Proceedings of ICCT2015

Virtual Personal Trainer via the Kinect Sensor

Xin Jin[1, 2, *], Yuan Yao[1], Qiliang Jiang[1], Xingying Huang[1], Jianyi Zhang[1], Xiaokun Zhang[1], and Kejun Zhang[1, *]

[1]Department of Computer Science and Technology
Beijing Electronic Science and Technology Institute, Beijing, 10700, China
[2]Information Technology Research Base of Civil Aviation Administration of China,
Civil Aviation University of China, Tianjin 300300, China
*Corresponding authors: {jinxin, zkj}@besti.edu.cn

Abstract: Our daily life is more convenient than ever before because of the great progress of science and technology. However, the body building or fitness are always ignored in modern daily life. In this paper, we propose a Virtual Personal Trainer to provide real time visually action guide and action assessment during the fitness time of users via the Microsoft Kinect Sensor. The user actions are captured by Kinect and compared to standard actions so as to give a fitness score representing how well users perform actions in real time. Users can see the skeletons of his/her own action and standard action in the screen together with videos. An action correctness guide will be displayed to guide the user to correct his/her actions. Users correct his/her actions to match the standard action in our pre-collected fitness action library. Thus our system can provide interpretable feedback on how people can improve their action. The whole system gives a new definition of the family gym in the future with the power of human-machine interactive.

Keywords: Virtual Personal Trainer; Kinect Sensor; Action Recognition; Action Assessment; Action Correctness

1 Introduction

The importance of exercising has been a common sense among people for a long time, but how to make them especially those white collars actively involved into those tiring practices is a big problem. Our daily life is more convenient than ever before because of the great progress of science and technology. However, the body building or fitness are always ignored in our daily life. People need going to fitness center or finding personal trainer to provide plans and instructions of their own fitness.

Current fitness guide products include book, video, mobile Apps etc., as shown in Figure 1, most of which do not provide real time fitness guide or even do not support any interaction with users. Typically, these products have several drawbacks. First, the whole purposes or main function of the products are too simple and dull, they only serves to give advice or some tips on the exercises without pertinence. Second, although some products aim at record the time or miles of how user walks and gives some values on how much calories has been consumed. Still, the accuracy is a problem. Last but not the least, they all neglected the comfortless of human-computer interacting.

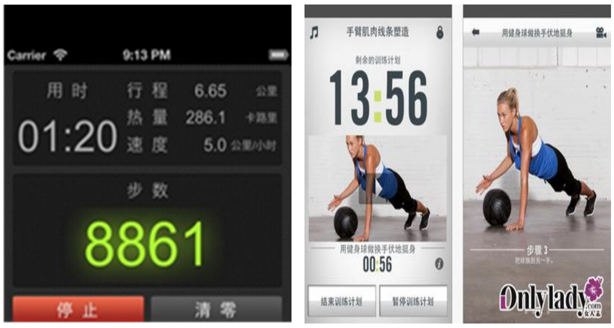


Figure 1 Typical Exercising Assistant Apps

With the virtual reality technology developing such incredible, people are eager to access their equipment with the least actions and fully liberate their hands and even fingers. The Kinect Sensor with its ability of capturing color images and depth images together, has been used in both industry and scientific research such as health care, personal entertainment, electronic commercials [1][2][3][4][5]. Thus we propose a system for personal fitness in the home using Microsoft Kinect Sensor so as to get better user experiences.

The Virtual Personal Trainer we build can provide real time visually action guide and action assessment during the fitness time of users. The user actions are captured by Kinect Sensor and compared to standard actions so as to give a fitness score representing how well users perform actions in real time. Users can see the skeletons of his/her own action and standard action in the screen together with video.

An action correctness guide will be displayed to guide the user to correct his/her actions. Users correct his/her actions to match the standard action in our pre-collected fitness action library. Thus our system can provide interpretable feedback on how people can improve their action. In addition, our system can be used for in social network for sharing scores of users. Our system can be also embedding in the online exercising and dancing game for competition, instead of current boring keyboard control.

system. The fitness score between her actions with the corresponding standard action from the database in real time. In the meanwhile, some correction guides are always shown to the user to guide her correcting her own

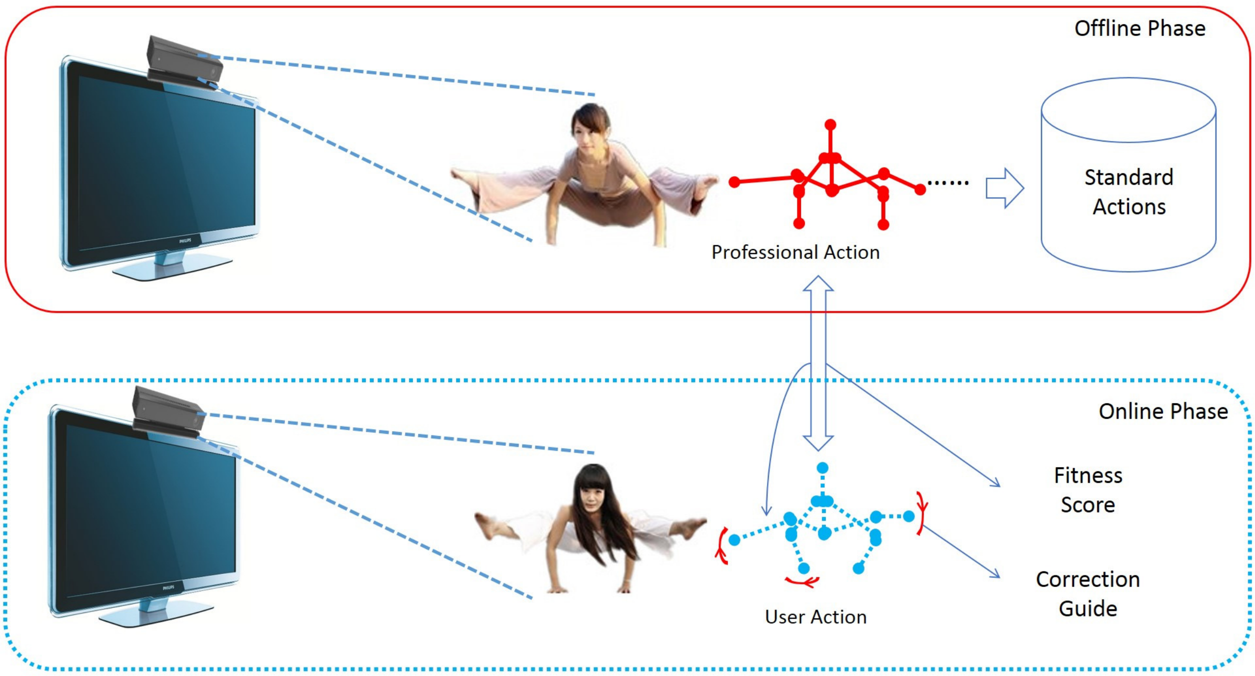


Figure 2 System Overview. Our system contains two phases: the offline phase and the online phase. In this example, a professional Yoga coach makes a standard Yoga action. We record it with the Kinect sensor as a skeleton and save it into our standard action database. A user makes the same action using our system and we compute the fitness score between her actions and the corresponding standard action from the database in real time. In the meanwhile, some correction guides are always shown to the user to guide her correcting her own actions.

Thus our system is indeed a personal and humane trainer just beside user during their home fitness. This project perfectly match the idea of world citizenship: improving our life. We hope our virtual personal trainer can help the sedentary white-collar to have more work-out, more and more people with disabilities to have personal rehabilitation instructors, people lack high cost of personal trainers to have another way for fitness. The virtual personal trainer is the beginning of healthy life.

The following paper is arranged as: An overview of our system is introduced in Section 2. The detail of our system is described in Section 3. The implementation results are shown in Section 4. We conclude our paper and make some discussion in Section 5.

2 Overview of Our System

In this section, we give an overview of our system. As shown in Figure 2, our system contains two phases: the offline phase and the online phase.

In the offline phase, professional athletes, coaches, etc. make standard exercising actions. We record it with the Kinect sensor as a skeleton and an action video. The skeleton and the corresponding action video are stored in our standard action database.

In the online phase, users make the same action using our actions, as shown in Figure 3.

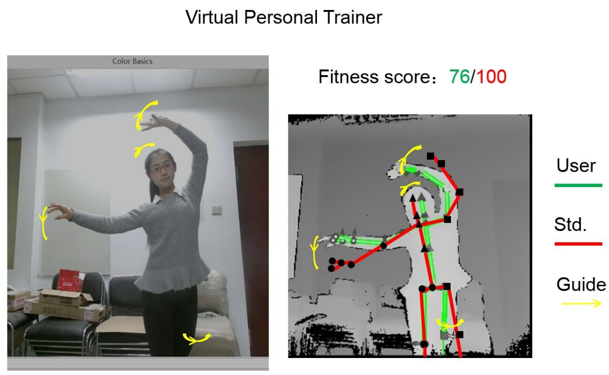


Figure 3 The proposed Virtual Personal Trainer. The left is the user video captured by the camera of the Kinect. The right is the real time skeleton of the user (shown in green lines) and the matched standard skeleton (shown in red lines). The fitness score is shown above the skeleton in real time. The action correction guide in 3D space is shown in both the user video and the skeleton video.

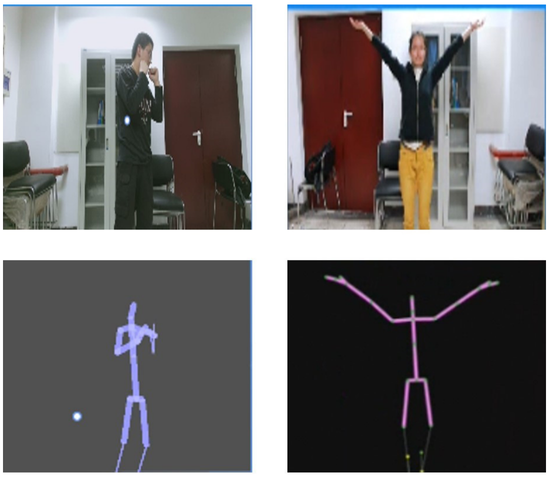
3 The Virtual Personal Trainer System

The detail of our system is described in this section. The action capture and the standard action database, action comparison are described below.

3.1 Action Capture and Standard Database

In both the offline and the online phases, we need to capture actions. The standard action and the user action are both capture by the Kinect sensor. There are some physical limitations of Kinect for Windows. The distance of 1.8~2.5 miters from user's location to the Kinect camera will be best.

One of our most important purpose is to satisfy different groups of customers. Hence, the capability of building a series of standard actions to use in the future is necessary. Various actions such as *jump*, *punch*, and *dance* should be distinguished. Two actions captured and the corresponding video shots are shown in Figure 4.



(a) punch (b) dance

Figure 4 Two actions of *punch* and *dance* captured and the corresponding video shots.

To make the actions in the database can be comparable with real ones, we use Gesture Builder to make some tags at those "key points", and how to define the key points in your action processes is depend on the user. Generally, more tags made, more accurate the comparison would be shown in results, as shown in Figure 5.

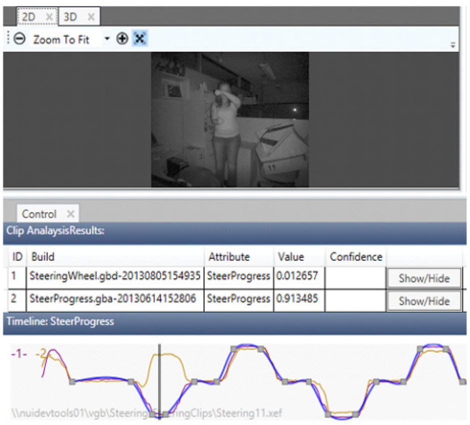


Figure 5 In an example of making the *rotating* actions database by Kinect, we can easily see how explicit it will gives us the processing data from the raw data. The two lines with different colors are the results of positive and negative samples processed by the Adaboost classifier, and how much they are close to each other can be explained as the confidence of the action.

3.2 Action Comparison

We propose a three-stage comparison method between the user action and the standard action. First, we use the Adaboost classifier to classify the user action into one of the action class in our standard action database. Then, we leverage a dynamic warping and 3D Euclidean distance based scheme to compute the fitness score between the user action and the standard action.

Action Classification. This stage is provided by the Kinect SDK, the Adaboost trigger can detect a discrete gesture. Boolean values of true/false are used to indicate when the user is performing a gesture [6]. We use this feature to classify the user action into one of the action class in our standard action database.

Dynamic Space-time Warping. After the action classification, for a same action, different person may use different time. Thus, we should warp the user action to use the same time to that the standard action use. The traditional Dynamic Time Warping (DTW) algorithm matches two 1D signals, equivalent to search the optimal path in a 2D matrix. We use the extended DTW algorithm the Dynamic Space-time Warping (DSTW) [7], which is used to warp the user action of length T_u with the standard action of length T_s equivalents to search the optimal path by dynamic programming.

Fitness Score. The last stage is to compute the fitness score. After warping the user action to the same length of the standard action, we compute the fitness score frame by frame in 3D Euclidean space. The skeleton obtained by the Kinect sensor contains 20 joints, as shown in Figure 6.

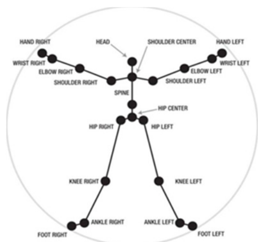


Figure 6 The skeleton joints of the Kinect

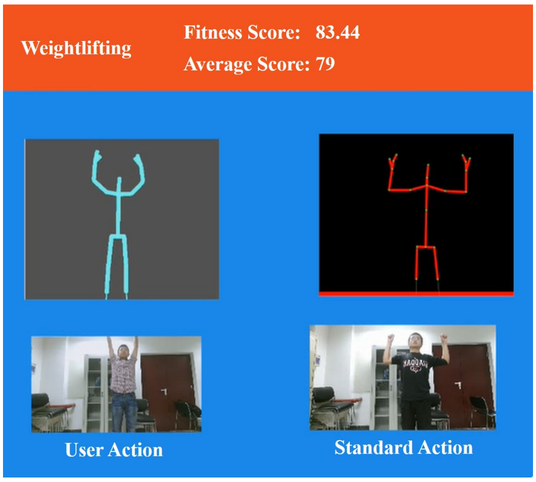
The fitness score F_s is computed by the 3D distances between the joints of the user action and the standard action:

$$F_s = \frac{1}{\sum_i M(S_i)} \sum_i M(S_i) e^{-\frac{D^2(S_i)}{2\sigma}} \tag{1}$$

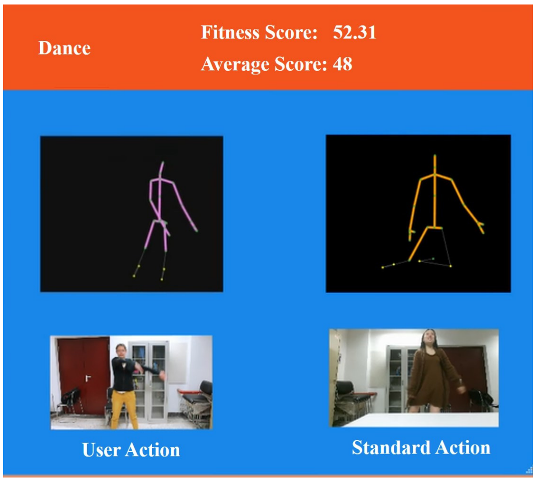
where, the s_i is the ith joint. $M(s_i)$ is the weight of each joint. The $D^2(S_i)$ is square of the 3D Euclidean distance between the corresponding joint of the user action and the standard action. Then the fitness score is mapped into $[1,100]$.

4 Results

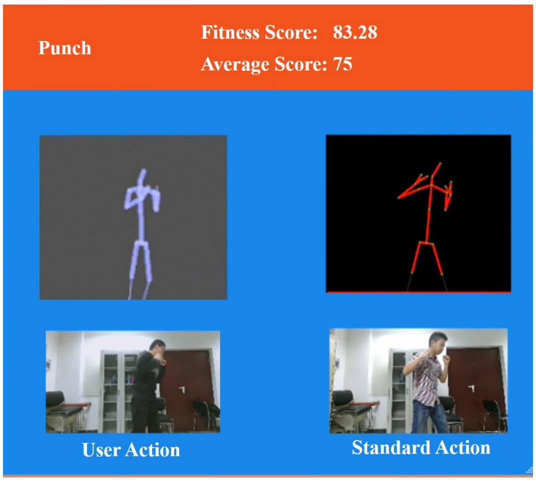
We test our system in several actions such as weightlifting, dance and punch. As shown in Figure 7, the fitness scores correspond well to the similarity of the two actions.



(a) Weightlifting



(b) Dance



(c) Punch

Figure 7 Demonstrations of our system.

5 Conclusion and Discussion

Our daily life has become more and more convenient with the progress of technology, people are losing more and more time and passion to get involved into physical exercise. Thus, it's the inevitable trend that many devices and apps will be developed to make exercising easier and accessible anywhere.

Based on the idea of making the life better and smart, our team presents a solution for building a virtual coach Kinect application. With this powerful tools, people can exercise anytime and anywhere they want. Although we have achieve our primary goal that make the comparisons between standard actions and real ones, the truly effective comparisons are always limited in the 2D ones due to some technical details of the machine learning algorithm used in the Kinect for Windows. We would discover more deeply in the future and try a new way to present the solution for 3D actions. Furthermore, to achieve a better user interface, we consider adding more element such as award mechanism and other plugins in the future.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.61402021, No.61402023), the Fundamental Research Funds for the Central Universities (No.2014XSYJ01, No.2015XSYJ25), the Science and Technology Project of the State Archives Administrator (No.2015-B-10), and Open Project Foundation of Information Technology Research Base of Civil Aviation Administration of China (No.CAAC-ITRB-201403)

Reference

[1] Massimo Camplani,Carlos Roberto Blanco, Luis Salgado, Fernando Jaureguizar, Narciso García. Advanced background modeling with RGB-D sensors through classifiers combination and inter-frame foreground prediction [J]. Machine Vision and Applications. 2014 (5)

[2] Da Wei Yu,Yue Long Hou,Xiao Li Wang. Research on Landscape Environment with 3D-Reconstruction and Volume Measurement of Fruit Tree Canopy Based on Kinect [J]. Advanced Materials Research. 2013 (788)

[3] Yue H,Chen W,Wu X, et al.Fast3D modeling in complex environments using a single Kinectsensor. Optics and Lasers in Engineering. 2014

[4] Benjamin John Southwell,Gu Fang.Human object recognition using color and depth information from an RGB-D Kinect Sensor. International Journal of Advanced Robotic Systems. 2013

[5] Shotton J,Fitzgibbon A,Cook M.Real-time human pose recognition in parts from single depth images. Proc.Computer Vision and Pattern Recognition. 2011

[6] Microsoft. The detection technology of the Visual Gesture Builder (VGB). https://msdn.microsoft.com/en-us/library/dn785523.aspx

[7] Yao, B.; Song-Chun Zhu, Learning deformable action templates from cluttered videos, IEEE 12th International Conference on Computer Vision. pp.1507-1514, Sept. 29 2009-Oct. 2 2009