Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3, 2761-2773
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i3.5873

Auxiliary motion training system based on wireless human pose computer vision estimation algorithm

Yaoze Gong[1], Jiankun Gong[2], Ting You[3*]
[1,2]Department of Media and Communication Studies, University of Malaya, Kuala Lumpur, 50603, Malaysia; gongyaozesdty@163.com (Y.Z.G.); jk_gong@um.edu.my (J.K.G.).
[3]Department of Information Technology, Shanghai Industrial and Commercial Polytechnic, Shanghai, China; andelating@yeah.net (T.Y.).

Abstract: With the advancement of science and technology, many technologies are gradually being used in various aspects. In terms of sports, computer vision has been comprehensively utilized, making training more effective. Human pose estimation reflects various information regarding the dynamics of the body by establishing the similarity between different parts of the human body, such as angles and positions. The purpose of this paper is to study a computer vision estimation algorithm based on wireless human pose and to combine it with image processing. By scanning human body information through the computer, basic information about the person and the corresponding images can be obtained, allowing for the estimation of human joint posture. In terms of methodology, compared with traditional training methods, the results showed that training under computer vision improved the accuracy of joint data by 16%, the accuracy of coaches' judgments on athletes by 15%, and the performance of students by 14.8%. In conclusion, it was also shown that the auxiliary training system based on wireless computer vision was beneficial to the training of athletes and could help trainees train more scientifically and effectively.
Keywords: *Auxiliary training system, Computer vision, Pose estimation, Visual estimation algorithms, Wireless human pose.*

1. Introduction

In the traditional sports mode, it is generally after a long period of training, or the coach's experience is used to judge whether the training is qualified. However, with the discovery of computer vision technology, intelligent training methods have been adopted one after another. According to the intuitive information transmitted in the computer, various information data of athletes can be directly learned. This technology can help coaches train better and give more reasonable training guidance.

The application of computer vision technology in physical education is still in its infancy. This paper mainly studies computer vision techniques in sports training. By detecting and tracking the motion characteristics of the moving human body, it can analyze the motion posture and provide training suggestions. An upgrade can be made in the previous sport mode. Judging by the results of the above research, it can be found that the research has great application potential in theory, science and engineering.

In this paper, an auxiliary training system under computer vision is introduced through the shortcomings of traditional training methods. The comparison is mainly made in three aspects: the accuracy of the coach's judgment on the athlete, the athlete's performance, and the accuracy of the joint data. It is found that the accuracy of joint data for traditional training is about 72%, but the accuracy of joint data for training under comp*uter vision is about 88%. The accuracy of the coach's judgment on

History: Received: 3 January 2025; Revised: 14 March 2025; Accepted: 18 March 2025; Published: 29 March 2025
* Correspondence: andelating@yeah.net

the athlete has also been improved, and the performance of the athlete has also been improved, which also shows the scientificity and efficiency of the auxiliary training system.

2. Related Work

The auxiliary training system is a system that helps sports athletes in training. It mainly uses technology to provide scientific training guidance to trainers, which plays an important role in athletes. Through an understanding of the problem of muscle damage in sports, Liu X developed a squat training assistance system with MPU9250 as the core. The system utilized embedded technology and the cloud structure of the recurrent neural network to analyze the data and monitor the deep training. The results were fed back to the mobile APP to achieve correct guidance for each step, which can truly display the user's recovery status

[1]. PCA and KPCA are an unsupervised learning method. They maximize the overall variance to the greatest extent possible while ignoring both intra- and inter-class information. In response to this problem, Chen S proposed a simple and efficient method to improve the performance of PCA and extended it to KPCA. The algorithm used an intra-class training sample based on linear interpolation. Experiments showed that the algorithm was effective [2]. Athletes are trained mainly to improve their performance and personal level. In training, coaches also have their own training methods to help athletes reach their goals. Li C had a different idea, and proposed a new view of training methods in sports based on computer technology. Through the use of computer data and technology, scientific training was provided, and a new training mode was given [3]. Physiological data played a very critical role in the athlete's training and competition. The system was divided into hardware layer, data processing layer, algorithm layer and interface layer. At the hardware level, Xiong D used Berkeley's Tricorder platform. The data processing layer integrated data from individual sensors into the adapter schema. The test results showed that the system could better help coaches to monitor and analyze the training and training status of athletes [4]. Scholars have carried out a series of researches on auxiliary training systems, but not through computer vision estimation algorithms.

In the estimation of human body posture, many scientists have also conducted in-depth research on it. Pei Yean L believed that through computer technology, the human body can be traced, a model shop can be established, and various information of the human body can be mapped in the 3D model. This problem was a class of unconstrained optimization problems with special orthogonal and conic intersections, which could make full use of the fundamental parameter space. A direct proof of local quadratic convergence to the optimal solution was given. In the process of 5-10 iterations, the algorithm can significantly reduce the error of attitude estimation, and can reach the global optimum in 5-10 iterations [5]. Head pose estimation has been a frequently discussed topic, and it has recently attracted attention. It has potential for human-computer interaction, augmented reality and driver assistance. However, most of the current work is carried out in a controlled environment and does not have good stability. In response to the above problems, Patacchiola and Cangelosi [6] proposed a new technology based on convolutional neural network, and introduced deep learning technology on this basis. The research results showed that the head pose could be effectively estimated using the CNN algorithm and the adaptive gradient method [6]. The Mirage algorithm is a linear time-domain algorithm capable of analyzing and solving multi-camera pose parameters. This method uses the reference camera pose to minimize the two-dimensional projection error between the reference and the real pixel coordinates, so as to obtain the pose. DincS had a comprehensive review of Mirage. Experiments were conducted with and without noise using simulation and actual data. Experimental results showed that Mirage could produce better fast and accurate results in various test environments [7]. The various methods used by scholars in pose estimation are not carried out under the wireless human pose computer. In this regard, this paper used the method of visual estimation to assist athletes' exercise training based on the wireless human posture computer.

3. Establishment of Auxiliary Training System

3.1. Human Pose Estimation Algorithm

The standard of the algorithm is mainly to see whether there is a model, and the algorithm mainly adopts the method of combining statistical learning and matching [8]. Based on the distribution of plane and space, the pose estimation of the human body can be divided into two types. Model-based human pose estimation algorithms mainly rely on related motion information and image segmentation techniques. And due to the limitations of the human body model, in general, a model can estimate a specific action well. Therefore, model-based approaches for motion analysis and behavioral understanding can be used for continuous or single images. The method is not constrained by the model and is not limited to pose estimation, and is suitable for the analysis and retrieval of a large amount of image data. Model-free pose estimation algorithms can be continuously trained and learned from indexed image data. Therefore, the effect of pose estimation is better. However, this requires more time to train, more samples, and better training algorithms.

Human motion analysis faces many problems [9]. The goal of human motion feature extraction is to extract the three-dimensional pose information hidden in the image from the image. The extraction of human motion features is based on the motion posture and delineation points, and is also an important link in the realization of human body posture estimation. The movements of the human body are complex, and there are problems such as mutual occlusion between the limbs, so it is a very difficult problem to extract the movement features of the human body in real time and accurately. The key to human pose estimation is to realize the relationship between human features and human poses. The posture parameters of the human body are mostly dozens or hundreds, which is a multi-parameter state variable. Therefore, how to select appropriate features, establish appropriate human model to describe these multi-dimensional state variables, and give the corresponding relationship between them is very important. The main content of this paper is to use wireless technology to realize the auxiliary training system of human pose estimation. The focus of this paper is object detection, feature extraction, and human pose estimation [10]. The implementation steps are shown in Figure 1.

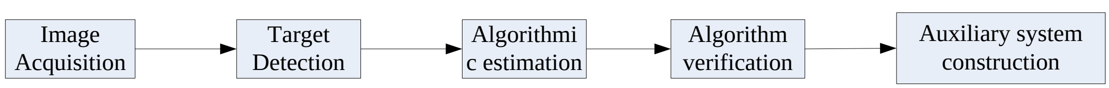

Figure 1.
Implementation steps.

3.2. Image Acquisition and Attitude Data Extraction

This article uses Kinect's 3-D sensor to collect data. The system can simultaneously acquire depth and color images, and process the acquired depth images to obtain the key parts of the human skeleton [11]. This article uses a golf ball as an example. The image data acquired by the Kinect sensor is input into the subsequent golf training system. Then, an algorithm verification platform is established based on the pose data of the five key points extracted by the Kinect sensor, and compared with the method described in this paper. Using Win7x86+VS2008+ Kinect+OpenNI1.5.2+OpenCV2.2.4, Kinect driver can be written on this platform and used in Xbox. This technology can capture moving images from different angles, and the frequency of motion and the number of frames can be captured very accurately [12]. When tracing the human skeleton, it needs to be positioned by a specific technique [13]. By calibrating important parameters and information, the specific technology can carry out the information of various positions in the past. The GetSkeletonJoint() function is used to obtain the above coordinate information, the Kinect camera coordinate system can be used, and ConvertRealWorldToProjective can be used to map the coordinates to the screen coordinate system to make it more intuitive. The information obtained by this method is stored for subsequent use.

3.3. Motion Detection

The main work of motion state detection is to record and analyze the motion state information. The method extracts the object to be searched according to the geometric and statistical properties of the object by processing the image. Effective extraction of human objects and features is an important part of pose estimation. Firstly, the feature extraction and common target detection are briefly introduced, and their advantages, disadvantages and application environment are analyzed.

The success or failure of analytical methods mainly depends on the features they use, such as color, texture, shape, edge gradient, spatiotemporal relationship, etc [14]. Color is a very commonly used image property. Texture and color are an integral feature that has the same mechanics, except that it is based on multiple pixels.

Shape contour feature: A basic visual feature of describing an image is the shape of the object. The shape outline has certain semantic information, which is a higher-level feature than the color feature and the texture feature.

Edge gradient characteristics: In the analysis and research of human motion characteristics, the gradient direction histogram and SIFT characteristics are commonly used [15]. Oriented gradient histogram is a feature descriptor used in image processing and computer vision. It mainly forms features by counting and calculating the gradient direction and size of the local area of the image, thereby obtaining the statistical information of the image gradient. In the specific implementation, the concept of graphics is introduced. Based on the basic unit, a histogram of gradient directions can be constructed according to the gradient of each pixel or the direction of the edge. Their histograms are then combined into a large block and normalized within the block to form a feature description. Compared with other feature description methods, HOG has good invariance and can process local squares in images. SIFT is a method based on LaplacianofGaussian approximation. The method has the characteristics of fast operation speed, rotation invariance and scale invariance. There are many operation steps in the SIFT algorithm. First, a scale space is constructed, which is an initialization operation. Then, by approximating with LaplacianofGaussian, the extreme points of the scale space are detected. Next, the edge and the extreme point with poor brightness are assigned to 128-dimensional direction parameters respectively, so that the operator's rotation remains unchanged. Finally, through the scale and rotation invariance, the feature keypoint descriptor of SIFT is obtained.

Region spatial properties [16]: The algorithm does not require modelling of the background, nor pre-detection and tracking. Therefore, it has been widely used in human motion recognition, scene recognition and other fields. In video images, the local spatial characteristics of space are used to describe the shape and motion characteristics of objects, and a relatively independent expression method of space and spatial characteristics is proposed. The motion analysis based on the local spatial characteristics of space usually adopts the optical flow method, which calculates the motion information between each frame according to the temporal change of each pixel in the image sequence and its correlation, thereby judging the correspondence between each frame. It is effective to use optical flow to capture a change in motion. The optical flow method is best at detecting the general shape and orientation of moving objects. In order to obtain an accurate and complete target surface profile, the optical flow method must be optimized, or other methods must be used [10]. The following is a detailed summary of two common target detection algorithms.

The working principle of background subtraction [17]: Background subtraction is a special frame difference method. Once the background model is established, a rough outline of the foreground object can be obtained by comparing and multiplying the existing image and the background model. The basic idea of background subtraction is: the background remains unchanged, or the background does not change with the number of frames. The grayscale difference between the pixels corresponding to the current frame and the background image can be compared to detect the shape of the object. Its use flow chart is shown in Figure 2.

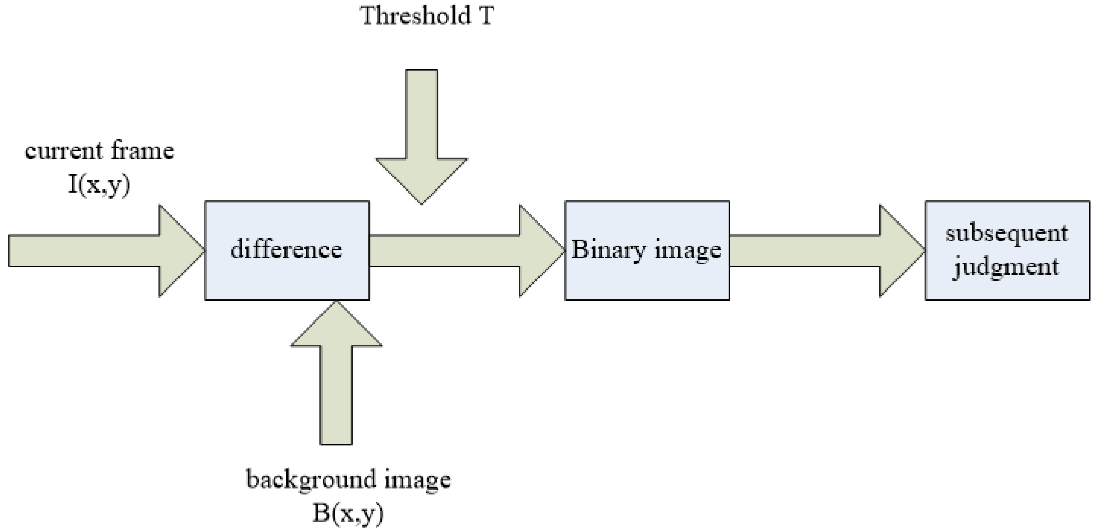

Figure 2.
Flow chart of background subtraction method.

The background image may be expressed in such a way that its order is expressed as to indicate the positional coordinates of the image, to indicate the frame number. The differential image may be obtained by the grayscale values of the respective frames and the grayscale of the background image. Noise in the process is removed by thresholding the image, and the specific formula is:

$$\mathrm{D(x,y)} = \begin{cases} 1, (\mathrm{I(x,y)} - \mathrm{B(x,y)}) \geq \mathrm{T} \\ 0, (\mathrm{I(x,y)} - \mathrm{B(x,y)}) < \mathrm{T} \end{cases} \quad (1)$$

The thresholded pixel points are eliminated by drawing points in the current picture frame within the prescribed interval of the threshold value. And this method can also act on the point capture of human motion, and eliminate the points outside the bounds in the picture to obtain the difference.

The method of frame number difference is to deal with the thresholding of points captured in motion scenes and the same number of frames in the case of similar background [18]. Through the obtained information, the state of human movement can be analyzed. The steps are similar to Figure 2, which can be expressed as:

$$D_k(x,y) = |I_k(x,y) - I_{k-1}(x,y)| \quad (2)$$

$$F_k(x,y) = \begin{cases} 0, D_k(x,y) \leq T \\ 255, D_k(x,y) \geq T \end{cases} \quad (3)$$

3.4. Model-Based Human Pose Estimation

Model-based human pose estimation methods play a key role in the pose estimation and tracking process [19]. The method is to build joint trees or other complex models of human targets. By capturing the data information under the motion of the human body, the model parameters of the response are established, the parameters of the human body model are obtained, and finally the purpose of human body pose estimation is realized. It is a top-down approach.

In the case of model parameters, the moving information points can be processed to control the required points within a certain range. There are three types of model construction, which are generally constructed through the shape of the human body. The human body shape is also an important parameter in the modeling of the human body, which affects the main points, bone points and joint points in the modeling, as shown in Figure 3.

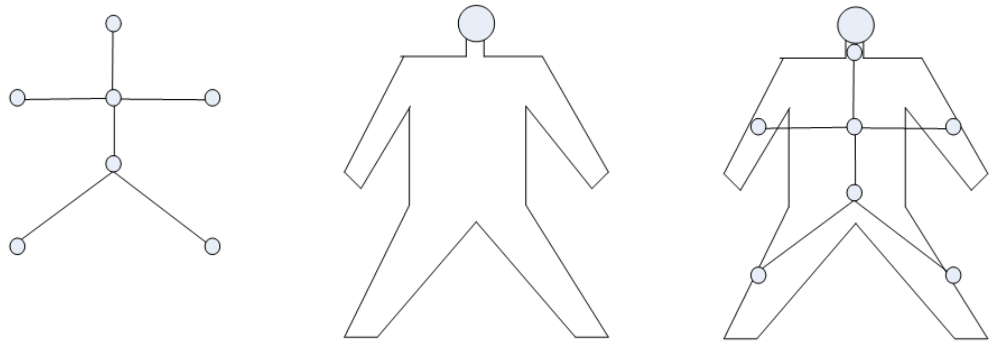

Figure 3.
Human body structure model diagram.

In the established model, the main thing is the key points in it, which is the most concise expression of human body structure [20]. There are two kinds of graph structure modes: one is tree graph structure mode, and the other is non-tree mode.

In human pose estimation, its main content is to map the obtained human body dynamic data, and compare the mapped map with the standard parameters to obtain the best model parameter points.

The goal is to make pose maneuverability and search range smaller. Human pose estimation based on wireless technology is to search and solve the space with more poses, the best data parameters can be obtained from the comparison. Many researchers take human behavior as the theoretical basis for the continuous evolution of time series relationships, and build a dynamic mathematical model through methods such as the Gaussian method to simulate a variety of situations. Its main method is to simulate some unreachable movements of the limbs in the state of motion of the human body.

The pose estimation algorithm of this model not only has high estimation accuracy, but also can use the existing knowledge for modeling, which has good universality. However, the disadvantage is that the posture parameters of the human body are multi-dimensional state variables, which must be searched and matched in high-dimensional planes. The real-time performance is poor, the optimization speed is slow, and the noise and errors gradually increase in the search process, and it is difficult to restore, resulting in the failure of attitude estimation.

Based on the attitude estimation in the model parameters, it is necessary to perform multiple queries on the data points, and the query time is relatively long. Therefore, it is difficult to determine its timeliness. In the case of no model parameters, it takes less time to accurately find reasonable data. However, the calculation process is still relatively complicated, and the various costs are relatively large. Through these methods, accurate information in the case of human motion can be obtained. The schematic diagram is shown in Figure 4.

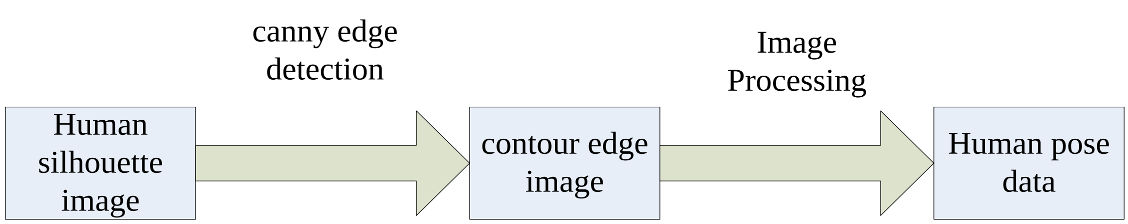

Figure 4.
Schematic diagram of pose estimation

First, through Canny's edge extraction, its edges are obtained. Then, the information obtained by computer vision is calculated, and the positions of key parts of the human body are obtained, so as to realize the estimation of human body posture. Its main contents are:

The contour lines are extracted from the human body. The image is smoothed with a Gaussian filter, which can effectively remove noise. There are two ways to implement Gaussian filtering. One is to use a one-dimensional Gaussian kernel to perform secondary weighting respectively. A one-shot convolution method is used, which is done at the core of a 2D Gaussian.

The one-dimensional Gaussian kernel function is:

$$\mathrm{K} = \frac{1}{\sqrt{2\pi\sigma}} \mathrm{e}^{\frac{x^2}{2\sigma^2}} \quad (4)$$

By determining the parameters, a one-dimensional kernel vector can be obtained.
The two-dimensional Gaussian kernel function is:

$$\mathrm{K} = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

By determining the parameters, a two-dimensional kernel vector can be obtained.

The first order difference computes the magnitude and direction of the gradient. Two matrices of partial derivatives in the two directions of X and y can be calculated by the finite difference of the first-order partial derivatives, and the gradient of the gray value of the image can be approximately represented by the first-order difference. The expressions of the image gray value gradient are shown in Formulas (6) to (10).

$$H_1 = \begin{vmatrix} -1 & -1 \\ 1 & 1 \end{vmatrix} \; H_2 = \begin{vmatrix} 1 & -1 \\ 1 & -1 \end{vmatrix} \quad (6)$$

$$\tau_1(m,n) = f(m,n) * H_1(x,y) \quad (7)$$

$$\tau_2(m,n) = f(m,n) * H_2(x,y) \quad (8)$$

$$\tau(\mathrm{m},\mathrm{n}) = \sqrt{\tau_1{}^2(\mathrm{m},\mathrm{n}) + \tau_2{}^2(\mathrm{m},\mathrm{n})} \quad (9)$$

$$\theta_\tau = tan^{-1} \frac{\tau_2(m,n)}{\tau_1(m,n)} \quad (10)$$

Among them, the Canny algorithm is used, where H_1 and H_2 represent the factors in the algorithm, and the main influence is the amplitude and trend of the edge.

Gradient magnitudes are subjected to non-maximum suppression. Only through the amplitude and movement of the edge, the position cannot be completely understood, and the situation of the edge needs to be controlled before it can be used as the real edge data. This step is a critical step in the detection, as shown in Figure 5. The principle of non-maximum suppression is described below.

DOI: 10.55214/25768484.v9i3.5873

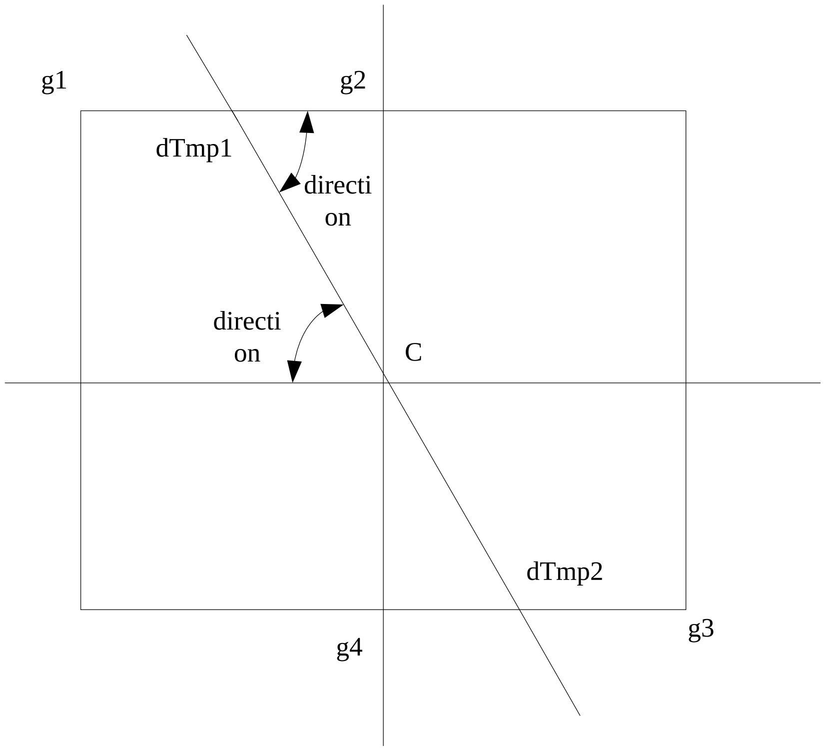

Figure 5.
Schematic diagram of non-maximum suppression.

The direction of point C can be seen from the slope of the straight line, and its edge is the covered area and the range of values. At the intersection point, the values of each area are compared to distinguish whether the point is an edge or not. When point C is below the intersection, it is the largest in the area, otherwise it is the boundary.

The graph at the high point has a small number of edges, and the graph at the low point forms connections to close the edges.

Automatic extraction of joint points: After an edge is obtained, it becomes a tightly connected pixel point, which contains a series of position information. In order to obtain ideal human joint data, this paper performs edge processing on it. In the human skeleton model, it has certain representativeness and characteristics, representing the head, shoulder, elbow and other parts respectively. These joint points are arranged in a certain order to form a person's joint points, as shown in Figure 6.

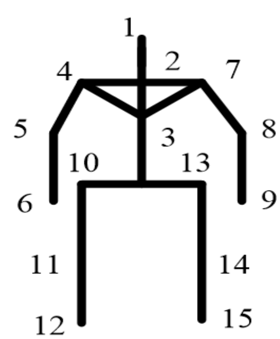

Figure 6.
Human 2D skeletal model.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 2761-2773, 2025
DOI: 10.55214/25768484.v9i3.5873

In Figure 6, the numbers are 1 to 15, and the corresponding relationship with each joint point of the human body is shown in Table 1. The length ratio constraint refers to the ratio of the length of each limb of the human body to the height. If the height of the human body is L, the length ratio constraints of the human body are shown in Table 2.

Table 1.
Correspondence of each joint point of the human body.

Human Skeleton Model Label	Corresponding node	Human Skeleton Model Label	Corresponding node
1	Head	9	Right wrist
2	Neck	10	Left hip
3	Chest	11	Left knee
4	Left shoulder	12	Left foot
5	Left elbow	13	Right hip
6	Left wrist	14	Right knee
7	Right shoulder	15	Right foot
8	Right elbow		

Table 2.
The relationship between the proportions of each bone and joint in the human body.

Limb joint labels	Relative skeleton length
1-2	0.1L
2-3	0.1L
2-4， 2-7	0.127L
4-5， 7-8	0.2L
5-6， 8-9	0.245L
10-11， 13-14	0.273L
11-12， 14-15	0.284L

At the edge, a horizontal line should be drawn from the bottom to the top, and then it should be translated in the direction of the straight line, because the soles of the feet are at the center point and are opposite. Thus, two intersection points are formed where the horizontal line and the edge of the outline meet. The intersection from left to right represents the turning point of the left foot and the right foot, respectively. At this time, the position coordinate on the horizontal axis is H1. A horizontal line is drawn on the horizontal line, and then moves in the direction of the straight line. When the level line and the edge of the outline intersect, this point is used as the head point. At this time, the position coordinate on the horizontal axis is H2. Then the height detected at this time is the difference between H2 and H1.

According to the length ratio of the human body, the lengths of the foot and knee joints are 0.284 L and 0.273 L, respectively. A horizontal direction on the sole of the foot is made and moved up 0.284 L in a straight line. Four intersection points between the horizontal plane of the human body and the boundary contour image of the body are obtained, and the joint points of the left knee and the right knee are drawn at the center points of the two intersection points. Then, the position of the foot joint point to the knee is calculated. If it is not within the human scale factor, other location information can also be calculated by calculating the nodes.

A vertical line is drawn on the left, moving from left to right. The intersection of the vertical line and the left side is left off, the length of the hand and elbow is 0.245 L, and the length of the elbow and shoulder is 0.2 L. A vertical line is drawn from the hand joint point to move it 0.245 L to the right. From top to bottom, the left elbow is took as the starting point, and moved 0.2 L to the right from the elbow, and took the vertical line as the left arm. Similarly, the key parts of the right hand can also be obtained.

4. Comparison between Auxiliary Sports Training System and Traditional Training

In this paper, VC2008 combined with Open CV image processing library is used to extract the joint data of athletes. The training system based on the wireless human pose computer vision estimation algorithm is compared with the traditional sports training, and the number of training is three times. It mainly compares the accuracy of the athlete's judgment by the coaches, the athlete's performance, and the accuracy of the joint data, to judge whether the auxiliary training system under computer vision can provide better scientific guidance.

The information of joint data can reflect the situation of athletes during training, and can be displayed more intuitively, which is of great benefit to coaches. The comparison chart of the joint data accuracy of traditional training (indicated by # in the Figure) and the accuracy of joint data based on computer vision (indicated by * in the Figure) is shown in Figure 7.

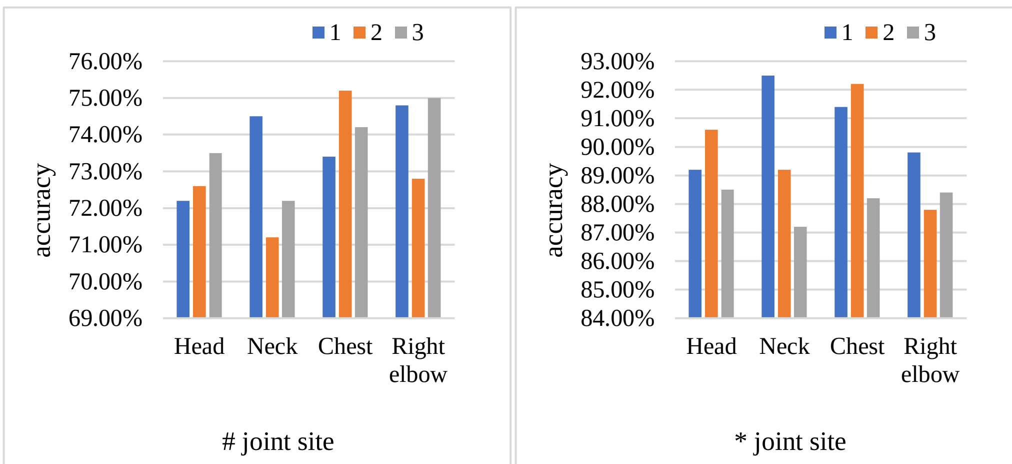

Figure 7.
Comparison of joint data accuracy.

From the two sets of comparison charts, it can be found that the accuracy of the training joint data based on computer vision has been greatly improved compared with the traditional training joint data. The accuracy of the joint data for traditional training was about 72%, but the accuracy of the joint data for training under computer vision was about 88%, and the accuracy was improved by 16%. It also shows that training under computer vision is more conducive to the scientific training of athletes.

The accuracy of the coach's judgment on the athlete is also an important reference in sports. It is also an injury to the athlete if the coach leaves the athlete with poor functioning. Coaches can also guide more effectively when they understand the athlete's body. Figure 8 shows a comparison chart of the accuracy of coaches' judgment on athletes.

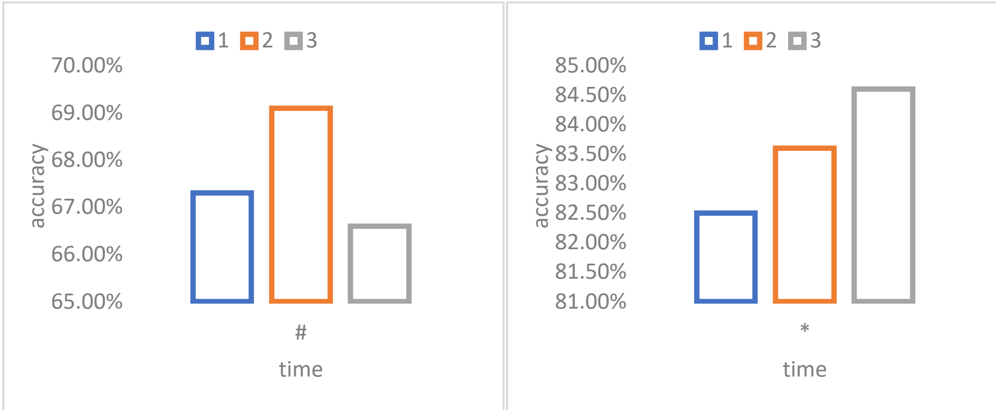

Figure 8.
Comparison of the accuracy of coaches' judgments.

Through two sets of data, it is found that compared with traditional training, the accuracy of the coach's judgment on the students is largely based on experience. The traditional training mode was only 67% accurate. The accuracy of the auxiliary system under computer vision has reached 82%, and the accuracy has increased by 15%, which also shows that the auxiliary training system based on computer vision is more conducive to training.

The accuracy of the training joints and the two main purposes of the coaches are to improve the performance of the players, which is the ultimate goal. The performances of the players' explosive power, balance, and sense of rhythm were compared, and the comparison chart is shown in Figure 9.

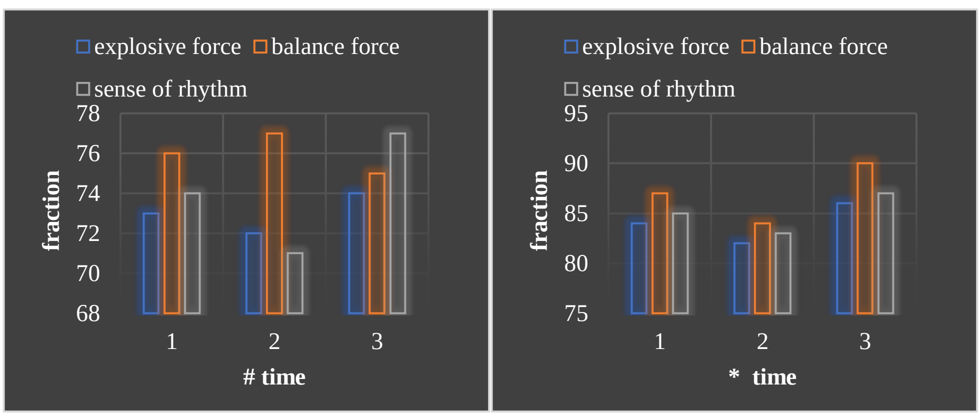

Figure 9.
Score comparison.

Through the results of the trainees, it was found that the trainees of traditional training generally scored 74 points. However, with the help of auxiliary training under computer vision, the student's score was 85 points, an increase of 11 points, and an increase of 14.8% in the score, which also shows the effectiveness of the system.

5. Conclusion

By adopting a new posture calculation method, the movement posture of the coach and the athlete can be adjusted very well, and suitable movement limb data can also be provided, which provides intuitive exercise analysis guidance for trainers. Through the description of the calculation method of human body pose, and on the basis of the above research, combined with the needs of the subject, a pose estimation method based on computational vision is proposed to realize the estimation of the two-dimensional pose parameters of the human body in the sequence pictures. Mainly through the accuracy of training joints, the accuracy of the coach's judgment on the athletes, and the students' performance are compared. The results show that the auxiliary training system based on computer vision is helpful to the training to a great extent, and it can also make the students train more scientifically. The inadequacy of the paper is that the matching method of computer vision is not explained, because of time and space reasons, and there is no detailed description of the human body pose estimation without the model, and it is simply taken. It is also hoped that further in-depth research by relevant scholars in the follow-up can also provide trainers with the best training methods and scientific training systems.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] X. Liu, Z. Lin, and Z. Lin, "An auxiliary system for rehabilitation training," in *Journal of Physics: Conference Series*, 2021, vol. 1920, no. 1, p. 012078, doi: https://doi.org/10.1088/1742-6596/1920/1/012078.

[2] S. Chen, X. Wu, and H. Yin, "KPCA method based on within-class auxiliary training samples and its application to pattern classification," *Pattern Analysis and Applications*, vol. 20, pp. 749-767, 2017.

[3] C. Li and J. Cui, "Intelligent sports training system based on artificial intelligence and big data," *Mobile Information Systems*, vol. 2021, no. 1, p. 9929650, 2021.

[4] D. Xiong, L. Yan, and P. Qiong, "An athletic training analysis system research based on physiological computation," *International Journal of Healthcare Information Systems and Informatics*, vol. 13, no. 2, pp. 54-67, 2018. https://doi.org/10.4018/IJHISI.2018040104

[5] P. Y. Lee and J. B. Moore, "Gauss-Newton-on-manifold for pose estimation," *Journal of industrial and management optimization*, vol. 1, no. 4, p. 565, 2005. https://doi.org/10.3934/jimo.2005.1.565

[6] M. Patacchiola and A. Cangelosi, "Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods," *Pattern Recognition*, vol. 71, pp. 132-143, 2017. https://doi.org/10.1016/j.patcog.2017.06.009

[7] S. Dinc, F. Fahimi, and R. Aygun, "Mirage: an O (n) time analytical solution to 3D camera pose estimation with multi-camera support," *Robotica*, vol. 35, no. 12, pp. 2278-2296, 2017. https://doi.org/10.1017/S0263574716000874

[8] K. Shirahata, Y. Coppens, T. Fukagai, Y. Tomita, and A. Ike, "GUNREAL: GPU-accelerated UNsupervised REinforcement and Auxiliary Learning," *International Journal of Networking and Computing*, vol. 8, no. 2, pp. 408-423, 2018. https://doi.org/10.15803/ijnc.8.2_408

[9] Y. Kawana and N. Ukita, "Occluded Appearance Modeling with Sample Weighting for Human Pose Estimation," *IEICE TRANSACTIONS on Information and Systems*, vol. 100, no. 10, pp. 2627-2634, 2017. https://doi.org/10.1587/transinf.2017EDP7088

[10] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," *The International journal of robotics research*, vol. 36, no. 2, pp. 142-149, 2017.

[11] J. Li, J. A. Besada, A. M. Bernardos, P. Tarrío, and J. R. Casar, "A novel system for object pose estimation using fused vision and inertial data," *Information Fusion*, vol. 33, pp. 15-28, 2017. https://doi.org/10.1016/j.inffus.2016.04.006

[12] D. Onoro-Rubio, R. J. Lopez-Sastre, C. Redondo-Cabrera, and P. Gil-Jimenez, "The challenge of simultaneous object detection and pose estimation: A comparative study," *Image and Vision Computing*, vol. 79, pp. 109-122, 2018. https://doi.org/10.1016/j.imavis.2018.09.013

[13] S. Urban, S. Wursthorn, J. Leitloff, and S. Hinz, "MultiCol bundle adjustment: a generic method for pose estimation, simultaneous self-calibration and reconstruction for arbitrary multi-camera systems," *International Journal of Computer Vision*, vol. 121, pp. 234-252, 2017. https://doi.org/10.1007/s11263-016-0935-0

[14] C. R. Upshaw, J. D. Rhodes, and M. E. Webber, "Modeling electric load and water consumption impacts from an integrated thermal energy and rainwater storage system for residential buildings in Texas," *Applied Energy*, vol. 186, pp. 492-508, 2017. https://doi.org/10.1016/j.apenergy.2016.02.130

[15] M. Hiroura *et al.*, "Construction of arteriovenous circulation system to gain of the close feeling to insert a catheter into human vessel," *Journal of Biorheology*, vol. 32, no. 2, pp. 56-64, 2018. https://doi.org/10.17106/jbr.32.56
[16] C. Laoudias, A. Nikitin, P. Karras, M. Youssef, and D. Zeinalipour-Yazti, "Indoor quality-of-position visual assessment using crowdsourced fingerprint maps," *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, vol. 7, no. 2, pp. 1-32, 2021. https://doi.org/10.1145/3433026
[17] N. H. Khan and A. Adnan, "Ego-motion estimation concepts, algorithms and challenges: an overview," *Multimedia Tools and Applications*, vol. 76, pp. 16581-16603, 2017. https://doi.org/10.1007/s11042-016-3939-4
[18] A. Liu, Y. Shi, P. Jing, J. Liu, and Y. Su, "Structured low-rank inverse-covariance estimation for visual sentiment distribution prediction," *Signal Processing*, vol. 152, pp. 206-216, 2018. https://doi.org/10.1016/j.sigpro.2018.06.001
[19] G. Wang, X. Chen, H. Guo, and C. Zhang, "Region ensemble network: Towards good practices for deep 3D hand pose estimation," *Journal of Visual Communication and Image Representation*, vol. 55, pp. 404-414, 2018. https://doi.org/10.1109/JSEN.2018.2873055
[20] F. Zheng and Y.-H. Liu, "SE (2)-constrained visual inertial fusion for ground vehicles," *IEEE Sensors Journal*, vol. 18, no. 23, pp. 9699-9707, 2018.