Voice-Based Bodyweight Training Support System Using Smartphone

Ruiyun Wang(✉), Shin Takahashi, Buntarou Shizuki, and Ikkaku Kawaguchi

University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan
wang@iplab.cs.tsukuba.ac.jp, {shin,shizuki,kawaguchi}@cs.tsukuba.ac.jp

Abstract. Bodyweight training has grown in popularity; it is desirable to be fit and strong. However, training can be dangerous if performed incorrectly. Several systems are used to correct pose during training. However, most require wearable sensors that may interfere with training, or an expensive depth camera. We offer a new form of training support using a smartphone camera and a server. We use a verbal interface to help users to correct their pose and to encourage them. We describe our new system and experimentally evaluate it.

Keywords: Training · Verbal interface · OpenPose · Bodyweight training

1 Introduction

Bodyweight training is any work performed against gravity, and includes pushups, pullups, and squats used to increase strength and stamina [1]. Bodyweight training has grown in popularity because it requires minimal or no equipment. However, errors in pose or training intensity compromise the possible benefits and may negatively affect the joints. To avoid injury, it is desirable to observe and evaluate poses assumed during training.

Few convenient and inexpensive methods are available to correct pose and ensure appropriate exercise intensity. Support systems using wearable sensors or depth cameras (such as Microsoft Kinect) are expensive. Body-attached sensors may interfere with training.

Here, we develop a new bodyweight training support system using only a smartphone and a server; users receive verbal corrections and encouragement during training. We use OpenPose [2] to obtain skeletal data from RGB (Red–Green–Blue) images taken by the smartphone camera; these data aid pose recognition and correction. Feedback is aural (delivered by the smartphone). We constructed a prototype and evaluated it experimentally.

2 Related Work

Training support is a popular research topic. Depth cameras, wearable sensors, RFID (Radio Frequency IDentification) technologies, and RGB cameras have

A. Marcus and E. Rosenzweig (Eds.): HCII 2020, LNCS 12202, pp. 370–379, 2020.
https://doi.org/10.1007/978-3-030-49757-6_26

been used to detect user movements during training. For example, Eyes-Free Yoga [3] (an accessible yoga exergame) uses Microsoft Kinect to enable low– or no–vision subjects to correct their poses based on skeletal tracking and verbal feedback. Lee et al. [4] developed a rehabilitation system based on the Kinect sensor to assist patients with movement disorders to perform "Tai Chi" exercises at home. The FitCoach [5] is a virtual fitness coach using an accelerometer and a gyroscope powered by a smartphone or smartwatch to evaluate the quality of user exercises. The differences in body movement strength and speed between exercise repetitions are evaluated. RunBuddy, developed by Hao et al. [6], measures the breathing rhythm during running using an acceleration sensor and a wireless earphone connected to a smartphone. The physiological state can be estimated by reference to the breathing rhythm. The TTBA, an RFID-based motion tracking system proposed by Ding et al. [7], uses a single dumbbell tag to recognize vertical and circular motion. Recently, an RGB camera has been used to read two-dimensional gestures and thus support training. Qiao et al. [8] developed a real-time gesture grading system employing a single RGB camera. The differences between the standard and user joint trajectories are compared and the gesture grade calculated and shown on a screen.

Unlike previous works, we focus on low cost and convenience; we develop a voice-based bodyweight training support system using a smartphone camera. We deliver real-time verbal feedback to users.

3 System Overview and Implementation

We use the GUI (Graphical User Interface) smartphone application to support bodyweight training (Fig. 1); we deliver verbal corrections and encouragement via a smartphone. For example, in Fig. 2, user pose (a) is captured and evaluated during training. The verbal correction is "*Do not bend your knee*". If the pose is then corrected, the user is told the time for which s/he is required to assume a high-quality pose. Successful training in a certain pose is followed by: "*Good job, go to the next pose*". Before training commences, smartphone imaging during training must be established. Pose support requires whole-body images; the smartphone must thus be 2 to 3 m distant from the user. Users could also confirm the correct poses through the description using texts and figures in the GUI application before training.

Figure 3 shows the system architecture; the steps are: Attribute Pose Data Acquisition, Pose Recognition, and Pose Evaluation. The smartphone camera images poses assumed during training, evaluates them with the aid of the server, and sends verbal feedback to the user. We used OpenPose (a real-time, two-dimensional pose estimation method) to track the skeleton in RGB images; OpenPose recognizes body joints and yields their X and Y positions; we derived rules for training pose recognition and evaluation. The attribute pose data are the distances from the joints to the center of the body and the angles between the vectors.

We constructed a prototype. The smartphone was a Samsung Galaxy S10 running Android Ver. 9; the software prototype was implemented using QT

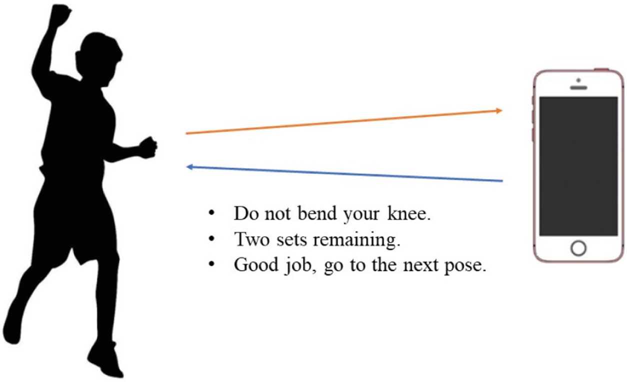

Fig. 1. Use scene.

for Android. We used Python API to run OpenPose in the server; all images were analyzed by NVIDA GeForce RTX 2060 at a speed of about 16.0 fps. The smartphone and the server were connected to the same local area network; a UDP connection was used to transfer images.

(a) (b) (c) (d)

Fig. 2. The four poses used in the evaluation experiment, in the order in which they were performed.

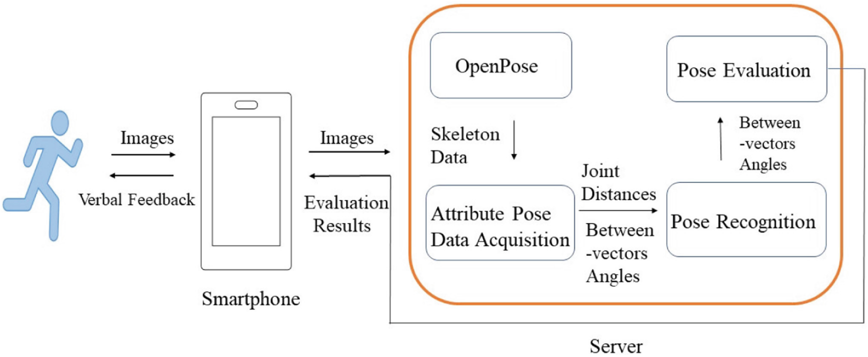

Fig. 3. The system architecture.

3.1 Attribute Pose Data Acquisition

We used OpenPose to acquire skeletal coordinates. OpenPose accepts RGB image inputs, and outputs the two-dimensional positions of 25 key anatomical points. We used the coordinates of 15 key points on the trunk (Fig. 4) to calculate attribute pose data (joint distances and angles between vectors); all data were employed for pose recognition. For pose evaluation, we used only pose-specific angles, thus, not all the angles. The joint distances and the between-vector angles used are shown in Fig. 4. The distances between the elbows, arms, knees, and ankles, and the center of the body, are indicated by red lines. We calculated 12 different angles: (n_0, n_1), (n_1, n_2), (n_2, n_3), (n_0, n_4), (n_4, n_5), (n_5, n_6), (n_7, n_8), (n_8, n_9), (n_9, n_{10}), (n_7, n_{11}), (n_{11}, n_{12}), (n_{12}, n_{13}).

3.2 Pose Recognition

Pose recognition employed the k-nearest neighbor algorithm [9], which is a simple but very accurate non-parametric method. The joint distances and between-vector angles served as the attributes.

3.3 Pose Evaluation

We evaluated pose accuracy, the time for which the pose was maintained, and the number of pose repetitions. For example, for pose (a) of Fig. 2, evaluation proceeds as follows:

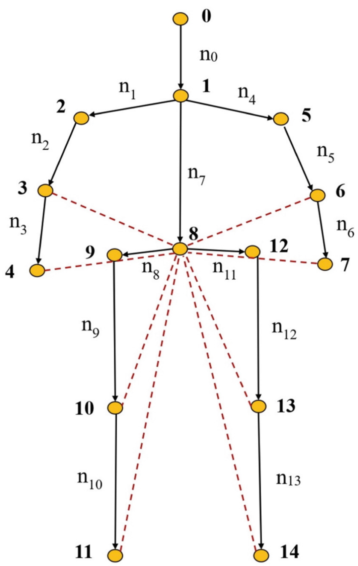

Fig. 4. Attribute pose data.

1. Check whether the angle between the thigh and the back (n_7, n_9) is 90°.
2. Check whether the angle between the shin and the thigh (n_9, n_{10}) is 180°.
3. If pose (a) was correct, was it maintained for 10 s?
4. Was pose (a) repeated four times?

In the implementation, we considered the recognition accuracy of OpenPose and set angle ranges.

4 Experiment 1: OpenPose Recognition Accuracy

To verify OpenPose recognition accuracy in terms of supporting bodyweight training, we performed a preliminary experiment using OpenPose to analyze images taken during training.

4.1 Participants

We recruited three graduate students (two males and one females) aged 24 to 27 years, of height 172 to 185 cm.

4.2 Procedure

The participants assumed bodyweight training poses and images were taken during training. We used OpenPose to identify and display key bodily points; we checked whether all points were correctly detected and displayed.

The smartphone camera was positioned vertically 2 and 3 m in front the participants. The three participants assumed six poses at two different distances; we collected 180 whole-body images totally. The participants first assumed three poses, including pose (c) of Fig. 2, while facing the camera. Next, they assumed three poses, including poses (a) and (b) of Fig. 2, with the camera to their right.

4.3 Results and Discussion

We calculated the percentage of the each key point that could be detected correctly in the 180 images. The recognition accuracies of the three poses when facing the camera are shown in Table 1. The accuracies when the camera was to the right are listed in Tables 2 and 3.

Recognition accuracy was high when participants faced the camera. When the camera was on the right, key left-side points were poorly detected because the camera could not see them. We assumed that bodyweight training poses were symmetrical; thus coordinates of the left side could be calculated and very accurate data could be obtained.

In addition, we also found that when the camera was closer to the participants (who then tended to fill the image), recognition accuracy decreased.

Table 1. The recognition accuracy of the three poses evaluated from the front.

Distance/m	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Average
2	100	100	100	97.8	100	100	96.7	99.2
3	100	100	100	100	100	100	100	100

Table 2. The right side recognition accuracy of the three poses evaluated from the right.

Distance/m	Head	Center	r-Sho	r-Elb	r-Wri	r-Hip	r-Kne	r-Ank	Ave
2	100	83.3	100	90	90	70	80	53.3	83.3
3	100	100	100	100	100	96.7	100	100	99.6

5 Experiment 2: System Evaluation

We explored whether the proposed system delivered useful and convenient training support.

Table 3. The left side recognition accuracy of the three poses evaluated from the right.

Distance/m	l-Sho	l-Elb	l-Wri	l-Hip	l-Kne	l-Ank	Ave
2	90	0	0	50	46.7	26.7	35.6
3	100	0	0	100	66.7	53.3	53.3

5.1 Participants

We recruited eight graduate students (P1–P8) (five males and three females between the ages of 23 and 27 years). Since the proposed body-weight training system was designed for the beginners in bodyweight training, all of the participants had little or no experience in bodyweight training.

5.2 Procedure

We first explained the purpose and flow of the experiment. We asked each subject to confirm that s/he was in good physical condition; we wished to be sure that all training tasks would be completed. All participants performed four types of bodyweight training (Fig. 2). Each training exercise required about 10 s; we scheduled four repetitions. The entire process required about 8 min. Figure 5 shows the experimental conditions during evaluation.

The flow of the experiment and the specific support method are described using the pose (a) in Fig. 2 as an example. Given the limitations of OpenPose, we accepted angles from 80° to 100° as 90° and 170° to 190° as 180°. Each participant pressed the GUI start button. The words: "*Please set the smartphone*" were spoken. Each participant trained while learning how to assume pose (a). If the pose was correct, a voice began to count the seconds. If the angle between the thigh and the back was incorrect, the voice said: "*Please raise your foot*". If the angle between the thigh and shin was incorrect, the voice said: "*Do not bend your knees*". If the pose remained incorrect, the second-count ceased until the user correctly adjusted his/her pose. After correction, the second-count recommenced. As each set was completed, the words "*One set*", "*Two sets*", "*Two sets remaining*", and "*The last set*" were vocalized. After all sets were completed, the voice said: "*Good job, go to the next pose*" and the user moved on. When all training was completed, the experiment was terminated using the words: "*Good job, the training is finished*".

After the experiment was completed, a questionnaire was administered (Table 4).

5.3 Results

The answers to questions Q1–Q5 are listed in Table 5. The answers to questions Q6 and Q7 were as follows (translated from the Japanese):

P1: *After becoming accustomed to all of the training actions, I think that the voice-only training support is convenient and effective.*
P2: *The proposed system very effectively supports bodyweight training, but I want more types of supported training actions.*
P3: *It is convenient because I can use it at home, but it seems that identification accuracy is poor. I wonder if training is in fact effectively supported. Also, the system cannot be used in a small room.*
P4: *It is very important to correct posture, but the whole body must be photographed; the system cannot be used in a small room.*

Table 4. The questionnaire.

No	Questions	Answers
Q1	Your gender and age	Free description
Q2	Do you usually do bodyweight training?	5·4·3·2·1 Always ⇔ Never
Q3	The system could support bodyweight training	5·4·3·2·1 Strongly agree ⇔ Strongly disagree
Q4	The system was convenient	5·4·3·2·1 Strongly agree ⇔ Strongly disagree
Q5	I want to use the system again	5·4·3·2·1 Strongly agree ⇔ Strongly disagree
Q6	Do you have any advice about this system	Free description
Q7	Please tell us your sense of use	Free description

Table 5. The questionnaire results.

	Q1	Q2	Q3	Q4	Q5
P1	23 Female	2	5	5	5
P2	23 Female	3	5	5	5
P3	26 Male	3	3	3	3
P4	25 Male	3	5	4	4
P5	24 Male	2	5	4	4
P6	24 Female	1	4	4	5
P7	25 Male	3	4	4	4
P8	27 Male	1	4	5	4
Average		2.25	4.38	4.25	4.25

Fig. 5. Experimental conditions during evaluation.

P5: *The voice can help me correct poses and I don't need to look at the screen, so it is useful. It would be better if there was a function to score after the training was completed.*
P6: *The verbal corrections increase my desire to assume a correct posture. Even when I become tired, it is easy to continue training a few seconds after the verbal notification.*
P7: *More training actions would be good. Also, the correct posture ranges should be adjusted by a user depending on the individual physical situation.*
P8: *Only simple postures are supported. It is difficult to support more complex postures.*

5.4 Discussion

On Q3, all participants awarded scores of 4 or more (average 4.38), indicating that posture evaluation and correction were effective. Q4 and Q5 explored system convenience; the average score was 4.25, indicating the system was easy to use. The answers to Q6 and Q7 showed that it was difficult to support more complex training poses; this is a topic for the future. The voices should convey more information. A video offering more specific descriptions of poses before training might be useful. Visual and verbal feedback could be combined. An overall training score was requested; we will soon implement this to encourage users to keep exercising. The score will be based on the differences between the user and standard poses; the training time; and the number of pose repetitions. As a user scores more highly, rewards will be given to enhance motivation.

Different voices should be used for correction. Also, the current verbal feedback works for poses with long, but not short, hold times. Brief comments are required to correct short wrong poses. Also, if the user found it difficult to finish a pose, that pose could not be skipped. The system must deliver feedback when a user cannot complete his/her current training. We will ask the user if s/he finds it difficult to complete the pose, and skip that pose if the answer is "Yes". In addition, there is a need to distinguish a wrong posture from the state of training cessation. Finally, we explored only posture evaluation; real-time posture identification is required in future.

6 Conclusion

We present a voice-based, bodyweight training support system using a smartphone and a server. We used skeletal data from OpenPose for pose recognition, evaluation, and correction. We developed a prototype and evaluated it experimentally. The system was convenient and effective.

In future, we will improve the voices used, deliver more accurate pose corrections, and enhance motivation. We will improve real-time recognition and finally develop a system supporting all facets of bodyweight training.

References

1. Harrison, J.S.: Bodyweight training: a return to basics. Strength Cond. J. **32**(2), 52–55 (2010)
2. Zhe, C., Tomas, S., Shih-En, W., Yaser, S.: Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, pp. 7291–7299. IEEE (2017)
3. Rector, K., Bennett Cynthia, L., Kientz Julie, A.: Eyes-free yoga: an exergame using depth cameras for blind & low vision exercise. In: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2013), Bellevue, Washington, pp. 12:1–12:8. ACM (2013)
4. Lee, J.-D., Hsieh, C.-H., Lin, T.-Y.: A preliminary study of using kinect-based physical rehabilitation system to perform Tai Chi exercises with FLS evaluation. Neuropsychiatry, Int. J. Clin. Skills **8**(1), 165–175 (2018)

5. Guo, X., Liu, J., Chen, Y.: FitCoach: virtual fitness coach empowered by wearable mobile devices. In: Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM 2017), GA, USA, pp. 1–9. IEEE (2017)
6. Hao, T., Xing, G., Zhou, G.: RunBuddy: a smartphone system for running rhythm monitoring. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2015), Osaka, Japan, pp. 133–144. ACM (2015)
7. Ding, F., Zhang, Q., Zhao, R., Wang, D.: TTBA: an RFID-based tracking system for two basic actions in free-weight exercises. In: Proceedings of the 14th ACM International Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet 2018), QC, Canada, pp. 7–14. ACM (2018)
8. Qiao, S., Wang, Y., Li, J.: Real-time human gesture grading based on OpenPose. In: 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), ShangHai, China, pp. 357–358. IEEE (2017)
9. Hastie, T., Tibshirani, R.: Discriminant adaptive nearest neighbor classification. IEEE Trans. Pattern Anal. Mach. Intell. **18**(6), 607 (1996)