tARget: Limbs Movement Guidance for Learning Physical Activities with a Video See-Through Head-Mounted Display

Ping-Hsuan Han
National Taiwan University

Jia-Wei Lin
National Taiwan University

Chen-Hsin Hsieh
National Taiwan University

Jhih-Hong Hsu
National Taiwan University

Yi-Ping Hung
Tainan National University of the Arts

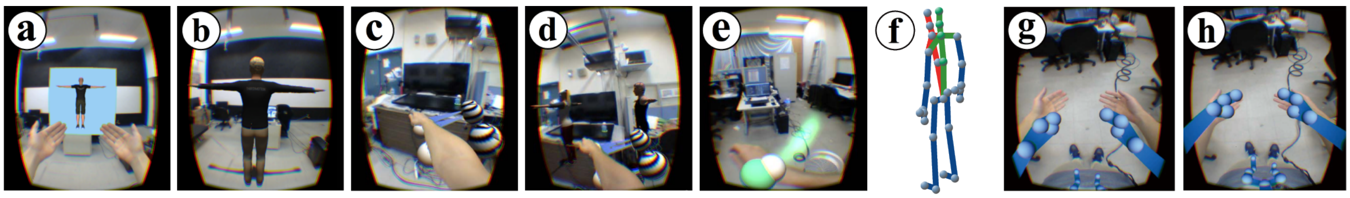

Figure 1: The design of visual guidance: (a) Video, (b) Coach-Surrounding, (c) Ball-Following, (d) Coach-Surrounding and Ball-Following. Additionally, (e) the design of movement hint for Ball-Following Guidance, (f) full-body skeleton detected by Kinect 2, (g) upper-body skeleton before calibration, (h) upper-body skeleton after calibration.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**;

KEYWORDS

Augmented Reality, Body Movement Guidance

ACM Reference format:
Ping-Hsuan Han, Jia-Wei Lin, Chen-Hsin Hsieh, Jhih-Hong Hsu, and Yi-Ping Hung. 2018. tARget: Limbs Movement Guidance for Learning Physical Activities with a Video See-Through Head-Mounted Display . In *Proceedings of SIGGRAPH '18 Posters, Vancouver, BC, Canada, August 12-16, 2018,* 2 pages.
https://doi.org/10.1145/3230744.3230776

1 INTRODUCTION AND MOTIVATION

In the aging society, people are paying more attention to having good exercise habits. The advancement of technology grants the possibility of learning various kinds of exercises using multi-media equipment, for example, watching instruction videos. However, it is difficult for users to learn accurate movements due to lack of feedback information.

To explore what augmented reality (AR) or virtual reality (VR) technologies can assist the users, lots of research on movement guidance and evaluation has been purposed and developed. Chua et al. [Chua et al. 2003] built a Tai Chi Chuan training system in virtual reality environment. They created a virtual coach in front of the user. Integrated with motion capture system, users can see their movement in VR and compare it to the coach's movement. However, these systems can only let the users see their avatars instead of themselves. LightGuide [Sodhi et al. 2012] used a projector hanging from the ceiling and projected visual information on users' hands to guide hand movements. It could only guide users in a fixed space with their hands being under the projection zone. OutsideMe [Yan et al. 2015] used Kinect to capture the skeleton, RGB and depth images of users. They enabled users to see their body movements as external observers through a video see-through head-mounted display (VST-HMD). AR-Arm [Han et al. 2016] showed semitransparent arms to indicate the correct movement of Tai Chi Chuan. Users could follow the virtual arms intuitively to achieve accurate arm movement.

In this paper, we present a full-body movement guidance system for learning physical activities with a VST-HMD. It contains a method for skeleton calibration and two interfaces for movement guidance: Coach-Surrounding Guidance and Ball-Following Guidance. We conducted a user study to evaluate the system on posture and movement learning.

2 APPROACH

For the VST-HMD, we combined Oculus Rift DK2 and two Logitech C310 cameras to build a binocular-cameras module like AR-Rift [Steptoe et al. 2014], and the distance between two lenses of the cameras is 6.4 cm. We applied camera calibration method [Zhang 2000] and rendered the hands as foreground by skin detection. These algorithms were implemented in shader for speeding up. Our system runs on a workstation equipped with i7 CPU and GTX980.

We detected the user's skeleton by Kinect 2 and modified the skeleton calibration based on [Han et al. 2016]. We found that when users bow their heads, Kinect cannot see their necks due to occlusion caused by the HMD, which causes the skeleton to lean forward (Figure 1(f)). Moreover, the upper-body skeleton cannot be displayed correctly because of this problem (Figure 1(g)). In this paper, we proposed a skeleton calibration method to solve this problem. Since the shoulders are not occluded by the HMD, we can find the correct position of the midpoint of the shoulders.

SIGGRAPH '18 Posters, August 12-16, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5817-0/18/08.
https://doi.org/10.1145/3230744.3230776

Table 1: The Distance Error (cm) of Postures and Movements

	(1) V	(2) CS	(3) BF	(4) CS+BF	RM-ANOVA	Tukey Post-hoc
Pos.1	17 (4)	11 (2)	8 (3)	7 (1)	F=22.63, p<0.001	1-2, 1-3, 1-4, 2-4
Pos. 2	28 (10)	26 (3)	19 (4)	19 (3)	F=6.45, p<0.05	1-3, 1-4
Pos. 3	17 (4)	18 (5)	14 (5)	14 (6)	F=1.18, p=0.334	N/A
Pos. 4	33 (11)	31 (8)	19 (4)	17 (2)	F=11.16, p<0.001	1-3, 1-4, 2-3, 2-4
Mov. 1	29 (10)	27 (10)	13 (4)	13 (5)	F=10.60, p<0.001	1-3, 1-4, 2-3, 2-4
Mov. 2	19 (8)	19 (6)	16 (9)	15 (8)	F=0.49, p=0.691	N/A
Mov. 3	22 (6)	21 (5)	19 (4)	18 (4)	F=2.11, p=0.118	N/A
Mov. 4	14 (3)	12 (2)	12 (3)	12 (3)	F=1.14, p=0.348	N/A

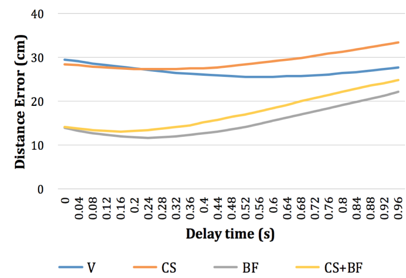

Figure 2: The Delayed Distance Error

Then the joint positions on the spine can be recalculated according to their heights because the hip, spine, and the midpoint of the shoulders should be on a straight line. As a result, we can calibrate the skeleton of upper-body (Figure 1(h)).

3 PRELIMINARY STUDY

We conducted a posture and movement guidance study with the following interfaces (Figure 1 (a) to (d)): (1) Video (V): baseline performance by watching video; (2) Coach-Surrounding (CS): eight virtual coaches surround the user; (3) Ball-Following (BF): there are ten white balls placed on the user's palms, wrists, elbows, knees and ankles, and another ten stripe balls guiding the correct joint positions. The stripe balls would turn green when the distance error is under 5 cm (Figure 1(e)); (4) Coach-Surrounding and Ball-Following (CS+BF): the combination of Coach-Surrounding and Ball-Following guidance. In this study, we recruited 9 participants (3 males and 6 females). Their ages range from 18 to 25 years old (mean = 21.0, SD = 2.4). In each interface, participants were asked to perform 4 postures and 4 movements chosen from CMU Motion Capture Database, which is simple and easy to understand. Movement 1 to 4 are movements performed by single arm, two arms, single leg, two legs, respectively. More details of the postures and movements are shown in the video sequence. Each posture task ends when the participant said they are done, and each movement task consists of 10 times of the movement. Participants were asked to finish the Video interface for the baseline performance. Then, half of them completed CS before BF, while the others in reverse order. After they finished, the last one was CS+BF, which is for observation of using the CS and BF interfaces simultaneously. In the end, we made an interview to investigate their preferences.

4 RESULTS AND DISCUSSION

The results are shown in Table 1. The significant differences in posture 1, 2, and 4 show the helpfulness of BF and CS+BF for learning postures. Nevertheless, among the movement tasks, the significant difference is only found in Movement 1 (single arm movements). Participants mentioned that it is difficult to follow the guiding balls on both arms at the same time, and they could not understand the guiding balls on the legs due to lack of depth information. Additionally, participants may need some time to react after seeing the guidance without feedforward. Taking Movement 1 for example, the distance error could be reduced if we calculate it after a short delay (as shown in Figure 2). The delay time that minimizes the distance error of interface (1) to (4) are 0.56, 0.28, 0.24, and 0.16, respectively, which may be viewed as the reaction time of users in each interface. In the interview, participants (7/9) prefer CS in the posture tasks, since they could understand the postures directly and finish them quickly. The finish time of postures in V and CS is about 10 seconds, while it takes about one minute in BF and CS+BF because the participants tend to adjust their postures carefully with the guiding balls. In the movement tasks, participants (8/9) prefer CS for the same reason in the posture tasks. Additionally, 4 participants mentioned that they can have better performance with CS+BF because they can see the rough movements from the virtual coaches and move accurately with the guiding balls.

5 CONCLUSIONS AND FUTURE WORK

In our user study, CS+BF has the best performance in both posture and movement tasks. The system can be applied to guide various physical activities such as dancing, physiotherapy, and martial arts. However, the video see-through cameras should be calibrated to the real position of user's eyes [Lai et al. 2016]. We will try to refine our method with RGBD cameras and conduct a further user study.

ACKNOWLEDGMENTS

This work was partially supported by the Ministry of Science and Technology of Taiwan under Grants MOST 104-2627-E-002-001 and 106-3114-E-369 -001.

REFERENCES

Philo Tan Chua, Rebecca Crivella, Bo Daly, Ning Hu, Russ Schaaf, David Ventura, Todd Camill, Jessica Hodgins, and Randy Pausch. 2003. Training for physical tasks in virtual environments: Tai Chi. In *Virtual Reality, 2003. Proceedings. IEEE*. IEEE, 87–94.

Ping-Hsuan Han, Kuan-Wen Chen, Chen-Hsin Hsieh, Yu-Jie Huang, and Yi-Ping Hung. 2016. Ar-arm: Augmented visualization for guiding arm movement in the first-person perspective. In *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, 31.

Chun-Jui Lai, Ping-Hsuan Han, Han-Lei Wang, and Yi-Ping Hung. 2016. Exploring Manipulation Behavior on Video See-Through Head-Mounted Display with View Interpolation. In *Asian Conference on Computer Vision Workshops*. Springer, 258–270.

Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 179–188.

William Steptoe, Simon Julier, and Anthony Steed. 2014. Presence and discernability in conventional and non-photorealistic immersive augmented reality. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*. IEEE, 213–218.

Shuo Yan, Gangyi Ding, Zheng Guan, Ningxiao Sun, Hongsong Li, and Longfei Zhang. 2015. OutsideMe: Augmenting Dancer's External Self-Image by Using A Mixed Reality System. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 965–970.

Zhengyou Zhang. 2000. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* 22, 11 (2000), 1330–1334.