IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 February 2023, accepted 17 February 2023, date of publication 22 February 2023, date of current version 10 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3247583

RESEARCH ARTICLE

Periodic Physical Activity Information Segmentation, Counting and Recognition From Video

SHENG-HSIEN CHENG[1], MUHAMMAD ATIF SARWAR[1], YOUSEF-AWWAD DARAGHMI[2], TSÌ-UÍ İK[1], (Member, IEEE), AND YIH-LANG LI[1], (Member, IEEE)

[1]Department of Computer Science, College of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
[2]Department of Computer Systems Engineering, Palestine Technical University-Kadoorie, Tulkarem P310, Palestine

Corresponding author: Tsì-Uí İk (cwyi@nctu.edu.tw)

This work was supported by the Center for Open Intelligent Connectivity from the Featured Areas Research Center Program within the Higher Education Sprout Project framework by the Ministry of Education (MOE), Taiwan. The work of Tsì-Uí İk was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 110-2627-H-A49-001, Grant MOST 110-2221-E-A49-063-MY3, and Grant MOST 111-2622-E-A49-009.

ABSTRACT The research on complex human body motion including sports and workout activity recognition is a major challenge and long-lasting problem for the computer vision community. Recent development in deep learning algorithms to track people's workout activity characteristics based on video sensors can be used to infer the human body pose for further analysis. Specifically, tracking complex body movements while performing multi-pose physical exercise helps individuals provide fine granularity feedback including activity repetition counting and activity recognition. Therefore, this research proposes a system that provides a repetition counter and activity recognition of physical exercise from video frames (extracted 3D human skeleton using VIBE) based on the deep semantic features and repetitive segmentation algorithm. The proposed system locates both ends of the activity's action and segments the activity into multiple unit actions which improves activity recognition, time intervals, # of sets, and other quantitative values of activity. The proposed system is evaluated on the physical activities dataset named ''NOL-18 Exercise'' through extensive experiments. The proposed system results show that the accuracy of the repetitive action segmentation is 96.27% with 0.23% time error, and action recognition reaches 99.06%. The system can be employed to fitness or rehabilitation centers and used for treating patients.

INDEX TERMS Activity analysis, activity segmentation, periodical information mining, activity recognition, repetitive action counting.

I. INTRODUCTION

Performing physical activity regularly is an essential part of people's daily life for strengthening the immune system and the cardiovascular system, balancing anxiety, and maintaining the mental and physical health [1], [2], [3], [4], [5], [6], [7], [8]. People from different age groups prefer to perform physical activities, e.g. push up, running, and biceps, in a periodical manner. However, regardless of the daily encouragement by the healthcare society, maintaining a workout regimen in our fast-paced lifestyle is challenging [9], [10], [11]. Therefore, this motivates us to explore new horizons that can encourage people of all ages to perform physical activities and track exercises repetitions and movement quality.

Recently, several sports gadgets and smartphone applications have become available to assist people in tracking workout activities and coaching. These applications and gadgets have broadly become available with off-the-shelf mobile devices specifically smartwatches or smart bands. Some of these gadgets even have complete ecosystems with the goal of increasing the motivation to work out and helping

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

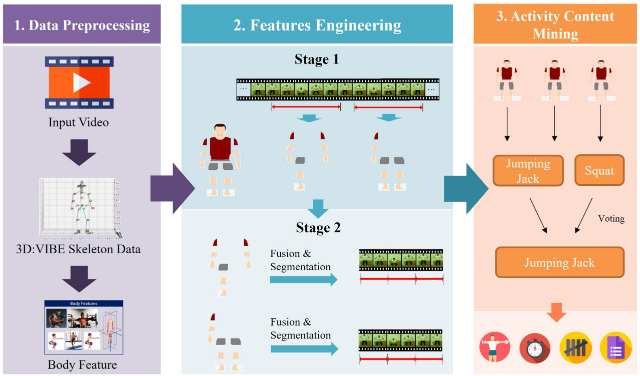

FIGURE 1. Prototype of repetitive action segmentation and recognition.

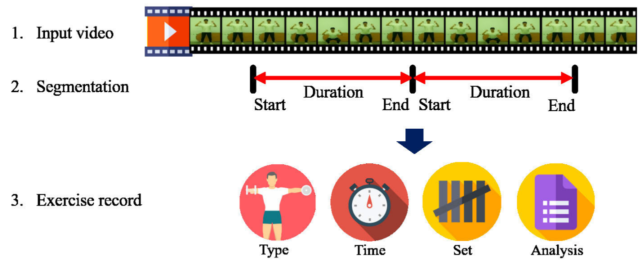

FIGURE 2. Periodic physical activity system overview.

users to keep on tracking their progress. These gadgets include feature positioning sensors, such as Global Position System (GPS), inertial measurement units (IMU) sensors, accelerometers, and gyroscopes fitness trackers. Most of these devices perform well with simple tracking, counting the number of daily steps, energy estimation, and sitting time for long-lasting exercises, such as running, cycling, and swimming. However, these devices usually have some concerns to cover a wider range of complex sports activities, especially for commercial reasons [12], [13]. Similarly, people are more concerned about how many times they have performed the exercise, and they need to maintain track of their progress. But, these devices can not provide feedback on a finer level of granularity [13], [14], [15].

To tackle the aforementioned shortcomings of the widely used gadgets and apps, this work proposes a novel hierarchical repetitive action segmentation and recognition system that counts, validates, and recognizes the repetition from the physical activity video which is commonly used in a workout regime as illustrated in figure 1. The proposed system works at a single frame level to detect and recognize the current movement on a finer level of granularity of the exercise execution that provides guidance to fitness enthusiasts. The system consists of three modules that are: The multi-features pre-processing module analyzes the physical activity video frame-by-frame to extracts the 3D skeleton information using 3D pose estimation model to find the unit action [16], [17], [18], [19] for complex sports activities, especially the multitude of movements in one step. The second module is the segmentation of the repetitive movements in the time domain based on feature selection and peak analysis. The third module is classifying the complex activities based on movement characteristics to know the activity type, counting the repetition, and other quantitative values.

The proposed system faces a major challenge which is the collection of large scale data that are accurately labelled, similar to [20], [21], [22], and [23]. This is only possible with extensive human effort including supervision of recording and analyzing the annotated data. Thus, we designed a data collection system that allows collection of high quality data with minimal supervision to identify the repetitive action and recognize it as illustrated in figure 2. First, the system labels the activity data into the planned, organized, and repetitive movement stages including contraction (C), relaxation (R), pause (P), and noise (N). Each movement stage is represented by the instance of a change of body joint position that looks like a quarter-cycle wave and each half-cycle is composed of several movement stages. The system labels each half-cycle w.r.t the starting and ending time of the movement and counts the repetitive movements per half-cycle. Second, the system labels each physical activity w.r.t starting time of the first quarter-cycle and ending time of the last quarter-cycle. Lastly, the system labels the exercise type, time, # of sets, counting, and and other quantitative values, such as joint angle, speed velocity, and displacement. The output of the data collection system is the NOL-18 activity dataset.

The accuracy of the proposed system is evaluated using the NOL-18 activity dataset.The data set was captured independently of subjects that performed physical activity, and it contains gym exercises and square dance movements in 211 videos of 18 different exercises with three different shooting angles. The repetitive action segmentation module is evaluated by IoU metrics and time error. The repetitive action segmentation is measured between the predicted and ground-truth value by the ratio $\geq$ 50% and the time error duration of the predicted value which must be within the range of $\pm 1/3$ of the ground truth value duration. The repetitive action segmentation module has an accuracy is 96.4% with a 0.20s time error on the NoL-18 dataset. Similarly, the action recognition module is evaluated on the NoL-18 dataset, and its accuracy is 99.06%. Our primary contributions are listed below

- The development of a periodical activity framework that is capable of counting the repetitions, activity recognition, activity time, # of sets, and other quantitative values including joint angle, speed velocity, and displacement, and validating them based on three main modules.
- Creating the NOL-18 dataset containing 211 videos of physical exercise including gym exercise and dancing movement which were collected by the Network Optimization Lab (NOL), NYCU, Taiwan.
- Developed framework extracted the skeleton information from video, identify the keypoint axis by PCA,

extracted the key feature for each body part like arm movement, machine learning model detect and recognize the activity based on the movement of the key features.

The rest of the paper is organized as follows: Section II discusses the related work about activity segmentation, object detection, and activity recognition. Section III presents the methodology and prototype design. Section IV-B describes the performance evaluation and implementation environment, and section V concludes this research.

II. RELATED WORK

Human activity recognition (HAR) with rich bodywork is an important sub-field of human-computer interaction (HCI). HAR covers an extensive range of tasks including activity recognition [24], repetition counter [25], [26], [27] and distinguishing activity from non-activity [28], [29]. The outcomes of these tasks have a wide range of potential in real-world applications including healthcare the and physical fitness. The most popular approach throughout the years for activity tracking and coaching for fitness is the use of commercially available wearable inertial sensor devices. For instance, Google's Wear [30] and Apple watch [31] collect the activity data and distinguish between moving, exercising, and standing positions, and support activity tracking naively through third-party apps. Also, motion sensors are used for a fitness tracking system that consists of separate lightweight heart rates, and motion trackers are worn on the wrist for boxing [32]. Nowadays, HAR is based upon many different modalities, such as internal motion units (IMU) [33], [34], [35], [36], electromagnetic waves [37], [38], sound [39] and video [40], [41].

Several studies have focused on using wearable devices including Heeven et al. [42] that proposed a deep learning-based network based on reflectance-type photoplethysmography (PPG) sensors to ensure accurate heart rate estimation even during intensive exercise. Junchang et al. [43] proposed Jerk-Kalman-based algorithm to estimate motion velocity, acceleration detection, and energy expenditure for human limb motion. Franca et al. [44] proposed stress detection system based on several machine learning algorithms based on wearable technologies and m-health solutions. Mokaya et al. [45] presented a MyoVibe system to determine muscle movement in high-motion physical exercises that sense vibration signals of muscles using the network of accelerometers and modeling the characteristics of muscles. Cheng et al. [46] developed a methodology to recognize physical activity without training by putting a sensor on the user's hip. Similarly, other authors developed activity recognition systems using the different wearable sensors, as shown in table 1.

There are other technologies also developed to recognize the complex physical movement in this race. For example, the Motion Capture (MOCAP) system consists of multiple cameras positioned to capture the 3D positions of reflective markers put on an actor's clothing in a specially designed room. The MOCAP system tracks human motion very accurately. But this system is not suitable for sports and workouts enthusiast because they do not prefer wearing tight suits with markers, and these systems are very expensive [50], [51]. Similarly, Microsoft Kinect devices have been used in cognitive assessment and rehabilitation. Kinect-based systems [52] could facilitate the proper performance of rehabilitation exercises at home that significantly minimize trips to a physical therapy center, which are costly and inconvenient for patients. Game consoles, Microsoft Kinect, and other systems employing various ways of user interaction have nowadays been established as a form of home-based exercise. This system monitors the patient's movements, repetitions, and most importantly body posture based on the recordings of the exercises in 3D form. However, certain constraints create some problems including occlusion [53], and complex recognition of exercise movement.

Up to now, the progress in the HAR field is dominated by the classical machine learning methods that perform the classification and segmentation tasks. These classical methods include random forest, hidden Markov models, naive Bayes classifiers, decision trees, and support vector machines. Further, the increasing available data via off-the-shelf devices, such as smartwatches, smartphones, and fitness trackers, increases the complexity of data, particularly activity patterns detection, which shifts the focus towards using computer vision (CV) and deep learning (DL) [54], [55]. The CV and DL work extremely well for HAR as compared to the classical methods of machine learning that require feature engineering.

Recently, Human body pose estimation for HAR is becoming most popular using CV and DL for complex activity classification. The recent advances in pose estimation that created a trans-disciplinary dialogue discourse allow the computer vision community to produce a vast space of solutions for workout activity recognition and repetition counter. Interestingly, HAR methods use images and videos to represent semantic-based human recognition. These methods identify the space and semantic-based features including, pose, poselet, attributes, and scene context based on the 3D and depth skeleton data. For instance, GymCam [56] is a single-camera-based solution that uses Object detection to extract the movement characteristics of a single person and then uses neural networks to identify the activity type and count the repetitions. However, it cannot count the time of each exercise and cannot provide the quality of the activity.

Pose Trainer [57] is a single camera-based application that detects and evaluates the human pose and provides personalized recommendations. However, it does not give information about activity repetition. Wolf and levy [58] proposed a CNN-based online solution that contains repetitive movements. Their solution counts the repetition variability in length and appearance from online and real-time videos, however, it does not measure the quality of each exercise, # of a set of each activity, time of activity. Runia et.al. [59] proposed a real-world repetition estimation model for activities such as push-

TABLE 1. Comparison of activity repetition methods.

Paper	Data Collection	Method	Strength	Weakness
Shoaib et al.[24]	IMUs	K-Nearest Neighbor etc.	13 Activities classified	Less-repetitive activities
Soro et al. [12]	CrossFit	Neural Network	10 full-body Exercises classified	No-Dancing Activity
O'reilly et al. [47]	5 wearable sensor	Random Forest	lower body exercises classified	No-Upper Body
Ebert et al. [48]	Accelerometer	Naive Bayes	8 bodyweight exercises	No Counting
Burns et al. [49]	IMUs	Convolutional RNN (CRNN)	Shoulder Exercises classified	No Full-Body
Um et al. [21]	forearm-worn wearable sensor	CNN	50 gym exercises	No Counting
Chang et al.[27]	accelerometers	naive Bayes	9 exercises classified and counting	miscount rate 5%
Seeger et al. [26]	accelerometers	Gaussian distribution	16 typical activity classified	multiple counting algorithms
Muehlbauer et al. [28]	smartphone	k-nearest neighbours	10 upper body exercises classified	multiple classical algorithms
Morris et al. [29]	IMUs	autoa -correlation	14 gym exercises classified	miscount rate 7%
Shen et al. [25]	smartphones and wearables	Conditional Random Fields	Weightlifting Exercise classified	multiple classical counting

ups, cutting a melon, etc. however, it does not provide a detailed analysis of each activity.

Interestingly, many other fitness tracker methods including [60], [61], and [62] proposed the exercise repetitions model for the different perspectives of patients' rehabilitation and fitness on a single activity. However, they do not cover the wider range of physical activity to count the activity repetitions and provide deep personalized analysis. Therefore, there is a need for a system that addresses the problem of complex activity at a single unit level of action to counter the repetition and recognize and validate the activity from the video.

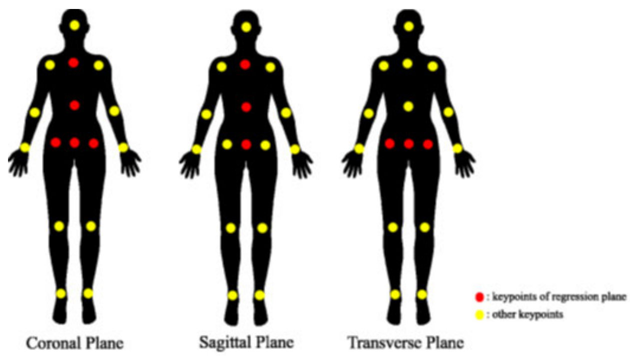

FIGURE 3. Human plane with joint points.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The system architecture and workflow are illustrated in figure 1. The proposed network is mainly divided into three modules including data pre-processing, repetitive action segmentation and counting, and action recognition.

A. DATA PRE-PROCESSING

Data pre-processing consists of multiple steps including Keypoint Normalization and Geometry Evaluation to smooth each keypoints of every frame of a single activity.

1) KEYPOINTS NORMALIZATION

A deep learning-based 3D pose estimation model called Video Inference for Body Pose and Shape Estimation (VIBE) [63] is used to extract the activity data of human skeleton joints as illustrated in figure 4. VIBE generates the 3D coordinates (x,y,z) data of 49 joints of the human body for each frame of video. Each 3D coordinate is tracked to smooth the body keypoint to avoid obscured frames. These 3D keypoints generalized because of different body length measurements, distance from the camera, and other relative factors [57] and generalized these keypoints based on torso's length. To calculate the torso's length, compute the average of the distance from the neck keypoints to the right and left hip keypoints. Principal component analysis (PCA) is applied on each joint keypoint to identify which keypoint axis is most visible throughout all the activity due to high mobility of human body parts, e.g. arms and legs and extract the first principal component of each joint to indicate its activity movement. Figure 5 illustrates the hands up 2 poses of the right arm after extracting the principal component.

2) PERSPECTIVE DETECTION

We record each activity at different shooting angles to resolve the ambiguity of which body joint participating in the movement. To obtain the real movement of each human joint, we convert the weak perspective camera parameters output by VIBE into the absolute coordinates in the camera coordinate system through the conversion matrix as described in 1. In equation 1, the left matrix is the transformation matrix, and the fourth column represents the weak perspective camera parameters, the middle array is VIBE coordinates and the left matrix represents to the camera coordinates. For example, hands-up two activities were recorded with three different angles, and we found the front view is the best perspective because of both arm's movements at a time.

$$\begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ 1 \end{bmatrix} = \begin{bmatrix} a+t_x \\ b+t_y \\ c+t_z \\ 1 \end{bmatrix} \tag{1}$$

3) GEOMETRY EVALUATION

We calculate the body features (body vector) from joint keypoints due to different body planes and frames perspectives. We define the human body coordinate system through the relative coordinates of VIBE output. Based on three planes of the human body in anatomical planes including the Coronal,

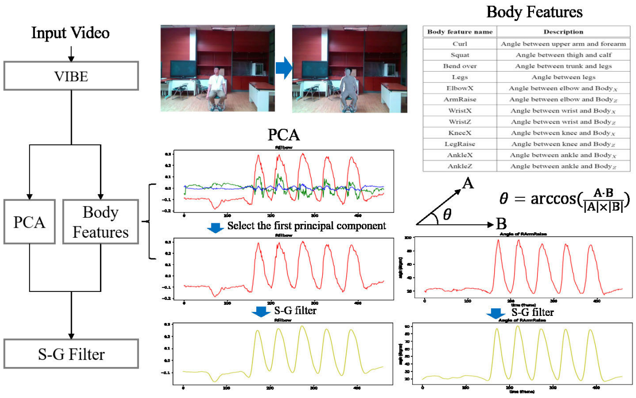

FIGURE 4. Data pre-processing flowchart.

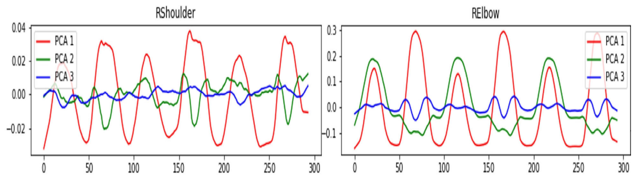

FIGURE 5. Hand up 2 key poses illustration.

Sagittal, and Transverse planes compute the three axes of the human body coordinate system (X is the human torso facing, Y is the right side of the human torso and Z is the downward side of the human torso). Based on these axes, measured the three-dimensional human body posture using the linear regression and regression plane as defined in equation 2 and the points used in each plane are illustrated in figure 3. In equation 2, line 1 sub-equation represents to linear regression where a and b are VIBE coordinates respectively. Similarly, lines 2 and 3 represent plane equation [64] and regression formula where a and b are VIBE coordinates and A is transformation matrix respectively. These equations help to find a body features which are 26 body features in total. For example, body features elbowX is represented by the angle between the elbow and X coordinate, biceps curl that can be found between the upper arm and forearm keypoints angle, etc. After the body feature identification, we smooth all body information by applying the Savitzky Golay (SG) filter.

$$\begin{aligned} y &= ax + b \\ z &= ax + by + c \\ x &= (A^T A)^{-}1A^T b \end{aligned} \tag{2}$$

B. STAGE 1: FEATURE ENGINEERING AND TIME INTERVALS

Stage 1 defines the selection procedure of the important features and filters out the time interval for each physical activity as illustrated in figure 6. The features are selected based on action peaks period and amplitude, as shown in the left side of figure 6, and the right side shows the time intervals of each exercise which is represented by the union of all intersections of the time interval set. The important features are selected based on the movement of keypoints. When an action occurs, the wave pattern will trend upward or downward resulting in a continuous wave with different amplitude levels. Generally, we observed in our action detection experiments the larger amplitude of a wave is a better representation of action. Amplitude is a key factor to decide whether the feature is important or should be filtered out as an unimportant feature. We define a reasonable amplitude threshold based on the two less movement physical activities, e.g., hip inversion and elbow external rotation, which was used to obtain displacement (D) and angular displacement (AD) threshold. These activities are evaluated based on three subjects and to determine the threshold values that are 15cm and 30 degrees as shown in table 2.

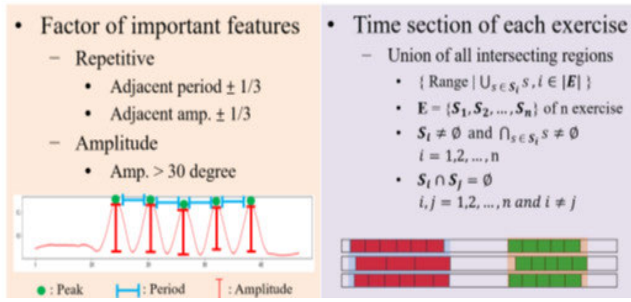

FIGURE 6. Selection criteria of feature engineering and time intervals.

TABLE 2. Amplitude threshold data sheet.

Body Information	User A	User B	User C
hips inward	D : 17.3cm AD : 35°	D : 24.2cm AD : 50°	D : 21.6cm AD : 45°
elbow external rotation	D : 15.4cm AD : 32°	d : 19.6cm AD : 42°	D : 18.1cm AD : 38°

C. STAGE 2: REPETITIVE ACTION SEGMENTATION AND COUNTING

Stage 2 describes the movement sequence and action segmentation. Movement sequence defines the physical activity into different movement states and action segmentation selects and divides the repetitive action w.r.t the movement states.

1) REPETITIVE MOVEMENT SEQUENCE

There are infinite possibilities for human physical activity, but exercise is a planned, organized, and repetitive physical activity. The movement of an exercise can be simplified as a quarter-cycle movement. Based on this concept, a movement sequence algorithm was designed with the goal to simplify and condense the body information into movement states. Movement states are used to represent the changes in the body information, namely contraction (C), relaxation (R), pause (P), and noise (N), as shown in figure 7. The body information is composed of body keypoints and features. To process the body information, we classify the input into movement blocks

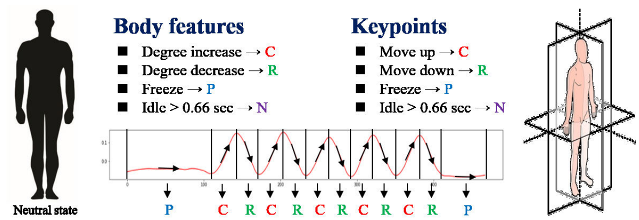

FIGURE 7. Movement sequence.

TABLE 3. Detection error angle threshold.

Body Feature	Length (cm)	Angle (degree)
Fore Arm	27	18
Upper Arm	27	18
Thigh	43	12
Calf	49	10
Average error		14.5

with continuous body movement and define the block as C, R, P, and N. The actual movement of the two adjacent blocks has different meanings and are defined individually.

In terms of body features the movement states define the pause state as when a person is standing upright and looking straight ahead (Neutral state) and the angle of all body features remains unchanged. When the angle becomes larger, it means that the movement is in the "going" stage, and this is called a contraction. Conversely, if the angle becomes smaller, it means the movement is in the "return" stage, and this is called relaxation. Similarly, we define the movement sequence in terms of joint keypoints when a person stands upright and looks straight ahead, the arms and legs hang down naturally is called a pause state. When arms and legs movement is upward, it means that the movement is in the "going" stage and called a contraction state. if the movement is downward, it means that the movement is in the "return" stage and defined as a relaxation.

In addition, 3D:VIBE pose evaluation has movement detection error which is 5.19 cm as mentioned in paper [63]. Therefore, when the keypoints movement is less than 5.19 cm, it cannot be judged which is classified as a pause. Although the detection error is 5.19 cm, so keypoints have different physical characteristics based on an error. Table 3 shows the body features evaluation based on the forearm, upper arm, thigh, and calf. The table shows that detection error causes an average of 14.5 degrees, so the body feature angle changes less than 14.5 degrees, which is defined as a pause. This evaluation uses some simple and fast actions to measure the reasonable duration of the movement state. Similarly, the reasonable duration of a movement state is 0.33 seconds as mentioned in table 4. Therefore, if the duration of a block is less than 0.33 seconds, it will be temporarily classified as idle first. If the duration of idle exceeds 0.66 seconds, it will be classified as Noise. If it is less than 0.66 seconds, it will be divided equally into the contraction and relaxation movement states.

TABLE 4. Action duration threshold in seconds.

	User A	User B	User C
Swing	0.62	0.92	0.40
Leg kick	0.83	0.38	0.53
Jumping	0.59	0.59	**0.33**
Walking	0.60	0.66	0.71

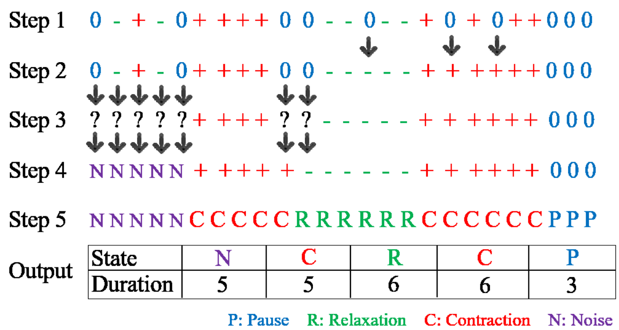

FIGURE 8. Movement sequence workflow.

2) MOVEMENT SEQUENCE ALGORITHM

The movement sequence algorithm is proposed based on the keypoints movement as described in algorithm 1. Firstly, we simplify each frame to find its movement sequence of body information by using the slope. If the slope is greater than the positive threshold value, it will be marked as "+" as shown in figure 8. If the slope is lower than the negative threshold value, it. will be marked as "−", and if the slope is between positive and negative threshold, it will be marked as "0". Secondly, we smooth the slope that does not break the threshold due to a slight pause in movement or detection error. If the previous and next value of the slope is marked with "0" the marker will be changed to a value that is consistent with the previous and the next value. Thirdly, we classify the consecutive values that do not reach the reasonable duration mentioned above as idle and mark them as "?". Fourthly, if the duration exceeds 0.66 seconds, it will be classified as noise and marked as N. If the duration does not exceed 0.66 seconds, the value of the slope is divided equally between the before and after values of the slope. Fifthly, we convert the blocks of the continuous value into the actual movement state, and finally, concentrate them at the state level.

Algorithm 1 Pseudo-Code of Movement Sequence Algorithm

Input: Body Information (BI)
Output: Movement Sequence (MS)
//Initialization
state ← convert_to_state(BI) `// Step 1`
state ← smooth(state) `// Step 2~4`
MS ← state_to_real(state) `// Step 5 & Output`

3) SINGLE ACTION SEGMENTATION

The action can be usually simplified as a half-cycle movement, and multiple half-cycle is considered a repetitive action. Repetitive action can be located by finding the pattern of repeated contractions and relaxations. The start time and the end time of each action can be known from each contraction and relaxation and are useful for the segmentation of repetitive actions. To know whether the two half-cycle movements are the same or not, we analyze the contraction-relaxation pattern by two factors including period and amplitude. In addition, if the action is more complex, there may be more than one movement state in a cycle. Two factors' descriptions are given below:

- Period: The period of the next action should be within the range of $\pm 1/3$ of the period of the current action.
- Amplitude: The amplitude of the next action must be within $\pm 1/5$ of the period of the current action.

4) SINGLE ACTION SEGMENTATION ALGORITHM

The single-action segmentation algorithm aims to identify the repetitive actions in movement sequences. Two adjacent actions are selected in the sequence and compared according to their periods and amplitudes. The principle of repetitive action selection is based on the states of movement sequence including contraction and relaxation. The single action starts from the "contraction" state and goes to the start of the next "contraction" state, and the action must contain two states including contraction and relaxation as illustrated in figure 9.

In step 1, we select two patterns from the index value 0 for comparison. The first pattern (red rectangle) starts from the pause state and continues until the next pause state and contains contraction and relaxation. The second pattern (blue rectangle) is found similarly to the first pattern, but the two adjacent sequence periods are not similar to meet the standard of repetitive action. In step 2, we select the two actions from the first pattern, the first action starts from the contraction state and ends before the start of the next contraction state and contains the contraction and relaxation states. while the second action has a similar pattern and period to meet the standard of repetitive action. Step 3, because the previous comparison was successful, the second action from the pattern is selected for comparison with the next action. If the action period is close to the adjacent action period and meets the standard of repetitive action, then we mark the action as the same action. In step 4, we do a further comparison for upcoming adjacent action and found the period that does not meet the standard. In step 5, we show the failure of the previous comparison, because the current index doesn't meet the standard criteria of action selection including contraction and relaxation and the comparison goes further until the action contains the contraction and relaxation state. Otherwise, the algorithm will interrupt as shown in step 6. The pseudo-code of single-action segmentation is described in algorithm 2.

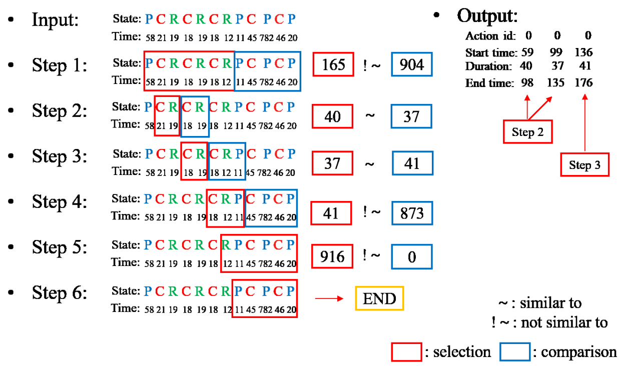

FIGURE 9. Example of single action segmentation algorithm.

Algorithm 2 Pseudo-Code of Single Action Segmentation

```
Input: Movement Sequence MS, Amplitude Amp, MS Duration MSD
Output: Repetitive Period RP //Initialization for each action i in exercise do
    pivot ← 0
    while pivot < length(MS) do
        action ← select an action from index pivot
        next_action ← select an action after action
        while action's Amp and MSD is similar to next_action do
            Append action into RP
            pivot ← action end index
            next_action ← select an action after action
        end
        else
            pivot ← pivot + 1
        end
    end
end
RP ← fix_common_miss_seg(RP)
```

D. ACTION RECOGNITION

Actions are composed of the movement of joints that controls the muscles and have different exercise effects. Therefore, we only analyze the joints that are mainly used for an activity to action recognition. In addition, the various others activity information obtained from each body features such as displacement, angular displacement, angular velocity, etc. are used to recognize each activity's action. The unimportant action information is discarded by the feature importance threshold. Finally, all information is synthesized into a vector and entered into the support vector machine (SVM) for action recognition. The SVM uses the radial basis kernel function, the loss function C is 22.5, and kernel function coefficient Gamma is 1/# of features. The same action of the previous stage that exceeds the moving window is identified by the action of the current stage. The same action of the previous phase is connected by the action recognition.

TABLE 5. Distribution of exercise types.

Exercise and Movement	#Angle	#Phases	List of Key poses
Biceps Curls	3	2	1: Standing Straight and relax arms down 2: Bend at the elbow and approach to shoulders.
Triceps Extension	3	2	1: Standing Straight, bending the elbows behind your head 2: Arms extended with palms facing the roof and elbows pointing forward.
Chest Press	3	2	1: Lie on the bench, bend elbows to 90-degree angle and feet flat on the floor. 2: Extend your arms directly over your shoulders, palms facing toward your feet.
Chest Fly	3	2	1: Start with your arms at chest level at your sides, elbows bent and pointing out. 2: Slowly exhale and lift your arms above your chest.
Shoulder Press	3	2	1: Sit on a bench, bend your elbows to shoulder height and feet flat on the floor. 2: Raise arms in full length directly over your head, and back to ear level
Seated Rowing	3	2	1: Adjust the seat and chest pad, and place feet against the foot holders. 2: Grab and pull the handles to squeeze shoulder blades together in the back.
Deadlift	3	2	1: Bend over and grab the bar, and bend knees until shins touch the bar. 2: Lift your chest up and straighten your lower back and stand up with the weight.
Pulldown	3	2	1: Sit at seat, put thighs under the leg pads, grasp the bar with arms fully extended. 2: Exhale the bar straight lower down to your chest.
Dumbbell Squat	3	2	1: Bending hips and knees and thighs are parallel to the floor. 2: Straighten knees and hips and rise back up to a standing position.
Moving Hand Squat	1	2	1: Bending hips and knees, moving hands from shoulder height to back and thighs are parallel to the floor. 2: Straighten knees, hands and hips and rise back up to a standing position.
Superman Side Squat	1	4	1: Bend body at 30 degree, bend right elbow at 45 degree and knee at 150 degree. 2: Straighten knees, hands and body and rise back up to a standing position. 3: Bend body at 30 degree, bend left elbow at 45 degree and knee at 150 degree. 4: Straighten knees, hands and body and rise back up to a standing position.
Hands Up 2	1	1	1: Raise arms in left and right to the height of shoulders. 2: Back to arms down position 3: Raise your both arms in front of body to the height of shoulders. 4: Back to arms down position
Hands Up 4	1	1	1: Raise arms in front of body to the height of shoulders. 2: Raise arms to overhead. 3: Down arms in left and right to the height of shoulders. 4: Back to arms down position
Hands Legs 4	1	1	1: Raise arms to the height of shoulders and bend elbow to 45 degree. 2: Move legs on step forward. 3: Move legs one step backward 4: Back to arms down position
Hand Circle	1	2	1: Moving right or left hand in 360 degrees. 2: Back to arms down position
Left Right Hands Up	1	2	1:Moving one hand in right hand to the height of shoulders and left at down position. 2: Moving one hand in left hand to the height of shoulders and right at down position.
Left Right Hands Up 2	1	2	1: Moving both hands in left and right direction to the height of shoulders. 2: Back to arms down position

IV. PERFORMANCE EVALUATION AND IMPLEMENTATION ENVIRONMENT

In this section, the proposed solution including the repetitive action segmentation and recognition is evaluated by modifying the accuracy method used in [65], [66], and [67]. The repetitive action segmentation is evaluated by the intersection-over-union (IoU) metric and the action recognition is evaluated by the accuracy metric.

A. DATASET

We recorded two activity datasets including NOL-18 Exercise and Timed Up and Go (TUG). NOL-18 Exercise dataset is used for the model training, validation, and testing meanwhile TUG is used for testing the model.

1) NOL-18 EXERCISE DATASET

The dataset was captured independently on multiple subjects and used for physical activity segmentation and recognition. The type of physical activities including gym exercise and movements contains 324 videos of 18 different activities. Videos are named based on the physical activity characteristics and with a resolution of 640×480 and a frame rate of 30 fps. The total length of videos is 3,173.56 seconds. Each video has at least one physical activity, and the number of activity repetition movements is inconsistent, ranging from 5 to 15, and each movement is performed by one person, with no interaction between more than one person. Each video is captured with three different backgrounds and shooting angles including front, 45 degrees, and side (90 degrees) views.

Table 5 shows the physical activities including gym exercise and movement with # shooting of Angle, # of phases, and descriptions of each activity phase. Each physical activity is labeled into two different levels. First, physical activity is labeled as a movement level where the movement sequence of each body feature is labeled to simplify and condense the information and define the movement starting time and ending time durations, as illustrated in the middle of the figure 10. Second, the action level labeling is divided into two sub-levels including feature and body level. At feature level labeling, the start time and end time of the repetitive action of each body are labeled as illustrated at the bottom

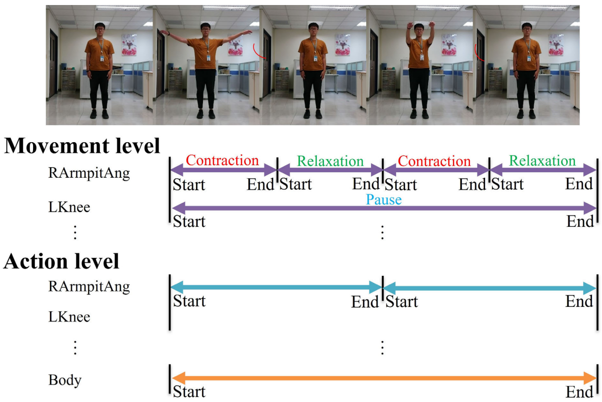

FIGURE 10. Labeling of different levels of physical activity.

TABLE 6. TUG dataset statistics.

Condition View Angle	Front	45 Degree	Side	**Total**
Normal	40	40	38	118
Medium	30	30	30	90
High	30	30	30	90
Total	100	100	98	**298**

of figure 10. Body level labeled the overall body movement actions. It segments the starting and ending time of the overall repetitive action as illustrated at the bottom of the figure 10. The NOL-18 Exercise dataset demo can be found at URL: https://hackmd.io/uc5R3jz1RMmcki-CKESzzQ.

2) TIMED UP AND GO (TUG) DATASET

TUG [68] is a method used to judge the risk of falling by analyzing the simple balance test from sitting to standing and walking. The subject sits on a chair without a backrest, and when she hears the "start" command, she gets up and walks forward at a constant speed to the target object at a distance of 3 meters, walk around the target object, walk back, and sit on a chair and camera at a fixed location and angle record the whole scenario and time. If the time is less than 15 seconds, it means that the subject has no risk of falling. If the time is between 15 seconds and 20 seconds, this indicates that the subject is at moderate risk of falling and needs to pay more attention in daily life. If the time is longer than 20 seconds indicate a high risk of falling and requires further examination to avoid danger. This dataset is self-collected for this research, including 3 shooting angles (front, 45 degrees, side) and 3 fall conditions (normal, medium, and high). There is a total of 298 videos with risks. Table 6 shows some detailed statistics for the TUG dataset.

B. EVALUATION OF REPETITIVE ACTION SEGMENTATION

The performance evaluation of repetitive action segmentation was evaluated based on the two factors that were used to predict the success rate of valid repetitive action by matching the groundtruth and predicted values.

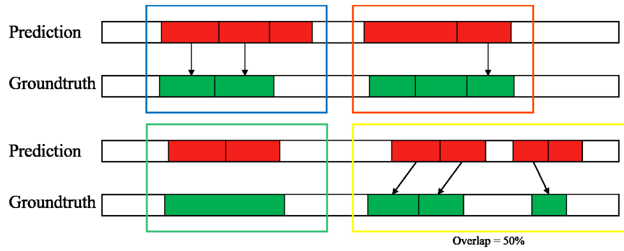

FIGURE 11. Predicted and groundtruth action matching diagram.

1) Duration: The time duration of the predicted action should be within $\pm 1/3$ of the time duration of the groundtruth action.
2) Intersection over Groundtruth (IoG): The proportion of the intersection of the predicted action and the groundtruth action must be $\geq 50\%$.

If the segmented action passes these two factors, then the action is considered valid. Figure 11 illustrates the different matching cases of groundtruth and prediction actions. Case 1: As in the blue rectangle, the starting time of predicted and groundtruth actions is the same and the ending time is different. The predicted actions overlapped with groundtruth by the one-to-one matching policy is more than 50%, which is considered a valid action. Case 2: in the green rectangle, the starting time of a predicted action and groundtruth action are the same. However, the ending time doesn't match with groundtruth, and the overlap between them is less than 50% w.r.t one-to-one that's considered an invalid action. Case 3: in the red rectangle, the starting time of the first predicted action does not match with its one-to-one groundtruth action, and that is considered an invalid action, the second predicted action matches with its one-to-one groundtruth action and ignores the second ground truth action. Case 4: in the yellow rectangle, the starting and ending times of the predicted action and ground truth do not match although the overlap ratio of both is equal to 50% and matches with each other based on the first come first served principle.

The movement sequence is evaluated using the NoL-18 dataset based on the accuracy and time error metrics. The movement sequence in algorithm 1 is evaluated by pairing sequences between the ground truth and predicted events in the time domain. The pairing sequence is evaluated based on equation 3 and 5 [65], [66], [67]. The movement sequence simplifies and condenses the information of each joint and represents the data in terms of state and time duration. The movement sequence is also evaluated by finding the accuracy of the total # of corrected movement pairs over the maximum duration number of true and predicted values. The movement sequence time error is evaluated by subtracting the absolute value of movement start time of prediction with the movement start time of true value then the movement ends time of prediction with movement end time of groundtruth value and adding them together. After addition, the total value is divided by the maximum duration number of groundtruth and predicted values. The accuracy and time error measured by equation 3 and 5 [65], [66], [67] are shows in table 7.

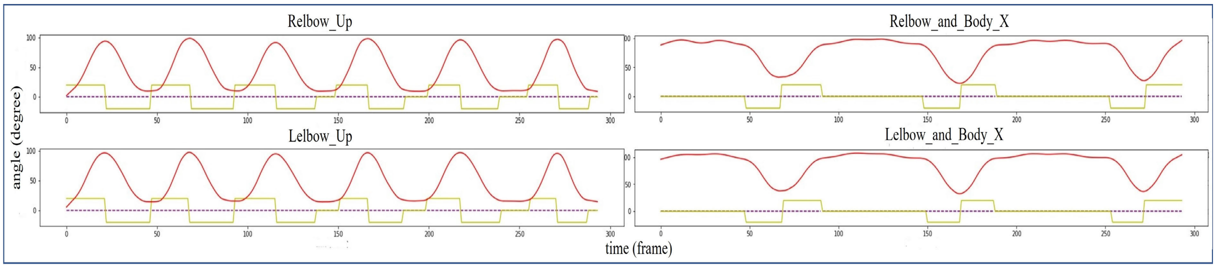

FIGURE 12. Hand up2 key poses movement.

TABLE 7. Evaluation of single repetitive activity movement sequence.

Exercise	Accuracy	Time Error RMS
Biceps Curl	99.36	0.08
Triceps Extension	92.05	0.10
Chest Press	100	0.00
Chest Fly	97.38	0.14
Shoulder Press	98.96	0.09
Seated Rowing	100	0.00
Deadlift	95.55	0.09
Pulldown	100	0.00
Dumbbell Squat	98.89	0.05
Moving Hand Squat	78.26	0.18
Superman Side Squat	84.08	0.28
Hands Up 2	100	0.00
Hands Up 4	98.38	0.12
Hands Leg 4	97.37	0.02
Hand Circle	99.38	0.06
Left Right Hands Up	97.51	0.10
Left Right Hands Up 2	91.12	0.39
Average	**95.65%**	**0.1%**

TABLE 8. Evaluation of single repetitive activity actions segmentation.

Exercise	# Reps	Accuracy	Time Error RAS
Biceps Curl	78	98.97	0.05
Triceps Extension	72	96.99	0.08
Chest Press	60	97.56	0.07
Chest Fly	48	96.95	0.67
Shoulder Press	78	96.65	0.12
Seated Rowing	36	100	0.00
Deadlift	78	94.65	0.25
Pulldown	36	100	0.00
Dumbbell Squat	84	98.45	0.06
Moving Hand Squat	48	95.65	0.50
Superman Side Squat	84	94.12	0.73
Hands Up 2	48	97.56	0.08
Hands Up 4	76	89.95	0.57
Hands Leg 4	76	95.83	0.09
Hand Circle	72	99.95	0.03
Left Right Hands Up	72	88.89	0.52
Left Right Hands Up 2	72	94.56	0.46
Average		**96.27%**	**0.23%**

The average accuracy rate is 95.65%, and the time error is 0.1 seconds.

Further, the repetitive action segmentation module is evaluated on the single activity video clips using the NoL-18 dataset based on the accuracy and time error metrics. The repetitive action segmentation is evaluated on the joint's information, which is represented by the start time and end time of the repetitive action and measured based on the corrected pairing sequence between the ground truth and predicted event in the time domain. The repetitive action segmentation is measured based on the equation 4 and 6 [65], [66], [67]. The accuracy, time error and # of each activity repetitions are shown in Table 8. The average accuracy rate is 96.27%, and the time error is 0.23 seconds.

$$Accuracy_{RMS} = \frac{\#\ of\ Correct\ Movement\ Pairs}{MAX(Num_{GT},\ Num_{Pred})} \tag{3}$$

$$Accuracy_{RAS} = \frac{\#\ of\ Correct\ Action\ Pairs}{MAX(Num_{GT},\ Num_{Pred})} \tag{4}$$

$$TimeError_{RMS} = \frac{\begin{pmatrix} \#\ of\ (abs(MST_{Pred} - MST_{GT}) \\ + abs(MET_{Pred} - MET_{GT})) \end{pmatrix}}{2 \times \#of\ Correct\ Action\ Pairs} \tag{5}$$

$$TimeError_{RAS} = \frac{\begin{pmatrix} \#\ of\ (abs(AST_{Pred} - AST_{GT}) \\ + abs(AET_{Pred} - AET_{GT})) \end{pmatrix}}{2 \times \#of\ Correct\ Action\ Pairs} \tag{6}$$

Similarly, the repetitive action segmentation module is evaluated on more than one activity video clips using the NoL-18 dataset on actions performed by the same subject. First, the subject performs one activity then with the same seconds of pause the subject performs the second activity. All these activities are also under the umbrella of the Nol-18 dataset. These activities are evaluated based on the accuracy and time error metrics. The accuracy and time error are evaluated using the equations 3, 4, 5 and 6 are shows in table 9. The average accuracy score with time error is shown in the table. The overall average accuracy of repetitive action segmentation is 94.01% with time error 0.175 seconds respectively.

The action recognition module is evaluated using the NoL-18 dataset based on the accuracy metric. Each activity is recognized by its wave characteristics using the Support vector machine (SVM) algorithm. As exercise is typically performed in a cycle and repetitive fashion, the majority of the frames tend to the inherent phases of their key poses. For example, figure 12 illustrated the key pose phases of the Hand Up 2 activity. The SVM predicts the most representative classes based on the characteristics of the phase to achieve state-of-the-art accuracy. Also, the SVM is validated using 10-fold cross-validation on the NoL-18 dataset and the average accuracy reaches 99.06% as mentioned in table 11.

The proposed method is compared with a similar repeated method as mentioned in 10. RepNet is an approach for estimating the period with which an action is repeated in a video. Similarly, we evaluate the proposed method on

TABLE 9. Evaluation of multiple repetitive activities actions segmentation.

Activity No. 1	# Reps	Activity No. 2	# Reps	Total Reps	Overall Accuracy	Time Error RAS
Triceps Extension	13	Deadlift	13	26	94.01	0.08
Moving Hand Squat	13	Left Right Hands Up legs	12	25	85.26	0.38
Biceps Curl	15	Dumbbell Squat	12	27	99.36	0.08
Deadlift	13	Dumbbell Squat	13	26	93.77	0.12
Left Right Hands Up	12	Hand Up 4	14	26	96.68	0.10
Superman Side Squat	14	Left Right Hands Up legs	11	25	89.08	0.28
Hand Circle	12	Hand Leg 4	14	26	96.37	0.10
Shoulder Press	13	Biceps Curl 4	14	27	93.61	0.15
Hand Circle	13	Hands Up 4	13	26	95.12	0.22
Triceps Extension	15	Hands Up 2	11	27	96.75	0.25
Average	**13.29%**		**12.7%**	**26.09%**	**94.01%**	**0.175%**

TABLE 10. Comparative analysis with related methods.

Model	Activity	Dataset	Accuracy
RepNet [69]	Single	NOL-18	88.32%
Ours	Single	NOL-18	**96.27%**
RepNet	Double	NOL-18	84.72%
Ours	Double	NOL-18	**94.01%**

TABLE 11. Performance evaluation.

Method	Activity	Dataset	Average Accuracy
SVM	Mixed	NOL-18	99.06%
SVM	Single	TUG	94.64%

another dataset called TUG to validate the test procedure. The TUG dataset collects the gait of weak and healthy people from multiple perspectives. The SVM learns the difference between weak and normal gait characteristics of fall risk and determines the risk of falling. The accuracy rate is evaluated, and the average accuracy rate reaches 94.64% as mentioned in table 11. Based on the evaluation, we can use this module to judge the gait characteristics of the elderly to determine whether there is a risk of falling and to receive treatment early to avoid falling.

V. CONCLUSION

This research proposed a system that provides a repetition counter and recognition of physical exercise from video frames based on the deep semantic features and repetitive segmentation algorithm. The system handles the inputs of real-time physical activity to segmentation and recognition based on different modules in the spatiotemporal domain. Our future work will include testing the proposed system in different fitness and rehabilitation center environment.

ACKNOWLEDGMENT

The work of Tsì-Uí İk was supported in part by the Ministry of Science and Technology, Taiwan under grants MOST 111-3114-H-A49-001, MOST 111-2622-E-A49-009, and MOST 110-2221-E-A49-063-MY3. This work was financially supported by the Center for Open Intelligent Connectivity from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

REFERENCES

[1] K. Wojcik, "Pattern recognition methods as a tool to build an automatic system for learning coordinated human motions," *IEEE Access*, vol. 8, pp. 141407–141421, 2020.

[2] W. Zhao, M. A. Reinthal, D. D. Espy, and X. Luo, "Rule-based human motion tracking for rehabilitation exercises: Realtime assessment, feedback, and guidance," *IEEE Access*, vol. 5, pp. 21382–21394, 2017.

[3] M. Seiffert, F. Holstein, R. Schlosser, and J. Schiller, "Next generation cooperative wearables: Generalized activity assessment computed fully distributed within a wireless body area network," *IEEE Access*, vol. 5, pp. 16793–16807, 2017.

[4] N. Alshurafa, W. Xu, J. J. Liu, M.-C. Huang, B. Mortazavi, C. K. Roberts, and M. Sarrafzadeh, "Designing a robust activity recognition framework for health and exergaming using wearable sensors," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 5, pp. 1636–1646, Sep. 2014.

[5] A. Mannini and S. S. Intille, "Classifier personalization for activity recognition using wrist accelerometers," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1585–1594, Jul. 2019.

[6] H. Monkaresi, R. A. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1153–1160, Jul. 2014.

[7] M. A. Nystoriak and A. Bhatnagar, "Cardiovascular effects and benefits of exercise," *Frontiers Cardiovascular Med.*, vol. 5, p. 135, Sep. 2018.

[8] F. J. Penedo and J. R. Dahn, "Exercise and well-being: A review of mental and physical health benefits associated with physical activity," *Current Opinion Psychiatry*, vol. 18, no. 2, pp. 189–193, 2005.

[9] R. A. Hamad, A. G. Salguero, M.-R. Bouguelia, M. Espinilla, and J. Medina-Quero, "Efficient activity recognition in smart homes using delayed fuzzy temporal windows on binary sensors," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 387–395, Feb. 2020.

[10] M. Zhang and A. A. Sawchuk, "Human daily activity recognition with sparse representation using wearable sensors," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 3, pp. 553–560, May 2013.

[11] S. R. Vanrell, D. H. Milone, and H. L. Rufiner, "Assessment of homomorphic analysis for human activity recognition from acceleration signals," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1001–1010, Jul. 2018.

[12] A. Soro, G. Brunner, S. Tanner, and R. Wattenhofer, "Recognition and repetition counting for complex physical exercises with deep learning," *Sensors*, vol. 19, no. 3, p. 714, Feb. 2019.

[13] B. Ferreira, P. M. Ferreira, G. Pinheiro, N. Figueiredo, F. Carvalho, P. Menezes, and J. Batista, "Deep learning approaches for workout repetition counting and validation," *Pattern Recognit. Lett.*, vol. 151, pp. 259–266, Nov. 2021.

[14] B. Ferreira, P. M. Ferreira, G. Pinheiro, N. Figueiredo, F. Carvalho, P. Menezes, and J. Batista, "Exploring workout repetition counting and validation through deep learning," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2020, pp. 3–15.

[15] T. Lan, Y. Zhu, A. R. Zamir, and S. Savarese, "Action recognition by hierarchical mid-level action elements," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4552–4560.

[16] A. Hussain, T. Hussain, W. Ullah, and S. W. Baik, "Vision transformer and deep sequence learning for human activity recognition in surveillance videos," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–10, Apr. 2022.

[17] A. Hussain, Z. A. Khan, T. Hussain, F. U. M. Ullah, S. Rho, and S. W. Baik, "A hybrid deep learning-based network for photovoltaic power forecasting," *Complexity*, vol. 2022, pp. 1–12, Oct. 2022.
[18] A. Hussain, K. Muhammad, H. Ullah, A. Ullah, A. S. Imran, M. Y. Lee, S. Rho, and M. Sajjad, "Anomaly based camera prioritization in large scale surveillance networks," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 2171–2190, 2022.
[19] S. Mekruksavanich, A. Jitpattanakul, K. Sitthithakerngkiet, P. Youplao, and P. Yupapin, "ResNet-SE: Channel attention-based deep residual network for complex activity recognition using wrist-Worn wearable sensors," *IEEE Access*, vol. 10, pp. 51142–51154, 2022.
[20] F. Lin, A. Wang, L. Cavuoto, and W. Xu, "Toward unobtrusive patient handling activity recognition for injury reduction among at-risk caregivers," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 3, pp. 682–695, May 2017.
[21] W. Tang and E. S. Sazonov, "Highly accurate recognition of human postures and activities through classification with rejection," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 1, pp. 309–315, Jan. 2014.
[22] H. Bi, M. Perello-Nieto, R. Santos-Rodriguez, and P. Flach, "Human activity recognition based on dynamic active learning," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 922–934, Apr. 2021.
[23] A. Dalton and G. O. Laighin, "Comparing supervised learning techniques on the task of physical activity recognition," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 1, pp. 46–52, Jan. 2013.
[24] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Complex human activity recognition using smartphone and wrist-Worn motion sensors," *Sensors*, vol. 16, no. 4, p. 426, 2016.
[25] C. Shen, B.-J. Ho, and M. Srivastava, "MiLift: Efficient smartwatch-based workout tracking using automatic segmentation," *IEEE Trans. Mobile Comput.*, vol. 17, no. 7, pp. 1609–1622, Jul. 2018.
[26] C. Seeger, A. P. Buchmann, and K. Van Laerhoven, "myHealthAssistant: A phone-based body sensor network that captures the wearer's exercises throughout the day," in *Proc. 6th Int. ICST Conf. Body Area Netw.*, 2011, pp. 1–7.
[27] K.-H. Chang, M. Y. Chen, and J. Canny, "Tracking free-weight exercises," in *Proc. Int. Conf. Ubiquitous Comput.* Cham, Switzerland: Springer, 2007, pp. 19–37.
[28] M. Muehlbauer, G. Bahle, and P. Lukowicz, "What can an arm holster Worn smart phone do for activity recognition?" in *Proc. 15th Annu. Int. Symp. Wearable Comput.*, Jun. 2011, pp. 79–82.
[29] D. Morris, T. S. Saponas, A. Guillory, and I. Kelner, "RecoFit: Using a wearable sensor to find, recognize, and count repetitive exercises," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2014, pp. 3225–3234.
[30] *Wear Os by Google*. Accessed: Nov. 19, 2021. [Online]. Available: https://wearos.google.com/#uniquely-you
[31] *Apple Watch Series 4*. Accessed: Nov. 19, 2021. [Online]. Available: https://www.apple.com/lae/watch/
[32] *Moov Fitness System*. Accessed: Nov. 17, 2021. [Online]. Available: https://welcome.moov.cc/
[33] J. Y. Xu, Y. Wang, M. Barrett, B. Dobkin, G. J. Pottie, and W. J. Kaiser, "Personalized multilayer daily life profiling through context enabled activity classification and motion reconstruction: An integrated system approach," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 177–188, Jan. 2016.
[34] J. Huang, S. Lin, N. Wang, G. Dai, Y. Xie, and J. Zhou, "TSE-CNN: A two-stage end-to-end CNN for human activity recognition," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 292–299, Jan. 2020.
[35] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-Worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, pp. 1–33, 2014.
[36] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, 2015.
[37] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao, "FEMO: A platform for free-weight exercise monitoring with RFIDs," in *Proc. 13th ACM Conf. Embedded Networked Sensor Syst.*, Nov. 2015, pp. 141–154.
[38] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.
[39] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell, "SoundSense: Scalable sound sensing for people-centric applications on mobile phones," in *Proc. 7th Int. Conf. Mobile Syst., Appl., Services*, Jun. 2009, pp. 165–178.
[40] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1473–1488, Nov. 2008.
[41] R. Poppe, "A survey on vision-based human action recognition," *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, 2010.
[42] H. Chung, H. Ko, H. Lee, and J. Lee, "Deep learning for heart rate estimation from reflectance photoplethysmography with acceleration power spectrum and acceleration intensity," *IEEE Access*, vol. 8, pp. 63390–63402, 2020.
[43] J. Zhang, S. Liu, and R. Zhu, "Motion velocity, acceleration and energy expenditure estimation using micro flow sensor," *IEEE Access*, vol. 7, pp. 75901–75909, 2019.
[44] F. Delmastro, F. D. Martino, and C. Dolciotti, "Cognitive training and stress detection in MCI frail older people through wearable sensors and machine learning," *IEEE Access*, vol. 8, pp. 65573–65590, 2020.
[45] F. Mokaya, R. Lucas, H. Y. Noh, and P. Zhang, "MyoVibe: Vibration based wearable muscle activation detection in high mobility exercises," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2015, pp. 27–38.
[46] H.-T. Cheng, F.-T. Sun, M. Griss, P. Davis, J. Li, and D. You, "NuActiv: Recognizing unseen new activities using semantic attribute-based learning," in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2013, pp. 361–374.
[47] M. A. O'Reilly, D. F. Whelan, T. E. Ward, E. Delahunt, and B. Caulfield, "Technology in strength and conditioning tracking lower-limb exercises with wearable sensors," *J. Strength Conditioning Res.*, vol. 31, no. 6, pp. 1726–1736, 2017.
[48] A. Ebert, M. Kiermeier, C. Marouane, and C. Linnhoff-Popien, "SensX: About sensing and assessment of complex human motion," in *Proc. IEEE 14th Int. Conf. Netw., Sens. Control (ICNSC)*, May 2017, pp. 327–332.
[49] D. M. Burns, N. Leung, M. Hardisty, C. M. Whyne, P. Henry, and S. McLachlin, "Shoulder physiotherapy exercise recognition: Machine learning the inertial signals from a smartwatch," *Physiological Meas.*, vol. 39, no. 7, Jul. 2018, Art. no. 075007.
[50] A.-T. Chiang, Q. Chen, S. Li, Y. Wang, and M. Fu, "Denoising of joint tracking data by Kinect sensors using clustered Gaussian process regression," in *Proc. 2nd Int. Workshop Multimedia Pers. Health Health Care*, Oct. 2017, pp. 19–25.
[51] A.-T. Chiang, Q. Chen, Y. Wang, and M. R. Fu, "Kinect-based in-home exercise system for lymphatic health and lymphedema intervention," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–13, 2018.
[52] W. Zhao, H. Feng, R. Lun, D. D. Espy, and M. A. Reinthal, "A Kinect-based rehabilitation exercise monitoring and guidance system," in *Proc. IEEE 5th Int. Conf. Softw. Eng. Service Sci.*, Jun. 2014, pp. 762–765.
[53] D. González-Ortega, F. J. Díaz-Pernas, M. Martínez-Zarzuela, and M. Antón-Rodríguez, "A Kinect-based system for cognitive rehabilitation exercises monitoring," *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 620–631, 2014.
[54] K. Vijayaprabakaran, K. Sathiyamurthy, and M. Ponniamma, "Video-based human activity recognition for elderly using convolutional neural network," *Int. J. Secur. Privacy Pervasive Comput.*, vol. 12, no. 1, pp. 36–48, Jan. 2020.
[55] J. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 3995–4001.
[56] R. Khurana, K. Ahuja, Z. Yu, J. Mankoff, C. Harrison, and M. Goel, "GymCam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 1–17, Dec. 2018.
[57] S. Chen and R. R. Yang, "Pose trainer: Correcting exercise posture using pose estimation," 2020, *arXiv:2006.11718*.
[58] O. Levy and L. Wolf, "Live repetition counting," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3020–3028.
[59] T. F. H. Runia, C. G. M. Snoek, and A. W. M. Smeulders, "Real-world repetition estimation by div, grad and curl," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9009–9017.
[60] P. Ghadekar, S. Bijawe, P. Akolkar, H. Pandey, M. Mahajani, and A. Shinde, "Real-time virtual fitness tracker and exercise posture correction," *Int. J. Next-Gener. Comput.*, vol. 12, Nov. 2021.
[61] L. Brennan, E. D. Zubiete, and B. Caulfield, "Feedback design in targeted exercise digital biofeedback systems for home rehabilitation: A scoping review," *Sensors*, vol. 20, no. 1, p. 181, Dec. 2019.

[62] G. Prabhu, N. E. O'Connor, and K. Moran, "Recognition and repetition counting for local muscular endurance exercises in exercise-based rehabilitation: A comparative study using artificial intelligence models," *Sensors*, vol. 20, no. 17, p. 4791, Aug. 2020.

[63] M. Kocabas, N. Athanasiou, and M. J. Black, "VIBE: Video inference for human body pose and shape estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5253–5263.

[64] *Coordinate System*. Accessed: Aug. 24, 2022. [Online]. Available: https://www.mathworks.com/help/aerotbx/ug/coordinate-systems-for-modeling.html

[65] H.-C. Chi, M. A. Sarwar, Y.-A. Daraghmi, K.-W. Lin, T.-U. Ik, and Y.-L. Li, "Smart self-checkout carts based on deep learning for shopping activity recognition," in *Proc. 21st Asia–Pacific Netw. Oper. Manag. Symp. (APNOMS)*, Sep. 2020, pp. 185–190.

[66] M. A. Sarwar, Y. A. Daraghmi, K. W. Liu, H. C. Chi, T. U. Ik, and Y. L. Li, "Smart shopping carts based on mobile computing and deep learning cloud services," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–6.

[67] M.-H. Lin, M. A. Sarwar, Y.-A. Daraghmi, and T.-U. Ik, "On-shelf load cell calibration for positioning and weighing assisted by activity detection: Smart store scenario," *IEEE Sensors J.*, vol. 22, no. 4, pp. 3455–3463, Feb. 2022.

[68] G. Sprint, D. J. Cook, and D. L. Weeks, "Toward automating clinical assessments: A survey of the timed up and go," *IEEE Rev. Biomed. Eng.*, vol. 8, pp. 64–77, 2015.

[69] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet, and A. Zisserman, "Counting out time: Class agnostic video repetition counting in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10387–10396.

SHENG-HSIEN CHENG is currently pursuing the master's degree with National Yang Ming Chiao Tung University, Taiwan. His research interests include artificial intelligence, deep learning, and computer vision.

MUHAMMAD ATIF SARWAR received the B.S. and M.S. degrees in computer science from COMSATS University Islamabad, Sahiwal Campus, Pakistan, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree with the EECS International Graduate Program, National Yang Ming Chiao Tung University, Taiwan. His research interests include artificial intelligence, deep learning, and computer vision. His current research to detect activity recognition and actions in a retailers store, sports, and exercise.

YOUSEF-AWWAD DARAGHMI received the B.E. degree in electrical and computer engineering from An-Najah National University, in 2002, and master's and Ph.D. degrees in computer science and engineering from National Chiao Tung University, Taiwan, in 2007 and 2014, respectively. He is currently an Associate Professor with the Computer Systems Engineering Department, Palestine Technical University-Kadoorie. His research interests include intelligent transportation systems, vehicular ad hoc networks, and blockchain. He served as a Technical Program Committee Member for the International Conference on Connected Vehicles and Expo (ICCVE), from 2012 to 2016, the International Conference on Intelligent Transportation Systems Telecommunications (ITST), from 2012 to 2018, the International Conference on Signal Processing (ICOSP), in 2015 and 2016, and the Asia Pacific Network Operation and Management Symposium (APNOMS), in 2015 and 2016. He received the Best Paper Award from the International Conference on Intelligent Transportation Systems Telecommunications, in 2012. He is a Reviewer of some highly distinguished journals, including IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, *IEEE Communications Magazine*, and *IEEE Network* magazine.

TSÌ-UÍ İK (Member, IEEE) received the B.S. degree in mathematics and M.S. degree in computer science and information engineering from National Taiwan University, in 1991 and 1993, respectively, and the Ph.D. degree in computer science from the Illinois Institute of Technology, in 2005.

He was a Senior Research Fellow at the Department of Computer Science, City University of Hong Kong. He is currently a Professor with the Department of Computer Science and the Director of the Institute of Computer Science and Engineering, National Yang Ming Chiao Tung University. His research interests include on intelligent applications, such as intelligent sports learning and intelligent transportation systems, mobile sensing, machine learning, deep learning, and wireless sensor and ad hoc networks.

Dr. İk was a recipient of the Outstanding Young Engineer Award by the Chinese Institute of Engineers, in 2009, and the Young Scholar Best Paper Award by IEEE IT/COMSOC Taipei/Tainan Chapter, in 2010. He received the Best Paper Award from ITST 2012. He received a Three-year Outstanding Young Researcher Grant from the National Science Council, Taiwan, in 2012. In 2020, he received the Sports Science Research and Development Award, MoE, Taiwan. In 2020 and 2021, his research works received the MOST Future Tech Award.

YIH-LANG LI (Member, IEEE) received the B.S. degree in nuclear engineering and the M.S. and Ph.D. degrees in computer science, majoring in designing and implementing a highly parallel cellular automata machine for fault simulation from National Tsing Hua University, Hsinchu, Taiwan. In 2003, he joined the Faculty of the Department of Computer Science, National Chiao Tung University (NCTU), Hsinchu, where he is currently a Professor. From 1995 to 1996 and from 1998 to 2003, he was a Software Engineer and an Associate Manager with Springsoft Corporation, Hsinchu, where he first completed the development of design rule checking (DRC) tool for the custom-based layout design and then established and led a routing team for developing a block-level shape-based router for the custom-based layout design. His current research interests include physical synthesis, parallel architecture, vehicle navigation, and deep learning. He joined the Technical Committee of the first CAD Contest in Taiwan and served as a Committee Member for ten years. He has also been serving as the Compensation Committee Member and the Independent Director of Board of Directors for AMICCOM Electronics Corporation, since 2012. He was a recipient of the Japan Society for the Promotion of Science Faculty Invitation Fellowship. He was the Contest Chair of the first CAD Contest at ICCAD, in 2012, and the Technical Program Committee Member for ASPDAC and DAC.