IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 1, 2021, accepted June 16, 2021, date of publication June 22, 2021, date of current version June 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091430

Human Motion Target Posture Detection Algorithm Using Semi-Supervised Learning in Internet of Things

LEI CHEN [1] AND SHANGBIN LI [2]
[1]Physical Education Department, Shenyang Ligong University, Shenyang 110159, China
[2]Physical Education Department, Harbin Engineering University, Harbin 150001, China
Corresponding author: Shangbin Li (sports@hrbeu.edu.cn)

This work was supported by the Scientific Research Funding of the Liaoning Provincial Department of Education under Grant LG201905.

ABSTRACT To address the problem that the traditional human motion attitude detection process is easy to ignore the data calibration, which leads to the problems of long running time, low accuracy and poor detection effect, a human motion target attitude detection algorithm based on semi-supervised learning in the Internet of things environment is proposed. Firstly, human motion target images are collected using the Internet of things (IoT), human motion attitude features are extracted based on the eight-star model, and multi-features are fused to form image blocks of 17-dimensional feature vectors. Then, random fern classifiers are optimized and semi-supervised learning is used to calculate a large number of uncalibrated data in time domain, spatial domain and data. The classifier is trained to complete image block classification. Finally, the classifier parameters are updated iteratively to complete the attitude detection of human motion target. The results show that the proposed algorithm has high accuracy in human motion attitude extraction and multi-feature fusion, and has a high correct classification rate for different feature poses, as high as 92.5%. The average F value of human motion attitude detection is 0.95, the overlap ratio is high and the time is short. The overall performance is good.

INDEX TERMS Semi-supervised learning, human motion posture, extraction, multi-feature fusion, classifier, detection.

I. INTRODUCTION

Human motion posture detection is mainly to describe the information about human motion, grasp the content expressed by human body and then further detect human behavior, which is highly practicable [1]. With the improvement of the people's quality of life, simple video monitoring can no longer meet the needs [2], [3]. So it is of great significance to find an efficient way to detect the human motion posture for response to various emergencies, especially for public places. Human motion posture detection is also widely used in sports, medical and other fields, and is the focus of the current research on artificial intelligence [4], [5].

The Internet of Things is a combination of radio frequency identification technology, sensor technology and artificial intelligence technology. The network environment created by this technology can well perceive the real world.

In the Internet of Things environment, the human body motion target posture is detected, the human target motion process is monitored in real time through the information network, and the motion posture information is shared, which can effectively obtain human body motion information, thereby realizing intelligent motion posture detection. However, due to the variety of environment types, there is a huge amount of information in human motion [6], and together with the interference from other external factors, it is difficult to precisely detect the human motion target posture in practice. Therefore, it is a point research issue to find an effective algorithm of human posture detection in the computer filed.

Machine learning is a hot technology emerging in recent years [7] and [8]. With strong data processing capability, it is quite popular in the field of human motion target analysis. Semi-supervised learning is a typical machine learning method, which is produced from the combination of supervised learning and unsupervised learning. With it data

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Khurram Khan.

processing is more accurate, operation is convenient and manpower can be saved, so it becomes more and more popular [9], [10]. Based on multimodal human motion data, Chao *et al.* [11]. constructed a human motion feature map and put forward a multi view semi-supervised learning framework, which was used to complete human movement identification. However, this method is time-consuming. Tao *et al.* [12] proposed a semi-supervised classification framework based on local and global consistency. This framework was used to process the reconstructed data graph structure, effectively improving the accuracy of semi-supervised learning, but the accuracy is not high. Yanchao *et al.* [13] analyzed the limitations of the existing active learning methods, proposed an active learning method based on deep learning, and used deep neural network to fully establish the relation between labeled samples and unlabeled samples. They obtained a good research result, but the analysis on image features was inadequate. Lan *et al.* [14] applied a small number of slice-level labeled samples and a large number of image-level labeled samples to semi-supervised learning training and completed the image target detection in combination of convolutional neural network, but the overall effect was unsatisfactory.

By virtue of the advantages of semi-supervised learning, this paper puts forwards the detection algorithm of human motion target posture based on semi-supervised learning in the environment of Internet of Things, expecting to efficiently complete human posture detection by fully using semi-supervised learning training. In order to improve the quality of human motion posture image information, this paper first extracts the motion posture features and fuses the multiple features. Semi-supervised learning can calculate a large number of unlabeled data, and with this advantage, the poor calculation effect caused by less labeling of the actual video and image data can be avoided. Meanwhile, this paper also designs a classifier and repeats training to complete the design of human motion posture detection algorithm. The experimental results show that the proposed algorithm has a superior performance and provides a reference for the human target research in the future. The main contributions of this paper are as follows:

(1) This paper fuses and processes the multiple features and forms image maps, providing an effective foundation for the following algorithm calculation;

(2) The training classifier based on semi-supervised learning gives full play of data processing and solves the problems of heavy computation burden and long time consuming of unlabeled data;

(3) This paper innovatively considers the three restrictions of time domain, space domain and data in the algorithm design, which plays crucial role in the training effect of classifier;

(4) Based on a large number of data, various types of human motion postures, this paper designs multiple verification indexes and tests the algorithm with them one by one, improving the credibility of the experimental results.

II. RELATED WORK

There are many researches on human target detection at home and abroad. In order to accurately obtain the human movement information in video, literature [15] first extracted the information at joints according to human body image, described posture changes, constructed the posture space-time feature descriptors, then encoded the descriptors respectively, and finally realized the human behavior recognition through weighted fusion. This method is time-consuming. Taking human joints as vectors, literature [16] built a set of human posture tracking system based on dual Kinect sensors and completed the human tracking with unscented Kalman filter. But the tracking is slow, which reduces the effectiveness of the algorithm. Literature [17] constructed a human feature matrix with skeleton, posture, joint position as feature points, and a comprehensive matrix by fusing the features of human interacting with objects, and finally completed the 3D complex human body recognition. The accuracy of this method needs to be improved. Literature [18] focused on the analysis of human behavior according to the changes of human joints and posture and used the weighted calculation method to extract the joints and postures with rich information. It gave time series modeling analysis, adopted support vector machine to process joint and posture information and completed recognition of human skeleton behavior. But the calculation process of this method is complex.

There are also many studies abroad. Literature [19] used convolutional neural network and long-term and short-term memory network to construct a deep network, learnt the long-term dependence of human behavior recognition under the multi perspective framework from video and proposed an image clustering method based on superimposed convolution automatic encoder. This generates clustering tags for auto-tagging and improves the recognition effectiveness of human behavior, but the feature fusion is insufficient. Literature [20] discussed a parallel multi-layer depth recognition architecture, which improved the human body feature extraction capacity of the algorithm by using the hierarchy function of deep learning network, but the integrity is not good. Literature [21] applied local sparse segmentation to global clustering before the approximate rank pool to get semantic images, summarized the human motion features in single or multiple images, and used short-term memory network for human movement recognition to complete the network construction of semantic images. The accuracy of this method is not high.

Aiming at the problems of the above-mentioned research, this paper designs the human body moving target pose detection algorithm in the Internet of Things environment, introduces semi-supervised learning, calculates a large amount of unlabeled data, and adds the three-layer restriction conditions of time domain, space domain and data, and efficient training The classifier improves the accuracy of algorithm detection.The results show that the proposed algorithm has high accuracy in feature extraction and multi-feature fusion,

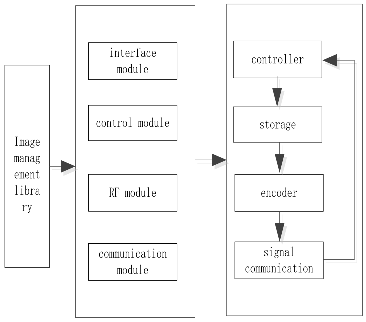

FIGURE 1. RFID human target recognition framework.

the correct classification rate is as high as 92.5%, the F value of human motion posture detection is high, the overlap ratio between the algorithm output and the calibration range is as high as 95%, and the average running time is as low as 1.1s, compared with other literature algorithms, the proposed algorithm has the advantage of being more efficient and accurate.

III. HUMAN MOTION POSTURE DESCRIPTION OPERATORS BASED ON FEATURE ANALYSIS

Computer vision detection technology is now the main way to realize the analysis of human motion posture. Visual detection obtains the human trajectory or contour information based on video and image, having a good analysis of the details of human body. Internet of Things (IoT) is an extension of computer technology, which uses wireless sensor network to connect and transmit information. It can collect human motion images with sensor and capture human motion target, to realize the intelligent human target recognition and detection [22], [23]. Therefore, this paper uses IoT to collect human motion target images and extract and fuse posture features, to provide accurate basis for the following human posture detection.

A. COLLECTION OF HUMAN TARGET IMAGES BASED ON IoT

Ratio Frequency Identification (RFID), is a typical non-contact IoT sensing technology [24], [25]. It can read a large number of data through wireless communication with long-distance identification and fast speed. It can be used for multi-target identification and can efficiently collect human motion target images [26], [27] to design RFID identification framework. It is shown in Figure1.

B. EXTRACTION OF HUMAN MOTION POSTURE FEATURES

After the human motion target images are collected by IoT, the human motion posture features [28], [29] are extracted based on eight-star model according to the stretching and amplitude of human motion posture. The establishment eight-star model is established based on the human motion target images which need to have significant contour features. With the head and limbs as the main parts, the centroid is selected to establish the eight-star model [30], [31].

Set the centroid coordinates of the human motion target (x_0, y_0), which can be expressed as:

$$x_0 = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$y_0 = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{2}$$

where, x_i and y_i represent the abscissa and vertical coordinate points respectively, and N represents the number of pixels.

On the basis of the collected human motion target contour, select the contour extreme points up, down, left and right, and calculate the Euclidean distance [32], [33] from each extreme point to the centroid, which can represent the features of human motion posture. The formula is as follows:

$$d = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \tag{3}$$

Considering that there are two contour boundary points in each of the four directions of human body, the human motion target has 8 contour boundary points, and the value range of i in the above formula is less than or equal to 8.

C. MULTIPLE FEATURES FUSION

In order to eliminate feature redundancy and effectively connect feature vectors, the human motion posture features are fused to improve the detection performance of the algorithm. Connect each extreme point with the centroid and form angles with the horizontal line as the boundary. 8 angle values can be obtained, and the formula is as follows:

$$\theta = \frac{180}{\pi} \arccos \frac{|x_i - x_0|}{d} \tag{4}$$

Assuming that the human target image is an ellipse, the eccentricity f can be used to represent the amplitude of human motion posture:

$$f = \sqrt{1 - \left(\frac{j}{k}\right)^2} \tag{5}$$

where, j and k represent the length of the major and minor axes of the ellipse respectively.

Based on the above analysis, the eight star model M is constructed, and the description operator is used to represent the human motion posture:

$$M = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, f) \tag{6}$$

where, The eight-star model is a 17-dimensional feature vector constructed by fusing the distance feature, angle feature and eccentricity feature of the four contour extreme points of the upper, lower, left and right part of human body to form image maps. The description operator is not so complex in computation and can help the algorithm to complete the posture detection quickly.

IV. DETECTION ALGORITHM OF HUMAN MOTION TARGET POSTURE BASED ON SEMI-SUPERVISED LEARNING

A. CLASSIFIER DESIGN

Image maps are formed after the fusion of the multiple features of human motion target posture. In order to complete the detection of various types of human motion postures, it is necessary to construct a classifier. In this section, we use a random fern classifier to classify and detect the image maps.

The image map of human motion posture is represented as A, the category set as Z, and the binary feature set of human motion posture that needs to be classified as $D\,(D_1, D_2, \ldots D_i, \ldots, D_R)$. And the binary feature set D_1 is closely related to the pixel values of the image maps at different positions [34], [35], which can be expressed as:

$$D_i = \begin{cases} 1, & \text{if } A_a \geq A_b \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where, a and b represent the positions of the image maps.

When binary features are used for calculation, it needs a lot of data feature values to guarantee the accuracy. In order to reduce the memory required by the algorithm, the classifier design is optimized and independent assumption is introduced to reduce the space required by storage, that is, to assume that the features are independent to each other [36], [37]. In this case, the correlation between features is completely ignored. To reasonably solve the problem, the features are classified as R/2 groups which are defined as ferns. Calculate the joint probability distribution among ferns:

$$P(D_i\,|Z) = \prod_{l=1}^{R/2} P(S_l\,|Z) \tag{8}$$

where, S_l represents the l-th fern. $P(.)$ is the probability function.

The estimate value of similarity of different types of human motion posture, namely the similarity of different ferns is calculated as follows:

$$P(Z\,|D_1, D_2, \ldots, D_R) = \frac{P\,(D_1, D_2, \ldots, D_R\,|Z)\,P\,(Z)}{P\,(D_1, D_2, \ldots, D_R)} \tag{9}$$

It can be seen from the above formula that the denominator does not depend on types, the following simplified formula can be used to solve the maximum likelihood estimate value $\hat{z}$ [38], [39], thus to complete the classifier design:

$$\hat{z} = \arg\max P\,(D_1, D_2, \ldots, D_R\,|Z) \tag{10}$$

B. TRAINING CLASSIFIER BASED ON SEMI-SUPERVISED LEARNING UNDER RESTRICTED CONDITIONS

In practice, the analysis data of human motion posture is generally unlabeled data, which cannot well reflect image features, but it needs a large amount work to manually label images sequence. Therefore, the introduction of semi-supervised learning is a kind of algorithm between supervised learning and unsupervised learning. There are some advantages in algorithm accuracy to use part of labeled data and a large number of unlabeled data for calculation [40], [41].

P-N Learning is now a typical semi-supervised learning algorithm. This method introduces the three restrictions of time domain, space domain and data, restricts the process of unlabeled data processing [42], [43] and constructs a semi-supervised learning framework to complete the classifier training and improve the classification performance.

1) TIME DOMAIN RESTRICTION

The time domain restriction is mainly used to ensure the correctness of human motion trajectory and collect the human motion track points [44]. The track points are collected in the forward direction to obtain the set of track points at the time t, that is, the forward human motion trajectory can be expressed as G_f^k:

$$G_f^k = (x_1, x_2, \ldots, x_t) \tag{11}$$

Conversely, the track points are connected backward to obtain the backward human motion trajectory at time t, which can be expressed as G_b^k:

$$G_b^k = \left(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_t\right) \tag{12}$$

The time domain trajectory reliability Q is used to express the time domain restriction, and the formula of Q is as follows:

$$Q = dis\left(G_f^k, G_b^k\right) \tag{13}$$

where, $dis\,(\cdot)$ is a distance function, which represents the distance between the forward trajectory G_f^k and backward trajectory G_b^k of human motion. The Euclidean distance is used to calculate between the start point of the forward trajectory and the end point of the backward trajectory, and the distance function can be accurately expressed as:

$$dis\left(G_f^k, G_b^k\right) = \left\| c_t - \hat{c}_t \right\| \tag{14}$$

where, c_t represents the forward track points at the center position at time t, and $\hat{c}_t$ is the backward tack points at the center position at time t.

2) SPACE DOMAIN RESTRICTION

The space domain restriction is proposed to meet the assumption of continuous motion of the target. A series of track points will be produced during human motion. Assuming that the target is moving continuously, samples are extracted near the motion trajectory as positive samples, and those deviating from the motion trajectory as negative samples [45], [46].

3) DATA RESTRICTION

Data restriction is mainly used to accurately label unlabeled samples, so as to improve the accuracy of classifier. The normalized cross correlation technique is generally adopted to measure the similarity between the feature vectors of unlabeled samples and training sample set, so as to avoid too many misclassification problems and improve classification accuracy. Assuming that the unlabeled sample is w_0, and the training sample set is w_1, the formula to calculate their correlation is as follows: (15), as shown at the bottom of the page, where, A_1 and A_2 represent the pixel values of human motion pose image blocks at different positions,a and b represent two position points on the image block. $\overline{A}_1(a_1, b_1)$ represents the average pixel value at point a1 and point b1 on the image map $\bar{A}_2(a_2, b_2)$ represents the average pixel value at point a2 and point b2 on the image mapand χ is the deviation coefficient.

Under the three restrictions, semi-supervised learning is used to train the classifier, and the process is as follows:

Step 1: initialize the posture data of the human motion target, set the classification parameters, and use the labeled data to initially train the classifier;

Step 2: give the sample feature space K and the motion posture category set J;

Step 3: respectively label the unlabeled samples and conduct classification and verification under the three restrictions to test whether they meet the classification assumptions. If yes, add the training sample set, and if no, label and verify the data again.

Step 4: use the constantly expanding training sample set to train the classifier repeatedly until the convergence is completed.

C. DETECTION ALGORITHM DESIGN OF HUMAN MOTION TARGET POSTURE

The process of the detection algorithm of human motion target posture based on semi-supervised learning is as follows:

Input: the image sequence of human motion target posture acquired through IoT.

Output: the detection results of human motion target posture.

Initialize the posture contour of human motion target and detect the human motion target posture based on semi-supervised learning, and the steps are:

1) Set the parameter values according to the position and scale of the first image, and give the time and space range of the pixel points;
2) Scan the human motion image with the given range to form a set of human motion postures;
3) In the set of human motion postures, determine the positive and negative samples according to the value of eccentricity;
4) Select the mode according to binary features and get the optimal feature mode as the image position to be compared;
5) In the optimal feature mode, train the classifier based on semi-supervised learning;
6) Iteratively update the parameters of the classifier until convergence to complete the human motion target posture detection.
7) End.

Based on the above analysis, the algorithm flow is used to show the process of detecting the posture of human moving targets. It is shown in Figure2.

TABLE 1. Experimental environment parameters.

Parameter name	Value
Processor	Intel Core i7 2.3 GHz
Memory	32GB
Operating system	Windows10
Computer vision library	Opencv

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND DATA SETS

This experiment used Matlab platform and the environmental parameters are shown in Table 1

Weizmann human behavior database and UCF motion database are selected as the data sources:

Weizmann human behavior database: a public human behavior database, including various human behaviors such as walking, running, jumping, bending and arm stretching. It has a high pixel resolution and its video frame transmission rate is up to 50frame/s.

UCF motion database: it covers various human motion postures, and with rich data types, diverse sampling scenes and more on-site samplings, it has high practical value.

Select 10 million frames of human posture videos each from the above two data sets, with a total of 20 million frames of videos, including walking, running and jumping. Half of them are used as training dada sets and the other half as test data sets. In actual video monitoring, there are always invalid videos. So in order to reduce the calculation of the algorithm, the video frames are preprocessed to filer invalid frames, and then the videos are edited at the at the frame rate of 25frame/s. Adjust the video image pixel to 227*227, to ensure the human motion posture image features with high definition.

B. EXPERIMENTAL STEPS

1. From Weizmann human behavior database and UCF motion database, 10 million frames of human motion posture videos are selected respectively, with a total of 20 million frames, including walking, running and jumping. Half of them are used as training data set to train the classifier, and the remaining half are used as test data set.

$$T(w_0, w_1) = \frac{\sum_{i=-a}^{a}\sum_{i=-b}^{b}\left[A_1(a_1+i, b_1+j) - \overline{A}_1(a_1, b_1)\right] \times \left[A_2(a_2+i, b_2+j) - \overline{A}_2(a_2, b_2)\right]}{(2a+1)(2b+1)\sqrt{\chi^2(A_1+A_2)}} \tag{15}$$

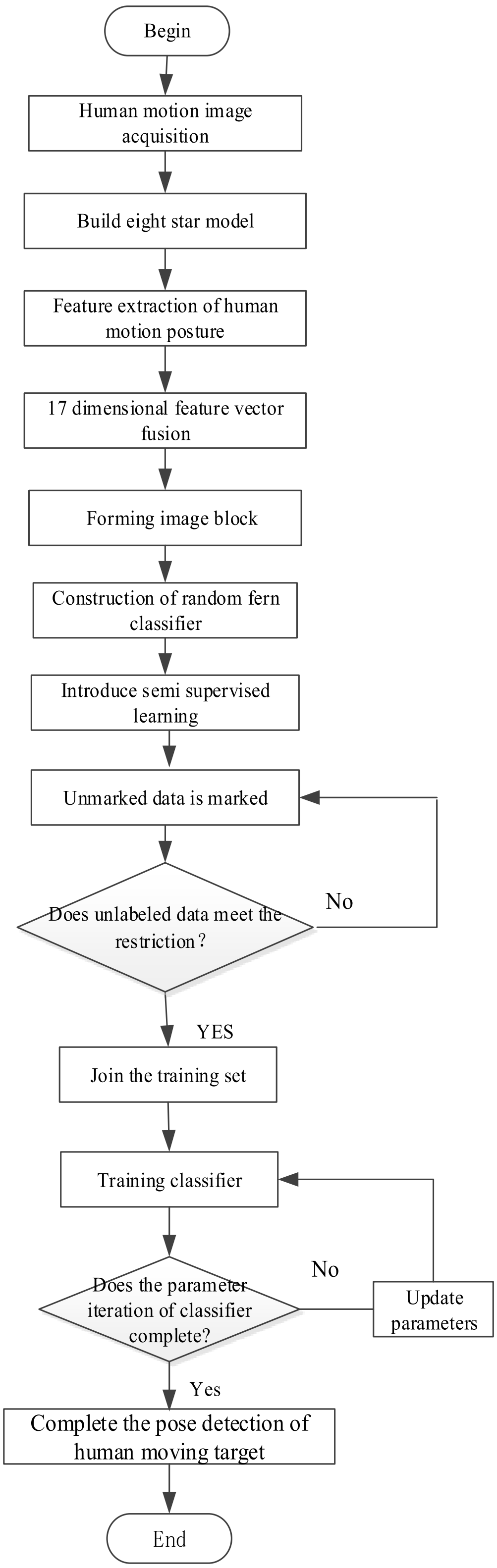

FIGURE 2. **Human body moving target posture detection process.**

2. In the actual video monitoring, invalid video frames are often easy to appear. In order to reduce the calculation amount of the algorithm, we first preprocess the video frames, filter the invalid frames, and then process the video clips according to the frame rate of 25frame / s, and adjust the video image pixels to 227 * 227, so as to ensure the characteristics of high-definition human motion posture image.

3. According to the parameters in Table 1, the experimental environment is built, and the experimental data set is input and operated;

4. The performance of this algorithm is verified by using different indicators.

C. ANALYSIS OF EXPERIMENTAL INDICATORS

Select the following experimental indexes to verify the algorithm in this paper, as follows:

(1) Extraction effect of human motion posture

In the text, the eight-star model is used to extract the human motion posture features. In order to verify the feature extraction effect, select the human motion posture images data of walking, running and jumping from the database, extract the feature values and draw the eight-star model to verify the feature extraction effect;

(2) Accuracy of multiple features fusion: the accuracy of motion posture features fusion is calculated by comparing the different numbers of human motion videos. The formula is as follows:

$$Precision = \frac{V_{true}}{V_{false} + V_{true}} \tag{16}$$

where, V_{true} represents the number of pixels of the video images correctly fused, and V_{true} is the number of pixels of the images wrongly fused.

(3) Correct classification rate of the postures with different features

In order to correctly detect the human motion postures, design a classifier to classify the image maps. The classification rate directly affects the detection accuracy of the algorithm. Compare the human motion posture classification rate of the algorithm in this paper with those of different algorithms.

(4) F-measure of human motion posture detection

F-measure is the weighted average of the recall rate and accuracy of the comprehensive evaluation. The formula is as follows:

$$\mathrm{F-measure} = \frac{Precision \times recall}{Precision + recall} \tag{17}$$

(5) Overlap ratio

To comprehensively evaluate the detection performance, the index of overlap ratio is proposed. Describe the overlap of the detection algorithm output and the calibrated range, and the overlap ratio can be taken as a key index to determine the detection effectiveness. The formula is as follows:

$$m = \frac{area\,(O \cap W)}{area\,(O \cup W)} \tag{18}$$

where, O represents the scan pixel domain of human motion target posture, and W is the calibrated domain of human motion target posture.

(6) Time consuming of the algorithm

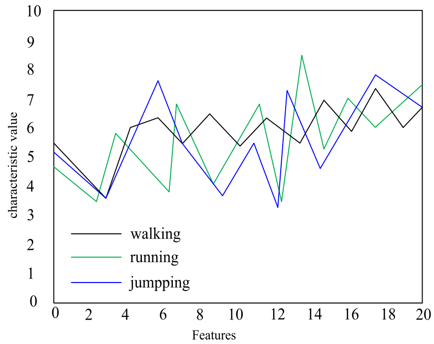

FIGURE 3. Eight star model of the proposed algorithm.

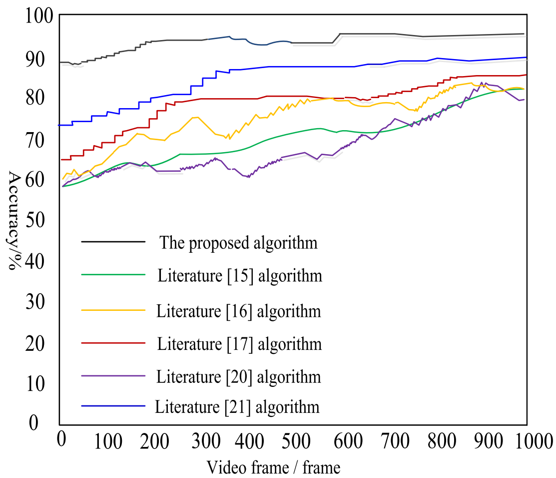

FIGURE 4. Comparison of multi feature fusion accuracy.

Compare the time consuming of this algorithm and that of other algorithms in literature, to verify the timeliness of the algorithm in this paper.

D. RESULTS

1) COMPARISON OF EXTRACTION EFFECT OF HUMAN MOTION POSTURE

Based on the image data of the human motion postures of walking, running and jumping, the algorithm in this paper is used to draw the eight-star model. It is shown in the Figure 3.

From the above drawing of the eight-star model, it can be seen that the features of the three human motion postures drawn with the algorithm in this paper have a better display effect, with obvious feature quantity and low redundancy. In that way it verifies the effectiveness of the use of eight-star model to extract motion posture features in this paper, and provides a good basic condition for the following posture detection.

2) COMPARISON OF ACCURACY OF MULTIPLE FEATURES FUSION

Select 10 million frames of human motion videos from the data sources for test, calculate the feature values of the multiple features fusion with different quantities of video frames, measure the fusion accuracy and compare the algorithm in this paper with algorithms in literature [15]–[17], [20], and [21]. It is shown in Figure4.

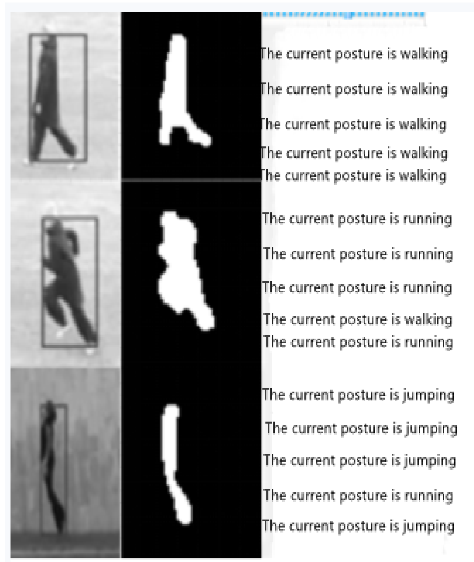

FIGURE 5. Classification of human motion posture.

TABLE 2. Comparison of correct classification rate (%).

Algorithm	Human motion posture walk	running	jump
The proposed algorithm	90.3	92.5	91.3
Literature [15]	89.3	88.5	87.2
Literature [16]	83.5	84.2	80.5
Literature [17]	76.3	76.5	70.2
Literature [20]	69.2	78.6	75.4
Literature [21]	63.2	73.5	65.0

From the analysis of Figure4, it can be seen that with the increase of video frames, the broken lines in the above algorithm are on the rise as a whole. The multiple features fusion accuracy of the algorithms in literature [15]–[17] and [20] is up to about 80%, and that of the algorithm in literature [21] is higher with a stable 90% after the video frame quantity reaches 400. It is obvious from the broken line figure that the accuracy of the algorithm in this paper fluctuates little, which rises slowly between 90% and 95%. Compared with the algorithms in other literatures, it has advantages both in stability and accuracy, further verifying the effectiveness of the use of eight-star model to extract and fuse the motion posture features in this paper.

3) COMPARISON OF CORRECT CLASSIFICATION RATE OF POSTURES WITH DIFFERENT FEATURES

This paper selects the image data of the motion postures of walking, running and jumping as the test objects and generates and classifies the human motion postures by computer operation.As shown in Figure5. It can be seen from the classification results that the calculation results of the proposed algorithm are more accurate.

Under different feature poses, the correct classification rates of different algorithms are calculated and compared, as shown in Table 2

TABLE 3. F value comparison of human motion posture detection.

Algorithm	Video frame / frame			
	250	500	750	1000
The proposed algorithm	0.96	0.94	0.98	0.92
Literature [15]	0.62	0.72	0.72	0.77
Literature [16]	0.75	0.81	0.74	0.82
Literature [17]	0.82	0.72	0.87	0.79
Literature [20]	0.70	0.83	0.83	0.91
Literature [21]	0.86	0.78	0.80	0.82

Form Table2 it can be seen that in the three different motion scenarios, the correct classification rate in different algorithms for running is high as whole, because running has rich features which are easier to be classified. The comparison data from different algorithms shows that the correct classification rates in literature [15] and [16] are 80% and above and those in literature [17], [20] and [21] are lower, less than 80%. However, the correct classification rates of the algorithm in this paper for the three motion scenarios are all more than 90%, with the highest 92.5%. Therefore, the construction of classifier and the classifier training based on semi-supervised learning in this paper have a good effect and can give higher classification accuracy.

4) COMPARISON OF F-MEASURE OF HUMAN MOTION POSTURE DETECTION

The higher the F-measure is, the more accurate and comprehensive the results of the algorithm are. To fully verify the advantage of the algorithm in this paper, the F-measures of human motion posture detection in different algorithms are calculated, the results of which are shown in Table3:

From the analysis of the F-measures comparison results of human motion posture detection in Table. 3, it can be seen that F-measure of human motion posture detection in the algorithm of this paper is always the highest, with the average of 0.95. In the algorithms in other literatures, when the quantity of video frames reaches 10 million in literature [20], its F-measure is 0.91, relatively higher but lower than the minimum F-measure in the algorithm in this paper. In literature [15], the F-measure in its algorithm is not up to 0.8, and those in literature [16], [17], and [21] are not up to 0.9. The comparison data shows that the human motion posture detection results of the algorithm I this paper are far better than those in other literatures.

5) COMPARISON OF OVERLAP RATIO

The overlap ratio is used to describe the overlap between the detection algorithm output and the calibrated. The higher the overlap is, more accurate the input results are. The algorithm in this paper is compared with the algorithms in other literatures to verify the performance of the algorithm. It is shown in Figure 6.

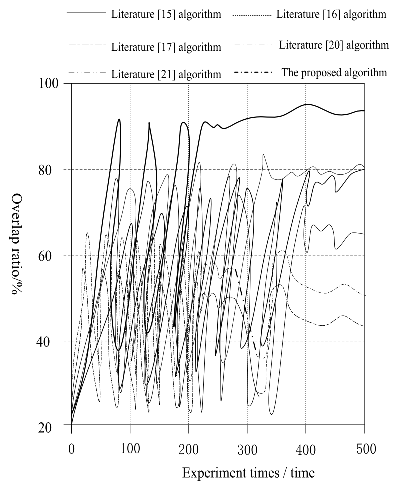

FIGURE 6. Comparison of overlap ratio.

According to the principle that the higher the overlap is, the detection performance of the algorithm is, it can be seen from the trend of the curve in Figure6 that the algorithm in this paper has some advantages in the higher overlap with the overlap ratio of 95%. In the algorithms in other literatures, the overlap ratios fluctuate greatly and need to be stabilized with a lot of experiments. And finally the higher overlap ratios are 80% in literature [16] and [20], 60% and above in literature [15], and not up to 60% in literature [17] and [21]. In this paper, semi-supervised learning is used to train the classifier and efficiently complete the labeling of the unlabeled data, which realizes the high overlap ratio between the detection algorithm output and the calibrated range and improves the detection performance of the algorithm.

6) COMPARISON OF ALGORITHM TIME CONSUMING

The time consuming of the proposed algorithm is compared with those in literature [15]–[17], [20] and [21], and the results are shown in Table4.

TABLE 4. Comparison of algorithm running time.

Algorithm	Video frame / frame			
	250	500	750	1000
The proposed algorithm	0.9	1.0	1.1	1.5
Literature [15]	1.2	1.8	2.3	2.5
Literature [16]	1.3	1.6	1.9	1.9
Literature [17]	1.0	1.5	1.9	2.3
Literature [20]	1.3	1.5	1.6	2.5
Literature [21]	1.6	2.3	2.6	3.0

By analyzing the time consuming of different algorithms in Table4, it can be seen that with the increase of video frames, the time consuming of the algorithm also increases. The average time consuming of the algorithm in this paper is 1.1s, followed by literature [16] and [17] with 1.7s, then literature [20] with 1.8s and literature [15] with 2.0s, and literature [21] has the longest time consuming with the average of 2.4s. Therefore, the algorithm in this paper has a faster running speed, and it uses semi-supervised learning to train a large number of unlabeled data, which greatly saves the algorithm running time.

E. DISCUSSION

In order to find an efficient human motion pose detection method, actively respond to various emergencies, and give full play to the advantages of human motion target analysis in the field of sports and medical treatment, this paper proposes a human motion target pose detection algorithm in the environment of Internet of things. The results show that the proposed algorithm can effectively avoid the interference of external factors, complete the body motion attitude detection with high accuracy, and the target image detection has high overlap ratio, and the algorithm runs in a short time.

The application of computer vision technology for human motion target pose detection can make the computer better understand human behavior, the research results will be applied in more and more fields. Aiming at the problem of human motion target analysis, some achievements have been made in the existing literature, Guoliang and Xiaoxiang [47] collected data of pedestrian behavior characteristics with the help of mobile phone built-in multi-source sensors, constructed classification model to match different behavior characteristics, and then identified human behavior; Carmona and Climent [23] used sub tensor projection and dense trajectories to recognize human motion, proposed a method based on tensor projection, and evaluated the results obtained by only using template based motion recognition. In order to achieve the most advanced recognition rate, different characteristics were fused to effectively complete human motion recognition; Qi *et al.* [48] trained a probability model using semi supervised learning of Gaussian mixture model, and used a series of probability feature vectors to describe each pose in a given motion. Based on the statistical data of all pose features, a motion feature descriptor was proposed. The existing research literature has made some achievements. The existing research literature has made some achievements. However, on the basis of the existing research, this algorithm optimizes the design of the classifier, and uses the semi supervised learning method to train the classifier under the three-layer constraints to complete the pose detection of human moving objects, which has high accuracy of feature extraction and feature fusion, and the correct classification rate of feature pose is high, It can provide data support for human motion behavior research and further promote the development of computer vision.

VI. CONCLUSION

In order to further delve the presentation information of human motion target, promote the further development of research on computer vision and give full play to its practical role in video monitoring and other fields, more than more scholars pay attention to the detection method of human motion posture. This paper puts forward the detection algorithm of human motion target posture based on semi-supervised learning in the environment of IoT. By virtue of the mature model theory of eight-star model, the human motion target posture features are accurately extracted and fused to improve the quality of images. This paper adopts the PN Learning algorithm in semi-supervised learning, sets the restrictions and fully takes the advantage of semi-supervised learning to unlabeled data calculation, reducing the calculation complexity and storage space of unlabeled data and improving the calculation efficiency. It uses semi-supervised learning to train the classifier and complete human motion target posture detection. The experiment uses Weizmann human behavior database and UCF motion database as the data sources and selects three different postures of walking, running and jumping to verify the performance of the algorithm in this paper in multiple experiments with a large number of data. The results show that, compared with the algorithms in other literatures, the algorithm in this paper has a good effect of feature extraction. The accuracy of multiple features fusion is about 95%, the correct classification rate of different feature postures is up to 92.5%, and the average F-measure of human motion posture detection is 0.95, with high overlap ratio and short time consuming, it has significant advantages. However, in the actual human motion video acquisition, there are many external factors, and the human motion posture will also produce a variety of complex situations. In the future research, we will deeply understand the problem of human motion posture under various conditions, explore the relevant influencing factors, obtain positive and useful information through research, and contribute to the research of computer human vision.

REFERENCES

[1] S. V. Mora and W. J. Knottenbelt, "Deep learning for domain-specific action recognition in tennis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 170–178.
[2] J. Yu, D. Tao, M. Wang, and Y. Rui, "Learning to rank using user clicks and visual features for image retrieval," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 767–779, Apr. 2015.
[3] R. Singh and V. Ranga, "Performance evaluation of machine learning classifiers on Internet of Things security dataset," *Int. J. Control Autom.*, vol. 11, no. 5, pp. 11–24, May 2018.
[4] T. Bates, K. Ramirez-Amaro, T. Inamura, and G. Cheng, "On-line simultaneous learning and recognition of everyday activities from virtual reality performances," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 3510–3515.
[5] B. Victor, Z. He, S. Morgan, and D. Miniutti, "Continuous video to simple signals for swimming stroke detection with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, vol. 3, no. 123, pp. 122–131.
[6] N. Baik, S. K. Sharma, and B. Khuntia, "Encrusted security for Internet of Things using MAC-OMURA," *Int. J. Control Autom.*, vol. 11, no. 2, pp. 45–54, Feb. 2018.

[7] J. Yu, M. Tan, H. Zhang, D. Tao, and Y. Rui, "Hierarchical deep click feature prediction for fine-grained image recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jul. 30, 2019, doi: 10.1109/TPAMI.2019.2932058.
[8] U. E. Rabab, I. Ahmed, M. I. Aslam, and M. Usman, "FPGA implementation of secure Internet of Things (SIT) algorithm for high throughput area ratio," *Int. J. Future Gener. Commun. Netw.*, vol. 11, no. 5, pp. 63–72, Sep. 2018.
[9] Y. Yin, Z. Cao, Y. Xu, H. Gao, R. Li, and Z. Mai, "QoS prediction for service recommendation with features learning in mobile edge computing environment," *IEEE Trans. Cognit. Commun. Netw.*, vol. 6, no. 4, pp. 1136–1145, Dec. 2020.
[10] M. H. Shaikh and J. Kim, "Privacy safeguarding techniques for Internet of Things based mobile enterprise systems," *Int. J. Adv. Sci. Technol.*, vol. 132, pp. 5–34, Nov. 2019.
[11] C. Tang, W. Wang, X. Wang, C. Zhang, and L. Zou, "Human behavior recognition based on multi view semi supervised learning," *Pattern Recognit. Artif. Intell.*, vol. 32, no. 4, pp. 376–384, 2019.
[12] Z. Tao and T. Zhenmin, "An improved semi supervised learning algorithm based on non negative low rank sparse graph," *Acta Electronica Sinica*, vol. 4, no. 39, pp. 158–164, 2017.
[13] L. Yanchao, X. Fu, C. Zhi, and L. Bo, "Adaptive active semi supervised learning method," *Acta Sinica Sinica*, vol. 31, no. 12, pp. 142–156, 2020.
[14] D. Lan, W. Di, L. Lu, and Y. Guo, "SAR target detection network based on semi supervised learning," *Acta Electronica Sinica*, vol. 42, no. 1, pp. 154–163, 2020.
[15] Z. Xiao, P. X. Dong, and W. J. Xuan, "Human behavior recognition method based on spatiotemporal characteristics of posture," *J. Comput. Aided Design Graph.*, vol. 30, no. 9, pp. 29–38, 2018.
[16] L. Qi, W. Xiangdong, and L. Hua, "3D human pose tracking method based on dual Kinect sensors," *J. Syst. Simul.*, vol. 32, no. 8, pp. 1446–1454, 2018.
[17] W. Yongxiong, Z. Yan, and L. Xuan, "3D complex human behavior recognition based on fusion of interactive information and energy features," *Mini Comput. Syst.*, vol. 8, no. 2, pp. 1828–1834, 2018.
[18] D. Chongyang, L. Kai, and L. Guang, "Research on human skeleton behavior recognition based on spatiotemporal weight pose motion features," *Acta Computa Sinica*, vol. 43, no. 445, pp. 31–42, 2020.
[19] Y. L. Hsueh, W. N. Lie, and G. Y. Guo, "Human behavior recognition from multiview videos," *Inf. Sci.*, vol. 517, pp. 275–296, May 2020.
[20] C. Dai, X. Liu, J. Lai, P. Li, and H.-C. Chao, "Human behavior deep recognition architecture for smart city applications in the 5G environment," *IEEE Netw.*, vol. 33, no. 5, pp. 206–211, Sep. 2019.
[21] S. A. Khowaja and S.-L. Lee, "Semantic image networks for human action recognition," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 393–419, Feb. 2020.
[22] D. Sudha and M. Ramakrishna, "Comparative study of features fusion techniques," in *Proc. Int. Conf. Recent Adv. Electron. Commun. Technol. (ICRAECT)*, vol. 1, no. 3. Washington, DC, USA: IEEE Computer Society, Mar. 2017, pp. 235–239.
[23] J. M. Carmona and J. Climent, "Human action recognition by means of subtensor projections and dense trajectories," *Pattern Recognit.*, vol. 81, pp. 443–455, Sep. 2018.
[24] L. Rosyidi, M. Asvial, and R. F. Sari, "Lightweight and energy-aware wireless mesh routing for autonomic Internet of Things network," *Int. J. Control Autom.*, vol. 12, no. 2, pp. 33–44, Feb. 2019.
[25] N. Singh, A. Hans, and S. Kaur, "Layer and RFID based security issues of Internet of Things," *Int. J. Grid Distrib. Comput.*, vol. 9, no. 10, pp. 301–310, Oct. 2016.
[26] C. Cao, Y. Zhang, C. Zhang, and H. Lu, "Body joint guided 3-D deep convolutional descriptors for action recognition," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1095–1108, Mar. 2018.
[27] Z. Li, D. Milutinović, and J. Rosen, "From reaching to reach-to-grasp: The arm posture difference and its implications on human motion control strategy," *Exp. Brain Res.*, vol. 235, no. 5, pp. 1–16, 2017.
[28] R. Moreira, A. Teles, R. Fialho, T. C. P. dos Santos, S. S. Vasconcelos, I. C. de Sá, V. H. Bastos, F. Silva, and S. Teixeira, "Can human posture and range of motion be measured automatically by smart mobile applications?" *Med. Hypotheses*, vol. 142, Sep. 2020, Art. no. 109741.
[29] H. Kainz, M. Hajek, L. Modenese, D. J. Saxby, D. G. Lloyd, and C. P. Carty, "Reliability of functional and predictive methods to estimate the hip joint centre in human motion analysis in healthy adults," *Gait Posture*, vol. 53, pp. 179–184, Mar. 2017.
[30] T. Qi, Y. Feng, J. Xiao, H. Zhang, Y. Zhuang, X. Yang, and J. Zhang, "A human motion feature based on semi-supervised learning of GMM," *Multimedia Syst.*, vol. 23, no. 1, pp. 85–93, Feb. 2017.
[31] F. Wrede and A. Hellander, "Smart computational exploration of stochastic gene regulatory network models using human-in-the-loop semi-supervised learning," *Bioinformatics*, vol. 35, no. 24, pp. 5199–5206, Dec. 2019.
[32] M. Lv, L. Chen, T. Chen, and G. Chen, "Bi-view semi-supervised learning based semantic human activity recognition using accelerometers," *IEEE Trans. Mobile Comput.*, vol. 17, no. 9, pp. 1991–2001, Sep. 2018.
[33] D. Suh, K.-S. Cho, and J.-H. Song, "A novel way of integrated risk awareness based on the Internet of Things for intelligent crime prevention," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 2, pp. 69–80, Feb. 2018.
[34] Q. Zheng, H. Delingette, and N. Ayache, "Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow," *Med. Image Anal.*, vol. 56, pp. 80–95, Aug. 2019.
[35] D. Slepev and M. Thorpe, "Analysis of p-Laplacian regularization in semi-supervised learning," *SIAM J. Math. Anal.*, vol. 51, no. 3, pp. 1520–1526, 2017.
[36] Z. Krkeljas, "Changes in gait and posture as factors of dynamic stability during walking in pregnancy," *Hum. Movement Sci.*, vol. 58, pp. 315–321, Apr. 2018.
[37] J. Pettersson and P. Falkman, "Human movement direction classification using virtual reality and eye tracking," *Procedia Manuf.*, vol. 51, no. 3, pp. 95–102, 2020.
[38] J.-S. Hung, P.-L. Liu, and C.-C. Chang, "A deep learning-based approach for human posture classification," in *Proc. 2nd Int. Conf. Manage. Sci. Ind. Eng.*, Apr. 2020, pp. 171–175.
[39] A. Wang, Y. Wang, and Y. Chen, "Hyperspectral image classification based on convolutional neural network and random forest," *Remote Sens. Lett.*, vol. 10, no. 11, pp. 1086–1094, Nov. 2019.
[40] O. Samanta, A. Roy, S. K. Parui, and U. Bhattacharya, "An HMM framework based on spherical-linear features for online cursive handwriting recognition," *Inf. Sci.*, vol. 441, pp. 133–151, May 2018.
[41] H. Ragb and V. Asari, "Multifeature fusion for robust human detection in thermal infrared imagery," *Opt. Eng.*, vol. 58, no. 4, pp. 1–10, 2019.
[42] Y. Peng, H. Tao, W. Li, H. Yuan, and T. Li, "Dynamic gesture recognition based on feature fusion network and variant ConvLSTM," *IET Image Process.*, vol. 2, pp. 521–533, Apr. 2020.
[43] J. Zhao, C. Zhang, D. Li, and J. Niu, "Combining multi-scale feature fusion with multi-attribute grading, a CNN model for benign and malignant classification of pulmonary nodules," *J. Digit. Imag.*, vol. 33, no. 4, pp. 869–878, Aug. 2020.
[44] K. Yiqun, C. Hong, and C. Fang, "Semi supervised hand pose estimation algorithm based on multi view projection," *J. UESTC*, vol. 48, no. 5, pp. 109–115, 2019.
[45] Z. Jun, F. Bin, and Y. Xinjun, "Human detection algorithm of fire infrared image based on U-Net network," *Laser Infr.*, vol. 49, no. 12, pp. 1483–1489, 2019.
[46] M. Kim, D.-G. Lee, and H. Shin, "Semi-supervised learning for hierarchically structured networks," *Pattern Recognit.*, vol. 95, no. 2, pp. 191–200, Nov. 2019.
[47] C. Guoliang and C. Xiaoxiang, "Pedestrian behavior recognition in indoor fire based on mobile multi-source sensor," *J. Tongji Univ., Natural Sci. Edition*, vol. 47, no. 3, pp. 414–420, 2018.
[48] T. Qi, Y. Feng, J. Xiao, H. Zhang, Y. Zhuang, X. Yang, and J. Zhang, "A human motion feature based on semi-supervised learning of GMM," *Multimedia Syst.*, vol. 23, no. 1, pp. 85–93, Feb. 2017.

LEI CHEN received the M.S. degree from Shenyang Sport University. She is currently an Associate Professor with Shenyang Ligong University. Her research interests include intelligence sports, physical education, and social sports.

SHANGBIN LI received the M.S. degree from Harbin Engineering University. He is currently a Professor with Harbin Engineering University. His research interests include artificial intelligence and sport engineering.