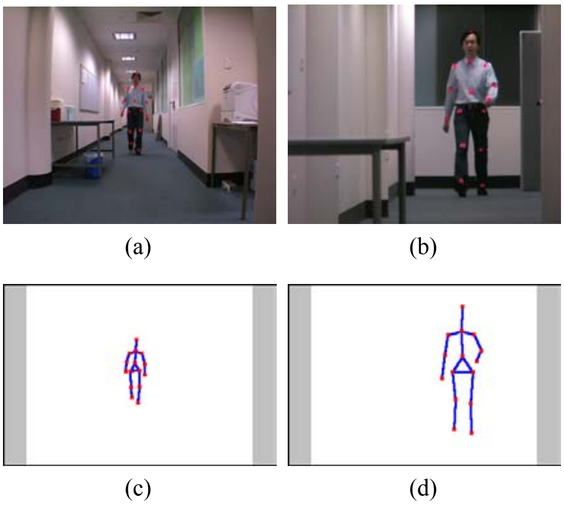


Fig.6 Input data. (a) Video 1; (b) Video 2; (c) 2D data 1; (d) 2D data 2

Test of key frames extraction

In the proposed system, key frames are defined as the frames that record the moment when both human feet are placed on the floor. The first step of the proposed system is to find all such frames.

Key frames are automatically extracted from the 2D data in the key frames extraction module. The extraction algorithm first determines the foot that remains motionless in each frame, and then extracts the stance foot from the motionless foot. Finally, the frames with two stance feet are chosen as the key frames.

The experiment results of the key frames ex-

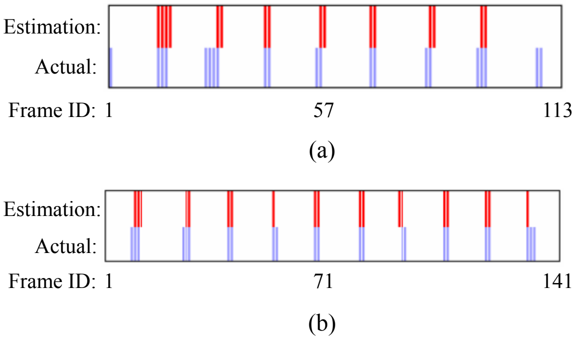


Fig.7 Key frames extraction results. (a) Key frames in input 2D data 1; (b) Key frames in input 2D data 2

traction algorithm are shown in Fig.7. The horizontal axis of the figure indicates the frame ID. If a frame is extracted as a key frame, a bar will be drawn: the red bars on the top are the extraction results by the proposed algorithm, while the blue bars at the bottom indicate the manually extracted key frames through the judgment of human eyes for comparison purpose.

It can be seen that the key frames extracted by the proposed algorithm and the manually extracted key frames are very close. The proposed algorithm may extract some frames by mistake. However, such frames are very close to the key frames. Such frames can be accepted as key frames since both feet are actually very close to the floor in them.

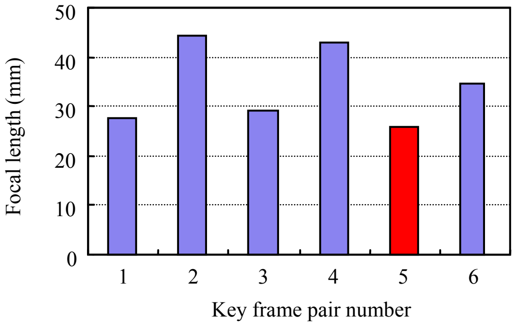
Test of camera model recovery

Camera model recovery is the second important module of the proposed system. The camera model used in this project is a fixed perspective camera with its orientation parallel to the floor. As discussed before, the only camera parameter to be recovered is the focal length.

The projected human height in each key frame is first calculated from the 2D data. Next the proposed algorithm is employed to calculate the focal length range for each pair of neighboring key frames using Eq.(4). Lastly, the focal length is approximated using Eq.(6) by choosing the upper bound of the smallest focal length range.

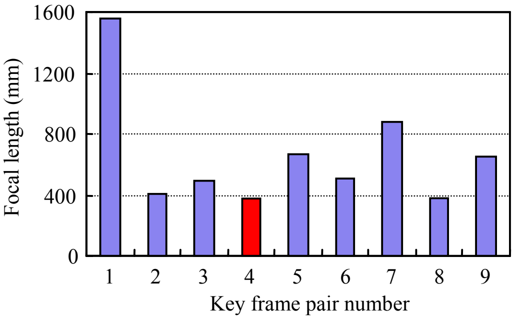
Fig.8 shows the experiment results. In Fig.8, the *i*th column bar represents the range of the focal length calculated for the *i*th pair of key frames. Each pair of such key frames indicates a foot step. For example in Fig.7, the 4th and 5th key frame consist of the first pair of such key frames in Video 1. When the human figure is not walking towards the camera, such pair may produce a very large range of focal length according to Eq.(4). However, the final result will not be affected as long as the video contains at least one frame in which the human figure is walking towards the camera.

It can be seen that the recovered focal length is very close to the actual focal length when using 35 mm film format as a standard. For Video 1, the recovered focal length is 25.914 mm while the actual focal length is 29.05 mm. For Video 2, the recovered focal length is 373.898 mm while the actual focal length is 400 mm.



Recovered focal length: 25.914 mm
Actual focal length: 29.05 mm

(a)



Recovered focal length: 373.898 mm
Actual focal length: 400 mm

(b)

Fig.8 Camera model recovery results. (a) Results in Video 1; (b) Results in Video 2

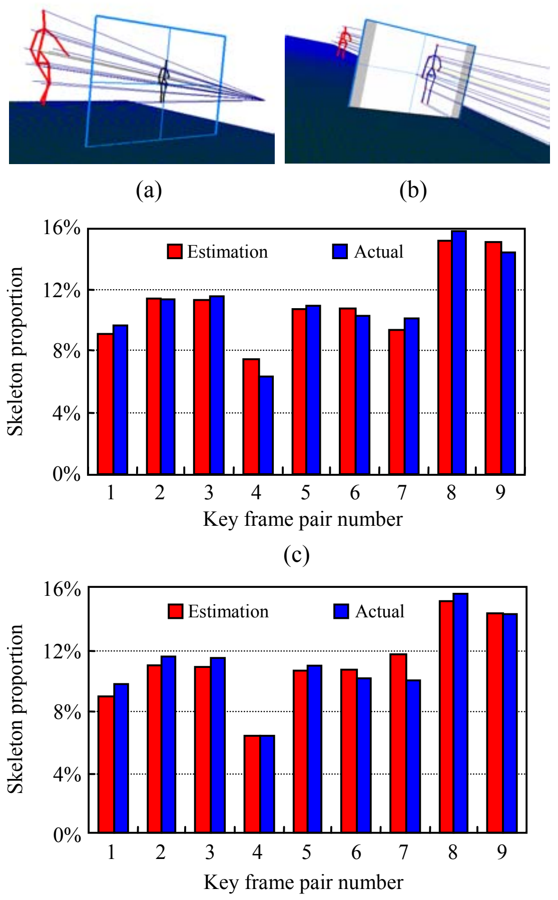
Therefore, the proposed camera recovery algorithm is tested to be very satisfactory, considering that it does not require any information on the image background or known 3D points.

Test of skeleton proportions estimation

Based on the recovered perspective camera model, the proposed algorithm will reconstruct the virtual human figure for each of the extracted key frames.

The virtual scene is first reconstructed, where the virtual floor is defined as the plane passing through the bottom of the 35 mm film and parallel to the orientation of the camera model. Next the virtual human figure is constructed from both feet towards the other three ends of the human body: head and both hands using the method discussed above. Once the virtual human figure is reconstructed for all key frames, the skeleton proportion can be calculated by dividing the length of each skeleton segment by the sum of them.

The experiment results are shown in Fig.9: (a) and (b) show the virtual scene examples used in the proposed system; (c) and (d) compare the skeleton proportions of the estimated virtual human figure against the actual human figure, where the horizontal axis represents the skeleton ID and the vertical axis represents the skeleton proportion. The definition of the skeleton segment ID is listed in Table 1.


(a) (b)

(c)

(d)

Fig.9 Skeleton proportion estimation results. (a) Virtual scene example for Video 1; (b) Virtual scene example for Video 2; (c) Skeleton proportion estimation results in Video 1; (d) Skeleton proportion estimation results in Video 2

Table 1 Skeleton segment ID

Skeleton ID	Description	Skeleton ID	Description
1	HP-AM	6	EB-WR
2	AM-NK	7	HP_1-HP_2
3	NK-HT	8	HP-KN
4	NK-SD	9	KN-AK
5	SD-EB		

It can be seen from Figs.9c and 9d that the maximum estimation error is 1.13% for Video 1 and 1.60% for Video 2. The error can be further minimized if the number of iterations u_1 is increased.

The detailed numerical evaluation of the three modules in the proposed system showed that the proposed algorithms have successfully estimated the skeleton proportions of the human figure directly from 2D data. The experimental results are highly satisfactory.

CONCLUSION

This paper proposed a novel system to recover the human skeleton proportions from 2D uncalibrated data. The proposed system extracts the key frames, recovers the perspective camera model, and then estimates the skeleton proportions of the human figure in the source video. The proposed method is tested on the 2D data from two real video sequences. The experiments achieved satisfactory results. The recovered human skeleton model is very close to the original and can be used in future research such as full body reconstruction and human motion reconstruction from monocular data.

References

Bacakoglu, H., Kamel, M., 1997. An Optimized Two-step Camera Calibration Method. Proceedings of IEEE International Conference on Robotics and Automation, **2**:1347-1352.

Barrón, C., Kakadiaris, I.A., 2001. Estimating anthropometry and pose from a single uncalibrated image. *Computer Vision and Image Understanding*, **81**(3):269-284. [doi:10.1006/cviu.2000.0888]

Barrón, C., Kakadiaris, I.A., 2003. On the improvement of anthropometry and pose estimation from a single uncalibrated image. *Mach. Vision Appl.*, **14**(4):229-236. [doi:10.1007/s00138-002-0088-8]

Batista, J., Araujo, H., De Almeida, A.T., 1998. Iterative Multi-Step Explicit Camera Calibration. IEEE International Conference on Computer Vision (ICCV'98). Bombay, India, p.709-714.

Cheung, K.M., Baker, S., Kanade, T., 2003. Shape-From-Silhouette of Articulated Objects and its Use for Human Body Kinematics Estimation and Motion Capture. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Cohen, I., Lee, M.W., 2002. 3D Body Reconstruction for Immersive Interaction. 2nd International Workshop on Articulated Motion and Deformable Objects.

D'Apuzzo, N., Gruen, A., Plankers, R., Fua, P., 2000. Least Squares Matching Tracking Algorithm for Human Body Modelling, XIX ISPRS Congress, Amsterdam, Netherlands.

Hilton, A., Gentils, T., 1998. Popup People: Capturing Human Models to Populate Virtual Worlds (Sketch). SIGGRAPH'98.

Ji, Q., Zhang, Y., 2001. Camera calibration with Genetic Algorithms. *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans*, **31**(2):120-130. [doi:10.1109/3468.911369]

Kakadiaris, I.A., Metaxas, D., 1995. 3D Human Body Model Acquisition from Multiple Views. ICCV'95: Proceedings of the Fifth International Conference on Computer Vision, p.618. [doi:10.1109/ICCV.1995.466881]

Kirtley, C., 1998. Walking, Normal and Pathological. Http://guardian.curtin.edu.au/cga/faq/obs.doc.

Lee, W.S., Gu, J., Magnenat-Thalmann, N., 2000. Generating Animatable 3D Virtual Humans from Photographs. *Computer Graphics Forum (Eurographics 2000)*, **19**(3).

Memon, Q., Khan, S., 2001. Camera calibration and three-dimensional world reconstruction of stereo-vision using neural networks. *International Journal of Systems Science*, **32**(9):1155-1159. [doi:10.1080/00207720010024276]

Peng, E., Li, L., 2005. Estimation of Human Skeleton Proportion from 2D Uncalibrated Monocular Data. Proceedings of Computer Animation and Social Agents (CASA 2005). Hong Kong.

Remondino, F., Roditakis, A., 2003. Human Figure Reconstruction and Modelling from Single Image or Monocular Video Sequence. 3-D Digital Imaging and Modelling 2003 (3DIM 2003), Proceedings, Fourth International Conference on, p.116-123. [doi:10.1109/IM.2003.1240240]

Starck, J., Hilton, A., 2003. Model-Based Multiple View Reconstruction of People. Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03), p.915. [doi:10.1109/ICCV.2003.1238446]

Taylor, C.J., 2000. Reconstruction of articulated objects from point correspondences in a single uncalibrated image. *Computer Vision and Image Understanding*, **80**(3): 349-363. [doi:10.1006/cviu.2000.0878]

Villa-Uriol, M., Sainz, M., Kuester, F., Bagherzadeh, N., 2003. Automatic creation of three-dimensional avatars. *Videometrics VII, Proceedings of the SPIE*, **5013**:14-25. [doi:10.1117/12.473109]

Zhao, J., Li, L., Kwoh, C.K., 2005. Posture Reconstruction and Human Animation from 2D Feature Points. *Computer Graphics Forum*, **24**(4):759-771. [doi:10.1111/j.1467-8659.2005.00900.x]