Gait Recognition Using Optical Motion Capture: A Decision Fusion Based Method

Li Wang [1], Yajun Li [2],*, Fei Xiong [3] and Wenyu Zhang [4]

[1] School of Physical Education, Sichuan Normal University, Chengdu 610101, China; 20060132@sicnu.edu.cn
[2] Department of Physical Education, Central South University, Changsha 410083, China
[3] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China; xiongf@bjtu.edu.cn
[4] Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China; wyzhang@ustb.edu.cn
* Correspondence: liyajun@csu.edu.cn

Abstract: Human identification based on motion capture data has received signification attentions for its wide applications in authentication and surveillance systems. The optical motion capture system (OMCS) can dynamically capture the high-precision three-dimensional locations of optical trackers that are implemented on a human body, but its potential in applications on gait recognition has not been studied in existing works. On the other hand, a typical OMCS can only support one player one time, which limits its capability and efficiency. In this paper, our goals are investigating the performance of OMCS-based gait recognition performance, and realizing gait recognition in OMCS such that it can support multiple players at the same time. We develop a gait recognition method based on decision fusion, and it includes the following four steps: feature extraction, unreliable feature calibration, classification of single motion frame, and decision fusion of multiple motion frame. We use kernel extreme learning machine (KELM) for single motion classification, and in particular we propose a reliability weighted sum (RWS) decision fusion method to combine the fuzzy decisions of the motion frames. We demonstrate the performance of the proposed method by using walking gait data collected from 76 participants, and results show that KELM significantly outperforms support vector machine (SVM) and random forest in the single motion frame classification task, and demonstrate that the proposed RWS decision fusion rule can achieve better fusion accuracy compared with conventional fusion rules. Our results also show that, with 10 motion trackers that are implemented on lower body locations, the proposed method can achieve 100% validation accuracy with less than 50 gait motion frames.

Keywords: decision fusion; gait recognition; kernel ELM; optical motion capture; sensor fusion

1. Introduction

Human identification using motion capture data has attracted much attention for its wide applications in authentication, surveillance, and medical applications [1]. Some commonly used identification methods, such as face recognition and fingerprint, are not preferred by some people who have sensitive privacy considerations [2]. Gait recognition does not require the privacy data of the target person, and therefore is one promising alternative that can be used for human identification. In addition, when the facial appearance of the target person is unavailable or changeable, it can be used for tracking the persons of interest in public security with the gait information of target persons [3].

In gait recognition, the system authenticates or classifies the target person by using her/his walking manner. The used gait data mainly include the following four modalities: camera image gait data [4], inertial acceleration sensor data [5], floor sensor data, and passive wireless signals [6] or wave radar data [7]. In addition, the recognition performance can be further improved by integrating the above four data modalities [8]. The optical

motion capture systems (OMCSs), such as the Vicon [9], can obtain highly precise gait motion data of the target persons, and have been widely used in sports training [10], animation [11], medical analysis [12], and robotics [13]. For example, in sports training, using an optical motion capture system can accurately record the body motions, which is helpful for the coach and the athletes to analyze the characteristics of the motions, and provide new training guides for improving training performance [14]. Similarly, in medical analysis, the OMCSs can be used to capture the motion behavior of patients with Parkinson's disease, and can be used as a diagnostic basis and in Parkinson treatment [15].

Existing gait recognition works mainly rely on sensor data or image data for classifying the object, and the recognition accuracy is not high enough for some applications with stringent accuracy requirements. The OMCSs can obtain high-precision gait motions, which may be helpful for improving the gait recognition performance, but its potentials in gait recognition has not been studied. Though it requires the implementation of optical motion trackers, the system does not require sensitive information like face image and fingerprint, thus it will be acceptable for some users with privacy concerns. In addition, OMCS is widely used in sports training, physical treatment, and visual effects production applications [11–15]. Existing work only support one player at one time, and it is necessary to develop gait recognition methods if we want to classify different persons in a multiple player scenario. In this way, the efficiency and capability of an OMCS can be greatly improved.

As such, in this paper we attempt to use an OMCS for human identification, and our goal is to achieve state of art identification accuracy with 10 optical trackers that are planted on the lower body locations of the target person, including thigh, lower leg, and foot, and both left and right body sides. The proposed method only requires the collection of several gait cycles of the participants, and the tracked body locations are chosen as the lower body part, which reduces the intrusion degree compared with other body parts such as chest and head. We propose a decision fusion-based gait recognition method, and compared with existing works, the proposed method has the following distinguished differences:

(1) We use an OMCS to record the gait trajectories of the persons and study the gait recognition problem with the obtained data, which has not been considered in existing works. Though OMCS requires the implementation of optical trackers, the system can obtain high-precision gait motion data, which is helpful to improve the classification accuracy when only using a relative smaller number of gait motions for human classification. In addition, the system does not require a large number of optical trackers, and using only 10 optical trackers can achieve very good classification accuracy, and even 100% accuracy when sample number is large enough.

(2) We use a powerful classifier, namely kernel extreme learning machine (KELM) [16], to conduct the classification process for the extracted feature data. Existing works mainly use SVM and random forest as the classifier [1]; in the considered problem of this paper, we show that KELM can achieve much higher classification accuracy and efficiency compared with the above two competitive classifiers.

(3) Instead of focusing on improving the classification accuracy of the base classifier, in this paper we highlight the importance of decision fusion of multiple motion frames, and we develop a first-classification-then-fusion method to achieve better classification performance, and the design of the decision fusion rules plays a vital role in improving the fusion performance. More specifically, we first conduct the classification process for each single motion frame, and then propose a reliability weighted sum (RWS) rule for combining the classification decisions of multiple motion frames. In the proposed RWS rule, we first transform the outputs of the KELM classifier into fuzzy decisions by using a membership transformation function, then compute the consistency matrix of all the fuzzy decisions by using a consistency degree measurement. With the obtained consistency matrix, we then use the Eigenvalue decomposition method to obtain the weight vector of the fuzzy decisions. A fuzzy decision with larger average consistency value to other fuzzy

decisions means that it is more reliable, and the corresponding reliability weight value will be relatively larger, otherwise it will be relatively smaller.

(4) The experimental results on the collected gait motion data with 76 participants demonstrate that the proposed rule outperforms several decision rules, including sum rule, belief rule, weighted belief rule, product rule, and majority voting rule. The results also show that the proposed method can achieve 100% classification accuracy with 10 motion optical trackers, which shows its potentials for its applications in authentication and multiplayer motion tracking applications.

The remainder of this paper is organized as follows: Section 2 gives a brief review to related work, the detailed process of the proposed decision fusion-based gait recognition method is shown in Section 3. Experimental results are shown in Section 4. Section 5 concludes the paper.

2. Related Work

Gait recognition has been extensively studied in for its wide applications in authentication, surveillance, training, and medical treatment [1–3]. In most existing methods, the gait recognition methodology mainly includes the following four steps: feature extraction, dimension reduction, and classification [1]. The gait motion data contains both the locations and velocity data; thus, it is more related to image gait data and acceleration sensor data. From the perspective of tracker implementation way, the acceleration sensor data is similar to the optical motion tracker used in this paper, since they all need to collected from the tracker sensors deployed on different locations of the human body. The difference is that accelerator cannot track the 3D positions of the gait, and it requires the system to collect sufficient data samples (e.g., more than 1000) to achieve a relative better classification performance [17]. For the optical motion tracker based system, its deployment cost is higher compared with inertial sensor system, but it can obtain high-quality and precise trajectories of the optical trackers. Since the precise gait motion trajectories are recorded, the proposed system can achieve good performance with very limited number of data samples or a few gait cycles. On the other hand, the image gait data can be regard as a 2D rejection of the 3D gait motions, and a set of gait images contain the trajectories of the monitored person. However, due to the limitation of image resolution, it is hard to extract precise 3-D precise trajectories from the image data without the assistance of range sensor, which decreases the recognition accuracy of the system [18].

In the feature extraction step, the system needs to extract expressive feature representations from the raw data. Although the quality of the extracted features has significant impacts on the classification accuracy of the gait recognition task, yet it is still not unified that which feature is the best choice that suits all situations. For image data, existing feature extraction methods can be categorized as model-based methods and model-free methods [1]. In model-based feature extraction, the features depicting the human body characteristics are extracted, typical features including stride length and step frequency [19], body distances [20], gait template [21], and velocity Hough transform [22]. According to the summary of Tang [23], the commonly used for extracting the acceleration data feature mainly include mean value, standard deviation, range, signal energy, spectral entropy, bandpower, etc. Some other features, including number of zero crossings, inter quartile range, average peak length, spectral edge frequency, are also used in gait recognition systems [24]. In this paper, we will also use relative distance feature and velocity features as in the input of the classifiers, which belongs to the methodology of model-based feature extraction. It has been shown that using relative distance of the body locations can achieve good classification performances since they are able to depict both body shape and motion actions [20]. In addition, experiments have shown that the walking speed is highly correlated to the age and gender of the target persons [25], thus it is also adopted in the proposed gait recognition method.

In the classification step, the most commonly used classifiers include support vector machine (SVM) [26], random forest (or Decision Tree Ensemble Classifier) [27]. The deep

learning based method are also popular for gait recognition, such as deep convolutional neural networks (CNN) [28], deep autoencoder-decoder [29], and long short-term memory (LSTM) network [30]. In this paper, we want to train the classifier with only a limited training data; thus, deep learning methods are not suitable in this situation since they need large enough data to train the model. We will use a kernel extreme learning machine (KELM) classifier to conduct the gait recognition process for single frame motion data, and we will prove its superiority in classification accuracy compared with SVM and random forest.

We emphasize that the gait recognition problem in this paper is different from the above existing works, in which the classifier usually requires a very large training data samples, and each sample is composed by several gait cycles to achieve good classification performance. In the proposed method, we only record several gait cycles for training the classifier for single motion classification, and 1–2 gait cycles for classification, which is efficient for data collection process. However, the available number of gait motion data is quite limited and cannot be solved by methods that require a larger number of training and test data. To fully use the decisions of each single motion frame, in this paper we first propose a method that first conducts frame-level classification, and then develop a decision fusion method to combine the decisions of the single frames, while in existing work the decision is directly made by the output of the classifier, and each classification requires a bunch of gait motion data frames. In the following section, we will illustrate the detailed steps of proposed gait recognition method using decision fusion.

3. Decision Fusion Based Gait Recognition

3.1. The Gait Motion Tracking System

We use optical motion trackers to record high-precision body motion trace to identify the target person. Before identification, the participants are required to wear a set of optical trackers, and then naturally and straightly walk through a flat test field with a length less than 5 m, which can be done in a very short period of time. There are 76 participants in total, with 46 females and 30 males, and their ages range from 20 to 60. The heights of the participants range from 144 to 178 cm, and the weights ranges from 42 to 115 kg. The sampling frequency of the body locations is 5 Hz, and there are 10 lower-body locations are recorded in each frame, which include thigh, knees, shin, ankle, and tiptoe, and both left and right sides are covered. The obtained walking lengths of the participants are different, which range from 2.37 to 4.15 m. An example of the recorded gait data is shown in Figure 1, in which 5 gait motion samples of one participant is plotted. Visually, we can see that it is hard to identify the target person without using a proper gait recognition method.
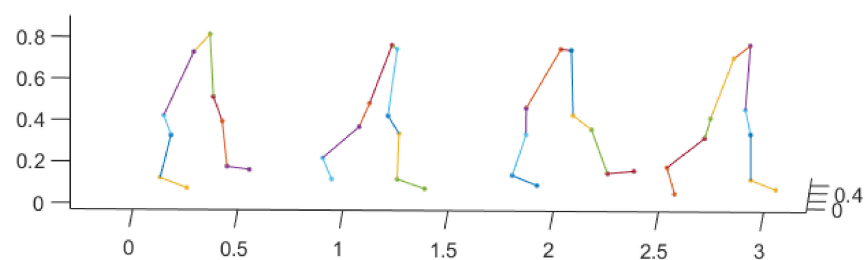
Figure 1. An example of the 3D gait motion track data.

In this paper, we consider a gait recognition problem using high-precision optical gait motion trackers. Before identifying the target, the system previously prepared a set of training dataset with N different persons, denoted as $\mathcal{T} = \{\{(\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1), c_1\}, \ldots; \{(\mathbf{x}_N, \mathbf{y}_N, \mathbf{z}_N), c_N\}\}$, where $\mathbf{x}_i = \{x_{i,1}, \ldots; x_{i,10}\}$, $\mathbf{y}_i = \{y_{i,1}, \ldots; y_{i,10}\}$, and $\mathbf{z}_i = \{z_{i,1}, \ldots; z_{i,10}\}$ denote the recorded 3D coordinates of the trackers of person i, and c_i denotes its label. Note that the training data must at least contain a complete walking cycle of the person. Given a new person to be identified, the system record T consecutive motion frames,

denote as $\mathcal{X} = \left\{ \left(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}, \mathbf{z}^{(1)} \right), \ldots; \left(\mathbf{x}^{(T)}, \mathbf{y}^{(T)}, \mathbf{z}^{(T)} \right) \right\}$, where $\mathbf{x}^{(t)} = \{x_1^{(t)}, \ldots; x_{10}^{(t)}\}$, $\mathbf{y}^{(t)} = \{y_1^{(t)}, \ldots; y_{10}^{(t)}\}$, and $\mathbf{z}^{(t)} = \{z_1^{(t)}, \ldots; z_{10}^{(t)}\}$ denote the 3D coordinates of t-th gait motion frame. When the classifier is properly trained, our problem becomes identifying the target person according to the input motion data $\mathcal{X}$. As shown in Figure 2, the proposed first-classification-then-fusion method mainly includes the following four steps:

Figure 2. The flowchart of the proposed gait recognition method.

(1) **Feature exaction**: For each person, the input raw data includes 10 motion tracks, and they cannot be directly used to classify the target object. In this paper, we want to identify the target person with a short gait motion capture trace; extracting the features from a trace recorded from motion tracker time series is not practical in this situation since each person may only has several gait cycles, and the number of collected gait motion is quite limited. As such, a feature exaction process that only extracts relative location distance and speed features from single frame data is proposed to obtain an expressive feature representation of the input data, and then the extracted features will be the input of the following identification process.

(2) **Unreliable feature calibration**: Though the OMCS can record high precise gait trajectory data, we observe that a recorded motion instance may be biased due to sensing failure or noise interference, and the corresponding the features of the biased motion data are also unreliable. Therefore, it is necessary to detect and calibrate the unreliable feature data, and relief their impact on the classification performance.

(3) **Classification**: In this paper, we will use a kernel extreme learning machine (KELM) to deal with the classification task for the feature data of each single gait motion frame. We will provide performance comparison results to demonstrate the advantages on classification accuracy and efficiency of KELM for the gait classification in the experimental section.

(4) **Decision fusion**: The motion frame number of different persons are also different due the variation of walking speed of the target person. With the obtained KELM outputs of all the frames, we then need to combine them into a unified one to obtain the final global decision. Compared with single frame classification, we expect a classification accuracy improvement after combining the decision of multiple frames, and the decision fusion rule will play a vital rule on the final fusion accuracy.

In the next subsections, we will give a detailed illustration of the above four steps, along with their mathematical models.

3.2. Feature Extraction

Given a gait motion data, we use the relative distance of the tracker locations as the feature of the input motion data. The reason is that the relative distance metric can depict both walking action and physical body shape characteristics, which can effectively distinguish the differences between gait motions of two persons. In frame (t), for two

trackers i, j with coordinates $(x_i^{(t)}, \ldots; x_i^{(t)})$ and $(x_j^{(t)}, \ldots; x_j^{(t)})$, their Euclidean distance can be computed by

$$d_{i,j}^{(t)} = \sqrt{\left(x_i^{(t)} - x_j^{(t)}\right)^2 + \left(y_i^{(t)} - y_j^{(t)}\right)^2 + \left(z_i^{(t)} - z_j^{(t)}\right)^2} \tag{1}$$

In this way, we can obtain a pairwise distance matrix $\mathbf{D} = \left[d_{i,j}\right]_{\{10 \times 10\}}$ that contains the relative distance metrics between the 10 optical tackers, in which $d_{i,i}^{(t)} = 0$ and $d_{i,j}^{(t)} = d_{j,i}^{(t)}$. Since $\mathbf{D}^{(t)}$ is symmetrical, and diagonal elements are all 0, we only use the elements of upper triangle or lower triangle of matrix $\mathbf{D}^{(t)}$ as the feature representations. Let $\mathbf{d}^{(t)} = \{d_1^{(t)}, \ldots, d_{K_d}^{(t)}\}$ be the one-dimensional vector that contains all the elements of upper triangle matrix of $\mathbf{D}$, then we can know the feature dimension is $K_d = 10 \times 9/2 = 45$.

It has been shown that gait speed can be used as features for human identification and age prediction [23], thus except the relative distances among the optical trackers, we also include the x-axis speed of each tracker into the features vector. For tracker i, the x-axis coordinates in frames (t) and $(t+1)$ are $x_i^{(t+1)}$ and $x_i^{(t)}$, then we can estimate the x-axis velocity as follows:

$$v_i^{(t)} = \frac{1}{\tau}\left(x_i^{(t+1)} - x_i^{(t)}\right) \tag{2}$$

where τ denotes the time period between two consecutive gait motion frames. In this paper, $\tau = 0.2$ s. In this way, we can obtain a velocity feature vector $\mathbf{v}^{(t)} = \{v_1^{(t)}, \ldots, v_{K_v}^{(t)}\}$, and $K_v = 10$. Combined with the relative distance features, we finally obtain the gait feature vector of motion frame (t) as $\mathbf{z}^{(t)} = \{\mathbf{d}^{(t)}; \mathbf{v}^{(t)}\}$. Note that, before the classification step, the features obtained need to be normalized since their magnitudes are different.

3.3. Unreliable Feature Calibration

Due to sensing failure or noise interference, the obtained track data may be biased, and even become outliers. Accordingly, the obtained features also will be unreliable, and will cause negative impacts on the classification performance of the gait recognition task; thus, the unreliable features need to be detected and calibrated. In this paper, we use a hypothesis test method to identify the unreliable features, in which we first estimated the probability density function (PDF) of each feature, then find the unreliable features that is larger or lower than the probability thresholds. Since the feature is irregularly distributed, and cannot be reasonably depicted by one specified distribution, thus we use the kernel density estimation method to estimate the PDFs of the features. Given a feature vector $\mathbf{z} = \{z_1, \ldots, z_L\}$, where L denotes the number of feature data. The estimated PDF at point z is estimated as follows [31]

$$f(z) = \frac{1}{L}\sum_{i=1}^{L}\mathcal{K}\left(\frac{z - z_i}{\rho}\right) \tag{3}$$

where $\rho > 0$ denotes that kernel parameter. In this paper, we use radial basis function as the kernel function, i.e., $\mathcal{K}\left(\frac{z-z_i}{\rho}\right) = \exp(-\frac{(z-z_i)^2}{\rho})$, and the bandwidth parameter ρ is computed by $\rho = \sigma(\frac{4}{3L})^{0.2}$ [32]. With the obtained PDF, we then can obtain the corresponding cumulative distribution function (CDF) by $\Phi(z) = \int_o^z p(z)\mathrm{d}z$. Let P_{th} be the probability threshold that decides whether the feature is unreliable or not. In other word, a feature data z will be regarded as unreliable when it satisfies the following condition:

$$\Phi(z) > P_{th} \text{ or } \Phi(z) < 1 - P_{th} \tag{4}$$

The above criterion requires us to compute the CDF $\Phi(z)$ every time, which is not efficiency enough. We can also compute the corresponding upper bound z_{ub} and lower

bound z_{lb} with respect to P_{th} and $1 - P_{th}$ by $z_{ub} = \Phi^{-1}(P_{th})$ and $z_{lb} = \Phi^{-1}(1 - P_{th})$ respectively. In this way, expression (4) is equivalent to

$$z > \Phi^{-1}(P_{th}) \text{ or } z < \Phi^{-1}(1 - P_{th}) \tag{5}$$

For a feature data $z_i^{(t)}$, if it is judged as unreliable, in this paper we simply calibrate it as follows $z_i^{(t)} = z_i^{ub}$ if $z_i^{(t)} > \Phi^{-1}(P_{th})$, and $z_i^{(t)} = z_i^{lb}$ if $z_i^{(t)} < \Phi^{-1}(1 - P_{th})$. Note that we do not delete the unreliable because the remained features are still reliable, and a classifier that is strong enough may still can classify the target with calibrated data. In the following multiple frame decision fusion process, the soft decision of a calibrated motion data is still useful for recognizing the target person.

Figure 3 shows an example of the detected unreliable features, in which (a) and (b) plot the tracker distance features and gait speed respectively, (c) and (d) plot the PDFs of (a) and (b), respectively. When the probability threshold $P_{th} = 99.9\%$, we will find 1 and 5 outliers in (a) and (b), respectively, and the outlier points are marked with red circles. In this situation, the corresponding features will be to calibrated as their mean values to relief their impacts on the classification performance.
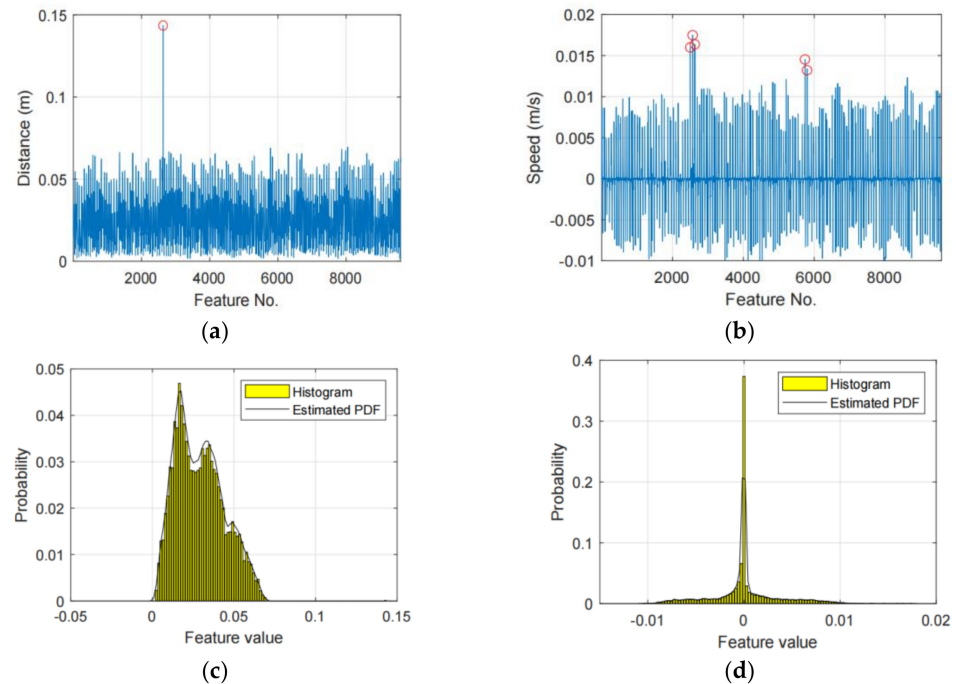

Figure 3. Two examples of unreliable feature, in which (**a,b**) plot the tracker distance features and gait speed, respectively, the unreliable features are marked with red circles, (**c,d**) plot the distribution of (**a,b**), respectively.

3.4. Classification with Single Gait Motion Frame

With the obtained features, we then use a pattern recognition classifier to identify the target person. The used classifier can be any proper classifier with acceptable classification performance, such as the commonly used support vector machine (SVM), and random forest. In the proposed decision fusion based gait recognition method, the classification performance of each single gait motion frame has critical influences on the following fusion accuracy, thus it is necessary to find a classifier with powerful classification capacities. However, from the test results, we found that the classification performances of SVM and random forest are not satisfactory to us. In this paper, we use KELM as the base classifier for its outstanding classification performance in both accuracy and efficiency. Let **Z** be the

features extracted from training motion data, then we can obtain the corresponding kernel Gram matrix $\mathbf{K} = \left[K_{i,j}\right]_{M \times M}$ by using the kernel function, in which $K_{i,j}$ is computed by

$$K_{i,j} = \mathcal{K}\left(\mathbf{z}_i, \mathbf{z}_j\right) \tag{6}$$

where $\mathcal{K}\left(\mathbf{z}_i, \mathbf{z}_j\right)$ denotes the kernel function with inputs $\mathbf{z}_i$ and $\mathbf{z}_j$. In this paper we use Gaussian kernel function, and $K_{i,j}$ is computed by

$$K_{i,j} = e^{\frac{(\mathbf{z}_i - \mathbf{z}_j)^2}{h}} \tag{7}$$

where $h > 0$ denotes the bandwidth function, usually it can be selected from $\left\{2^{-9}, \ldots, 2^0, \ldots, 2^9\right\}$. In the training process of KELM classifier, our goal is to obtain the output weight matrix $\boldsymbol{\beta}$, which is computed by [33]

$$\boldsymbol{\beta} = \left(\mathbf{K} + \frac{1}{\lambda}\right)^{-1} \mathbf{C} \tag{8}$$

where λ denotes the regularization parameter, and it can be selected from $\left\{10^{-9}, \ldots, 10^0, \ldots, 10^9\right\}$. Note that parameters h and λ can be set by different trials, and the values can be chosen as the one with maximal classification performance. Now, given a new input feature data $\mathbf{z}^{(t)}$, we can obtain the corresponding output of the KELM classifier as follows:

$$\boldsymbol{o}^{(t)} = \mathcal{K}\left(\mathbf{z}^{(t)}, \mathbf{Z}\right)^T \boldsymbol{\beta} = \mathcal{K}\left(\mathbf{z}^{(t)}, \mathbf{Z}\right)^T \left(\mathbf{K} + \frac{1}{\lambda}\right)^{-1} \mathbf{C} \tag{9}$$

Note that $\boldsymbol{o}^{(t)}$ is a vector that contains continuous predictions of target data, and we need an additional decision making process to if we want to know the final discrete predicted labels, namely the hard decisions.

Remark 1. *A relative larger value of a KELM output means the probability that the target belongs to the corresponding class will be higher. Apparently, a class with maximal output will be regarded the hard decision of the classifier. For a KELM classifier, if an output $o_i^{(t)}$ is closer to -1, then it is more probable that the target does not belong to the corresponding class c_i. On the other hand, if $o_i^{(t)}$ is closer to 1, then it will be more probable that it belongs to the class c_i.*

Since we want to combine the decisions of the multiple motion frames, the output will be transformed to fuzzy decisions, and details will be introduced in the following subsection.

3.5. Decision Fusion of Multiple Motion Frames

To combine the decisions of the consecutive motion frames, in this paper we propose a reliability-weighted sum rule (RWS) that adjusting the fuzzy decisions by considering the differences among the fuzzy decisions. In general, a decision is relatively more consistent to other decisions, it is more reliable, otherwise more unreliable. In RMS, the obtained output data of KELM classifier are first transformed to fuzzy decisions by using a fuzzy membership function. More specifically, for frame t, the fuzzy membership that the target person belongs to class c_i is defined as follows:

$$\mu_i^{(t)} = \exp\left[-\left(\frac{o^{(t)}}{\bar{o} + \gamma \sigma_o^{(t)}}\right)^2\right] \tag{10}$$

where $\bar{o}$ and $\sigma_o^{(t)}$ denote the average value and the standard deviation of the $o^{(t)}$. Parameter γ is used for adjusting the discriminative degree of the obtained membership values. A larger value of γ will produce a larger span of the membership, and the discriminative degree is also higher. In this paper, we set $\gamma = 0.5$ in default.

Remark 2. *The above fuzzy decision means that, for one classifier output vector, a relative larger output of one class will be transformed to a larger membership compared with other classes, otherwise it will be relatively smaller. In particular, in KELM, we set the class label that the person belongs to as +1, and other class labels as −1. For example, in a classification task with 5 classes, the decision label vector when the person belongs to class 2 is $c = [-1, +1, -1, -1, -1]^T$. Given a new gait instance that belongs to class 2 and suppose its output vector is $o = [-0.9, \, 0.2, -0.4, -0.8, -0.95]^T$, then the fuzzy decision computed by Equation (11) will be $\mu = [0.0005, 0.6913, 0.2284, 0.0027, 0.0002]$, we can see that a relative larger output value will produce a larger value of fuzzy membership.*

With the above membership transformation process, we can obtain the fuzzy decisions of all the motion frames, which are denoted as $\mu^{(1)}, \ldots, \mu^{(T)}$. One can directly combine the fuzzy decisions by using some classical decision fusion rules, such as the sum rule, product rule, and majority voting rule [34]. However, the reliabilities of the decisions are not considered in the above rules, which may decrease the accuracies of the final fusion results. If we can know a reasonable reliability for each fuzzy decision, then the impacts of the misclassified decisions can be reduced, and the classification accuracy of the global fusion results can be improved.

As such, in this paper we propose a reliability estimation method by using the consistency degrees among the fuzzy decisions. In belief function theory, the consistency degree between two basic belief assignments (BBAs) $m_1 = \{m_1(\omega_1), \ldots, m_1(\omega_M)\}$ and $m_2 = \{m_2(\omega_1), \ldots, m_2(\omega_M)\}$ is defined as follows [34]

$$\varphi(m_1, m_2) = \sum_{\omega_i \cap \omega_j \neq \varnothing} m_1(\omega_i) m_2(\omega_j). \tag{11}$$

Following the above definition, we define the consistency degree between two fuzzy decisions $\mu^{(1)} = \{\mu_1^{(1)}, \ldots, \mu_M^{(1)}\}$ and $\mu^{(2)} = \{\mu_1^{(2)}, \ldots, \mu_M^{(2)}\}$ as follows:

$$\psi(\mu^{(1)}, \mu^{(2)}) = \sum_{\omega_i \cap \omega_j \neq \varnothing} \mu_i^{(1)} \mu_j^{(2)}. \tag{12}$$

Since we do not have compound classes in the above Equation, thus the consistency degree equals to the inner product of $\mu^{(1)}$ and $\mu^{(2)}$, as given by

$$a_{1,2} = \psi(\mu^{(1)}, \mu^{(2)}) = \sum_{i=1}^{M} \mu_i^{(1)} \mu_i^{(2)} = <\mu^{(1)}, \mu^{(2)}> \tag{13}$$

If the obtained consistency value $\psi(\mu^{(1)}, \mu^{(2)})$ is relative larger, then we can know that $\mu^{(1)}$ and $\mu^{(2)}$ are more similar with each other, otherwise they are more conflicting with each other.

Remark 3. *For a complex classification task with multiple classes, there may exist several outputs with relative larger fuzzy membership values. For example, for a fuzzy decision $\mu^{(1)} = [0.1, 0.1, \, 0.7, \, 0.85, \, 0.1]^T$, in which the memberships of class c_3 and c_4 are much larger than the other 3 classes, and both of them are probable to be results. Given another output vector $\mu^{(2)} = [0.1, 0.1, \, 0.2, \, 0.85, \, 0.1]^T$, in which the membership of class c_4 is much larger than other classes. If the target belongs to class c_3, we can see that, when the reliability weights of the two fuzzy decisions are the same, fuzzy decision $\mu^{(2)}$ will impose a higher negative impact to the final decisions compared with $\mu^{(1)}$. As such, it is necessary to allocate a relative smaller reliability weight to $\mu^{(2)}$ to avoid misclassification risks.*

Remark 4. *According to the above definition of decision consistency degree, we can see that, for two fuzzy decisions $\mu^{(1)}$ and $\mu^{(2)}$, if $\mu^{(1)} \succcurlyeq \mu^{(2)}$, i.e., the memberships of $\mu^{(1)}$ are all larger than the corresponding memberships of $\mu^{(2)}$, we have $\psi(\mu^{(1)}, \mu^{(t)}) \geq \psi(\mu^{(2)}, \mu^{(t)})$ for any*

$\boldsymbol{\mu}^{(t)} \succcurlyeq 0$. *According to this property, we can see that, if one fuzzy decision has more than one relative larger memberships, its corresponding consistency degree will be more probable larger than a fuzzy decision with only one relative larger membership. For example, given two fuzzy decisions* $\boldsymbol{\mu}^{(1)} = [0.1, 0.7, 0.3, 0.2, 0.2]^T$ *and* $\boldsymbol{\mu}^{(2)} = [0.1, 0.7, 0.8, 0.2, 0.2]^T$, *then for* $\boldsymbol{\mu}^{(t)} \succcurlyeq 0$, *we have* $\psi\left(\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(t)}\right) \geq \psi\left(\boldsymbol{\mu}^{(2)}, \boldsymbol{\mu}^{(t)}\right)$. *This is because* $\boldsymbol{\mu}^{(2)}$ *indicates that the classifier indicates that both class 2 and 3 are very probable to be the target class, and its consistency degree will be larger.*

With above consistency computation method, we can obtain a consistency matrix $\mathbf{A} = [a_{t,k}]_{T \times T}$ that contains pairwise consistency values of among the fuzzy decisions, in which $a_{t,k} = \psi\left(\boldsymbol{\mu}^{(t)}, \boldsymbol{\mu}^{(k)}\right)$. It can be expected that, for a fuzzy decision $\boldsymbol{\mu}^{(t)}$, if its consistency degrees $\mathbf{a}_t = [a_{t,1}, \ldots, a_{t,T}]$ to other fuzzy decisions are relative lager values compared with other fuzzy decisions, then we can see it is more consistent to other fuzzy decisions, and its reliability degree should be higher. To achieve this goal, we use the eigenvalue decomposition method (EDM) [35] to compute the reliability weight of each fuzzy decision. In EDM, we want to compute the eigenvalues $\lambda_1, \ldots, \lambda_T$ and eigenvectors $\mathbf{w}_t, \ldots \mathbf{w}_T$ of $\mathbf{A}$, which satisfies the following condition:

$$\mathbf{A}\mathbf{w}_t = \lambda_t \mathbf{w}_t \tag{14}$$

We can see that one eigenvalue corresponds to a unique eigenvector. The above EDM problem can be properly solved by using the well-known Singular Value Decomposition (SVD) method [35]. When all the eigenvalues and eigenvectors are obtained, we use the eigenvector $\mathbf{w}^s$ that with the maximal eigenvalue $\lambda_{max} = \max\{\lambda_1, \ldots, \lambda_T\}$ as the decision weight vector. Note that the obtained eigenvector $\mathbf{w}^s$ can not be directly used as reliability weight if it not normalized. Let $r = \left[r^{(t)}\right]_{1 \times T}$ be the normalized weight vector, and it is computed by

$$r^{(t)} = a + b \frac{w_t^s - w_{min}^s}{w_{max}^s - w_{min}^s} \tag{15}$$

where a and b denote the lower bound and upper bound of the normalized weight respectively, w_{max}^s and w_{min}^s denote the maximal and minimal value of $\mathbf{w}^s$, respectively. In this paper, we set $a = 0.6$ and $b = 0.4$; thus, we have $r^{(t)} \in [0.6, 1]$.

Remark 5. *It has shown that eigenvector* $\mathbf{w}^s$ *can be used as the representation of the importance of each vector in consistency matrix* $\mathbf{A}$ *[36,37]. More specifically, a relative larger value of* $\mathbf{a}_t$ *will produce a relative larger eigenvalue of* w_t^s. *With this property, eigenvector* $\mathbf{w}^s$ *can be used as the reliability degree of the fuzzy decisions. As mentioned above, a fuzzy decision with larger average consistency value corresponds to a relative larger eigenvalue, and it is more reliable compared with other fuzzy decisions.*

Next, we can combine all fuzzy decisions into a unified global one by using the obtained reliability weights, as given by

$$\boldsymbol{\mu}^g = \frac{1}{T} \sum_{t=1}^{T} r^{(t)} \boldsymbol{\mu}^{(t)} \tag{16}$$

At last, the final decision is made by choosing the class with maximal global membership value, as given by

$$c^g = \max\left\{\mu_1^g, \ldots, \mu_N^g\right\} \tag{17}$$

The above fusion process is suitable for classifying the feature data with different number of gait motion frames. It can be expected that, the fusion accuracy will be increased if the number of the fused decisions (or motion frame number T) is increased. In general, only several gait cycles (e.g., $T \geq 3$) will be sufficient to achieve robust fusion accuracy.

The detailed process of the proposed RWS rule is illustrated in the Algorithm 1. We first train the KELM classifier by using Equation (8). Note that, in the proposed method, only KELM is required to be trained, and RWS rule does not need to be trained and it can be directly used for combining the fuzzy decisions. Given a data with several new motion frames, we obtain the outputs of the KELM classifier, and transform the outputs into fuzzy decisions $\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(T)}$. Then, we compute the consistency matrix $\mathbf{A}$ by using the obtained fuzzy decisions. Subsequently, the eigenvalue decomposition process is conducted to matrix $\mathbf{A}$, and the obtained eigenvector $\mathbf{w}^s$ with the maximal eigenvalue is used for representing the reliability values of the fuzzy decisions. Subsequently, we normalize the eigenvector $\mathbf{w}^s$ into a suitable interval and obtain the reliability vector $\boldsymbol{r} = \left[r^{(t)} \right]_{1 \times T}$, and combine the fuzzy decisions by using a weighted sum combination operation. Finally, the classification result of all fuzzy decisions is chosen as the class with maximal global fuzzy membership.

Algorithm 1 The proposed RWS decision fusion rule.

Input: Motion frame data $\left(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}, \mathbf{z}^{(1)} \right), \ldots; \left(\mathbf{x}^{(T)}, \mathbf{y}^{(T)}, \mathbf{z}^{(T)} \right)$, RBF kernel parameter h, KELM regularization parameter λ;
Output: Classification result c^g;
1:    **for** $t = 1, \ldots, T$
2:        Compute classification output $\boldsymbol{o}^{(t)}$ by using Equation (10);
3:        Compute fuzzy decisions $\boldsymbol{\mu}^{(t)}$ by using Equation (11);
4:    **end for**
5:    Compute fuzzy decision consistency matrix $\mathbf{A} = \left[a_{t,k} \right]_{T \times T}$ by using Equation (14);
6:    Compute eigenvalues $\lambda_1, \ldots, \lambda_T$ and eigenvectors $\mathbf{w}_t, \ldots \mathbf{w}_T$ of consistency matrix $\mathbf{A} = \left[a_{t,k} \right]_{T \times T}$ by using eigenvalue decomposition;
7:    Find the eigenvector $\mathbf{w}^s$ with the maximal eigenvalue $\lambda_{max} = \max\{\lambda_1, \ldots, \lambda_T\}$;
8:    Compute decision reliabilities $\boldsymbol{r} = \left[r^{(t)} \right]_{1 \times T}$ by using Equation (16);
9:    Compute global fuzzy decision $\boldsymbol{\mu}^g$ by using Equation (17);
10: Obtain the final classification result c^g by using Equation (18).

3.6. A Toy Example for Illustrating the Proposed RWS Rule

In this subsection we present a toy example to give a better understanding of the proposed RWS rule. Consider a gait motion recognition problem with 5 possible persons, and suppose that we have 10 consecutive motion frames, and the corresponding fuzzy decisions are shown in Table 1. In this example, the memberships of class 2 are randomly generated from interval (0.1, 0.8), and the memberships of other classes are randomly generated from interval (0.1, 0.4). We can see that, except $\boldsymbol{\mu}^{(2)}$, the membership values of class 2 in other fuzzy decisions are actually not very large. In particular, we can see that in fuzzy decisions $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(3)}$ and $\boldsymbol{\mu}^{(6)}$, the classes with largest membership values are not class 2. It can be expected that their reliability degrees will be relative smaller than other fuzzy decisions.

Next, we compute the corresponding decision consistency matrix of the fuzzy decisions in Table 1, and the results are shown in Table 2, in which $\mathbf{A}_i$ and $\mathbf{A}^{(i)}$ denote the i-th column and i-th row, respectively. The corresponding fuzzy decisions are shown in Table 3. As expected, we can see the reliabilities of fuzzy decisions $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(3)}$ and $\boldsymbol{\mu}^{(6)}$ are the three smallest of the 10 fuzzy decisions, and reliability of $\boldsymbol{\mu}^{(2)}$ is the largest. In particular, we can see that the reliabilities of $\boldsymbol{\mu}^{(9)}$ and $\boldsymbol{\mu}^{(10)}$, respectively. From this example, we can see that the obtained reliability weight of one fuzzy decision can reasonably reflect its overall consistency to other fuzzy decisions. Finally, with the obtained reliability weights, the global fuzzy decision can be obtained by Equation (17), which is $\boldsymbol{\mu}^g = [0.11, 0.27, 0.16, 0.14, 0.14]$, which shows that the final decision is class 2.

Table 1. The fuzzy decisions.

	Class 1	Class 2	Class 3	Class 4	Class 5
$\mu^{(1)}$	0.23	0.24	0.34	0.13	0.40
$\mu^{(2)}$	0.32	0.71	0.39	0.23	0.32
$\mu^{(3)}$	0.10	0.29	0.19	0.39	0.18
$\mu^{(4)}$	0.19	0.52	0.31	0.26	0.34
$\mu^{(5)}$	0.14	0.40	0.36	0.31	0.13
$\mu^{(6)}$	0.13	0.33	0.37	0.19	0.23
$\mu^{(7)}$	0.16	0.57	0.13	0.31	0.37
$\mu^{(8)}$	0.20	0.36	0.11	0.35	0.19
$\mu^{(9)}$	0.22	0.50	0.15	0.11	0.19
$\mu^{(10)}$	0.26	0.58	0.36	0.33	0.14

Table 2. The consistency matrix of the fuzzy decisions.

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
$A^{(1)}$	0.40	0.53	0.28	0.44	0.34	0.35	0.40	0.29	0.31	0.42
$A^{(2)}$	0.53	0.92	0.46	0.72	0.59	0.54	0.69	0.50	0.57	0.76
$A^{(3)}$	0.28	0.46	0.31	0.39	0.34	0.30	0.39	0.31	0.27	0.42
$A^{(4)}$	0.44	0.72	0.39	0.59	0.47	0.44	0.57	0.41	0.44	0.60
$A^{(5)}$	0.34	0.59	0.34	0.47	0.43	0.38	0.44	0.35	0.34	0.52
$A^{(6)}$	0.35	0.54	0.30	0.44	0.38	0.36	0.40	0.30	0.31	0.46
$A^{(7)}$	0.40	0.69	0.39	0.57	0.44	0.40	0.60	0.43	0.44	0.57
$A^{(8)}$	0.29	0.50	0.31	0.41	0.35	0.30	0.43	0.34	0.31	0.44
$A^{(9)}$	0.31	0.57	0.27	0.44	0.34	0.31	0.44	0.31	0.36	0.46
$A^{(10)}$	0.42	0.76	0.42	0.60	0.52	0.46	0.57	0.44	0.46	0.66

Table 3. The reliabilities of the fuzzy decisions.

	$\mu^{(1)}$	$\mu^{(2)}$	$\mu^{(3)}$	$\mu^{(4)}$	$\mu^{(5)}$	$\mu^{(6)}$	$\mu^{(7)}$	$\mu^{(8)}$	$\mu^{(9)}$	$\mu^{(10)}$
Reliability (Unnormalized)	0.26	0.44	0.24	0.36	0.29	0.27	0.35	0.26	0.27	0.37
Reliability	0.37	1.00	0.30	0.70	0.48	0.39	0.66	0.35	0.39	0.75

4. Experimental Results

In this section, we test the performance of the proposed method by comparing it with some acknowledged baseline methods. The experiment mainly includes two parts: the first one is the classification performance comparison of the single motion frame, and the results of KELM, SVM, and random forest will be provided. The other one is the performance comparison results of the proposed RWS rule and other fusion rules, including sum rule, belief rule, weighted belief rule, product rule, and majority voting rule. As mentioned before, the system has 10 motion trackers, which are implemented on the lower body locations, and each body location has one or two trackers. For one body side, there are 5 trackers, and the trackers are implemented at the following three locations: thigh (one tracker), lower leg (one tracker at ankle, one tracker at shank), foot (one tracker at ankle, and one tracker at tiptoe). Since the system requires us to plant optical trackers on human

bodies, it is better to reduce the number of trackers as much as possible. As such, we also want to clarify that whether it is possible for us to implement only one tracker on the lower leg and foot locations, thus we choose three trackers each body side, and six in total. With the above considerations, we will test the recognition performances of single motion frames and multiple motions with 10 and 6 motion trackers. All the experiments are conducted in a Windows 10 system with Intel i7 CPU, and 16 GB RAM, and the algorithm is developed in MATLAB 2020a platform.

4.1. Results of Single Motion Frame Classification

In this subsection, we test the classification performance of KELM on single motion frames. We first use 9595 motion frames as the training dataset, and the number of each of consecutive motion frames of each participant ranges from 82 to 179, and average number of each participant is about 122, and the corresponding number of gait cycles are about 3–5. We use another 4787 motion frames for the 76 participants, and the number of consecutive motion frames of each participant ranges from 45 to 89, or 1–3 gait cycles. We can see that test data size is about the 1/2 of the training data.

Except KELM, we also test the performance of support vector machines (SVM) and random forest, which are used in [26] and [27], respectively. The classification performance comparison results along with the parameter settings that can achieve the maximal classification accuracies are shown in Tables 4 and 5. In Table 4, all the data collected from 10 motion trackers are used, while in Table 5, only 6 selected trackers are used, and they are located at front thigh, front knee, and ankle, both left foot and right foot. Since the collected data from 6 selected motion trackers are just a part of the data from all the 10 trackers, we can expect that its classification accuracy will be lower compared with the result of 10 motion trackers. For the random forest classifier, the number of template classification trees is set as 400, and the maximal tree split is set as 1500 or 2000. For the SVM classifier, the used kernel is RBF kernel, and the bandwidth parameter is $h = 2^{-1}$ and $h = 2^{-2}$ in Tables 1 and 2, respectively. In KELM, the used kernel is also RBF kernel, the kernel bandwidth parameter and regularization parameter in Table 1 are set as $h = 2^3$ and $\lambda = 10^4$, respectively, and in Table 2 are $h = 2^1$ and $\lambda = 10^4$, respectively.

Table 4. Classification accuracy with all 10 motion trackers.

Classifier	Accuracy (%)	Train Time(s)	Parameter Settings
Random Forest [27]	50.44	74.62	Number of Trees = 400, Maxsplit = 2000
SVM [26]	56.17	27.88	RBF Kernel, $h = 2^{-3}$
KELM	83.55	2.46	RBF Kernel, $h = 2^4$, $\lambda = 10^5$

Table 5. Classification accuracy with 6 selected motion trackers.

Classifier	Accuracy (%)	Train Time(s)	Parameter Settings
Random Forest [27]	36.42	45.77	Number of Trees = 400, Maxsplit = 1500
SVM [26]	43.80	18.19	RBF Kernel, $h = 2^{-2}$
KELM	68.17	2.61	RBF Kernel, $h = 2^1$, $\lambda = 10^7$

From Tables 4 and 5, we can observe that, KELM achieves much better performances in both classification accuracy and training time compared with both SVM and random forest. More specifically, in Table 1, the classification of KELM achieves 27.38% and 33.11% higher accuracy compared with SVM and random forest, respectively. In Table 2, KELM achieves 24.37% and 31.75% classification accuracy improvement compared with SVM and random forest, respectively. In addition, SVM and random forest require more than 10 and

30 times of training time compared with KELM. The above results also indicate that, by adding more motion trackers, the classification accuracy of single motion frame will be increased. However, this practice will increase the deployment cost, and a relative lower classification accuracy for single motion frame will also have good fusion accuracy after combining multiple decisions of the frames.

4.2. Results of Multiple Decision Fusion

In this subsection, we demonstrate the classification performance of the proposed RWS rule by comparing it with several well-known decision fusion rules. The details of the compared rules as illustrated as follows:

(1) **Majority voting rule** [38]: Majority voting is a commonly used fusion rule for hard decisions, thus it is widely used in sensor fusion for its advantage in low data transmission amount. In this rule, we first obtain the hard decision of each motion frame, and then the final decision is made by choosing the class with maximal number of hard decisions. In this paper, if two class has the same voting number, the final decision is made by choosing the one with maximal average membership.

(2) **Sum rule** [38]: In this rule, the final fuzzy decision is simply computed by adding all the motion frame decisions, and the decision reliability is not considered. Given T motion frame fuzzy decisions $\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(T)}$, the global membership of class c_i is computed as

$$\mu_i = \frac{1}{T} \sum_{t=1}^{T} \mu_i^{(t)} \tag{18}$$

(3) **Product rule** [38]: similar to Naive Bayes fusion, in product rule the final fuzzy decision is obtained by the product of all the frame decisions. More specifically, the un-normalized global membership of class c_i is computed as

$$\mu_i = \prod_{t=1}^{T} (\mu_i^{(t)} + \delta) \tag{19}$$

where $\delta > 0$ is a very small constant to eliminate the influence of the memberships that are close to 0. In our experiment, we set $\delta = 10^{-3}$.

(4) **Belief rule** [34]: Depmster-Shafer evidence theory is widely used in dealing with multiple decision fusion problems. The belief fusion rule is derived by combining multiple basic belief assignments (BBAs) by using the Dempster's combinational rule, and it is also applicable in the multiple motion frame decision fusion problem. The un-normalized global BBA of class c_i can be computed by

$$\mu_i = \prod_{t=1}^{T} \frac{1}{(1 - \mu_i^{(t)} + \delta)} \tag{20}$$

(5) **Reliability-weighted belief rule** [34]: The obtained decision reliability can also be used in computing the global BBA from the decisions of multiple motion frames. Similar to Equation (18), the un-normalized global BBA on class c_i is given by

$$\mu_i = \prod_{t=1}^{T} \frac{1}{(1 - r^{(t)} \mu_i^{(t)} + \delta)} \tag{21}$$

Next, we test the classification accuracy of the fusion results of the proposed RWS rule and the baseline fusion rules with different KELM parameter settings. In addition, since the number of the test motion frames influences the classification, we also test the classification accuracies with increasing ratio of test motion frames, and ratio increases from 0.1 to 1. We test the fusion accuracies with the 10 trackers and 6 selected trackers, and the obtained results are shown in Figures 4 and 5, respectively.
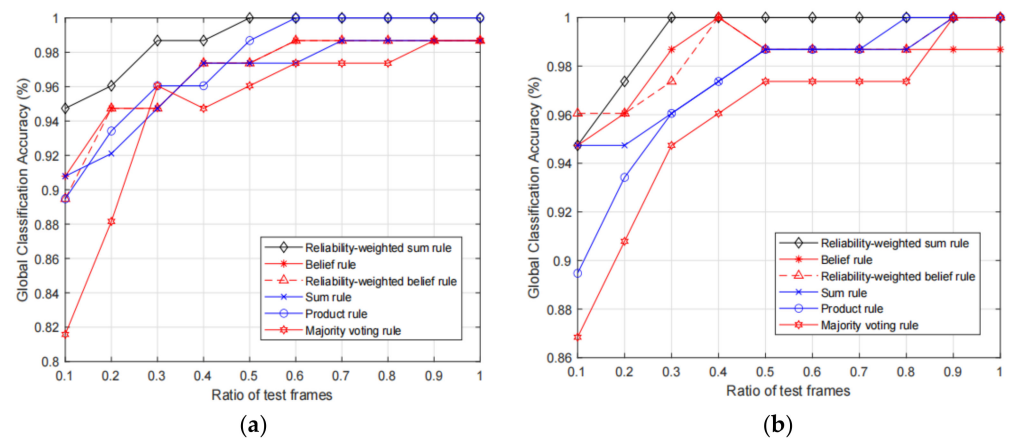
Figure 4. Comparison of fusion accuracies with 10 optical motion trackers, in which (**a**) and (**b**) show the results with original input features and calibrated input features, respectively.
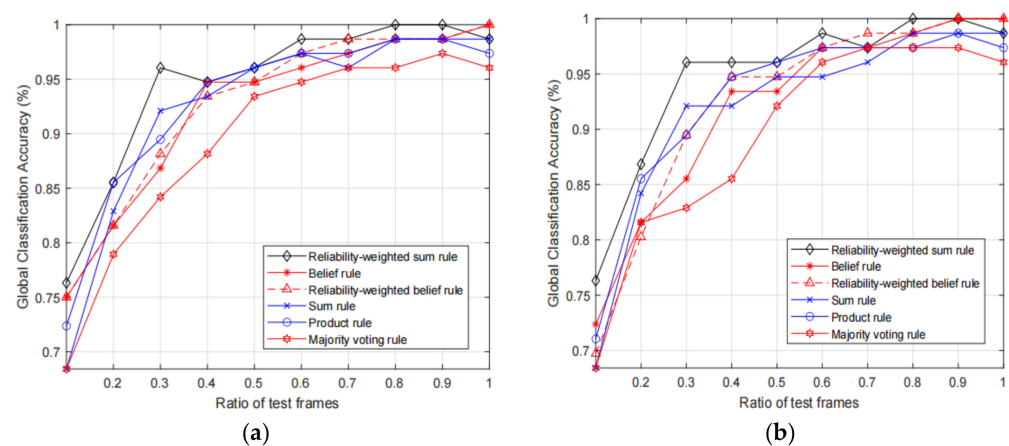

Figure 5. Comparison of fusion accuracies with 6 optical motion trackers, in which (**a**) and (**b**) show the results with original input features and calibrated input features, respectively.

From the two figures, we can observe that the proposed RWS rule can achieve higher fusion accuracy compared with other rules. More specifically, we can see that, given the same ratio of the test frames, in most times the proposed RWS can achieve higher accuracies. However, it does not mean that the proposed rule will always be better than other rules. For example, in Figure 5b, we can see that belief rule and reliability-weighted belief rule achieve 100% accuracy, while the proposed rule only achieves 98.68%. As expected, the fusion accuracies of 10 trackers are much higher than the accuracies of 6 selected trackers. Therefore, if the deployment cost is acceptable, we can add more trackers to increase the classification accuracy. By comparing the results of Figure 4a,b we can observe that results with calibrated feature data are slightly better than the results with original non-calibrated data. This is because the classification performance of single motion frame is influenced by the unreliable features, and the performance will be increased if the unreliable features are calibrated.

5. Conclusions

In this paper, we have studied the gait recognition problem by using optical motion capture data, and we proposed a first-classification-then-fusion method, which includes the following four steps: feature extraction, unreliable feature detection, classification of single motion frame, and decision fusion of multiple frames. In particular, we proposed an RWS decision fusion rule to combine the fuzzy decision of the gait motions. The experimental results on 76 participants show that KELM achieves much higher classification accuracy and training efficiency compared with SVM and random forest in the single motion frame

classification task, and they demonstrate that the proposed method RWS achieves better fusion accuracy compared with several existing fusion rules. Particularly, our results show that, with high-precision 3-D gait motion tracking data, the recognition method can achieve 100% accuracy when the full data of 10 optical trackers are used.

It has to admit that, although the proposed method can achieve 100% recognition accuracy, the relative high implementation cost on both optical trackers and the capture cameras limits its application scenarios. On the other hand, our results indicate that the performance of other gait recognition systems, such as video surveillance, can be further improved if range sensor and depth sensor is integrated to enable measuring the distances of the captured persons. It has to be recognized that, though the proposed method can achieve very high recognition accuracy, it requires the implementation of optical motion trackers, which limits its practical application scenarios.

Our future work is applying the proposed system in multiplayer motion tracking scenarios and test the performance with more complex trajectories. In this way, the system can support several players in the tracking field at the same time, which will greatly enhance its capability and efficiency.

Author Contributions: Conceptualization, L.W. and Y.L.; methodology, L.W., W.Z. and Y.L.; software, W.Z. and F.X.; validation, W.Z. and F.X.; formal analysis, L.W. and W.Z.; investigation, W.Z. and F.X.; resources, L.W. and Y.L.; data curation, L.W. and Y.L.; writing—original draft preparation, L.W. and W.Z.; writing—review and editing, Y.L.; visualization, Y.L.; supervision, Y.L. and F.X.; project administration, Y.L.; funding acquisition, F.X. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wan, C.; Wang, L.; Phoha, V.V. (Eds.) A survey on gait recognition. *ACM Comput. Surv.* **2018**, *51*, 1–35.
2. Malik, M.N.; Azam, M.A.; Ehatisham-Ul-Haq, M.; Ejaz, W.; Khalid, A. ADLAuth: Passive authentication based on activity of daily living using heterogeneous sensing in smart cities. *Sensors* **2019**, *19*, 2466. [CrossRef]
3. Khamsemanan, N.; Nattee, C.; Jianwattanapaisarn, N. Human identification from freestyle walks using posture-based gait feature. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 119–128. [CrossRef]
4. Wang, Y.; Du, B.; Shen, Y.; Wu, K.; Zhao, G.; Sun, J.; Wen, H. EV-gait: Event-based robust gait recognition using dynamic vision sensors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 6358–6367.
5. Sun, F.; Mao, C.; Fan, X.; Li, Y. Accelerometer-based speed-adaptive gait authentication method for wearable IoT devices. *IEEE Internet Things J.* **2018**, *6*, 820–830. [CrossRef]
6. Wang, W.; Liu, A.X.; Shahzad, M. Gait recognition using wifi signals. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 363–373.
7. Jiang, X.; Zhang, Y.; Yang, Q.; Deng, B.; Wang, H. Millimeter-Wave Array Radar-Based Human Gait Recognition Using Multi-Channel Three-Dimensional Convolutional Neural Network. *Sensors* **2020**, *20*, 5466. [CrossRef] [PubMed]
8. Zou, Q.; Ni, L.; Wang, Q.; Li, Q.; Wang, S. Robust gait recognition by integrating inertial and RGBD sensors. *IEEE Trans. Cybern.* **2017**, *48*, 1136–1150. [CrossRef] [PubMed]
9. Sutherland, D.H. The evolution of clinical gait analysis part III–kinetics and energy assessment. *Gait Posture* **2005**, *21*, 447–461. [CrossRef] [PubMed]
10. van der Kruk, E.; Reijne, M.M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *Eur. J. Sport Sci.* **2018**, *18*, 806–819. [CrossRef]
11. Holden, D. Robust solving of optical motion capture data by denoising. *ACM Trans. Graph.* **2018**, *37*, 1–12. [CrossRef]
12. Ali, A.; Sundaraj, K.; Ahmad, B.; Ahamed, N.; Islam, A. Gait disorder rehabilitation using vision and non-vision based sensors: A systematic review. *Bosn. J. Basic Med. Sci.* **2012**, *12*, 193. [CrossRef]

13. Zhang, J.; Li, P.; Zhu, T.; Zhang, W.A.; Liu, S. Human Motion Capture Based on Kinect and IMUs and Its Application to Human-Robot Collaboration. In Proceedings of the 2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM), Shenzhen, China, 18–21 December 2020; pp. 392–397.

14. Noiumkar, S.; Tirakoat, S. Use of optical motion capture in sports science: A case study of golf swing. In Proceedings of the 2013 International Conference on Informatics and Creative Multimedia, Kuala Lumpur, Malaysia, 4–6 September 2013; pp. 310–313.

15. Abtahi, M.; Borgheai, S.B.; Jafari, R.; Constant, N.; Diouf, R.; Shahriari, Y.; Mankodiya, K. Merging fNIRS-EEG brain monitoring and body motion capture to distinguish parkinsons disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 1246–1253. [CrossRef]

16. Iosifidis, A.; Tefas, A.; Pitas, I. On the kernel extreme learning machine classifier. *Pattern Recognit. Lett.* **2015**, *54*, 11–17. [CrossRef]

17. Yang, G.; Tan, W.; Jin, H.; Zhao, T.; Tu, L. Review wearable sensing system for gait recognition. *Clust. Comput.* **2019**, *22*, 3021–3029. [CrossRef]

18. Singh, J.P.; Jain, S.; Arora, S.; Singh, U.P. Vision-based gait recognition: A survey. *IEEE Access* **2018**, *6*, 70497–70527. [CrossRef]

19. BenAbdelkader, C.; Cutler, R.; Davis, L. Stride and cadence as a biometric in automatic person identification and verification. In Proceedings of the Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA, 21 May 2002; pp. 372–377.

20. Yang, K.; Dou, Y.; Lv, S.; Zhang, F.; Lv, Q. Relative distance features for gait recognition with Kinect. *J. Vis. Commun. Image Represent.* **2016**, *39*, 209–217. [CrossRef]

21. Balazia, M.; Sojka, P. Gait recognition from motion capture data. *ACM Trans. Multimed. Comput. Commun. Appl.* **2018**, *14*, 1–18. [CrossRef]

22. Cunado, D.; Nixon, M.S.; Carter, J.N. Automatic extraction and description of human gait models for recognition purposes. *Comput. Vis. Image Underst.* **2003**, *90*, 1–41. [CrossRef]

23. Tang, C.; Phoha, V.V. An empirical evaluation of activities and classifiers for user identification on smartphones. In Proceedings of the 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems, Niagara Falls, NY, USA, 6–9 September 2016; pp. 1–8.

24. Connor, P.; Ross, A. Biometric recognition by gait: A survey of modalities and features. *Comput. Vis. Image Underst.* **2018**, *167*, 1–27. [CrossRef]

25. Garofalo, G.; Argones Rúa, E.; Preuveneers, D.; Joosen, W. A systematic comparison of age and gender prediction on imu sensor-based gait traces. *Sensors* **2019**, *19*, 2945.

26. Wang, F.; Yan, L.; Xiao, J. Human gait recognition system based on support vector machine algorithm and using wearable sensors. *Sens. Mater.* **2019**, *31*, 1335. [CrossRef]

27. Guo, Q.; Jiang, D. Method for walking gait identification in a lower extremity exoskeleton based on C4.5 decision tree algorithm. *Int. J. Adv. Robot. Syst.* **2015**, *12*, 30. [CrossRef]

28. Alotaibi, M.; Mahmood, A. Improved gait recognition based on specialized deep convolutional neural network. *Comput. Vis. Image Underst.* **2017**, *164*, 103–110. [CrossRef]

29. Yu, S.; Jia, S.; Xu, C. Convolutional neural networks for hyperspectral image classification. *Neurocomputing* **2017**, *219*, 88–98. [CrossRef]

30. Wang, X.; Yan, W.Q. Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory. *Int. J. Neural Syst.* **2020**, *30*, 1950027. [CrossRef] [PubMed]

31. Zhang, W.; Zhang, Z.; Chao, H.C.; Tseng, F.H. Kernel mixture model for probability density estimation in Bayesian classifiers. *Data Min. Knowl. Discov.* **2018**, *32*, 675–707. [CrossRef]

32. Xu, X.; Yan, Z.; Xu, S. Estimating wind speed probability distribution by diffusion-based kernel density method. *Electr. Power Syst. Res.* **2015**, *121*, 28–37. [CrossRef]

33. Zhang, W.; Zhang, Z.; Wang, L.; Chao, H.C.; Zhou, Z. Extreme learning machines with expectation kernels. *Pattern Recognit.* **2019**, *96*, 106960. [CrossRef]

34. Zhang, W.; Zhang, Z. Belief function based decision fusion for decentralized target classification in wireless sensor networks. *Sensors* **2015**, *15*, 20524–20540. [CrossRef]

35. Herviou, L.; Bardarson, J.H.; Regnault, N. Defining a bulk-edge correspondence for non-Hermitian Hamiltonians via singular-value decomposition. *Phys. Rev. A* **2019**, *99*, 052118. [CrossRef]

36. Ho, W.; Ma, X. The state-of-the-art integrations and applications of the analytic hierarchy process. *Eur. J. Oper. Res.* **2018**, *267*, 399–414. [CrossRef]

37. Liu, Y.; Eckert, C.M.; Earl, C. A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Syst. Appl.* **2020**, 113738. [CrossRef]

38. Mangai, U.G.; Samanta, S.; Das, S.; Chowdhury, P.R. A survey of decision fusion and feature fusion strategies for pattern classification. *IETE Tech. Rev.* **2010**, *27*, 293–307. [CrossRef]