ISSN 2616-5775 Vol. 6, Issue 6: 55-61, DOI: 10.25236/AJCIS.2023.060608

Analysis of Human Joint Mechanics Based on High-Precision Unmarked Motion Capture System

Ma Luheng

School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, 230027, China

Abstract: *The analysis of human joint mechanics is of great value in rehabilitation assessment and competitive sports training. There are many limitations in the application of the existing motion capture system, which makes the usage scenarios relatively limited. To solve this problem, this paper proposes a multi-camera 3D motion capture system based on reprojection optimization, and combines the foot mechanics data to obtain the human joint torque. First, build a multi-camera-based motion capture system to obtain human body motion data, reconstruct key points to obtain 3D data, and then propose an accurate key point algorithm based on weight-based reprojection optimization to obtain precise coordinates of key points. Finally, the coordinates of the key points combined with the mechanical data of the foot are calculated by OpenSim to obtain the mechanical data of the human joints. The experimental results show that the capture accuracy of key points is 5% higher than the average accuracy of the traditional least square method, and the analysis of joint mechanics data conforms to the laws of motion, which can be used to guide rehabilitation evaluation and competitive sports.*

Keywords: *machine vision; reprojection optimization; human body mechanics analysis*

1. Introduction

Introducing the analysis of joint torque into competitive sports is of great significance for the precise control of athletes during training. As the technical conditions for sports data collection and processing mature, the role of data-driven precision training[1] in athlete training has gradually become prominent. Through the method of data collection, data processing, and analysis feedback, athletes can guide daily training, achieving accurate load quantification, accurate loss warning, and accurate training effectiveness evaluation. Similarly, accurate training feedback in rehabilitation training can better guide doctors to develop more accurate and personalized plans for patients.

Motion capture is an important way to achieve precision sports training, and its application in sports training and rehabilitation training can enable the quantification of training data. Guidance training has shifted from experience to a scientific, quantitative, and data-based data analysis model[2]. Motion capture also has very important application value in the prediction of sports injury, which can effectively predict the possible injuries caused by posture issues during sports, thereby effectively reducing the risk of sports injury.

Currently, motion capture systems have been applied to sports training, event assistance, rehabilitation training, and other aspects. However, in China, there are problems such as complex machine structures, expensive prices, and bias values that have led to the lack of widespread promotion of motion capture systems, which still require detailed research.

In this article, the use of a camera without marking points combined with a pressure sensor for motion data collection makes the device simpler and less restrictive. To some extent, it solves the problem of complex machine structure and high price in the current application of motion capture in sports. The contributions of this article are as follows:

1) High precision 3D motion acquisition and subsequent key point position optimization. Through weight based re projection optimization and Levenberg Marquardt algorithm[3], gradient descent is performed to calculate the position of key points, improving accuracy; (High precision 3D dynamic and inverse dynamic analysis based on visible light vision)

2) A mechanical analysis of human joints based on unmarked motion capture system and mechanical data feedback is proposed.

ISSN 2616-5775 Vol. 6, Issue 6: 55-61, DOI: 10.25236/AJCIS.2023.060608

2. High precision 3D motion capture and continuous posture estimation method based on multi vision

Under visible light vision conditions, there is a problem of occlusion of key points in optical capture systems. To solve this problem, multiple cameras can be used to collect images from multiple angles. By placing a group of cameras in space, a camera calibration algorithm is used to restore the content captured by the camera to three-dimensional coordinates.

The motion capture process is as follows:

1) Complete the layout of the camera, calibrate the camera, and obtain the internal and external parameters of the camera;

2) The person being collected completes the specified action and synchronously maintains data from multiple cameras;

3) The image is subjected to distortion processing, and the pixel coordinates of joint points are extracted using an open source attitude estimation library;

4) According to the pixel coordinates of the same joint point in different cameras, combined with the internal and external parameters of cameras at different positions, the three-dimensional coordinates of the human joint point are reconstructed, and accurate three-dimensional coordinates are obtained using weight based re projection optimization.

2.1. High precision 3D motion capture method based on multi vision

Before performing motion capture, it is necessary to calibrate the internal and external parameters of the camera. In this experiment, the coordinate system of camera 1 is selected as the world coordinate system, so the translation matrix T of camera 1 is zero, and the rotation matrix R is E.

There is a conversion relationship between the coordinate system in the camera and the coordinate system in the image as follows:

$$Z_c\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f_x&0&u_0\\0&f_y&v_0\\0&0&1\end{bmatrix}\begin{bmatrix}X_c\\Y_c\\Z_c\end{bmatrix} \tag{1}$$

$\begin{bmatrix}f_x&0&u_0\\0&f_y&v_0\\0&0&1\end{bmatrix}$ is the internal parameter matrix of the camera, which can be obtained by calibrating the camera. (X_c,Y_c,Z_c) is the position of the point in the camera coordinate system, and u, v is the pixel coordinates of the point in the image coordinate system.

The coordinate system transformation between different cameras and between cameras and the world coordinate system can be calculated by the following formula:

$$\begin{bmatrix}X_c\\Y_c\\Z_c\\1\end{bmatrix}==\begin{bmatrix}R_{3*3}&T_{3*1}\\0^T&1\end{bmatrix}\begin{bmatrix}X_m\\Y_m\\Z_m\\1\end{bmatrix} \tag{2}$$

In the above formula, R and T respectively represent the rotation direction and displacement distance of the world coordinate system relative to each coordinate axis of the camera coordinate system. Together, they are the external parameters of the camera.

The values of external parameters can be obtained through camera calibration. In the study, Zhang Zhengyou calibration method was used[4] to calibrate cameras 1, 2, 3, and 3, 4 in pairs. Based on the transitivity of matrix transformation, a transformation matrix corresponding to the world coordinate system for cameras 2, 3, and 4 can be obtained.

2.2. Optimization method for human posture estimation based on reprojection optimization

This article uses the open source program OpenPose[5] to extract the pixel coordinates of human joint points in the image. This project is an open source library developed by Carnegie Mellon University based on convolutional neural networks and supervised learning. The 2D pose detection method PAF (Part Affinity Fields) in the image is used, that is, the partial affinity field. Using the bottom-up method,

ISSN 2616-5775 Vol. 6, Issue 6: 55-61, DOI: 10.25236/AJCIS.2023.060608

first detect each joint point, and then connect them with the individual

Compared to other open source programs, it has the characteristics of high recognition accuracy. OpenPose outputs pixel coordinates in joint point order for each frame, so when matching key points between multiple cameras, we only need to match by row.

In binocular stereo vision, if two cameras with identical internal parameters are placed in parallel, the main optical axes of the two cameras are parallel, the imaging planes coincide, and there is a point $P(x_c, y_c, z_c)$ in space, forming a pixel point P on each of the two cameras $P_r(X_r, Y_r)$, $P_t(X_t, Y_t)$, then:

$$\begin{cases} X_r = f\frac{x_c}{z_c} \\ X_t = f\frac{(x_c - B)}{z_c} \\ Y = f\frac{y_c}{z_c} \end{cases} \tag{3}$$

The Y coordinates of the two image coordinates are equal, but there is only a difference in the X coordinate. Let's make D = X_t - X_r, called parallax, can then obtain the three-dimensional coordinates of the spatial point P:

$$\begin{cases} x_c = \frac{BX_r}{D} \\ y_c = \frac{BY_r}{D} \\ z_c = \frac{Bf}{D} \end{cases} \tag{4}$$

From this, preliminary results of three-dimensional reconstruction can be obtained.

2.3. Weight based reprojection optimization

Considering that parallel binocular cameras may block certain parts of the human body at some angles, resulting in inaccurate 3D key point positions, we use the beam adjustment BA (Bundle Adjustment) algorithm[6] to optimize the spatial position of key points.

As shown in the figure, due to the phase error problem, the reconstructed key point X_1 in camera P_2, P_3 pixel coordinates obtained after reprojection are different from the pixel coordinate positions extracted in chapter 2.2, and there is an error, namely, re projection error. By minimizing the re projection error, the spatial position of key points is further optimized.

Command point X_j in camera P_i the normalized coordinate system of the captured image is:

$$k(u_{ij}^T, 1)^T = K_i^{-1} x_{ij} \tag{5}$$

The coordinates under the normalized coordinate system of the graph after re projection are:

$$k'(v_{ij}^T, 1)^T = K_i^{-1} P_i X_j \tag{6}$$

Where K_i^{-1} is a constant term used to convert homogeneous coordinates to non homogeneous coordinates, and k and k' are constant terms that are independent of camera internal parameters in the calculation. It can be obtained that the reprojection error is:"

$$e_{ij} = u_{ij} - v_{ij} \tag{7}$$

Minimize the sum of all reprojection errors, i.e.:

$$min \sum_{i=1}^{n} \sum_{j=1}^{m} e_{ij}^2 \tag{8}$$

Considering that different cameras have different shooting angles for the same key point, and the presence of occlusion problems can lead to a large deviation in the data of a certain camera, so key points taken by different cameras should have different confidence levels. In this article, the confidence level p is introduced from the calculation results of OpenPose. Therefore, the final optimization function is:

$$min \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} e_{ij}^2 \tag{9}$$

For this optimization model, the LM (Levenberg Marquardt) algorithm is used and based on this, the sparse nature of the BA model is utilized for calculation. Using the Levenberg Marquardt algorithm for gradient descent, P (X, Y, Z) converges through iteration. The spatial coordinates of the feature points are obtained. In the Levenberg Marquardt algorithm, each iteration is to find an appropriate damping

ISSN 2616-5775 Vol. 6, Issue 6: 55-61, DOI: 10.25236/AJCIS.2023.060608

factor λ, When λ At an early age, the algorithm became the optimal step size calculation formula of the Gaussian Newton method, λ When it is large, it is transformed into the optimal step size calculation formula of the gradient descent method. Levenberg Marquardt algorithm is a nonlinear optimization method between Newton's method and gradient descent method. It is not sensitive to over parameterization problems, and can effectively handle redundant parameter problems, greatly reducing the chance of cost functions falling into local minima. These characteristics enable Levenberg Marquardt algorithm to be better applied to such optimization problems.

The 3D trajectory of the processed key points is not smooth enough, and the joint point data has jitters or abnormal mutations. In order to reduce the impact of noise signals on subsequent calculations, it is necessary to filter the key point data. Median filtering is the optimal filtering under the "minimum absolute error" criterion. Median filtering can effectively filter noise, especially while protecting information edges. Moreover, the median filtering algorithm is easy to implement, and if the window size is selected appropriately, its calculation speed is also very fast, which is conducive to real-time data processing by the system. Subsequent experiments also verified that the filtering algorithm only eliminated all abnormal points in the data and smoothed the data. The data did not lag or oversmooth due to filtering, and the maximum program retained the authenticity of the data.

3. Inverse Dynamics Tracking Analysis Method Based on Human 3D Pose

In this article, we use the human musculoskeletal model in OpenSim[7], an open source program from Stanford University, and scale it based on it, adjusting it to be basically consistent with the parameters of the tested person. By combining the position data of human key points and pressure data, we perform inverse dynamic calculations in OpenSim to obtain joint dynamics data.

3.1. Mechanical modeling of human lower limbs

A musculoskeletal model is one in which multiple bones are connected by various joints, in which the muscles attach to the bones and drive joint movements through the forces generated by the muscles. OpenSim uses a person's height, weight data, and individual muscle characteristics data to establish a generic model. To obtain a personalized model, it is necessary to scale the generic model to adapt to the different body characteristics of the subject. The model scaling is based on laboratory test key point data, and scales the length and mass of each link according to the ratio between the experimental data and the human body key points in the general model. The least square method is used to control the error between the marked points in the experiment and the theoretical arguments in the model during the scaling process.

$$\min \sum_{i \in \text{ markers}} w_i \left(x_i^{\text{exp}} - x_i^{\text{model}} \right) \tag{10}$$

Where x_i^{exp} is the experimentally measured coordinate of the joint point i, x_i^{model} is the coordinate of the joint point in the model, w_i is the weight.

The comparison of human musculoskeletal models after scaling the standard model based on a real human model is shown in Figure 1.

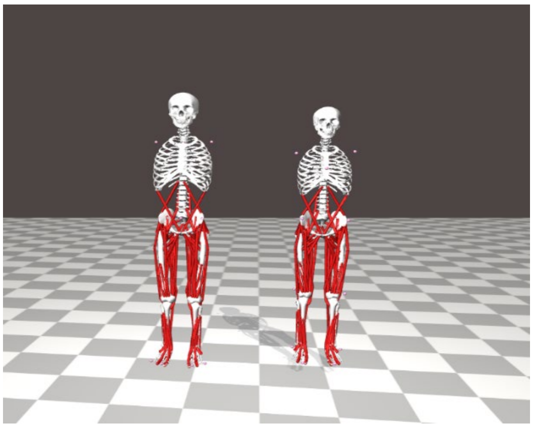

Figure 1: Comparison of the front (right) and rear (left) manikin before and after zooming.

ISSN 2616-5775 Vol. 6, Issue 6: 55-61, DOI: 10.25236/AJCIS.2023.060608

3.2. Inverse dynamics solution

Given the movement status of the body and the external forces, combined with the human body mechanics model, the forces and moments of each joint are calculated using inverse dynamics solutions. There are two processes involved.

The first process is called Inverse Kinematics IK[8], in which the motion velocities of each joint are calculated. Kinematics is the study of motion without considering the forces and moments that generate motion, so when doing kinematic analysis, it is not necessary to know the mass and inertia of an object. The goal of inverse kinematics is to find a model joint angle that best reproduces the experimental kinematics of a particular object. The experimental kinematics data used in inverse kinematics here is based on high-precision joint points in the experiment. Inverse kinematics traverses every frame of an action, and then calculates a set of joint angles to allow the model to "best" match experimental data. OpenSim finds the "best match" by solving weighted least squares optimization problems that minimize labeling errors. The marking error is defined as the distance between the experimental marking point and the corresponding virtual marking point. Each marker point has a weight value that represents the degree to which the marker error term is minimized in the least squares problem. In each frame, the inverse kinematics tool finds a generalized coordinate vector (such as joint angle) q, minimizing the weighted sum of marker point errors, expressed as:

$$min_q[\sum_{i \in markers} w_i \left\| x_i^{exp} - x_i(q) \right\|^2] \tag{11}$$

Where x_i^{exp} is the position of the key points of the human body captured in the experiment, $x_i(q)$ is the location of the corresponding virtual marker point i (depending on q), w_i is the weight corresponding to the marker point i. Each key point should be given a different weight, which represents the degree of fitting of the key points during model motion. The higher the weight, the better the fitting effect. "We choose to give higher weights to the knees and ankles that we focus on."

The second process is called Inverse Dynamics ID (Inverse Dynamics)[9]. The solution is to calculate the force and torque output of each joint in reverse according to Newton's second law, given the motion status of the body and the external forces. The inverse dynamics results can be obtained by combining pressure data with IK solution results. The solution equation is as follows:

$$M(\dot{q}) + C(q,\dot{q}) + G(\dot{q}) = \tau \tag{12}$$

Where $q, \dot{q}, \ddot{q} \in R^N$ is a generalized vector of position, velocity, and acceleration, $M(q) \in R^{N\times N}$ Is the mass matrix of the entire system, $C(q,\dot{q}) \in R^N$ is the Coriolis force, $G(q) \in R^N$ is the gravitational vector, $\tau \in R^N$is the unknown quantity to be solved.

4. Experiment and Result Analysis

4.1. Experimental setup

This experimental study has designed a collection device for synchronously acquiring human body images and foot pressure. In the experiment, four sets of cameras have a resolution of 720 × 540. An RGB camera with a refresh frequency of 120 Hz is used as an image collector, and an oscilloscope is attached as a signal synchronization device between cameras. A pressure sensor with a collection frequency of 20 Hz is used to obtain pressure data. In the experiment, multiple cameras were used to focus on the person being captured, and to enable clear imaging of the person in each camera. In the experiment, we collected walking data from a healthy male.

The mechanical calculation in this article is based on the method in OpenSim developed by Stanford University, using Gait2354_ Simbody, a human musculoskeletal model, is calculated by merging the collected body data.

To verify the accuracy of human posture capture, three-dimensional reconstruction of image information is performed, and the length of limbs in the method is calculated for each frame. Theoretically, the length of the limb remains unchanged, but due to system errors, the limb length fluctuates in the capture result. In order to characterize the accuracy of captured data, this experiment uses the average absolute percentage error MAPE as an evaluation indicator. The formula is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{t-t_i}{t}\right| \times 100\% \tag{13}$$

ISSN 2616-5775 Vol. 6, Issue 6: 55-61, DOI: 10.25236/AJCIS.2023.060608

Where: n represents the total number of video frames; $\dot{t_i}$ represents the limb length calculated for the ith frame; t represents the true value of the limb. The smaller the MAPE error value, the smaller the measurement error, and the more accurate the measurement result.

In order to analyze the validity of the mechanical calculation results, a healthy adult's walking gait is divided into periodic stages, and the joint angle and joint torque changes in each stage are qualitatively analyzed to evaluate the effectiveness of the solution results.

4.2. Analysis of experimental results

To verify the accuracy of motion capture results, this study extracted data and compared them with real data to calculate their MAPE errors. Through the calculation and analysis of MAPE errors for the same motion capture result using various methods, the experimental results show that the accuracy of this method is significantly improved compared to mainstream methods in terms of motion capture accuracy, with the error percentage ranging from two digits to one digit, and the motion capture accuracy reaching millimeter level in the area of 7 meters * 7 meters. Significant improvement in method accuracy.

The magnitude of MAPE errors in various optimization methods is shown in Table 1.

Table 1: Calculation Table of MAPE Error Values.

Comparative method	Left forearm	Right forearm	Left thigh	Right thigh
Triocular least square method	13.05%	13.03%	13.21%	13.12%
Quadratic least square method	10.57%	10.41%	10.61%	10.37%
Reprojection optimization method	5.21%	5.17%	5.29%	5.24%

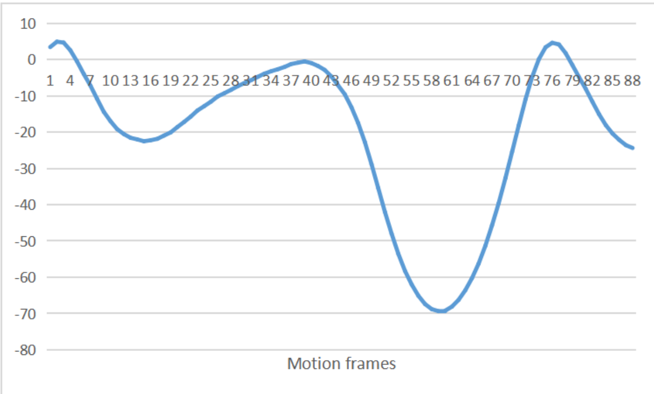

Figure 2: Variation of knee joint angle with motion.

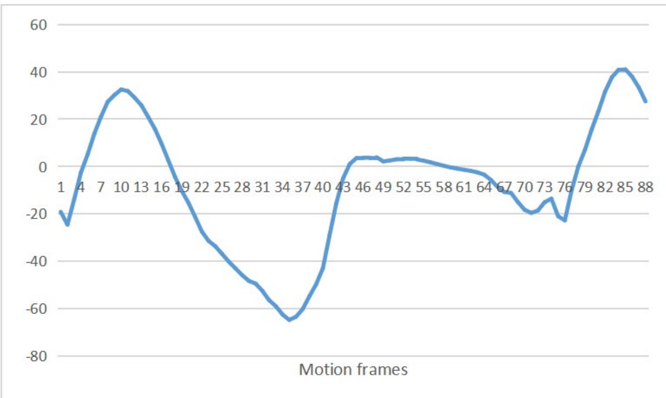

Figure 3: Variation of knee joint torque with motion.

According to the gait cycle theory[10], each walking cycle can be divided into two stages: the support phase and the swing phase. The support phase is the entire process of contact between the foot and the ground, starting with initial landing. The swing phase refers to the period during which the limb moves forward without contact between the foot and the ground (moving in the air). The swing phase begins at the moment the foot is lifted off the ground. And the variation of joint angle and torque with time is

ISSN 2616-5775 Vol. 6, Issue 6: 55-61, DOI: 10.25236/AJCIS.2023.060608

obtained by solving as shown in the Figure 2 and Figure 3:

For the analysis of the right knee dynamics data in Figure 2 and Figure 3, it can be divided into support phases from frames 1-46, and swing phases from frames 47-75; During the foot following ground period (starting at frame 1), the knee joint angle is minimal and gradually increases, at which point the knee joint has started to exert force; During the landing period of the foot, the bending angle of the knee joint reaches the local maximum, and the force of the knee joint gradually reaches the local maximum; During the mid support and heel off ground periods (19-40 frames), the knee joint angle decreases to 0 and the knee torque reaches the maximum value; 40-46 frames can be divided into a toe off ground period, during which the knee angle gradually increases, while the knee force gradually decreases, until the toe off ground period is the minimum value of 0; During the swing phase, the angle of the knee joint gradually decreases, and the joint does not exert force at this time, with the torque around 0.

According to the analysis of knee joint angle and torque changes, the obtained data changes can basically reflect the joint angle and torque at each stage of the gait cycle.

5. Conclusions

In this study, posture estimation methods are used to capture three-dimensional joint points of the human body. By using multiple cameras, high-precision human motion capture is achieved, which to some extent overcomes the hardware limitations of acquisition. A high-precision optimization strategy was proposed to improve the accuracy of three-dimensional data of human joint points. After the implementation of this study, data comparison found that the accuracy of the optimized three-dimensional reconstruction algorithm was improved by 5% compared to the least square method. In addition, a human body mechanics analysis based on high-precision human motion data is proposed. This method combines the OpenSim human body model and calculation method, and uses the human motion data obtained by this method to analyze the gait of subjects. This method conforms to the generality of walking gait rules, and can better reflect the stress situation during walking.

In the future, this method will continue to be improved and applied to the analysis of other motions, expanding its scope of application.

References

[1] Ghasemzadeh H, Loseu V, Jafari R. Wearable coach for sport training: A quantitative model to evaluate wrist-rotation in golf [J]. Journal of Ambient Intelligence and Smart Environments, 2009, 1(2): 173-184.

[2] Johnson W R, Mian A, Donnelly C J, et al. Predicting athlete ground reaction forces and moments from motion capture [J]. Medical & biological engineering & computing, 2018, 56(10): 1781-1792.

[3] Moré J J. The Levenberg-Marquardt algorithm: implementation and theory [M]//Numerical analysis. Springer, Berlin, Heidelberg, 1978: 105-116.

[4] Zhang Z. A flexible new technique for camera calibration [J]. IEEE Transactions on pattern analysis and machine intelligence, 2000, 22(11): 1330-1334.

[5] Osokin D. Real-time 2d multi-person pose estimation on cpu: Lightweight openpose [J]. arXiv preprint arXiv:1811.12004, 2018: 2-5.

[6] Triggs B, McLauchlan P F, Hartley R I, et al. Bundle adjustment—a modern synthesis [C] // International workshop on vision algorithms. Springer, Berlin, Heidelberg, 1999: 298-372.

[7] Delp S L, Anderson F C, Arnold A S, et al. OpenSim: open-source software to create and analyze dynamic simulations of movement [J]. IEEE transactions on biomedical engineering, 2007, 54(11): 1940-1950.

[8] D'Souza A, Vijayakumar S, Schaal S. Learning inverse kinematics[C]//Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180). IEEE, 2001, 1: 298-303.

[9] Nguyen-Tuong D, Peters J, Seeger M, et al. Learning inverse dynamics: a comparison[C]//European symposium on artificial neural networks. 2008 (CONF): 2-4.

[10] Trivino G, Alvarez-Alvarez A, Bailador G. Application of the computational theory of perceptions to human gait pattern recognition [J]. Pattern Recognition, 2010, 43(7): 2572-2581.