Viewpoint-invariant exercise repetition counting

Yu Cheng Hsu1,2*, Tsougenis Efstratios2 and Kwok-leung Tsui1,3*

Abstract

Counting the repetition of human exercise and physical rehabilitation is common in rehabilitation and exercise training. The existing vision-based repetition counting methods less emphasize the concurrent motions in the same video, and counting skeleton in different view angles. This work analyzed the spectrogram of the pose estimation cosine similarity to count the repetition. Besides the public datasets. This work also collected exercise videos from 11 adults to verify that the proposed method can handle concurrent motion and different view angles. The presented method was validated on the University of Idaho Physical Rehabilitation Movements Data Set (UI-PRMD) and MM-fit dataset. The overall mean absolute error (MAE) for MM-fit was 0.06 with off-by-one Accuracy (OBOA) of 0.94. As for the UI-PRMD dataset, MAE was 0.06 with OBOA 0.95. We have also tested the performance in various camera locations and concurrent motions with 57 skeleton time-series videos with an overall MAE of 0.07 and OBOA of 0.91. The proposed method provides a view-angle and motion agnostic concurrent motion counting. This method can potentially use in large-scale remote rehabilitation and exercise training with only one camera.

Keywords: Camera, Repetition counting, Exercise

Introduction

Repetitive exercise and training are omnipresent in daily life and are especially common in sport and rehabilitation training. Early research [1] has surveyed that disobedience to the exercise program is a factor that deteriorate the training outcome. Some researchers [2, 3] believe that lack of motivation and feedback is why the users do not comply with the designated exercise plan. Engaging users by providing more feedback and motivating results could increase the adherence of the users [4, 5], and consequently, enhance the training outcome.

As technology advances, training and rehabilitation programs might not necessarily occur in primary care. The concept of the home exercise program (HEP) and remote rehabilitation with information technology assistance draws researchers’ and medical professionals’ attention. This approach relieves the medical area’s burden and enables flexibility for people who encounter difficulty getting medical resources. Providing repetition counting feedback is common feedback in the HEP. Previous works [6–9] focused on HEP systems also implemented several feedbacks, such as repetition counting. Users can be more engaged in the exercise and training program given the counting feedback. Users could know how many repetitions have been done and their speed for completing a cycle.

Still, counting exercise repetition is a less studied than the prosperous-growing field in human activity recognition. Dwibedi et al. [10] pointed out that searching for suitable video from large-scale public datasets is strenuous. No specific keyword and annotations cater to this particular need. The root cause of this phenomenon is that annotating video or signal across the temporal domain is labor-intensive and monotonous work. Due to these reasons, the repetition counting task usually encounters the lack of sufficient annotated data.

This research aims to solve a counting on concurrent human exercise motions. This research provides a novel repetition counting method using skeleton data based on any camera, including the 3D depth camera. The
proposed method first calculates the self-similarity of the sequence of skeleton data generated from pose estimation methods. Subsequently, we analyzed the spectrogram of the self-similarity and estimated the count.

The primary contribution of this work is that it offers (1) view-angle and motion agnostic vision-based counting methods. This work takes the skeleton data from the camera and counts through the temporal similarity. (2) Counting multiple people concurrent repetition in the same frame. The proposed method takes the skeleton data as the input. Current pose estimation methods can estimate multiple people in a frame, so the proposed method could leverage these advantages and infer the repetition for each person with different repetition frequencies in the same video. The current state of the art [10, 11] usually has the limitation of counting only one motion in the videos. In contrast, counting concurrent movement becomes possible by counting skeleton data. This work has strong application in group exercise training and human factors work-study.

The current research works can be categorized using these two aspects, the feature used for counting and counting methods, listed in Table 1. In terms of features used for counting, the current study can be categorized into 1. Crafted features, 2. Self-similarity features. The other aspect is the counting method being implemented. Current counting methods can be categorized into three approaches 1. Peak detection, 2. Frequency transformation and 3. Neural network methods. Recent research works can be classified as a combination of these two aspects (Fig. 1).

In the feature perspectives, only pose estimation features can identify several different human movements in the video and can estimate multiple periodic movements in the videos. Current pose estimation [17–22] will output a connected graph of the human joints. Figure 1 Example of pose estimation output. Previous studies [13, 14] leverage these ideas to count exercise repetitions, but this work uses handcrafted features for different exercises. This disadvantage makes this approach hard to generalize all periodical movements. Self-similarity features for counting were only used in previous research [10]. However, the self-similarity features are popular and proved to be view-angle agnostic in human motion recognition fields [23–25]. Dwibedi et al. [10] study cannot count people exercising in the same video together as it does not consider such a situation.

From the counting method perspective, most methods adopt either frequency method or peak counting methods. Peak detection methods are intuitive but subject to the noise in the signal. Frequency domain counting methods are more robust to the signal noise and with fewer tuning parameters. In addition, Dwibedi et al. [10] work using neural network methods from the self-similarity features to count the repetition has drawn tremendous success and opened another approach to solve this challenge.

The proposed algorithm is a self-similarity feature with a frequency-based counting method. From the previous literature, we can observe that this method has the advantages of fewer tuning parameters and view-angle agnostics.

**Method**

**Data source**

Our method was trained and tested using two public datasets, the UI-PRMD dataset [26] and MM-fit [14]. These datasets give each video a label indicating how many repetitions are in the video. In addition to these two datasets, an additional dataset was collected to verify the claim of viewpoint-invariant and concurrent motion.

**UI-PRMD dataset**

UI-PRMD dataset [26] is composed of ten motions commonly used in rehabilitation, including (1) deep squat, (2) hurdle step, (3) inline lunge, (4) side lunge, (5) sit to stand, (6) standing active straight leg raise, (7) standing shoulder abduction, (8) standing shoulder extension, (9) standing shoulder internal-external rotation, and

| Table 1 Counting method used in the previous research, method using pose estimation is in bold |
|----------------------------------------------------------|
| Crafted features | Self-similarity features |
|-----------------|-------------------------|
| Peak detection  | [12–14]                 | NA                      |
| Frequency domain counting | [15, 16]          | This study              |
| Neural network  | NA                      | [10]                    |
standing shoulder scaption. Ten healthy subjects repeated each exercise 10 times. A Kinect v2 camera was put in front of the subject during the data collection. It is worth pointing out that subjects were allowed to use their dominant side to perform the tasks, so some might use the right side, and some may use the left side to complete tasks 2, 3, 4, 6, 7, 8, 9, and 10.

**MM-fit dataset**

The MM-fit dataset [14] consists of ten types of commonly used workouts for home training (1) squats, (2) push-ups, (3) shoulder press, (4) lunges, (5) dumbbell rows, (6) sit-ups, (7) triceps extensions, (8) biceps curls, (9) lateral raises, (10) and jumping jacks. There are ten participants in the dataset, and each participant was asked to perform a set of exercises consisting of several exercises with ten repetitions. An RGB depth camera was put in front of the participant during the training. This dataset adopted OpenPose for 2D pose estimation and the method developed by Martina et al. [20] for 3D pose estimation.

**Additional dataset**

We acknowledged that both public datasets fixed the camera location and did not test concurrent motions in the same videos. 11 adults were recruited and asked to squat, lateral shoulder raises, and lunge for ten repetitions with different camera angles or concurrently to demonstrate the proposed idea work in the various scenarios. Total 57 skeleton time series were extracted from the videos, and the detailed decomposition was listed in Table 2. Videos were recorded by iPad Pro 5th generation with a frame rate of 30 Hz, and the pose estimation was done by OpenPose [21]. Please refer to the supplementary animated GIF picture for a typical example of identifying the motions of two people. The counting label for each label was labeled by two student annotators, and the kappa agreement is 1.00 between two student annotators. This research was approved by the Human Subjects Ethics Sub-Committee, City University of Hong Kong (Ref. 3-2-201803_02).

**Outline of the proposed method**

The proposed method relied on the existing pose estimation algorithms to extract the skeleton location of each frame. Though the proposed method relied on the pose estimation, it did not rely on a specific type of skeleton format. Consequently, 3D pose skeleton formats like Kinect or 2D skeleton in Openpose are all valid input for this method.

After obtaining skeleton data, this work proposed a new perspective of processing the skeleton graph time series to construct the temporal self-similarity.

$\text{self cosine similarity} = \cos(\theta) = \frac{X_{t_0} \cdot X_{t_i}}{||X_{t_0}|| ||X_{t_i}||}$ (1)

Figure 2 displays the output similarity corresponding to time 0.

**Constructing spectrogram**

Observing the strong periodicity, analyzing the frequency patterns, and counting through the frequency
domain patterns should be reasonable. We started to investigate the spectrogram of cosine similarity for $t = 0$ with respect to the rest of the time point $i, \forall i \leq t$. The sequence is demonstrated in Fig. 2. The spectrogram was done by calculating the fast Fourier transform (FFT) on a fixed sliding window with window size $w$. For signals $x$ in the sliding window segment, the FFT response:

$$\text{FFT}(x)_k = \sum_{n=0}^{w-1} x(n) \exp\left(-\frac{i2\pi n}{w}\right), k = 0, \ldots, w - 1$$

Figure 3 illustrates the spectrogram of the self-cosine similarity in Fig. 2. The yellow line in Fig. 3 revealed a strong signal at the frequency of around 0.25 along with the whole exercise.

**Counting out of the time**

We believe the most dominant frequency in the spectrogram is the frequency corresponding to the repeating motion. The local frequency $f_m$ in the sliding window $m$ is the frequency with the highest amplitude in such a sliding window.

$$f_m = \arg\max_k \| \text{FFT}(x)_k \|$$

The estimation of the repetition counts $\hat{c}$ was defined as the integral of local frequency over time, where $\delta t$ is the time length in that FFT window.

$$\hat{c} = \sum_{t=t_0}^{t_f} f_t \delta t$$

Runia et al. [15] adopted the same counting method, which emphasized counting non-stationary signals.

**Consideration of hyperparameters**

The only hyperparameter pair in this method was the FFT spectrogram’s sliding window size and sliding steps. There are two factors to take into consideration. First, the sliding window should be wide enough to envelop at least a cycle of repetition. Therefore, the lower bound of the window size should be the frame rate times the maximum length of a cycle. This lower bound is at 90 in the public dataset used in this study. Besides, the wider the window, the better the frequency resolution. As a trade-off, wider windows and more giant sliding window steps resulted in lower time resolution. Both public datasets’ skeleton sequences are around 400 to 800 frames, so we choose the sliding window size 128, 256 for testing. We have evaluated and discussed the different combinations of sliding window sizes and steps in the result and discussion.

**Model evaluation**

There are two evaluation metrics commonly used [10, 11, 15] in repetition counting.

**Mean absolute error (MAE) of the count**

MAE of the count is the average normalized absolute difference between the predicted count on the i-th video $\hat{c}_i$ and the ground truth $c_i$ of the i-th video. This metric was used in the previous research works as the metric can be interpreted as the percentage of the counting difference compared with the ground truth.

$$\text{MAE of the count} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{c}_i - c_i|}{c_i}$$

**Off-by-one accuracy (OBOA)**

OBOA first labels the video is correctly classified if the absolute difference between the predicted count and the ground truth is less or equal to one. Otherwise, it marks the prediction as misclassified. The OBOA is the
accuracy of using such a definition. OBOA provides a brief idea of how accurate the algorithm is and favors a precise algorithm with a tolerance of a few extremely miscounting cases.

\[
\text{OBOA} = \frac{1}{n} \sum_{i=1}^{n} n_i, \quad n_i = \begin{cases} 1 & \text{if } |\hat{c}_i - c_i| \leq 1 \\ 0 & \text{else} \end{cases}
\]

**Results**

Overall, the proposed method can reach 0.94 to 0.95 OBOA with MAE 0.06 with the best hyperparameter combinations.

**MM-fit dataset**

Table 3 indicated the MAE and OBOA for the proposed method and the baseline demonstrated in [14]. The MAE and OBO error was reported with the same testing scheme as Stromback et al. [14]. The proposed method yielded the same or significantly better in each task. In addition to the baseline method, the proposed method performs more stably across different types of motion. In contrast, the baseline method varies significantly among different kinds of exercise. This might be due to the baseline method requiring adjusting the hyperparameter specific to the motion.

The MM-fit dataset also provided 2D pose estimation data using Openpose [21]. We have listed the performance of the MM-fit dataset using 2D pose estimation format as the input in Table 4.

**UI-PRMD dataset**

Table 5 indicates the MAE and OBOA for the proposed method testing on the UI-PRMD dataset. All the data were used to report the result. To the best of our knowledge, there is no research work adopted this dataset for motion counting.

**Additional dataset**

Table 6 was obtained using 256 window sizes with step 1. Table 6 indicated that the proposed method was able to count with different view angles and concurrent motions. The boxplots (Fig. 4) of the MAE suggest that at least 75% of the pose estimation time series were miscounted within one repetition for any situation. Only a few video clips showed extremely incorrect estimations.

**Discussion**

**Effect of the hyperparameters**

Tables 3 and 5 have demonstrated the effect of hyperparameters. We could observe that a window size of 256 was more desirable than 128 (MANOVA, \( p < 0.001 \)). Window size with length 256 performed better than window size with 128 in overall MAE and OBOA. The only exception was jumping jacks in the MM-fit dataset. We suspected the primary reason was that performing jumping jacks took significantly less time than the rest of the exercise (student t-test, \( p < 0.001 \)). The average frame length of jumping jacks was 309 frames with a standard deviation of 51, and in contrast, the rest of the exercise took 637 frames with a standard deviation of 163. Consequently, a window with 256 frames might not provide

**Table 3 MAE of count and OBOA for motion counting MM-fit dataset, squats (Sq), push-ups (Pu), dumbbell shoulder press (Sp), lunges (Lg), dumbbell rows (Dr), situps (Su), tricep extensions (Te), bicep curls (Bc), lateral shoulder raises (Sr), and jumping jacks (Jj)**

| Windows steps | 128 | 256 |
|---------------|-----|-----|
| 1             |     |     |
| 2             |     |     |
| 4             |     |     |
| 8             |     |     |
| 16            |     |     |

| Windows steps | 128 | 256 |
|---------------|-----|-----|
| 1             |     |     |
| 2             |     |     |
| 4             |     |     |
| 8             |     |     |
| 16            |     |     |

| MAE for each motion | 0.10 | 0.08 | 0.08 | 0.08 | 0.08 | 0.06 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |

| Overall MAE       | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 |

| OBOA              | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.94 | 0.94 | 0.93 | 0.92 | 0.91 | 0.90 | 0.90 | 0.90 | 0.90 |
sufficient time resolution to estimate the frequency accurately. For the rest of the exercise, 256 window size offered a higher frequency resolution than 128 windows to count the repetition more accurately than 128. For the same reason, a sliding window with size 512 is not suitable for all of the exercises in the public datasets.

Sliding window step size does not affect the repetition counting result (MANOVA, $p = 0.305$). Table 3 showed that MAE increased and OBOA decreased while increasing the step size in the MM-fit dataset. Interestingly, this effect cannot be observed in the UI-PRMD dataset (Table 5). In any case, the result shown in Tables 3 and 5 indicated that sliding window steps made a minor difference in repetition counting.

**Robustness of the proposed method**
The robustness of our proposed method has been tested. The robustness is vital for the skeleton-based method as

**Table 4 MAE of count for motion counting MM-fit dataset (2D), squats (Sq), push-ups (Pu), dumbbell shoulder press (Sp), lunges (Lg), dumbbell rows (Dr), situps (Su), tricep extensions (Te), bicep curls (Bc), lateral shoulder raises (Sr), and jumping jacks (Jj)**

| Windows steps | 128 | 256 |
|---------------|-----|-----|
| 1             |     |     |
| 2             |     |     |
| 4             |     |     |
| 8             |     |     |
| 16            |     |     |

**MAE for each motion**

|                | 128 | 256 |
|----------------|-----|-----|
| Sq             | 0.11| 0.06|
| Pu             | 0.10| 0.13|
| Sp             | 0.10| 0.04|
| Lg             | 0.10| 0.04|
| Dr             | 0.07| 0.06|
| Su             | 0.22| 0.17|
| Te             | 0.07| 0.06|
| Bc             | 0.07| 0.05|
| Sr             | 0.08| 0.04|
| Jj             | 0.04| 0.13|

**Overall**

|                | 128 | 256 |
|----------------|-----|-----|
| MAE            | 0.09| 0.08|
| OBOA           | 0.87| 0.94|

**Table 5 MAE of count and OBOA for motion counting UI-PRMD dataset, deep squat (Dq), hurdle step (Hs), inline lunge (Il), side lunge (Sl), sit to stand (Ss), leg raise (Lr), shoulder abduction (Sa), shoulder extension (Se), shoulder internal-external rotation (Sr), and shoulder scaption (Sc)**

| Windows steps | 128 | 256 |
|---------------|-----|-----|
| 1             |     |     |
| 2             |     |     |
| 4             |     |     |
| 8             |     |     |
| 16            |     |     |

**MAE for each motion**

|                | 128 | 256 |
|----------------|-----|-----|
| Dq             | 0.07| 0.05|
| Hs             | 0.10| 0.01|
| Il             | 0.10| 0.04|
| Sl             | 0.14| 0.07|
| Ss             | 0.12| 0.02|
| Lr             | 0.15| 0.11|
| Sa             | 0.08| 0.05|
| Se             | 0.11| 0.12|
| Sr             | 0.08| 0.08|
| Sc             | 0.05| 0.06|

**Overall**

|                | 128 | 256 |
|----------------|-----|-----|
| MAE            | 0.10| 0.06|
| OBOA           | 0.81| 0.94|
pose estimation performance is subject to occlusion from the environments and the view-angle. Consequently, noisy pose estimation or missing frames were expected in the real-world application: two scenarios, 1. A zero-mean Gaussian noise with different standard deviation levels and 2. missing data were added to the public datasets to test the robustness of the proposed method. The standard deviation of the Gaussian noise was set on different percentages of the corresponding axis and joint. The missing data was simulated by replacing the frames with interpolation from the nearest frames. Figures 5 and 6 showed how these scenarios affected the proposed method.

Figure 5 shows the performance of the proposed method while Gaussian noises were present. The Gaussian noise follows a zero-mean Gaussian distribution with the standard deviation 0% to 100% of the standard deviation of the corresponding joint axis. The MAE increased while the noise increased. Figure 5 revealed that a wider sliding window (window size $= 256$) is more robust to the Gaussian noise. The MAE did not increase significantly within increasing 30% of standard deviation 256 sliding window size (MANOVA, $p = 0.97$).

Our method is barely affected by the missing data. Figure 6 shows the performance of the proposed method while some parts of the frames are missing. This scenario could be treated as the pose estimation method failing to recognize the person while there was a person in the frame. In each simulation, a percentage of frames were replaced by a zero vector to simulate the missing. Subsequently, these missing frames were imputed by interpolating the nearest frame before the loss and after the loss. Under this imputation method, the proposed method remained nearly constant MAE (MANOVA, $p > 0.99$). The primary reason is that interpolating the missing data did not significantly alter the periodical patterns of the skeleton, so the performance of the proposed method did not vary while missing data was present.

### Counting in different types of skeleton formats
We have also illustrated counting in different types of skeleton formats. MM-fit data is in either a 3D COCO skeleton format (18 joints, 3D) or a 2D COCO skeleton format (18 joints, 2D). UI-PRMD dataset uses Microsoft Kinect skeleton format, which has 22 3D joint locations. The data collected in this study was in 2D COCO

| Description               | MAE (SD)   | OBOA |
|---------------------------|------------|------|
| Camera location           |            |      |
| Front                     | 0.11 (0.15)| 0.85 |
| Left                      | 0.04 (0.05)| 1.00 |
| Right                     | 0.078 (0.13)| 0.89 |
| Concurrent/single motion  |            |      |
| Single person             | 0.05 (0.09)| 0.94 |
| More than one             | 0.11 (0.15)| 0.85 |
| Exercise                  |            |      |
| Squat                     | 0.06 (0.06)| 0.92 |
| Lateral shoulder raise    | 0.11 (0.20)| 0.81 |
| Lunge                     | 0.06 (0.05)| 1.00 |

![MAE breakdown boxplot of additional dataset](image)
The novelty of the proposed work
This work provides a new perspective on analyzing human repetition counting. There is a new finding and a breakthrough in this research. There is no research investigating the frequency domain pattern of pose estimation results to the best of the author’s knowledge. We verify that the repetitive exercise which shows periodical patterns in the video could be analyzed through Fourier transform. This research tackles counting different periodical exercises in the same video through the finding. This task is less addressed in similar research.

Limitation and future research direction
This method assumes the motion is a simple back and forth exercise in which the similarity of skeleton data behaves like periodical signals. Exercise with more complicated procedures might not demonstrate this property. Nonetheless, most rehab or exercise training does not consist of this kind of complex exercise. Second, we assume the most dominant frequency corresponds to the exercise. Minor motion counting will need further investigation on the spectrogram. The last limitation is that we rely on motion detection and recognition repetition motion to perform the proposed algorithm. We might investigate more exercises in different contexts to understand how these limitations affect the proposed method.

Conclusion
We have presented a motion repetition counting method using skeleton data. This research provides a solution to concurrent repetition counting in human exercise, which is less addressed in the previous study. The proposed method has been verified on the public datasets and a few conveniently collected videos with desirable accuracy in different view angles, motions, and concurrent motions. The proposed method possesses the potential in rehabilitation and exercise training to monitor progress and provide remote rehabilitation and exercise training.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1007/s13755-023-00258-3.

Below is the link to the electronic supplementary material: Supplementary file 1 (GIF 5363 kb).

Author contributions
YCH conducted the analysis and writing of the report and data collection. YCH, TE, and KT contributed to the study design and review of the manuscript.
Funding This work is funded by National Key Research and Development Program of China, Ministry of Science and Technology of China: 2019YFE0109600 and Innovation and Technology Fund of Innovation and Technology Commission of Hong Kong: MHP/081/19.

Data availability The data supporting this study's findings are available on request from the corresponding author, YCH. The data are not publicly available due to the privacy of research participants.

Code availability The code for this work is available on https://github.com/YuChengHSU/repetition-counting.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval This research was approved by the Human Subjects Ethics Sub-Committee, City University of Hong Kong (Ref. 3-2-201803_02). All of the participants were well-informed and consent to participate the experiment.

Consent for publication Written informed consent for publication was obtained from all of the participants in our experiment.

Author details
1 School of Data Science, City University of Hong Kong, Tat Chee Rd, Kowloon, Hong Kong, China.
2 AI Lab, Hospital Authority, Argyle St., Kowloon, Hong Kong, China.
3 Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

Received: 18 March 2022  Accepted: 2 November 2023

Published online: 01 December 2023

References
1. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Manual Ther. 2010;15(3):220–8.
2. Heath G, Howe EE, Kahn EB, Ramsey LT. Increasing physical activity; a report on recommendations of the task force on community preventive services. Atlanta: CDC; 2001.
3. Standage M, Duda JL, Ntoumanis N. A model of contextual motivation in physical education: using constructs from self-determination and achievement goal theories to predict physical activity intentions. J Educ Psychol. 2003;95(1):97.
4. Garcia-Garcia FE, Boccherrini-Gallardo M, Rossa-Sierra A, Cortes-Chavez F. Lessons learned. IEEE J. Biomed. Health Inf. 2015;20(1):201–12.
5. Triandafilou KM, Tsoupikova D, Barry AJ, Stoykov N, Thielbar KN, Smeulders AWM. Real-world repetition estimation and validation through deep learning. In: International conference on image analysis and recognition. 2020. https://doi.org/10.1007/978-3-030-50347-5_1.
6. Roosink M, Robitaille N, McFadyen BJ, Hebert LJ, Jackson PL, Bouyer LJ, Mercier C. Real-time modulation of visual feedback on human full-body movements in a virtual mirror: development and proof-of-concept. J. Neuroeng. Rehabil. 2015;12(1):1–10.
7. Thangali A, Scarloff S. Periodic motion detection and estimation via time-space-time sampling. In: 2005 7th IEEE workshops on applications of computer vision (WACV/MOTION'05), vols. 1, 2. IEEE, 2005. p. 176–82.
8. Levy O, Wolf L. Live repetition counting. In: Proceedings of the IEEE international conference on computer vision, 2015. https://doi.org/10.1109/ICCV.2015.346.
9. Garcia-Garcia FE, Boccherrini-Gallardo M, Rossa-Sierra A, Cortes-Chavez F. Automatic 3d human-interpretable feedback models for fitness training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 10387–96.
10. Dwivedi D, Ayar Y, Tompson J, Sermanet P, Zisserman A. Counting out time: Class agnostic video repetition counting in the wild. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 10387–96.
11. Levy O, Wolf L. Live repetition counting. In: Proceedings of the IEEE international conference on computer vision, 2015. https://doi.org/10.1109/ICCV.2015.346.
12. Levy O, Wolf L. Live repetition counting. In: Proceedings of the IEEE international conference on computer vision, 2015. https://doi.org/10.1109/ICCV.2015.346.
13. Dwivedi D, Ayar Y, Tompson J, Sermanet P, Zisserman A. Counting out time: Class agnostic video repetition counting in the wild. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 10387–96.
14. Strömbäck D, Huang S, Radu V. Mm-fit: Multimodal deep learning for automatic exercise logging across sensing devices. Proc. ACM Interact. Mobile Wearable Ubiquit. Technol. 2020;4(4):1–22.
15. Martinez J, Hossain R, Romero J, Little JJ. A simple yet effective baseline for 3d human pose estimation. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 2640–9.
16. Biasioli A, Ahuja N. Extraction and analysis of multiple periodic motions in video sequences. IEEE Trans. Pattern Anal. Mach. Intell. 2007;29(7):1244–61.
17. Sun K, Xiao B, Liu D, Wang J. Deep high-resolution representation learning for human pose estimation. In: CVPR, 2019.
18. Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X, Liu W, Xiao B. Deep high-resolution representation learning for visual recognition. In: TPAMI (2019).
19. Yuan Y, Chen X, Wang J. Object-contextual representations for semantic segmentation. In: Proceedings of European conference on computer vision (ECCV), Glasgow, UK, 2020.
20. Martinez J, Hossain R, Romero J, Little JJ. A simple yet effective baseline for 3d human pose estimation. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 2640–9.
21. Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y. Openpose: realtime multi-person 2d pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. 2021. https://doi.org/10.1109/TPAMI.2019.2929257.
22. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG,Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X. Tensorflow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX conference on operating systems design and implementation; 2016.
23. Runia THF, Snoek CGM, Smequiparians ACM. Real-world repetition estimation by div, grad and curl. In: 2018 IEEE/CVF conference on computer vision and pattern recognition; 2018.
24. Triandafilou KM, Tsoupikova D, Barry AJ, Thielbar KN, Stoykov N, Kamper DG. Development of a 3d, networked multi-user virtual reality environment for home therapy after stroke. J. Neuroeng. Rehabil. 2018;15(1):1–13.
25. Ofir E, Kurillo G, Obdržálek S, Bajczy R, Jimison HB, Pavel M. Design and evaluation of an interactive exercise coaching system for older adults: lessons learned. IEEE J. Biomed. Health Inf. 2015;20(1):201–12.
26. Ishii S, Yokokubo A, Luimula M, Lopez G. Exersive: physical exercise recognition and counting algorithm from wearable robots to positioning. Sensors. 2021;21(1):91.
27. Fierau M, Zarif M, Pilea SC, Olaru V, Sminchisescu C. Aife: Automatic 3d human-interpretable feedback models for fitness training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. p. 9919–28.