IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 September 2022, accepted 17 September 2022, date of publication 5 October 2022, date of current version 8 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3212366

RESEARCH ARTICLE

A High-Performance Gait Recognition Method Based on n-Fold Bernoulli Theory

QING ZHOU, JARHINBEK RASOL, YUELEI XU, ZHAOXIANG ZHANG, AND LUJUAN HU
Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Yuelei Xu (xuyuelei@nwpu.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant D5000210767.

ABSTRACT Gait feature recognition refers to recognizing identities by collecting the characteristics of people when they walk. It shows the advantages of noncontact measurement, concealment, and nonimitability, and it also has good application value in monitoring, security, and company management. This paper utilizes Kinect to collect the three-dimensional coordinate data of human bones. Taking the spatial distances between the bone nodes as features, we solve the problem of placement and angle sensitivity of the camera. We design a fast and high-accuracy classifier based on the One-versus-one (OVO) and One-versus-rest (OVR) multiclassification algorithms derived from a support vector machine (SVM), which can realize the identification of persons without data records, and the number of classifiers is greatly reduced by design optimization. In terms of accuracy optimization, a filter based on n-fold Bernoulli theory is proposed to improve the classification accuracy of the multiclassifier. We select 20000 sets of data for fifty volunteers. Experimental results show that the design in this paper can effectively yield improved classification accuracy, which is 99.8%, and reduce the number of originally required classifiers by 91%-95%.

INDEX TERMS Gait characteristics, Kinect v2, Bernoulli theory, least-squares support vector machine.

I. INTRODUCTION

Human gait is characterized by repetitive movements of different body joints. Unique features extracted from body joint motions are utilized to identify a person [1]. Gait is generated unconsciously by an individual and is one of the most difficult biological characteristics to replicate [2]. Compared with physiological feature recognition technology (such as those concentrating on human faces, fingerprints and irises), gait recognition has the advantages of being noninvasive, hard to conceal and not requiring explicit cooperation from the subjects [3]. Therefore, it has great application value in human-computer interaction [4], security, clinical diagnosis [5] and other fields.

Gait recognition is a challenging noncontact biometric technology that has received much attention from researchers. The classic process of gait recognition is shown in Figure 1. Both the training and verification procedures include four steps: data collection, data preprocessing, feature extraction and classification [6].

(1) Data collection: The quality of the collected gait data affects the performance of trained gait recognition models. Typical gait capture sensors include multiple cameras, professional motion capture systems (such as the Vicon Optical Motion Capture system), or cameras with depth sensors (such as Kinect).

(2) Data preprocessing: Data preprocessing mainly includes image noise reduction and human contour extraction. Optical-based data acquisition methods usually use background subtraction to obtain the outlines of pedestrians [7]. Periodic extraction of data can also be performed at this stage to divide the gait sequence into several gait cycles to reduce data redundancy [8].

(3) Feature extraction: Gait features can be generated by manual extraction and machine learning. Manually extracted features are easily generalized to different datasets, while machine-learned features are usually dataset specific.

(4) Classification: A classifier based on gait features can be used for identity recognition. Traditional classifiers

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

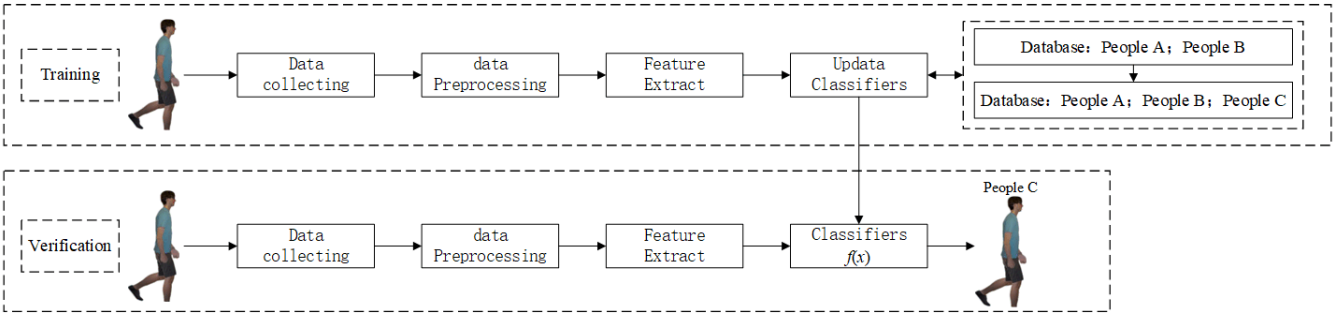


FIGURE 1. Gait recognition process.

(such as k-nearest neighbors, KNN) and modern classifiers (such as deep neural networks) have achieved success in facial recognition, handwriting recognition, and speech recognition [9], [10].

In the past two decades, 2D-based gait recognition methods have been applied more than 3D model-based methods. 2D-based methods use human silhouettes as raw input data, and they are usually divided into model-based and model-free methods [11], where the former is sensitive to human shape changes and the latter is computationally expensive [12]. Kinect v2 is an integrated depth camera developed by Microsoft that can provide accurate 3D human skeleton information, which is insensitive to the human body and appearance. Over the past decade, many approaches based on Kinect gait biometrics have been devised with promising results [13], [14], [15], [16]. However, traditional methods cannot achieve high accuracy, while deep learning methods are computationally intensive and have high requirements on the operating environment [17].

In this paper, we try to design a new multiclassifier using machine learning algorithms combined with knowledge of probability theory. We chose machine learning algorithms rather than deep learning algorithms to reduce computation, which is more convenient in embedded applications. the spatial information of human joints is collected by Kinect v2. The main contributions of this paper include the following:

(1) A set of customized human gesture recognition datasets suitable for Kinectv2 was constructed, including a total of 20,000 sets of data after screening;

(2) We skillfully combine the SVM-derived OVO and OVR algorithms. On one hand, we solve the problem that the OVO classifier set cannot identify samples that do not exist in the training set. On the other hand, we minimize the number of classifiers in the classifier set by optimizing the design, the classification speed can be increased by more than 13 times, and high-speed classification is achieved under the premise of ensuring accuracy.

(3) Furthermore, we have added a new classification rule to the algorithm, which greatly improves the classification accuracy by utilizing the n-fold Bernoulli algorithm in probability theory. The classification accuracy is as high as 99.8%.

In the second part of the article, the related work of Kinect-based gait recognition technology is elaborated. The main work of this paper is described in detail in Section III. First, the overall process of the proposed algorithm is introduced. In Subsection A, the knowledge about human joint point extraction based on Kinect v2 is briefly introduced. Data denoising, feature selection and feature extraction are introduced in Subsection B. In Subsection C, we elaborate on how to design a high-speed and high-accuracy multiclassifier based on the OVO and OVR multiclassification algorithms derived from SVM. In addition, we improve the proposed algorithm through multiple Bernoulli theory and optimization design to further improve the accuracy and speed. In Section IV, the classification speed and accuracy of the proposed algorithm are evaluated through experiments, and in Subsection B, we compare the performance of the proposed algorithm and state-of-the-art algorithms. The conclusions of this paper are summarized in the fifth section.

II. RELATED WORK

Generally, gait recognition methods can be divided into 3D data-based and 2D data-based methods [18]. The 2D data-based gait recognition methods generally recognize the human silhouette captured by 2D cameras, which is common in video surveillance. Gait recognition methods based on 2D data dominate the field of gait recognition and are usually divided into model-based and model-free methods [11], [19]. Model-free methods refer to analyzing changes in human motion by generating binary segmented images of gait sequence contours. Li and Chen [20] developed well-performing gait energy images (GEIs) by fusing foot and head energy images from contour sequences. There are other methods, such as active energy images [21], gait flow images [22], and structural gait energy images [20]. Although these methods are not computationally expensive, the recognition effect is affected by viewing angle and distance factors. Model-based methods refer to constructing a model by estimating changes in parameters of different parts of the human body in the video. BenAbdelkader et al. [23] extracted features by deriving 3D models from 2D images to compute step size and walking speed. Yam et al. [24] identified a person by calculating the change in motion between

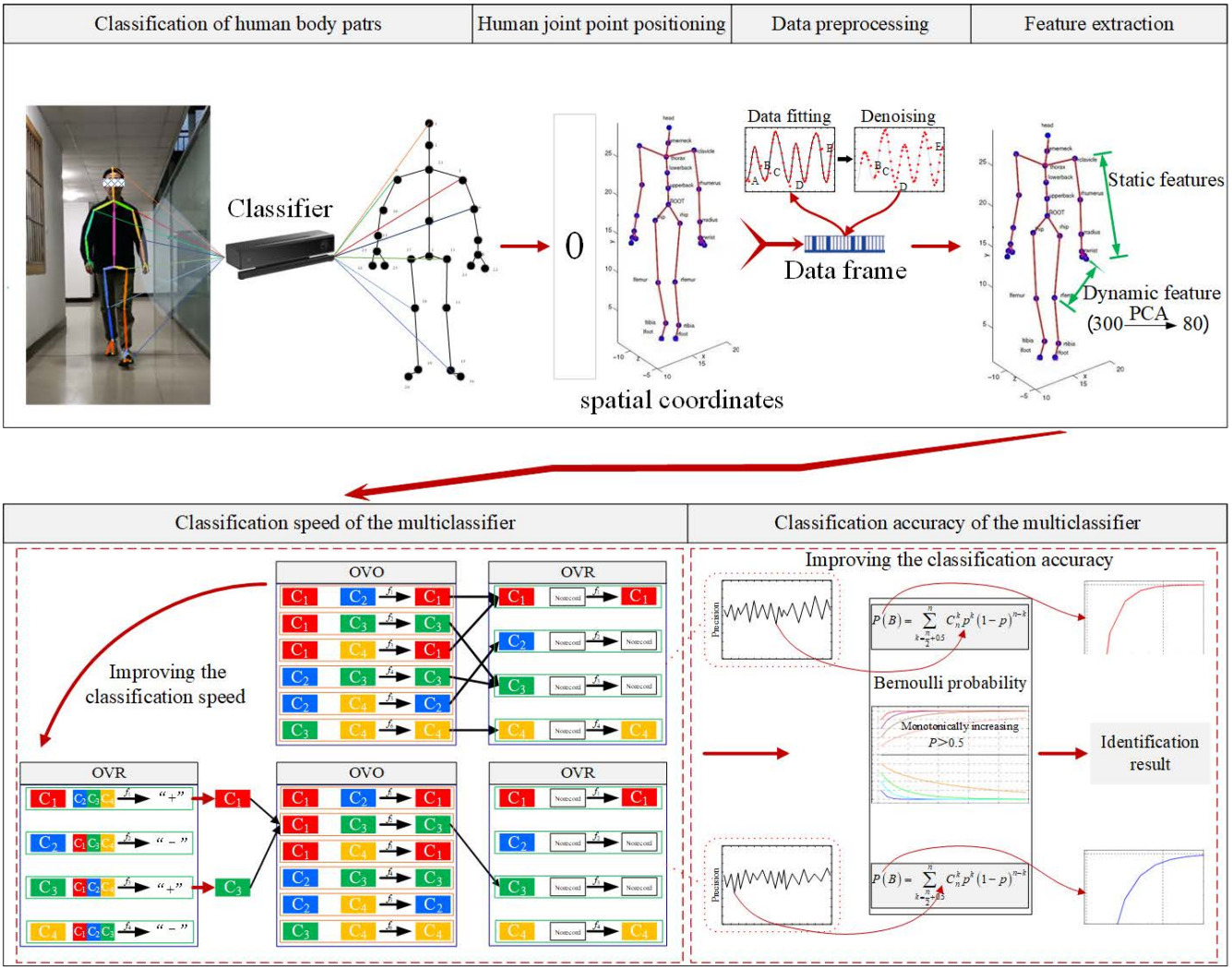


FIGURE 2. The overall idea of the proposed approach.

walking and running. In summary, most of these methods construct models based on 2D images, which are computationally intensive. Therefore, they are not good methods for many practical applications [12]. In addition, factors such as clothing, viewing angles, and external environments in 2D images also affect the recognition rate of gait recognition approaches [25]. Therefore, gait recognition studies based on 3D information have gradually attracted the attention of researchers.

Kinect v2 is an integrated depth camera developed by Microsoft that can provide accurate 3D joint point data information of the human body. Initially, Kinect rapidly expanded its application range due to its low cost, portability, and convenience of data access, such as fall detection in hospitals [26], [27] and gesture signal interpretation for emergency response [28]. Currently, Kinect sensors are used in outdoor environments [29], [30], [31], [32]. Kinect can generate 3D models of human skeletons in real time without any labels attached to the human body [33], and the measurement error of depth information is small [34], [35]. Therefore, the 3D data of the Kinect skeleton model can be utilized for gait recognition. Clark et al. [36], conducted experiments on kinectmatic, postural, and spatiotemporal analyses to validate the usefulness of the Kinect sensor for gait analysis.

Many viable approaches based on Kinect gait biometrics have been devised during the last decade to aid in the successful recognition of individuals [14], [15], [37]. Preis et al. [13] pioneered model-based gait identification using the Kinect sensor, including eleven handcrafted static and two dynamic features, as well as rule-based, decision tree (DTree), and Nave Bayes classifiers. Using the K-means clustering method, temporal features of eighteen angles derived from selected body joints were retrieved to analyze gait variables [38]. The multilayer perceptron (MLP) architecture was used by Andersson and Araujo [37]; however, the performance of the K-nearest neighbors (KNN) and support vector machine (SVM) classifiers outperformed the MLP architecture. Banerjee et al. [39] used the fuzzy C-means clustering algorithm to identify elderly individuals based on the gait sequences obtained by Kinect. Sabir et al. [40]

proposed a three-dimensional gait recognition approach using the spatiotemporal variations in relative angles among various skeletal joints and changes in the measured distance between limbs and land. Ahmed et al. [41] extracted features based on the generated triangles using the triangle area and the three angles of each triangle. Yang et al. [15] calculated the average and standard deviation of relative distance features from selected body joints over the frames of a gait cycle. Sun et al. [42] extracted static and dynamic features to train the standard KNN classifier by extracting features from selected body joints in 2018. In 2019, researchers designed a deep neural network with 3 hidden layers, the first attempt to apply deep learning to the Kinect-based gait recognition task [43]. Its frame-by-frame prediction may cause prediction mistakes when a frame is similar to another person's gait pattern. Another study calculated the average and standard deviation of handcrafted features over 30 frames by extracting joint relative distance and joint relative angle data [44]. The uniform kernel cannot be utilized to extract hierarchical features using the CNN architecture because it was trained with handcrafted features, and the model suffers from overfitting. Bari et al. [45] designed two deep neural networks to realize Kinect-based gait recognition. In the ANN approach, a deep learning neural network architecture based on hand-engineered Joint Relative Cosine Dissimilarity and Joint Relative Triangle Area features; in the CNN approach, KinectGaitNet is utilized with a uniform kernel for the convolution to extract distinctive features [16]. Both methods require a high number of model parameters because the neural network design utilizes a multilayer network methodology and needs to be accelerated by GPU to achieve good recognition effects that are not suitable for embedded systems. In summary, these studies generally suffer from speed or accuracy deficiencies.

III. GAIT RECOGNITION METHOD

The general idea of the proposed method is shown in Figure 2. First, Kinect V2 automatically extracts the joint points of the human body, establishes a spatial coordinate system, and calculates the spatial coordinate information of the joint points of the human body. After the data are fitted and denoised, the static and dynamic features of the human body are extracted, and feature dimensionality reduction is achieved through principal component analysis (PCA) [46]. Then, the OVO classifier set->OVR multiclassifier algorithm is designed. On this basis, the speed of the classifier can be further improved by designing the OVR classifier set->OVO classifier set->OVR classifier algorithm, and the classification accuracy of the classifier can be greatly improved by utilizing n-fold Bernoulli theory.

A. HUMAN JOINT POINT EXTRACTION BASED ON KINECT v2

The frame data acquired by Kinect v2 contain a set of 25 joint points of the human body, as shown in Figure 3.

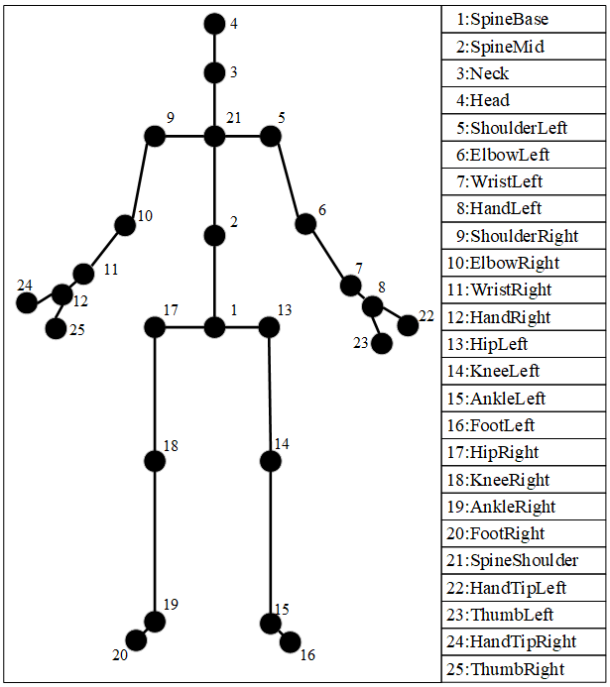


FIGURE 3. Twenty-five joint points of the human body.

The data contain spatial coordinate information for each joint. The human height measurement range is 0.5-4.5 m, and up to 6 people can be simultaneously tracked. 0-5 means the tracked body index, and -1 (0xFF) means no body is found. The processing speed is 30 frames per second [35]. Human joint extractions tracked by the Kinect sensor show an accuracy of less than 2 mm [47].

Kinect utilizes actively emitted infrared to generate 3D depth images without illumination conditions. To reduce the calculation, the human body is separated from the background by the background subtraction method, and only the human body image is processed in the subsequent calculation [48]. By analyzing the depth-of-field image of the human body, the specific parts of the human body can be judged. The following formula is used:

$$f_\theta(I, x) = d_1\left[x + \frac{u}{d_1(x)}\right] - d_1\left[x + \frac{v}{d_1(x)}\right] \quad (1)$$

where x represents the pixel value, $d_1(x)$ denotes the depth value of x in the image, $\theta = (u, v)$, u and v represent the offsets, and $1/d_1(x)$ is used to normalize the offsets and scale the size of the body.

A decision forest can be viewed as a combination of multiple decision trees [49]. A tree is trained on an image of a body part containing the corresponding label and keeps updating the tree until the correct classification of the body part is determined. Then, the regions most likely to describe each body part are selected. For example, if the "head" classification has the greatest probability, the region is judged as a "head". Finally, the relative position of the classifier node is calculated.

B. PREPROCESSING AND FEATURE EXTRACTION

1) DATA PREPROCESSING

The data collected by Kinect contain some noise. During the preprocessing stage, this noise must first be removed. Generally, when a person walks, the movements of their body parts exhibit a certain degree of regularity, and their trajectory can be easily fitted and predicted by artificial intelligence algorithms [8, 50]. Here, the least-squares support vector regression (LSSVR) algorithm is used to fit the collected feature parameters [51]. The HHD (the distance between two hands) is taken as an example, as shown in the following formula:

$$y_{HHD}(x) = \sum_{k=1}^{L_{HHD}} \alpha_k K(x, x_k) + b \tag{2}$$

where x_k represents the SVM, K represents the kernel function, α_k represents the optimal solution of the dual problem of the SVR algorithm, and b represents the offset. $x \in X_{HHD}$, $X_{HHD} = \{1, 2, \cdots, L_{HHD}\}$. L_{HHD} is the length of the HHD data. Assuming that the true value corresponding to sample x is y_{HHD}, we now utilize a threshold ε. If $|y_{HHD}(x) - y_{HHD}| > \varepsilon$, then it is considered to be noisy data, and x is removed. Figure 4 shows that this method is used to effectively eliminate noise data. A, B, C, D, and E are noise points in the figure.

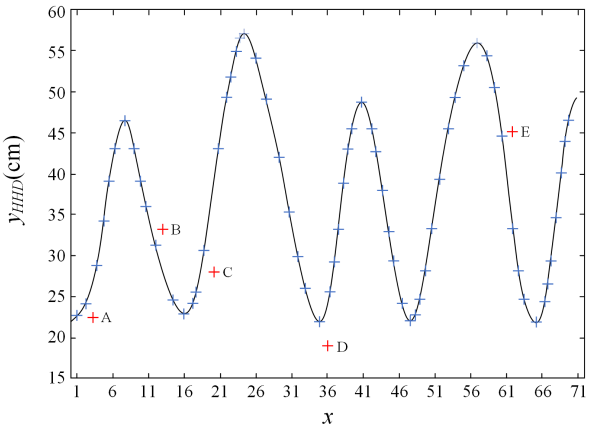


FIGURE 4. Eliminating noise points.

2) FEATURE EXTRACTION

The Euclidean distances between the various joint points of the human body are selected as features. The features are less affected by the Kinect placement position and pose when represented by relative parameters, thus improving the robustness of the algorithm. The gait features used in this article mainly include two types: static features and dynamic features. Static features refer to features that are unchanged, such as height and body length, as shown in Table 1. Dynamic features refer to features that change significantly while walking, such as the distance between two hands. The combination of these two types of features can further improve the model's ability to express human features, thereby improving the accuracy of recognition.

TABLE 1. Gait features.

	Features	Feature interpretation
Static features	HandLen	Length of upper limb
	LegLen	Length of lower limb
	Height	Body height

The length of the upper limb (HandLen) is taken as an example. Figure 3 shows that the left arm of the upper limb includes four points (5, 6, 7, and 8), and the right arm includes four points (9, 10, 11, and 12). Therefore,

$$\begin{aligned} HandLen &= D(p_5, p_6, p_7, p_8) + D(p_9, p_{10}, p_{11}, p_{12}) \\ &= \sqrt{d(p_5, p_6)} + \sqrt{d(p_6, p_7)} + \sqrt{d(p_7, p_8)} \\ &\quad + \sqrt{d(p_9, p_{10})} + \sqrt{d(p_{10}, p_{11})} + \sqrt{d(p_{11}, p_{12})} \end{aligned} \tag{3}$$

LegLen, Height and other static features can be obtained in the same way.

For a dynamic feature, the Euclidean distance between two points is calculated. A total of 300 combinations are available for the 25 points in Figure 3. We utilize the PCA algorithm to reduce the feature dimensionality to reduce computation [46], and redundant data are eliminated on the premise of ensuring the recognition accuracy of the model. Because there is no clear mathematical relationship between the feature dimensionality and target recognition rate, we obtain statistical records of the feature dimensionality and recognition rate through experiments. Then, on the premise of ensuring a high recognition rate, the PCA algorithm corresponding to the feature dimensionality with a low numerical value is selected. Experiments prove that the best recognition effect can be obtained by mapping 300-dimensional features to 80 dimensions. The results are shown in Table 2.

TABLE 2. The relationship between the feature dimensionality and the recognition rate.

Feature dimensionality	Recognition rate
100	95.36
90	95.12
85	95.23
80	95.09
75	94.25
70	93.56

C. DESIGN OF A HIGH-ACCURACY MULTICLASSIFIER

1) MULTICLASSIFIER BASED ON AN SVM

Kinect can track and lock onto human bodies in its field of view and assign an ID to each human body [52]. The recognition algorithm can track and recognize multiple people simultaneously. Gait recognition is a multiclassification problem that can be transformed into multiple dual-classification problems; finally, dual classifiers can be used to solve the

multiclassification problem [53]. In this paper, we propose a multiclassification algorithm based on the OVO and OVR classification strategies in SVM [54].

The joint data of 50 individuals were selected as test samples, and all samples of each category in the test set were separately passed through the (OVO or OVR) classifier *f(x)*. The classifier accuracy was recorded, as shown in Figure 5. The abscissa in the figure represents the category, and the ordinate represents the accuracy. When *f(x)* is the OVO classifier set, the lower accuracy limit is 91%; when *f(x)* is the OVR classifier set, the upper and lower accuracy limits are 94.6% and 76%, respectively, and the overall accuracy is lower than that of the OVO classifier set. However, when the samples of each class are only passed through the OVR classifier corresponding to each class, the accuracy is greatly improved, and the lower limit of the accuracy reaches 96.5%. If the OVO classifier set is used alone, each classifier in the OVO classifier set yields classification results (1 or - 1) [55], which are the categories contained in the training set samples. When the tested person does not exist in the training set, the OVO classifier set also gives a classification result, that is, a category included in the training set samples; obviously, this is incorrect. However, if the OVR classifier alone is used as the classifier, its accuracy cannot be guaranteed.

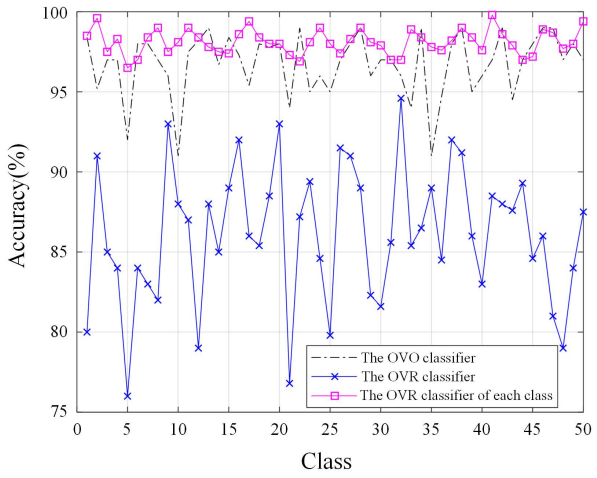


FIGURE 5. Accuracies achieved for various samples.

Therefore, according to the above analysis, by combining the OVO and OVR classification strategies, we design a classifier with high classification accuracy that can distinguish whether the tested category exists in the training set. For the OVR classifier, we set the class to which the training samples belong as the positive class and mark the Norecord class as the negative class (representing no data records). After sample xi passes through the OVO classifier set, a category is produced; then, the sample passes through the OVR classifier corresponding to the determined class. If a positive class is obtained after passing through the OVR classifier, the result class is the input class; otherwise, the result is the Norecord class. We call this multiclassification algorithm the OVO classifier set→OVR classifier algorithm, as shown in Figure 2. Not only can this algorithm identify people without data records, but it also addresses the problem that OVO classification may yield multiple outcomes. The accuracy of this method is analyzed below.

Let D = "the sample passes through the OVO classifier set", and E = "the sample passes through the OVR classifier corresponding to category C", where class C represents the class marked after the sample passes through the OVO classifier set. The classification accuracy is set as P.

$$p = P(ED) = P(D)\,P(E|D) \tag{4}$$

The corresponding OVR classifier can be selected only after event D occurs [27], so the value of $P(E|D)$ is the accuracy when *f(x)* is the OVR classifier corresponding to each class. Therefore, the classification accuracy of the samples in each class is equal to the multiplication of the upper two lines in Figure 5.

Suppose that the number of samples belonging to the i-th category (i-th person) is n_i, and the accuracy of the OVO classifier set for the i-th category is dp_i. The classification accuracy of the i-th OVR classifier for all samples in the i-th class is dep_i, and the total number of test samples is N. Then, the classification accuracy of the i-th category achieved by using the algorithm designed in this article is $p_i = dp_i \times dep_i$. The total test accuracy is

$$p_{total} = \frac{\sum_{i=1}^{50} n_i p_i}{N} \tag{5}$$

2) IMPROVING OF CLASSIFICATION ACCURACY OF THE MULTICLASSIFIER

Suppose that the classification accuracy of the SVM classification function *f(x)* on dataset s is p, and n samples $X = \{x_1, x_2, \cdots, x_n\}$ belong to class C. It can be proven that when p>0.5 and n is an odd number, the probability of more than 50% of the samples being classified as class C by the classifier *f(x)* is greater than p. The proof is as follows.

Let event A_k="k samples are classified into class C" (k=0, 1, 2,..., n) and event B="the samples classified as class C account for more than 50% of the total sample"; which class these samples are classified into is an independent event. Then, from the n-fold Bernoulli probability formula [57], we can obtain:

$$P(A_k) = C_n^k p^k (1-p)^{n-k} \tag{6}$$

The following formula is established:

$$P(B|A_k) = \begin{cases} 1, & k > \frac{n}{2} \\ 0, & k \le \frac{n}{2} \end{cases} \tag{7}$$

Applying the total probability formula, we obtain:

$$P(B) = \sum_{i=0}^{n} P(A_k)\,P(B|A_k) \tag{8}$$

Combining formulas (6) and (7), we obtain that when $k > \frac{n}{2}$,

$$P(B) = \begin{cases} \sum_{k=\frac{n}{2}+0.5}^{n} C_n^k p^k (1-p)^{n-k}, & n \text{ is odd} \\ \sum_{k=\frac{n}{2}+1}^{n} C_n^k p^k (1-p)^{n-k}, & n \text{ is even} \end{cases} \tag{9}$$

To analyze the relationships between $P(B)$, p, and n, the value of p is varied from 0 to 1 in this paper with an interval of 0.1. For each p, the value of n is varied from 2 to 50, and $P(B)$ is calculated; the results are shown in Figures 6-8.

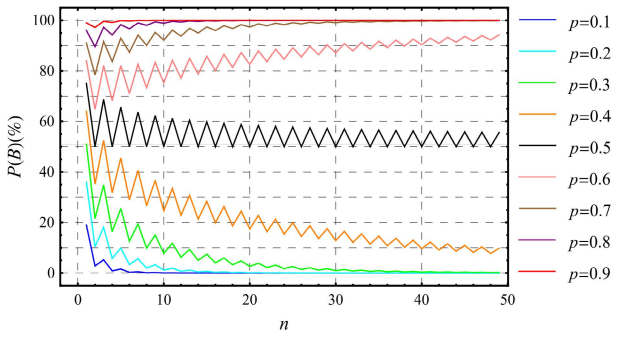


FIGURE 6. The accuracy varying with the number of test samples *n*.

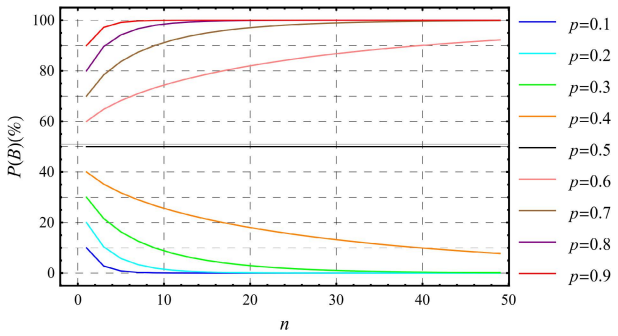


FIGURE 7. The accuracy varying when the number of test samples *n* is odd.

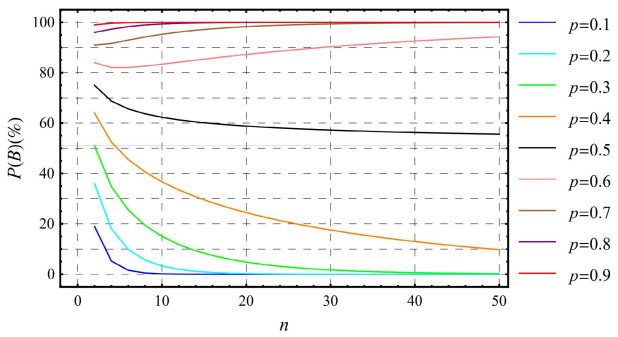


FIGURE 8. The accuracy varying when the number of test samples *n* is even.

Figure 6 shows that when $p > 0.5$, the accuracy increases in a zigzag shape with increasing n. When $p < 0.5$, the accuracy decreases in a sawtooth shape. With the increase in n, the sawtooth waveform becomes gentler, and the accuracies achieved when n is odd and even gradually approach the same limit. Figure 7 is a waveform diagram when n is an odd number. If $p > 0.5$, as n increases, $P(B)$ exhibits a monotonically increasing trend and eventually tends to 1, and $P(B) > p$. When $p = 0.5$, $P(B)$ is always equal to 0.5. If $p < 0.5$, then $P(B)$ decreases monotonously as n increases and eventually tends to 0, and $P(B) < p$. Figure 8 is a waveform diagram when n is an even number. If $p < 0.5$, $P(B)$ exhibits a monotonically decreasing trend as n increases and finally tends to 0. If $p > 0.5$, there is an integer N such that when $p < 0.5$, $P(B)$ decreases when $n < N$ and and increases when $n > N$ and finally tends to 1.

According to the above analysis, when n are odd numbers, the trend curve is smooth, and the classification effect will be better. Therefore, we set n as odd. The class of each sample classified by $f(x)$ is recorded, and we count the categories obtained and mark the category with the largest number as the final category.

Here, for formula (5), we solve for the minimum value $p_{\min}$ of p_{total}, where $N = \sum_{i=1}^{50} n_i$.

$$p_{\min} = \min_{1 \le i \le 50} (p_i) \tag{10}$$

$$p_{total} = \frac{\sum_{i=1}^{50} n_i p_i}{N} \ge \frac{p_{\min} \sum_{i=1}^{50} n_i}{N} \ge p_{\min} \tag{11}$$

Therefore, $p > p_{\min}$. According to formula (4), it is easy to obtain that $p_{low} = 88.78\% > 0.5$. According to formula (9) and Figures 7, it has been proven that when n is odd and $p > 0.5$, $P(B)$ is a monotonically increasing function. Thus,

$$\begin{aligned} P(B) &= \sum_{k=\frac{n}{2}+0.5}^{n} C_n^k p^k (1-p)^{n-k} \\ &\ge \sum_{k=\frac{n}{2}+0.5}^{n} C_n^k p_{\min}^k (1-p_{\min})^{n-k} \end{aligned} \tag{12}$$

where n is an odd number. When n is set from 3 to 15, the waveform of $P(B)$ is obtained, as shown in Figure 9:

The accuracy reaches 96.5% when n=3, and the accuracy is as high as 99.85% when n=7. The classification accuracy is greatly improved.

3) IMPROVING THE CLASSIFICATION SPEED OF THE MULTICLASSIFIER

When there are many categories to be classified, the number of OVO classifiers increases explosively, which will increase the classification time and reduce the classification efficiency [54]. To address this issue, the OVR classifier set is added before the OVO classifier set → OVR classifier algorithm because when samples pass through the OVR classifier set, only a part of the OVR classifiers obtain positive classification results, and these classes are the most likely

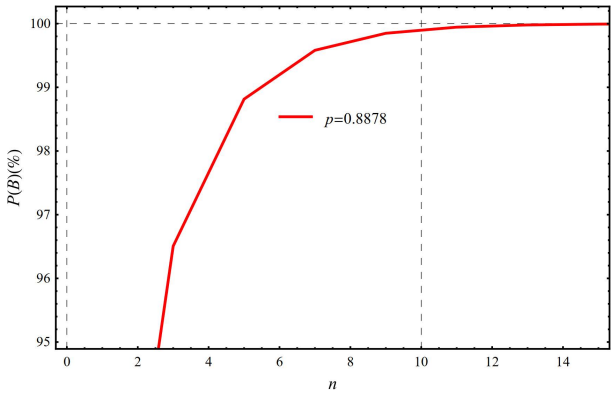


FIGURE 9. The trend curve of *P(B)* versus *n* when *p*=88.78%.

classes of the sample. Figure 5 shows that each OVR classifier has a higher classification accuracy for its corresponding class. The positive class after the sample passes through the OVR classifier set has a high probability of containing the correct class of the sample [58]. We pass all samples of each class through all OVR classifiers, record the number of positive results as n after each sample passes through the OVR classifier set, and depict the result in Figure 10. The figure shows that for approximately 24% of the samples, $n \geq 2$; for approximately 2.82% of the samples, $n \geq 10$. For passed samples, the larger n is, the smaller the number of samples. The number of the following OVO classifiers is the number of combinations of n groups of positive samples in each batch. Most of these samples are concentrated in the range when n is relatively small, so the number of combinations is also greatly reduced, so it can greatly reduce the required number of classifiers and greatly increase the classification speed. This article refers to this multiclassification algorithm as the OVR classifier set → OVO classifier set → OVR classifier algorithm, as shown in Figure 2. In the experimental section, we compare the classification efficiency of the OVR classifier set → OVO classifier set → OVR classifier algorithm with that of the OVO classifier set → OVR classifier algorithm in detail.

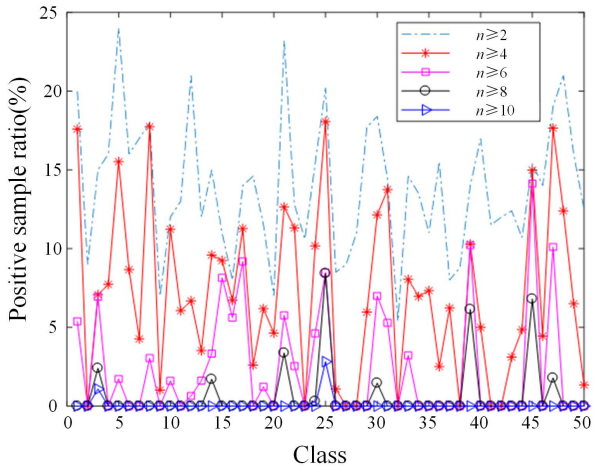


FIGURE 10. The ratio of positive classes after all samples of each class pass the OVR classifier set (the ratio of the number of positive classes greater than or equal to n to the total number of samples after passing through the OVR classification set).

IV. EXPERIMENTAL RESULTS

Fifty volunteers are selected for data collection. Each volunteer is required to walk back and forth in front of the Kinect 5 times, and on average, at least 80 sets of data can be collected each time. This article selects 20000 sets of data. The gait recognition system is divided into three modules: the sample collection module, where the main function is to collect samples; the training module, where the main function is to process the original samples and train the processed samples to generate a classifier; and the classification module, where the main function is to use the classifier generated by the training module to classify the samples. The overall process is shown in Figure 11.

FIGURE 11. Recognition software and the actual recognition process.

Additionally, we compare the performance of our proposed method with current state-of-the-art methods. For fairness, we compare the recognition results on two publicly available Kinect-based benchmark gait datasets. The UPCV gait dataset was released by Kastaniotis et al. [3] and includes gait sequences of 15 males and 15 females. There were five walking sequences for each participant, and each gait sequence consisted of approximately 55-120 frames. The video frame rate is 30 fps. Another dataset collected by Andersson and Araujo recorded skeleton-based gait biometric data for 164 individuals [37]. Participants were directed to walk along a semicircular path, and an X-Box 360 Kinect sensor was used to capture gait sequences. Each walking sequence consists of 6-12 gait cycles, approximately 500-600 frames. In this paper, the Kinect gait biometric dataset is denoted as the KGB dataset.

The performance of the proposed gait recognition method is evaluated in terms of recognition accuracy, precision, recall, and F-score. After feeding each sample to the model

and receiving a classification, the predicted versus actual classification can be counted in a table called a confusion matrix.

TABLE 3. The confusion matrix.

	Negative (predicted)	Positive (predicted)
Negative (actual)	True negative (TN)	False positive (FP)
Positive (actual)	False negative (FN)	True positive (TP)

The formulas for these four indices are: Accuracy = (TP + TN) / (TP + TN + FP + FN), Precision = TP / (TP + FP), Recall = TP / (TP + FN), F-score = (2 × Precision × Recall) / (Precision + Recall). In our study, we calculate the accuracy, precision, recall, and F-score and report the average for each class.

A. ALGORITHMIC CLASSIFICATION ACCURACY AND SPEED VERIFICATION

We analyze the accuracy that the proposed method can achieve in practical applications. We compare the accuracies achieved using the OVO classifier set, OVO classifier set → OVR classifier, OVR classifier set → OVO classifier set → OVR classifier algorithm on the same dataset. Tenfold cross-validation is used to record the accuracy of each algorithm for each fold. All algorithms are compared on the same dataset (including the training set and test set). The accuracy achieved on each fold is recorded, as shown in Figure 12.

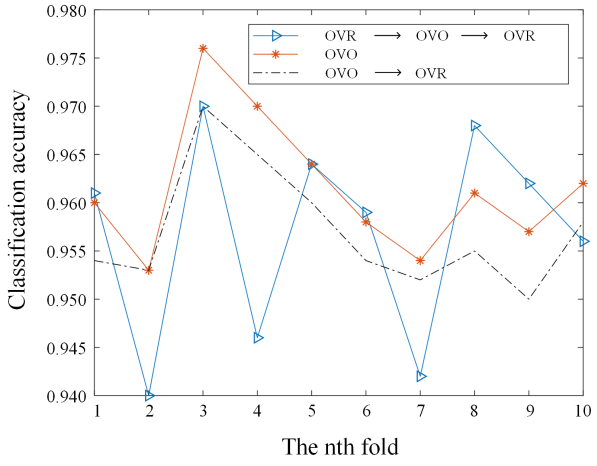


FIGURE 12. Comparison among the classification accuracies of the different algorithms when the classes of the test set are included in the training set.

When the test set sample classes are included in the training set sample classes, the OVO classifier set has higher accuracy than the OVO classifier set → OVR classifier algorithm, and the overall accuracy of the OVO classifier set → OVR classifier algorithm exceeds 88.78%. In Figure 12, the accuracy of the OVR classifier set → OVO classifier set → OVR classifier algorithm is higher than that of the OVO classifier set → OVR classifier algorithm in some intervals. This is because after the OVR classifier set is added to the front, the number of OVO classifiers after a sample passes through the OVR classifiers is reduced, and the reduction in this number may increase the accuracy of the OVO classifier set.

All the classes contained in the above test dataset exist in the training set. In practical applications, the algorithm should also be able to identify people without data records (marked with Norecord). Therefore, we add samples belonging to classes that do not exist in the training set to the test set. The test accuracies achieved on each fold are compared, as shown in Figure 13.

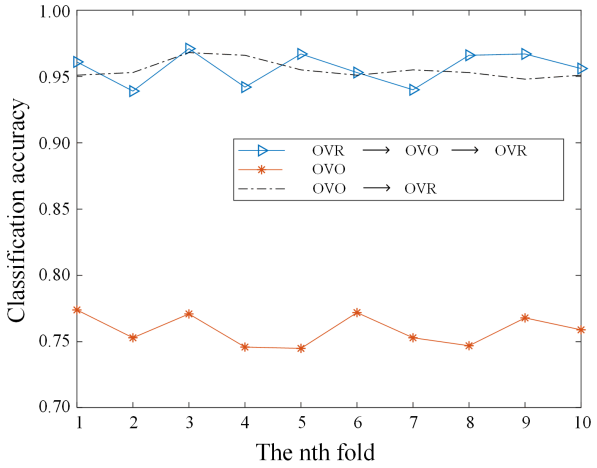


FIGURE 13. Comparison among the classification accuracies of the different algorithms when the test set contains classes that do not exist in the training set.

Comparing Figure 12 with Figure 13, it can be seen that the accuracy of the OVO classifier set is greatly reduced, while the accuracies of the two remaining multiclassifiers do not change much. This is because the OVO classifier set cannot recognize samples that do not exist in the training set; when some samples belonging to classes that do not exist in the training set are added to the test set, the OVO classifier set yields incorrect results for these samples [59]. After adding the OVR classifier set at the rear, this shortcoming of the OVO classifier set can be compensated, which also demonstrates the rationality of such a design. Figure 13 shows that the OVR classifier set -> OVO classifier set -> OVR classifier algorithm achieves high recognition accuracy for samples without data records.

From Figure 13, the average test accuracy in the tenfold cross-validation scenario is 95.68%. Here, the Bernoulli probability method is further used to improve the model accuracy. Values of n from 1 to 21 are taken consecutively, during the recognition process of a person, the recognition results of consecutive n frames are counted, and if more than 50% of the frame recognition results belong to a certain class, it is judged as that class; tenfold cross-validation is performed for each value of n, the resulting test accuracy is recorded, and the graph is drawn. To compare the difference between the actual and theoretical analyses, a theoretical change curve is

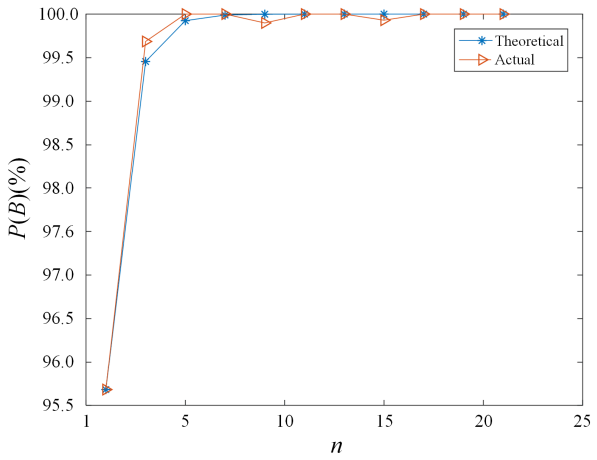


FIGURE 14. Differences between the theoretical and practical classification accuracies achieved after applying Bernoulli probability.

constructed according to formula (9), where P=0.9568. The results are shown in Figure 14.

Figure 14 shows that some deviation occurs between the actual curve and the theoretical curve; the maximum is 0.3%, and the test result is very close to that of the theoretical analysis. When $n = 3$, the accuracy exceeds 99.5%, and as n increases, the recognition accuracy gradually tends toward 100%. Taking the recognition time factor into account, this article selects $n = 5$, and the accuracy is 99.8%.

We further evaluate the performance of the proposed recognition method using cumulative matching feature (CMC) curves. CMC curves are the most popular evaluation metrics for person re-identification methods, it can show the accuracy of recognition of the algorithm for multiple chances, the probabilities are typically expressed visually through the CMC curve. The CMC curve of this method on this dataset is shown in Figure 15. The rank-1 recognition accuracy is 99.80%. The rank-4 recognition accuracy reaches 100%. Therefore, our proposed method can achieve high accuracy.

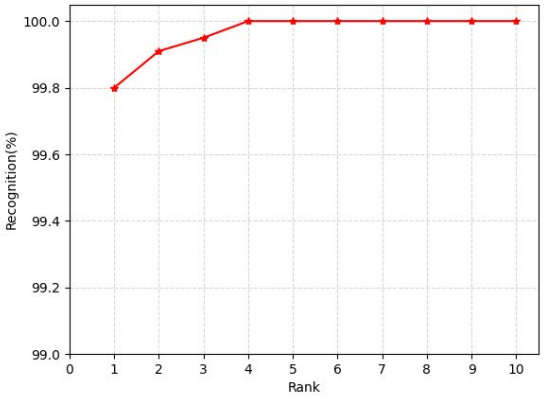


FIGURE 15. CMC curve of the proposed method on our datasets.

To test the speed of the proposed algorithm, this paper divides all the datasets into 10 equal parts, taking 9 of them as the training set and 1 as the test set. We record the number of classifiers that obtain positive results after each sample passes through the OVR classifier set->OVO classifier set->OVR classifier, and Figure 16 is obtained.

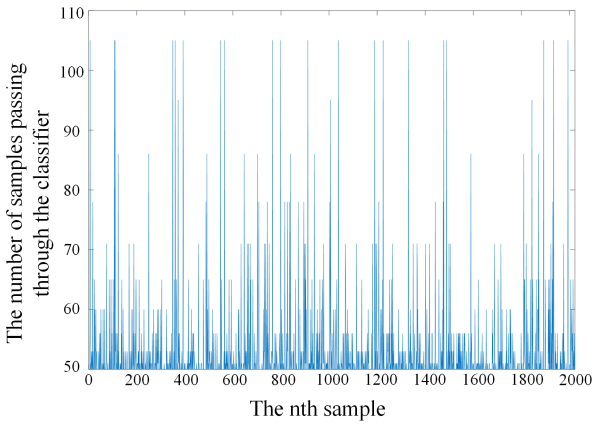


FIGURE 16. The number of classifiers passed through by the sample when tested.

Figure 16 shows that the maximum number of classifiers is 106. After subtracting the 50 OVR classifiers before the OVO classifier, according to the formula $\frac{C(C-1)}{2}$, it can be inferred that up to 11 classifiers become positive after passing through the frontal OVR classifier set. If the frontal OVR classifier set is not used, each sample must pass through at least $\frac{50(50-1)}{2} = 1225$ classifiers during the classification process. To confirm the gains brought about by this design, this paper compares versions of the proposed algorithm with the OVR classifier set and without the OVR classifier set. Using tenfold cross-validation [60], the running time of each fold is recorded for comparison purposes, as shown in Figure 17.

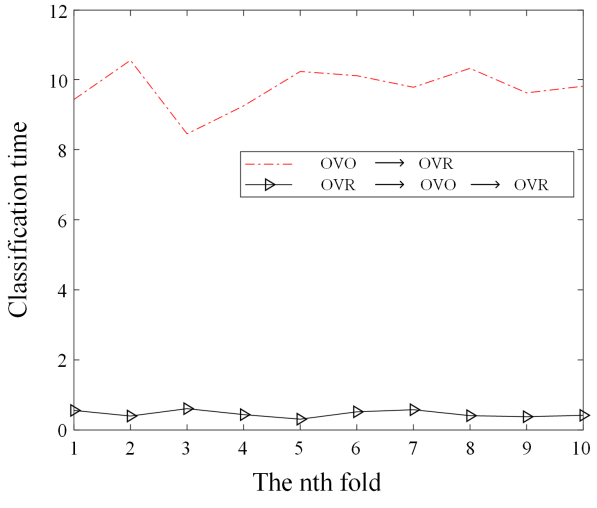


FIGURE 17. Comparison between the classification times of the two algorithms.

Figure 17 shows that the algorithm with the OVR classifier set added in the front takes only 0.61 seconds to complete

each fold, while the algorithm without the OVR classifier set in the front takes at least 8.45 seconds to complete each fold. Therefore, the gain brought by this method in practical applications is at least a 13.85-fold improvement.

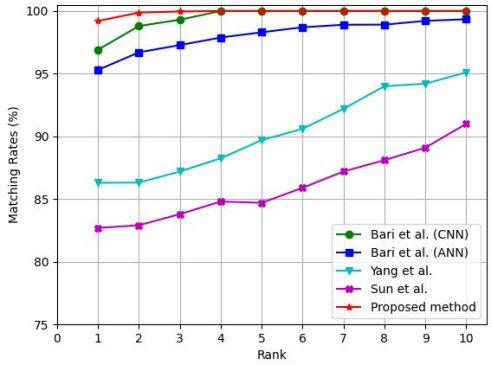
The above results are obtained by running on the CPU. To facilitate comparison with other algorithms (including deep learning algorithms), we rerun the algorithm on the GPU. The system configurations for determining the inference time are an Intel Core i7-8700 CPU of 3.20 GHz, 16 GB of RAM, and an NVIDIA GeForce GTX 1080 GPU. The time of identification of a person is 0.21 ms. It is noteworthy that the application of the n-fold Bernoulli algorithm proposed in this paper is the statistics and judgment of the results of 5 consecutive frames, which hardly increases the computational complexity of the algorithm. But the disadvantage is that the program will not output the final judgment result until after the first 5 frames to identify a person. In practical applications, it will be more affected by the sampling speed of the sensor. In general, the proposed algorithm is a multi-classification algorithm with excellent performance.

B. COMPARISON WITH RELATED WORKS

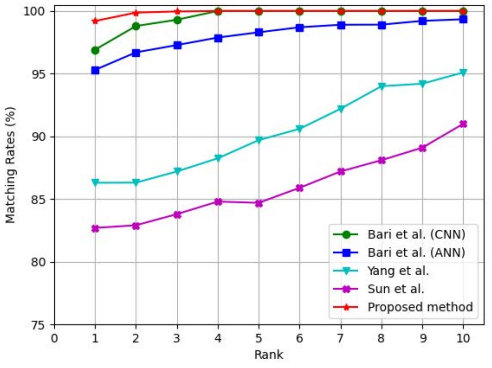
To demonstrate the efficacy of our strategy, the proposed gait recognition method's recognition performance is compared with that of state-of-the-art approaches. As mentioned above, current gait recognition methods for Kinect-based gait recognition are roughly divided into two categories: machine learning-based methods and deep learning-based methods. The works of Ball et al. [38], Preis et al. [13], Ahmed et al. [41], Sun et al. [42], and Yang et al. [15] can be classified into the former. Bari et al. [16], [45] proposed a deep learning neural network to achieve good recognition performance. We compare the proposed method with these classic methods. To be fair, the platforms are run on the same public datasets based on the same configuration of hardware.

We evaluate the performance of the proposed model using the CMC curve. The CMC curve of the algorithm on the UPCV dataset is shown in Figure 18a. The recognition accuracy at Rank-1 is 99.2% and reaches 100% at rank-4. The CMC curve of the algorithm on the KGB dataset is shown in Figure 18b. The recognition accuracy of ranking 1 is 99.81% and reaches 100% at rank-5.

The performance of the proposed method is compared with methods [13], [15], [16], [42], [45]. The proposed method achieves 100% accuracy at rank-4 on the UPCV dataset and 100% accuracy at rank-5 on the KGB dataset. In previous research, the algorithm of Bari et al. (CNN) performed the best, achieving 100% accuracy at rank-4 on the UPCV dataset and 99.98% accuracy at rank-10 on the UPCV dataset. The overall performance is slightly lower than that of the proposed algorithm. None of the other algorithms achieved 100% on rank-10. In summary, the proposed method achieves the highest recognition rate at each rank. The CMC scores are better at each rank, the normalized area under the curve is higher, and the equal error rate is lower than in prior research.



(a)



(b)

FIGURE 18. Performance comparison of the CMC scores of the proposed method with the state-of-the-art methods on the UPCV and KGB datasets. (a) On the UPCV dataset. (b) On the KGB dataset.

The performance comparisons of the proposed method and previous methods on the UPCV and KGB datasets are shown in Table 4 and Table 5, respectively.

TABLE 4. Performance comparison of the proposed method with previous methods on the UPCV dataset.

Methods	Accuracy	Precision	Recall	F-Score
Preis et al.	78.00	74.27	73.41	70.43
Sun et al.	82.67	80.50	80.19	79.67
Yang et al.	86.67	85.48	83.76	83.08
Bari et al. (ANN)	95.30	94.40	94.02	93.27
Bari et al. (CNN)	96.91	96.66	96.17	96.02
proposed method	99.20	98.79	99.17	98.71

The proposed method achieves the best recognition performance on both benchmark datasets. On the UPCV dataset, the recognition accuracy, precision, recall and F-score are 99.20%, 98.79%, 99.17% and 98.71%, respectively. On the KGB dataset, the recognition accuracy, precision, recall and F-score are 99.81%, 99.83%, 99.79% and 99.80%, respectively. Furthermore, our proposed method not only achieves higher recognition accuracy but also secures higher precision, recall, and F-score on both datasets.

TABLE 5. Performance comparison of the proposed method with previous methods on the KGB dataset.

Methods	Accuracy	Precision	Recall	F-Score
Preis et al.	75.46	77.70	75.34	73.71
Sun et al.	79.76	80.12	79.32	75.74
Yang et al.	94.88	94.67	95.02	93.92
Bari et al. (ANN)	98.08	98.00	98.26	97.81
Bari et al. (CNN)	99.33	99.36	99.35	99.33
proposed method	99.81	99.83	99.79	99.80

TABLE 6. Running time comparison of the proposed method with previous methods.

Methods	Inference time (ms)
Bari et al. (ANN)	95.36
Bari et al. (CNN)	0.38
proposed method	0.21

We compare the running time of the proposed algorithm and the two best algorithms on the same computer configuration. The results are shown in Table 6. This shows that our algorithm has a faster inference speed than the previous two algorithms.

V. CONCLUSION

Identity recognition is a multiclassification problem. The binary classification strategy based on OVO and OVR can be used to achieve multiclassification. Although the OVO classification strategy can achieve higher accuracy than the OVR classification strategy, when many categories are present, the number of classifiers in the OVO classifier set increases explosively, consuming considerable time and leading to low classification efficiency. Moreover, the OVO classifier cannot identify samples without data records. To solve these two problems, this paper designs a multiclassification algorithm combining OVR and OVO classifiers. The algorithm solves the abovementioned problems effectively; the algorithm is capable of identifying samples without data records, and its classification speed is improved by a large margin.

Although the above design realizes the recognition of samples without data records and improves the classification speed of the algorithm, it also leads to a decrease in classification accuracy. To solve this problem, this paper designs an accuracy improvement method based on the characteristics exhibited by Kinect when tracking the human body. Kinect assigns each person in its field of view an ID, and all samples extracted from this ID belong to the same person's samples. Based on this, this article continuously tests multiple samples, counts the number of marked classes for all samples, and selects the class with the largest number of samples as the target class. According to the experimental verification, when the number of consecutive test samples reaches a certain value, the accuracy is close to 100%.

ACKNOWLEDGMENT

(Qing Zhou and Jarhinbek Rasol contributed equally to this work.)

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[2] T. Inoue, M. Chikano, S. Awai, and T. Konno, "Gait recognition with 2D pose infomation using a surveillance camera," in *Proc. IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech)*, Mar. 2021, pp. 351–355, doi: 10.1109/LIFETECH52111.2021.9391846.

[3] D. Kastaniotis, I. Theodorakopoulos, C. Theoharatos, G. Economou, and S. Fotopoulos, "A framework for gait-based recognition using Kinect," *Pattern Recognit. Lett.*, vol. 68, pp. 327–335, Dec. 2015, doi: 10.1016/j.patrec.2015.06.020.

[4] M. Roccetti, G. Marfia, and A. Semeraro, "Playing into the wild: A gesture-based interface for gaming in public spaces," *J. Vis. Commun. Image Represent.*, vol. 23, no. 3, pp. 426–440, 2012, doi: 10.1016/j.jvcir.2011.12.006.

[5] J. Behrens, C. Pfüller, S. Mansow-Model, K. Otte, F. Paul, and A. U. Brandt, "Using perceptive computing in multiple sclerosis—The short maximum speed walk test," *J. NeuroEng. Rehabil.*, vol. 11, no. 1, p. 89, May 2014, doi: 10.1186/1743-0003-11-89.

[6] X. Xing, K. Wang, T. Yan, and Z. Lv, "Complete canonical correlation analysis with application to multi-view gait recognition," *Pattern Recognit.*, vol. 50, pp. 107–117, Feb. 2016, doi: 10.1016/j.patcog.2015.08.011.

[7] A. Cioppa, M. V. Droogenbroeck, and M. Braham, "Real-time semantic background subtraction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 3214–3218.

[8] B. Han, Y. Jiao, G. Liu, L. Zhang, Q. Zhu, Y. Yan, and J. Fei, "Kinematic & dynamic models of human lower extremity during the gait cycle," in *Proc. 6th Int. Conf. Control, Autom. Robot. (ICCAR)*, Apr. 2020, pp. 568–573.

[9] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, May 2018, doi: 10.1109/TNNLS.2017.2673241.

[10] S. S. Rajput and K. V. Arya, "CNN classifier based low-resolution face recognition algorithm," in *Proc. Int. Conf. Emerg. Frontiers Electr. Electron. Technol. (ICEFEET)*, Jul. 2020, pp. 1–4.

[11] R. J. Liao, S. Q. Yu, W. Z. An, and Y. Z. Huang, "A model-based gait recognition method with body pose and human prior knowledge," *Pattern Recogn*, vol. 98, Feb. 2020, Art. no. 107069, doi: 10.1016/j.patcog.2019.107069.

[12] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A review of vision-based gait recognition methods for human identification," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, Dec. 2010, pp. 320–327.

[13] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien, "Gait recognition with Kinect," in *Proc. 1st Int. Workshop Kinect Pervasive Comput.*, New Castle, U.K., 2012, pp. 1–4.

[14] F. Ahmed, P. P. Paul, and M. L. Gavrilova, "DTW-based kernel and rank-level fusion for 3D gait recognition using Kinect," *Vis. Comput.*, vol. 31, nos. 6–8, pp. 915–924, Jun. 2015, doi: 10.1007/s00371-015-1092-0.

[15] K. Yang, Y. Dou, S. Lv, F. Zhang, and Q. Lv, "Relative distance features for gait recognition with Kinect," *J. Vis. Commun. Image Represent.*, vol. 39, pp. 209–217, Aug. 2016.

[16] A. S. M. H. Bari and M. L. Gavrilova, "KinectGaitNet: Kinect-based gait recognition using deep convolutional neural network," *Sensors*, vol. 22, no. 7, p. 2631, Mar. 2022.

[17] C. Liu, X. Yue, J. Zhang, and K. Shi, "Active disturbance rejection control for delayed electromagnetic docking of spacecraft in elliptical orbits," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 3, pp. 2257–2268, Jun. 2022.

[18] J. Luo and T. Tjahjadi, "Gait recognition and understanding based on hierarchical temporal memory using 3D gait semantic folding," *Sensors*, vol. 20, no. 6, p. 1646, Mar. 2020, doi: 10.3390/s20061646.

[19] B. Jawed, O. O. Khalifa, and S. S. N. Bhuiyan, "Human gait recognition system," in *Proc. 7th Int. Conf. Comput. Commun. Eng. (ICCCE)*, Sep. 2018, pp. 89–92.

[20] X. Li and Y. Chen, "Gait recognition based on structural gait energy image," *J. Comput. Inf. Syst.*, vol. 9, no. 1, pp. 121–126, 2013.
[21] E. Zhang, Y. Zhao, and W. Xiong, "Active energy image plus 2DLPP for gait recognition," *Signal Process.*, vol. 90, no. 7, pp. 2295–2302, 2010.
[22] T. H. W. Lam, K. H. Cheung, and J. N. K. Liu, "Gait flow image: A silhouette-based gait representation for human identification," *Pattern Recognit.*, vol. 44, no. 4, pp. 973–987, Apr. 2011.
[23] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2002, pp. 372–377.
[24] C. Yam, M. Nixon, and J. N. Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recognit.*, vol. 37, no. 5, pp. 1057–1072, 2004.
[25] J. Luo and T. Tjahjadi, "View and clothing invariant gait recognition via 3D human semantic folding," *IEEE Access*, vol. 8, pp. 100365–100383, 2020, doi: 10.1109/ACCESS.2020.2997814.
[26] C. Zhang, Y. Tian, and E. Capezuti, "Privacy preserving automatic fall detection for elderly using RGBD cameras," in *Proc. Int. Conf. Comput. Handicapped Persons*. Berlin, Germany: Springer, 2012, pp. 625–633.
[27] E. E. Stone and M. Skubic, "Fall detection in homes of older adults using the Microsoft Kinect," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 290–301, Jan. 2015.
[28] Y. Maret, D. Oberson, and M. Gavrilova, "Real-time embedded system for gesture recognition," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 30–34.
[29] A. Al-Kaff, F. M. Moreno, A. de la Escalera, and J. M. Armingol, "Intelligent vehicle for search, rescue and transportation purposes," in *Proc. IEEE Int. Symp. Saf., Secur. Rescue Robot. (SSRR)*, Oct. 2017, pp. 110–115.
[30] H. I. Osman, F. H. Hashim, W. M. Diyana Wan Zaki, and A. B. Huddin, "Entryway detection algorithm using Kinect's depth camera for UAV application," in *Proc. IEEE 8th Control Syst. Graduate Res. Colloq. (ICSGRC)*, Aug. 2017, pp. 77–80.
[31] H. Wang, C. Zhang, Y. Song, and B. Pang, "Information-fusion based robot simultaneous localization and mapping adapted to search and rescue cluttered environment," in *Proc. 18th Int. Conf. Adv. Robot. (ICAR)*, Jul. 2017, pp. 511–517.
[32] R. Banerjee, A. Sinha, and K. Chakravarty, "Gait based people identification system using multiple switching Kinects," in *Proc. 13th Int. Conf. Intellient Syst. Design Appl.*, Dec. 2013, pp. 182–187.
[33] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. CVPR*, Jun. 2011, pp. 1297–1304.
[34] K. Khoshelham, "Accuracy analysis of Kinect depth data," in *Proc. ISPRS Workshop Laser Scanning*, 2011, vol. 38, no. 1, pp. 1–6.
[35] A. Schmitz, M. Ye, R. Shapiro, R. Yang, and B. Noehren, "Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system," *J. Biomech.*, vol. 47, no. 2, pp. 587–591, 2014.
[36] R. A. Clark, B. F. Mentiplay, E. Hough, and Y. H. Pua, "Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives," *Gait Posture*, vol. 68, pp. 193–200, Feb. 2019.
[37] V. O. Andersson and R. M. Araujo, "Person identification using anthropometric and gait data from Kinect sensor," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 425–431.
[38] A. Ball, D. Rye, F. Ramos, and M. Velonaki, "Unsupervised clustering of people from 'skeleton' data," in *Proc. 7th Annu. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, Mar. 2012, pp. 225–226.
[39] T. Banerjee, J. M. Keller, and M. Skubic, "Resident identification using Kinect depth image data and fuzzy clustering techniques," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2012, pp. 5102–5105.
[40] A. T. Sabir, M. H. Ahmed, A. K. Faeq, and H. S. Maghdid, "Human gait identification using Kinect sensor," *Kurdistan J. Appl. Res.*, vol. 2, no. 3, pp. 142–146, Aug. 2017.
[41] M. H. Ahmed, A. T. Sabir, and H. S. Maghdid, "Kinect-based human gait recognition using triangular gird feature," in *Proc. 1st Int. Conf. Adv. Res. Eng. Sci. (ARES)*, Jun. 2018, pp. 1–6.
[42] J. Sun, Y. Wang, J. Li, W. Wan, D. Cheng, and H. Zhang, "View-invariant gait recognition based on Kinect skeleton feature," *Multimed. Tools Appl.*, vol. 77, no. 19, pp. 24909–24935, 2018.
[43] C. Nattee and N. Khamsemanan, "A deep neural network approach for model-based gait recognition," *Thai J. Math.*, vol. 17, no. 1, pp. 89–97, 2019.
[44] T. Huynh-The, C.-H. Hua, N. A. Tu, and D.-S. Kim, "Learning 3D spatiotemporal gait feature by convolutional network for person identification," *Neurocomputing*, vol. 397, pp. 192–202, Jul. 2020.
[45] A. S. M. H. Bari and M. L. Gavrilova, "Artificial neural network based gait recognition using Kinect sensor," *IEEE Access*, vol. 7, pp. 162708–162722, 2019.
[46] K. Xi and C. Cai, "Feature selected based on PCA and optimized LMC," in *Proc. MATEC Web Conf.*, vol. 336, 2021, pp. 1–6, doi: 10.1051/matecconf/202133606034.
[47] S. T. Pöhlmann, E. F. Harkness, C. J. Taylor, and S. M. Astley, "Evaluation of Kinect 3D sensor for healthcare imaging," *J. Med. Biol. Eng.*, vol. 36, no. 6, pp. 857–870, 2016.
[48] A. Djerida, Z. Zhao, and J. Zhao, "Background subtraction in dynamic scenes using the dynamic principal component analysis," *IET Image Process.*, vol. 14, no. 2, pp. 245–255, Feb. 2020, doi: 10.1049/iet-ipr.2018.6095.
[49] P. T. Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery," *Sensors*, vol. 18, no. 2, p. 18, Dec. 2017, doi: 10.3390/s18010018.
[50] F. S. Chagas, L. D. P. D. Farias, M. Bozza, and P. F. F. Rosa, "A tenacity learning algorithm for humanoid robot locomotion based on the human gait cycle," in *Proc. IEEE 29th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2020, pp. 519–524.
[51] S. Zhang, T. Zhou, L. Sun, W. Wang, and B. Chang, "*LSSVR* model of G-L mixed noise-characteristic with its applications," *Entropy*, vol. 22, no. 6, p. 629, Jun. 2020, doi: 10.3390/e22060629.
[52] G. Akbari, M. Nikkhoo, L. Wang, C. P. C. Chen, D.-S. Han, Y.-H. Lin, H.-B. Chen, and C.-H. Cheng, "Frailty level classification of the community elderly using Microsoft Kinect-based skeleton pose: A machine learning approach," *Sensors*, vol. 21, no. 12, p. 4017, Jun. 2021, doi: 10.3390/s21124017.
[53] M. Rucco, F. Giannini, K. Lupinetti, and M. Monti, "A methodology for part classification with supervised machine learning," *Artif. Intell. Eng. Design, Anal. Manuf.*, vol. 33, no. 1, pp. 100–113, Feb. 2019, doi: 10.1017/S0890060418000197.
[54] H. Liu, W. Zheng, G. Sun, Y. Shi, Y. Leng, P. Lin, R. Wang, Y. Yang, J.-F. Gao, H. Wang, K. Iramina, and S. Ge, "Action understanding based on a combination of one-versus-rest and one-versus-one multi-classification methods," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–5.
[55] Y. Zheng, X. Li, Y. Si, W. Qin, and H. Tian, "Hybrid deep convolutional neural network with one-versus-one approach for solar flare prediction," *Monthly Notices Roy. Astronomical Soc.*, vol. 507, no. 3, pp. 3519–3539, Sep. 2021, doi: 10.1093/mnras/stab2132.
[56] J. Rohmer, "Uncertainties in conditional probability tables of discrete Bayesian belief networks: A comprehensive review," *Eng. Appl. Artif. Intell.*, vol. 88, Feb. 2020, Art. no. 103384, doi: 10.1016/j.engappai.2019.103384.
[57] Z. Su, H. Ji, and Y. Zhang, "A Poisson multi-Bernoulli mixture filter with spawning based on Kullback–Leibler divergence minimization," *Chin. J. Aeronaut.*, vol. 34, no. 11, pp. 154–168, Nov. 2021, doi: 10.1016/j.cja.2020.11.015.
[58] H. Faris, M. Habib, M. Faris, M. Alomari, and A. Alomari, "Medical speciality classification system based on binary particle swarms and ensemble of one vs. rest support vector machines," *J. Biomed. Informat.*, vol. 109, Sep. 2020, Art. no. 103525, doi: 10.1016/j.jbi.2020.103525.
[59] K. Hassine, A. Erbad, and R. Hamila, "Important complexity reduction of random forest in multi-classification problem," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 226–231.
[60] J. Pohjankukka, T. Pahikkala, P. Nevalainen, and J. Heikkonen, "Estimating the prediction performance of spatial models via spatial k-fold cross validation," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 10, pp. 2001–2019, Oct. 2017, doi: 10.1080/13658816.2017.1346255.

QING ZHOU received the Ph.D. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2019. Since March 2019, he has been a Research Associate with the Perception and Collaboration Laboratory, Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, China. His research interests include machine learning, electronic technology, visual measurement, navigation, and control.

JARHINBEK RASOL is currently pursuing the Ph.D. degree with the Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, China. His research interests include unmanned aerial vehicle, computer vision, and image processing.

YUELEI XU is currently a Professor with the Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, China, where he is also the Leader of the Perception and Collaboration Laboratory. His research interests include image processing, pattern recognition, object detection, and brain-like computation.

ZHAOXIANG ZHANG is currently an Associate Professor with the Perception and Collaboration Laboratory, Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, China. His research interests include transfer learning, causal inference, and advanced methods of synthetic aperture radar (SAR) processing.

LUJUAN HU received the bachelor's degree from the North University of China, Taiyuan, China, in 2020. She is currently pursuing the master's degree with Northwestern Polytechnical University, Xi'an, China. Her research interests include track fusion and deep learning.