Mathematical Biosciences and Engineering

MBE, 20(6): 10135–10152.
DOI: 10.3934/mbe.2023444
Received: 27 January 2023
Revised: 19 March 2023
Accepted: 22 March 2023
Published: 29 March 2023

http://www.aimspress.com/journal/MBE

Research article

Human motion recognition based on Nano-CMOS Image sensor

Shangbin Li and Yu Liu*

Physical Education Department, Harbin Engineering University, Harbin 150001, China

* **Correspondence:** Email: liuyu6666666@163.com.

Abstract: Human motion recognition is of great value in the fields of intelligent monitoring systems, driver assistance system, advanced human-computer interaction, human motion analysis, image and video processing. However, the current human motion recognition methods have the problem of poor recognition effect. Therefore, we propose a human motion recognition method based on Nano complementary metal oxide semiconductor (CMOS) image sensor. First, using the Nano-CMOS image sensor to transform and process the human motion image, and combines the background mixed model of pixels in the human motion image to extract the human motion features, and feature selection is conducted. Second, according to the three-dimensional scanning features of Nano-CMOS image sensor, the human joint coordinate information data is collected, the state variables of human motion are sensed by the sensor, and the human motion model is constructed according to the measurement matrix of human motions. Finally, the foreground features of human motion images are obtained by calculating the feature parameters of each motion gesture. According to the posterior conditional probability of human motion images, the recognition objective function of human motion is obtained to realize human motion recognition. The results show that the human motion recognition effect of the proposed method is good, the extraction accuracy is high, the average human motion recognition rate is 92%, the classification accuracy is high, and the recognition speed is up to 186 frames/s.

Keywords: recognition; human motion; Nano-CMOS Image sensor; feature extraction; gray image

1. Introduction

The behavioral motion of the human is an activity controlled by the brain on the surface of the body; it is a coordinated and balanced posture between the various parts of the body and a way of

interaction between the body and its surroundings [1–2]. However, since the human brain has excellent filtering and comprehension of the messages transmitted by body motions, machines nowadays do not yet have the relevant capabilities [3]. In addition, the understanding of human motion by machines and the establishment of methods for characterizing and recognizing human motion performance can help people to make new investigations into the visual cognitive mechanism of the brain [4]. Human motion recognition is the use of computers to analyze visual information in images or videos to obtain information about human motion and describe what the human is trying to communicate. Human motion recognition, as one of the high-level technologies in computer vision processing systems, has received a lot of attention in recent years not only from researchers but also from related vendors and companies, and has a wide range of applications and development prospects in many fields such as intelligent surveillance systems, driver assistance systems, advanced human-computer interaction, human motion analysis, and content-based image and video processing. With the continuous development of computer and information sensing technology, people put forward higher requirements for human motion analysis, expecting that more accurate human motion information can be directly obtained through new technical means. Therefore, human motion analysis has become one of the key technologies in the field of computer vision. Complementary metal oxide semiconductor (CMOS) image sensor is a typical solid-state imaging sensor, and with the development and application of nanotechnology, CMOS devices have been developed toward the nanometer level, thus giving birth to the Nano-CMOS image sensor with high integration, integrated A/D converter inside the chip, digital signal output, simple peripheral circuit, no phase drive pulse required for operation, and cheap price, and so on. These advantages make it very suitable for human motion recognition and ensure the quality and speed of data acquisition in the process of human motion recognition.

In domestic and international research, Yuan et al. [5] proposed a new constructive model of human motion, which not only allows the source to obtain the node characteristics of the distant side of human running, but also enables to obtain the dynamic skeleton map according to the various limb relationships generated during the motion, and verifies the feasibility of the method according to the correlation between the skeleton map and human motion, thus enhancing the adaptive capability and accuracy of the model. However, the method suffers from the problem of poor recognition effect. Feng et al. [6] demonstrated the superiority of the method by performing a large amount of data analysis on NTGB+ D and NTGB+D-120, and the preliminary construction of a motion framework using a scale space-time map combined with a neural network. However, the method suffers from the problem of low recognition rate, which makes it difficult to achieve the desired application effect. Hong [7] uses an unsupervised deep learning algorithm to build a deep learning model to achieve the extraction of motion features. This method does not require prior knowledge. Experiments on a group of challenging UCF motion data verify the effectiveness and feasibility of this method. However, it is found in the experiment that this method has classification accuracy, which leads to poor recognition quality.

This paper proposes a human motion recognition method based on Nano-CMOS image sensor to improve the quality of human motion recognition, and the main contributions of this paper are as follows: (1) The Nano-CMOS image sensor is used to transform and process the human running motion images to extract human motion features and perform feature selection, which addresses the problem of poor feature extraction accuracy and recognition rate caused by the traditional method due to poor data quality. (2) The goal of human motion recognition is achieved based on the posterior conditional probability of human motion images and the recognition objective function. This process solves the problems of poor recognition quality, low classification accuracy and recognition speed due

to too many feature parameters of traditional methods. (3) The performance of the recognition method in the paper is verified by applying several data sets and experimental metrics, which improves the reliability of the method and lays a solid foundation for the study of human motion recognition.

2. Related works

At present, there are many research achievements in the field of human motion recognition [8]. Li et al. [9] had done many experiments on NTGBD and Kinetics and proposed a multi-attention-based convolutional network, which can automatically learn spatial and temporal characteristics based on the convolutional network of multi-attentional spatiotemporal maps to achieve the prediction of human motion. Through several experiments with large datasets NTGBD and Kinetics, the method is proven to have high recognition ability. However, this method has the problem of low accuracy of feature extraction, and the actual application effect is poor. Xiao et al. [10] established a new motion behavior recognition algorithm based on kinematic curves, where an image stream feature map was generated by combining RGB images with a depth map and introduced into a dual-stream convolutional neural network for human behavior recognition. The results show that the method has superior performance in performing motion behavior recognition, with an accuracy of 91.85% compared to traditional recognition techniques. Liu [11] discussed the design and implementation of a human motion capture system based on Micro-electromechanical systems (MEMS) and Zigbee network, the system can be installed in various parts of the human multiple sensor nodes to obtain human motion information, and the use of sensor network technology to aggregate and upload this data to the host computer, the paper uses sensor network technology to obtain more complete and accurate human motion information. Chen et al. [12] proposed a collaborative multi-radar human motion recognition model based on migration learning and integration learning, and proposed a migration learning-based ResNeXt network model to address the problem of too few samples, which requires shorter cycles and obtains higher accuracy compared with typical convolutional neural networks from small to large. Tao et al. [13] proposed distance correlation coefficients based on Pearson weights to identify current human motion model categories, and five common human motion patterns were identified from the output data of inertial module units and flexible pressure sensors using the DBMK algorithm. Luo and Ning [14] investigated a high dynamic dance motion recognition method based on video multi-feature fusion, using directional gradient histogram features, optical flow direction histogram features and audio features extracted from dance videos to characterize dance movements after considering all features of dance movements. Lee et al. [15] propose a real-time interface control framework based on contactless gesture recognition by processing the original signal, detecting the electric field perturbation triggered by the gesture movement near the capacitive sensor using adaptive thresholding, and extracting significant signal frames to cover the real signal interval to complete the gesture recognition.

Through the above analysis of related literature, it can be seen that the existing research has achieved certain results. However, the process of human motion recognition is limited by the quality of data, resulting in poor recognition results. Therefore, this paper proposes a human motion recognition method based on Nano-CMOS image sensor. The Nano-CMOS image sensor is used to acquire human motion images and joint coordinate information data, and construct human motion models to lay the foundation for recognition research. The recognition target function is calculated by the posterior conditional probability calculation to complete the human motion recognition. The results show that the average motion recognition rate of the proposed method is 92%, the recognition speed

is up to 186 frames/s, and it has high extraction accuracy and classification accuracy.

3. Methodology

In this paper, human motion recognition is based on Nano-CMOS image sensors to construct a recognition framework in Figure 1.

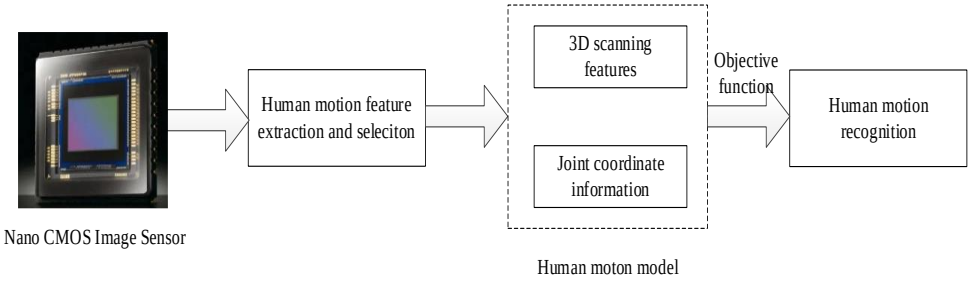


Figure 1. Framework for human motion recognition.

According to Figure 1, In this paper, human motion features are extracted based on the Nano-CMOS image sensor, and then the human motion model is constructed by combining 3D scanning features and joint coordinate information. On this basis, human motion recognition is completed by the construction and calculation of the objective function.

3.1. Extraction and selection of human motion features based on Nano-CMOS image sensor

The Nano-CMOS image sensor is an open quantum mechanical regime with steric transport and spatio-temporal irreversibility [16]. Figure 2 represents the CMOS image sensor transport model on different physical layers.

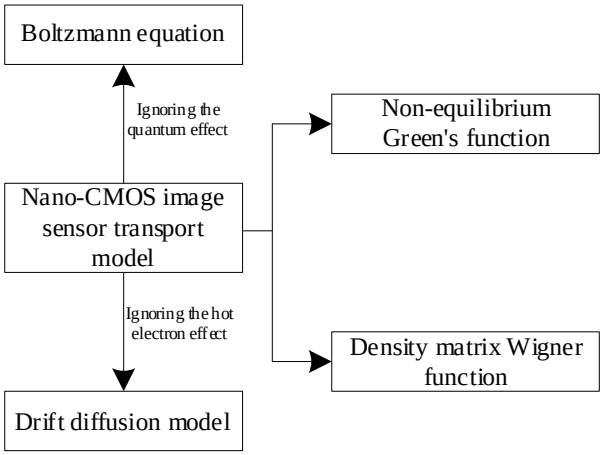


Figure 2. Nano-CMOS image sensor transport model.

Analysis of Figure 2 shows that the Nano-CMOS image sensor transport process involves the

Boltzmann equation, the Nano-CMOS image sensor transport model, and the drift-diffusion model, while the Nano-CMOS image sensor transport model consists of the non-equilibrium Green's function as well as the density matrix Wigner function. For the transport problem of CMOS devices, the non-equilibrium Green's function can be viewed as the current response of the device system to external excitation (e.g., applied bias voltage) and internal excitation (e.g., scattering by various types of particles). Under approximation and simplification conditions, the non-equilibrium Green's function model can generate the transport equations of the Wigner function and density matrix as a way to ensure the transport quality and efficiency of Nano-CMOS image sensors. Neglecting the quantum effect as well as the hot electron effect by the Boltzmann equation, the drift-diffusion model transmits discrete signal levels multiplexed to a single output and uses it to extract human motion features, which can ensure feature extraction accuracy and efficiency.

In CMOS image sensors, the electron transport is three-dimensional. Therefore, the recognition of human motion can be effectively achieved by using CMOS image sensors as a feature extraction tool.

Using the double convolution principle to threshold the human motion segmentation [17–18], the features of the human motion are extracted and the wavelet transform of the human motion image is implemented on a theoretical basis. The equation is

$$\Phi_{(a,b)}(k,l) = \| \partial(a,b) \cdot \ell(k,l) \| \tag{1}$$

where $\partial(a,b)$ is the human motion image, $\ell(k,l)$ is the working group of Nano-CMOS image sensors.

The human motion images are classified by support vector machine, and the classified images are normalized [19]. Suppose that at any time T, the running probability of any action (a,b) in the human motion image is expressed as

$$p\left(H_{ab}^{T}\right) = \sum_{s=K,L} p_s\left(H_{ab}^{T} |\ \gamma_{ab,s}^{T}\right) \tag{2}$$

where H_{ab}^{T} is the value of human motion image at time T. $p_s\left(H_{ab}^{T} |\ \gamma_{ab,s}^{T}\right)$ is the probability of action (a,b) in human motion image at any time T. At any time T, the mixed model of action (a,b) in the human motion image can be expressed as

$$p_b\left(H_{ab}^{T}\right) = \sum_{k} \vartheta_{ab,b,k}^{T} \times \eta\left(H_{ab}^{T}, \varepsilon_{ab,b,k}^{T}, \Sigma_{ab,b,k}^{T}\right) \tag{3}$$

where k refers to the number of human motion image models in the Nano-CMOS image sensor transport model. $\vartheta_{ab,b,k}^{T}$ is the weight information of human motion image in the Nano-CMOS image sensor transport model at any time T. $\varepsilon_{ab,b,k}^{T}$ is the characteristics of human motion transport model at time T. $\Sigma_{ab,b,k}^{T}$ is the threshold of the transport model at time T.

After obtaining the hybrid model of the human motion image, the difference between the human action image and the background of the human motion image is calculated [20] to complete the extraction of the features of the human motion image. The equation is

$$p_{\chi}(y) = Q_h \sum_{i=1}^{n} k\varphi\left[b\left(x_i\right) - \chi\right] \tag{4}$$

where Q_h is the motion feature of human motion. χ is the threshold of human motion image. y is the center pixel of human motion image.

Using the Nano-CMOS image sensor, the human running motion image is transformed and processed to obtain the pixel probability of any pixel in the human motion image, and the human motion features are extracted by combining the background mixing model of the pixels in the human motion image.

To remove redundant features and extract more effective human motion features, and lay the foundation for subsequent human motion recognition. Based on the extracted human motion features, the Information Gain is used for feature selection. The Information Gain measures the features based on the information size of the features, and if a human motion feature has a greater impact on the recognition elimination of the original dataset, it means that the feature is more beneficial to the accurate recognition of the dataset, and it can be recognized as a valid feature. The calculation equation is as follows:

$$f\left(d, p_{\chi}(y)\right) = Q(d) - Q\left(d \middle| p_{\chi}(y)\right) \tag{5}$$

where $f\left(d, p_{\chi}(y)\right)$ denotes the information gain of eigenvalue $p_{\chi}(y)$ for dataset d. $Q(a)$ denotes the information entropy. $Q\left(d \middle| p_{\chi}(y)\right)$ is the conditional entropy.

3.2. Construction of human motion model

Based on the human motion characteristics, the human motion model is designed and shown in Figure 3.

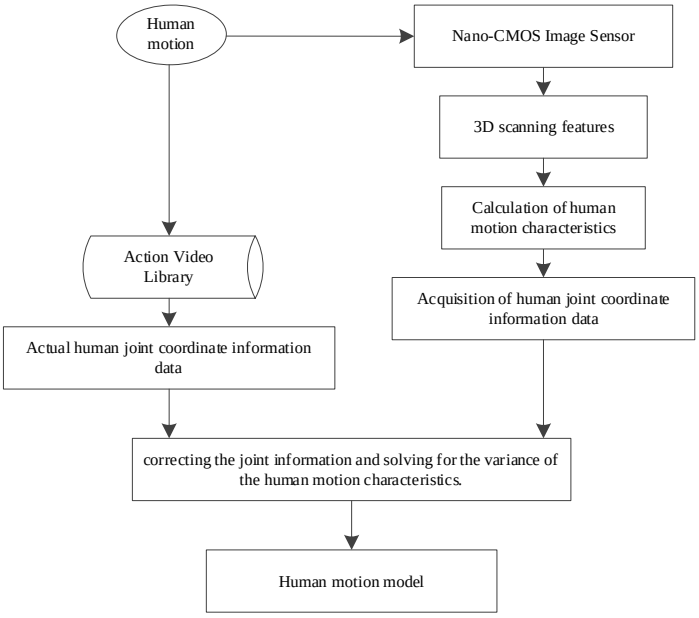


Figure 3. Human motion model.

Analyzing the results in Figure 3, it can be seen that the human motion characteristics are calculated based on the 3D scanning characteristics of the Nano-CMOS image sensor, and the human joint coordinate information data are collected and compared with the actual human joint coordinate information data in the motion video library to correct the joint information, the variance of the human motion characteristics is solved, and a human motion model is constructed based on the measurement matrix of human behavioral actions.

In this paper, Nano-CMOS image sensor is used to collect human motion information data, and human motion is detected, which is expressed as

$$\hat{\Omega}_{k+1} = \phi\Omega_k \tag{6}$$

where ϕ is the structural parameter of human motion. Ω_k is the covariance difference of human motion in frame k image.

According to the detection results of human motion, the state variables of human motion are predicted [21], namely

$$X_k = Jg_{k-1} + \varsigma u_k + U_k \tag{7}$$

where X_k is the dynamic information of human motion. g_{k-1} is the dynamic vector of human motion. J is the transformation matrix. ς represents the transfer matrix. u_k is the recognition coefficient. U_k is the error value of Nano-CMOS image sensor.

When CMOS image sensor is applied, the corresponding mathematical model must be established to determine its software and hardware parameters, transmission matrix and test matrix. In this paper, Taylor's expansion is used to express human motion information [22]. If i is used to represent a key point of human motion, the motion rate of this point on the X axis is

$$X_i(k+1) = X_i(k) + \frac{\Delta T^2}{2} X_i \tag{8}$$

where X_i represents the state variable of a joint during human motion. k stands for motion state. ΔT represents the time of sample collection.

According to various actions in the process of human motion, the human motion characteristics are calculated. The equation is:

$$\mu'_{k+1} = (1-\xi)\mu'_K + \xi\mu_k \tag{9}$$

$$\sigma'_{k+1} = (1-\xi)\sigma'_k + \xi\sigma_k \tag{10}$$

where ξ is the covariance of human motion characteristics. μ_k is the k-th frame of human motion image. μ'_K represents human motion characteristics. σ_k represents the eigenvector of human motion. σ'_k represents the perceived variance of human motion.

At the end of the sampling of the nodal targets for human motion, the mathematical model of the Nano-CMOS image sensor is established by the following equation

$$Z(k) = \varsigma X_i(k) + \Theta(k) \tag{11}$$

where $\Theta(k)$ represents the covariance matrix of Nano-CMOS image sensor to recognize human motion.

Using the mathematical model of the Nano-CMOS image sensor [23], information on the coordinates of all joint points of the human action is obtained and a human motion model is constructed, denoted as

$$Z_i(k+1) = \varsigma X_i(k+1) + \Theta(k+1) \tag{12}$$

3.3. Algorithm of Human motion recognition

Using the human motion model constructed above to complete the human motion recognition. The specific process is:

Input: test samples and training samples of human motion images to be recognized.

Output: human motion recognition results.

Initialize the parameters related to the Nano-CMOS image sensor, construct a human motion recognition model, and perform human motion image analysis based on the Nano-CMOS image sensor starting from $t=1$. The specific description is:

(1) Calculate the feature parameters of each action pose during each human motion, the equation is

$$l_{ij} = \frac{\left| l_{ij} - \min\left\{ l_{ij} \right\} \right|}{O_{edge}}, (a,b) \in \Lambda \tag{13}$$

where O_{edge} represents human motion image. Λ is (i,j) centered 3 × 3 window. $\min\left\{ l_{ij} \right\}$ represents the characteristic parameter of each motion.

(2) Using the characteristic parameters of human motion, the human motion matrix is obtained and the foreground features of the human action image are calculated using the following equation

$$R = R^* + \sigma_u \tag{14}$$

where R^* represents the classified mean value of human motion. σ_u is the dynamic vector of human motion image.

(3) Based on the acquired feature vectors of human motion, a perceptual model of human motion characteristics is constructed using a Nano-CMOS image sensor.

$$R^s = \bar{R}^s + \sigma_u^s \tag{15}$$

where S represents the covariance matrix of human motion. R^s and $\bar{R}^s$ are the action parameter during human motion. σ_u^s is the dynamic parameter of human motion image.

Using R^ε to represent the features of human motion, the state variables of human motion images are extracted using the human motion feature perception model constructed above [24] to obtain the feature vectors of human motion, and finally the feature components $\bar{R}^\varepsilon$ and G^ε of human motion are obtained, and the human motion feature model is transformed [25] to obtain

$$\left(R^\varepsilon\right)^s = \frac{\left(\bar{R}^\varepsilon\right)^s + \left(G^\varepsilon\right)^s}{u} \tag{16}$$

(4) The recognition probability of human motion images is calculated by multi-objective classification of human motion images [26], which can be expressed as

$$p\left(\psi_i \middle| s\right) = \frac{p\left(s \middle| \psi_i\right) p\left(\psi_i\right)}{\sum_{i=1}^{2} p\left(s \middle| \psi_i\right) p\left(\psi_i\right)} \tag{17}$$

where when $i=1$, $p\left(s \middle| w_i\right)$ is the target model of human motion image. When $i=2$, $p\left(s \middle| w_i\right)$ is the background model of human motion image. ψ_i is the eigenvalue of human motion image. $p(\psi_i)$ is the feature probability of human motion image.

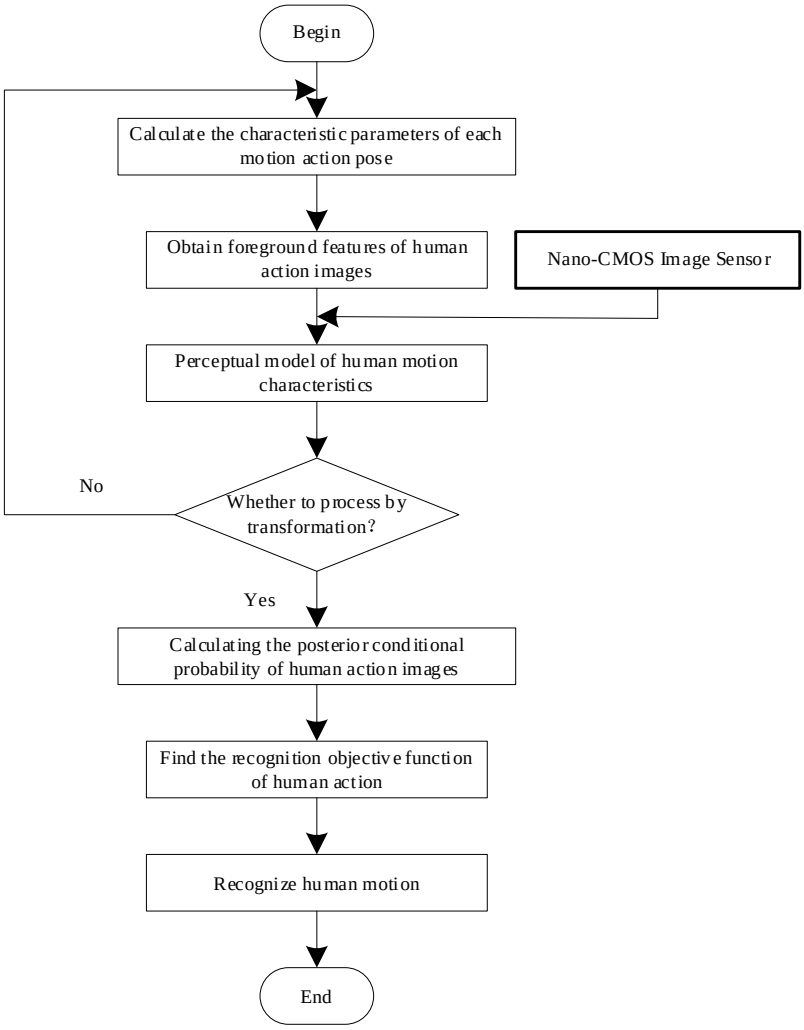


Figure 4. The recognition process of human motion.

(5) The size of the previous human motion image is the initial value, and the peak human motion information is obtained using the uniform shift method [27], and finally the objective function of human motion recognition is obtained, expressed as

$$Y = \beta_i \sum_{i=1}^{N} R_Z\left(U^{i+1}, V^{i+1}\right) \tag{18}$$

where β_i is the human motion image feature. R_Z is the feature vector of human motion image. $\left(U^{i+1},V^{i+1}\right)$ is the human motion recognition coordinate.

According to the above process, the recognition process of human motion is obtained, as shown in Figure 4. To sum up, the foreground features of human motion images are obtained by calculating the feature parameters of each motion gesture. The perception model of human motion features is used to transform the human motion feature model. According to the posterior conditional probability of human motion images, the recognition objective function of human motion is calculated to achieve human motion recognition.

4. Experimental analysis

4.1. Data sets

In this paper, C-MHAD data set (https://personal.utdallas.edu/~kehtar/C-MHAD.html) and MSR Daily Activity 3D data set (https://sites.google.com/view/wanqingli/data-sets/msr-action3d) are selected as experimental data sets. C-MHAD data set: the data acquisition device is applied to the commercially available Shimmer3 wearable inertial sensor, which can be connected to a laptop computer via Bluetooth, operating at 50HZ, and the laptop computer will capture video at 15 frames per second, with a resolution of 640 × 80 per frame, containing depth information of human motion, skeleton joint position information, RGB video sequences and motion inertial information, a total of 27 types of motion, all captured along the line of sight direction. MSR Daily Activity 3D data set: including RGB video data during human motion, motion depth information and skeleton joint position information, there are 16 types of motions, and each movement is captured in a real environment with background objects.

The experimental dataset is divided into training set and test set. In which, one thousand images in the C-MHAD dataset are used as the training set and the remaining 200 images are used as the test set. Two thousand images in the MSR Daily Activity 3D dataset are used as the training set and the remaining 200 images are used as the test set. To verify the performance of the proposed method in recognizing human motion, a human motion monitoring platform was built, including three inertial sensor nodes, a laptop computer, and a wireless receiver node. Each inertial sensor is composed of a three-axis gyroscope and an accelerometer with a sampling frequency of 100 Hz. After the Nano-CMOS image sensors acquire the human motion data, the acquired data are transmitted to the data processing software of the laptop through the wireless receiver node, and the data transmission mode of the wireless receiver node is set to Zigbee protocol. Eight Nano-CMOS image sensors are strapped to eight parts of the human, as shown in Figure 5.

The sensor used in this experiment is a VGA CMOS image sensor OV7680 of 1/10-inch (2.54 mm) from Howell Technologies. The OV7680 has a unique non-linear micro-lens shift technology, which can shorten the distance between the sensor and the lens and improve the pixel quality. The sensor has low noise, low power consumption, and wide range, and is capable of accomplishing a module size of 4.5 mm × 4.5 mm × 3.0 mm. To improve the data quality, the experimental dataset is pre-processed by de-noising, and then the feature data of different datasets are normalized and the data format is unified. The collected data and the data from the above-mentioned public data set are used as experimental data sources. After the experimental dataset was set up, it was divided into a training set and an experimental

set. The 500 images in the C-MHAD dataset were used as the training set and 100 images were used as the experimental set. The 1000 images in the MSR Daily Activity 3D dataset were used as the training set and 200 images were used as the experimental set. In the human data set obtained from the acquisition, 100 images were selected as the training set and 20 images were used as the test set.

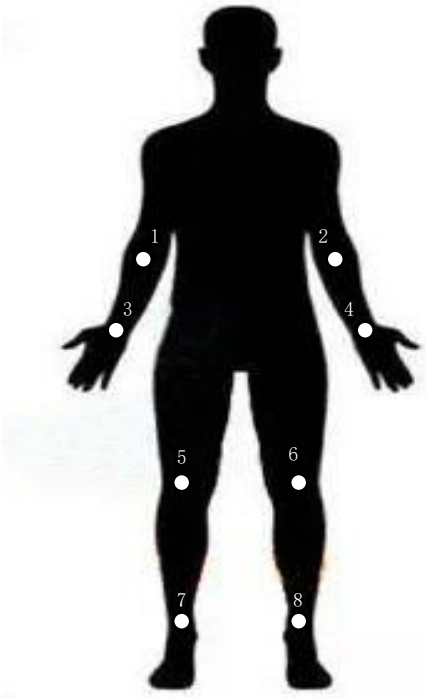


Figure 5. Sensor layout.

4.2. Evaluation metrics

To highlight the performance of the methods in the paper, HMSR-MEMS [11], THMRTL [12], HMMRMF [13], HDDMR [14], RICMGR [15] and the proposed method are introduced for comparison, and the effectiveness of the application of the different methods is verified by comparing different experimental indicators.

Human motion recognition effect: In the C-MHAD data set to obtain the standing long jump action and walking action, MSR Daily Activity 3D data set to obtain the skateboarding action and dancing action, according to the sensor arrangement, the sensor is tied to the human body to recognize the human motion in different positions, the higher the recognition success probability, the better the recognition effect.

The equation for this indicator of human motion feature extraction accuracy is as follows:

$$r = \frac{r_1}{r_2} \times 100\% \tag{19}$$

where r_1 represents the number of correctly extracted human motion features. r_2 represents the total amount of human motion characteristics.

The calculation equation of human motion recognition rate is as follows:

$$D = \frac{d_i}{d_j} \times 100\% \tag{20}$$

where d_i represents the number of correctly recognized human motion. d_j is the total amount of human motion.

The calculation equation of the accuracy of human motion classification is as follows:

$$K = \frac{k_2 - k_1}{k_2} \times 100\% \tag{21}$$

where k_1 refers to the number of human motions incorrectly classified. k_2 represents the number of human motion categories.

The calculation equation of human motion recognition speed is as follows:

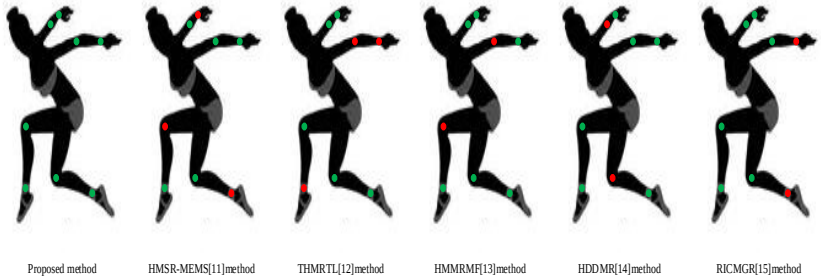
$$Z = \frac{O}{T} \times 100\% \tag{22}$$

where O represents the number of human motion video frames. T represents the recognition time.

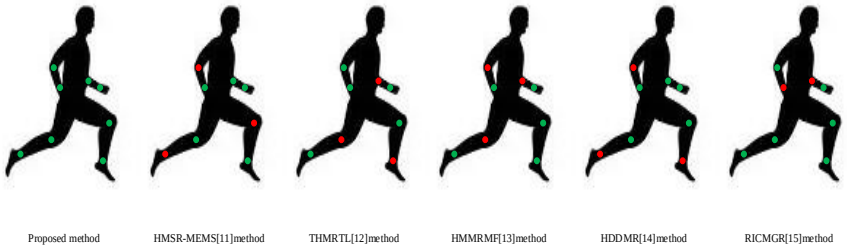
4.3. Results and discussion

The human motion recognition effect is shown in Figure 6.

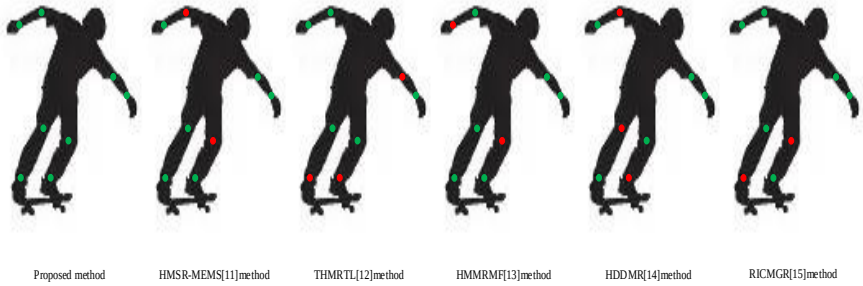
According to Figure6. The green circle means successful recognition, the red circle means failed recognition. According to the analysis of the results in Figure 6. For the standing long jump motion. HMSR-MEMS [11] method shows three recognition failures. THMRTL [12] method shows three recognition failures. HMMRMF [13] method shows two recognition failures. HDDMR [14] method shows two recognition failures. RICMGR [15] method shows two recognition failures. The proposed method did not show any recognition failure. For running motion. HMSR-MEMS [11] method shows three recognition failures. THMRTL [12] method shows three recognition failures. HMMRMF [13] method shows three recognition failures. HDDMR [14] method shows three recognition failures. RICMGR [15] method shows two recognition failures. The proposed method did not show any recognition failure. For skateboard motion. There are two recognition failures in HMSR-MEMS [11] method. There are three recognition failures in THMRTL [12] method. There are two recognition failures in HMMRMF [13] method. There are three recognition failures in HDDMR [14] method. There are two recognition failures in RICMGR [15] method. The proposed method did not show any recognition failure. For dance motion. There are three recognition failures in HMSR-MEMS [11] method. There are three recognition failures in THMRTL [12] method. There are three recognition failures in HMMRMF [13] method. There are two recognition failures in HDDMR [14] method. There are three recognition failures in RICMGR [15] method. The proposed method did not show any recognition failure.



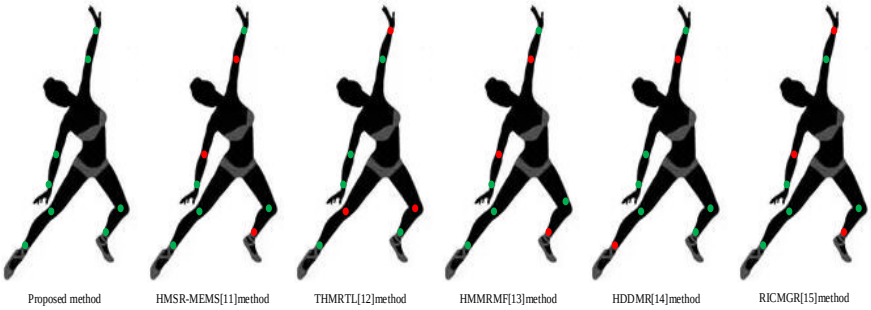
(a) Standing long jump motion



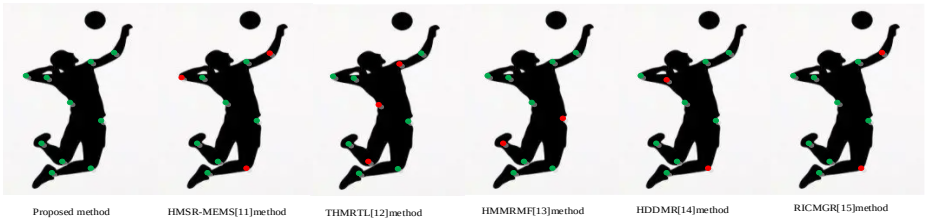
(b) Running motion



(c)Skateboarding motion



(d) Dance motion



(e) Shooting motion

Figure 6. Comparison of human motion recognition results using different algorithms.

To sum up, only the proposed method can recognize human motion at all monitoring points, and the other five methods have two or three motion recognition failures. Therefore, the proposed method has better recognition effect.

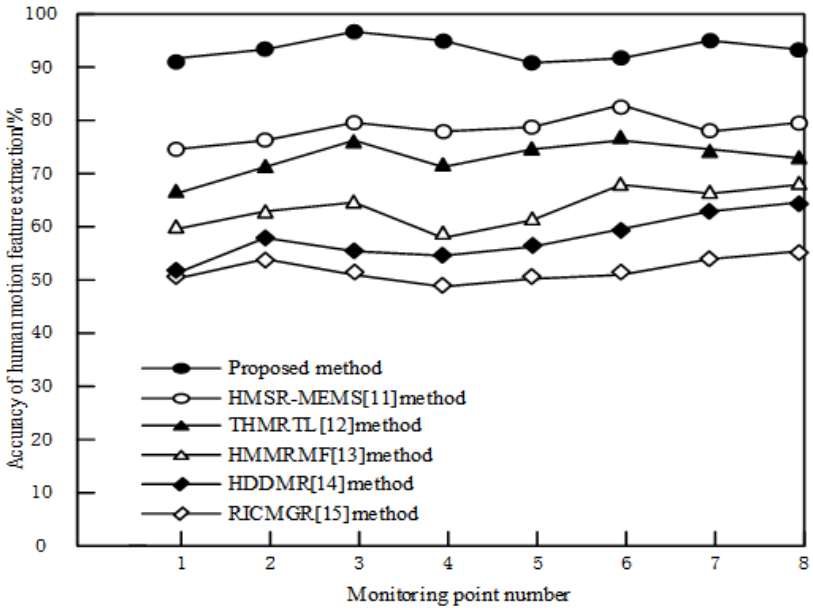


Figure 7. Comparison of accuracy of human motion feature extraction.

Analysis of the results in Figure7 shows that the proposed method reaches a maximum accuracy of 96% for human motion feature extraction when the monitoring point number is 3. When the monitoring point number is 6, the human motion feature extraction accuracy in HMSR-MEMS [11] method reaches a maximum of 83%. When the monitoring point number is 6, the human motion feature extraction accuracy in THMRTL [12] method reaches a maximum of 76%. When the monitoring point number is 6 and 8, the human motion feature extraction accuracy in HMMRMF [13] method reaches a maximum of 65%. When the monitoring point number is 8, the human motion feature extraction accuracy in HDDMR [14] method reaches a maximum of 60%. When the monitoring point number is 2, the human motion feature extraction accuracy in RICMGR [15] method reaches a maximum of 53%. The feature extraction accuracy of the proposed method is 13, 20, 31, 36 and 43% higher than the method of HMSR-MEMS [11], the method of THMRTL [12], the method of HMMRMF [13], the method of HDDMR [14], and the method of RICMGR [15], respectively, and can control the accuracy of motion feature extraction above 90%, while the other five methods have poor performance in terms of human motion feature extraction accuracy.

Analysis of the data in Table 1 shows that the average human motion recognition rate of the proposed method is 92%, which is 21, 22, 34, 32 and 35% higher than the methods of HMSR-MEMS [11], THMRTL [12], HMMRMF [13], HDDMR [14], and RICMGR [15], respectively. It indicates that the proposed method has a higher recognition rate in recognizing human motion compared with the other five methods, and can provide higher technical support for human motion analysis.

Table 1. Comparison of human motion recognition rate (unit: %).

Monitoring point number	Proposed method	HMSR-MEMS [11] method	THMRTL [12] method	HMMRMF [13] method	HDDMR [14] method	RICMGR [15] method
1	92	77	68	65	60	52
2	95	72	69	62	56	57
3	90	82	72	64	63	55
4	89	70	71	53	65	53
5	92	69	70	57	61	69
6	96	68	70	57	60	56
7	93	67	69	53	57	60
8	92	66	71	52	59	51
Average	92	71	70	58	60	57

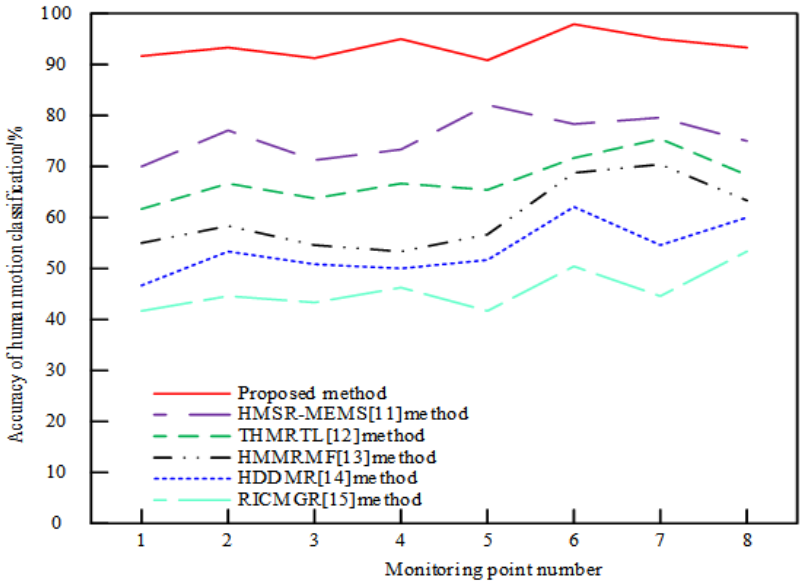


Figure 8. Comparison of accuracy of human motion classification.

According to the data in Figure 8 shows that the proposed method reaches a maximum human motion classification accuracy of 97%, which is 19% higher than the method of HMSR-MEMS [11], 18% higher than the method of THMRTL [12], 27% higher than the method of HMMRMF [13], 27% higher than the method of HDDMR [14], and 50% higher than the method of RICMGR [15]. Therefore, the performance of the proposed method in human motion classification accuracy is far higher than that of the other five methods, and accurate classification can ensure human motion.

Analyzing the data in Figure9, it can be seen that the proposed method reaches a maximum human motion recognition speed of 186 fps, which is 27 fps higher than the method in HMSR-MEMS [11]; 66 fps higher than the method in THMRTL [12]; 64 fps higher than the method in HMMRMF [13]; 71 fps higher than the method in HDDMR [14]; 91 fps higher than the method in RICMGR [15]. In summary, the proposed method can improve the recognition speed to more than 180 frames/s when recognizing human motion, and the higher recognition speed ensures the quality of human motion images.

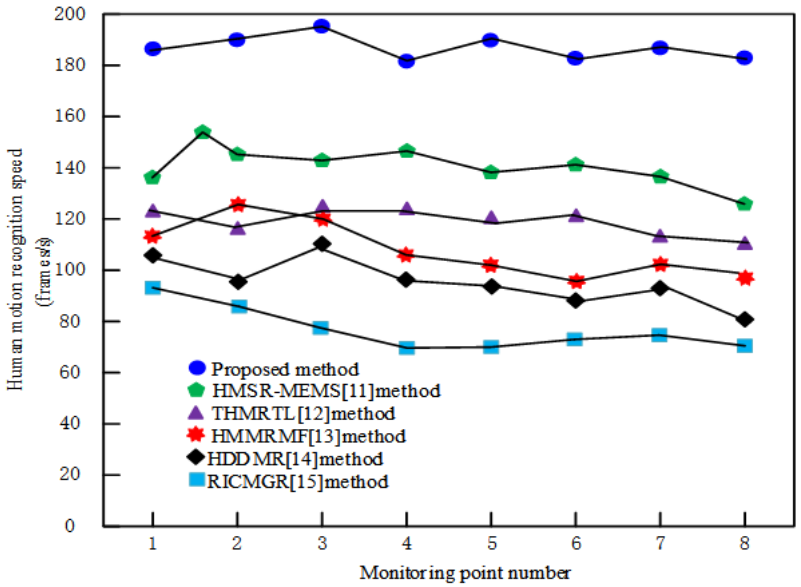


Figure 9. Comparison of human motion recognition speed.

5. Conclusions

In this paper, Nano CMOS image sensor is applied to human motion recognition to improve its effectiveness of human motion recognition. The results show that the method has good performance in human motion recognition. The maximum accuracy of human motion feature extraction is 96%, the average human motion recognition rate is 92%, the maximum accuracy of human motion classification is 97%, and the maximum human motion recognition speed is 186 frames/s. This proves that the proposed method has higher performance in human motion recognition. However, there are still some shortcomings in this study, the sample size of the current research work is not large, especially for the Nano-CMOS image sensor acquisition of the sample data is not enough. The sample data capacity needs to be expanded for further analysis in future studies to provide more effective data support and further application possibilities for the application of Nano-CMOS image sensors in human motion recognition.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. G. Yan, M. Hua, Z. Zhong, Multi-derivative physical and geometric convolutional embedding networks for skeleton-based action recognition. *Comput. Aided Gemo. D.*, **86** (2021), 101964. https://doi.org/10.1016/j.cagd.2021.101964
2. H. Su, W. Qi, Y. Schmirander, S. E. Ovur, S. Cai, X. Xiong, A human activity-aware shared control solution for medical human-robot interaction, *Assembly Auton.*, **42** (2022),388–394. https://doi.org/10.1108/AA-12-2021-0174

3. Y. Xue, Y. Yu, K. Yin, P. F. Li, S. X. Xie, Z. J. Ju, Human In-Hand motion recognition based on multi-modal perception information fusion, *IEEE Sens. J.*, **22** (2022), 6793–6805. https://doi.org/10.1109/JSEN.2022.3148992
4. Y. Liu, H. Zhang, D. Xu, K. J. He, Graph transformer network with temporal kernel attention for skeleton-based action recognition, *Knowl.-Based Syst.*, **240** (2022), 108146. https://doi.org/10.1016/j.knosys.2022.108146
5. Y. Yuan, B. Yu, W. Wang, B. H. Yu, Multi-filter dynamic graph convolutional networks for skeleton-based action recognition, *Procedia Comput. Sci.*, **183** (2021), 572–578. https://doi.org/10.1016/j.procs.2021.02.099
6. D. Feng, Z. C. Wu, J. Zhang, T. Ren, Multi-Scale Spatial Temporal Graph Neural Network for Skeleton-Based Action Recognition, *IEEE Access*, **9** (2021), 58256–58265. https://doi.org/10.1109/ACCESS.2021.3073107
7. Y. Q. Hong, Visual human action recognition based on deep belief network, *Comput. Sci.*, **48** (2021), 400–403. https://doi.org/10.11896/jsjkx.210200079
8. H. Su, W. Qi, J. Chen, D. Zhang, Fuzzy approximation-based Task-Space control of robot manipulators with remote center of motion constraint, *IEEE Trans. Fuzzy Syst.*, **30** (2022), 1564–1573. https://doi.org/10.1109/TFUZZ.2022.3157075
9. X. Y. Li, X. W. Hao, J. G. Jia, Y. F. Zhou, Human action recognition method based on multi-attention mechanism and spatiotemporal graph convolution networks, *J. Comput. Aided Des. Comput. Graph.*, **33** (2021), 1055–1063. https://doi.org/10.3724/SP.J.1089.2021.18640
10. Z. T. Xiao, L. Zhang, W. Wang, Human action recognition based on kinematic dynamic image, *J. Tianjin Polytechnic Univer.*, **40** (2021), 53–59. https://doi.org/10.3969/j.issn.1671-024x.2021.01.010
11. Q. Liu, Human motion state recognition based on MEMS sensors and Zigbee network, *Comput. Commun.*, **181** (2022), 164–172. https://doi.org/10.1016/j.comcom.2021.10.018
12. P. Chen, S. Guo, H. Li, X. Wang, G. L. Cui, C. S Jiang, et al., Through-wall human motion recognition based on transfer learning and ensemble learning, *IEEE Geosci. Remote. S.*, **19** (2022), 1–5. https://doi.org/10.1109/LGRS.2021.3070374
13. Z. Tao, Z. Hao, Y. Lei, Human motion mode recognition based on multi-parameter fusion of wearable inertial module unit and flexible pressure sensor, *Sensor. Mater.*, **34** (2022), 1017–1031. https://doi.org/10.18494/SAM3755
14. W. Luo, B. Ning, High-dynamic dance motion recognition method based on video visual analysis, *Sci. Programming-Neth*, **2022** (2022), 1–9. https://doi.org/10.1155/2022/6724892
15. H. Lee, J. K. Mandivarapu, N. Ogbazghi, Y. S. Li, Real-time interface control with motion gesture recognition based on non-contact capacitive sensing, preprint, arXiv: 2201.01755.
16. S. Chen, K. Xu, Z. Mi, X. H. Jiang, T. F. Sun, Dual-domain graph convolutional networks for skeleton-based action recognition, *Mach. Learn.*, **111** (2022), 2381–2406. https://doi.org/10.1007/s10994-022-06141-8
17. X. Ji, Q. Zhao, J. Cheng, C. F. Ma, Exploiting spatio-temporal representation for 3D human action recognition from depth map sequences, *Knowl.-Based Syst.*, **227** (2021), 107040. https://doi.org/10.1016/j.knosys.2021.107040
18. W. Ding, C. Ding, G. Li, K. Liu, Skeleton-based square grid for human action recognition with 3D convolutional neural network, *IEEE Access*, **9** (2021), 54078–54089. https://doi.org/10.1109/ACCESS.2021.3059650

19. M. F. Tsai, C. H. Chen, Spatial temporal variation graph convolutional networks (STV-GCN) for skeleton-based emotional action recognition, *IEEE Access*, **9** (2021), 13870–13877. https://doi.org/10.1109/ACCESS.2021.3052246
20. H. Xia, X. Gao, Multi-scale mixed dense graph convolution network for skeleton-based action recognition. *IEEE Access*, **9** (2021), 36475–36484. https://doi.org/10.1109/ACCESS.2020.3049029
21. K. B. de Carvalho, V. T. Basílio, A. S Brando, Action recognition for educational proposals applying concepts of social assistive robotics, *Cogn. Syst. Res.*, **71** (2022), 1–8. https://doi.org/10.1016/j.cogsys.2021.09.002
22. Y. Kong, Y. Wang, A. Li, Spatiotemporal saliency representation learning for video action recognition, *IEEE T. Multimedia*, **24** (2022), 1515–1528. https://doi.org/10.1109/TMM.2021.3066775
23. H. Wang, B. Yu, K. Xia, J. Q. Li, X. Zuo, Skeleton edge motion networks for human action recognition, *Neurocomputing*, **423** (2021), 1–12. https://doi.org/10.1016/j.neucom.2020.10.037
24. R. D. Brehar, M. P. Muresan, T. Marita, C. Vancea, M. Negru, S. Nedevschi, Pedestrian street-cross action recognition in monocular far infrared sequences, *IEEE Access*, **9** (2021), 74302–74324. https://doi.org/10.1109/ACCESS.2021.3080822
25. J. Xie, W. Xin, R. Liu, L. Sheng, X. Liu, X. Gao, et al., Cross-channel graph convolutional networks for skeleton-based action recognition, *IEEE Access*, **9** (2021), 9055–9065. https://doi.org/10.1109/ACCESS.2021.3049808
26. A. Avp, B. Apa, A. Iao, Comparison of action recognition from video and IMUs, *Procedia Comput. Sci.*, **186** (2021), 242–249. https://doi.org/10.1016/j.procs.2021.04.144
27. R. Xia, Y. Li, W. Luo, LAGA-Net: Local-and-global attention network for skeleton based action recognition, *IEEE T. Multimedia*, **24** (2022), 2648–2661. https://doi.org/10.1109/TMM.2021.3086758