https://doi.org/10.2352/ISSN.2470-1173.2018.09.IRIACV-125

Recognition and reproduction performance of hand motions with HMD-based motion learning method

Shin Kinoshita, Yoshihiko Nomura, Ryota Sakamoto, Tokuhiro Sugiura
Department of Mechanical Engineering, Graduate School of Engineering, Mie University; Tsu, Japan

Abstract

Particular motions are important to play sports with high performance. The particular motions are mastered by learning motions, and visual information is considered to be effective for understanding and learning motions. In recent years, HMD with VR has been introduced as a new tool for learning motions with visual information. An advantage of the HMD-based motion learning method is that it enables learners to switch their observation view. Here, this research investigates basic view characteristics of observing and reproducing particular dynamic motions, which would be necessary to develop some methods for switching observation view properly. An experiment was conducted in order to study the basic view characteristics. As for the observation view factor, we prepared two factor levels, one was the front mirror view, and the other the rear camera view. In the experiment, a subject recognized and reproduced some reference dynamic motions on real time with each of the two views. The experimental results revealed that the reproduction performance with the rear camera view was significantly better than that with the front mirror view in the case of the depth-directional motions, compared with the other case of the depth-uncorrelated motions. It should be noted that the difference in the motion reproduction may become crucial for learners in particular as the motion velocity increases. It is supposed that the observation with the front mirror view requires some mental transformation operation when the learners reproduce motions. In selecting the observation view, it is required to minimize the mental transformation operation. The requirement is expected to be satisfied with the rear camera view, provided that occlusions are not crucial for learners to observe reference motions.

Introduction

Mastering some particular motions such as those in ball games, martial arts, and dances is quite important to play them with high performance. To acquire motion skills, we have learned motions by observing and imitating expert motions through videos and photos. Many researchers have studied such kind of vision-based motion learning method and demonstrated that visual information is effective to understand and learn motions.

In recent years, the vision-based motion learning methods have been changed as a result of technological progress of virtual reality (VR) technologies: the progress replaced the conventional device such as videos and photos with head mount displays (HMDs) to provide visual information.

For example, Bailenson et al. compared the VR-based motion learning method with the conventional video-based one. Then, it was proved that learners acquired more immersive feeling and recognized reference motions more accurately with the VR method than with the conventional one [1]. Roosink et al. demonstrated that VR has a potential for motion learning from the standpoint of the recognition accuracy [2]. Thus, it is considered that the HMD-based motion learning method with VR is superior to the conventional ones from the following points.

1. Interactivity

The sensors built in HMDs detect their wearer's movements, and the movement information can be transmitted to the motion learning system as feedback signals. Then, the wearer is able to get visual information on which the movements are reflected.

2. Immersivity

HMDs provide binocular stereopsis to their wearer. The feature encourages the wearer to recognize particular motions better: in particular, it is supposed to be effective when the wearer focuses on depth-directional motions, which are generally difficult to be recognized through conventional videos.

3. View switching

It is difficult to change view angle of reference motions on videos, although applying a proper view for motions is considerably crucial. While, HMDs enable the wearer to switch their view to observe reference motions.

Among the three advantages of VR on motion learning, this research focuses on the third feature of view switching. The reason is that it is quite a difficult task to master motions in the case that the motions are composed of instantaneously successive postures and just a few duration time is given to observe each of the postures. In the case, applying a proper observation view is particularly important to get learners easier to recognize the dynamic motions. Also, the proper views differ depending on many factors such as postures and movements of the reference, the level of proficiency of the learners, etc. Based on these features, it is supposed that implementation of view-switching is a key function to develop the HMD-based motion learning methods with VR in the future. Here, this research investigates basic view characteristics of observing and reproducing particular motions that would be necessary to develop a view-switching function.

Observation view

Classification

In this chapter, two specific learner's views for observing reference motions presented by an avatar are introduced.

1. Rear camera view

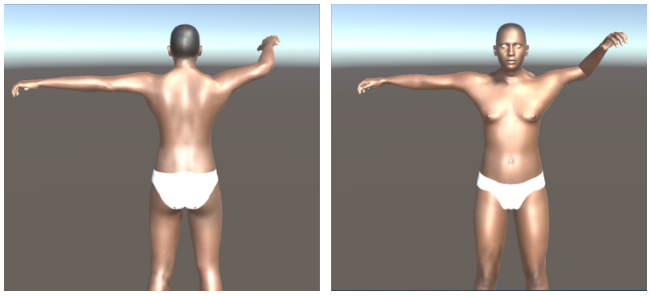
Ordinarily, we observe the motions of another person, i.e., an avatar in the situation of this paper, at a frontal position about them: it is called the front camera view in the followings. On the contrary, the rear camera view is a specific view with which learners directly observe avatar's reference motions at a rear position (Fig. 1 (a)). This view allows learners to observe reference motions in the same coordinate system as the learner-centered one:

it means that, different from the front camera view, neither the right-and-left inversion nor the front-and-back inversion does not occur, and the learner does not need to perform mental transformations. However, it has a disadvantage that learners suffer difficulties when observing reference limb parts occluded by their trunk.

2. Front mirror view

The front mirror view is another view with which learner can see horizontally inverted reference motions in a frontal position (Fig. 1 (b)). This view has learners observe reference motions in the mental mirror-transformed coordinate system: although the front-and-back inversion as in the front camera view is remained, the right-and-left inversion is removed. It, in most cases, enables learners to observe reference motions without any occlusions of limb parts behind their trunk as in the front camera view. Fraser et al. adopted this view for their proposed motion learning system, YouMove [3].

While the front camera is usually applied to videos and movies for motion learning. We, however, consider that the rear camera is superior to the front camera on a situation that learners are trying to master motions except that body parts were not occluded by other ones. This is based on an assumption that the front camera involves the front-and-back inversion. These inversions could impose some mental transformations on learners to recognize the moving direction. That's why, it is predicted that observation with the front mirror view requires more time to recognize and reproduce presented motions than the rear camera.



(a) Rear camera view (b) Front mirror view

Figure 1. View classification

Experiment

An experiment was conducted in order to investigate the performance differences of motion recognition and reproduction in relation to the observation views in VR environments. As the observation view factor, the front mirror view and the rear camera were prepared as explained in the preceding section. One subject participated in the experiment as a learner: an HMD, controller and tracker were equipped with them. The learner was able to observe an avatar as a reference through the HMD. The subject was instructed to recognize and reproduce the reference motions in real time. The reproduced motions were measured and evaluated to compare the performances between the two views.

Reference motions

Some right hand strokes were selected as reference motions. For deciding the reference motions, a moving area of the right hand was set, considering the motion range limitation of the subject's arm. The *x*-coordinate of the left- side boundary of the area was set at the same value as that of the learner's shoulder position in order to avoid the occlusion of the right hand behind the avatar's head and trunk (Fig. 2).

The reference motions were comprised of six strokes. The minimum-jerk straight-line trajectories were employed: it is assumed to be the most general model representing human motions. That is, each of the stroke motion trajectory, $P_R(t)$ was generated by the following equations.

$$P_L(t) = (x_L(t), y_L(t), z_L(t)), \tag{1}$$

$$x_L(t) = x_{start\_i} + (x_{end\_i} - x_{start\_i}) \cdot (6\tau^5 - 15\tau^4 + 10\tau^3), \tag{2}$$

$$y_L(t) = h_{stroke}, \tag{3}$$

$$z_L(t) = z_{start\_i} + (z_{end\_i} - z_{start\_i}) \cdot (6\tau^5 - 15\tau^4 + 10\tau^3). \tag{4}$$

Let us denote the start position and end position on the i^{th} stroke by P_{start_i} and P_{end_i} as in the followings.

$$P_{start\_i} = (x_{start\_i}, h_{stroke}, z_{start\_i}), \tag{5}$$

$$P_{end\_i} = (x_{end\_i}, h_{stroke}, z_{end\_i}). \tag{6}$$

The next stroke's start position, $P_{start\_(i+1)}$, is identical with P_{end_i}. The start position of the 1^{st} stroke and the end point of each stroke were randomly given inside the moving area. The variable, h_{stroke}, is a constant to give the right 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.

$$\tau = \frac{t}{T_i}, \tag{7}$$

$$T_i = \frac{\sqrt{(x_{end\_i} - x_{start\_i})^2 + (z_{end\_i} - z_{start\_i})^2}}{v_i}. \tag{8}$$

Where T_i is the duration time to perform the i^{th} stroke motion, and v_i is the average speed on the i^{th} stroke. Six levels of speeds were employed; they were 0.1, 0.2, 0.3, 0.4, 0.5 and 0.6 [m /sec].

Experimental device and software

To constitute a VR environment, HTC Vive HMD was adopted. 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 his 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 specific 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 connected to the PC.

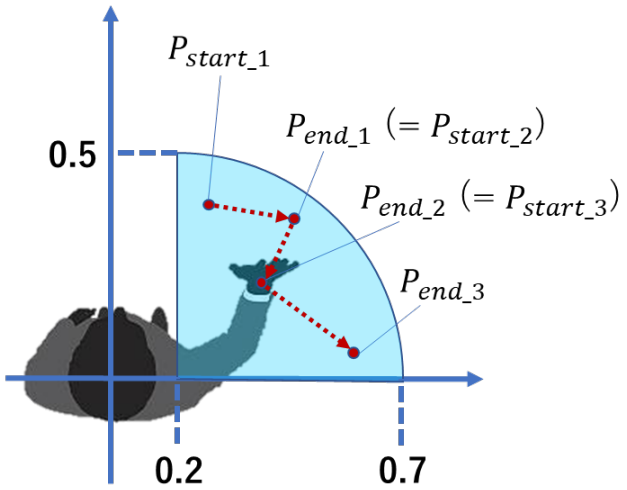


Figure 2. Movable area: The reference motions were presented as the reference's right hand moved inside the above blue-colored area. And the reference motions were comprised of some stroke motions. As an example, the start position and the end position of each stroke are shown, randomly given inside the blue-colored area.

Figure 3. 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 reference with each observation view. The information board showed the reference's right hand position and the learner's actual right hand position as 3-dimensional coordinate values. Here, the learner was asked to do experimental trials: each of the trials was composed of following experimental steps.

Step 1: Initial position matching

The learner saw the reference avatar who was static and was keeping a particular posture. Also, the learner was able to see the reference's right hand position and the learner's actual right hand position as 3-dimensional coordinate values. Then, the learner reproduced the static posture by matching the coordinate values. This step realized the accurate initial position matching of the reference avatar's and learner's right hand.

Step 2: Recognition and reproduction of reference motions

After finishing the initial position matching step, the learner pulled and held a trigger of the Vive controller on the learner left 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 right hand and the learner 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 reference motions, it stopped its right hand movement at the end position of the motions. As soon as the learner observed the finish of the reference motions and completed motion reproduction, the learner stopped pulling the trigger. Thus, the recognition and reproduction were finished.

The learner experienced 54 trials for each of the two observation views.

Evaluation methods

While the learner kept pulling the trigger of the controller, the learner's right hand position and the reference's right hand position were recorded. Using the time-series data on both positions, the reproduction accuracy and the time delay were evaluated. Here, the latter was defined as a phase error, and the former as a position error.

Phase error

Phase error represents the delay time of the learner motion from the reference motion. The phase error, E_{phase}, is calculated by the following equations.

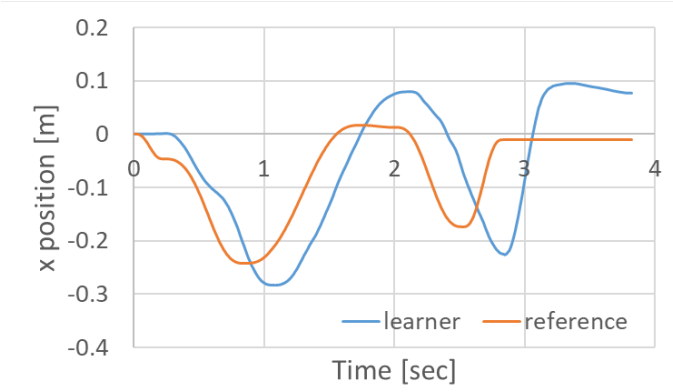
$$E_{phase} = T_s \cdot \hat{j}_{delay} \tag{9}$$

$$\hat{j}_{delay} = argmin_{j_{delay}}\{{E_{x\_pos}}^2 + {E_{z\_pos}}^2\} \tag{10}$$

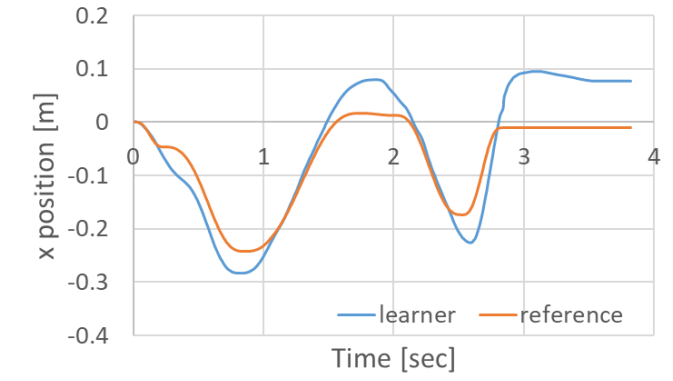
$$E_{x\_pos} = \sqrt{\frac{\sum_{j=0}^{n}(x_L^{\,j+j_{delay}} - x_R^{\,j})^2}{n}} \tag{11}$$

$$E_{z\_pos} = \sqrt{\frac{\sum_{j=0}^{n}(z_L^{\,j+j_{delay}} - z_R^{\,j})^2}{n}} \tag{12}$$

Where T_s is the sampling time of each record, $\hat{j}_{delay}$ is an estimated gap index that represents the learner position delay from the reference. The variables, E_{x_pos} and E_{z_pos}, are horizontal-directional and depth-directional RMSE between the learner and reference positions, respectively. They are calculated using the sampled data ($j = 0, 1, \ldots, n$) in a stroke. The variables, x_L^j and z_L^j, are the j^{th} learner's x- and z-coordinate in a stroke. The variables, x_R^j and z_R^j, are the j^{th} reference x- and z-coordinate in a stroke.



(a) The measured x positional trajectories of the learner and reference



(b) The adjusted x positional trajectories of the learner and reference

Figure 4. An operation of trajectory adjustment: (a) shows the measured x positional trajectories that were recorded from Vive tracker. (b) shows the adjusted x positional trajectories. In the adjustment operation, the learner x positional trajectories in (a) was shifted to minimize E_{x_pos} and E_{z_pos}. Consequently, the time delay of the learner motion from the reference one is cancelled.

Position error

Position error totally represents the differences between the learner's and reference's right hand position in a stroke. Here, since the position error is separated into x- and z-position error, it is expected that the recognition and reproduction characteristics are different depending on the reference motion directions. By the condition of $j_{delay} = \hat{j}_{delay}$, the ill-effect of the time delay on the RMSEs between the learner and reference positions were removed, and they were denoted by $E_{x\_pos|\hat{j}_{delay}}$ and $E_{z\_pos|\hat{j}_{delay}}$: with, and are shown in Fig. 4.

Experimental results

The evaluation values were calculated by using the learner's and reference's right 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.

Phase error

The mean values of phase error with the front mirror view and the rear camera view were 393 and 308 [ms], respectively. The t-test reveals a significant difference ($t = 2.50$, $p = 0.013 < 0.05$). It is assumed to arise from the fact that the front-and-back inversion occurs in the front mirror view while not in the rear camera view.

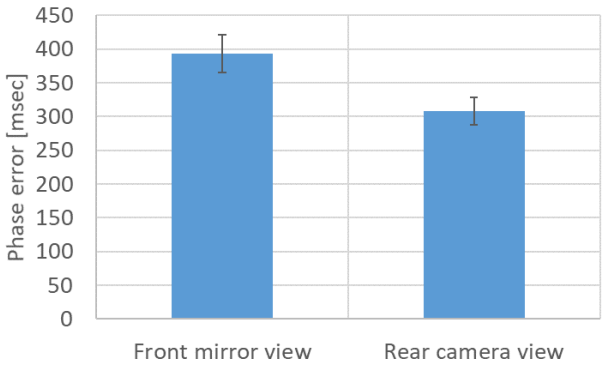


Figure 5. Mean phase errors (error bar: standard error)

x-position error

The mean values of the optimized x-position error in the front mirror view and the rear camera view were 0.029 and 0.027 [m], respectively. The t-test reveals no significant difference ($t = 1.68$, $p = 0.093$). It means that there is no difference on reproduction accuracies between the front mirror view and the rear camera view.

z-position error

The mean values of the optimized z-position error in the front mirror view and the rear camera view were 0.043 and 0.031 [m], respectively. The t-test reveals a significant difference ($t = 5.60$, $p = 4.53 \times 10^{-8} < 0.001$). It means that the reproduction accuracy in the rear camera view is significantly better than that in the front mirror view.

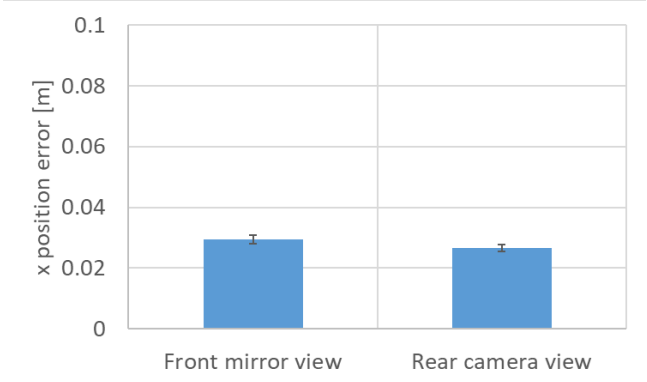


Figure 6. Mean x-position errors (error bar: standard error)

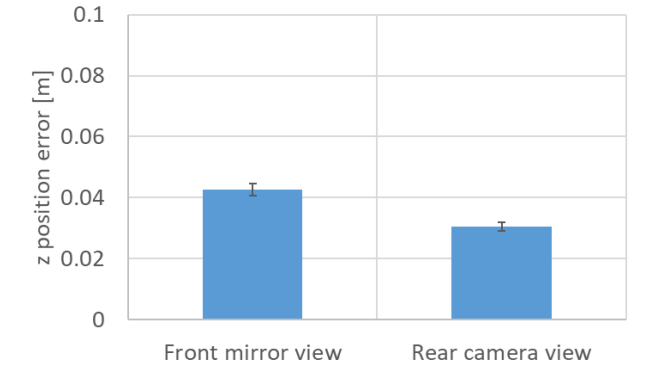


Figure 7. Mean z-position errors (error bar: standard error)

Conclusion

The experimental results show a possibility that the motion recognition with the front mirror view causes significant time-delay and the accuracy decreases of reproducing the reference motions compared to the rear camera view. It seemed to be caused by the front-and-back inversion while the learner was recognizing and reproducing the reference's motions with the front mirror view. The results suggest that the rear camera is effective in some situations. It is assumed that, when the learner observing the reference motions, non-necessity of any mental rotations is important to recognize and reproduce them immediately and accurately.

The learners are not able to observe some body parts with just one observation view, when the body parts being occluded behind other parts. At the case, switching the observation view provide a solution, it seems to have an advantage on motion learning methods with VR systems. In particular, the rear camera is better than the ordinary front camera view as long as any body parts are not occluded. In the future, the authors continue to examine the multi-view characteristics by psychophysical experiments with enough large sample size, i.e., enough number of subjects.

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] M. Roosink, N. Robitaille, B. J. McFadyen, L. J. Hébert, P. L. Jackson, L. J. Bouyer, C. Mercier, "Real-time modulation of visual feedback on human full-body movements in a virtual mirror: development and proof-of-concept", Journal of NeuroEngineering and Rehabilitation 2015 12:2.

[3] F. Anderson, T. Grossman, J. Matejka, G. Fitzmaurice, "YouMove: Enhancing Movement Training with an Augmented Reality Mirror", UIST '13 Proceedings of the 26th annual ACM symposium on User interface software and technology Pages 311-320.

[4] J. V. Stone, "Object recognition: View-specificity and motion-specificity", Vision Research, 39 (1999), 4032-4044.

[5] M. Lesourd, J. Navarro, J. Baumard, C. Jarry, D. L. Gall, F. Osiurak, "Imitation and matching of meaningless gestures: distinct involvement from motor and visual imagery", Psychological Research (2017) 81:525–537.

[6] M. V. Elk, O. Blanke, "Imagined own-body transformations during passive self-motion", Psychological Research (2014), 78(1), 18-27.

[7] A. B. Yu1, J. M. Zacks, "How Are Bodies Special? Effects Of Body Features On Spatial Reasoning", Q J Exp Psychol (Hove). 2016 June, 69(6): 1210–1226.

[8] S. Thorpe, D. Fize, C. Marlot, "Speed of processing in the human visual system.", Nature 381, 520–522 (06 June 1996)

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

[10] E. H. van Asseldonk, M. Wessels, A. H. Stienen, F. C. van der Helm, H. van der Kooij, "Influence of haptic guidance in learning a novel visuomotor task.", J Physiol Paris. 2009 Sep-Dec, 103(3-5):276-85.

[11] G. Macauda, G. Bertolini, A. Palla, D. Straumann, P. Brugger, B. Lenggenhager, "Binding body and self in visuo-vestibular conflicts", European Journal of Neuroscience, pp. 1–8, 2014

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

Author Biography

Shin Kinoshita. He is a graduate student in the Dept. of Mechanical Eng. Grad. School of Eng. at Mie University. He was given his Bachelor Degree at Mie University in 2015. 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.