# User-friendly 3D Object Manipulation Gesture using Kinect

Jaemin Soh
KAIST
Daejeon, South Korea
revelati83@gmail.com

Yeongjae Choi
KAIST
Daejeon, South Korea
yjchoi@paradise.kaist.ac.kr

Youngmin Park
KAIST
Daejeon, South Korea
creathrow@gmail.com

Hyun S. Yang
KAIST
Daejeon, South Korea
yang@paradise.kaist.ac.kr

## Abstract

With the rapid development and wide spread of the virtual reality technology, we can easily find VR systems in various places such as school, library and home. Translation and rotation, the most frequently used gestures for manipulating objects in the real world, need to be implemented in order to make a user feel comfortable while manipulating 3D objects in the systems. In this paper, we propose a set of user-friendly 3D object manipulation gestures and develop a recognizer using Kinect for the VR systems. The usefulness and suitableness of the proposed system is shown from the user study performed.

**CR Categories:** H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented and virtual realities;

**Keywords:** human computer interaction, manipulation gesture, virtual reality, kinect

## 1. Introduction

With the development of the VR technology, there has been rapid improvement in the interaction technology for the VR environment. It is known that people can have more sense of reality when they can have more control of virtual world. Since interaction technology provides people with control of virtual world, it is required to be intuitive and generally acceptable. If it is not, people will feel less sense of reality in virtual world.

In this paper, we aim to provide more intuitive interaction to VR systems which are widely spread over such places where background is not fixed and computing resource is not enough as school, library and home. One of the VR systems is exemplified in [Lee et al. 2010] where users can study a foreign language or science using their body as a way of interaction with the system.

For the past decades, 2-dimensional image understanding technology has been widely used for the recognition of body postures and gestures. But, there have been a lot of difficulties achieving good recognition performance using 2-dimensional image only, due to the lack of information, especially when there is complex background in the image. Recently, it is easy to get depth information as well as RGB-color information using Kinect. The introduction of the depth information results in recognition performance improvement for the VR systems.

In this paper, we propose a set of user-friendly 3D object manipulation gestures and develop a recognizer using Kinect for the VR systems.

## 2. Related Works

Recently, Kinect has gained a great deal of attention from many researchers and developers due to its affordability and commodity. 'Kinect your Metro'[Greg 2012] is an middleware architecture for a user to interact with metro style app using Kinect sensor. It is potentially useful to some applications such as an image gallery. But it lacks flexibility since it only allows a very limited set of simple gestures like left and right hand up. 'Kinect Treatment of Windows 7'[Evoluce 2010] demonstrates a two-hand gesture controller. The demonstration shows that the gesture controller can be one of the useful alternative methods for classic input devices but it is only suitable for desktop environment. Lee presents image manipulation interface using fingertips and palm tracking [Lee et al. 2012]. But the interface is not fit for virtual systems since it doesn't consider 3D object manipulation.

Fusion4D[Keika and Roberto 2011] is a natural interface for manipulating 3D objects. The interface allows a user to move, rotate, and scale objects directly in a 3D virtual space using hand gestures and speech commands. But speech recognition is not always available in crowded and noisy environments such as a school and library. Therefore, we argue that it is necessary to develop an intuitive gesture-based interface for such places.

## 3. Manipulation Gesture Design

Translation and rotation, the most frequently used gestures for manipulating objects in the real world, need to be implemented in order to make a user feel comfortable manipulating 3D objects in the systems. Therefore, we design a set of intuitive manipulation gestures for them. If all the information of the posture can be obtained, the perfect manipulation gesture can be implemented. However, in the previous popular system, the information is often incomplete due to the lack of sensing quality. A manipulation gesture suitable for these systems of insufficient clues needs to be designed.

### 3.1 Constraints

To design the manipulation gesture, such things as the system input, environment and information quality need to be considered. In this chapter, some constraints are listed involved with these considerations.

The first constraint of the popularized systems is the unconstrained background. To acquire the posture information from a person, the segmentation from the background has to be preceded. The segmentation using traditional 2D color cameras is difficult with a complicated or human-like background. Kinect can help

solve this problem. When segmentation is done based on the scene depth information acquired by Kinect, the result of the segmentation can be robust to the complex texture. Moreover, the SDK provided by the manufacturer offers the high-quality position information of each body part.

The second constraint is the prohibition of voice recognition system. Just as the first constraint, the environment of the system is open space. Noisy sounds make the voice recognition do not work.

The third constraint is about the additional equipment. In order to popularize the system, extra tools such as hand-held or wearable equipment must be reduced or prohibited because they increase the cost and make the users tired.

3.2 Candidate Gestures

With the positions of the body parts obtained using Kinect, gesture sets can be designed in two directions. One is making the gesture as similar to the real world gesture as possible, and the other is making as high recognition performance as possible with an artificial gesture.

The former uses the positions of two hands. Imitating the rotation and translation of the real world gesture, the change of two hands' positions affects the virtual objects. Since the user cannot grasp the virtual objects, the selection method needs to be implemented. An example of the latter is like this. A user stretches his/her arms forward and moves left hand up and right down. Then, the selected object rotates clockwise with the axis of user's looking direction. This gesture does not make an occlusion.

3.3 Proposed Gesture

To implement the manipulation, it is needed to analyze the environment where the system will be installed and the suitability of the each gesture set. In this paper, the gesture is selected for the system which is suitable for home or the public places like schools and libraries.

With the constraints mentioned in 3.1, the first aim of design is the clarifying of each movement; selection, translation, rotation. To achieve this, each movement has its own mode which is shown in Table I. Only admitted movement has been handled in a particular mode. The change of mode obeys the Figure 1 and Table II.
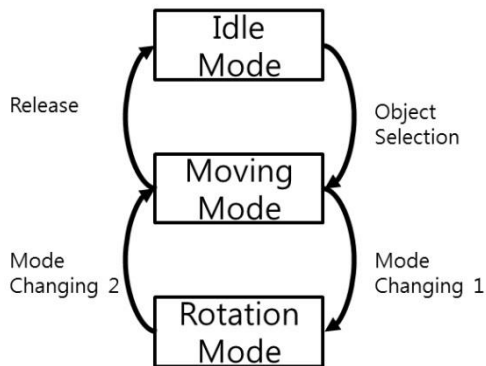


**Figure 1.** *Mode Change*

**Table I.** *Manipulation Gesture Set*

| Mode | Movement | Meaning |
| --- | --- | --- |
| Idle Mode | None | No movements affects the object |
| Moving Mode | Move the hand which uses the object selection | The selected virtual objects follows the movement of the hand |
| Rotation Mode | Spread and move two hands like rotating a real object | The virtual object is on the center of two hands and this is rotated by the movements of hands |

**Table II.** *Mode Change Gesture*

| Previous mode | Gesture | Next mode |
| --- | --- | --- |
| Idle Mode | Hold left or right hand for 1 second touching an object (do not press the other hand) | Moving Mode |
| Moving Mode | Press two hands for 1 second | Rotation Mode |
| Rotation Mode | (After rotation movement) Press two hands for 1 second | Moving Mode |
| Moving Mode | (After rotation and translation movement) Hold the hand which selected an object for 1 second | Idle Mode |

## 4. Gesture Recognition Process

For the development of manipulation gesture designed in the chapter 3, the positions of two hands need to be obtained every frame. The position changes of two hands have to affect the movement of virtual objects realistic.

4.1 Coordinate

The coordinate systems of the real world and the virtual world have to be concerned to achieve the correct manipulation of the virtual objects with real world's movements. Without aligning the coordinates, the complicated translation from one coordinate to the other is needed and it can easily have some side effects. If coordinates of two worlds are at least parallel, these effects can be reduced. Also, the ratio of the distance has to be concerned. Some objects move too much against the user's intention without the distance ratio adjustment. The most effective solution is making the two coordinates accorded using the calibration methods like [Zhang 2000][Tsai 1987].

4.2 Position

In this paper, a calibration method is used to make the real and virtual coordinates accorded. The calibration acquires the camera intrinsic and extrinsic parameters of the real camera, and then these parameters are applied to the virtual camera's parameter. The positions for translation are from this accorded coordinates.

According to the design of chapter 3, the selection (idle mode → moving mode) and release (moving mode → idle mode) are caused by touching a virtual object for 1 second, and the selected virtual object is moved by the movement of the user's hand in 'Moving Mode'.

## 4.3 Rotation

For the rotation of the object, the differences of positions are used instead of the absolute positions of the hands. If the Cartesian x, y, z style coordinate is used, the rotation problem is easy to fall down the gimbal lock effect. Therefore, the Axis-Angle style expression is used like Figure 2; The perpendicular vector of the planes formed by the movements of hands is used as 'Axis' and the conversion of the amount of hands' movements is used as 'Angle'.
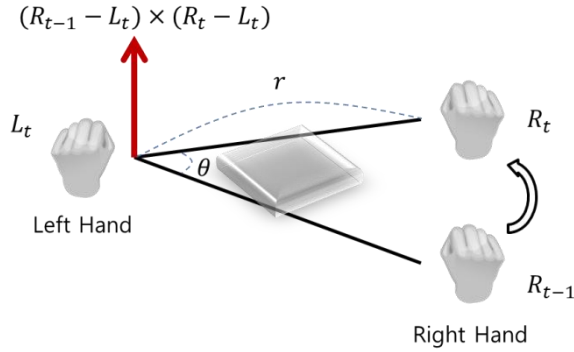


**Figure 2.** *Rotation calculation from hand position change*

The movements of left and right hands between time t and t-1, there are four positions in a three-dimensional space. Four positions cannot make one plane in general. Therefore, we calculate the difference between current and previous right hand positions $(R_t - R_{t-1})$ with the current left hand position $(L_t)$ and the difference between current and previous left hand positions $(L_t - L_{t-1})$ with the current right hand position $(R_t)$ independently. They make own planes and own axis-angle representations. Due to the duplication effect, the angle values are divided by 2. Followings are the formulations of these.

Right hand rotation Axis: $(R_{t-1} - L_t) \times (R_t - L_t)$

$$\text{Angle:} \quad \frac{|R_t - R_{t-1}|}{|R_t - L_t|}/2 \tag{1}$$

Left hand rotation Axis: $(L_{t-1} - R_t) \times (L_t - R_t)$

$$\text{Angle:} \quad \frac{|L_t - L_{t-1}|}{|L_t - R_t|}/2 \tag{2}$$

The rotation is achieved by the conversion from (1) and (2) to the rotation matrix or quaternion representation and application to the selected object.

## 5. Result Analysis

We had a survey of participants' feedback to evaluate our proposed gestures.

### 5.1 Survey



**[Manipulation Gestures]**

You can select / release an object by holding it in seconds.

You can move selected object by moving your hand.

You can enter and escape rotation mode by push your hands to forward and staying.

When you are in rotation mode, you can rotate selected object by changing the positions of your hands like in real world.

**[Evaluation]**

How much easier to learn manipulation gestures?

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| (Very Difficult) | | (Medium) | | (Very Easy) |

How much intuitive are the manipulations?

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| (Not intuitive) | | (Medium) | | (Very Intuitive) |

How much accurate are the manipulations?

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| (Not accurate) | | (Medium) | | (Very Accurate) |

Are you willing to use this system for manipulating 3D object?

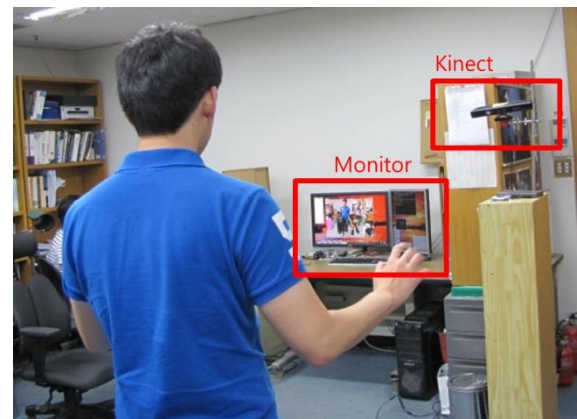| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| (Definitely No) | | | | (Definitely Yes) |

**Figure 3**. *Questionnaire*



**Figure 4.** *System configuration*

All participants performed moving, rotating, scaling 3D objects in the virtual environment and answer the questionnaire of Figure 3. 3D objects in virtual environment are augmented to the monitor which is positioned in front of the participant. Kinect is also positioned in front of the user and captures user images. The system configurations are shown in Figure 4.

Six graduate students who are not familiar with gesture interface or virtual reality participated for this experiment. They were instructed manipulation gestures once.

A participant sees the augmented scene of the real world and virtual objects. The user selects an object and makes it translate or rotate. Figure 5 is the scene captured from monitor.
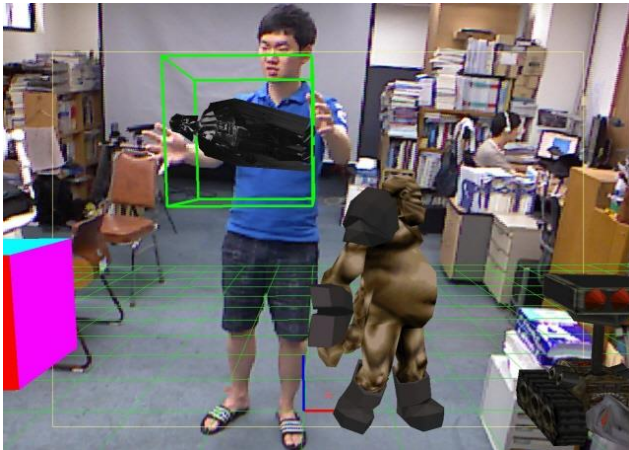


**Figure 5.** *User experience scene*

5.2 Analysis

Participants evaluate convenience, intuitiveness, accuracy, usefulness of proposed system into 5 classes. The experiment results are shown in Figure 6.

The results show that proposed manipulation gestures are intuitive and easy to learn but not so accurate. Especially, participants feel difficulty when they want to rotate an object slightly. This is due to self or partial occlusions which can be improved by adopting multiple Kinect sensors. Participants also point out that inconsistency of view point between user and augmented display is obstructive to manipulate objects accurately.

## 6. Conclusion

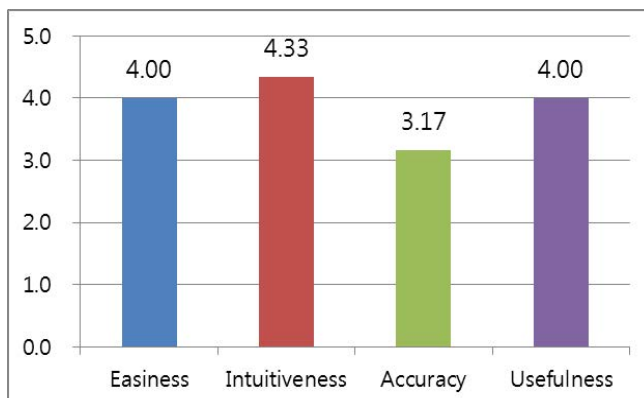With the development of the virtual reality technology, VR systems which need the way of 3D virtual object manipulation will be widely spread and the needs for much more intuitive and easier ways of manipulation will increase.

In this paper, we propose a set of user-friendly 3D object manipulation gestures and develop a recognizer using Kinect for the VR systems. The usefulness and suitableness of the proposed system are shown from the user study performed.

The result of the user study suggests that user feedback needs to be improved. First, it can be studied how users can be notified of the current state more intuitively. Second, it can be surveyed how much the proper change is for users' comfortableness while users translate or rotate virtual objects.

### Acknowledgement

### References

R. Y. TSAI. 1987. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation.*

ZHENGYOU ZHANG. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

S. LEE, J. KO, S. KANG, AND J. LEE. 2010. An immersive elearning system providing virtual experience. *IEEE International symposium on Mixed and augmented reality.*

EVOLUCE. 2010. Kinect Treatment of Windows 7. http://www.youtube.com/watch?v=2HkKcFKzorQ

KEIKA KEIKO MATSUMURA, ROBERTO SONNINO. 2011. FUSION4D. http://www.interlab.pcs.poli.usp.br/fusion4d/

LEE UNSEOK, AND JIRO TANAKA. 2012. Hand Controller: Image Manipulation Interface Using Fingertips and Palm Tracking with Kinect Depth Data. *Proc. Asia Pacific Conf. on Computer Human Interaction.*

GREG DUNCAN. 2012. Kinect your Metro. http://channel9.msdn.com/coding4fun/kinect/Kinect-your-Metro

**Figure 6.** *Result of user study*