

Retraction

Retracted: Research on Human Pose Capture Based on the Deep Learning Algorithm

Journal of Control Science and Engineering

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Journal of Control Science and Engineering. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] W. Xie, X. Gu, X. Li et al., "Research on Human Pose Capture Based on the Deep Learning Algorithm," *Journal of Control Science and Engineering*, vol. 2022, Article ID 7215500, 5 pages, 2022.

Research Article

Research on Human Pose Capture Based on the Deep Learning Algorithm

Wenbin Xie ¹, Xueyun Gu ¹, Xiuming Li ², Luqi Zheng ¹, Guodong Li ¹,
Haisheng Lu ¹ and Huakang Li ¹

¹Beibu Gulf University, Qinzhou 535011, China

²Guangxi Research Institute of Mechanical Industry, Nanning 537000, China

Correspondence should be addressed to Wenbin Xie; 1351140114@xzyz.edu.cn

Received 15 May 2022; Revised 2 June 2022; Accepted 10 June 2022; Published 28 June 2022

Academic Editor: Jackrit Suthakorn

Copyright © 2022 Wenbin Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A method based on the deep learning algorithm is proposed to accurately capture the posture of the human body. It is one of the important means to improve athletes' competitive level in modern sports to accurately analyze the posture of sports training by technical means. Aiming at the application demand of using artificial intelligence technology to accurately analyze and predict the motion training posture, a motion posture analysis and prediction system based on deep learning is designed in this paper. Based on the Arduino embedded development board and equipped with multiple IMU sensors, the scheme established a system to collect accurate human movement data such as speed and acceleration by using stepper motors and obtained accurate human movement data. The experimental results show that these models have been trained with H3.6 m data sets. The sampling frequency was reduced to 25 Hz, and the joint angles were converted into exponential graphs. When the time window covers approximately 1 660 ms, the loop network will be initialized to 40 frames, equivalent to 1 600 ms. For each action, a separate pretrained recursive model is used. It is proved that the method based on deep learning can reduce the prediction error of fine-tuning specific movements and effectively classify and predict the movements not included in the original training data.

1. Introduction

The development of human pose estimation has been more and more close to reality, such as gait analysis, human-computer interaction, video monitoring, and other fields, and human pose estimation has a broad application prospect. Current mainstream human pose estimation algorithms can be divided into traditional methods and deep learning-based methods [1]. The traditional method is to design a 2D human body part detector based on the graph structure and deformation part model, establish the connectivity of each part using graph model, and estimate human body pose by optimizing the graph structure model based on the relative constraints of human kinematics. Although the traditional method has a high time efficiency, the extracted features are mainly artificially set HOG and SHIFT features, which cannot make full use of the image information, resulting in the algorithm being subject to the

different appearance, perspective, occlusion, and inherent geometric ambiguity in the image. The human pose estimation method based on deep learning mainly uses the convolutional neural network (CNN) to extract human pose features from images. Compared with the traditional method of artificial design features, the CNN can not only obtain features with richer semantic information, furthermore, multiscale and multitype human node feature vectors and the entire contextual of each feature can be acquired under different sensory fields, and the dependence on the structural design of the component model can be eliminated. Then, coordinate regression of these feature vectors to reflect the current attitude can be applied to the specific situation [2]. Different from the traditional method of explicitly designing feature extractors and local detectors, it is easier to construct the CNN during deep learning. Meanwhile, CNN models dealing with sequence problems can be designed, such as the recurrent neural network (RNN). By analyzing

continuous multiframe images, the changing rules of human body posture can be obtained. Thus, a more accurate topological structure was established for each node in the human body posture [3].

Wang et al. Before deep learning was applied to human pose estimation, most of the methods based on the graphic structure were used to deal with human pose. Therefore, a deformable component model based on graphic structure method emerges. These methods require local detectors and can only model a subset of all connections between human nodes. Although the efficiency is relatively high, it is greatly affected by the factors such as the figure blocking the shooting Angle and the illumination of the image, so the representation ability is limited. But at the characters' shooting angle, the image is influenced by factors such as illumination, and the said ability is limited. In addition, this traditional method relies on manually setting features in feature extraction, such as directional gradient histogram, scale-invariant feature transformation, and edge feature, which requires a lot of labor, time, and energy [4]. Li et al. With the development of artificial intelligence technology, many deep learning models have been proposed, such as the convolutional neural network, generative adversarial network, autoencoder, and recursive neural network. These models have achieved superior results in the field of image processing than traditional nondeep learning methods. Significant achievements have been made in image segmentation, target detection, image recognition, and other fields [5]. Yang et al. proposed a large number of human pose estimation methods and provided some public data sets of human pose estimation. Therefore, human pose estimation based on deep learning has become the main research direction [6]. Based on this, an algorithm based on deep learning is still proposed in this paper, and combined with mixed coding, it can effectively reduce the prediction error of fine-tuning specific actions and can effectively classify and predict the actions not included in the original training data.

2. Exercise Data Acquisition Process

To obtain accurate human movement data, a set of the human movement data acquisition system based on the IMU core sensor is designed. DMP is a unique hardware feature of IMU devices that can compute quaternions from sensor readings. The IMU takes data directly from the secondary sensor, allowing the embedded processor to process the sensor fusion data without the intervention of the system application processor. A MPU6050IMU monitoring program is developed based on the processing language. The original data obtained from the accelerometer and gyroscope were fused by DMP, and Euler angle information was extracted from quaternion representation to calculate the yaw, pitch, and roll motion of IMU. Software first obtains the orientation value from XBee Pro, and then it calculates the relative position of IMU using the length and orientation value of each part of the human body. Two IMU sensors are used to measure the motion of human joints. The original acceleration data from IMU contain a lot of unfiltered noise

and are prone to significant errors due to fluctuations in short-term measurement. Therefore, the joint motion data are not directly derived from the integration of IMU acceleration. An IMU relative motion data processing algorithm based on human motion analysis is developed which is based on the shape of human joints [7].

2.1. Acquisition of Angular Velocity Data. We are required to use the stepper motor to ensure accuracy of IMU data. The stepper motor NEMA 23 YH57BYGH56-401A specifications are as follows: the stepper angle is 1.8° , and the control accuracy is $\pm 5\%$; the rated current is 2.8 A; the phase resistance is 0.9Ω , and the control accuracy is $\pm 10\%$. The phase inductance is 2.5 MH, and the control precision is $\pm 10\%$. The holding torque is 1.2 Nm. We are required to maintain the minimal axial radial and axial clearance, and a rotor disk less than 450 g is connected to the stepper motor. When verifying the angular velocity data measured by IMU, the stepper motor with relatively high accuracy is used for the experiment, which is more controllable and reliable. The IMU was installed on the rotor disk for testing, and the Arduino UnoR3 was used to control the speed and the number of steps the stepper moved during the test [8].

2.2. Acquisition of Motion Direction and Position Data. Human motion analysis equipment has been developed to analyze human joint movements. Therefore, it is necessary to test its accuracy for measuring human joint movements. The human motion analysis device is worn on the shoulder, and two IMUs are placed on the upper and lower arms. During the test, the subjects were asked to lower their hands and repeatedly bend their elbows five times; then, they were asked to straighten the hand to its original position in the plane. At the same time of object movement, human motion analysis equipment was used to collect motion data, and GoProHero3 equipment was used to record a high resolution video with an effective photo resolution of 12 M and a pixel and frame rate of 47 fps on the 2 D plane. Finally, the video is postprocessed. Analysis was performed using motion analysis software MaxTRAQ2D to analyze the elbow movement. MaxTRAQ2D is video-based motion tracking software for extracting kinematic characteristics from standard AVI video files. Through manual and automatic tracking, users can view angles and distances between the points frame by frame [9].

2.3. Graphical Display Platform. With the help of the ArduinoUnoR3 program developed, the computer can receive yaw angle, pitch angle, and roll angle values from the two IMUs, and we need to use the obtained values to simulate the direction of the IMUs in real time and design software to display them in real time. In the software package, cuboids are used to represent IMU, where the different sides of cuboids have different colors so that users can distinguish directions. Table 1 shows the side colors corresponding to the IMU axis.

The software package represents the IMU as two cubes in the program and uses the limb structure of the human body

TABLE 1: Corresponding relationship between the color and the IMU axis.

Rectangle box color	Innovative marketing axial
cyan	Positive X
violet	Negative X
red	Positive Y
Dark blue	Negative Y
green	Positive Z
yellow	Negative Z

to show the movement of human joints. The left cuboid is used to simulate the orientation of the first IMU, and the right cuboid is used to simulate the orientation of the second IMU. Yaw, pitch, and roll values are displayed in the upper part of the software package to accurately track the direction of the IMU [10]. To display the motion of IMU and human joints in 2D mode, the software package was improved to more accurately represent the motion of various parts of the human body and joints on all axes. The relative motion between IMU sensors is calculated using the direction of two IMUs.

3. Motion Analysis Software Design

In the first section, accurate velocity and angular velocity information are obtained based on the embedded system in order to accurately identify the law of motion data. This section establishes a model based on time coding to accurately identify human movement patterns. The following three variants of the recognition model are used: symmetric coding S-TE, time scale coding C-TE, and structural coding H-TE.

3.1. Data Processing and Presentation. The MoCap skeleton in the Cartesian space is selected; that is, the frame at time t consists of $f_t = [f_{x,i,t}, f_{y,i,t}, f_{z,i,t}]_{i=1:N_{\text{joints}}}$, where N_{joints} is the number of joints. To standardize the model, the joint angles are converted to Cartesian coordinates of the standardized mannequin. The joint position is centered at the origin of the coordinate system, while preserving the global rotation of the bone, ignoring the translation. Data sets connected to the matrix $F_{t:(t+\Delta t-1)} = [f_t, f_{t+1}, \dots, f_{t+\Delta t-1}]$ can be determined within Δt time windows. This data set consists of the input frame window $F_{(t-\Delta t+1):t}$ and output frame window $F_{(t+1):(t+\Delta t)}$ of each time step $t \in [\Delta t, (T - \Delta t - 1)]$, where T is the sampling time length [11].

3.2. Time Encoder. The coding-decoding framework is used to calculate the projection of high-dimensional input data onto low-dimensional graphics and predict the output data based on this projection. The high-dimensional input data $x \in R^N$ is optimized by the autoencoder, as shown in equation (1).

$$\min_{f,g} \|x - f(g(x))\|. \quad (1)$$

Among them, the encoder $y = g(x)$ maps input data to the low-dimensional space $y = R^M$ and $N > M$, while the decoder $\hat{x} = f(y)$ maps back to the input space $\hat{x} \in R^N$, and functions f and g are represented by a symmetric multilayer perceptron. An alternative approach is used in the system to capture temporal correlations of human movement data rather than static representations of the human posture. We assume $x \in R^N$ to be the observed value of time t , and the optimization function of the time encoder is shown in equation (2).

$$\min_{f,g} \|X_{(t+1):(t+\Delta t)} - f(g(X_{(t-\Delta t+1):t}))\|. \quad (2)$$

Here, the encoder $y = g(X_{(t-\Delta t+1):t})$ maps input data to the low-dimensional space $y = R^M$ ($N \times \Delta t > M$) and the decoder $\hat{X}_{(t+1):(t+\Delta t)} = f(y) \in R^{N \times \Delta t}$ for mapping back to the data space.

In this application, the size of the input and output matrices is $3 \times N_{\text{joints}} \times \Delta t$, and the encoder $y = g(F_{(t-\Delta t+1):t})$ maps the input data to the low-dimensional space $y = R^M$ ($3 \times N_{\text{joints}} \times \Delta t > M$), and the decoder $\hat{F}_{(t+1):(t+\Delta t)} = f(y) \in R^{3 \times N_{\text{joints}} \times \Delta t}$ maps back to the data space.

4. Experimental Results and Analysis

We need to develop a program for Arduino to take signals from the DMP of the InvenSense MPU6050 IMU and send them to the computer via XBeePro wireless serial communication. The embedded DMP is located in the IMU and can divert the calculation of the motion processing algorithm from the host processor. DMP captures data from accelerometers and gyroscopes and provides the integrated motion fusion output. To display and plot the data received by XBee, a computer program based on a processing language is developed to read the data using a serial communication port. The subject was placed among four Kinect cameras, and the trunk movement of the subject was recorded using customized software. A set of 3D point cloud information can be obtained by stitching together each frame image of the four cameras using the transformation matrix from the calibration step. By comparing the point cloud at each frame, the movement change of the person under test can be accurately deduced from it. To deduce the movement changes, geomagic Studio2012 computes the set macro data created. In summary, the recording and deducting steps of experimental data are as follows: loading each point cloud exported from customized software, constructing 3D grid, filling the grid, and smoothing the grid with grid diagnostic tools. We need to ensure the accurate analysis of body movement, and the analysis scope should be limited; that is, a boundary box should be placed around the initial image to limit the volume calculation to the field of the human torso.

The database obtained in the experiment contains 2,235 records from 144 different subjects performing a variety of complex actions. Since many of the records were sampled at 120 Hz and others were sampled at 60 Hz, the previous test sampling was reduced to 60 Hz. For evaluation, a

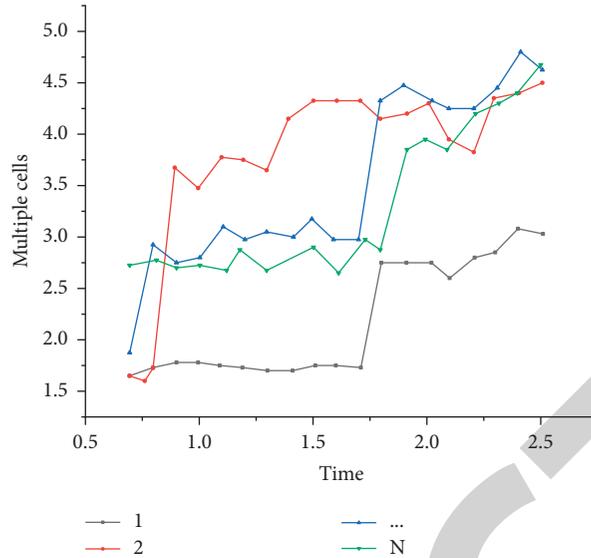


FIGURE 1: Average posture of multiple units.

TABLE 2: Comparison of classification errors.

DS	Method	80 ms	160 ms	320 ms	560 ms	1000 ms	1600 ms
H3.6 M	ERD	0.39	0.44	0.53	0.6	0.67	0.7
	S-RNN	0.34	0.38	0.45	0.52	0.59	0.63
	LSTM3L	0.27	0.33	0.42	0.49	0.57	0.62
	S-TE	0.36	0.37	0.37	0.38	0.43	0.45
	C-TE	0.21	0.23	0.27	0.34	0.42	0.45
	H-TE	0.21	0.22	0.25	0.3	0.36	0.39
CMU	S-TE	0.3	0.31	0.33	0.35	0.37	0.37
	C-TE	0.18	0.21	0.25	0.3	0.33	0.35
	H-TE	0.18	0.12	0.24	0.28	0.31	0.33

preprocessed H3.6 m dataset was used to train the current model with a time window of 100 frames and 1 660 ms. Human pose estimation based on deep learning relies on a large number of data sets to train the model. The larger the sample size and the more diverse the data, the more beneficial it is to develop a robust human pose estimation model.

4.1. Evaluation Indicators. Different data sets have different characteristics and different task requirements. The commonly used two-dimensional human pose estimation mainly includes the following: the percentage of correct parts (PCP) is an evaluation index of early attitude estimation, which is used to evaluate the positioning accuracy of a limb. If the two ends of a limb are within the threshold of the corresponding truth endpoint, the limb is correctly positioned. The percentage of correct keypoints (PCK) is used to evaluate the accuracy of human body keypoint positioning. If the candidate node falls within the threshold pixel of the real node, the candidate node is correct. The key is flat.

Average precision of keypoints (APK). After the predicted attitude is assigned to the true attitude by PCK evaluation, the average accuracy of positioning accuracy of

each node is obtained by APK. Object keypoint similarity (OKS) is an evaluation index of multiperson pose estimation. The similarity between the truth value and the predicted human body points is calculated.

4.2. Feature Visualization. Figure 1 shows the average postures of multiple units in the excited H-TE interlayer. To reduce noise, posture and network activity are considered only when the s-cell output exceeds 0.8.

4.3. Action Classification Effect Display. The whole movement sequence is classified, rather than a single movement sequence. The prediction ability of the three models (S-TE, C-TE, and H-TE) was compared with the recently proposed ERD classification prediction algorithm, as shown in Table 2 and Figure 2.

These models were trained with h3.6 m datasets. The sampling frequency was reduced to 25 Hz, and the joint angles were converted into exponential graphs. When the time window covers approximately 1 660 ms, the loop network will be initialized to 40 frames, equivalent to 1 600 ms. For each action, a separate pretrained recursive

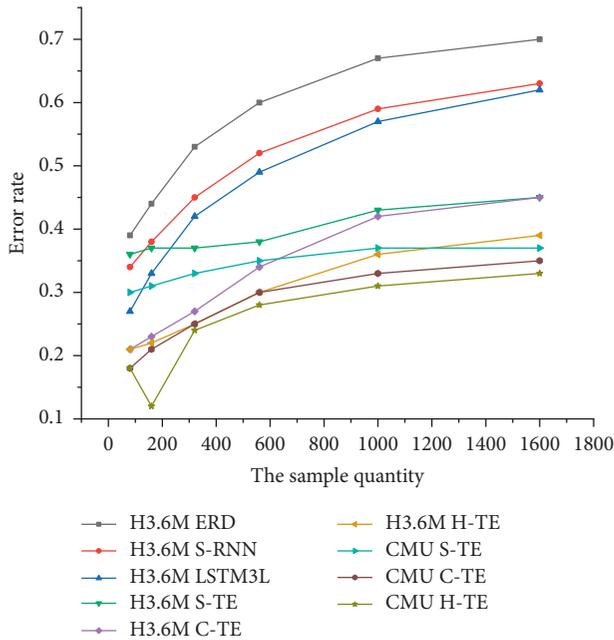


FIGURE 2: Error data graph.

model is used. Although LSTM3L performs better than some of the models in terms of initial prediction, the time encoder C-TE performs better at 160 ms or longer predictions. Because human motion is a complex nonstationary motion, it is difficult for the circular network to make short-term prediction, but this model can infer the future prediction framework. In most predictions, the symmetric time encoder S-Te and the convolution time encoder Ci-Te are superior to the hierarchical time encoder H-Te, indicating that the structural priors are beneficial to motion prediction. The mixed coding method can reduce the prediction error of fine-tuning specific movements and classify and predict the movements not included in the original training data effectively.

5. Conclusions

The neural network based on deep learning has a large amount of computation, and compared with the classification task and detection task, the human pose detection requires a higher resolution output feature map, which will greatly increase the computation of the algorithm. Improving the network, it often meets the problem of improving the accuracy, but by increasing the amount of calculation and reducing the time efficiency, a lightweight network optimization attitude estimation algorithm can be adopted. The lightweight network is combined with attitude estimation, which can simplify the attitude estimation network and improve the time efficiency while ensuring the algorithm accuracy.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Undergraduate Innovation and Entrepreneurship Training Program (no. S202111607103) and Research Foundation for Advanced Talents (no. 2019KYQD06) of Beibu Gulf University.

References

- [1] Y. Q. Chen, J. L. Guo, H. Yang, Z. Q. Wang, and H. L. Liu, "Research on navigation of bidirectional a* algorithm based on ant colony algorithm," *The Journal of Supercomputing*, vol. 77, no. 2, pp. 1958–1975, 2021.
- [2] S. Jiang, P. Lin, Y. Chen, and C. Tian, "Research on heterogeneous fiber grating spectrum separation based on transfer matrix and adaptive algorithm," *Optical Engineering*, vol. 58, no. 9, p. 1, 2019.
- [3] J. Virgili-Llop, C. Zagaris, R. Zappulla, A. Bradstreet, and M. Romano, "A convex-programming-based guidance algorithm to capture a tumbling object on orbit using a spacecraft equipped with a robotic manipulator," *The International Journal of Robotics Research*, vol. 38, no. 1, pp. 40–72, 2019.
- [4] Y. B. Wang, X. R. Liu, Z. M. Jing, and Y. F. Liu, "Research on optical code assignment algorithm based on pce domain," *Optical Fiber Technology*, vol. 50, pp. 302–308, 2019.
- [5] L. Li, T. Martin, and X. Xu, "A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders," *Applied Ergonomics*, vol. 87, no. 4, Article ID 103138, 2020.
- [6] L. Yang, H. Wang, B. Huo, F. Li, and Y. Liu, "An automatic welding defect location algorithm based on deep learning," *NDT & E International*, vol. 120, no. 7, Article ID 102435, 2021.
- [7] G. Li, F. Liu, A. Sharma et al., "Research on the natural language recognition method based on cluster analysis using neural network," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9982305, 13 pages, 2021.
- [8] J. Jayakumar, B. Nagaraj, P. Ajay, and P. Ajay, "Conceptual implementation of artificial intelligent based E-mobility controller in smart city environment," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5325116, 8 pages, 2021.
- [9] X. Liu, J. Liu, J. Chen, and F. Zhong, "Degradation of benzene, toluene, and xylene with high gaseous hourly space velocity by double dielectric barrier discharge combined with Mn₃O₄/activated carbon fibers," *Journal of Physics D: Applied Physics*, vol. 55, no. 12, Article ID 125206, 2022.
- [10] R. Huang, S. Zhang, W. Zhang, and X. Yang, "Progress of zinc oxide-based nanocomposites in the textile industry," *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 3, pp. 281–289, 2021.
- [11] H. Xie, Y. Wang, Z. Gao, B. P. Ganthia, and C. V. Truong, "Research on frequency parameter detection of frequency shifted track circuit based on nonlinear algorithm," *Nonlinear Engineering*, vol. 10, no. 1, pp. 592–599, 2021.