https://doi.org/10.2352/ISSN.2470-1173.2019.2.ERVR-180

Both-hands motion recognition and reproduction characteristics in front/ side/ rear view

Tatsunosuke Ikeda, Yoshihiko Nomura, Ryota Sakamoto,
Department of Mechanical Engineering, Graduate School of Engineering, Mie University; Tsu, Japan

Abstract

In this paper, a "learning by observation" method, which is most commonly employed motion for learning, is examined. In the observation-based learning method, learners generally observe, recognize, and reproduce a model-performed reference motions only from one direction. Subjects can observe the model from various directions: the orientation of the model's trunk doesn't accord with that of the subjects when viewing from a direction other than from behind the model. It prevents the subjects from learning the model's reference motions easily because the subjects need to rotate the model mentally (it is called the "mental rotation"). On the other hand, when viewing from behind the avatar in order to avoid the mental rotation cost, subjects would occasionally encounter occlusion problems. Therefore, we have studied perceptual characteristics of various observation views through a psychophysical experiment. Two kinds of physical values were employed for evaluating subject's responses. One is the delayed time for reproduced motion onset, and the other is the error rate of reproduced motion direction. The results suggest that the perception suffers ill-effects from the mental rotation in two ways: the amount of the mental rotation increases the delayed time, and the presence of the mental rotation does the directional errors.

Introduction

Motion skills in sports such as swimming, throwing, dancing and writing are acquired through some learning methods. Observing motions with video images, photographs, computers, etc. is one of the representative learning methods. In this research, we focused on "motion learning by observation" using HMD (Head-Mount Display) [1]. In the learning method, usually, when observing 3D-CG images with an HMD, learners can observe model's motions with various views from surrounding positions. Thus, it is necessary to select some appropriate directions to observe the motions [2]. Incidentally, in the rear view where subjects observe the model from behind the model, a problem may arise that the model's arm movement hides behind their back. It is called the "*occlusion problem*." On the other hand, in the front view observed from the front of the model, learners need to rotate the expert in their mind (called the "mental rotation" [3]). Therefore, in this study, psychophysical experiments were conducted in order to investigate reproduction and perception characteristics for various observation views.

Experiment

An experiment was conducted in order to investigate the performance differences of motion recognition and reproduction in relation to the observation views. As the observation view factor, "the front view", "the side view" and "the rear view" were prepared as explained in the following section. An HMD, controller and tracker were equipped with subjects. The subjects were able to observe models as reference through the HMD. The subject were instructed to recognize and reproduce the reference motions in real time. The reproduced motions were measured and evaluated to compare the performances between the three views.

Subjects

Twelve people participated in this experiment (aged 21 – 24); they were recruited from Mie university. They have some HMD experiences before.

Observation view

Classification

In this chapter, three specific learner's views for observing reference motions presented by a model are explained.

1. Front view

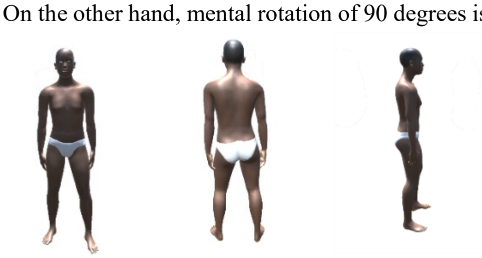
The front view is a view to observe the model's reference motions directly in front (Fig.1 (a)). It is the most natural view in the sense that we are used to observe the motion of the object with our own eyes. As a feature of this view, arm motions are less likely to be occluded by the torso.

2. Rear view

The rear view is a view to observe the model's reference motions from behind the model (Fig.1 (b)). As a feature of this view, it is not necessary for the learner to reverse the movement in the left-and-right fore-and-aft direction of the model as in the front view, which has learners recognize reference motions with mental rotations. However, there is a disadvantage that arm movement may be occluded by the model's back, which has learners suffer from difficulty in recognizing the movement of the models.

3. Side view

The side view is a view to observe the model's reference motions from the side (Fig.1 (c)). As the side view, there are two sides on the lateral side, left side and right side. It is considered that there is no difference between the left and right, and, therefore, only the view from the right side was taken. As a feature of this view, movement of limbs is less occluded by the trunk than in the rear view. On the other hand, mental rotation of 90 degrees is necessary.



(a) Front view (b) Rear view (c) Side view

Figure 1. View classification

Reference motion

As a simple motion, the reaching task, i.e., the horizontal planar straight-strokes were presented as a reference motion. In this experiment, in addition to the above-mentioned view factor, movement directions, movement distances, and movement speeds were employed as factors. We prepared four levels of movement direction from the initial position for both hands; they were right direction (-90 degrees), forward direction (0 degree), left direction (90 degrees), and backward direction (180 degrees) (Figure.1). In addition, from the initial positions of both hands, 4 levels of movement distances were employed; they were 0.02, 0.04, 0.08, and 0.16 [m]. 4 levels of speeds were employed; they were 0.05, 0.1, 0.2, and 0.4 [m/sec].

The reference motions were comprised of the minimum-jerk straight-line trajectories: it is assumed to be the most general trajectory representing human motions [4]. That is, the stroke motion trajectory of both-hands, $\mathrm{P}\,(t)$ was generated by the following equations.

$$\boldsymbol{P}(\mathrm{t}) = [x(t), y(t), z(t)] \quad (1)$$

$$\boldsymbol{P}_{\mathrm{start_i}} = (x_{start_i}, h_{stroke}, z_{start_i}) \quad (2)$$

$$\boldsymbol{P}_{\mathrm{end_i}} = (\,x_{end_i},\ \mathrm{h_{stroke_i}},\ \ \mathrm{z_{end_i}}). \quad (3)$$

$$x(\mathrm{t}) = x_{start_i} + \left(x_{end_i} - x_{start_i}\right) \cdot (6\tau^5 - 15\tau^4 + 10\tau^3), \quad (4)$$

$$y(t) = h_{stroke} \quad (5)$$

$$z(t) = z_{start_i} + \left(z_{end_i} - z_{start_i}\right) \cdot (6\tau^5 - 15\tau^4 + 10\tau^3) \quad (6)$$

Let us denote the start position and end position on the i^{th} stroke by $\boldsymbol{P}_{start_i}$ and $\boldsymbol{P}_{\mathrm{end_i}}$ as in the followings.

The next stroke's start position, $\boldsymbol{P}_{start_{i+1}}$, is identical with $\boldsymbol{P}_{end_i}$. The start position of the 1^{st} stroke was fixed. The end point of each stroke were decided following the above explained conditions . The variable, h_{stroke}, is a constant to give the both hand height: 1.4 [m] was set as h_{stroke} from the ground in this experiment. The variable, τ, is a parameter to give the instantaneous position at the elapsed time, t, from which the reference started to perform the current stroke and is defined as follows.

$$\tau = \frac{t}{T_i} \quad (7)$$

$$T_i = \frac{\sqrt{\left(x_{end_i} - x_{start_i}\right)^2 + \left(z_{end_i} - z_{start_i}\right)^2}}{v_i} \quad (8)$$

Where T_i is the time duration to perform the i^{th} stroke motion, and v_i is the average speed on the i^{th} stroke motion.

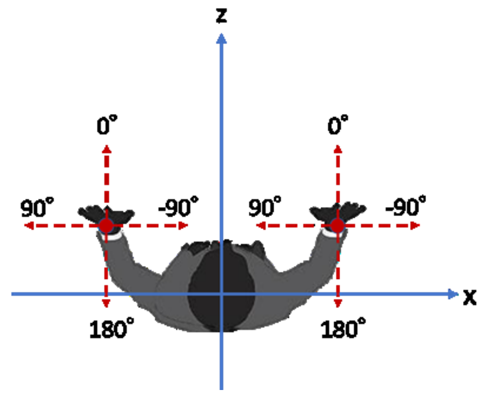


Figure 1. Top view of the model: red broken lines show moving directions.

Experimental device and software

In this research, HTC Vive HMD was used. It has SteamVR tracking, G-sensor, gyroscopes and proximity sensors, so that they measure wearer's head motions, and enhance wearer's immersivity by feedbacking the motions for displaying stereo images. The HMD display provides the resolution of 1080 × 1200 pixel image, the refresh rate of 90 Hz and the field of view of 110 degrees. Also, the learner wore a Vive tracker on their right wrist and held a Vive controller in their left hand. The tracker was used for measuring learner's hand motions in the experiment, and the controller was for learner's signaling completion of their experimental operations.

The software for the experiment was developed using Unity game engine. The experimental software was executed on a Windows desktop PC, and transmitted images to the HMD.

Figure 2. Experimental device: HTC Vive head mount display, tracker and controller were used in the experiment

Experimental procedure

At the beginning of the experiment, the learner put on the experimental devices. Then, the learner was to see a model's reference motions with each of the three observation views. The information board in the HMD screen showed the model's both hand positions and the subject's actual both hand positions as 3-dimensional coordinate values. Here, subjects were asked to do experimental trials: each of the trials was composed of the following experimental steps.

Step 1: Initial position matching

The subject saw the reference model who was static and kept a particular posture. By using the information bord, the subject was able to see the model's both hands position and the subject's actual both hands position as 3-dimensional coordinate values. Then, the subject reproduced the static posture by matching both of the coordinate values. This step realized the accurate initial position matching of the reference model's both-hands with the subject's ones.

Step 2: Recognition and reproduction of reference motions

After finishing the initial position matching step, the subject pulled and held a trigger of the Vive controller grasped in the subject right hand at their arbitrary timing. Then, the 3-dimensional coordinate values became invisible and the motion recognition and reproduction procedures started. That is, the reference began to move their both hands and the subject also started to recognize and reproduce the presented motions as early as possible.

Step 3: End of recognition and reproduction

When the reference finished reproducing model's reference motions, it stopped the both hands movements at the end position of the motions. As soon as the subject recognized the finish of the model's reference motions, they stopped pulling the trigger. Thus, the recognition and reproduction were finished.

The subject experienced 32 trials for each of the three observation views.

Evaluation methods

While the subject kept pulling the trigger of the controller, the subject's both-hand positions and the model's both-hand positions were recorded. Using the time-series data of both the positions, the reproduction accuracy and the time delay were evaluated. Hereafter, the latter was defined as a delayed time, and the former as a directional error.

Delayed time

The delayed time t_d was defined as the difference between the rising times of the learner and model because subject-reproduced distances aren't necessarily coincident to model's distances (Figure.3). We let the rising time be the time when moving distance reaches 10% the full stroke.

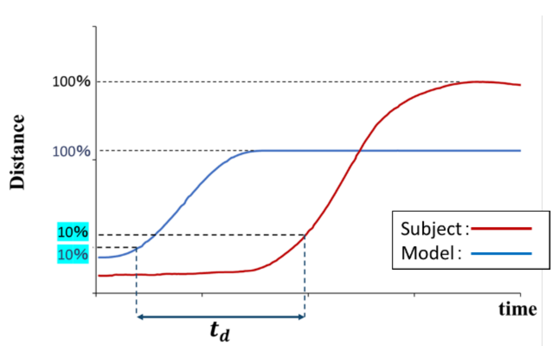


Figure 3. Example for time series data of distances.

Directional error

We defined the answer was correct if the angular differences between the learner-reproduced and the model's directions were within ±45 degrees.

Experimental results

The evaluation values were calculated by using the learner's and model's both-hand positions. Also, in order to judge whether the evaluation values are different depending on the observation views or not, a *t*-test was applied to the experimental data.

Delayed time

The mean values of delayed time with the rear view, the side view, and the front view were 523, 558, and 803 [ms], respectively.

The t-test reveals a significant difference between the rear view and the side view (t = 3.05, p = 0.0038 < 0.01). The difference between the side view and the front view was also significant (t = 14.85, $p = 3.0 \times 10^{-43} < 0.001$). The difference between the rear view and the front view was significant (t = 17.9, $p = 2.55 \times 10^{-59} < 0.001$).

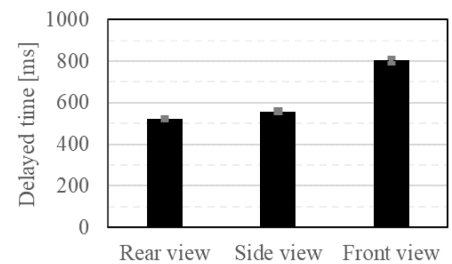


Figure 4. Mean delayed time (error bar. standard error)

Directional error rate

The directional error rate with the rear view, the side view, and the front view were 0.12, 0.17, and 0.18, respectively. The t-test reveals a significant difference between the rear view and the side view (t = 4.72, $p = 6.82 \times 10^{-6} < 0.001$). The difference between the side view and the front view is not significant (t = 1.47, p = 0.136 > 0.05). The difference between the rear view and the front view is significant (t = 6.19, $p = 3.03 \times 10^{-9} < 0.001$).

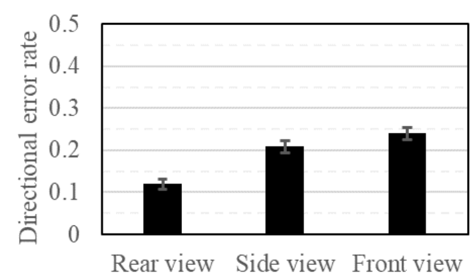


Figure 5. Rate of the directional error (error bar. standard error)

Conclusion

The rear view is the best in the front, side, and rear views for observing models. The rear view takes only about half time of the front view. The rear view improved the directional error rate to about half the other views.

This work was supported by KAKENHI (Grant-in-Aid for scientific research (B) 19H02929 from JSPS).

Reference

[1] J. Bailenson, K. Patel, A. Nielsen, R. Bajcsy, S. Jung, G. Kurillo, “The Effect of Interactivity on Learning Physical Actions in Virtual Reality”, Journal Media Psychology Volume 11, 2008 - Issue 3.

[2] A. E. N. Hoover, L. R. Harris, “The role of the viewpoint on body ownership”, Exp Brain Res (2015) 233:1053–1060

[3] R. N. Shepard, J. Metzler, “Mental Rotation of Three-Dimensional Objects”, Science, New Series, Vol. 171, No. 3972. (Feb. 19, 1971), pp. 701-703.

[4] Y. Nomura, H. Fukuoka, R. Sakamoto, T. Sugiura, “Motion Instruction Method Using Head Motion-Associated Virtual Stereo Rearview”, Technologies (2017), 5(4), 77

Author Biography

Tatsunosuke Ikeda. He is a graduate student in the Dept. of Mechanical Eng. Grad. School of Eng. at Mie University. He was given his Bachelor’s Degree at Mie University in 2018. His research interests lie in the area of human motion recognition and reproduction characteristics from visual information, specially, those by the virtual reality technology.

Yoshihiko Nomura, Ph.D. He is a Professor in the Dept. of Mechanical Eng. Grad. School of Eng. at Mie University where he has been a faculty member since 1997. During the period, he had also served as an Executive Vice President in Educational Development from 2007 to 2011. His research interests lie in the area of mechatronics and their application to intelligent robots, ranging from theory to design to implementation.