

Article

Intelligent Performance Evaluation in Rowing Sport Using a Graph-Matching Network

Chien-Chang Chen ¹, Cheng-Shian Lin ^{1,*}, Yen-Ting Chen ¹, Wen-Her Chen ², Chien-Hua Chen ^{2,3} and I-Cheng Chen ²

- ¹ Department of Computer Science and Information Engineering, Tamkang University, New Taipei City 25137, Taiwan; ccchen34@mail.tku.edu.tw (C.-C.C.); a80598059@gmail.com (Y.-T.C.)
² Office of Physical Education, Tamkang University, New Taipei City 25137, Taiwan; 137540@mail.tku.edu.tw (W.-H.C.); 132098@mail.tku.edu.tw (C.-H.C.); masa@mail.tku.edu.tw (I.-C.C.)
³ Department of Physical Education and Sport Sciences, National Taiwan Normal University, Taipei City 11718, Taiwan
 * Correspondence: 157446@o365.tku.edu.tw

Abstract: Rowing competitions require consistent rowing strokes among crew members to achieve optimal performance. However, existing motion analysis techniques often rely on wearable sensors, leading to challenges in sporter inconvenience. The aim of our work is to use a graph-matching network to analyze the similarity in rowers' rowing posture and further pair rowers to improve the performance of their rowing team. This study proposed a novel video-based performance analysis system to analyze paired rowers using a graph-matching network. The proposed system first detected human joint points, as acquired from the OpenPose system, and then the graph embedding model and graph-matching network model were applied to analyze similarities in rowing postures between paired rowers. When analyzing the postures of the paired rowers, the proposed system detected the same starting point of their rowing postures to achieve more accurate pairing results. Finally, variations in the similarities were displayed using the proposed time-period similarity processing. The experimental results show that the proposed time-period similarity processing of the 2D graph-embedding model (GEM) had the best pairing results.

Keywords: OpenPose; graph neural network



Citation: Chen, C.-C.; Lin, C.-S.; Chen, Y.-T.; Chen, W.-H.; Chen, C.-H.; Chen, I.-C. Intelligent Performance Evaluation in Rowing Sport Using a Graph-Matching Network. *J. Imaging* **2023**, *9*, 181. <https://doi.org/10.3390/jimaging9090181>

Academic Editor: Hocine Cherifi

Received: 19 July 2023

Revised: 21 August 2023

Accepted: 25 August 2023

Published: 31 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, many studies have been conducted that analyzed the performance of individual players in sports domains [1–8]. While there are many studies that focus on motion analysis, most require athletes to wear sensors to acquire their posture data; however, this method easily introduces noise due to friction during motions [1,2,8]. Therefore, the video-based contactless approach, which aims to acquire whole posture data using a simple camera without placing sensors on body parts, has made it easier and more convenient for scholars to analyze a variety of postures in many exciting sports domains [3,4,6,7].

Video-based motion analysis has witnessed numerous applications, including human-computer interaction systems [3], human action understanding systems [4], medical assistance systems [9], and human pose estimation [10], all based on deep learning and recognition models. The OpenPose system [10] has proven to be valuable in capturing the skeletal joints of individuals, even in complex postures, using a simple camera without the need for specialized hardware like the Kinect device [3,9]. However, as most previous applications are relevant only for recognizing simple postures, the issue of how to analyze the complex postures of sports players has been neither addressed nor effectively analyzed.

Several previous works [10–22] have utilized the OpenPose model [10] for various applications. Qiao et al. [11] used a series of coordinate trajectories of joint points to draw a curve that was used to determine whether Tai Chi movements were standard. Tsai et al. [12]

estimated the depth distance between a person and a lens in a single image using the OpenPose model to capture the coordinates of the human body keypoints. Nakai et al. [13] proposed a prediction method for a basketball shooting system, which detected human body keypoints using the OpenPose model and predicted basketball free throw shooting using a logistic regression model. In addition, several applications based on deep learning have also been proposed. Toshev et al. [14] proposed a state-of-the-art approach based on deep neural networks (DNNs) for human pose estimation. Xiao et al. [15] proposed a simple baseline method based on ResNet architectures [16] as the backbone network for human pose estimation and tracking. Zhang et al. [17] proposed a golf analysis system that includes a human detection subsystem and a performance analysis subsystem. In the human detection subsystem, they applied the OpenPose model, human tracking, and an LSTM deep learning model to detect golf players' postures, while the performance analysis subsystem scored the comparison results of the golf players' postures. Theagarajan et al. [18] proposed an automatically generated visual analytics and player statistics system for soccer players based on convolutional neural networks (CNNs) and deep convolutional generative adversarial networks (DCGANs). These systems can successfully perform relevant applications using system models, such as predicting basketball shots, assessing the standardness of sports movements, etc. However, they still experience a decrease in accuracy when used in complex environments. For instance, object appearance variations and lighting changes, factors present in a complex environment, make it challenging for the systems to accurately capture and analyze information from images [19–21]. Moreover, the issue of pairing athletes to work together to enhance the performance of a sports team has not been addressed.

This study proposed a video-based performance analysis system for analyzing the performance of rowing pairs using a graph-matching network, which first detected and acquired human joint points using the OpenPose system from a rowing video. Then, the detected human joints were represented as graph structures that extracted the features of the rowing posture process. The rowing posture feature of each video frame was extracted using the graph-embedding model (GEM) and graph-matching network (GMN) model. Then, a video was developed to compare the baseline processes for two video sequences, and the performances of the rowing pairs were measured by calculating the rowing posture similarities in the pair using the GEM and GMN models. Finally, the proposed time-period similarity processing method was used to distinguish the degree of similarity changes in the players in the video segments. Experiments were carried out using a dataset of over 15 test rowing players, and the results show that the proposed approach can effectively evaluate the performances of rowing pairs and provide good suggestions for coaches when creating player groupings and training programs.

In the sport of rowing, consistency in the rowing movements among team members has a significant impact. Well-coordinated rowing motions contribute to achieving outstanding performance. Therefore, similarity in rowing postures among athletes has great significance. The purpose of this study is to analyze paired rowers using a graph-matching network to enhance the performance of a team sport. This study was organized into the following sections. Section 2 gives a brief review of related works, including the OpenPose system [10] and graph neural networks (GNNs) [18–22]. Section 3 provides the details of the proposed approach. Section 4 presents the experimental results, and Section 5 offers concluding remarks.

2. Related Works

This section briefly reviews some works related to this study, the OpenPose system [10] and graph neural networks (GNNs) [22–26], which are described in the following two subsections, respectively.

2.1. The OpenPose System

The OpenPose system, developed at Carnegie Mellon University, is a widely used, deep learning-based, real-time pose estimation system [10]. This system is capable of detecting multi-person poses in real-time by leveraging the COCO, MPII, and body25 models, with the number of human joints being determined based on the specific model used. Among these models, the OpenPose system with the body25 model yields more accurate keypoint results compared with the COCO and MPII models. Figure 1 illustrates detection results obtained using the OpenPose system with the body25 model. Each detected keypoint in the resulting image contains three pieces of information: the X coordinate, Y coordinate, and C confidence.

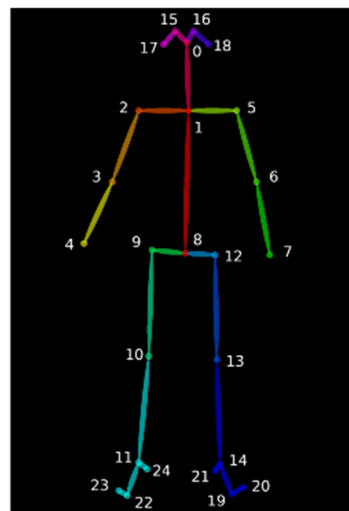


Figure 1. Keypoints detected using the OpenPose system and body25 model. Adapted from [10].

The OpenPose system is primarily built upon the human posture evaluation algorithm, with a core component known as part affinity fields (PAFs). This method uses a bottom-up detection approach, where it initially identifies joint point positions on the human body and subsequently extends to form the complete skeleton. As a result, the number of people being detected has minimal impact on the computational time required by OpenPose.

2.2. Graph Neural Networks

Deep learning has been extensively applied in image and sound recognition tasks, predominantly in Euclidean space recently. However, existing analysis methods like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) encounter challenges when dealing with non-Euclidean spaces, such as matching human joint points. Consequently, researchers have introduced deep learning models based on graph neural networks (GNNs) [22–26] to tackle non-Euclidean spaces.

Although Sperduti and Starita initially proposed using neural networks for graph analysis in 1997 [20], it was not until 2005 that Gori et al. [25] presented a complete GNN architecture. Li et al. [26] further introduced the graph-embedding model (GEM) and graph-matching network (GMN), based on the GNN framework for graph-pairing tasks. GEM primarily embeds each graph into a low-dimensional vector using message transfer among adjacent nodes, allowing for distance calculation and similarity measurement between graphs. The graph-matching network (GMN), built upon the three-layer architecture of GEM, enhances accuracy in graph pairing by introducing a cross-graph message exchange mechanism alongside the node message propagation within the graph. Additionally, GNN has been extensively applied in various practical applications and research areas, including human movement recognition [27,28], identity analysis using motion [29], football prediction analysis [18,30], and similarity-based pairing tasks [31,32].

This study aimed to develop a video-based approach to analyze the performance of rowing poses of rowing pairs. To enable effective automatic comparison of rowing poses, the OpenPose system was used to extract robust rowing pose features and convert them into a graph structure that is robust to external factors, such as appearance (e.g., size, shape, and color) and lighting conditions. Subsequently, the GEM and GMN models were utilized to analyze the similarities in rowing postures between each pair of rowers.

3. The Proposed Approach

This section presents details of the proposed approach for analyzing the performance of rowing pairs. As shown in Figure 2, the proposed approach consists of three main steps: feature extraction, baseline comparison analysis, and rowing performance measurement. The details of the proposed approach are described as follows.

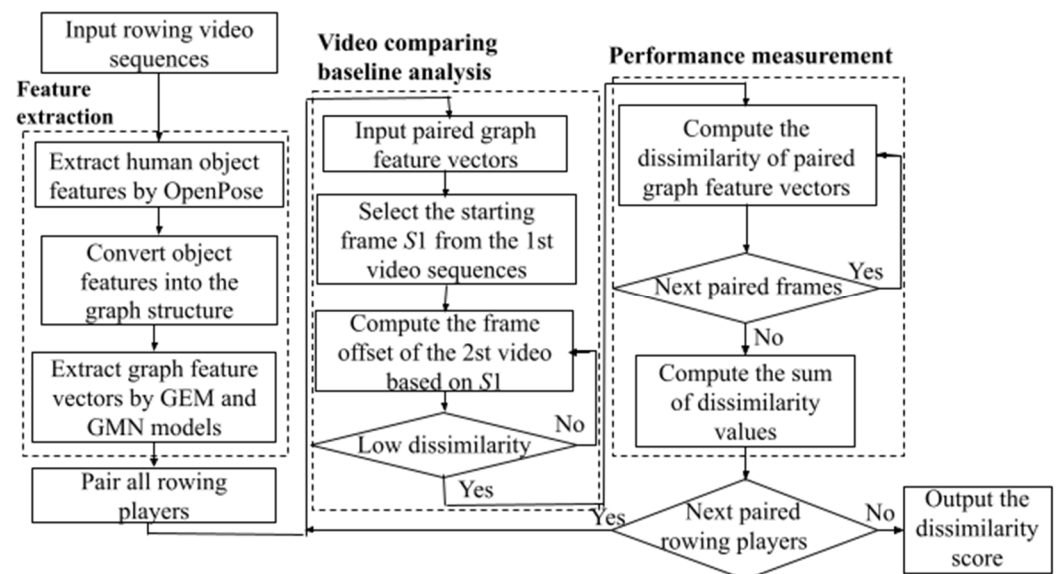


Figure 2. The proposed analysis approach for rowing player pairs.

3.1. Feature Extraction

This study used a video-based performance analysis system to analyze the performance of paired rowers, that is, the human body joints of the rowing pose features were detected and extracted using the OpenPose system. Furthermore, analyses were conducted to identify similarities in the rowing postures of paired rowers.

Let a video sequence contain L frames $f_1, \dots, f_L$, where each frame $f_i, i \in [1, L]$ contains an image of $M \times N$ pixels. In order to extract human object features of the rowing pose in each video frame, the proposed approach first utilizes the OpenPose system with the body25 model to detect the human joint points of rowing poses. The human joint points from the OpenPose system consist of three values, namely, X coordinates, Y coordinates, and confidence C ($0.0 \leq C \leq 1.0$). Therefore, the set of body joints in the i th frame is obtained $\{(X_i, Y_i)\}_{i=1}^B$, where B is the number of human body joints. Figure 3a–i shows the detected human joint points in the consecutive rowing poses from the catch position to the recovery position during a rowing period.

After obtaining the rowing pose features of the human body joints, the features are presented as a graph structure in a graph-matching network (GMN) for further performance measurement. Figure 4 shows an example of two key rowing poses when rowing, the catch pose and the finish pose, which correspond to the detected graph structure of the rowing pose, as shown in Figure 3a and 3f, respectively, where the joints of the human body are represented with vertices, and the connections in the joint points of the human body are represented with edges.

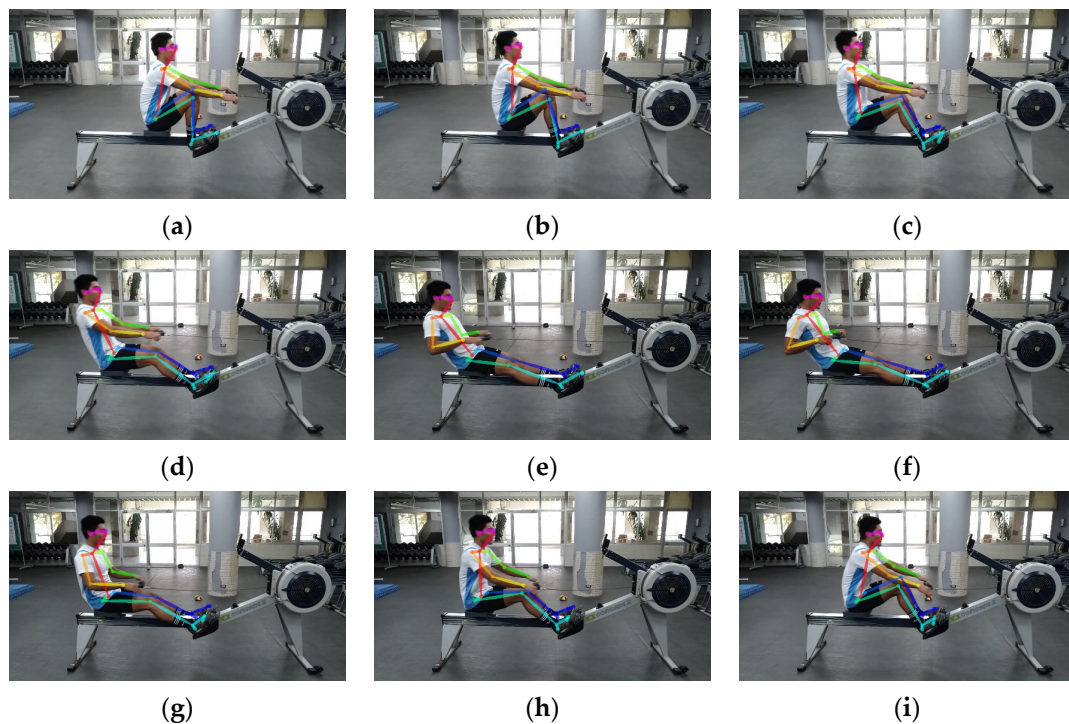


Figure 3. The human joint points detected during a rowing period using the OpenPose system. (a–i) The consecutive rowing poses from the catch position to the recovery position during a rowing period.

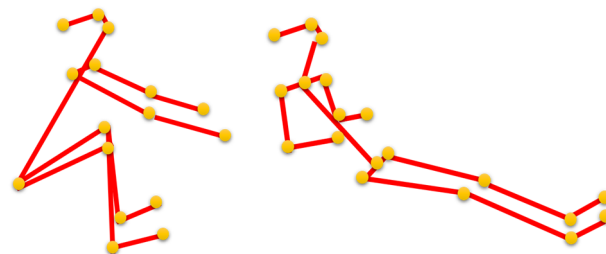


Figure 4. Illustration showing the detected rowing pose graph structures, which correspond to Figure 3a and 3f, respectively.

3.2. Baseline Comparison Analysis

To assess the performance of each pair of rowers, this study examined similarities between their rowing postures. The proposed system utilized a graph-matching network model, which contains a graph-embedding model (GEM) and a graph-matching network (GMN) [10], in order to extract the rowing posture feature vector from each video frame. Then, the similarities in the two rowing poses of each pair of video frames were calculated, which further demonstrated the performance of the paired rowers.

3.2.1. The Input Data for the Model

In order to acquire the graph feature vectors (*GF*), twenty-five joint points were represented by graph structures, and for each rowing posture, the detected *X* and *Y* coordinates of human joints were input into the GEM and GMN models. Li et al. [26] proposed that the input space of GEM and GMN models required one-dimensional data, that is, the *X* and *Y* coordinates are respectively input into the graph-matching network. However, their study found that one-dimensional inputs were unable to distinguish between two rowing poses. Therefore, this research improved previous work by proposing a two-dimensional input space for the coordinates of human joint points.

Let the human joint point graph structures be $G_i (V_i, E_i)$, $i \in [1, L]$, where V_i are the i th human joint point nodes and E_i is the set of edges that link the human joint point nodes. For the GEM and GMN models, the detected human joint points are formed as one-dimensional inputs $\{X_i\}_{i=1}^L$ and $\{Y_i\}_{i=1}^L$, respectively [26]. The proposed two-dimensional input data are $Input2D = \{(X_i, Y_i)\}_{i=1}^L$. Therefore, the one-dimensional and two-dimensional human joint data were inputted into the GEM and GMN models to acquire each rowing pose feature for each frame, respectively.

3.2.2. The Graph-Embedding Model (GEM)

The GEM and GMN models include a graph encoder layer, a propagation layer, and an aggregator layer [24]. In the graph encoder layer, feature encoding is performed on each node and edge of the graph, as follows:

$$h_i = MLP(x_{Vi}), \quad (1)$$

$$e_{ij} = MLP(x_{Eij}), \quad (2)$$

where x_{Vi} , x_{Eij} , and MLP are the node feature vector, the edge feature vector, and the neural network of the hidden layer, respectively. It should be noted that if there is no available message for the edge and node, x_{Vi} and x_{Eij} are set to 1 [24].

To exchange node values in the propagation layer, each transfer generates a value, and each node h_i^t receives values from adjacent nodes through the edges, resulting in the value of a new node h_i^{t+1} , which is defined as follows:

$$Value_{j \rightarrow i} = f_{Value}(h_i^t, h_j^t, e_{ij}), \quad (3)$$

$$h_i^{t+1} = f_{node}(h_i^t, \sum_{j \in E} Value_{j \rightarrow i}), \quad (4)$$

where f_{value} is a function based on MLP concatenation, f_{node} is any core network of MLP , $LSTM$, or GRU , h_j^t is the starting point of node value transfer, h_i^t is the end point of receiving node values, and e_{ij} is the edge that the node value passes through.

The aggregator layer aggregates nodes to obtain a graph feature vector GF , which is computed using:

$$GF = MLP\left(\sum_{i \in V} \sigma\left(MLP_{gate}(h_i^T)\right)\right) \odot MLP(h_i^T) \quad (5)$$

where $\odot$ is the Hadamard product and T is the times of message passing. After the T round of value passing, the feature vectors of the rowing postures for each frame are obtained $\{GF_i\}_{i=1}^L$.

To calculate the dissimilarity in two rowing postures between each pair of frames for a pair of players, the proposed approach uses the Euclidean distance metric to calculate the dissimilarity score ($DISSIM_{SCORE}$) between two rowing feature vectors $p1$ and $p2$, i.e.:

$$DISSIM_{SCORE(p1,p2)} = \sum_{i=1}^L \frac{\sqrt{(GF_i^{p1} - GF_i^{p2})^2}}{L}, \quad (6)$$

where $p1$ and $p2$ denote the number of paired rowers. In this study, 2 out of the 15 rowers were used for each similarity analysis. It was noted that lower value of the $DISSIM_{SCORE}$ denoted higher similarity between the rowing postures of two rowers, and it was more suitable for them to work together to improve the overall rowing performance, and vice versa.

3.2.3. The Graph-Matching Network (GMN)

This study also extracted feature vectors from each video frame using a graph-matching network (GMN). The main difference between the GEM and GMN models is that the propagation layers in the GMN model exchange node messages in the graph and add a cross-graph message exchange mechanism. Let two graphs be $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$. Each graph $G = (V, E)$ represents a set of nodes V and edges E . Each node $i \in V$ is associated with a feature vector x_i , and each edge $(i, j) \in E$ is associated with a feature vector x_{ij} . In the propagation layer, the computation is defined as:

$$Value_{j \rightarrow i} = f_{Value}(h_i^t, h_j^t, e_{ij}), \text{ for all } (i, j) \in E_1 \cup E_2 \quad (7)$$

$$Value_{j \rightarrow i} = f_{Value}(h_i^t, h_j^t, e_{ij}), \text{ for all } (i, j) \in E_1 \cup E_2 \quad (8)$$

$$h_i^{t+1} = f_{node}(h_i^t, \text{sum}(Value_i), \text{sum}(\mu_i)), \quad (9)$$

where $\text{sum}(Value_i)$ represents the sum of all values from node j to node i and $\text{sum}(\mu_i)$ represents the sum of all values from node j' to node i , respectively. f_{match} is a function of pairing cross-graph node value exchange, which is defined as,

$$a_{j \rightarrow i} = \frac{\exp(s_h(h_i^t, h_j^t))}{\sum_{j'} \exp(s_h(h_i^t, h_{j'}^t))}, \quad (10)$$

where $s_h()$ is the Euclidean squared distance metric. In order to pair the most similar nodes in a graph with another graph, the difference value for all nodes should be computed, i.e.,

$$\mu_{j \rightarrow i} = a_{j \rightarrow i} \times |h_i^t - h_j^t|, \quad (11)$$

and the total cross-graph value is:

$$\sum_j \mu_{j \rightarrow i} = h_i^t - (\sum_j a_{j \rightarrow i}) \cdot h_j^t, \quad (12)$$

where $a_{j \rightarrow i}$ is the attention weight and $\sum_j \mu_{j \rightarrow i}$ is the difference measured between h_i^t and its closest neighbor in another graph.

Finally, the aggregator layer aggregates the nodes under each graph to obtain graph feature vectors $GF1$ and $GF2$, which are defined as follows:

$$GF1 = f_G\left(\left\{h_i^T\right\}_{i \in V_1}\right), \quad (13)$$

$$GF2 = f_G\left(\left\{h_i^T\right\}_{i \in V_2}\right). \quad (14)$$

Finally, the dissimilarity of the pair of rowing postures is computed using Equation (6).

It should be noted that, in order to accurately measure the similarity in each pair of rowing postures between two video sequences, the proposed system must measure the rowing poses in the same frame baseline; hence, this study proposed a video that compares baseline processing. First, the starting video frame S_1 in the first video sequence V_1 was selected; then, the corresponding lowest dissimilarity video frame was searched based on the starting video frame's S_1 of V_1 . The starting video frame of the second video S_2 was obtained using:

$$S_2 = \min_{-100 \leq f_i \leq 100} DISSIM(V_1(GF_{S_1}), V_2(GF_{S_1} + (GF_{S_2} + f_i))), \quad (15)$$

where GF is the rowing feature vector, f_i is the number of frame offsets, and $DISSIM()$ is the dissimilarity function computed using the Euclidean distance.

3.3. Time-Period Similarity Processing

The purpose of the proposed time-period similarity processing (TPS) is to clearly distinguish the degree of similarity changes between the rowing postures of the paired rowers in the video segments.

In this study, the proposed approach performed segment similarity analyses in units of every 100 frames. The time-period similarity TPS was calculated as follows:

$$TPS = \frac{\sum_{n=0}^{np-1} DISSIM(X)}{np} \quad (16)$$

where np is the average value of the segment of every 100 frames and $DISSIM(X)$ is the dissimilarity measurement for each pair of rowing postures.

4. Results and Analysis

This section presents the experimental results of the proposed method. In this experiment, 15 high school rowers, who underwent training and practice in rowing techniques, used indoor rowing machines to simulate one minute of the actual rowing process [33,34], and each rower maintained a moderate rowing speed. Finally, we collected test videos from these 15 rowers to form the dataset used in our experiments. The test video was recorded at a speed of 60 frames per second (fps), and the resolution of each frame was 1920×1080 . All experiments were run on a computer equipped with an Intel Core i9-10900k CPU, 16 G RAM, and NVIDIA GeForce RTX 2080 GPU and analyzed using the Python 3 software development tool.

4.1. The Rowing Posture Similarity Analysis

This study trained the graph neural networks 5000 times. Five rounds of node messages were exchanged each time, and the trained model was used to evaluate rowing posture similarities in the rowers. In the experiment, 15 rowers were numbered from 0 to 14, and after pairing and grouping, four models were used to calculate similarities in the rowing postures of the 2 rowers in each group. The models were as follows: 1D input of the graph-embedding model, 2D input of the graph-embedding model, 1D input of the graph-matching network, and 2D input of the graph-matching network.

Tables 1 and 2 show the results of the rowing posture similarity analysis for all rower pairs using the 1D and 2D graph-embedding models (GEMs), respectively. In the dissimilarity matrix, which used the squared Euclidean distance defined in Equation (6) for similarity calculation, the red numbers indicate the lowest similarity in each column, while the green numbers indicate the highest similarity in each column. Furthermore, the lower the value, the greater the resemblance between the rowing postures. This indicates a higher suitability for assigning these rowers to the same group to further optimize their team's performance in the sport.

Table 1. The dissimilarity matrix among all player pairs using 1D input graph-embedding model ($\times 10^7$). The green numbers indicate the most similar pairing results, while the red numbers represent the least similar pairing results.

No.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0		2.93	2.40	4.89	2.11	2.83	2.32	4.15	3.26	2.84	3.85	4.03	2.83	3.29	1.43
1	2.93		2.28	3.21	2.65	2.79	2.39	3.17	2.90	3.01	3.36	3.11	3.50	3.19	2.63
2	2.40	2.28		3.39	2.39	2.87	2.68	3.37	3.05	3.13	3.51	2.72	3.21	3.05	2.38
3	4.89	3.21	3.39		3.09	3.25	2.34	2.43	3.33	3.67	2.54	1.23	5.14	3.36	3.92

Table 1. *Cont.*

No.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
4	2.11	2.65	2.39	3.09		2.77	1.40	4.85	4.32	2.66	4.45	2.86	2.25	3.38	2.71
5	2.83	2.79	2.87	3.25	2.77		2.55	2.56	2.44	2.31	3.73	3.20	2.87	3.16	2.47
6	2.32	2.39	2.68	3.25	1.40	2.55		3.09	2.95	1.53	2.90	2.75	3.02	2.64	1.85
7	4.15	3.17	3.37	2.43	4.85	2.56	3.09		2.08	2.70	1.90	2.32	5.06	2.62	2.48
8	3.26	2.90	3.05	3.33	4.32	2.44	2.95	2.08		2.68	2.61	3.77	3.41	2.92	2.22
9	2.84	3.01	3.13	3.67	2.66	2.31	1.53	2.70	2.68		3.66	3.61	3.35	2.64	2.17
10	3.85	3.36	3.51	2.54	4.45	3.73	2.90	1.90	2.61	3.66		2.50	6.98	3.84	2.55
11	4.03	3.11	2.72	1.23	2.86	3.20	2.75	2.32	3.77	3.61	2.50		5.04	3.12	3.62
12	2.83	3.50	3.21	5.14	2.25	2.87	3.02	5.06	3.41	3.35	6.98	5.04		3.46	3.15
13	3.29	3.19	3.05	3.36	3.38	3.16	2.64	2.62	2.92	2.64	3.84	3.12	3.46		2.51
14	1.43	2.63	2.38	3.92	2.71	2.47	1.85	2.48	2.22	2.17	2.55	3.62	3.15	2.51	

Table 2. The dissimilarity matrix among all player pairs using 2D input graph-embedding model ($\times 10^7$). The green numbers indicate the most similar pairing results, while the red numbers represent the least similar pairing results.

No.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0		10.14	8.01	16.40	6.97	10.22	8.24	13.67	11.70	10.31	11.33	15.09	8.86	11.63	4.77
1	10.14		7.95	10.89	9.13	10.47	8.44	10.93	10.59	11.43	10.58	12.19	11.48	11.70	9.71
2	8.01	7.95		11.03	7.62	10.65	9.31	11.32	10.85	11.71	10.70	10.46	9.91	10.90	8.15
3	16.40	10.89	11.03		9.96	11.48	7.97	8.01	11.62	13.20	7.76	3.37	17.37	11.48	14.75
4	6.97	9.13	7.62	9.96		9.59	4.34	15.88	14.43	9.09	13.21	10.38	7.01	11.18	9.50
5	10.22	10.47	10.65	11.48	9.59		9.64	9.31	9.49	9.32	12.34	13.29	9.37	123.1	9.52
6	8.24	8.44	9.31	7.97	4.34	9.64		11.13	11.05	4.65	8.35	10.54	11.08	9.49	7.38
7	13.67	10.93	11.32	8.01	15.88	9.31	11.13		7.04	9.76	5.73	8.39	16.02	9.29	8.60
8	11.70	10.59	10.85	11.62	14.43	9.49	11.05	7.04		10.38	8.15	15.20	10.99	10.74	8.13
9	10.31	11.43	11.71	13.20	9.09	9.32	4.65	9.76	10.38		12.15	15.28	11.26	10.22	8.00
10	11.33	10.58	10.70	7.76	13.21	12.34	8.35	5.73	8.15	12.15		8.22	20.63	12.44	7.56
11	15.09	12.19	10.46	3.37	10.38	13.29	10.54	8.39	15.20	15.28	8.22		18.73	12.36	14.59
12	8.86	11.48	9.91	17.37	7.01	9.37	11.08	16.02	10.99	11.26	20.63	18.73		11.22	11.05
13	11.63	11.70	10.90	11.48	11.18	123.1	9.49	9.29	10.74	10.22	12.44	12.36	11.22		9.02
14	4.77	9.71	8.15	14.75	9.50	9.52	7.38	8.60	8.13	8.00	7.56	14.59	11.05	9.02	

4.2. Rowing Posture Analysis

Figures 5 and 6 show the posture similarity degree changes in the 1D and 2D inputs for the graph embedding models (GEMs) and graph-matching network (GMN), respectively, during the one-minute rowing process. The green solid line in the figure represents the most similar rower pair including No. 0 and No. 14, while the red dotted line represents the most dissimilar rower pair including No. 0 and No. 3. As shown in Figures 5a and 6a, the differences in rowing postures between two rowers are not quite distinguished in the 1D input for GEM and GMN. However, the improved 2D input for GEM and GMN can slightly distinguish differences, as shown in Figures 5c and 6c. Furthermore, our proposed time-period similarity analysis approach can significantly distinguish temporal differences in rowing motions, as shown in Figures 5b,d and 6b,d. Moreover, Figures 5c,d and 6c,d

show that in the initial 500 frames, the posture similarities between rowers No. 0 and No. 3, which had the worst similarity originally, were relatively similar during this period. However, their differences in posture similarities became larger after about 500 frames. In contrast, rowers No. 0 and No. 14 maintained high similarities after 500 frames.

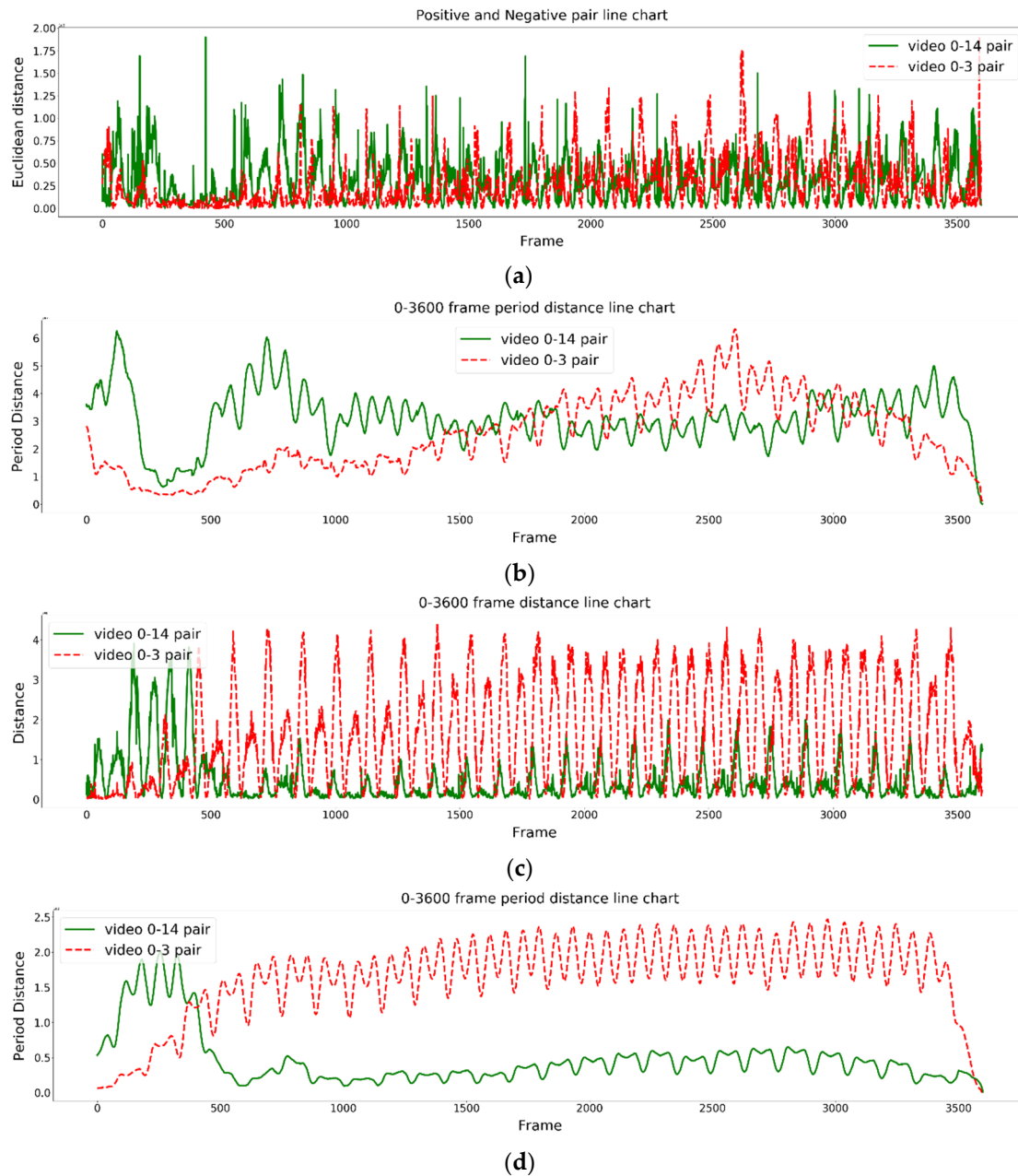


Figure 5. Experimental results of the 1D and 2D graph-embedding models (0–3600 frames): (a) 1D input GEM, (b) the time-period similarity results of (a), (c) 2D input of GEM, and (d) the time-period similarity results of (c).

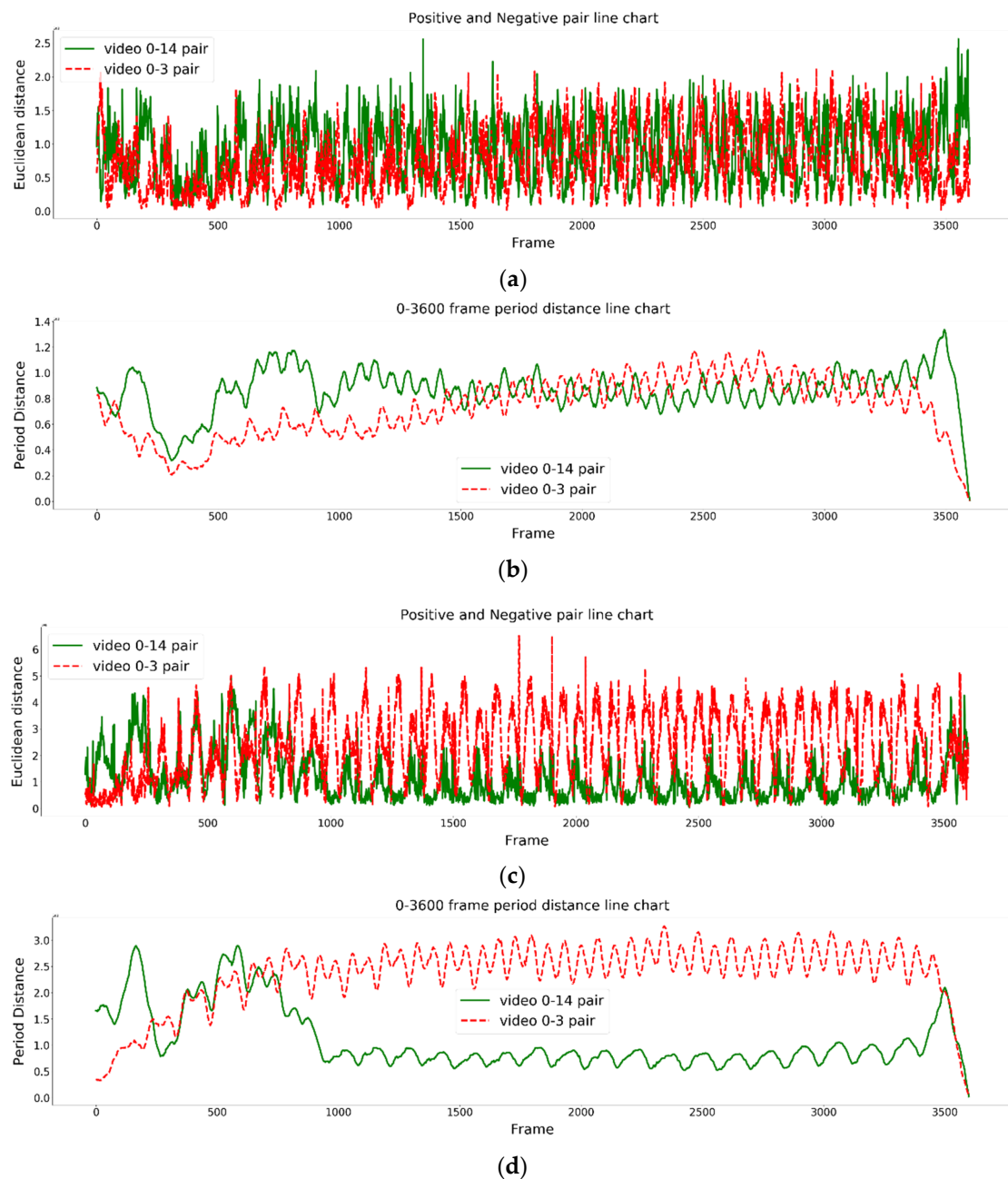


Figure 6. Experimental results of the 1D and 2D graph-matching networks (0–3600 frames): (a) 1D input GMN, (b) the time-period similarity results of (a), (c) 2D input GMN, and (d) the time-period similarity results of (c).

Figure 7 shows the similarities in the rowing posture results for two pairs of rowers regarding the 2D input for the graph-embedding model (GEM) at frame numbers 0–3600. The green solid line in the figure represents the rowing pair with the most similar rowing postures, while the red dotted line represents the rowing pair with the most dissimilar rowing postures. Figure 7a shows the rowing posture similarities in paired rowers using the 2D input GEM, while Figure 7b shows the proposed time-period similarity analysis results of Figure 7a. It can be found that the rowing postures of the two rowers were not stable before frame number 500 at the beginning of rowing; however, in the middle part of the video, the rowing postures of the two rowers became stable, as marked by the green line.

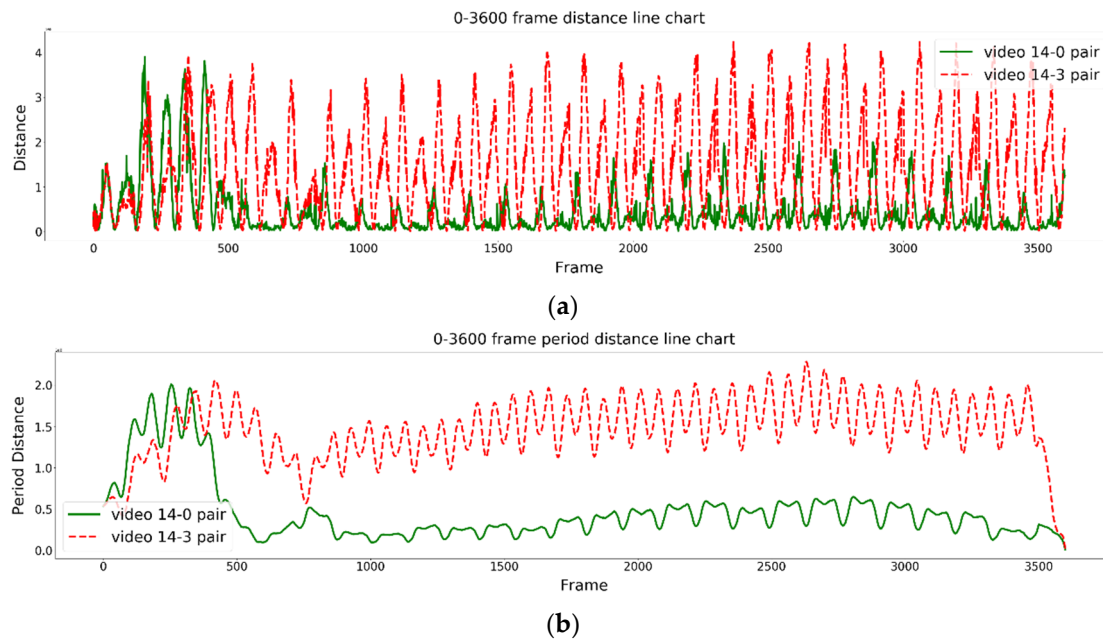


Figure 7. Experimental results for rower numbers 14, 0, and 3. (0–3600 frames): (a) 2D input GEM, which (b) is the proposed time-period similarity results of (a).

4.3. Visual Validation

In order to verify the effectiveness of the approach proposed in this study, skeletons of the rowing postures for the two groups of rowers were superimposed, their posture similarities were visualized for comparison, and 0~3600 frames were used for each group. Group 1 consisted of rowers No. 0 and No. 3 and had the lowest rowing posture similarities, while Group 2 consisted of rowers No. 0 and No. 14 and had the highest rowing posture similarities.

Figure 8 shows the results of superimposing the skeletons of rowers No. 0 and No. 3 with the lowest rowing posture similarities, in which the red is rower No. 0 and the black is rower No. 3. Figure 8a–e shows that at the beginning of rowing, i.e., frame numbers 0~500, the rowing postures of the skeletons were matched because the rowing rhythms of the two rowers were almost the same. It should be noted that the initial motions in the rowing posts of rower No. 0 and rower No. 3 and their rhythms are very similar. However, as time progressed and physical exertion increased, their rowing poses and rhythms could not be consistently maintained, resulting in a gradual dissimilarity in their rowing poses over time. This observation is in line with the prediction results of the rowing posture analysis using the 2D GMN analysis model in the previous section [26]. In addition, Figure 8f–j shows that the rowing postures of the two rowers were already quite different; hence, the skeletons in the rowing postures did not match after frame number 500.

Figure 9 shows the results of superimposing the skeletons for rower No. 0 and No. 14, which had the highest rowing posture similarities, where red is rower No. 0 and black is rower No. 14. Figure 9a–e shows that the rowing postures of the two rowers were not well matched at the beginning of rowing, i.e., frame numbers 0~500. However, the skeletons of their rowing postures began to match after frame number 500, as shown in Figure 9f–j. It should be noted that the initial motion of rower No. 0's rowing pose was standard, but their rowing rhythm could not be consistently maintained. On the other hand, rower No. 14's rowing pose at the beginning of rowing was not quite standard, but their rowing rhythm was stable. Due to this, rowers No. 0 and No. 14 were not well-matched at the beginning of rowing. However, as the rowing rhythm was adjusted, the rowing poses of the two individuals gradually became more similar. The results of this experiment show that the similarity values obtained using this research method were all in line with the results of the

visualized superimposed skeleton comparison. The final similarity matrix can be used to judge whether any two rowers are suitable to be grouped to increase the rowing speed.

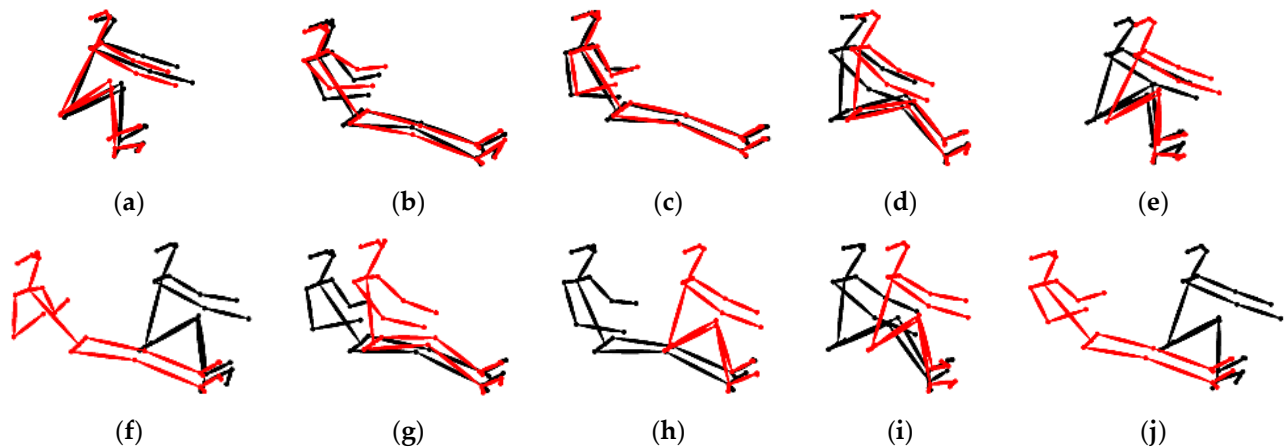


Figure 8. Rower numbers 0 and 3: (a–e) results comparing every 30 frames after frame number 0 and (f–j) results comparing every 30 frames after frame number 500.

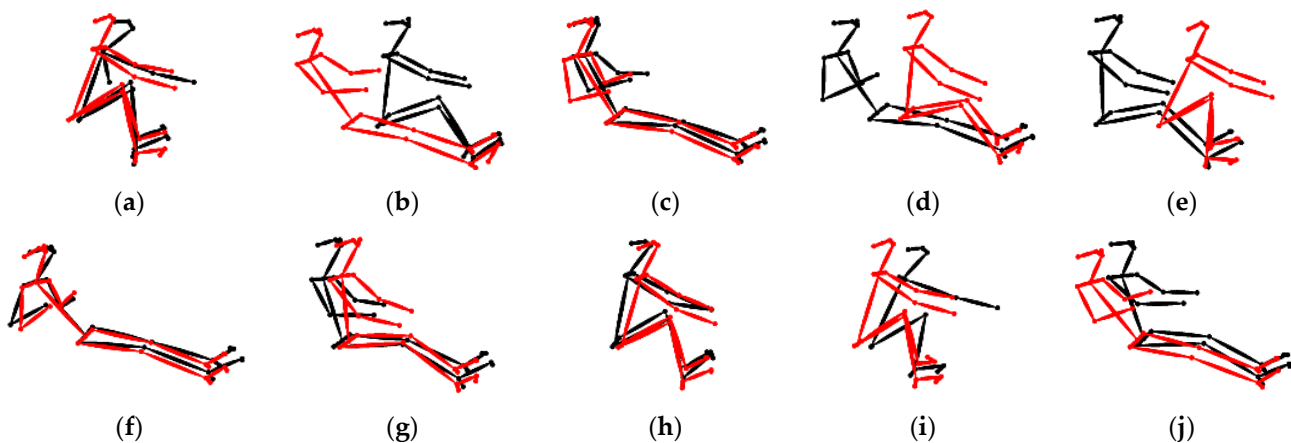


Figure 9. Rower numbers 0 and 14. (a–e) Results comparing every 30 frames after frame number 0 and (f–j) results comparing every 30 frames after frame number 500.

5. Conclusions

This study presented an efficient approach for evaluating the performance of rowing pairs using a graph-matching network. The proposed approach used the OpenPose system to obtain the rowers' postures during the rowing process and acquire the positions of human joints. Afterward, the 2D coordinates of the detected human joints were input into GEM and GMN models to extract the feature vectors of each rowing posture. The proposed video baseline analysis results calculate the performance of the rowing pairs. Furthermore, the similarities in each pair of rowing postures were efficiently calculated using the GEM and GMN models. The proposed time-period similarity analysis clearly distinguished the degree of similarity changes in rowers over time. This experiment was carried out to demonstrate that the proposed approach could effectively evaluate the performance of rowing pairs and provide good suggestions for coaches when creating rower grouping and training.

The main limitation of the proposed approach is that this study only simulates indoor rowing using indoor rowing machines, which may impose several limitations when applied to actual outdoor rowing scenarios. The primary strength of this study lies in the novel approach using graph-embedding models and graph-matching networks for rowing

posture analysis. Despite its focus on rowing, this system can also be applied to other sports, thereby enhancing overall team performance. In the future, identifying how to use 3D posture features and integrating a 3D graph neural network model to improve the proposed system merit further study.

Author Contributions: Conceptualization, C.-C.C., W.-H.C., C.-H.C. and I.-C.C.; methodology, C.-C.C. and C.-S.L.; software Y.-T.C.; writing—original draft preparation, C.-S.L.; writing—review and editing, C.-S.L. and C.-C.C.; supervision, C.-C.C. and I.-C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Reinschmidt, C.; Van-den-Bogert, A.J.; Nigg, B.M.; Lundbrg, A.; Murphy, N. Effect of Skin Movement on the Analysis of Skeletal Knee Joint Motion During Running. *J. Biomech.* **1997**, *30*, 729–732. [[CrossRef](#)] [[PubMed](#)]
2. Fujimori, Y.; Ohmura, Y.; Harada, T. Wearable motion capture suit with full-body tactile sensors. In Proceedings of the IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009.
3. Ren, Z.; Meng, J.; Yuan, J.; Zhang, Z. Robust hand gesture recognition with kinect sensor. In Proceedings of the ACM International Conference on Multimedia, Scottsdale, AZ, USA, 28 November–1 December 2011.
4. Wei, S.E.; Tang, N.C.; Lin, Y.Y.; Weng, M.F.; Liao, H.Y.M. Skeleton-augmented human action understanding by learning with progressively refined data. In Proceedings of the ACM International Conference on Human Centered Event Understanding from Multimedia, Orlando, FL, USA, 7 November 2014.
5. Igor, J.; Julio, C.G.; Francesco, C.; Jaime, S.; Francesc, C.; Luka, M.; Ivan, K.; Sergej, O.; Jesús, O.; Bernardo, R.; et al. The 360° Performance System in Team Sports: Is It Time to Design a “Personalized Jacket” for Team Sports Players? *Sports* **2021**, *9*, 40.
6. Chen, C.C.; Chang, C.; Lin, C.S.; Chen, C.H.; Chen, I.C. A Video based basketball shooting prediction and pose suggestion system. *Multimed. Tools Appl.* **2023**, *82*, 27551–27570. [[CrossRef](#)]
7. Duan, C.; Hu, B.; Liu, W.; Song, J. Motion Capture for Sporting Events Based on Graph Convolutional Neural Networks and Single Target Pose Estimation Algorithms. *Appl. Sci.* **2023**, *13*, 7611. [[CrossRef](#)]
8. Yuan, B.; Kamruzzaman, M.; Shan, S. Application of Motion Sensor Based on Neural Network in Basketball Technology and Physical Fitness Evaluation System. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5562954. [[CrossRef](#)]
9. Chang, Y.J.; Chen, S.F.; Huang, J.D.A. Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Res. Dev. Disabil.* **2011**, *32*, 2566–2570. [[CrossRef](#)] [[PubMed](#)]
10. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 7291–7299. [[CrossRef](#)] [[PubMed](#)]
11. Qiao, S.; Wang, Y.; Li, J. Real-Time human gesture grading based on OpenPose. In Proceedings of the IEEE International Conference on Image and Signal Processing, BioMedical Engineering and Informatics, Shanghai, China, 14–16 October 2017.
12. Tsai, Y.S.; Hsu, L.H.; Hsieh, Y.Z.; Lin, S.S. The real-time depth estimation for an occluded person based on a single image and OpenPose method. *Mathematics* **2020**, *8*, 1333. [[CrossRef](#)]
13. Nakai, M.; Tsunoda, Y.; Hayashi, H.; Murakoshi, H. Prediction of basketball free throw shooting by OpenPose. In Proceedings of the International Conference on New Frontiers in Artificial Intelligence, Yokohama, Japan, 12–14 November 2018.
14. Toshev, A.; Szegedy, C. DeepPose: Human pose estimation via deep neural networks. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
15. Xiao, B.; Wu, H.; Wei, Y. Simple baselines for human pose estimation and tracking. In Proceedings of the European International Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
17. Zhang, B.; Chen, C.Y.; Chan, L.C.; Fok, W. Intelligent Sports performance Scoring and Analysis system Based on Deep Learning Network. In Proceedings of the International Conference on Artificial Intelligence and Big Data, Chengdu, China, 28–31 May 2020.
18. Theagarajan, R.; Pala, F.; Zhang, X.; Bhanu, B. Soccer: Who Has the Ball? Generating Visual Analytics and Player Statistics. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

19. Smedt, Q.D.; Wannous, H.; Vandeborre, J.P.; Guerry, J.; Saux, B.L.; Filliat, D. SHREC'17 track: 3D Hand gesture recognition using a depth and skeletal dataset. In Proceedings of the International Conference on Eurographics Workshop on 3D Object Retrieval, Lyon, France, 23–24 April 2017.
20. Devineau, G.; Xi, W.; Moutarde, F.; Yang, J. Convolutional neural networks for multivariate time series classification using both inter- and intra- channel parallel convolutions. In Proceedings of the International Conference on Reconnaissance des Formes, Image, Apprentissage et Perception, Paris, France, 25–28 June 2018.
21. Hou, J.; Wang, G.; Chen, X.; Xue, J.H.; Zhu, R.; Yang, H. Spatial-temporal attention Res-TCN for skeleton-based dynamic hand gesture recognition. In Proceedings of the European International Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
22. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Trans. Neural Netw.* **2008**, *20*, 61–80. [[CrossRef](#)] [[PubMed](#)]
23. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *32*, 4–24. [[CrossRef](#)] [[PubMed](#)]
24. Sperduti, A.; Starita, A. Supervised neural networks for the classification of structures. *IEEE Trans. Neural Netw.* **1997**, *8*, 714–735. [[CrossRef](#)] [[PubMed](#)]
25. Gori, M.; Monfardini, G.; Scarselli, F. A New Model for Learning in Graph Domains. In *International Joint Conference on Neural Networks*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2005; pp. 729–734.
26. Li, Y.; Gu, C.; Dullien, T.; Vinyals, O.; Kohli, P. Graph Matching Networks for Learning the Similarity of Graph Structured Objects. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019.
27. Yan, S.; Xiong, Y.; Lin, D. Spatial Temporal Graph Convolutional Network for Skeleton-Based Action Recognition. In Proceedings of the International Conference on the Association for the Advance of Artificial Intelligence, New Orleans, LA, USA, 21–23 February 2018.
28. Monisha, M.; Mohan, P.S. A Real-Time Embedded System for Human Action Recognition Using Template Matching. In Proceedings of the International Conference on Electrical, Instrumentation and Communication Engineering, Karur, India, 27–28 April 2017.
29. Cho, Y.; Yoon, K. PaMM: Pose-Aware Multi-Shot Matching for Improving Person Re-Identification. *IEEE Trans. Image Process.* **2017**, *27*, 3739–3752. [[CrossRef](#)] [[PubMed](#)]
30. Tiwari, E.; Sardar, P.; Jain, S. Football Match Result Prediction Using Neural Networks and Deep Learning. In Proceedings of the Reliability, Infocom Technologies and Optimization, Noida, India, 4–5 June 2020.
31. Edouard Sarlin, P.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learn Feature Matching with Graph Neural Networks. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
32. Li, J.; Li, J. An Improved Pose Matching Method Based on Deep Learning. In Proceedings of the IEEE International Conference on Civil Aviation Safety and Information Technology, Kunming, China, 17–19 October 2019.
33. Von Zitzewitz, J.; Wolf, P.; Novakovic, V.; Wellner, M.; Rauter, G.; Brunschweiler, A.; Riener, R. Real-time rowing simulator with multimodal feedback. *Sports Technol.* **2008**, *1*, 257–266. [[CrossRef](#)]
34. Hamilton, N.; Weimar, W.; Luttgens, K. *Kinesiology: Scientific Basis of Human Motion*, 12th ed.; McGraw-Hill: New York, NY, USA, 2012.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.