

An Empirical Evaluation of Virtual Hand Techniques for 3D Object Manipulation in a Tangible Augmented Reality Environment

Taejin Ha* and Woontack Woot
GIST U-VR Lab., 500-712, S. Korea

ABSTRACT

In this paper, we present a Fitts' law-based formal evaluation process and the corresponding results for 3D object manipulation techniques based on a virtual hand metaphor in a tangible augmented reality (TAR) environment. Specifically, we extend the design parameters of the 1D scale Fitts' law to 3D scale and then refine an evaluation model in order to bring generality and ease of adaptation to various TAR applications. Next, we implement and compare standard TAR manipulation techniques using a cup, a paddle, a cube, and a proposed extended paddle prop. Most manipulation techniques were well-modeled in terms of linear regression according to Fitts' law, with a correlation coefficient value of over 0.9. Notably, the throughput by ISO 9241-9 of the extended paddle technique peaked at around 1.39 to 2 times higher than in the other techniques, due to the instant 3D positioning of the 3D objects. In the discussion, we subsequently examine the characteristics of the TAR manipulation techniques in terms of stability, speed, comfort, and understanding. As a result, our evaluation process, results, and analysis can be useful in guiding the design and implementation of future TAR interfaces.

KEYWORDS: Augmented reality, tangible user interface, virtual hand technique, 3D object manipulation, empirical evaluation, Fitts' law

INDEX TERMS: H.5.1 [Information Interfaces and Presentation]: Multimedia information systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User interfaces—Input interaction styles;

1 INTRODUCTION

Tangible augmented reality (TAR) [1] combines the intuitiveness of tangible user interfaces (TUI) [2], a physical object with physical constraints that restrict how it can be manipulated and used in an augmented reality (AR) environment. In a TAR environment, a user normally in an egocentric view of AR can interact easily with virtual objects by exploiting the TUI-based direct manipulation with virtual hand techniques in the following applications: furniture placement, urban planning, AR books, education, maps, and many other desktop-based AR applications.

The central aspect of TUI-based AR manipulation is 3D object selection and translation/rotation methods using a manipulation prop. Standard techniques within the TUI/AR community have

used paddle- [3], cup- [4], or box-shaped [5] props. 3D object selection is enabled by a collision/proximity process between the prop and a 3D object, generally on a 2D plane. 3D object translation/rotation usually is completed using a tilting, dropping, or hiding gesture of the prop.

However, many manipulation techniques have been arbitrarily implemented in many AR applications without formative evaluation or guidelines [6-7]. Some papers have studied the evaluation of TAR interaction techniques [8-12]; they mostly focus on selection methods without including translation/rotation for manipulation and interaction of 3D objects. Also, it is necessary to compare widely used TAR manipulation techniques, besides comparisons with the traditional keyboard, mouse, and wand, because there are some special characteristics for TAR with respect to an egocentric view with a real camera image and a tangible manipulation prop with fiducial-purpose tracking.

Therefore, in this paper, for the formative evaluation of 3D object manipulation techniques in 3D space in a tangible AR environment, we discuss Fitts' law-based [13] empirical comparison evaluations. This method is successful and grand, using an evaluation model for many existing pointing techniques. We adopt this evaluation technique to test 3D object selection and manipulation of an AR object, which are the most basic and frequently used manipulation techniques in TAR. For that, we conducted an evaluation design using a cup, paddle, cube, and an extension of the paddle. Through the performance and questionnaire-based subjective evaluations, we discuss the characteristics of manipulation techniques for 3D positioning of a 3D object considering stability, speed, comfort, and understanding. We then provide design suggestions for future implementation of TAR applications.

The contents of this paper are as follows. Related work is reviewed in section 2, including 3D object manipulations and an evaluation model. We then explain the technical details of the 3D object manipulation method in section 3. The evaluation process, comparison technique implementation, and evaluation results are then explained in section 4. Finally, a discussion and conclusion are presented in section 5.

2 RELATED WORK

2.1 3D Object Manipulation Technique

In a 3D user interface, 3D object selection and manipulation methods in egocentric techniques generally are categorized into laser pointing (virtual ray metaphor) and physical interaction (virtual hand metaphor) methods [14-15].

For the laser pointing, an object at a distance can be selected. However, variation with a small angle of manipulation in the prop can make a significant difference at a far distance, so it is difficult to select a small object. In order to solve this problem, spotlight [16], aperture-based [17], image-plane-based [18], or depth ray [19] methods may be used. However, the most critical limitation of the ray-casting technique is the lever arm. In fact, one DOF rotation around the ray axis is independently controllable; however, other manipulation depends on the other axis. In order to

* e-mail: tha@gist.ac.kr

† e-mail: wwoo@gist.ac.kr

This research was supported by the CTI development project of KOCCA, MCST in S.Korea.

solve this problem, the “fishing reel” metaphor [20] is proposed, but it is still an open problem for a high degree of free manipulation.

For 3D object selection and movement, a virtual hand technique is widely used, especially within the reach area [15], through direct mapping between the real and virtual hand or nonlinear mapping [21]. An interaction in VR can be very similar to real life, sharing the same spatial reference.

Even though most VR techniques can be applied to AR, there are special characteristics for AR with respect to an egocentric view with a real camera image and a tangible manipulation prop with fiducial purpose tracking. Previous studies on frequently used TAR manipulation techniques have used a stick- [3] or a cup-shaped prop [4] with a camera tracking fiducial purpose on top of them. In this case, 3D object selection is enabled by collision/proximity between the prop and a 3D object, only on a 2D plane. 3D object translation/rotation can be completed using a tilting or dropping gesture of the prop with respect to the specific height based only on the 2D plane. In [5], the fiducial markers are attached to each side of a cubical manipulation prop. To trigger a 3D object selection event, collision detection between an augmented short virtual line tip on the prop and a 3D object must be conducted for a specified period of time; to release the 3D object, the user also must hide the marker with a hand for a specific period of time. In section 3, we will study each manipulation technique in more detail.

2.2 Evaluation Model

Fitts’ law is a mathematical model used in cognitive science research to anticipate time to point along a 1D scale [13]. In this law, an index of difficulty (ID : bits unit) is used to determine the level of difficulty of a task based on two factors as shown in Equation 1: a target amplitude (A), referring to the distance from an initial position to a target position, and a target width (W). If the distance of the two positions is longer (higher A) or the target width narrower (lower W), the ID of the task is higher. In the conducted experiments, the movement times (MT) were measured repeatedly according to the various ID s. We used Shannon formulation for better fit [22] and linear regression to find the intercept (a) and slope (b) parameters of Equation 2.

$$ID = \log_2 \left(\frac{A}{W} + 1 \right) \quad (1)$$

$$MT = a + b \times ID \quad (2)$$

There have been trials to extend the 1D scale to multivariate dimensions. ID is a combination of multiple dimensions. This claim was explored in 2D by varying the angle (θ), width (W), and height (H) [23]. For 3D, this was extended to include the additional dimension (D) [24] in desktop VR environments and to accommodate movement/approach angles [25] in volumetric displays. Equation 3 shows the baseline 3D evaluation model. In our case, we deal with a cube-shaped 3D object, so that W , H , and D have the same value. The removal of the direction constraint makes the amplitude remain on the one-dimensional scale, as shown in Equation 1. Fewer parameters can bring generality and ease adaptation in various applications [23].

$$ID_{\min} = \log_2 \left(\frac{A}{\min(W, H, D)} + 1 \right) \quad (3)$$

ISO 9241-9 recommends an enhanced efficiency measure, called a throughput (TP : bit/sec unit) [26], as shown in Equation 4, which introduces an effective target width (W_e) considering error computation [27]. In Equation 6, W_e contains a standard deviation

of the difference between the ideal target position and the user’s final position. For a 3D environment, we define the ideal target position as shown in Equation 8.

$$TP = \frac{ID_e}{MT} \quad (4)$$

$$ID_e = \log_2 \left(\frac{A}{W_e} + 1 \right) \quad (5)$$

$$W_e = 4.133SD \quad (6)$$

$$SD = \sqrt{\frac{\sum_{i=1}^n \left[(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2 \right]}{n-1}} \quad (7)$$

$$x_i = A \cos(v) \cos(u), y_i = A \cos(v) \sin(u), \\ z_i = A \sin(v), \text{ where } v = u \quad (8)$$

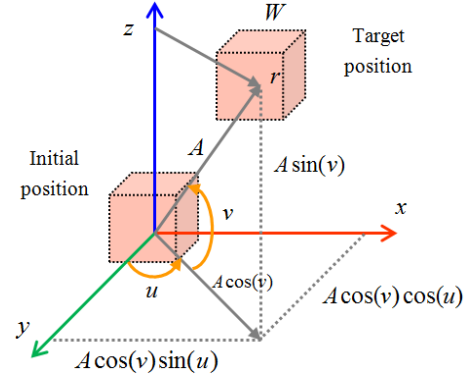


Figure 1. Arrangement of the 3D objects in 3D space.

3 3D OBJECT MANIPULATION TECHNIQUES IN A TANGIBLE AUGMENTED REALITY ENVIRONMENT

In this section, we provide a theoretical examination for the CUP, PADDLE, CUBE, and ExPADDLE techniques in detail. Note that some variations of implementation are possible for the purpose of the application.

CUP Technique. The CUP manipulation technique involved a picking and dropping gesture and was used for the 3D object selection and translation/rotation manipulations based on a 2D plane. The cylinder-shaped cup prop had a height of constant number HEIGHT and a radius RADIUS. A fiducial marker was attached to the top of the cup. The selection condition was such that the 3D object was not in the CUP prop and the position of the CUP was close to the boundary of an object, as shown in Equation 9. For the dropping condition, the 3D object was in the CUP prop and the vertical height was more than T_HEIGHT based on the plane, as shown in Equation 10, where \mathbf{M} is the transformation matrix.

Also during the translation and rotation, a position property from \mathbf{M}_{cup} is obtained from the frame most recently inserted into a local coordinate transformation matrix of the 3D object. Additionally, the rotation matrix of the 3D object is established using a relative angle from a rotation matrix of \mathbf{M}_{cup} in the previous frame to the rotation matrix in the current frame. This relation is shown in Equation 11, where \mathbf{T} and \mathbf{R} are the translation and rotation matrices of \mathbf{M} respectively. The notation convention used in this paragraph is applied in the following techniques.

Figure 2 shows the coordinate relations between the 3D object (\mathbf{M}_{obj}), the fiducial ($\mathbf{M}_{\text{fiducial}}$), and a CUP prop (\mathbf{M}_{cup}) based on a world coordinate ($\mathbf{M}_{\text{world}}$). \mathbf{M}_{cup} is simply translated from the fiducial coordinate ($\mathbf{M}_{\text{cup}} = [\mathbf{T}] \mathbf{M}_{\text{fiducial}}$).

Selection: if $\text{distance}(\mathbf{M}_{\text{cup}}, \text{boundary}(\mathbf{M}_{\text{obj}})) < \text{RADIUS}$
and there is no selected object (9)

Release: if $\text{height}(\mathbf{M}_{\text{fiducial}}) > T_HEIGHT$ and there is a
selected object (10)

Rotation and translation: $[\mathbf{RT}]\mathbf{M}_{\text{obj}(t)} =$
 $[\mathbf{R}]\mathbf{M}_{\text{obj}(t-1)}([\mathbf{R}]\mathbf{M}_{\text{cup}(t-1)}^{-1}[\mathbf{R}]\mathbf{M}_{\text{cup}(t)})[\mathbf{T}]\mathbf{M}_{\text{cup}(t)}$ (11)

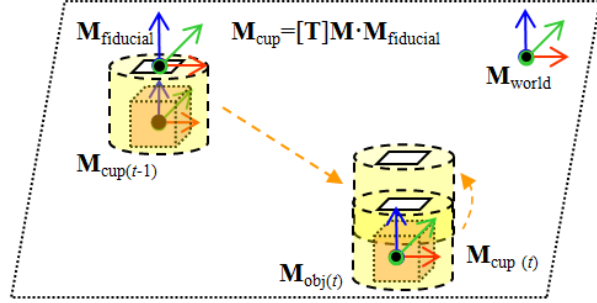


Figure 2. CUP technique in the 2D arrangement of 3D objects (2D_CUP).

PADDLE Technique. In the PADDLE manipulation technique, a user can conduct 3D object selection and manipulation based only on the 2D plane by grasping the prop and using the picking and tilting gesture (Figure 3). The 3D object can be selected and positioned on the prop when they are close together, as shown in Equation 12. The user can then move the prop to the target position and drop the 3D object onto the 2D plane using the tilting gesture, as shown in Equation 13. Rotation and translation are similar with the CUP technique except $\mathbf{M}_{\text{paddle}}$ equals $\mathbf{M}_{\text{fiducial}}$.

Selection: if $\text{distance}(\mathbf{M}_{\text{paddle}}, \text{boundary}(\mathbf{M}_{\text{obj}})) < \text{RADIUS}$
and there is no selected object (12)

Release: if $\text{degree}(\mathbf{M}_{\text{paddle}}) > \epsilon$ until an item slides away on
the marker and there is a selected object (13)

Rotation and translation: $[\mathbf{RT}]\mathbf{M}_{\text{obj}(t)} =$
 $[\mathbf{R}]\mathbf{M}_{\text{obj}(t-1)}([\mathbf{R}]\mathbf{M}_{\text{paddle}(t-1)}^{-1}[\mathbf{R}]\mathbf{M}_{\text{paddle}(t)})[\mathbf{T}]\mathbf{M}_{\text{paddle}(t)}$ (14)

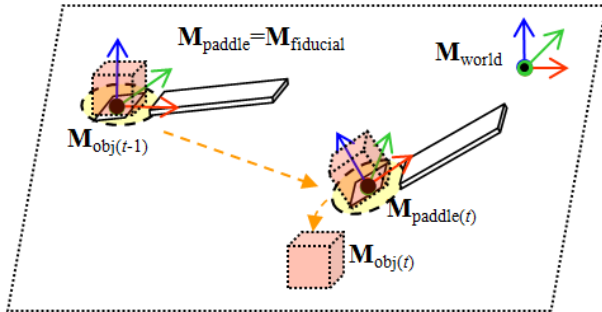


Figure 3. PADDLE technique in the 2D arrangement of 3D objects (2D_PADDLE).

CUBE Technique. To select the 3D object, a user must create and maintain a collision state for a specific period of time (N_FRAME frames) between the 3D object and a virtual line tip augmented on the front of the cubical manipulation prop, as shown in Equation 15. Specifically, the 3D object release event can be triggered when the marker is continuously hidden by a hand for more than F frames. In Equation 17, the rotation depends on $\mathbf{M}_{\text{fiducial}}$. On the other hand, translation is applied by $\mathbf{M}_{\text{tip}(t)}$.

Selection: if $\text{distance}(\mathbf{M}_{\text{tip}}, \text{boundary}(\mathbf{M}_{\text{obj}})) < \epsilon$ for
 N_FRAME frames and there is no selected object (15)

Release: if any marker is not detected for N_FRAME
frames and there is a selected object (16)

Rotation and translation: $[\mathbf{RT}]\mathbf{M}_{\text{obj}(t)} =$
 $[\mathbf{R}]\mathbf{M}_{\text{obj}(t-1)}([\mathbf{R}]\mathbf{M}_{\text{fiducial}(t-1)}^{-1}[\mathbf{R}]\mathbf{M}_{\text{fiducial}(t)})[\mathbf{T}]\mathbf{M}_{\text{tip}(t)}$ (17)

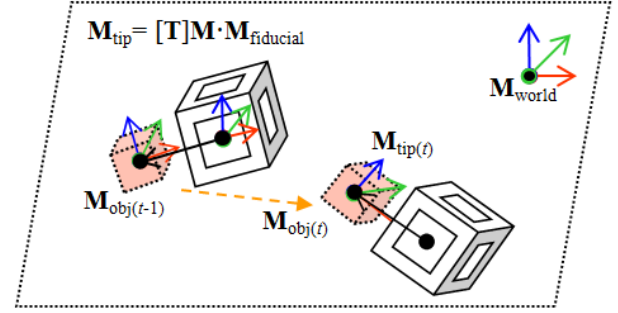


Figure 4. CUBE technique in the 3D arrangement of 3D objects (3D_CUBE).

ExPADDLE Technique. The ExPADDLE technique, and extended version of the PADDLE technique, features (i) a virtual spherical selection region, (ii) a multiple tracking pattern printed box, a button for discrete input, and (iv) a mouse handle for grasping it.

The explanation of each feature is as follows: (i) a virtual spherical selection volume of the manipulation prop expands the selection area and represents visual feedback; (ii) a multiple tracking pattern printed box, an extended marker tracking for PADDLE, which is attachable to the conventional mouse (this is for 3D absolute position tracking of the prop at arbitrary camera viewpoints with more stable tracking); (iii) discrete input from the mouse button can be used to trigger the event functions instantly (e.g., 3D object selection or release); and (iv) finally, a handle grasp is to prevent the user from hiding the fiducial markers.

For ExPADDLE, if the user presses and holds a button on the manipulation prop (discrete input type) in the 3D object selection state, then the drag state is initiated. 3D translation/rotation values of the sphere of the tracked prop are converted continuously into translation/rotation input values for the 3D object (e.g., continuous input type). When the user releases the button, the drop state completes the translation/rotation changes of the 3D object.

Selection: if $\text{distance}(\mathbf{M}_{\text{sphere}}, \text{boundary}(\mathbf{M}_{\text{obj}})) < \text{RADIUS}$
and a button pushed and there is no selected object (18)

Release: if a button released and there is a selected object (19)

Rotation and translation: $[\mathbf{RT}]\mathbf{M}_{\text{obj}(t)} =$
 $[\mathbf{R}]\mathbf{M}_{\text{obj}(t-1)}([\mathbf{R}]\mathbf{M}_{\text{sphere}(t-1)}^{-1}[\mathbf{R}]\mathbf{M}_{\text{sphere}(t)})[\mathbf{T}]\mathbf{M}_{\text{sphere}(t)}$ (20)

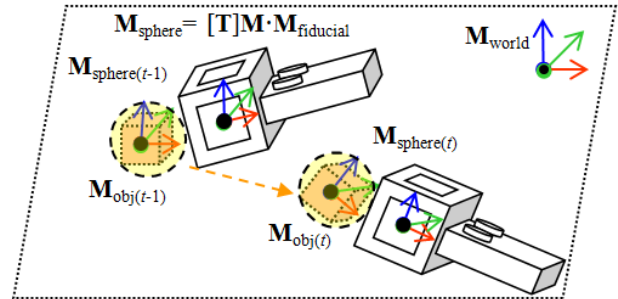


Figure 5. ExPADDLE technique in the 3D arrangement of 3D objects (3D_ExPADDLE).

4 EVALUATION PROCESS AND RESULT

4.1 Evaluation Process

A user evaluation was conducted in an indoor environment with consistent lighting. A general-purpose USB camera was used and

fixed onto a camera arm. Thirty images were captured per second with a 640x480 pixel resolution. Auto-exposure and auto-focus were set as a default. The computer was equipped with a 2.4 GHz CPU and 4GB of memory. The osgART [28] library was used to support scene graph structure-based rendering methods and computer vision-based fiducial marker tracking functions. As shown in Figure 6(a), a monitor, a camera, and the direction of the user’s view were aligned. The viewing angle of the camera in front of the user was toward the page, and the entire page area was visible within the camera input image, as shown in Figure 6(b).

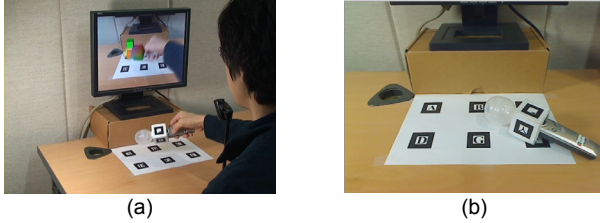


Figure 6. Experimental environment: (a) experimental space and (b) input image.

The independent variables involved in the evaluation centered on the index of difficulty (ID) of the tasks, including target width (W), amplitude (A), and angle, as shown in Table 1. The dependent variables were the completion time (ms), error (mm), number of successful trials, and efficiency measure TP. The ID was composed of a total of 16 step trials in a random sequence. The scales used for the 3D object were 10, 20, 30, and 40 mm, and the between distances were 60, 120, 180, and 240 mm based on the normal page size. Also, the angles of the target 3D object were changed to 0, 30, and 60 degrees. In the 2D arrangement of the 3D objects, there were 4 different techniques for comparison, 2 of which were comparable in the 3D arrangement of the 3D objects. In addition, we survey user subjective ratings using Likert scales in term of the stability, speed, comfort, and understanding.

Table 1. Independent variables in experiment

Independent variables	Properties
Width	10, 20, 30, 40 (mm)
Amplitude	60, 120, 180, 240 (mm)
Angle	0, 30, 60 (degrees)
Trial	Random 16 trials of total 48 trials (4 widths, 4 amplitudes, 3 angles)
Block	1 block
2D arrangement	2D_CUP
	2D_PADDLE
	2D_CUBE
	2D_ExPADDLE
3D arrangement	3D_CUBE
	3D_ExPADDLE

To reduce any learning effect due to the order in which the manipulation techniques were conducted, the experiments were executed at random in the 2D and 3D arrangements, respectively. Specifically, we distributed the trials randomly to ensure good coverage of all conditions. The number of experiments for each selection technique exceeded 320, and we tried to establish a correlation value (R^2) between MT and an ID higher than 0.9.

For the statistical analysis, we ran Levene’s test to assess the equality of variance. We then ran the ANOVA (Analysis of Variance) test for the comparison of the 4 kinds of 2D-based manipulation. In the post-hoc analysis, we performed the Scheffe adjustment for equal variance assumption and the T3 adjustment

of Dunnett for unequal variance assumption. We also ran the independent t-test for the comparison of the 2 kinds of 3D-based manipulation. On the other hand, for the sample data rectification from the collected results of each trial, the TP was calculated and placed in descending order. Fifty percent of the trials were then selected, based on the mean of the distribution and extracted as samples. If the translated 3D object occurred within this bound of the target 3D object, then the trial was extracted as a sample. Error was computed using the Euclidian distance between the center position of the translated 3D object established by the user and the ideal target position. A successful trial was counted when a collision occurred between the translated 3D object and the target 3D object, as previously mentioned. Another indication was that the final distance was closer than 3-sigma within from the average.

4.2 Implementation

For the common evaluation condition, we implemented and modified the comparison programs. In these experiments, “pattMultimarker” of the ARToolKit library [29] was used as a base marker and a total of six 40 mm markers were arranged in a 2 x 3 formation with a 100 mm displacement. The markers attached to the manipulation prop were also derived from the ARToolKit example.

In the implementation of ExPADDLE and CUBE, four markers were attached on each side of the cube next to the top and bottom. For the PADDLE and CUP implementations, a marker was attached to the upper side of both the stick and cup, respectively. The marker size was fixed at 30 mm based on the experimental estimation of a minimum distance of proper tracking of markers from the camera’s viewpoint when a user can stretch out his or her arm while grasping the manipulation prop. There was also a 5 mm white margin around the marker to distinctly identify the black rectangles of the marker well. As a result, each side of each marker was 40 mm in length, specifically, with two 5 mm margins and a 30 mm marker.

To confirm a selection, both a visual effect with green cube at the initial state and a red cube at selection/translation, and a sound effect for selection and release, were provided and were the same in all comparison experiments. An explanation of the implementation details for each technique is provided in the following paragraphs.

CUP Technique. The cylinder shaped cup prop had a height of 80 mm (HEIGHT value) and a 30 mm radius. A 40 mm marker was attached to the top of the CUP with a 20 mm RADIUS value. For the dropping condition, the T_HEIGHT value was set to 110 mm.

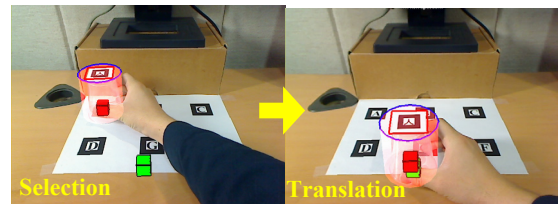


Figure 7. Selection and translation using a manipulation prop: CUP technique in the 2D arrangement of 3D objects (2D_CUP).

PADDLE Technique. For the PADDLE technique, we set the selection area to 20 mm (RADIUS value). For the dropping condition, a user tilts the PADDLE until an item slides away from the 20 mm radius of marker.

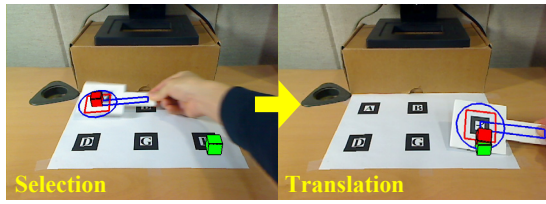


Figure 8. Selection and translation using a manipulation prop: PADDLE technique in the 2D arrangement of 3D objects (2D_PADDLE).

CUBE Technique. The 3D object selection is enabled when the tip of virtual line is inserted continuously in the 3D object for more than 15 frames (N_FRAME frames). The 3D object release event also can be triggered when the marker is continuously hidden by a hand for more than 15 frames.

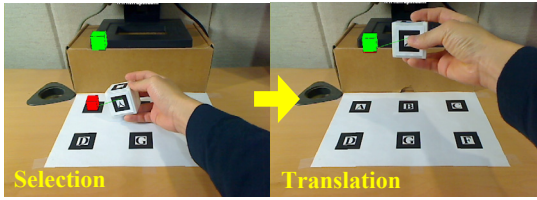


Figure 9. Selection and translation using a manipulation prop: CUBE technique in the 3D arrangement of 3D objects (3D_CUBE).

ExPADDLE Technique. The extended version of the PADDLE prop has a detachable cubical box printed with multiple markers attached to the front of a conventional mouse. A transparent plastic sphere with a 20 mm radius (RADIUS) and 5 mm thickness also was attached to the front of the box to provide physical feedback when the virtual sphere made contact with the 2D plane. A yellow virtual sphere model was registered on the plastic sphere in AR.

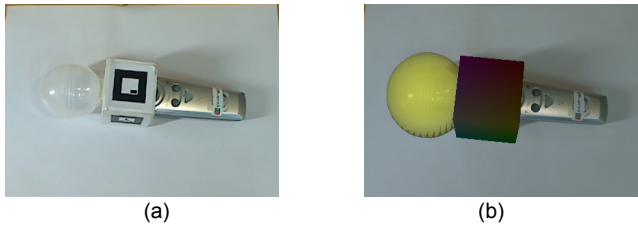


Figure 10. ExPADDLE manipulation prop: (a) a detachable cubical box printed with multiple markers and (b) a yellow virtual sphere model were registered on the plastic sphere in AR.

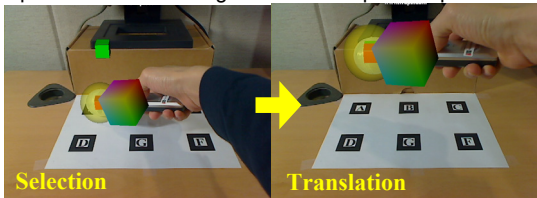


Figure 11. Selection and translation using a manipulation prop: ExPADDLE in the 3D arrangement of 3D objects (3D_ExPADDLE).

4.3 Evaluation Results

A total of 20 participants, 15 male and 5 female, participated in our evaluation and had an average age of 30.05 (min 24, max 42, STD 4.87). Of the 20 participants, 11 had a university degree in computer science, 16 had previous experience in AR, and every

participant was dominantly right-handed. Also, all participants had sufficient knowledge and physical ability to perform basic computer skills, specifically clicking a button on a mouse and dragging and dropping a file or folder into the GUI-based desktop environment. The total evaluation times varied between 20 minutes and 50 minutes. Before the evaluation, participants were given an explanation and demonstration of each technique and were then given time to practice the techniques. We requested the participants to translate a selected 3D object to the target position as quickly and as accurately as possible.

2D Arrangement. In the condition of 3D objects arranged on a 2D plane, there was a significant difference between the complete time and the error between manipulation techniques through the comparisons (MT: $F_{(3,421)}=73.376$, $p<0.01$, Error: $F_{(3,421)}=3.051$, $p=0.028$). As shown in Table 3, 2D_ExPADDLE required the least amount of completion time. This was followed by the 2D_PADDLE and 2D_CUP, although it should be noted that there was not a significant difference between them as shown in Table 2. Every manipulation technique had roughly 10 mm in error. The 2D_PADDLE had below-average error, but it produced no noticeable significant effects with the other techniques as shown in Table 2. Table 3 shows the TP efficiency measure for each manipulation technique. The 2D_ExPADDLE showed the highest TP, followed by the 2D_PADDLE and the 2D_CUP.

We observed that as ID (bit) of the task increased, the completion time MT (ms) also tended to increase linearly, as shown in Figure 12(a). Most of the manipulation techniques were well-modeled with linear regression according to Fitts' law with a correlation coefficient value of over 0.9, except for the 2D_CUBE.

In the user subjective ratings, the 2D_PADDLE and 2D_ExPADDLE had more preference than the 2D_CUP and 3D_CUBE in the Stability item. And in the Speed item, the 2D_ExPADDLE shown remarkably higher favor than other techniques. In the Comfort, the 2D_CUP was most preferred, followed by the 2D_ExPADDLE, 2D_PADDLE, and 2D_CUBE. In the Understanding, the 2D_ExPADDLE, 2D_PADDLE, and 2D_CUP were preferred than the 2D_CUBE, but it was hard to arrange in order.

3D Arrangement. In the 3D space arrangement of 3D objects, the 3D_ExPADDLE required more than 50% less time than the 3D_CUBE. The error of both techniques, however, was almost the same (10 mm). The comparisons also showed a significant difference in time between the 3D_ExPADDLE and 3D_CUBE ($t_{(154,262)}=15.651$, $p<0.01$); but again, there was no significant difference in error ($t_{(211)}=0.277$, $p=0.782$). Table 3 shows that the TP of the 3D_ExPADDLE was about 2 times higher than the 3D_CUBE.

As shown in Figure 12(b), the 3D_ExPADDLE was well-modeled with linear regression according to Fitts' law, with a correlation coefficient value of over 0.9, while the 3D_CUBE was not. In the user subjective ratings, the 3D_ExPADDLE obtained higher preferences than the 3D_CUBE in most items as shown in Figure 13.

Table 2. Significance probability: p-value (* notes significant level: $p<0.05$, and ** for $p<0.001$).

		MT(ms)	Error(mm)
2D_CUP	2D_PADDLE	0.985	0.155
	2D_CUBE	**	0.998
	2D_ExPADDLE	**	1.000
2D_PADDLE	2D_CUBE	**	0.048*
	2D_ExPADDLE	**	0.127
2D_CUBE	2D_ExPADDLE	**	1.000
3D_ExPADDLE	3D_CUBE	**	0.782

Table 3. Quantitative evaluation results for the 2D and 3D arrangement of the 3D objects. (* notes positive score of the evaluation items in each arrangement; standard deviation values are shown in parentheses.)

Technique	MT(ms)	Error(mm) for success trial	The number of successful trials (total 320)	TP (bit/s)
2D_CUP	2911.047 (1063.679)	12.302 (9.153)	Bounding: 182, 3-sigma: 268	0.685
2D_PADDLE	2826.667 (835.371)	9.833 (7.728)*	Bounding: 255, 3-sigma: 304*	0.797
2D_CUBE	3868.143 (1306.063)	12.898 (9.195)	Bounding: 195, 3-sigma: 300	0.553
2D_ExPADDLE	1750.989 (532.616)*	12.753 (10.123)	Bounding: 172, 3-sigma: 297	1.110*
3D_CUBE	4058.553 (1218.802)	10.368 (8.363)	Bounding: 187, 3-sigma: 289	0.540
3D_ExPADDLE	1939.118 (656.275)*	10.045 (8.636)*	Bounding: 210, 3-sigma: 304*	1.128*

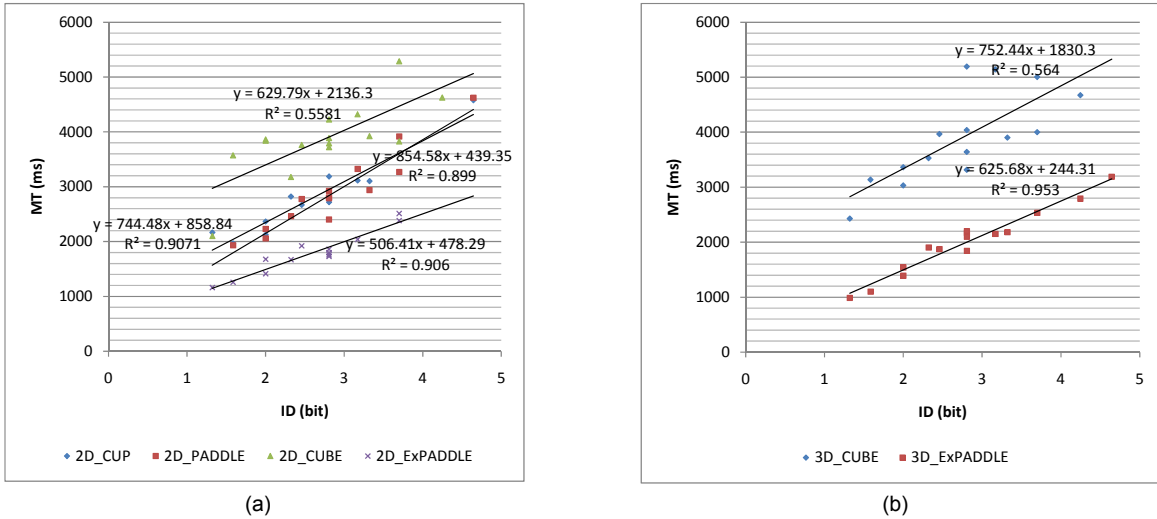


Figure 12. Graph shows a tendency of MT (ms) to ID (bit) with linear regression: (a) 2D arrangement and (b) 3D arrangement of the 3D objects (least-squares linear regression is used to find the intercept and slope parameters of Fitts' law).

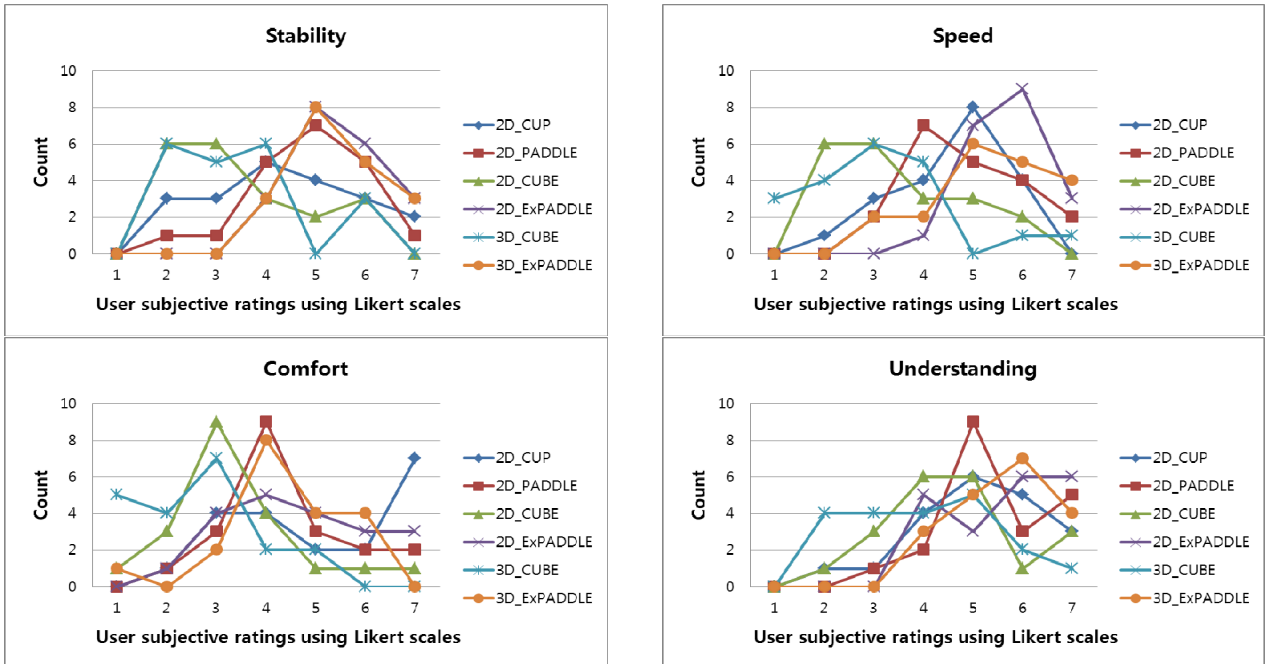


Figure 13. User subjective ratings using Likert scales (1 ~ 7 score; 1 being most negative and 7 being most positive) in term of stability, speed, comfort, and understanding. A total of 20 participants participated in our evaluation.

5 DISCUSSION AND CONCLUSION

5.1 Discussion

In this section, we summarize valuable comments and our opinions on the tangible AR manipulation techniques through performance evaluation, observations, and informal interviews through the experiments.

Stability and speed. Quantitative evaluation results in Table 2 show no noticeable significant differences in error (mm) for the success trial, but there are remarkable differences in the movement time (MT) and the number of successful trials.

In ExPADDLE, a user selects a 3D object with the virtual selection volume rendered on the physical sphere. The fixed-sized (20 mm radius) transparent plastic sphere could be physically constrained to properly release a 3D object (especially a small object) on the 2D plane. This kind of situation does not occur in the air of 3D space, so the number of successful trials was significantly enhanced from the 2D-based (bounding: 172, 3-sigma: 297) to 3D-based (bounding: 210, 3-sigma: 304) experiment. However, this constraint of ExPADDLE did not cause a noticeable problem in the user-subjective ratings as shown in Figure 13.

In the case of CUBE, the virtual line does not make a remarkable effect in the 2D-based (bounding: 195, 3-sigma: 300) and 3D-based (bounding: 187, 3-sigma: 289) experiments, unlike the ExPADDLE, because it might have no physical constraints.

On the other hand, the manipulation prop tracking using a single fiducial marker at an arbitrary camera viewpoint could fail. The CUP and PADDLE cases occasionally caused unstable camera tracking when picking up, dropping, and tilting the 3D object, because the angle of the marker on top of the prop suddenly changed; therefore users carefully selecting/releasing and translating a 3D object could exhibit longer task completion times in the quantitative evaluation results and lower preferences in the user-subjective ratings rather than ExPADDLE.

One of interesting things was that participants scored CUP and PADDLE similarly in the user-subjective ratings, although CUP had the lowest number of successful trials (bounding 182, 3-sigma: 268) and PADDLE had the highest score (bounding: 255, 3-sigma: 304) in the quantitative evaluations. For PADDLE, users move the prop inward toward the camera to release the 3D object; this might make more recognizable marker poses and more successful trials. The visual feedback of the selection volume-based techniques (e.g., CUP) caused temporary visual confusion between the selection volume and the 3D object, so this might have influenced the low score in the number of successful trials.

For CUBE, which has multiple markers, marker occlusion by hand occurred frequently because the participant was required to grasp the cube prop itself. The dwell time we set was a 500 ms delay (15 frames at 30 FPS) for selection and release. Occasionally, in the selection state, if the marker is not recognized continuously during the 500 ms period, the 3D object cannot be selected. As a result, most participants commented that every manipulation seemed to be latent and similar to the use of a single marker. Some participants also commented that the addition of a handle to the cube would be helpful to avoid hiding the marker.

Comfort. At first, when using CUP, participants were positive while translating the 3D object because they were able to rest their hands on the cup prop while grasping the prop. The Comfort item in the user-subjective ratings shows this tendency.

In the 3D layout case, with CUBE, many participants complained that continually holding their hand in the air caused fatigue in the wrist and arm. Also, for the small 3D objects, it was difficult and uncomfortable to maintain the selection state for a period of time.

For the ExPADDLE, although most participants were positive regarding instant manipulations, some participants found that it was uncomfortable and increased fatigue to press and hold the button continually during the drag state.

On the other hand, much positive feedback was collected from the participants regarding the volume-based selection. Accuracy of depth perception was not perfectly necessary in the volume-based selection, while in the CUBE, accurately maintaining a virtual line tip position into the 3D object in 3D space for the selection was required. This has been reported in several related experiments [30].

Understanding. The CUP and PADDLE seemed easy to understand due to the simple touch action used for 3D object selection. In the PADDLE, for example, some participants described the picking and dropping gestures as being like a shovel or transferring food with a scoop.

For the CUBE, the 3D object translation seemed to be easy and accurate because of the occlusion of the line tip into the 3D object. However, it is difficult to understand the need to hold the same hand posture continuously for a period of time to select the 3D object and hide the marker with a hand for the duration to release the 3D object.

On the other hand, with ExPADDLE, participants were able to know immediately the selection/release time by pushing a button, because it was similar to the drag-and-drop technique using a mouse input device in a desktop computer environment.

Here we conclude our discussion and suggestions for future design of a computer vision-based tangible AR manipulation technique using fiducial markers for tracking.

- The comparison techniques we examined follow intuitive pointing, dragging and dropping, scooping, and picking up behaviors to enhance usability through promoting the user's understanding: these obtained positive opinions.

- From the observations and experiment results, the multiple markers could enable proper prop tracking at arbitrary camera viewpoints. Also, a handle grasp could help a user avoid hiding the fiducial markers with hands.

- In ExPADDLE, we used a button to provide a discrete input to trigger the event function required in the manipulation. This kind of instant triggering could help rapid manipulation. Also, by expanding the 3D object selection area to a virtual volume area on the prop, the selection can be made easier if there is no visual confusion between the selection volume and the 3D object.

5.2 Conclusion and Future Work

In this paper, we examine 3D object selection and manipulation methods specifically for a tangible AR environment. We adopted Fitts' law and conducted comparison evaluations with conventional techniques. For 3D objects arranged on a 2D plane and in 3D space, the

throughput (TP) of the extended paddle technique peaked at around 1.39 to 2 times higher than the other techniques. In discussion, we summarize valuable comments and our opinions for virtual hand metaphor-based manipulation techniques in the TAR environment in terms of stability, speed, comfort, and understanding. We then provide design suggestions with respect to the intuitive manipulation metaphor, a multiple tracking pattern, a handle grasp, and a virtual selection volume.

Our evaluation results and discussion can be helpful in 3D user interface of various tangible AR applications, such as AR-based gaming, virtual pop-up books, pamphlets, school education, advertisements, broadcasting, exhibitions, theme parks, or in various other domains.

As a future study, we plan to extend these experiments considering the following issues:

- Adding several variations with different shapes, local pose, density, and color of the 3D block, stereo input images, or other factors. Also, for participants with no AR experience, we will compare learning effects between a gesture- and a button-based manipulation interaction.

- Visual and physical feedback of the selection volume would be beneficial for depth perception in an AR environment. To support this notion, we need to conduct more in-depth experiments with visual or physical feedback. Also, 3D depth perception enhancement through the use of shadow or vibration feedback will be analyzed for more accurate 3D object selection and manipulation.

- Props with no handles vs. props with handles are a tradeoff between using the fingers and using the fingers and wrist. In some cases, with a handle, users can move a 3D object using their wrists and fingers faster than using only fingers, or vice versa. In future in-depth experiments, we will show the benefit of the handle in small movements and larger movements, where a user wearing HMD must turn his head in large AR environments.

- In order to elaborate on the Fitts's law model, we need to prevent any side effects from fiducial-marker-based tracking. Its reliability, robustness, accuracy, and other factors might have an influence on the overall result.

REFERENCES

- [1] M.Billinghurst, H.Kato, I.Poupyrev, "Collaboration with Tangible Augmented Reality Interfaces," HCII, pp. 2001.
- [2] H.ISHII, B.ULLMER, "Tangible bits: towards seam- less interfaces between people, bits and atoms," ACM CHI, pp.234-241, 1997.
- [3] H.Kato, M.Billinghurst, I.Poupyrev, K.Imamoto, K.Tachibana, "Virtual Object Manipulation on a Table-Top AR Environment," ISAR, pp. 111-119, 2000.
- [4] H.Kato, K.Tachibana, M.Tanabe, T.Nakajima, Y.Fukuda, "MagicCup: A Tangible Interface for Virtual Object Manipulation in Table-Top Augmented Reality," the Augmented Reality Toolkit Workshop, pp.75-76, 2003.
- [5] G.Lee, C.Nelles, M.Billinghurst, G.J.Kim, "Immersive Authoring of Tangible Augmented Reality Applications," ISMAR, pp.172-181, 2004.
- [6] M.Billinghurst, "Usability testing of augmented/mixed reality systems," ACM SIGGRAPH ASIA Courses, 2008.
- [7] A.Dünser, R.Grasset, M.Billinghurst, "A survey of evaluation techniques used in augmented reality studies," ACM SIGGRAPH ASIA Courses, 2008.
- [8] J.Looser, M.Billinghurst, R.Grasset, A.Cockburn, "An evaluation of virtual lenses for object selection in augmented reality," the Computer Graphics and interactive Techniques in Australia and Southeast Asia, 2007.
- [9] H.Benko, S.Feiner, "Balloon Selection: A Multi-Finger Technique for Accurate Low-Fatigue 3D Selection," IEEE 3D UI, pp.79-86, 2007.
- [10] S.Myojin, H.Kato, S.Nishida, "Evaluation of a Cup-shaped Interface in Tabletop AR Environments," IEEE 3DUI, pp.137-138, 2009.
- [11] J.Chastine, K.Nagel, Ying Zhu, M.Hudachek-Buswell, "Studies on the Effectiveness of Virtual Pointers in Collaborative Augmented Reality," IEEE 3DUI, pp.117-124, 2008.
- [12] A.H.Mason, M.A.Walji, E.J.Lee, C.L.MacKenzie, "Reaching movements to augmented and graphic objects in virtual environments," ACM CHI, pp.426-433, 2001.
- [13] P.M.Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," Journal of Experimental Psychology, 47, pp.381-391, 1964.
- [14] I.Poupyrev, S.Weghorst, M.Billinghurst, T.Ichikawa, "Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques," Eurographics, 17, 3, pp.41-52, 1998.
- [15] DA Bowman, E Kruijff, JJ LaViola, I Poupyrev, "3D User Interfaces Theory and Practice," Addison-Wesley, 2004.
- [16] J.Liang, M.Green. "JDCAD: A Highly Interactive 3D Modeling System," Computers and Graphics, 18, 4, pp.499-506, 1994.
- [17] A.Forsberg, K.Herndon, R.Zeleznik, "Aperture based selection for immersive virtual environments," ACM UIST, pp.95-96. 1996.
- [18] J.Pierce, A.Forsberg, M.Conway, S.Hong, R.Zeleznik, "Image plane interaction techniques in 3D immersive environments," ACM I3D, pp.39-43, 1997.
- [19] T.Grossman, R.Balakrishnan, "The design and evaluation of selection techniques for 3D volumetric displays," ACM UIST, pp.3-12. 2006.
- [20] D.Bowman, L.Hodges, "An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments," ACM I3D, 1997.
- [21] I.Poupyrev, M.Billinghurst, S.Weghorst, T.Ichikawa, "The go-go interaction technique: non-linear mapping for direct manipulation in VR," ACM UIST, pp.79-80, 1996.
- [22] IS.MacKenzie, "Fitts' law as a research and design tool in human-computer interaction," HCI, 7, pp.91-139, 1992.
- [23] IS.MacKenzie, W.Buxton, "Extending Fitts' Law to two-dimensional tasks," ACM CHI, pp.219-226, 1992.
- [24] C.Ware, K.Lowther, "Selection using a one-eyed cursor in a fish tank VR environment." ACM TOCHI, 4, 4, pp.309-322, 1997.
- [25] T.Grossman, R.Balakrishnan, "Pointing at Trivariate Targets in 3D Environments," ACM CHI, 6, 1, pp.447-454, 2004.
- [26] SA.Douglas, AE.Kirkpatrick, IS.MacKenzie "Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard," ACM CHI, pp.215-222, 1999.
- [27] RW.Soukoreff, IS.MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI," IJHCS, 61, pp.751-789, 2004.
- [28] J.Looser, R.Grasset, H.Seichter, M.Billinghurst, "OSGART - a pragmatic approach to MR," ISMAR, 2006.
- [29] ARToolKit, <http://www.hitl.washington.edu/ARToolKit>
- [30] L.Vanacken, T.Grossman, K.Coninx, "Exploring the effects of environment density and target visibility on object selection in 3D virtual environment," IEEE 3DUI, pp.115-122, 2007.