Pedestrian Attitude Estimation and Recognition Algorithm Based on RF Data

Jie Zhang,1 Dejie Sun,2 Xin Li,2 Gang Li,2 and Yangsen Wei2

1School of Information Science and Engineering (School of Software), Yanshan University, Qinhuangdao 066000, China
2School of Mathematics and Information Science and Technology, Hebei Normal University of Science and Technology, Qinhuangdao 066004, China

Correspondence should be addressed to Dejie Sun; sdj3877@hevttc.edu.cn

Received 20 January 2022; Revised 13 February 2022; Accepted 14 February 2022; Published 20 March 2022

Academic Editor: Xin Ning

Copyright © 2022 Jie Zhang 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.

Pedestrian detection plays a vital role in the estimation of human posture, especially the pedestrian detector, which provides the position information of human body in the image. This paper proposes a pedestrian gesture recognition algorithm using radio frequency data based on RFID technology. Its goal is to serve as a starting point for research in the field of human gesture recognition. The position of pedestrians in the input picture can be adaptively adjusted by learning the transformation parameters of samples during network training, and the local features of each block can be trained by multiple cross entropy functions at the network’s end, allowing the sample local area to be fully utilized as the network’s training. More accurate DFL and attitude feature recognition can be achieved using the measured received signal strength information and channel state information. The pedestrian gesture recognition algorithm based on RF data proposed in this paper improves pedestrian recognition accuracy, especially in noisy environments, such as fuzzy or occlusion, according to experimental results.

1. Introduction

In the field of computer vision, pedestrian detection and attitude estimation is a frontier direction and research hotspot. Automatic detection of pedestrians in image sequences or video streams is primarily accomplished through intelligent analysis, which estimates the position and direction information of each part of the pedestrians in the image [1]. In intelligent vehicle safety systems, intelligent monitoring, motion analysis, human-computer interaction, and other fields, pedestrian detection and attitude estimation are widely used. The core problem of video surveillance and retrieval of specific targets is pedestrian re-recognition. Its goal is to find the target in videos’ shot by cameras in various positions during various time periods while specific pedestrians are present [2]. This subject, however, is facing a huge challenge due to a number of issues such as pedestrian posture, camera angle, and shooting picture resolution. Pedestrian re-recognition is usually divided into two categories: feature expression and distance measurement. Automatic classification from video sequences is referred to as video behavior recognition, and the category is usually human behavior, such as walking and jogging. The biggest difference between video analysis and image analysis is that video sequences contain extra time information, which usually requires much more computation. Traditional and existing mainstream pedestrian attitude estimation algorithms based on deep learning [3, 4] cannot meet the real-time and accuracy requirements of task processing. Pedestrians’ re-recognition algorithm pays more attention to the study of single-frame images and global features of images. When it is applied to actual video, it cannot effectively use the information contained in video sequences and deal with background noise interference such as blurring caused by pedestrian movement.

How to realize the attitude estimation and re-recognition of pedestrians, discover and deal with the abnormal dangerous behaviors of pedestrians in time, and realize the retrieval and tracking of target figures, so as to improve the safety early warning ability of public places, has become one of the hot research topics in industry and academia. Video behavior recognition is a technique for automatically
detecting and classifying ongoing activities in videos. It has a wide range of uses, including monitoring, online video, and motion analysis. Surveillance video abnormal behavior is the focus of attention in a few specific scenes. By automatically detecting abnormal behavior, staff can quickly determine whether dangerous behavior and its characteristics exist in the target area, preventing the spread of vicious activities and worse consequences [5]. RFID is a technology that uses radio frequency signals to automatically identify and obtain information about target objects. The RFID systems are widely used in a variety of industries because of their benefits such as large data volume, high confidentiality, strong anti-interference ability, quick identification time, and low cost [6]. In comparison to existing device-free localization (DFL) technologies based on camera, ultrawideband radar, infrared, and ultrasonic technologies, DFL technology based on wireless sensor networks has become a research hotspot in the DFL field at the moment due to its advantages of low cost, good versatility, and the ability to locate through walls and smoke [7]. Wireless positioning technology, particularly indoor wireless positioning technology, has gotten a lot of attention as a result of the popularity and development of mobile computing devices, and it has a bright future in areas like wireless navigation, warehouse management, logistics tracking, personal entertainment, mobile office, and employee management. The algorithm for pedestrian attitude estimation and recognition is investigated in this paper using RF data.

The RFID technology is a kind of automatic identification technology. During the development of automatic identification technology, there are many identification technologies, all of which have their own advantages [8]. But RFID technology has the most competitive advantage, so it has received extensive attention. In recent years, the RFID system product market has become one of the fastest growing markets in the information industry. It uses radio frequency signals to realize contactless information transmission through spatial coupling and identifies specific objects through the transmitted information, which can work in various harsh environments without manual intervention [9]. Compared with other recognition technologies, RFID has strong advantages in recognition distance, recognition speed, multitarget recognition, and moving object recognition [10]. Aiming at the research of pedestrian recognition, this paper proposes a pedestrian recognition algorithm based on RFID technology. Using the multitask pedestrian posture estimation network proposed by us, we set up a multitask regional suggestion network. According to the detected pedestrian posture and bounding box, we got the suggestions of local region division of pedestrian image, reduced the interference of background noise and solved the problem of pedestrian image alignment, preprocessing with a human body detector and foreground highlighting, determining the approximate position and size of the human body while simultaneously removing clutter from the background, and estimating the appearance model of human body parts based on prior segmentation and appearance transformation. The impact of random factors on positioning estimation accuracy is reduced by using historical time-domain data. The ray tracing method is used to investigate the differences among individuals in a wireless communication environment in order to improve the system’s positioning accuracy. According to the research, this method has a high recognition rate.

2. Related Work

Reference [11] improves the accuracy of the result of human posture estimation by reducing the requirements in the initialization process by adding more constraints, such as the similarity of the color of pants on two legs, the similarity of skin color of arms and face, and the similarity of clothes color of arms and upper body. Spotorno et al. [12] added the constraints of the non-tree-like probability graph model to the graph structure model, which improved the accuracy of pose estimation to some extent. Beck et al. [13] proposed a tree structure model based on the object context based on prior knowledge of other objects, which improved the accuracy of human pose estimation. Song et al. [14] used computer graphics to reconstruct the human skeleton from data collected by the nodes of human joints and analyze the human body’s movement posture. Reference [15] using the ant colony algorithm for feature selection, the feature vector with the best classification ability is obtained. Finally, the feature vector is used to create a support cross product model to classify human posture. Reference [16] employs a cyclic neural network and a twin neural network to learn the interactive relationship between multiple frames of video images in order to learn video-level discrimination features for pedestrian reidentification. Spindle Net, proposed in the literature [17], divides the human body into seven local regions by using the skeleton key point structure, then extracts features from these seven regions on different feature scales and then connects them, and finally obtains a pedestrian re-recognition distinguishing feature that aggregates local features and global features of multiple feature scales. Reference [18] proposed a coarse-to-fine attitude estimation algorithm, which is mainly used to find the matching with the given reference attitude in the image. Reference [19] proposed a novel feature extraction method, and the auxiliary system improved the recognition accuracy. Benziane et al. used a binary tree as a recognition model, which has a good performance in simple gesture recognition [20].

The HOG method is used by Mi et al. [21] to determine the position information of the human body in the image, and the random tree or random forest method is used to identify and classify the human body in the image. Zhenbing et al. design a feature selection mechanism that extracts global optimal features from a variety of features, such as shape context, edges, colors, and gradients, quickly completes the feature extraction process, and then performs pedestrian detection [22]. Dai et al. use time domain similarity constraints to find the best candidate matching samples after selecting multiple similar samples from the pretrained template library [23]. The descriptors of motion characteristics are described using optical flow technology, and the closest human posture sequence is found from the motion sample database as the final output result, according to the
principle of motion matching. In a low-resolution video shot from a long distance, the seeding method can accurately estimate human posture. To reduce the acquisition error of the acceleration sensor, reference [24] analyzes the error of sensor data and uses an arithmetic mean limiting filtering algorithm. The sensor’s output is then converted to acceleration output. The spatial angle coordinate transformation algorithm is used to analyze and calculate the data.

Based on the previous research, this paper proposes a pedestrian recognition algorithm based on RFID technology. The algorithm adopts dynamic complementary binary tree search method, makes full use of the conflict information obtained, effectively reduces the data transmission amount in the decision process, and improves the recognition efficiency of human posture. It also solves the problem of low pedestrian detection accuracy, which is caused by the fact that different structures have the same characteristics and pixels in opposite directions are mapped to the same interval when the target and the background have a high color contrast. In comparison to other algorithms in different image databases, simulation results show that this algorithm not only uses a human detector and a foreground highlighting algorithm to reduce component space but also improves the accuracy of human posture estimation.

3. Methodology

3.1. RFID. In recent years, automatic identification system has received extensive attention in various fields. For example, logistics management, cargo tracking, and sports timing. [25]. The automatic identification system is mainly divided into the following categories: barcode system, optical feature identification system, biometric identification system, smart card identification system, and RFID system. A typical RFID system includes three parts: reader, electronic tag, and application software system. The reader is a device for collecting information, which can read and write data. It consists of antenna, RF transceiver module, signal processing unit, control unit, and interface circuit. The tag information read by the reader is managed and transmitted by computer and network system. The electronic tag has an antenna and a chip inside, each tag has a unique electronic code, and the tag is attached to the object to identify the target object. The tag can be a “card” or other forms of devices. The application software system is a host computer, which mainly stores and manages data. It communicates with RFID readers distributed everywhere through various interfaces and acquires the electronic tag information captured by RFID readers in real time.

RFID research focuses on standardization, tag cost reduction, key technologies, and system applications because it is a relatively large and complex technology. Standardization is defined as the identification of problems in products, processes, or services, as well as the provision of working languages that can be observed together in order to promote technical cooperation and eliminate trade barriers [26]. The success of RFID technology application is dependent on the label cost. IC chip, antenna, and package make up the majority of RFID tags. Anticollision technology, security technology, antenna technology, frequency selection technology, low power consumption technology, and so on are some of the most important RFID technologies. RFID has a wide range of potential applications. It is now widely used in areas such as public transportation, identification, supply chain management, and anticounterfeiting. RFID is a noncontact automatic identification technology that uses radio frequency to identify targets and exchange data. It is primarily used to quickly identify targets in the radiofrequency area.

In RFID system, computer communication network is the equipment for data management and communication transmission. In the working process of the system, the reader usually emits RF energy in an area to form an electromagnetic field, and the range depends on the transmission power. When the tag passes through this area, it is triggered to send the data stored in the tag or rewrite the data stored in the tag according to the instruction of the reader/writer. The reader-writer can receive the data sent by the tag or send data to the tag and can interface with the computer communication network through the standard interface to realize the communication and transmission of data. When multiple electronic tags arrive at the effective radio frequency area at the same time, the tags will respond to the reader’s instructions and send signals at the same time, which will cause the reader to fail to receive data correctly and identify the tags correctly, resulting in conflicts [27]. A reliable anticollision algorithm is needed to solve the problem that readers cannot correctly identify tags due to data conflicts when identifying multiple electronic tags. The system composition of RFID is shown in Figure 1.

The reader and the electronic tag are separated by several meters to tens of meters in the far field of electromagnetic radiation. Backscattering modulation technology is used to create RFID systems at the moment. In some large applications, the data collected by multiple readers must be uniformly processed before being made available to users for decision-making and processing, necessitating the use of a networked RFID system. The remote control and processing of the networked RFID system is also possible. We can only meet the demands of real-time data collection and decision-making with the help of a networked RFID system. The energy exchange between the reader and tag antennas is similar to a transformer structure in a low-frequency RFID system, which is realized by quasi-static field coupling. At this point, alternating magnetic fields complete the communication between the reader and the tag. A strong electromagnetic field is produced by the reader’s antenna. Because the wavelength of the magnetic field is far greater than the distance between the reader and the tag because of its low frequency, a portion of the magnetic lines pass through the tag antenna coil at a distance from the reader antenna [28]. All circuit functions are sequenced by the clock so that data in the memory can be transmitted to the reader-writer at the correct time. Before the tag is installed in the identification object, the data in the memory is the unique code specified by the application system. When data is read from the memory, the encoder encodes it, and the modulator receives it and transmits and reflects it to the reader-writer via the antenna circuit. The controller is in charge of
decoding the signal received by the antenna and writing it to memory when writing data.

The antenna coil of the tag and its internal circuit form a resonant circuit whose resonant frequency is the frequency of the transmitted signal, and the tag receives the information from the reader through the resonant circuit. When the tag needs to transmit data back to the reader, the transmitted data stream changes the parameters of the resonant circuit of the electronic tag, which changes the impedance and phase of the primary circuit coupled with the magnetic field. This process is a modulation process [29]. The reader obtains the load modulation signal by detecting the voltage of the transformed impedance and then extracts the data returned by the tag through demodulation and related signal processing. Due to the limitation of application occasions, RFID tags usually need to be attached to the surfaces of different types and shapes of objects, and even embedded into objects. RFID tags require high reliability as well as low cost. In addition, the tag antenna and reader antenna are responsible for receiving energy and transmitting energy, respectively. These factors put forward strict requirements for antenna design. At present, the research on RFID antenna mainly focuses on the antenna structure and the influence of environmental factors on antenna performance.

3.2. Pedestrian Attitude Estimation. The task of pedestrian posture estimation is mainly to recover the process of human body posture formed by the position of pedestrian key points in a given image or video. The traditional method is to match people’s postures based on template matching and to match people’s postures in the image with a priori template as a whole. Human pose estimation mainly studies the positions of the detected parts of the human body in the image, and the direction and scale information of each part, which can be divided into two types: 2D and 3D according to the different pose estimation results. At present, the popular methods of human posture estimation at home and abroad can be divided into two types: model-based posture estimation and model-free posture estimation.

A simple algorithm is the model-based estimation method of human posture. The matching templates are usually images from a well-known source. Change the size and direction of the image to be detected in the template, then match the image to the template, calculate their similarity, and finally find the feature with the highest similarity to the image to be detected in the template. The following are the steps to take: First, a pedestrian is detected, and the detected rectangular frame is divided into a foreground and background region. The foreground and background in the foreground are then segmented further. Finally, a single frame image is used to estimate the human posture using the figure structure model. This method is primarily used to estimate human posture when the upper body is upright, but it can also be used to estimate human posture across the entire body. The machine learning method is commonly used in the model-free attitude estimation method. The algorithm for detecting pedestrians and estimating their attitudes is divided into three parts: training the classifier, detecting pedestrians, and estimating human posture using the pedestrian detector. The flow frame is shown in Figure 2.

The idea of template matching method is clear and simple, and the computational complexity is low. However, due to the relatively stable results obtained by the model-based human posture estimation method, it is the mainstream method of human posture estimation at present. However, this method has great dependence on the segmentation results and prior information of the image, and because the human body model is a kind of nonrigid object with great flexibility, usually an algorithm can only estimate the attitude of one kind of behavior. When using the model-free attitude estimation method, a large number of training samples need to be trained, which takes a lot of training time. We can get a relatively stable training algorithm and better optimization parameters. However, the model-free attitude estimation method is not limited by the model like the model-based attitude estimation method, so it is more flexible to use.

The representation model of human posture should not only include global considerations but also include sufficient local feature extraction modules. In this paper, a two-stage approach from coarse to fine is adopted to estimate human posture. That is, the rough human posture on a large level
is output first, and then, the details of each joint and the overall posture are embodied by using this result. Attitude recognition accuracy is one of the important indexes to judge whether attitude recognition is good or bad. Generally, the accuracy of gesture recognition of target without equipment is the success rate of estimating the actual gesture of the target.

Because the human body is a nonrigid object, each joint has a very high degree of flexibility. An upright human body may do different actions, which have a significant impact on the appearance of the human body. Therefore, when we detect and estimate pedestrians with different movements, we need a very flexible model to meet the different appearances brought by different movements. Pedestrian detection and re-recognition are the two main components of a pedestrian re-recognition application system. The video image data obtained by the pedestrian re-recognition system from the surveillance video was cut out by the pedestrian detection network and sent to the pedestrian re-recognition model trained by the pedestrian re-recognition data set for feature extraction to form the image feature library to be queried in the practical application process. We need to search the obtained image feature database to see if a pedestrian target of interest in the surveillance video is out of date.

3.3. Pedestrian Attitude Estimation and Recognition Based on RF Data. The background environment of pedestrians may be varied and beyond our control. These background environments may be simple. It may also be a complicated background. These complex backgrounds will have a great influence on pedestrian detection. Sometimes, the external lighting conditions will also affect pedestrian detection, and the resolution in dim environment is greatly different from that in bright environment. Even sometimes, the complex background color is very similar to the color of human clothes, and the shape of objects in the background is also very similar to that of human body, which easily leads to false detection or missed detection, which makes human body detection very difficult. In this paper, the target detection algorithm can be applied to the regression clustering of key points of pedestrian prediction in the pedestrian attitude estimation algorithm and can also be applied to the detection of pedestrian targets in the pedestrian re-recognition algorithm. The commonly used target detection algorithms are divided into single-stage and double-stage ones.

If real coordinate values are used directly for coordinate regression of key points in network training, the network will converge slowly or not at all, so the heat map of coordinate points is commonly used as the training label. A two-dimensional Gaussian surface is essentially what a heat map is. The human posture key point prediction network’s final prediction output is N feature maps, each of which represents a Gaussian surface and represents the prediction of N key point types. The heat maps of all of the network’s final key points are superimposed and synthesized into a single image. The processing ability of abnormal positive and negative samples will be poor because the triple loss function judges the distance between a positive and a negative sample and the anchor sample in the training process. At the same time, the triple loss function updates the gradient with a triple pair, which lacks finer constraints between positive and negative samples.

Adopt the bottom-up analysis method. That is, the method of detecting all the joint points and the connected areas of the joint points in the picture, and then judging which part belongs to the same pedestrian instead of the traditional top-down method. That is, firstly, by detecting the presence of pedestrians in the picture, the key points of each person’s bones will be detected. Through two network branches, detection and association matching of joint points are carried out simultaneously. In the traditional method of using the triple loss function, an anchor sample and a
positive sample from the positive sample set are chosen first, followed by a negative sample from the negative sample set, and finally three samples form a triple pair. The ranking function’s training sample pair selects one anchor sample and a number of positive samples from the positive sample set, followed by a number of negative samples from the negative sample set. Because the negative samples in the negative sample set may come from different categories, the training will involve comparing the relationships between samples from multiple categories, which is more stable than the strategy of optimizing only one pair of positive and negative samples at a time using the triple loss function. Simultaneously, it makes full use of the information shared by positive and negative samples, and to some extent, it solves the problem of sample imbalance. The key point of accurate prediction in the attitude estimation algorithm network is that the previous stage’s confidence diagram and the affinity field output by the branches are replicated as the input for the next stage’s update prediction, and the formula is as follows:

\[ c_1(t) = c_{11} + (c_{1T} - c_{11}) \frac{t}{T}, \]

\[ c_2(t) = c_{21} + (c_{2T} - c_{21}) \frac{t}{T}. \]

At each stage, a weighted loss function \( P \) is added to each branch for supervision.

\[ P^t = \left(P^t_{g_1}, P^t_{g_2}, \ldots, P^t_{g_d}\right), \]

where \( P^t_{g_i} = P^t_{g_i}(1 + 0.5\eta), i = 1, 2, \ldots, d, \)

and \( \eta \) is the corresponding predicted value. Add the weighted losses of the two branches together as the final loss function \( p \):

\[ p = \begin{cases} k & \sigma^2 < \sigma^2_{g_i}f(P^t_{g_i}) > f_d \\ 0 & \end{cases}. \]

The confidence diagram of key points adopts Gaussian distribution function, and the places affected by two key points take larger values at the same time.

The density of virtual labels (expressed by \( n \)) and the appropriate threshold value are selected when excluding small probability positions. The larger the \( n \), the denser the virtual tags, the more information used to assist in the positioning of the tags to be positioned, and the higher the positioning accuracy in theory. When \( n \) reaches a certain value, however, the positioning accuracy does not improve. Because there is little difference between the signal strength and location information of adjacent virtual tags when the virtual tags are dense to a certain extent, more similar virtual tags will be found at the end of the algorithm to assist in the positioning of the tag to be positioned, which will not only improve the positioning accuracy but also increase the amount of calculation. The model is optimized better and better as the number of model parameters or complexity increases, and the objective function gets lower and lower until it approaches zero. However, at some point during the training process, the cross-validation error will rise, and the model’s generalization error will rise as well. The more data you have, the more difficult it is to optimize your model. As a result, more data is frequently associated with models that perform better.

It is assumed that the linearly separable training sample set is formula (6), where \( x_i \in R^d \), the category to which the feature belongs \( y_i \in \{-1, 1\} \), and the linear discriminant function in the dimension space is formula (7).

\[ h(x) = w \cdot x + b, \]

\[ \begin{cases} h(x_i) \geq 1 & \text{When } y_i = 1, \\ h(x_i) \leq -1 & \text{When } y_i = -1, \end{cases} \]

which is

\[ y_i(w \cdot x_i + b) \geq 1. \]

After normalizing the discriminant function, after using the classifier to classify all the training samples, the correct classification result can be obtained.

\[ p = \begin{cases} k & \sigma^2 < \sigma^2_{g_i}f(P^t_{g_i}) > f_d \\ 0 & \end{cases}. \]

Obviously, the classification interval is the maximum interval, which is \( 2||u|| \).

Because the self-encoder can try to copy the input to the output after training and then obtain the main feature information of the original data, but it cannot effectively eliminate the influence of noise in the data. These influences can be more effectively eliminated in the encoding process of a sparse self-encoder by adding a sparsity penalty, and a set of overcomplete basis vectors can be found to represent the input vectors. The advantage of an overcomplete basis vector is that it is more effective at detecting the structure and pattern hidden in data. Because the target only affects a few links in the monitoring area, the number of affected links is small when compared to the total number of links. As a result, sparse components are chosen to represent the input data. A thorough examination of the algorithm’s two key parameters reveals that the difficulty lies in determining the appropriate threshold value for excluding small probability positions. The algorithm for selecting the appropriate threshold has been described in the algorithm, and the analysis results show that the positioning environment must be tested first to select the optimal threshold of signal strength. If some factors interfere with the positioning environment, the transmitted signal strength may also change, and this optimization threshold must be reselected.
4. Result Analysis and Discussion

In order to verify the effectiveness of this method, the performance of this detection algorithm is tested on the data set. The default hyperparameter is 600,000 training steps. The initial learning rate is 0.001, and the learning rate drops to 0.1 times at 500,000 iterations, the momentum is initially set to 1.0, and the weight attenuation is set to 0.0008. At the same time, mosaic data enhancement and cross-validation are used to train the model, analyze and predict the possible situation before the experiment starts, and then verify that the inference accords with the final result through experiments. The loss function has poor stability when there is only one positive sample and one negative sample. The reason for this is that the loss function’s gradient backpropagation value is completely dependent on the size comparison between a positive and a negative sample. However, because the range of distance difference between them in the feature space is large in the case of large within-class variance, convergence is difficult. As a result, prior to using the loss function, a sample screening strategy will be implemented to screen the sample pairs that are difficult to train, with the sample pairs with a general difficulty being retained for training.

In order to determine the required number of training samples, this paper first adjusts the number of training samples with the preset network and obtains the loss and accuracy rate through the test samples. In this paper, we train a posture in one round, as shown in Figure 3. It can be seen that with the increasing number of training samples, the loss value keeps decreasing, the accuracy of test samples keeps increasing, and the later training samples begin to converge and tend to be stable, so as to determine the model parameters of the network.

The semantic segmentation of pedestrians is carried out by using the algorithm network in the article. The algorithm can effectively distinguish the pedestrian area from the background area and can also effectively extract the general appearance of the pedestrian, providing intuitive and effective image features for pedestrian recognition. In the process, the network tends to position the pedestrian in the center of the training sample and at the same time removes a part of the background area that is not related to classification. This makes the final network processing such as pedestrian images have a large difference in distance and a large difference in pedestrian posture. The large sample played a good role in adaptation.

The sample data in this paper tested the changes of loss under 20 kinds of learning rates and finally selected 4 more representative learning rates of 0.01, 0.05, 0.08, and 0.012 to lose convergence trends, and the number of iterations was 200, as shown in Figure 4.

It can be seen from the Figure 4 that its loss changes with the learning rate, and the improvement of learning rate accelerates the training speed of neural network. According to many experiments and comprehensive consideration of time spent, the learning rate of this training data is 0.8, and the training effect is good and stable. In order to prove that the method proposed in this paper has a relatively accurate detection effect in the pedestrian detection database, the test error balance curve is used to show the influence of two kinds of errors, miss detection rate and false detection rate, on the test effect. In this paper, the complementary aggregation of multidimensional local features of image sequences can obviously make up for the influence of occlusion noise on pedestrian distinguishing features, strengthen the expression of effective features of pedestrians, and then obtain more complete pedestrian reidentification distinguishing features. The method in this paper is compared with the methods in literature [13] and literature [14], and the results are shown in Figure 5.

It can be seen from the Figure 5 that the method proposed in this paper has the lowest missed detection rate, which proves that the method proposed in this paper reduces the missed detection rate and improves the accuracy of pedestrian detection. In order to improve the effect of attitude recognition, this paper introduces wavelet transform to process the original data. The wavelet transform has obvious advantages in analyzing time-varying signals, because it can simultaneously perform local analysis in time domain and frequency domain. Wavelet algorithm has the advantages of good filtering effect and less loss of signal details. It is widely used in signal processing and image processing, while denoising and compression of signals and images are the most widely used in these two fields. Because the selection of orthogonal basis in orthogonal wavelet is close to the actual signal itself, the noise can be easily separated by wavelet transform. Therefore, wavelet analysis has many advantages in denoising and compression. Figure 6 shows the comparison results of detection time between this method and the methods in [13, 14].

In terms of operation time, our method is faster than the other two methods, as shown in Figure 6. Our method network is relatively light, and while the operation time will increase as the number of people in the picture grows, the rate of growth is slow. However, because the calculation time of the method in the literature increases linearly with the
number of people in the picture, it is not suitable for real-world crowded scenes and can only be used for scenes with fewer people and low real-time requirements. To summarize, the pedestrian attitude estimation algorithm proposed in this paper can be used in real-world scenarios with sparse and dense crowds, and it can meet the requirements of real-time operation and high detection accuracy.

For the training of local feature generation network, the supervision information is the feature of pedestrians in a single picture, and this single-frame feature can be trained to make it approach the center of pedestrian features in feature space. The video-level features are aggregated with the local quality scores of pictures, and the use of triplet loss in supervised learning training is helpful to extract more useful local feature information from single-frame features, further enhance the interclass differences and reduce intraclass differences, thus improving the recognition accuracy of network models. The method of training features of single frame level and video level jointly can extract more robust video level discrimination features. Figure 7 is the comparison result of the pedestrian attitude estimation error between this method and the methods in [13, 14].

The Figure 7 shows that, when compared to the other two methods, this method has less error in estimating pedestrian attitude. This method’s effectiveness and applicability are demonstrated. Mask layout extraction is distinct from category labels and border attributes extraction. The extraction of mask layout is based on convolution’s pixel-level correspondence, whereas the extraction of category labels and border attributes requires the full connection layer’s connection feature classification. Calculating the distance difference between every two samples as a back-propagation gradient over and over will result in training redundancy and a longer training time. Because the size comparison of a pair of positive and negative samples in different sample pairs can happen multiple times, a large increase in the number of positive and negative samples in a sample pair has little benefit for training. This method can reduce background noise and false targets, brighten the spot at the real target position, and improve positioning accuracy to a degree. Finally, the algorithm extracts foreground coordinates, which further weakens the influence of false targets and background noise. Their average positioning error has been reduced to some extent, and the root mean square error has also been correspondingly reduced, and the positioning effect has been greatly improved. We made an experiment on the time required for pedestrian gesture recognition. Figure 8 shows
the comparison result of the recognition time between this method and the methods in [13, 14].

It can be seen that the method used in this paper takes the least time to recognize pedestrian posture, which proves the superiority of this method. Pedestrian attribute recognition can also be regarded as a multilabel classification problem. This pedestrian attribute recognition algorithm uses multitask learning to train and learn and reduces the region of interest. Through human segmentation, the main features will be focused on the human features themselves, and the influence of background areas in feature representation will be weakened. The hot spot map of key points obtained by human posture estimation is used to assist in supervising and guiding the learning direction and distinguishing local clues. The ranking function has strong adaptability in different network structures, has excellent performance in different pedestrian re-identification data sets, and has strong robustness in training for extreme positive and negative sample pairs.

The research shows that the attitude estimation algorithm adopts the bottom-up thinking by changing the steps of detecting correlation. That is, all the key points in the video frame are detected first, and the method of matching key points can get rid of the influence of pedestrian detection results. Compared with the traditional attitude estimation algorithm, it is faster and more robust in complex multiperson scenes. In addition, this algorithm extracts the features of the whole pedestrian globally, instead of extracting the features based on local features before, thus realizing the unity among the face, hand, and human body. In different image databases, this algorithm can not only reduce the space of components but also improve the accuracy of human posture estimation. It meets the requirements of real-time operation and higher detection accuracy.

5. Conclusions

At present, with the rapid development of wireless technology, mobile computing devices, and the Internet, people’s demand for pedestrian attitude estimation and recognition is increasing day by day. RFID technology is an automatic identification technology that uses radiofrequency signals for information communication. Its principle is to use radiofrequency signals to realize automatic identification of identified objects. In recent years, the RFID technology has been paid more and more attention by countries all over the
world, the application scale of RFID products has been continuously expanded, and RFID related technologies and theories have been further enriched and improved. Based on the introduction of existing algorithms, this paper improves and innovates related algorithms. A pedestrian attitude estimation and recognition algorithm based on RF data is proposed. Through theoretical analysis and experimental verification, this paper deeply studies pedestrian attitude estimation and recognition based on RF data. By changing the state of the label itself, the label is divided into several areas, so that readers can communicate with tags in different areas with different powers. That is, the reader uses different power to identify tags at different distances, instead of using the maximum power to identify all tags in the region as in the traditional algorithm, thereby greatly reducing the identification energy consumption of the algorithm. The experimental test and simulation results show that accurate data feature extraction is beneficial to capture the complete input space, reduce the over-fitting of the network and enhance the generalization ability of the network. In this algorithm, human detector and foreground highlighting algorithm are used to reduce the search space of components, and at the same time, the accuracy of classification and the accuracy of human posture estimation and recognition are improved.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors do not have any possible conflicts of interest.

References

[1] K. I. Tanaka, J. Ueno, and S. Chiba, “Estimation of a boundary to distinguish between radioactive materials and non-radioactive materials around a main steam line and a feed water line in a biological shielding wall of a BWR for decommissioning,” Progress in Nuclear Energy, vol. 93, no. 11, pp. 371–385, 2016.

[2] E. Isioma and Q. H. Anabel, “Privacy attitudes and concerns in the digital lives of older adults: Westin’s privacy attitude typology revisited,” American Behavioral Scientist, vol. 62, no. 10, pp. 1372–1391, 2018.

[3] J. Kong, H. Wang, X. Wang, X. Jin, X. Fang, and S. Lin, “Multi-stream hybrid architecture based on cross-level fusion strategy for fine-grained crop species recognition in precision agriculture,” Computers and Electronics in Agriculture, vol. 185, p. 106134, 2021.

[4] Q. Liu, L. Cheng, A. L. Jia, and C. Liu, “Deep reinforcement learning for communication flow control in wireless mesh networks,” IEEENetwork, vol. 35, no. 2, pp. 112–119, 2021.

[5] A. B. Mabrouk and E. Zagrouba, “Abnormal behavior recognition for intelligent video surveillance systems: a review,” Expert Systems with Applications, vol. 91, pp. 480–491, 2018.

[6] K. Nguyen, Y. H. Chuang, and R. F. Yu, “Roles of socio-physical environments on air quality control policy with respect to knowledge, attitude and intention,” Journal of Cleaner Production, vol. 288, no. 2, 2021.

[7] S. B. Chen, X. Yi, and B. Luo, “Action-based pedestrian identification via hierarchical matching pursuit and order preserving sparse coding,” Cognitive Computation, vol. 8, no. 5, pp. 797–805, 2016.

[8] Y. Jin, Y. Zhang, Y. Cen, Y. Li, V. Mladenovic, and V. Voronin, “Pedestrian detection with super-resolution reconstruction for low-quality image,” Pattern Recognition, vol. 115, no. 3, p. 107846, 2021.

[9] M. Ullah and B. Sarkar, “Recovery-channel selection in a hybrid manufacturing-remanufacturing production model with RFID and product quality,” International Journal of Production Economics, vol. 219, pp. 360–374, 2020.

[10] Y. Ding, “Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022.

[11] J. Souza, D. Oliveira, and D. Costa, “Electroen instrumentation in Lofstrand: dynamic force and attitude monitoring,” IEEE Latin America Transactions, vol. 18, no. 5, pp. 818–824, 2020.

[12] N. Spotorno, S. Hall, D. J. Irwin et al., “Diffusion tensor MRI to distinguish progressive supranuclear palsy from α-synucleinopathies,” Radiology, vol. 293, no. 3, pp. 646–653, 2019.

[13] M. J. Beck, J. M. Rose, and S. P. Greaves, “I can’t believe your attitude: a joint estimation of best worst attitudes and electric vehicle choice,” Transportation, vol. 44, no. 4, pp. 753–772, 2017.

[14] D. Song, R. Tong, J. Chang, X. Yang, M. Tang, and J. J. Zhang, “3D body shapes estimation from dressed-human silhouettes,” In Computer Graphics Forum, vol. 35, no. 7, pp. 147–156, 2016.

[15] V. D. Nguyen, H. Van Nguyen, D. T. Tran, S. J. Lee, and J. W. Jeon, “Learning framework for robust obstacle detection, recognition, and tracking,” IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 6, pp. 1633–1646, 2017.

[16] J. Ren and J. Han, “A new multi-scale pedestrian detection algorithm in traffic environment,” Journal of Electrical Engineering & Technology, vol. 16, no. 2, pp. 1151–1161, 2021.

[17] Y. Shen and R. S. Kwak, “Country image and product attitude: an estimation of switching costs for the Korean wave,” Journal of Korea Trade, vol. 23, no. 8, pp. 45–52, 2019.

[18] Y. N. Kim, J. H. Park, and M. H. Kim, “Spatio-temporal analysis of trajectory for pedestrian activity recognition,” Journal of Electrical Engineering and Technology, vol. 13, no. 2, pp. 961–968, 2018.

[19] A. Adnane, Z. A. Foith, and M. Mohammed, “Real-time sensor fault detection and isolation for LEO satellite attitude estimation through magnetometer data,” Advances in Space Research, vol. 61, no. 4, pp. 1143–1157, 2018.

[20] L. Benziane, A. E. Hadri, and A. Seba, “Attitude estimation and control using linearlike complementary filters: theory and experiment,” IEEE Transactions on Control Systems Technology, vol. 24, no. 6, pp. 2133–2140, 2016.

[21] S. Mi, H. Han, C. Chen, J. Yan, and X. Guan, “A secure scheme for distributed consensus estimation against data falsification in heterogeneous wireless sensor networks,” Sensors, vol. 16, no. 2, p. 252, 2016.

[22] Z. QIU, H. QIAN, and G. WANG, “Adaptive robust cubature Kalman filtering for satellite attitude estimation,” Chinese Journal of Aeronautics, vol. 31, no. 4, pp. 194–207, 2018.
[23] Q. Dai and L. F. Sui, "Quaternion constrained cubature Kalman filter attitude estimation based on Gaussian mixture model," *Yuhang Xuebao/Journal of Astronautics*, vol. 39, no. 8, pp. 913–919, 2018.

[24] R. P. Patera, "Attitude estimation based on observation vector inertia," *Advances in Space Research*, vol. 62, no. 2, pp. 383–397, 2018.

[25] A. Chiella, B. Teixeira, and G. Pereira, "Quaternion-based robust attitude estimation using an adaptive unscented Kalman filter," *Sensors*, vol. 19, no. 10, p. 2372, 2019.

[26] X. Tong, Z. Li, G. Han et al., "Adaptive EKF based on HMM recognizer for attitude estimation using MEMS MARG sensors," *IEEE Sensors Journal*, vol. 18, no. 8, pp. 3299–3310, 2018.

[27] M. Kiani, A. Barzegar, and S. H. Pourtakdoust, "Entropy-based adaptive attitude estimation," *Acta Astronautica*, vol. 144, no. 3, pp. 271–282, 2018.

[28] L. Xie, O. Elnahas, Q. Zhao, and Z. Quan, "Data-driven RF transmit power calibration for wireless communication systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 721–725, 2020.

[29] X. Yang, W. Ni, W. Yan et al., "3-D electromagnetic-model-based absolute attitude estimation using a deep neural network," *Remote Sensing Letters*, vol. 12, no. 10, pp. 1015–1024, 2021.