

## *Retraction*

# **Retracted: Feature Recognition of Human Motion Behavior Based on Depth Sequence Analysis**

### **Complexity**

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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

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

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

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

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

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

### **References**

- [1] G. Zhao, "Feature Recognition of Human Motion Behavior Based on Depth Sequence Analysis," *Complexity*, vol. 2021, Article ID 4104716, 10 pages, 2021.

## Research Article

# Feature Recognition of Human Motion Behavior Based on Depth Sequence Analysis

**Guangyong Zhao** 

*School of Physical Education and Health, Linyi University, Linyi, Shandong 276005, China*

Correspondence should be addressed to Guangyong Zhao; [zhaoguangyong@lyu.edu.cn](mailto:zhaoguangyong@lyu.edu.cn)

Received 20 April 2021; Revised 24 May 2021; Accepted 24 June 2021; Published 5 July 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Guangyong Zhao. 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.

The current research on still image recognition has been very successful, but the study of action recognition for video classes is still a challenging topic. In this work, we propose a random projection-based human action recognition algorithm to address the lack of depth information in color information (RGB video frames) that is not easily affected by environmental factors such as illumination and the lack of ability to recognize actions along the direction of view. A network structure is designed to take the obvious advantage of long- and short-term memory networks for controlling and remembering long sequences of historical information. The network structure in this paper is constituted by multiple memory units. At the same time, this paper constructs the spatial features, temporal features, and depth features of the three recognition stream outputs into a feature matrix, whose feature matrix is divided into multiple temporal segments according to the temporal dimension, then inputs them into the network layer in order, and achieves the fusion of the feature matrix in this paper according to their correlation characteristics on the temporal axis. Here, we proposed the concept of random batch projection operators. This basically uses as much sublimation information as possible to improve projection accuracy by randomly selecting several subdependencies as projections defined during projection. A compressed sensing design of human motion acceleration data for low-power body area networks is proposed, and the basic idea and implementation process of compressed sensing theory for human motion data compression and reconstruction in wireless body area networks are introduced in detail.

## 1. Introduction

At present, human behavior action recognition based on the random projection algorithm is an emerging branch in the field of pattern recognition, whose basic idea is to collect the motion signals of human activities through wearable inertial sensors and transmit them to remote data processing centers through wireless communication technology and then to classify and recognize them according to some pattern recognition algorithms after feature extraction and selection. Human behavior movement is an external activity that is manifested through the domination of human thoughts, a gesture of movement, or stillness in which the head, elbows, arms, legs, feet, and other human parts coordinate with each other and is used as one of the ways of interaction between humans and the environment [1]. While the human brain has outstanding ability to filter and understand the

information conveyed by human movements through vision, machines still do not have this outstanding performance at present [2]. Thus, how to improve the performance of machines for intelligent recognition of human actions has become one of the popular research studies in the field of machine vision [3]. In addition, studying the mechanism of machine understanding of human actions and constructing the expression features and recognition algorithms of visual behaviors are beneficial to new discoveries and research on the mechanism of visual cognition of the human brain. Many new human action recognition methods have been proposed and implemented [4]. This is not only to provide experimental evidence for research related to higher-order semantic information of the brain but also to feedback and promote scientific theories through experiments and has a great scientific significance to promote the accelerated development of the artificial intelligence field [5]. Human

behavior recognition is not only theoretically important but also widely applied in real life. With continuous innovations in communication technology and hardware technology, the human behavior recognition has been enhanced by artificial intelligence techniques, such as virtual reality, video analysis, identification, physical interaction, human and machine interaction, intelligent surveillance, medical diagnosis, and so on. It has the potential for wide application [6–10].

The recognition of human behavior is very active in human behavior recognition research. In the worldwide propaganda of the extremist terrorist network, the importance of the action recognition technology in the video analysis field which can effectively prevent the diffusion of various terrorist obscene videos on the online basis increases [11]. At the same time, in 2020, in the global new coronavirus outbreak situation, action recognition technology in the field of intelligent monitoring also highlights the importance. Intelligent monitoring can not only detect the body temperature of the human body but also intelligently monitor and identify the behavior of the quarantined personnel, playing a role in preventing suspected virus carriers from escaping or spreading the virus. Such intelligent monitoring capabilities will greatly save the country's financial, material, and human resources [12]. At present, although the research on still image recognition has achieved great success, the behavioral action recognition of video class is still a challenging topic, as shown in Figure 1. In the field of behavior action recognition for RGB video class, RGB video is not only rich in detail information and easy to obtain, such as the optical flow information in RGB video, but also lacks depth information and is vulnerable to environmental factors such as lighting changes. It would be challenging to use only the rich detail information of RGB to model spatio-temporal channels and to estimate the real 3D spatial movements, such as those along the direction of view (perpendicular to the image plane), for which there is a lack of certain recognition ability. Under the circumstance that various environmental factors are relatively complex ring scenes and the rich detailed information of RGB is not distinguishable for human actions, taking the depth information in the scene becomes a popular idea of the human action recognition method [13]. Earlier the depth information in the acquired scenes was captured by laser sensors. However, the depth values of many regions in the depth information captured by the laser sensor were unknown, so there were no corresponding depth values on many pixels in the synchronized color information [14, 15]. In November 2011, a cheap, high-quality depth camera based on the principle of infrared scattering structured light appeared to capture the depth information in the scene. The idea of human motion recognition has become a reality.

Compared with the traditional RGB video, the depth video captured by the depth camera is not affected by external environmental factors such as light changes and can be used normally even in dark or no natural light environment, showing significant performance advantages in human action estimation in three-dimensional space and

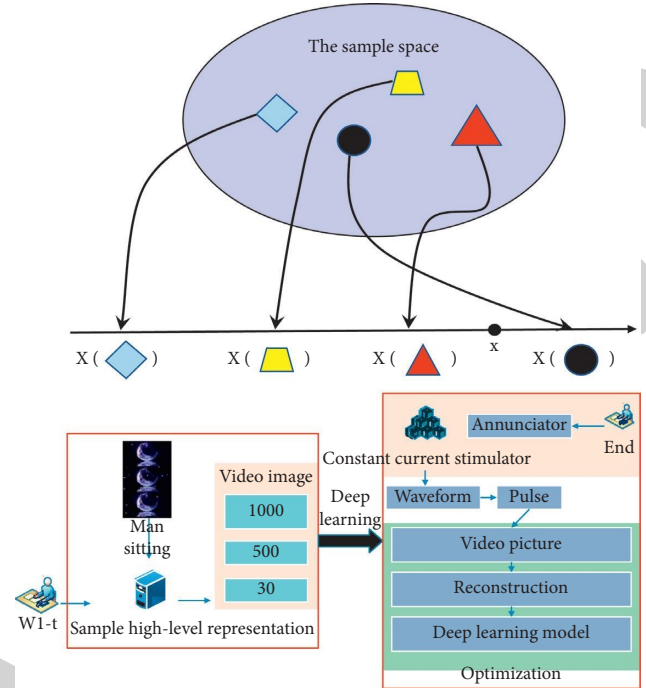


FIGURE 1: The operating process of image recognition.

having a certain degree of mitigation in the work tasks such as target detection and segmentation in the field of traditional visible video research. The adoption of depth video sequences in the field of human behavioral action recognition facilitates effective modeling and description of the correlation information between temporal motion and spatial representations. Although depth information is not affected by changes in ambient lighting, there is a large gap between the output resolution of current depth cameras and that of visible cameras, which are generally low in resolution and lack rich detail description information about objects in the scene, retaining only the relative position information in 3D space. Therefore, there are limitations in using RGB video or depth video for action recognition. Based on the above research background and combined with the actual scene application requirements, this paper adopts the current mainstream convolutional neural network, fuses the color information (RGB video frames) and optical flow information obtained from RGB video sequences, and obtains the depth information from the synchronized depth video sequences to build a multistream convolutional neural network for human action recognition in video class. The method used in this paper is optimized and improved by combining the advantages and shortcomings of existing methods.

The rest of this article is organized as follows. Section 2 is devoted to the analysis of related works; the distributed random projection algorithm is optimized in Section 3; in Section 4, a simulation experiment of the human movement action recognition under random projection is carried out; Section 5 mainly discusses the results; Section 6 summarizes the full text.

## 2. Related Works

In the field of human action recognition, as in the development of various neural network fields, domestic and foreign researchers have recently focused on the use of convolutional neural networks in the research methods of human action recognition. However, the performance results achieved by convolutional neural networks in the field of human action recognition have not increased significantly compared with other fields [16–18]. As described above, the new effective human action recognition method has been focused on improving and optimizing the original method. Recently, Zhou [19] proposed that the conventional recognition method of human motion is a recognition method with feature extraction by hand. In 2019, Dreisigmeyer [20] stored the motion history image (MHI) and the motion energy image (MEI) of human behavior features statically and then used the Marxian distance principle to match and classify the stored motion models with the MEI and MHI computed from the new test video. In more early work, Koopman et al. [21] proposed a human action recognition method for detecting spatio-temporal feature points, mainly by tracking the trajectories of spatial feature points for human action recognition, while showing that the models of the holistic representation method are difficult to match accurately and effectively and have certain drawbacks. The local representation method is to use the local region of human action pose in the scene to represent the human behavior features. In 2005, they proposed the spatio-temporal interest points (STIPs) behavior action recognition method, which laid the foundation for the later local representation method of human action recognition, whose local representation method of human action recognition will follow three steps: spatio-temporal interest points detection, local descriptor extraction, and local descriptor aggregation and representation. Sadeghi proposed an efficient dense and scale-invariant spatio-temporal interest point detection method, whose interest point detector has advantages in terms of repeatability, accuracy, and speed compared with the previous ones [22].

Compared with traditional methods, the random projection algorithm generally outperforms them and solves the difficulty of manually extracting robust features under different environmental changes and also allows efficient training of high-dimensional network models with robustness using huge data sets. Due to the above advantages, deep learning is gradually becoming the mainstream development direction in the field of video-based human action recognition. In 2019, Wei et al. [23] proposed a human behavioral action recognition model based on three-dimensional convolutional neural network (3D CNN), which adds a temporal dimension to the original 2D CNN, and this method outperforms the 2D CNN method in certain performance [24]. They proposed a cross-database approach for behavior recognition on a 3D Markov random field (MRF)-based framework using motion, appearance, and saliency to obtain the confidence of each pixel in the video as a foreground region and use it for action

recognition. It combined a CNN network and LSTM networks to construct a new network model long-term recursive convolutional network (LRCN), which is end-to-end trainable and is a suitable method for human behavioral action recognition for large-scale visual learning. In 2019, Ke et al. [25] proposed a human behavioral action recognition method based on acceleration signal and evolved radial basis function (RBF) neural network, which integrated the recursive genetic algorithm (HGA) into the radial basis function (RBF) neural network for training and achieved better recognition results.

## 3. Distributed Random Projection Algorithm

*3.1. Distributed Optimization Algorithms for Multi-Intelligent Systems.* Although the distributed subgradient algorithm has wider applications than the gradient method, it still inevitably has to make concessions in the convergence rate compared with the gradient method. How to further improve the convergence rate of the distributed subgradient algorithm is the focus of this paper. In the previous chapter, we mentioned the concept of batch random projection operator, which essentially utilizes as much information of subconstraints as possible to improve the projection accuracy by randomly selecting multiple subconstraints as the projection set at the time of projection. Most of the existing distributed subgradient literatures use only the information of individual current moment subgradient in the algorithm design, while the information of subgradient before the current moment is not utilized.

Consider the following multi-intelligent system optimization problem for switching networks with random time delays containing  $n$  intelligences:

$$\min f(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{f_i(x) + g_i(x)}{2}, \quad (1)$$

$$x \in X = \bigcup_{i=1}^n X_i,$$

where the communication network of the system at moment  $k$  is denoted as  $G(k) = (V, E(k))$ ,  $V$  is the vertex set containing  $n$  individuals,  $E(k)$  is the edge set, its corresponding weight adjacency matrix is  $A(k)$ , the  $i$ -th row and  $j$ -th column element of  $A(k)$  is denoted as  $a_{ij}(k)$ , the local cost function of individual  $i$  is denoted as  $f_i(x)$ ,  $x \in Rm$ , and the overall constraint set  $X$  is given by local constraints  $X_i$  for all individuals where  $i = 1, \dots, n$  is obtained by taking the intersection set, i.e.,  $X = \bigcap_{i=1}^n X_i$ . The cost function  $f_i(x)$  and constraint  $X_i$  of an individual are known only by itself and not shared with other individuals. We assume that the local constraints  $X_i$  of any individual  $i$  are intersected by a certain number of subconstraints, i.e.,  $\forall i \in V$ ; there exists a series of subconstraints  $X_{j_i}$ ,  $j \in I_i$  such that  $X_i = \bigcap_{j \in I_i} X_{j_i}$ , where  $X_{j_i}$  is a closed convex set and  $I_i$  is called the index set of  $i$ , which contains the ordinal numbers of all subconstraints of individual  $i$ . For example, if individual  $i$  has 5 subconstraints, then  $I_i = \{1, 5\}$  and  $X_i = \bigcap_{j=1}^5 X_{j_i}$ . According to the previous work [26], the overall constraint and sequence can be referred to the assumptions as follows.

*Assumption 1.* The overall constraint set  $X$  is bounded, i.e., there exists a constant  $C_B > 0$  such that

$$\bigoplus_i x < C_B. \quad (2)$$

*Assumption 2.* Let both  $\{\gamma_k\}$  and  $\{\alpha_k\}$  be sequences of nonnegative convergence to zero satisfying

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \eta_i = \frac{\bigoplus_i (f_i(x) - g_i(x))}{\bigoplus_i (f_i(x) + g_i(x))}. \quad (3)$$

Also assume that the default communication time between two individuals is 1 second, which is not included in the delay. Assume that the positive integer  $b_{ij}$  is the delay value on edge  $(i, j)$ , and if  $(i, j)$  has no delay, then  $b_{ij} = 0$ . Unlike the fixed delay model, we assume that the delay value  $b_{ij}$  for any delay edge  $(i, j) \in E(k)$  is a bounded random positive integer that obeys a uniform distribution from 1 to a positive integer  $B$ , i.e., the time consumed by the message sent by  $j$  to  $i$  varies from  $2s$  to  $(B + 1)s$  (as shown in Figure 2).

*3.2. Distributed Subgradient Batch Random Projection Algorithm.* We propose a distributed multistep gradient randomized projection algorithm to solve the problem. At moment  $k$ , we introduce the variable  $v_i(k)$  for any individual  $i$ . The initial value of the algorithm is chosen as  $x_i(0) \in X$ ,  $v_i(0) = 0$ , and the algorithm takes the following concrete form, which has been discussed in the previous work [27]:

$$\begin{aligned} v_i(i+1) &= \Gamma_i a_{ij}(i) + \alpha_i d_i(k), \\ y_i^0(i) &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{a_{ij}(i) + \varphi_{ij}}{a_{ij}(i) - \varphi_{ij}}, \\ y_i^p(i) &= \lim_{n \rightarrow \infty} P_{X_i^p} \sum_{i=1}^n a_{ij}(i), \quad p = 1, \dots, n, \end{aligned} \quad (4)$$

$$x_i(k+1) = y_i b(i),$$

where  $\alpha_k$  is the iteration step size and  $\gamma_k$  is the weighting parameter.

For network expansion with random time delay, we likewise assume that the edges of the network with time delay throughout the switching process  $k_{+\infty} = E(k)$  can be arranged in the order  $(i_1, j_1), (i_2, j_2), \dots, (i_t, j_t)$ , and the random time delays of the corresponding edges are  $U(1, B)$ ,  $h = 1, \dots, t$ . At each time delay edge  $(i_h, j_h)$ ,  $h = 1, \dots, t$  are increased by  $B$  virtual individuals. In this way, any edges  $(i_h, j_h)$  with random time delays are replaced with new paths  $(i_h, dB)$ . Also add the local cost function  $f_d B_i h_{jh} = 0$ ,  $h = 1, \dots, t$ , set the state  $x_{jh}(k)$ , and the local constraints of these virtual individuals are also set to be the same as the local constraints of individual  $x_{jh}$ , and the total number of individuals added is  $b = t_B$ . Thus, the original problem with random time delay is equivalent to the following new problem without time delay.

Let the set of network nodes for problem be  $V = V(t_h = 1 \{d_{1i} h_{jh}, \dots, dB_{ih} j_h\})$  with the switching network graph  $G(k) = (V, E(k))$ , the adjacency matrix corresponding to  $G(k)$  is  $Q(k)$ , and the elements of the  $i$ -th row and  $j$ -th column of  $Q(k)$  are noted as  $q_{ij}(k)$ . At the moment of  $k$ , regardless of the specific value of the time delay of the edge  $(i_h, j_h)$ , we expand the dimension according to the maximum time delay and get  $Q(k)$  after expanding  $A(k)$  to  $B$  dimensions (first expand the dimension of the first edge  $(i_1, j_1)$ , then the matrix  $A(k)$  becomes  $Q_1(k)$ , then expand the dimension of  $(i_2, j_2)$  to get  $Q_2(k)$ , and so on until the last edge  $(i_t, j_t)$  and finally the matrix  $Q_t(k)$  is the adjacency matrix  $Q(k)$  of the network graph  $G(k)$  obtained after complete dimensionality expansion of the graph  $G(k)$ ).

*3.3. Distributed Approximation of Random Noise.* Can the convergence rate of the subgradient algorithm be improved even further? Theoretically, it is possible to do so if no restrictions are imposed on the optimization problem, but due to the limitations of the general multi-intelligent system model, the subgradient information is not shared among individuals, and the improvement of the subgradient method has always started from the improvement of the subgradient information of individuals themselves. In fact, when the system considers some practical factors such as time delay and noise, many literatures have improved the stability of the system by increasing the communication bandwidth or adding quantizers to support the effective execution of the algorithm, which can be found in the precious work [28]. Most of the current literatures studying the distributed optimization problem of multi-intelligent systems with meaningful noise consider the effect of noise present in the communication process of individual information  $x_i(k)$ , and some literatures have also studied the case of noise interference in the objective function  $F$  [29]. This chapter focuses on the effect of individual states on the multistep approximate subgradient stochastic projection algorithm in the presence of random noise in the communication process and conducts convergence analysis and compares the convergence of the three algorithms in this paper in the presence of random noise and the degree of influence of random noise on the algorithm in different ranges through numerical simulations.

For any  $k \geq s \geq 0$ , the transfer matrix of  $Q(k)$  is noted as  $\Phi(k, s) = Q(k)Q(k-1) \dots Q(s)$ , and note that  $\varphi_i(k) = y_{1i}(k) - y_{0i}(k)$ , and pushing the  $k$  moments back to the zero moment, we have

$$\begin{aligned} v_i(k+1) &= \lim_{n \rightarrow \infty} \left( \prod_{i=0}^{n-1} \gamma_T \right) \sum_{i=1}^n [\Phi(k, 0)] \\ &+ \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} \left[ \prod_{i=1+1}^{n-1} \gamma_T \right] \alpha_\Gamma \left[ \lim_{n \rightarrow \infty} \sum_{j=1}^{n+b} [\Phi(k, j+1)] \right]. \end{aligned} \quad (5)$$

Noting that  $v_i(0) = 0$ , the above equation takes the paradigm, noting that  $\iota = \max\{\gamma_1, 1\}$ , and combined with the fact that  $\gamma_k$  is a decreasing sequence yields

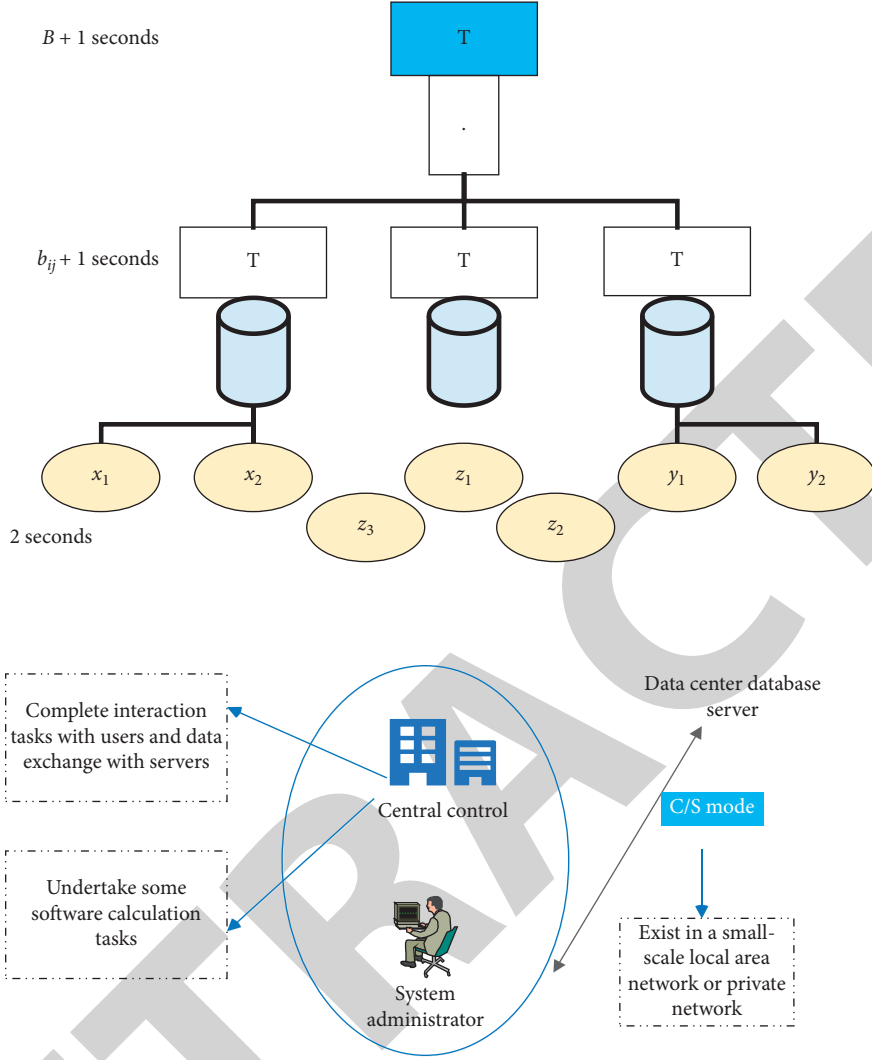


FIGURE 2: Random time delay model: there are various possible times for a message to travel from  $j$  to  $i$ .

$$\begin{aligned}
 \|v_i(k+1)\| &\leq \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} \left[ \prod_{i=\Gamma+1}^{n-1} \gamma_T \right] \alpha_T \\
 &+ \left( \prod_{i=0} \gamma_T \right) \lim_{n \rightarrow \infty} \sum_{i=1}^n [\Phi(k, 0)] \quad (6) \\
 &\leq \frac{\gamma}{\gamma-1} \left[ \lim_{n \rightarrow \infty} \sum_{j=1}^{n+b} [\Phi(k, j+1)] \right].
 \end{aligned}$$

By Lemma and  $X \subset X_{\omega p_{ii}}$ ,  $p = 1, \dots, b$ , for any  $\forall x \in X$ , the following relation holds:

$$\|v_i(k+1) - x_i\| \leq |y_i^b(k) - x_i| \leq \frac{\gamma}{\gamma-1}. \quad (7)$$

## 4. Human Movement Action Recognition under Random Projection

**4.1. Data Preprocessing for Deep Motion.** Although color information is rich in detail and easily accessible, color information lacks depth information that is not easily affected by environmental factors such as illumination and lacks the ability to recognize actions along the line of sight direction in the previous work [30]. In order to solve the above problem, this paper combines the optical flow information and the color information with color information. In this paper, we investigate three independent flow detection networks (space flow, time flow, and depth flow) based on LES net 101 and create characteristic features, time features, and depth features from three detection flow

networks as feature matrices and their features by LSTM. The time stream network channel data preprocessing is shown in Figure 3. Fusion is performed and finally classified into softmax layers with complete connectivity. Color information is susceptible to environmental factors such as lighting, but color information is rich in detail and easy to obtain. Therefore, in this paper, the RGB video series color information is used as input of the spatial streaming network channel, and the abundant detail information is utilized sufficiently. Many references show that RGB images can be used as input for spatial streaming to achieve very good performance. The recognition model of this paper is performed for the commonly published data set UTD-MAD-, and the spatial streaming network inputs each frame of the RGB video series sampled by the data set UTD-MAD.

**4.2. Time Stream Network Channel Data Preprocessing.** In this paper, in order to test the effect of different sampling strategies on the human action recognition effect of the spatial flow network, therefore, the experiments in this paper are tested and compared using three different sampling methods for RGB video sequences, as shown in Figure 4. The maximum sampling method is adopted because the detection performance is not improved even if the number of frames of the sampling space of the spatial streaming network is increased. Conversely, the length of each image of RGB in the UTD-MAD data set does not exceed 4 seconds, and the number of frames in each video series does not exceed 70 frames. Therefore, it is not desirable to test the number of interval frames of the video. In this experiment, the RGB video contained in the data set is tested with the full sampling method, and the first 3 frames and the last 3 frames of each RGB video series are deleted. In the beginning and end of the picture, the subjects were almost resting, and since only the small movement of the limbs was made, the motion characteristics were not affected so that the first and last frames were deleted and the situation was not moved.

In the process of spatial streaming network learning and modeling, in order to prevent the problem of overfitting, four data enhancement methods such as horizontal flip, angular rotation, staggered transformation, and translational transformation are used simultaneously for RGB video frames in this paper. The data enhancement techniques not only increase the variability of the samples and enlarge the size of the input data but also enhance the generalization ability of the network model, which can prevent the overfitting of the network to a certain extent. In order to test the impact of these data enhancement techniques on the performance of human action recognition on the Res Net101 spatial flow network model, five data enhancement techniques were used to compare the accuracy of human action recognition in this paper. As can be seen from the table, the recognition accuracy of the spatial flow network model in this paper is degraded to different degrees when any one of the data enhancement techniques is missing, so four data enhancement techniques will be used simultaneously for the spatial flow network input data in this paper.

**4.3. Deep Streaming Network Channel Data Preprocessing.** The importance of optical flow information in describing the motion information of human actions, as well as the simplicity and practicality of optical flow information in scientific experimental work, makes it the preferred motion feature representation in the field of human behavior recognition compared with the previous work [31]. In order to provide optical flow information to the network framework of this paper, there are different choices of optical flow algorithms in this paper. Most of the research studies use two optical flow algorithms, Blox or tv-l1, but there are still several differences in their optical flow algorithms. Experimental results show that the performance of tv-l1 is superior to that of the two optical flow algorithms. Therefore, in this paper, the optical flow information is extracted from the RGB video sequence using the tv-l1 optical flow algorithm in the unspecified case. The specific objective functions of the tv-l1 optical flow algorithm are shown by the following equation:

$$E_{\mu} = \oint_L \Delta U dx + \frac{1}{2\theta} (U - V)^2 dy. \quad (8)$$

As shown in Figure 5, in this paper, the horizontal and vertical optical flows are calculated, and the horizontal and vertical components of the tv-l1 optical flow are adjusted to adjust all the values of optical flow accuracy of less than 0.4 and all values of more than 8 frames. The network of this document needs to be linear transforming of optical flow information before it is used as the input of the temporal flow network channel of this document and simultaneously generates optical flow into two Gree scale images. In order to effectively utilize the motion information of the RGB video sequence, the temporal network channel of this paper has a horizontal optical flow of 10 frames and a vertical optical flow of 10 frames and 20 dense optical flow images according to continuity on the operation time axis.

## 5. Results and Discussion

The number of channels of the first convolutional layer of the original Res Net101 network is 3. However, the input of the temporal network in this paper is a stacked 20-layer dense optical flow image, so the number of channels of the input features is 20, which is not consistent with the number of channels of the temporal network in this paper. Therefore, the cross-modal cross-training method will be applied to this network by first calculating the average weights of the 3 channels in the first convolutional layer and then replicating the calculated average weights into 20 copies and using them as the weights of the 20 channels in the first layer of the input layer of the temporal network, while keeping the weight parameters of the other layers of the temporal network unchanged. The depth map can be used to represent the variation characteristics of the distance from each part of the human body to the sensor. A three-sided projection of the depth sequence frame has been obtained by three orthogonal Cartesian planes to represent the motion characteristics of an action. Due to the computational simplicity, a similar

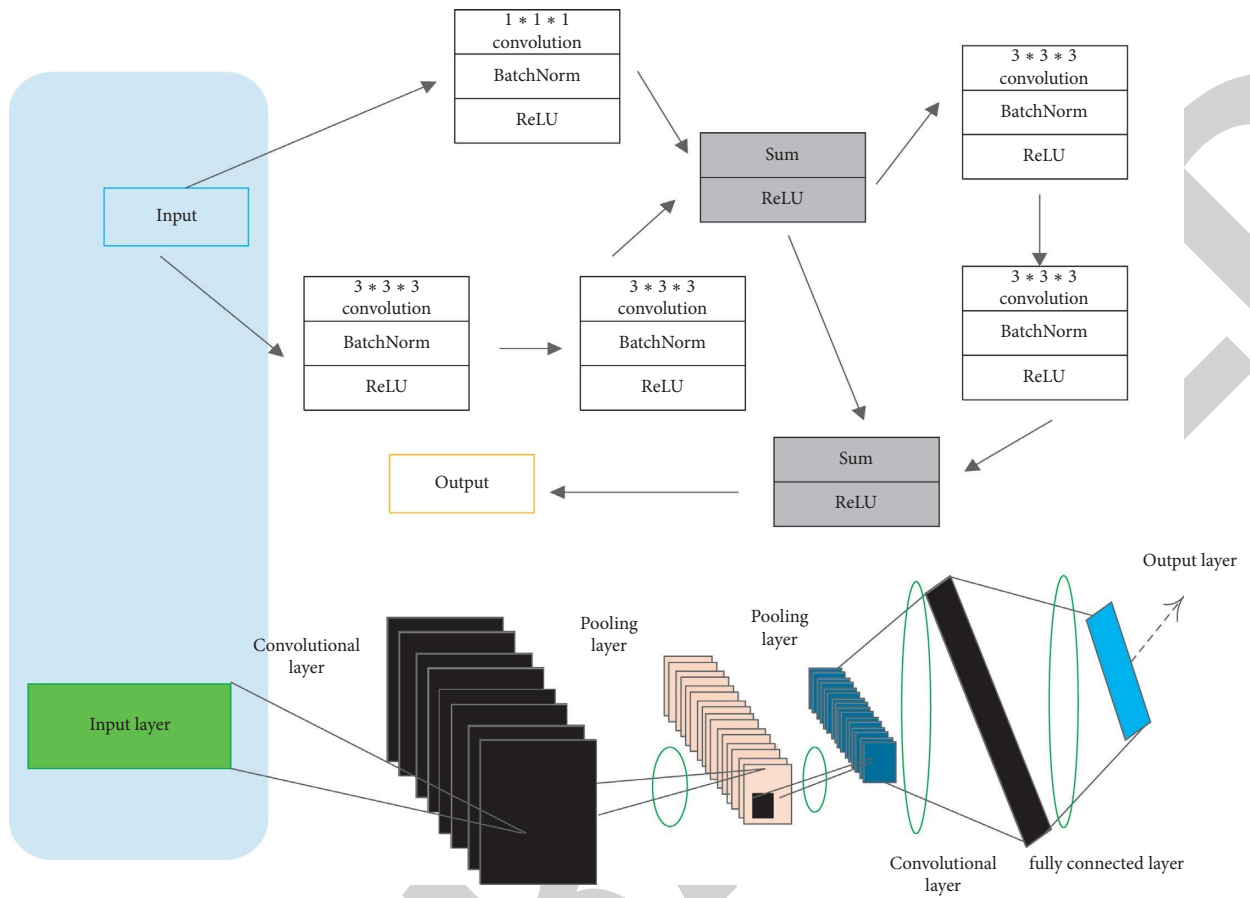


FIGURE 3: Time stream network channel data preprocessing.

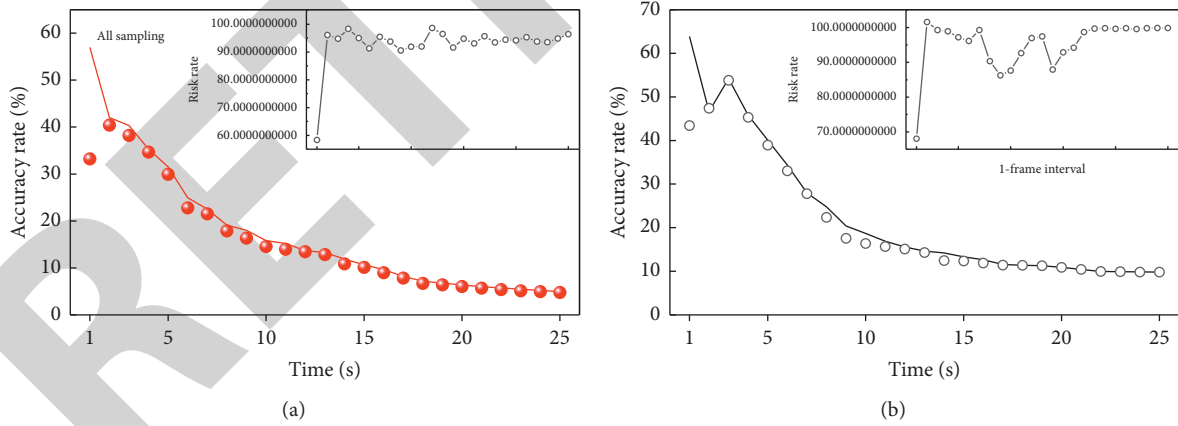


FIGURE 4: Continued.



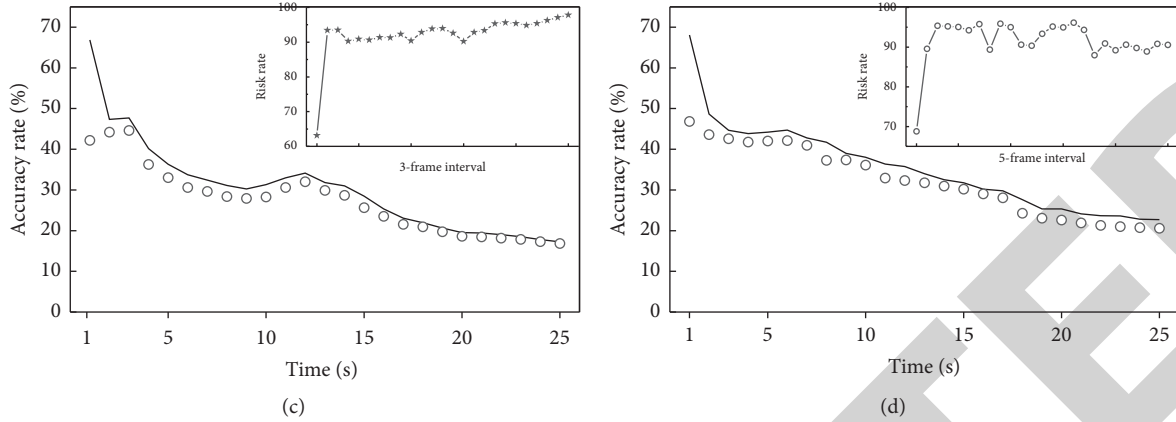


FIGURE 4: Accuracy of action recognition with different sampling methods.

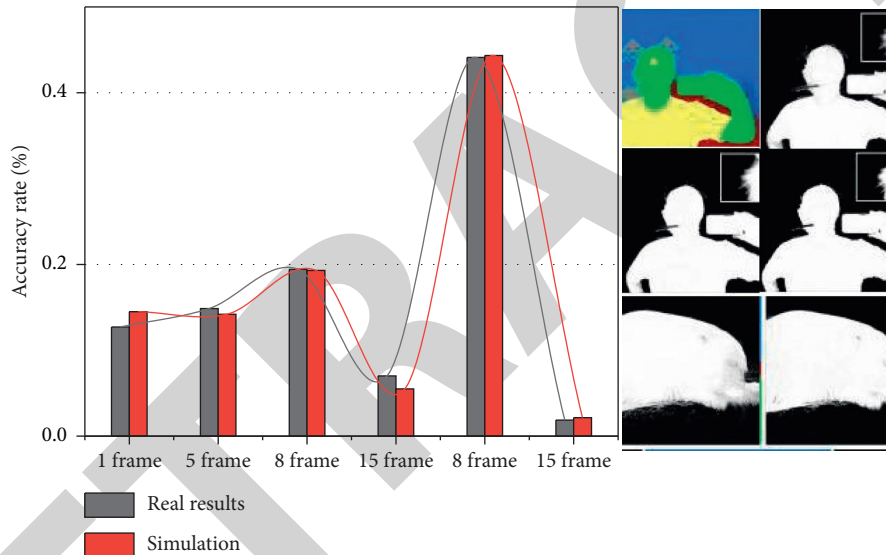


FIGURE 5: Optical flow diagram corresponding to two consecutive RGB images.

approach is taken in this paper. The method in this paper is to generate the two-dimensional projection maps of the depth sequence frame in its three planes for the front view  $f$ , the side view  $s$ , and the top view  $t$ , respectively, whose projection maps are shown in Figure 6. In the  $x$ -axis direction, the number of the recognition points can reach to 20000. And in the previous work [12–14, 26], the number can be less than 3000, which means that by this method, we can generate a recognition map over 10 times compared with the traditional single-correlation motion system. As an example, for different positions in  $x$  and  $y$ , the planes would be quite different, and when the  $x$  position is in about 550 cm and  $y$  position is in about 20 cm, the intensity of the conjunction would be around the max value.

In the work, instead of using principal component analysis (PCA) for dimensionality reduction, kernel entropy component analysis (KECA) was used for dimensionality reduction. KECA was chosen because principal component analysis is based on second-order statistics and is optimal

only for Gaussian distributions, while KECA is based on information theory and can handle data with non-Gaussian distributions, as shown in Figure 7. A bounding box is set to extract the nonzero regions in each DMM as foreground; finally, the features of motion are efficiently captured from the three projection views by DMM computation. By the inputting of the 3-D recognition system, a larger network of the RGB image can be obtained. Based on the multi-operation, the input of the temporal network in this paper is a stacked 20-layer dense optical flow image, so the number of channels of the input features is 20, which is much faster than the number of channels of the temporal network in the previous work [26]. Conversely, every RGB image in the utd-mad dataset is not exceeded for four seconds, and the number of frames in each video series does not exceed 70 frames. Therefore, you do not want to test the number of interval frames in the video. In this experiment, the RGB video contained in the dataset is tested with a complete sampling method, removing the first three frames and the

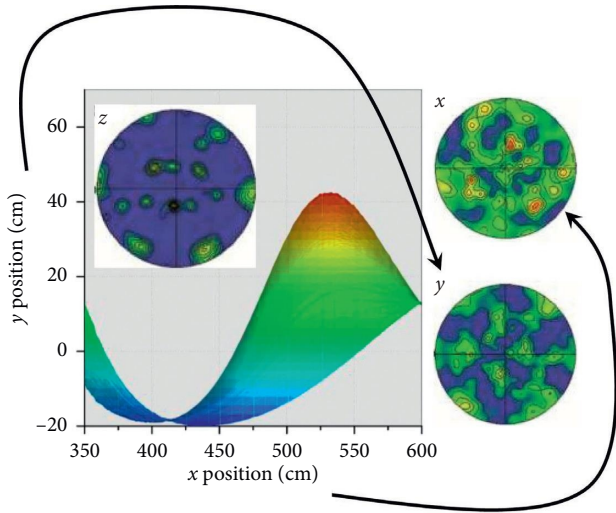


FIGURE 6: Extraction frame of front view, side view, and top view.

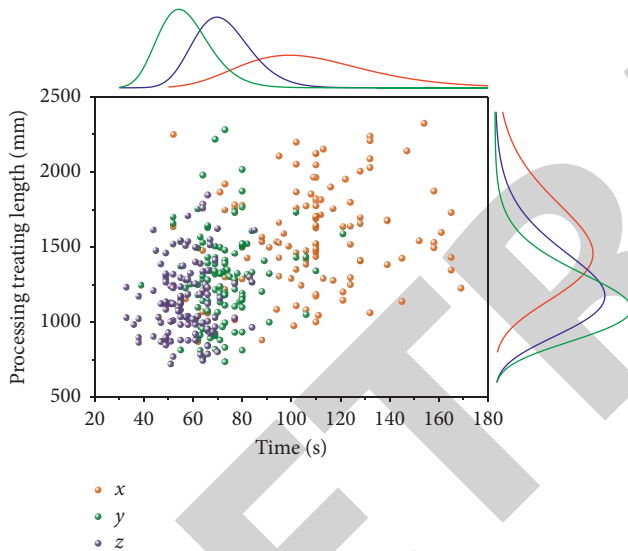


FIGURE 7: Processing information with a short time series.

last three frames of each RGB video series. In addition, to reduce the intraclass variability, all DMMs are resized to a fixed size using bitrivial interpolation in this paper; the reason being that the DMMs of different action video sequences may have different sizes. Note that the HOG features are not extracted by DMMs here as in the previous work, and thus the computational effort of the network model is reduced in this paper, and the fixed size of each DMM is set to half of the average of all sizes. As the proposed feature descriptors have high dimensionality, the dimensionality reduction technique is used.

The feature recognition of human motion behavior in the field of human behavioral action recognition facilitates effective modeling and description of the correlation information between temporal motion and spatial representations. In real application, it is not capable enough to handle the long-term dependency problem. At the beginning and at

the end of the photo, the motif was mostly rested, with only small movements of the limbs so that it did not affect the characteristics of the movement, and the first and last frames were removed, and the situation was not moved.

## 6. Conclusion

In this paper, three-dimensional projection of the depth sequence frame through three orthogonal planes is obtained to represent the behavior characteristics of the action. Since the calculation is simple, the same method is adopted in this article. The method of this paper is to generate a two-dimensional projection map in three planes of the depth sequence frame for front view  $f$ , side view  $s$ , and top view  $t$ . The optical flow information input to the global flow network channel is extracted from the RGB video sequence using the tv-l1 optical flow algorithm, and the horizontal and vertical components of the tv-l1 optical flow are adjusted, and the optical flow value is DMM calculation. The depth motion map of the depth action sequence, which effectively depicts the motion characteristics of three orthogonal planes orthogonal to the projection diagram, can be obtained by DMM calculation. Compressed sensing design of human motion acceleration data for the body area network of low-power consumption is proposed, and the basic concept and implementation process of compressed sensing theory for data compression and reconstruction of human motion in wireless body area network are described in detail. By this method, we can generate a recognition map over 10 times compared with the traditional single-correlation motion system. Besides, the input of the temporal network in this paper is a stacked 20-layer dense optical flow image, so the number of channels of the input features is 20, which is much faster than the number of channels of the temporal network in the previous work.

## Data Availability

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

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## References

- [1] C. Liu, M. Li, B. Lu et al., "High-sensitivity crack-based flexible strain sensor with dual hydrogen bond-assisted structure for monitoring tiny human motions and writing behavior," *Organic Electronics*, vol. 88, Article ID 105977, 2021.
- [2] A. Emmert, R. Franke, I. F. Brandes et al., "Comparison of conductive and convective warming in patients undergoing video-assisted thoracic surgery: a prospective randomized clinical trial," *Thoracic and Cardiovascular Surgeon*, vol. 65, no. 5, pp. 362–366, 2017.
- [3] X. Xu, S. Wu, J. Cui et al., "Highly stretchable and sensitive strain sensor based on polypyrrole coated bacterial cellulose fibrous network for human motion detection," *Composites Part B: Engineering*, vol. 211, no. 1, Article ID 108665, 2021.

- [4] J. Wang, J. D. Lee, M. Mahdavi, M. Kolar, and N. Srebro, "Sketching meets random projection in the dual: a provable recovery algorithm for big and high-dimensional data," *Electronic Journal of Statistics*, vol. 11, no. 2, pp. 4896–4944, 2017.
- [5] S. P. Sahoo and S. Ari, "On an algorithm for human action recognition," *Expert Systems with Applications*, vol. 115, pp. 524–534, 2019.
- [6] B. Cheng, S. Chang, H. Li et al., "Highly stretchable and compressible carbon nanofiber-polymer hydrogel strain sensor for human motion detection," *Macromolecular Materials and Engineering*, vol. 305, no. 3, Article ID 1900813, 2020.
- [7] W. Sun and Q. Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3185–3195, 2018.
- [8] T. Yellamraju and M. Boutin, "Clusterability and clustering of images and other "real" high-dimensional data," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1927–1938, 2018.
- [9] M. Murakami, J. K. Tan, H. Kim et al., "Human motion recognition using directional motion history images," *Proceedings of International Conference on Artificial Life and Robotics*, vol. 25, pp. 779–782, 2020.
- [10] A. G. Zebadua, P.-O. Amblard, E. Moisan, and O. J. J. Michel, "Compressed and quantized correlation estimators," *IEEE Transactions on Signal Processing*, vol. 65, no. 1, pp. 56–68, 2017.
- [11] R. Heckel, M. Tschannen, and H. Bölcskei, "Dimensionality-reduced subspace clustering," *Information and Inference: A Journal of the IMA*, vol. 6, no. 3, pp. 246–283, 2017.
- [12] N. Tang, C. Zhou, D. Qu et al., "Strain sensors: a highly aligned nanowire-based strain sensor for ultrasensitive monitoring of subtle human motion (small 24/2020)," *Small*, vol. 16, no. 24, Article ID 2070132, 2020.
- [13] H. Wang, J. Li, X. Yu et al., "Cellulose nanocrystalline hydrogel based on a choline chloride deep eutectic solvent as wearable strain sensor for human motion," *Carbohydrate Polymers*, vol. 255, no. 2, Article ID 117443, 2021.
- [14] M. Mukhopadhyay and D. B. Dunson, "Targeted random projection for prediction from high-dimensional features," *Journal of the American Statistical Association*, vol. 115, no. 532, pp. 1998–2010, 2020.
- [15] Z. Gao, P. Wang, H. Wang, M. Xu, and W. Li, "A review of dynamic maps for 3D human motion recognition using ConvNets and its improvement," *Neural Processing Letters*, vol. 52, no. 2, pp. 1501–1515, 2020.
- [16] D. Garca-Gil, S. Ramirez-Gallego, S. Garca, and F. Herrera, "Principal components analysis random discretization ensemble for big data," *Knowledge Based Systems*, vol. 150, pp. 166–174, 2018.
- [17] T. Li, J. Li, A. Zhong et al., "A flexible strain sensor based on CNTs/PDMS microspheres for human motion detection," *Sensors and Actuators A: Physical*, vol. 306, Article ID 111959, 2020.
- [18] C.-M. Vong, C. Chen, and P.-K. Wong, "Empirical kernel map-based multilayer extreme learning machines for representation learning," *Neurocomputing*, vol. 310, pp. 265–276, 2018.
- [19] X. Zhou, "Wearable health monitoring system based on human motion state recognition," *Computer Communications*, vol. 150, pp. 62–71, 2020.
- [20] D. W. Dreisigmeyer, "A quasi-isometric embedding algorithm," *Pattern Recognition and Image Analysis*, vol. 29, no. 2, pp. 280–283, 2019.
- [21] R. Koopman, S. Wang, and A. Scharnhorst, "Contextualization of topics: browsing through the universe of bibliographic information," *Scientometrics*, vol. 111, no. 2, pp. 1119–1139, 2017.
- [22] M. Sadeghi, R. Toosi, and M. A. Akhaee, "Blind gain invariant image watermarking using random projection approach," *Signal Processing*, vol. 163, pp. 213–224, 2019.
- [23] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognition*, vol. 92, pp. 64–81, 2019.
- [24] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2106–2116, 2017.
- [25] Q. Ke, J. Zhang, W. Wei et al., "A neuro-heuristic approach for recognition of lung diseases from X-ray images," *Expert Systems with Applications*, vol. 126, pp. 218–232, 2019.
- [26] M. M. Sardeshmukh and D. M. T. Kolte, "Human activity recognition using motion feature and two stage classification," *Digital Image Processing*, vol. 8, no. 4, pp. 111–116, 2016.
- [27] H.-L. Yang, M.-Z. Huang, and Z.-Q. Cai, "Research on human motion recognition based on data redundancy technology," *Complexity*, vol. 2021, Article ID 5542892, 6 pages, 2021.
- [28] I. R. Vega and S. Sarkar, "Statistical motion model based on the change of feature relationships: human gait-based recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1323–1328, 2003.
- [29] M. O. Yujian, H. Zhenjie, C. Xingzhi, L. Jiuzhen, C. Chen, and H. Juan, "Structural feature representation and fusion of behavior recognition oriented human spatial cooperative motion," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 45, no. 12, pp. 13–17, 2019.
- [30] L. Liu, L. Shao, and P. Rockett, "Boosted key-frame selection and correlated pyramidal motion-feature representation for human action recognition," *Pattern Recognition*, vol. 46, no. 7, pp. 1810–1818, 2013.
- [31] H. Zhou, M. You, L. Liu, and C. Zhuang, "Sequential data feature selection for human motion recognition via Markov blanket," *Pattern Recognition Letters*, vol. 86, pp. 18–25, 2017.