AI Fitness Coach at Home using Image Recognition

Haoran Ji (rane0430@gmail.com)
University of Tokushima: Tokushima Daigaku

Karungaru Stephen Githinji
University of Tokushima: Tokushima Daigaku

Terada Kenji
University of Tokushima: Tokushima Daigaku

Research

Keywords: Image Processing, Deep learning, Body Recognition

Posted Date: September 29th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2047283/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
AI Fitness Coach at Home using Image Recognition

Ji Haoran¹,∗ · Stephen Karungaru² · Kenji Terada³

Abstract Recently the number of people exercising at home has increased especially due to the COVID19 pandemic. Therefore, the need for no contact exercise instructions is in great demand since physical access to the gym is limited or discouraged. To meet this demand, many online exercise instruction videos are available. However, the systems are both passive and have no real time feedback to aid the user. In this work, we propose an AI based fitness monitoring system (AI Fitness Coach) that can offer real time guidance during exercise. The AI Fitness Coach, consists of a pose recognition unit, a fitness movement analysis unit, and a feedback unit. The user captures their pose by a fixed camera. The pose recognition unit processes the captured image and outputs the recognition result to the fitness movement analysis unit. After the results are processed by the fitness movement analysis unit, advice is output from the device through video or voice. On comparison to existing methods, the proposed method results are at par and encouraging.

Keywords Image Processing · Deep learning · Body Recognition

Ji Haoran¹,∗
Graduate School of Technology and Science, Tokushima University, Tokushima, 770-8501, Japan
Tel.: +81-80-6116-6868
Fax: +123-45-678910
E-mail: rane0430@gmail.com

Stephen Karungaru²
Graduate School of Technology and Science, Tokushima University, Tokushima, 770-8501, Japan

Kenji Terada³
Graduate School of Technology and Science, Tokushima University, Tokushima, 770-8501, Japan

1 Introduction

Human-computer interaction systems exchange different information between humans, machines, and the environment and carry out understanding and feedback. The major research issue and direction for human-computer interaction is how to more effectively actualize human-computer communication and understand humans. As its core technology, human posture image recognition technology has been successful in human posture recognition and recently in human motion recognition [1]. However, its function is limited to understanding the physical posture or pose of the human body. During this COVID19 period, the demand for more specialized human-machine interaction technology has increased especially for at-home fitness training. Even before coronavirus infection, most sports enthusiasts may not have been aware of the huge risks of using shared sports equipment. According to a study by Victor Tam [2], the number of bacteria in a gym is 362 times that of the toilet seat. Hence the at-home fitness setup is more desirable. To this end, human-computer technology has been applied for online fitness coaches recently. However, additional research is required to achieve meaningful human-computer interaction, such as giving good feedback. Currently, only a few professional fitness training using zoom have online guidance from fitness coaches. When people take fitness courses at home, the fitness software is only a demonstration video with no fitness guidance. In Fig. 1, a problem tree with all sub-issues and issues in the current fitness setup is illustrated. The core problem is “Many people are doing the wrong fitness”.

An AI fitness coach, an artificial intelligence coaching system with computer vision technology as the core, can help users with fitness training as a human coach. In
In this work, we investigate whether artificial intelligence can replace the coaching industry and how computer vision technology can empower it.

2 Related Works

Traditional feature-based methods and deep learning-based methods are the two methodologies for human action recognition. Many traditional image characteristics have been found to function well, including 3D-SIFT [3] and dense trajectories (DT) [4]. IDT [5] increases the performance of DT features even further by taking camera motion into consideration. However, it's expensive to extract dense trajectories and the video descriptors that are required. Conversely, numerous deep learning systems [6][7][8] analyze actions using human skeletal information and have outperformed traditional methods [4][5]. STGCN [8], for example, is a dynamic skeleton model that learns the spatial and temporal patterns of images automatically. Deep learning algorithms, on the other hand, require vast amounts of data to build a stronger model, which restricts their performance on modest quantities of data.

The primary methods used to compare action sequences for the evaluation of human activities are calculating distance errors [9] or dynamic temporal warping (DTW) [10]. The former has a proclivity for overlooking specific action features. A constant stance repeated in a series with visually comparable activities but different temporal evolution might turn out to be a better match to the reference sequence than a constant pose repeated in a sequence with visually similar actions but different temporal history. By warping the two sequences, DTW attempts to solve this difficulty. Enhancing the similarity of local stances DTW, on the other hand, has a hard time estimating movement similarity. Peaks and plateaus have slight temporal changes in their dynamics. Compared with the above methods, we propose a spatial skeleton coding by combining the traditional feature approach and the skeleton-based approach. An Openpos[11] based function library is developed to achieve the effect of detecting fitness.

3 Proposed method

Fig. 2 shows a diagram of the proposed system, AI Fitness Coach. It consists of the pose recognition unit, fitness movement analysis unit, and feedback unit. The user captures their pose by a camera. The pose recognition unit processes the captured image and outputs the recognition result to the fitness movement analysis unit. After the results are processed by the fitness movement analysis unit, the advice is output from the device through video or voice.

4 Overview

In monitoring fitness movements, there are four possible targets as shown in Fig. 3: (a) static single person, (b) static multi people, (c) dynamic single person, and (d) dynamic multi people. The static multi-person fitness move unit can monitor some poses in yoga classes, dynamic single person image monitoring unit can not only monitor sports (golf, baseball, fitness, etc.) but can also be used for medical rehabilitation. Dynamic multi-person image monitoring unit can monitor program rehearsal and multiplayer dance programs.

4.1 System Design: Pose recognition Unit

The pose recognition unit is designed to recognize 12 body segments shown in Fig 4. Each body segment has a tag from “a” to “l”, and a name.
We use OpenPose to extract the body segments. OpenPose will detect the human body and output a JSON file, in which 25 key points are included, and the correct body segment is obtained by concatenating two adjacent key points through the developed database.

### 4.2 System Design: Development of the function library

The posture library is a function library written in Python. The library accept to the JSON file from OpenPose and a tuple \((s1, s2, min, max)\). The library evaluates the tuple and returns the result. OpenPose output the 2D location of human body point in a large array. In order to pick up the location of desired body part, specifying the correct index is necessary. For example, for the location of right shoulder, index value 6 for x-coordinate and value 7 for y-coordinate must be specified. This makes the program code complicated and therefore easy to make mistakes. The posture library defines functions to get location data for all the body segment correctly.

### 4.3 System Design: Pose analysis Unit

OpenPose output the 2D parts location of the human body into a large array. To pick up the location of the desired body part, specifying the correct index is necessary. For example, for the location of the right shoulder, index value 6 for x-coordinate and value 7 for y-coordinate must be specified. Finally, we calculate the angle between two body segments using the eq. 1

\[
\cos \theta = \frac{A \cdot B}{\|A\| \|B\|}
\]  

### 4.4 System Design: Feedback Unit

After the Pose analysis Unit has processed the information and attained the corresponding values, the system will detect the output according to the angle of the preset body segment values and compare them, and finally judge whether the action is correct or not, and give the corresponding feedback, in the case of correct action, the system will output "correct".

### 4.5 System Configuration

Fig 5 shows the configuration of the prototype. The prototype uses a well-known open-source package OpenPose. OpenPose can recognize body segments from a photo image.
a JSON file. Based on the JSON file, the result is given through the Analysis Program finally outputs tips. Fig 6 is the image provided to OpenPose. It not only can output JSON files but also can output various results, such as processed image files, CSV files, etc.

The JSON is an array of 75 numbers. Every three numbers in the array form a tuple \((x, y, z)\), so the array element "pose key point 2d" contains 25 tuples in total.

We call this tuple a key point that corresponds to a position of the body. In each keypoint, \(x\) and \(y\) denote the 2D location of the key point in the image, and \(z\) corresponds to its accuracy.

### 4.6 Experiments and Results

Fig 7 shows the main page where the list of fitness poses is displayed. The user can select the desired pose by clicking a button. After selecting a pose, the user can enter the target time duration of the pose as shown in Fig 8. In this case, the user wants to keep this pose for 30 seconds. Then the user can start the function by clicking the “start” button. The user has to place the device so that its camera can capture the user’s pose. The device gives the user a preparation time and before starting processing as shown in Fig 9.

The device judges the correctness of the user’s pose once per every second and count down from 30 seconds to zero while it judges the pose as correct, or gives proper tips if it judges the pose as incorrect for three continuous seconds. The device can also display and record the recognized result as shown in Fig 10. The user can play back and check their pose after fitness.

Fig 11 is the success result graph given by the system. When successful, the system gives the success feedback. Fig 12 is the graph of the failure results given by the system. When the action fails, the system gives the appropriate advice and has the option to return or retry.

### 5 Ai fitness coach for Wallsit

A wall sit [12] is an exercise done to strengthen the quadriceps muscles. The exercise is characterized by the two right angles formed by the body, one at the hips and one at the knees. The person places their back against a wall with their feet shoulder-width apart and a short distance from the wall. Then, keeping their back against the wall, they lower their hips until their knees form right angles. This is a very intense workout for the quadriceps muscles, and it can be very painful to hold this position for extended periods. It requires...
The definition of wall sit is shown in Fig 13.

The prototype of the pose library does not support special segmentation. Therefore, the rules of niches will be simplified. If any of the conditions is negative, the pose is incorrect. Some special cases are shown in Fig 14.

There are several special cases. When the program starts, it will remind the user of the correct posture, which requires the user’s active cooperation. The current program can’t monitor the wall performed on the ground. For example, if the user does the pose like shown in Fig 15, The system will judge the pose correct.

### 6 Ai fitness coach for Plunk

The plank (also called a front hold, hover, or abdominal bridge) is an isometric core strength exercise that involves maintaining a position similar to a push-up for the maximum possible time. The most common plank is the forearm plank which is held in a push-up-like position, with the body’s weight borne on forearms, elbows, and toes. Many variations exist such as the side plank and the reverse plank. The plank is commonly practiced in Pilates and yoga, and by those training for boxing and other sports. On the plank version, the function library of the first version is still used. This pose affects 16 body parts of all function libraries. Fig 16 is the CSV file of plank.

<table>
<thead>
<tr>
<th>Part</th>
<th>Part2</th>
<th>Min</th>
<th>Max</th>
<th>Tip status</th>
<th>Tip message</th>
</tr>
</thead>
<tbody>
<tr>
<td>RightOpenLeg</td>
<td>RightLowerLeg</td>
<td>150</td>
<td>200</td>
<td>Success</td>
<td>Right leg must be straight</td>
</tr>
<tr>
<td>RightOpenLeg</td>
<td>RightLowerLeg</td>
<td>150</td>
<td>200</td>
<td>Success</td>
<td>Right arm over than 180 degree</td>
</tr>
<tr>
<td>LeftOpenLeg</td>
<td>LeftLowerLeg</td>
<td>150</td>
<td>200</td>
<td>Success</td>
<td>Left arm must be straight</td>
</tr>
<tr>
<td>RightOpenArm</td>
<td>RightUpperArm</td>
<td>90</td>
<td>150</td>
<td>Success</td>
<td>Right arm over than 150 degree</td>
</tr>
<tr>
<td>LeftOpenArm</td>
<td>LeftUpperArm</td>
<td>90</td>
<td>150</td>
<td>Success</td>
<td>Left arm over than 150 degree</td>
</tr>
<tr>
<td>RightShoulder</td>
<td>RightUpperArm</td>
<td>90</td>
<td>150</td>
<td>Success</td>
<td>Left shoulder over than 90 degree</td>
</tr>
<tr>
<td>RightShoulder</td>
<td>RightUpperArm</td>
<td>90</td>
<td>150</td>
<td>Success</td>
<td>Left shoulder less than 90 degree</td>
</tr>
<tr>
<td>LeftShoulder</td>
<td>LeftUpperArm</td>
<td>90</td>
<td>150</td>
<td>Success</td>
<td>Left shoulder over than 150 degree</td>
</tr>
<tr>
<td>Neck</td>
<td>Neck</td>
<td>150</td>
<td>200</td>
<td>Success</td>
<td>Neck needs straight</td>
</tr>
</tbody>
</table>

Fig. 16: csv file of plank
7 Ai fitness coach for System Expansion

As proposed in section 4.2 this system has good scalability because of the pre-definition of each body line segment through the function library in advance. Developers and users can implement new actions in a short time by calling the function library and setting up the csv file. For example, if we want to edit a kneeling lunge now, we only need to call the function library as shown in Fig17 The system will be able to detect the action automatically, but it is recommended to set it by the developer or the user.

Fig. 17: CSV for kneeling lunge

8 Evaluation

8.1 Accuracy of Wall sit

Table 1 is the test result of the first pose, with 4 test subjects, and 76 tries the test. The test was conducted from the front, sides, and diagonal sides. The poses tested were correct "wall-sit", standing pose, Left knee over ankle, Right knee over ankle, and double knee over ankle. We attained an accuracy rate of 85% for the correct pose, 100% for standing, 57% for left knee over ankle, 66% for right knee over ankle, and 77% for double knee over ankle.

In Table 1, the angle of the image also affects the accuracy rate, with the accuracy rate of 66% in the front, 97% on both sides, and 74% for the diagonal. The accuracy rate is reduced for the wrong pose, while the accuracy rate is high in the standing or correct pose. When the image is input on both sides of the subject, the accuracy is high, but on the front, because of the function of the subject detection system and some visual problems, the accuracy will be reduced.

8.2 Accuracy of Plank

In Table 3, the test results of the second pose "plank" are shown. 76 tries test for two people. The poses were still sampled from four angles, front, both sides of the body, and diagonal. In Table 4, the result of the accuracy for the correct pose is 69%, and the accuracy for the wrong pose is 71%. Similarly, we can calculate the accuracy rate of each angle. The accuracy for the front side is 54%, the accuracy for both sides of the body is 85%, and the accuracy rate of the diagonal side is 68%

8.3 Accuracy on public datasets

For action recognition and evaluation, we evaluated the performance of our method on a fitness action dataset. In Table 5 as a complement, we also used the evaluation dataset of Accuracy of Efficient Fitness Action Analysis based on Spatio-Temporal Feature Encoding [13]. In comparison, we can see from the dataset that
Front view of Efficient Fitness Action Analysis Based on Spatio-Temporal Feature Encoding has higher accuracy, while the detection accuracy of side view of our study is higher, and in Table 6 we add Diag view to Front view and side view.

Table 5: Accuracy of Efficient Fitness Action Analysis Based on Spatio-Temporal Feature Encoding action recognition with different feature dimensions in front view and side view [13]

<table>
<thead>
<tr>
<th>Feature Dimensions</th>
<th>Front</th>
<th>Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>8x8</td>
<td>87.63%</td>
<td>67.25%</td>
</tr>
<tr>
<td>12x12</td>
<td>94.71%</td>
<td>83.19%</td>
</tr>
<tr>
<td>16x16</td>
<td>95.78%</td>
<td>87.33%</td>
</tr>
<tr>
<td>20x20</td>
<td>96.85%</td>
<td>89.20%</td>
</tr>
<tr>
<td>24x24</td>
<td>97.27%</td>
<td>90.64%</td>
</tr>
<tr>
<td>28x28</td>
<td>96.80%</td>
<td>91.51%</td>
</tr>
<tr>
<td>32x32</td>
<td>96.85%</td>
<td>91.77%</td>
</tr>
</tbody>
</table>

Table 6: Accuracy of our system

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>diag</th>
<th>side</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>64.5%</td>
<td>61.5%</td>
<td>88.5%</td>
</tr>
<tr>
<td>wrong</td>
<td>54%</td>
<td>76.5%</td>
<td>97%</td>
</tr>
</tbody>
</table>

8.4 Discussion

The prototype has sufficient accuracy if the camera direction is inside vision. Its accuracy is low if the camera direction is the front vision or diagonal vision. Based on the test results, it can be seen that the “wall-sit” accuracy of the wrong pose will be reduced, while the accuracy of the standing pose or correct pose will be very high. When the image is input to both sides of the subject, the accuracy of the result is very high, but in the front, because of the function of the character detection system and some visual problems, the accuracy will be reduced. Because the second pose “plank” needs to detect more points, the accuracy rate has declined. In terms of angle, the accuracy rate is still high on both sides of the body, the diagonal side is medium, and the front is the lowest.

8.5 Limitation of prototype

Currently, the prototype only can process two fitness poses “wall-sit” and “Plank”. It also can only do limited functions as: Input: one static image output: message of “success” or “failure” together with tips messages. The prototype is designed for execution on a PC and not on a smartphone because the original OpenPose requires a high-performance graphics card. However, we can implement it also on a smartphone, if we use another version of OpenPose android-demo which is adapted to execution on a smartphone. The prototype has another limitation in that it does not support the special segments, such as (Horizontal Line and Perpendicular Line)

9 Conclusion

In this work, we propose an artificial intelligence-based fitness monitoring system (AI fitness coach) that provides real-time guidance during exercise. Posture recognition unit, Fitness movement analysis unit and Feedback unit all work well.

The experimental results proofs the effectiveness and viability of the proposed method. They show that the system can work normally, and the development of the function library is effective, which can greatly save the time of developers. Because of the function library, developers who do not know how to code can use csv files to edit and add or adjust movement poses, thus effectively joining the development of contact-less fitness.

The results of the proposed method are identical and encouraging compared to the existing methods.

10 Future works

The developed prototype runs well, but a lot of work and improvements are needed to optimize its terminal adaptability to achieve the following proposed solutions:

1. Replace the human body detection system, in order to achieve higher accuracy, and reduce the calculation time because the current human body detection system calculation is huge, resulting in slow operation.

2. Develop more functions based on this system, not only in the field of fitness, but also in the field of medical rehabilitation and social welfare.

3. Deep learning by CDBN(Convolutional deep belief network)[14] on the basis of the prototype and add more actions to improve the system.

Conflict of interest

The authors declare that they have no conflict of interest.
References

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- bio.pdf