

The Research and Application of an Algorithm about Human Joint Points Tracking based on a Video

Hanxing Xie

School of Reliability and System Engineering, Beijing
University of Aeronautics and Astronautics
RSE
Beijing, China
xhx@buaa.edu.cn

Sha Qin

School of Reliability and System Engineering, Beijing
University of Aeronautics and Astronautics
RSE
Beijing, China
sjjiitt@163.com

Xiaoxiao Li

School of Reliability and System Engineering, Beijing
University of Aeronautics and Astronautics
RSE
Beijing, China
guohailtx@163.com

Weilin Li

School of Reliability and System Engineering, Beijing
University of Aeronautics and Astronautics
RSE
Beijing, China
lwl890228@126.com

Abstract—This paper presents an algorithm about human joint points tracking based on a video. The method requires locating manually the joints' position of the first frame, and tracking automatically all the joints starting from the second frame. Firstly, segment human lower limb marking the feature points based on two frame difference and threshold method. Then tracking feature points by using block matching method. Finally, the information of each joint' location and the corresponding angle can be obtained. In order to ensure the accuracy of tracking, the scope of tracing must be limited by using human lower limb segmented by image and length model of each joint. This method has high accuracy and speed. Comparing the manual position and the computer programming tracking, its accuracy can be as high as 96%, and the actual error is less than 0.4cm.

Keywords- *feature tracking; joint position location; motion capture; Image segmentation; Image analysis*

I. INTRODUCTION

In recent years, the research field of computer image processing technology, such as kinematic analysis, intelligent control, virtual reality technology and especially human motion analysis, has been more and more widely. The main aspect of human motion analysis is human posture analysis, which is a hot research field of rehabilitation therapy, lower extremity exoskeleton and physical training. And the joint position location is the key point of human motion analysis. Generally, there are two kinds of the traditional method of joint position location: the first one, adding sensors to each joint of human respectively. In the process of human motion, the joints position in the space will be fed back to computer continually by sensors, thus the computer can accurately obtain the human motion posture at each moment [1, 2]. This method requires expensive motion capture devices, and its applications subject to considerable limitations. The

second method is to analyze the image sequence containing moving body, which can be located joints movement information manually, and the automatic tracking of joint points can also be achieved by programming [3, 4]. However, the manual position location method is not realistic to achieve since there are hundreds frames of a 10 seconds video, and takes a lot of work. Currently, the broader application is to perform the joints tracking automatically, and dynamically show the curves information of joints motion intuitively. Nafi Ur Rashid et al. [5] had proposed a method that could classify and identify vehicles in video, and James Black et al. [6] had used multiple camera systems to monitor and track the object. Regardless of which method, the problems of image segmentation, occlusion handling and arithmetic speed must be faced. As the method of lower limb joint points automatic positioning proposed by Lu-Ming Yang et al., the lower limb joint points can be located from the human motion image sequence without joint markers. Although without any marker, this method would increase the complexity of positioning procedure during tracking if the predicted value of joint angle deviates from true value. For the joint point automatic tracking method, the problems are mainly accuracy and speed of tracking.

The tracking algorithm, presented by this paper, requires the background of the used video no change and the feature points at human joints are necessary. By comparing several sets of videos, it has been found that the round-white feature points were not prone to distortion during human motor processes. Therefore, the feature points used here are white and round. They can be directly pasted at the joints while shooting videos. This method is simple and easy to operate, as well as no area restriction. This mentioned algorithm is characterized by high accuracy and speed.

II. THE GENERAL IDEA OF JOINT POINTS TRACKING ALGORITHM

The general idea of this paper is shown in Fig.1. To begin with, the input image sequence is initialized by human body calibration, and the aim of calibration is to obtain the specific length model of each joint, define the tracking range and make the results more accurate. The details of calibration process are described in Section 3.1. After the calibration, manually locating the joint point of the first frame image is needed. Then the computer will tracking the joint points automatically from the second frame. This procedure involves image processing, image segmentation and block matching process. The specific tracking procedure sees Section 3.2 and 3.3. This tracking algorithm is simple and fast, also high accuracy. Finally, the continuous information of each joint position are acquired, and the corresponding angular information are calculated and outputted.

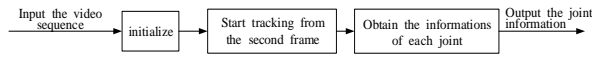


Figure 1. The general idea of tracking algorithm

III. THE TRACKING OF JOINT POINTS

A. The joint point extraction and the human calibration

The first frame information of joint points is obtained by manually, and the human calibration is performed. In order to ensure the accuracy of tracking, this algorithm need to make human calibration in the first place due to the difference of each person's height and each limb length. As shown in Fig.2, the Fig.2 (a) is a certain frame image of video, and (b) is the coordinate axis corresponding to the image of (a), where the four points of $S1$, $S2$, $S3$, $S4$ are the coordinate axis information of hip, knee, ankle and toe, respectively. And the lengths of $S1$ to $S2$, $S2$ to $S3$ and $S3$ to $S4$ represent the relative lengths of thigh, shank and sole, respectively. Since all joints of human lower limbs are rigid bodies, the length of each human joint can be considered as constant.

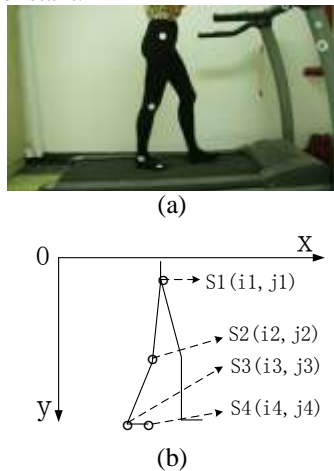


Figure 2. Human calibration

The approach of human body calibration is to take approximately 10 frames image as a training sequence, and manually acquire the information of joint points, then successively obtain the relative length of each joint and accordingly the length models of joints will be obtained.

The relative length models of thigh, shank and sole, which are calibrated in this test, are respectively $(h11, h12)$, $(s11, s12)$ and $(f11, f12)$, and the unit is pixel.

B. Image segmentation

The tracking of joint point information begins from the second frame. First of all, processing the image segmentation. The human lower limb containing feature points and background will be segmented to improve the accuracy of tracking, and also to avoid tracking in the ineffective area outside human body to increase the speed of tracking. In this paper, the background is segmented with the combine of frame difference and threshold method [8, 9]. At present, the commonly used method of background segmentation is difference method, i.e. taking a difference between the previous frame and the current frame. Then set a threshold. It is assumed to be the foreground (i.e. the moving object) when the difference is greater than the threshold, otherwise the background. Difference method typically include background difference and frame difference. The former is able to extract the moving object more completely, but it is closely related to the quality of background image. The latter has good adaptability to environment, whereas it cannot extract the complete moving object. And the threshold method is to analyze the pixel value of human body in the histogram and to segment the image by setting a threshold. For without losing the human lower limbs, which containing the feature point, firstly a rectangular search box must be obtained by frame difference. Then by combining with threshold method, the search scope will be further restricted on the human body. After several sets of experiments, it has been proved that when the frame difference method uses a threshold of 0.2 and the threshold method uses a threshold of 0.14, the segmentation between human body and background is the most obvious. For the purpose of more convenience during tracking, the feature of the feature points require to be eliminated. And in this paper, the elimination method of the white feature points on human body is to seek the connected domain.

The flowchart of image segmentation is shown in Fig.3.

The specific algorithm is as follows:

First, it is assumed that $f(t)$ and $f(t+1)$ represent the t th and the $t+1$ th piece of image, respectively. The size of feature block, required to tracking, is $m*n$.

1) Transforming $f(t)$ and $f(t+1)$ into grayscale image, then $f'(t)$ and $f'(t+1)$ will be obtained after median filtering;

2) There is $BW(t) = |f'(t) - f'(t+1)|$, then setting the threshold as 0.2 and the $BW'(t)$ will be derived after transforming the $BW(t)$ into binary image;

3) Searching in the order of row then column in the $BW'(t)$ to seek the rows where the first and the last pixel value larger than 0.5 are located, and denoted by $cormin$ and $cormax$, respectively.

4) Searching in the order of column then row in the $BW'(t)$ to seek the columns where the first and the last pixel value larger than 0.5 are located, and denoted by $rowmin$ and $rowmax$, respectively.

5) Transforming $f'(t+1)$ into binary image by setting the threshold as 0.14, in this way the $f''(t+1)$ will be derived.

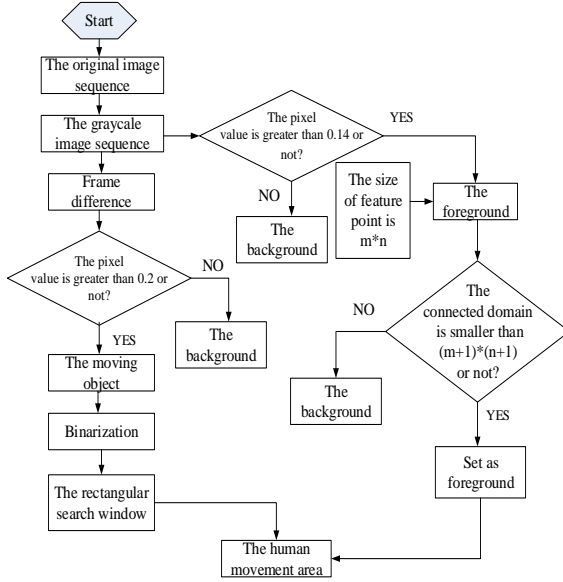


Figure 3. The flowchart of image segmentation

6) Searching the connected domains of $f''(t+1)$. When the number of pixels is less than $(m+1)*(n+1)$, the connected domain will be set to 0, means the feature points are foreground.

7) Then the searching window, i.e. the row $cormin$ to $cormax$ and the column $rowmin$ to $rowmax$ which values are 0, is the movement human body containing feature points, denoted by $area(t+1)$.

The searching window $area(t+1)$ is shown in Fig.4, i.e. the rectangular box of human body parts.



Figure 4. The result of image segmentation.

C. The tracking method

The tracking method used in this paper is block matching [10, 11], and the matching criterion [12] used here is based on the relevant metrics of cross-correlation function (CCF). As shown in Fig.5, it is assumed that

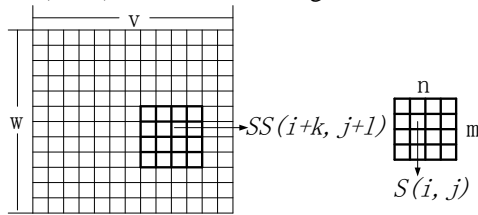


Figure 5. The block matching method

$S(i, j)$ indicates the feature blocks in $f(t)$ while $SS(i+k, j+l)$ indicates the feature blocks required to track in $f(t+1)$, with k and l are the relative displacements of the

matched block required to track to the original feature block.

Then the CCF is written as

$$CCF(k, l) = \frac{\sum_{i=1}^m \sum_{j=1}^n S(i, j) SS(i+k, j+l)}{\left(\sum_{i=1}^m \sum_{j=1}^n S^2(i, j) \right)^{1/2} \left(\sum_{i=1}^m \sum_{j=1}^n SS^2(i+k, j+l) \right)^{1/2}}$$

The purpose of block matching method is to find the appropriate k and l , which allow the relevant metrics of CCF to achieve the maximum.

Since the several feature points have the similar features (all of them are round and white), it will appear the confusing status of tracking while tracking the feature points by using the common block matching method. To avoid this drawback in this paper, the method of prior calibration is brought to. First, extracting approximately 10 frames image and manually acquiring the position coordinates of each feature point in each frame image, then calculating the lengths of thigh, shank and sole, respectively. Thus there will be 10 sets of lengths, and the length models will be obtained after making a statistics of them. During tracking, these length models will appoint each feature point. This method requires to obtain a joint point firstly. Because the amplitude of swing of hip joint is the minimum, first of all using block matching method to obtain the location information of hip joint. Then the location information of knee joint, ankle joint and toe will be obtained respectively in proper order according to the length model of thigh, shank and sole.

The specific flowchart of overall algorithm is shown in Fig.6.

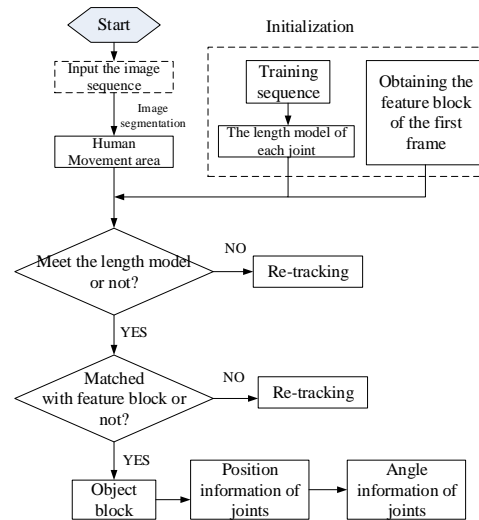


Figure 6. The flowchart of overall algorithm

The tracking algorithm is as follows:

Step 1: initialization.

Calibrating the length models of human joints first, and taking about 10 frames image as training sequence to acquire the position of feature points of each frame, next calculating the lengths of thigh, shank and sole in each frame. The length models of thigh, shank and sole will be respectively $(h11, h12)$, $(s11, s12)$ and $(f11, f12)$ after making a statistics of the 10 sets of length information.

And then acquiring the position information of feature points in the first frame manually so that the initial feature blocks of hip, knee, ankle and toe are $S1(i1,j1)$, $S2(i2,j2)$, $S3(i3,j3)$ and $S4(i4,j4)$, respectively.

Step 2: tracking

The searching window $area(t+1)$ of $f(t+1)$ can be obtained with the above-mentioned image segmentation method. Firstly, using block matching method in $area(t+1)$ to acquire the $SSI(i1+k1,j1+l1)$, which is matched with $S1(i1,j1)$ in $f(t+1)$, i.e. seeking out the $k1$ and $l1$ to allow the $CCF(k1,l1)$ to achieve the maximum. Then tracking the knee joints in $area(t+1)$, and calculating the distance between point $(i1+k1,j1+l1)$ and $(i2+k2,j2+l2)$ before seeking the $k2$ and $l2$. If $hl1 < hl < hl2$, the most appropriate $k2$ and $l2$ making the $CCF(k2,l2)$ maximal will be gotten by using block matching method. In a similar way, if the length between knee and ankle, i.e. the one in the tracking area of $area(t+1)$ is in the scope of $(sl1,sl2)$, the most appropriate $k3$ and $l3$ making the $CCF(k3,l3)$ maximal will be gotten by using block matching method as well. In the same way, the $k4,l4$ and $CCF(k4,l4)$ will all be sought out by block matching method if the length between ankle and toe is in the scope of $(fl1,fl2)$. Now it has obtained the new coordinate point of each feature point and each $SS(i+k,j+l)$ will instead of each initial feature block $S(i,j)$ correspondingly. Then making $t = t+1$ and not stopping circulating the step 2 until the last frame of the image sequence. It cannot be exceeded the image size during tracking, i.e. $-w < k < w$ and $-v < l < v$, $0 < m < w$ and $0 < n < v$ respectively based on the assumption that the size of $area(t+1)$ is $w*v$ and the feature block is $m*n$.

Step 3: output the information.

According to the information obtained from step 2 of joint points of each frame image, the angles of knee and ankle can be calculated by Cosine Theorem.

IV. THE APPLICATION OF EXPERIMENT

This algorithm is implemented by programming with MATLAB R2009b. The video used in experiment is shot by the camera of Pentax K-r and the experimental subject is a member of our research group. The results of experiment are shown in Fig.7 and Fig.8, where Fig.7 is the stick-figure of right limb movement and Fig.8 is the changing curves of joint angles of knee and ankle.

In order to validate the accuracy of tracking, manually extracting the positional information of any 20 frames image which are continuous, and then making a comparison between the data of this and the tracking of experiment. The following Fig.9, Fig.10, Fig.11 and Fig.12 are the results. The abscissa axis and ordinate axis in the figures represent the abscissa value and ordinate value of each joint point, respectively. For easily distinguishing, the abscissa values, obtained by manually and program tracking in each frame, have both made the same offset while the ordinate values are constant. Therefore, the difference values between the manually locating and program tracking after offset is

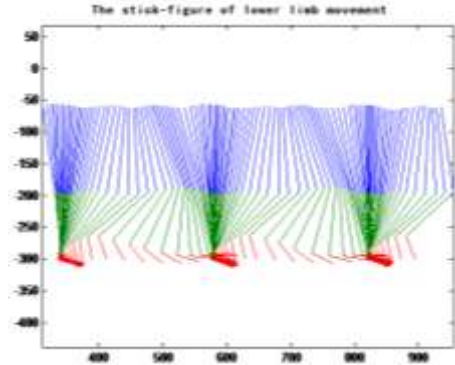


Figure 7. The stick-figure of right limb movement

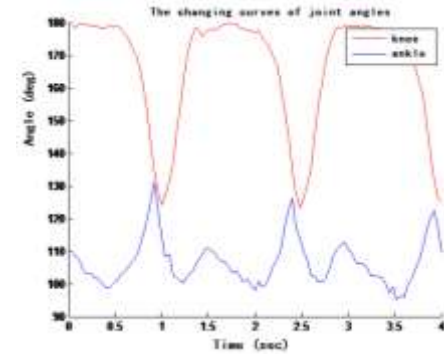


Figure 8. The changing curves of joint angles.

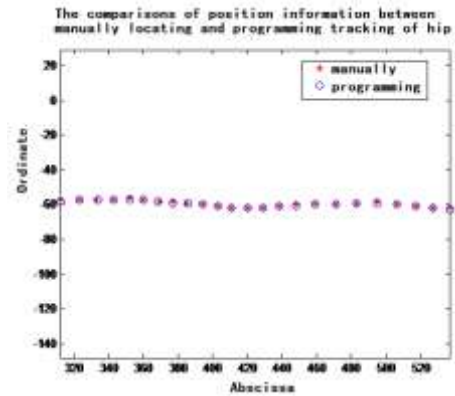


Figure 9. The comparisons of position information between manually locating and programming tracking of hip.

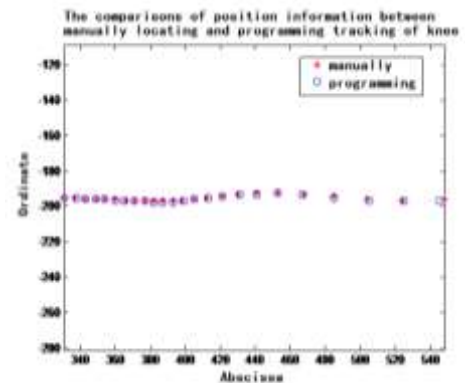


Figure 10. The comparisons of position information between manually locating and programming tracking of knee.

equal to the one before offset. In this video, the diameter of round feature point at the hip joint is 5cm, and the

corresponding size of pixel block is 14×14 , where a pixel value is approximately 0.36cm and each pixel block area is about 0.13cm^2 . Consequently, the errors of ± 1 produced by the abscissa and ordinate values with the programming tracking are accepted. The following figures are successively the comparisons of position information between manually locating and programming tracking of hip, knee, ankle and toe.

It can be seen from Fig.9, at the hip joint, the abscissa and ordinate errors of the difference value between manually obtaining and programming tracking are all in ± 1 . The results obtained by the two tracking are nearly the same, means that the accuracy is approximately 100%.

As shown in Fig.10, at the knee, only the difference value of coordinate of the last pair of points, which are obtained by programming tracking and manually locating, has exceeded the range of ± 1 , and the other points are all in the allowable range of error. That is to say, the accuracy is approximately 96%.

Similarly in Fig.11, at the ankle, the errors are accepted although the difference values of the 19th to 24th are larger than others, then the accuracy is larger than 99%.

In Fig.12, at the toe, only the last pair of points' coordinate difference value has exceeded the range of ± 1 , i.e. the accuracy is approximately 96%.

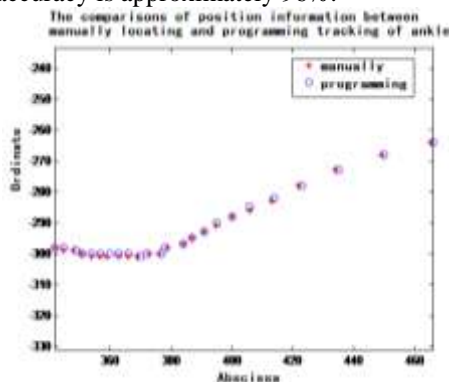


Figure 11. The comparisons of position information between manually locating and programming tracking of ankle.

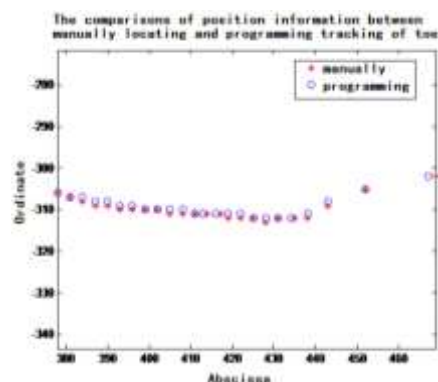


Figure 12. The comparisons of position information between manually locating and programming tracking of toe.

V. CONCLUSIONS

This paper has researched the motion capture method of human lower limbs. And the automatic tracking algorithm has been implemented by programming with MATLAB.

This method has the advantages of simple, fast and exact and its accuracy rate can reach more than 95%. What's more, the difference value between automatic tracking and manually locating is not exceeding 0.4cm. This can fully meet the actual requirements. The method in this paper is mainly applied to the exoskeleton project in order to obtain the position and angle information of human, then the relevant kinematics analysis of biomechanics can be performed. And it lays the foundation for further research on the exoskeleton. The major defect of this method is just unable to adapt well to the blurring generated by high-speed movement. However, this problem will be solved with a higher collection speed of camera. Therefore, the higher requirement of camera is required correspondingly. This method is only able to obtain two-dimensional information and the human body must be labeled with feature points. It is not adapted well to the occlusion problem. With further in-depth research, it will make improvement at later stage.

REFERENCES

- [1] John Y. Goulermas, Andrew H. Findlow, Christopher J. Nester, et al.: An Instance-Based Algorithm with Auxiliary Similarity Information for the Estimation of Gait Kinematics from Wearable Sensors. IEEE TRANSACTIONS ON NEURAL NETWORKS, SEPTEMBER 2008:1574-1582.
- [2] Che-Chang Yang, Yeh-Liang Hsu, Kao-Shang Shih, etc.: Real-Time Gait Cycle Parameters Recognition Using a Wearable Motion Detector. Proceedings of 2011 International Conference on System Science and Engineering, Macau, China, June 2011:498-502.
- [3] Chung-Lin Huang, Chung-Chin Lin: Model-Based Human Body Motion Analysis for MPEG IV Video Encoder [A]. Proc. Of Inter. Conference on Information Technology: Coding and Computing [C]. 2001:435-439.
- [4] Wei-Yi Wang, Ming Zhang, Rui-Hua Jin, Ai-Lan Chen: The Human Joint Tracking Based on Automatic Annotation of Feature Points. Computer Measurement & Control. 2007. 15(10):1379-1381.
- [5] Nafi Ur Rashid, Niluthpol Chowdhury Mithun, Bhadhan Roy Joy, S. M. Mahbubur Rahman: Detection and Classification of Vehicles from a Video Using Time-Spatial Image. 6th International Conference on Electrical and Computer Engineering, December 2010:502-505.
- [6] James Black, Dr. Tim Ellis: Multi Camera Image Tracking. Proceedings 2nd IEEE Int. Workshop on PETS, Kauai, Hawaii, USA, December 9, 2001.
- [7] Lu-Ming Yang, Yan Gui, Qing-Dong Zeng: The Human Lower Limb Joints Location Based on the Model of Appearance. Computer Engineering and Applications, 2008.44(10):67-70.
- [8] Electronic Publication: Digital Object Identifiers (DOIs): Article in a journal: [8] Rui Zhang, Sizhu Zhang, Songyu Yu: Moving Objects Detection Method Based on Brightness Distortion and Chromaticity Distortion. IEEE Transactions on Consumer Electronics, AUGUST 2007:1177-1185.
- [9] Li-Xia Xue, Yan-Li Luo, Zuo-Cheng Wang: The Adaptive Detection Method of Moving Object Based on Frame Difference. Application Research of Computer, 2011.28(4):1551-1559.
- [10] Lei Zhang: The Motion Estimation Algorithm Research of Fast Block Matching in Video Coding. Chengdu: The University of Electronic Science and Technology, 2008.
- [11] Peng-Wan Jiang: The Adaptive Block Matching Algorithm Based on Image Characteristics. Wuhan: The Huazhong University of Science and Technology, 2008.
- [12] Li-Jun Luo, Cai-Rong Zhou, Xi-Qi Gao, Zhen-Ya He: The Motion Estimation Algorithm of Block in Video Coding. 1997(12):2-8.