Introducing Gestural Operation of the Viewpoint in a 3D Virtual Space of a Multiple Perspectives System

Kasumi Tarukawa
Graduate School of Science and Technology
Keio University Japan
tarukawa@mos.ics.keio.ac.jp

Tomoo Inoue
Faculty of Library,
Information and Media Science
University of Tsukuba
inoue@slis.tsukuba.ac.jp

Ken-ichi Okada
Graduate School of Science and Technology
Keio University, JST Japan
okada@mos.ics.keio.ac.jp

Abstract

Proper spacial information recognition promotes tasks dealing with space design. This paper presents the design, implementation, and evaluation of an interactive tabletop and vertical display system that supports spacial information recognition in tasks. This research focuses on two kinds of viewpoints, the overhead view and the individual view. The proposed system shows both viewpoints. We use a tabletop display for the overhead view and a vertical display for the individual view. The system allows the user to move the object intuitively in each viewpoint by the introduction of the gesture operation. It was investigated whether the gestural operation of the object representing the user him/herself in the virtual space could be compared with the physical object manipulation in the individual view. Statistically significant difference was not found between the two, but the gestural operation in the individual view gave better impression to the users.

1. Introduction

Making a minute plan before putting ideas into action is important in various tasks. There are many systems that support a user to making a plan. Among them, tasks dealing space design need proper spacial information recognition. In this paper, we focus on two different viewpoints needed for spacial information recognition. They are the overhead view and the individual view. In tasks dealing space design, being aware of the overhead view and the individual view at the same time is essential.However, there are few systems to support the simultaneous use of two different types of view. Even they show multiple viewpoints, there is not the individual view, or the user can operate an object only from one side. In operation of specific object that provide the individual view, there is a need to consider two processes that make operation difficult. These processes are "where the object is facing," and "to where move the object." Hence, this paper proposed gesture operation on the individual view. We show the overhead view on a table-shaped touch panel, and the individual view on vertical display. Also we introduce Microsoft Kinect for gesture recognition [1,2]. The user can easily rotate or move the object, using simple gesture by matching of the gaze direction of the target and the operator. We present an analysis of 14 user evaluations about operation of an object from each viewpoint. Accordingly, we concluded the proposed system allows the user to move the object intuitively. The system promotes spacial information recognition helping space design.

2. Spacial information presentation

Suitable spacial information recognition promotes tasks dealing with spacial design, such as planning of urban design, theater, and sports strategy. For example, in making sports strategy, we arrange the player's position and make effective formation. Moreover, we check the viewpoint of a player in the field. In this way, to make plan accommodate the real space, the space information such as a balance of the entire and a sense of distance between objects is necessary. If you correctly recognize spacial information and imagine the space, you can make a plan more suitable to the actual situation even when you cannot actually reproduce the space.
In this paper, we focus on two different viewpoints needed for spacial information recognition.

2.1. Use of two viewpoints

We use frequently the overhead view and the individual view, for spacial information recognition. In this section, we will explain each the two viewpoints.

CollabTech 2012 , August 27-29, 2012, Hokkaido, Japan.

2.1.1. Overhead view. When a user checks the whole spacial information, he uses a viewpoint looking down from sky. This viewpoint is called “the overhead view”. Presenting spacial information from the overhead view is effective in sports strategy board, maps, floor plans of the room, and so on. Overhead view is used frequently, such as when reviewing the overall balance between multiple objects and their current location.

2.1.2. Individual view. In this study, the perspective of an object in space referred to as "the individual view". It is also important to be aware of the viewpoint from the eyes of an object that exist in space. Presentation of spacial information from the individual view is effective in action game to proceed in the eyes of a hero, camera from the eyes of an actor, and so on. The user can see the perspective of the object that actually exists in space, it is possible to proceed more aware of the actual situation. In this paper, we call the object, which shows its own viewpoint in a display for the individual view, “the viewpoint object.”

In a variety of tasks, being aware of the overhead view and the individual view at the same time is essential. For example, in room design, the user arrange interiors from the overhead view, then check viewpoint from the eyes of resident at the individual view in order to check the appearance of the interior.

2.2. Gap between viewpoints

Figure 1 indicates the same position viewed by two viewpoints in Google Street View[3]. The top image is the overhead view and the bottom image is the individual view. Even in the same location in space, the given image is also very different between the overhead view and the individual view.

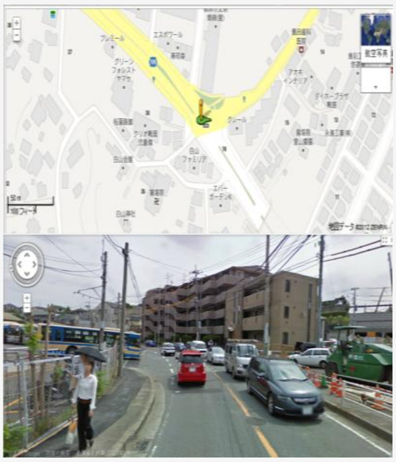

Figure 1. Given images from two viewpoints.

Therefore, space design system needs to display both the overhead view and the individual view. But there are few systems to support the simultaneous use of two different types of view. By focusing too much on the perspective of one or the other, the user cannot be aware of other viewpoints. As a result, there is a risk of not be able to recognize accurately the spacial information and task progress becomes inefficient. Also, for viewpoints display system for space design, linking all objects through all viewpoints and supporting intuitive operation in each viewpoint are important. However, current space design system has not been able to achieve intuitive operation. This is because the user’s operation is possible only from the overhead view. When the user looks the individual view and tries to move the viewpoint object, he must operate it from the overhead view. After determine the direction at the individual view, it is difficult to change direction in the overhead view. There is a need to consider two processes, “where the object is facing,” and “to where move the object.” This is the gap on operation of the viewpoint object between two viewpoints. This gap interferes with intuitive operation. For example, in making sports strategy, when a user see the individual view and attempt to move a player, who is the viewpoint object, the user must judge where the player facing before operate in the overhead view. The user judge the direction by comparing the sequence of objects between two viewpoints. Moreover, the user must think where to move the player in the overhead view. Thus, although the user can use multiple perspectives, their work might not proceed efficiently. Limited input operation makes operation of the objects difficult.

2.3. Related works

A tabletop interface is a touch panel of the table type. It is suitable for cooperative work because multiple users can work together around the table. Warcraft III[4] andThe Sims[5] use specific tabletop called Diamond Touch[6]. Diamond Touch can identify the operator. Also Diamond Touch can recognize the multiple points of operation, and can read a variety movement of hand. The systems can operate the game in a variety of gesture utilizing features of Diamond Touch.

We mentioned two kinds of viewpoint capturing the virtual space in Section 2.1. Then, we introduce how to present the perspective of various systems. First, we introduce methods to recognize the spatial information in the perspective of workers themselves. User can recognize spatial information intuitively in these systems. One research uses HMD display[7]. The user wears the HMD, and sees the digital graphics overlaid on the real space. The purpose of this method is to help user's work in real space by adding digital graphics. CAVE[8] is a space surrounded by the display. It can give the user a sense as if he is actually in the virtual space by changing the virtual space presented according to the movement of workers. Second, we introduce methods to recognize the spatial information in the perspective of an object in the space. Users perceive spatial information indirectly by manipulating the object. Reality is less than the first methods, but users can share their view with another user easily in the second methods.

One example is Google Street View[3]. A map is showed in PC display. User drags the object and put on where he wants to see the landscape of a street on the map. Then, another view shows the landscape depending on the orientation of the doll.

Besides, there is a technique to present spatial information from the perspective behind a manipulatable character in the virtual space. This technique provides a lot of information compared with the method above, but it is difficult to check the viewpoint of the object itself

Recently, Sekai Camera[9] show digital graphics overlaid on the real space by recognizing AR marker with iPad or iPhone.

There are some researches that support spacial information recognition by showing virtual images from multiple perspectives. One of them is Extending the Multi-user Tabletop Interface by Direction-aware Physical Objects for Producing Activity of Theatrical Performance, which uses tabletop as miniature stage and PC display showing virtual space[10]. It is possible to adjust the perspective to capture the virtual stage by the operation on the tabletop. Thus, a user can verify the virtual stage from various angles. Florian Geyer et al created AffinityTable that supports collaborative design activities[11]. This system uses tabletop as a main workspace and an additional vertical display which is used to support reflection. MERL (Mitsubishi Electric Research Laboratories) uses wall-mounted displays and a tabletop, create a system that can be checked overhead view of the multiple perspectives at the same time[12]. The tabletop shows the overhead view. And three wall-mounted displays shows the overhead view from the multiple perspectives, such as way to the destination and overhead view shifted a little angle. This system allows the user to simultaneously confirm an image that includes a variety of information and perspectives from different positions depending on the application. These two systems show overhead views from a different direction. However, these systems do not use individual view and they cannot support space design from two kinds of viewpoint. Andy et al created a map system[13], in which a user sets the doll on the map on a tabletop, he can see how the terrain of the point of view of the doll. Therefore, they can see not only the overhead view, but also the individual view of objects in space and better understanding of space in more detail is possible. Movement of the viewpoint is done by moving the doll on the map, but it does not support an operation from the individual view. The gap of two viewpoints on operation still remains.

3. System design

In this study, we aim to provide an environment that shows overhead view and individual view. This system allows a user to move viewpoint object more easily and intuitively. As noted in section 2, in the conventional space design system, there is a problem of manipulation of viewpoint objects when the user is looking the individual view. The reason is that the system allows the user to operate it only on the overhead view. In order to reduce the gap of operation in the information space, we introduce gestural manipulation that can make intuitive operation while looking at the individual view.

For this new way of operation, we take following points into consideration.

(1) The gaze direction of the target of the operation must match the operator's gaze direction.

When performing the operation of the viewpoint object on the overhead view while looking at the individual view, it is necessary to consider the correspondence of spacial information from the individual view to overhead view.

This process sometimes causes difficulty in object manipulation especially when the user tries to move the viewpoint object continuously. There is a mismatch of the gaze direction of the target and the operator. Although the user wants to move forward on the individual view, he cannot necessarily see the direction as forward from himself on the overhead view, actual operation input device. To determine the course of the overhead view, it is necessary to confirm the orientation of the viewpoint object, and the placement of other objects. Gesture manipulation can operate the viewpoint object intuitively while looking at the individual view, because gaze direction of the target and the operator are match. Then, the gap between the two perspectives is resolved. Therefore, the user can determine the direction of the operation intuitively, such as “If the user wants to advance the viewpoint object, he does gesture for moving forward.”

(2) The gesture operation promoted in this paper is easy to remember.

Even though the judgment of the course of the viewpoint object can be intuitive, if gesture itself is hard to remember for the user, easy object manipulation is impossible. Therefore, in order to achieve easy manipulation of the viewpoint object, operation must be set to simple gestures that the user can remember easily.

In the proposed system, there are only three types of gestures: putting one’s palms together, extending one’s arm, raising one’s arm, and complex gesture is not used at all.

4. Implements

We create a system to support the planning of soccer tactics as an application. The system aims to support the planning process of the optimal positioning of each player in a variety of situations. In soccer, being aware of both the overhead view and the individual view is important. By using the overhead view, a user can be aware of the overall balance of formation and a sense of distance between the players. By using the individual view, the user is able to image the situation in the position of actual players. Hence, more real tactical planning becomes possible. In this section, we explain how to realize the functions and features of this system.

4.1. System architecture

As shown in Figure 2, the proposed system contributes the work environment that 3 users simultaneously conduct the work with DiamondTouch (DT) [6], which is one of tabletop interface. DT individually recognizes each user's operation. Also this system use a vertical display and Microsoft Kinect as a sensor. Kinect is developed by PrimeSense[14], and equipped with the depth camera and the RGB camera. It always measures the height and the position of the player, allows manipulation by gestures.
DT shows the overhead view, and the vertical display shows the individual view. By placing these devices as shown in Figure 2, the system creates a work environment that users can operate objects and recognize their information from multiple perspectives.

Figure 2. System architecture

4.2. Object operation

The function of this system is setting of perspectives to each object, rotation of the viewpoint, and movement of the viewpoint.If a user sets an object for viewpoint objects on the overhead view shown by DiamondTouch, perspective of the object is displayed on the vertical display as the individual view. Figure 3 shows the displays of each view.

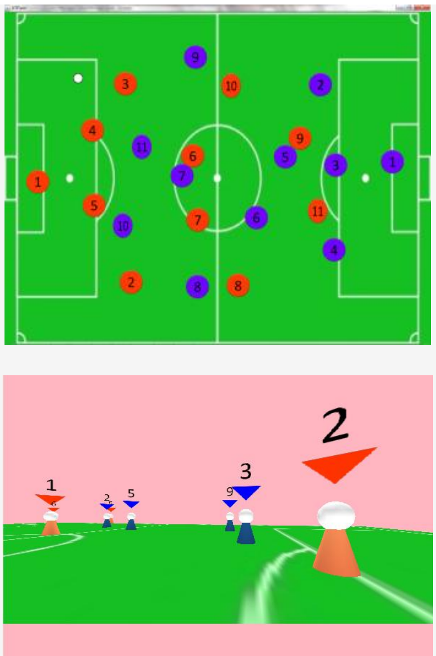

Figure 3. Displays of each view

On the overhead view, a sector is drawn around the viewpoint object as Figure 4. The user can determine which object being displayed its viewpoint as an individual view. That Sector represents a field of view. Orientation of the viewpoint object can be set by touching the edge of the sector. Movement of the object can be set by touching and dragging itself. Operations can also move to objects other than viewpoint object.
On the individual view, operation of viewpoint is possible by gesture. Detailed method of gesture operation described in the next section. All operations are linked between the viewpoints.With these features, this system is to achieve the operations of viewpoint object easily, and promote the recognition of spacial information.

Figure 4. Viewpoint objects on the overhead view

4.3. Gesture recognition

This system detects gesture operations by using the Kinect. There are three types of gestures. Putting one's palms together, extending one's arm, and raising one's arm. In this section, we describe what gestures are assigned and how each gesture is recognized. Kinect is placed at the bottom of the vertical display. All the gestures are performed while looking at the front of the body Kinect. Therefore, the user can operate the object while looking the individual view at the front.We use OpenNI[15] as framework released by OpenNI[16] organization. To detect gestures we use the skeleton tracking program of NITE[17]. NITE is the middleware libraries for treat raw sensor data. And we have installed SensorKinect[18] as a device driver for Kinect. We save the information of the skeleton 25 times a second, to increase the accuracy of the operation. Gesture operation is possible for translation and rotation of the viewpoint object.

,

- **4.3.1. Putting one's palms together.** When the distance between the left and right hands was within 30mm, it judges with having fit the palm. This gesture operation used to choose whether to move or rotate a viewpoint object.

- **4.3.2. Extending one's arm.** When a user issues a certain distance of his right hand in front of the body in the horizontal direction, it judges with extending the arm. After choosing the type of operation, the user determines the direction of the move and rotates by extending his arm.

Direction of the operation is changed continuously in conjunction with the movement of the arm. When moving an object, the orientation of the object is consistent with the direction of movement. So, feel like operator is moving the space actually.

- **4.3.3. Raising one's Arm.** When the user's left hand above the shoulder, it judges with raising the arm. Select the movement of the viewpoint object, and this operation can be moved without changing the orientation of the object.

5. Evaluation

We evaluate to assess the usefulness of the gesture operation of viewpoint object in an environment of multiple perspectives.In this experiment, the participants actually operate the viewpoint object in two conditions such as with and without the system. Then, through the results, we verified if this system could provide effective support in operating viewpoint object.The participants were asked to manipulate the viewpoint object in two ways to move courses on virtual space. Course is presented by the image capture of consecutive individual perspective as Figure 5. Capture images are the individual view that appears in the middle of the movement of viewpoint object through the course. This capture images is the point in the middle of when we move the viewpoint object actually. In two consecutive images, it is possible to guess the direction the viewpoint object has moved from the difference between the placement of objects in the image before and after. Participants will actually get to move the viewpoint object in accordance with that guess. One course consists of 5-7 pieces of captured image. In two consecutive images, so participants can guess the course, some of the objects that are displayed in the previous image were displayed in the after image.

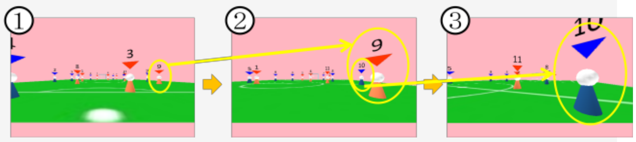

Figure 5. Capture images

We chose 14 participants. Some of them don't have any experience in soccer and some play soccer occasionally. 7 of 14 use the system for first time, then conducted without the system second time. The rests did in reverse order. Time limit is not provided. Before the experiment, the participants were explained way of touch operation on the overhead view and gesture operation on the individual view. While moving the course, each participant was asked to make sign when the object passed the location that was displayed in capture image. In order to measure quantitatively whether correct direction to move.we quantitatively analyze the deviation of the coordinates as error. The deviation is difference between a coordinate that we captured and a coordinate that the participant indicated when operating the object. We show an example calculation of the error.

Table 1. coordinates example

	Correct coordinates		User's answer	
	x	y	x	y
Point 1	906	319	916	329
Point 2	709	197	736	157
Point 3	462	217	497	201

Table 1 shows correct coordinates and one user's coordinates in the question presented by Figure 5. We calculated the total error by adding the error of each point. We get the error of this user at Point1 in this way.

$$\text{Error at Point1} = \sqrt{|916 - 906| + |329 - 319|}$$

After solving the problem of six, questionnaire was conducted. The questionnaire evaluate ease of operation remember, rotation and translation of the viewpoint object, and the simplicity of the problem. The questionnaire is a five-step evaluation.

5.1. Result

The result of this experiment is shown in the Table 2. 5 means "very agree", and 1 means "very disagree".Figure 6 shows the proposed and conventional method of total error average in each question.

Table 2. Impression of operation

	Tabletop[point]	Gesture[point]
Easy to remember how to operation	4.5	4.2
Easy to move the object	3.6	3.9
Easy to rotate the object	3.3	4.2
Easy to solve the ploblems	3.5	4.0

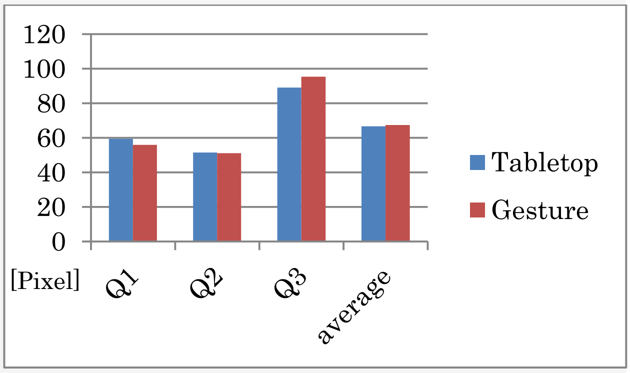

Figure 6. Error of operation

5.2. Consideration

Looking at Table 2, Tabletop manipulation has a high reputation in fields of ease of operation remembered. However, gesture manipulation obtained 4.2 point, is easy enough to remember. Why tabletop manipulation got high point is because of the spread of device that using the touch panel in recent years. The participants frequently use touch device in life and they are accustomed to the touch operation.Gesture manipulation got higher point than tabletop manipulation at the ease of rotation and movement of viewpoint object. The reason is the gaze directions of object and operator are matched, and therefore the user can operate viewpoint object intuitively. From this, the proposed system achieved eliminating the gap between the two viewpoints of the operation. Gesture operation also got higher point than tabletop manipulation in the ease of solving the questions. Therefore, proposed system was able to achieve easy object manipulation.
As gesture operations were sensuous, there was concern the accuracy falls in gesture manipulation comparing to tabletop manipulation. However, when we look at Figure 6, the conventional method and proposed method has become almost equal error. This allows the user to perform operations also table top and the same accuracy in the proposed method.
By the results of these experiments, proposed system allows the user to rotate or move viewpoint object more easily, and with an accuracy equivalent to operation from the tabletop.

6. Conclusion

We use two kinds of viewpoints for spacial information recognition frequently. The overhead view and the individual view give entirely different spacial information even they indicate the same location in the space. However, there are few systems to support the simultaneous use of two different viewpoints. Although they show multiple viewpoints, operation is limited to only from the overhead view. The processes that the user has to translate information from one view to another caused interfere with intuitive operation. That is the gap between viewpoints. Then, this paper proposes a new space design system. We use touch and gesture interaction to operate the objects. In the system, the user can accuracy manipulates all the objects on the overhead view. Moreover, when the user attempt to operate the viewpoint object, the gaze direction of the target and the operator are match. The user can rotate or move the viewpoint object by simple gesture. The proposed system allows the user to move any object intuitively. We presented a user evaluation that assesses the usefulness of the gesture operation. By the results of the experiments, proposed system allows the user to rotate or move viewpoint object more easily, and with an accuracy equivalent to operation from the tabletop.

7. Acknowledgement

This work was supported in part by a Grant-in-Aid for Scien-tific Research(A), 2012, from the Ministry of Edu-cation, Culture, Sport, Science and Technology in Japan.

References

[1] Microsoft, http://www.microsoft.com/

[2]Kinect, http://www.xbox.com/en-US/kinect

[3]Google Street View, http://maps.google.co.jp/

[4] BLIZZARD. http://www.blizzard.com/war3/.

[5] THE SIMS 2. http://www.thesims.jp/thesims2/.

[6] P. Dietz, D. Leigh, "DiamondTouch: A Multi-User Touch Technology", In *Proceedings of UIST '01*, ACM, NY, 2001, pp. 219–226.

[7] Tobias Franke, Svenja Kahn, Manuel Olbrich, Yvonne Jung, "Enhancing Realism of Mixed Reality Applications through Real-Time Depth-Imaging Devices in X3D", , In *Proceedings of Web3D '11*, 2011, pp.71-79.

[8] Mark J. Prusten, Michelle McIntyre, and Marvin Landis. "3D workflow pipeline for CAVE virtual environments", In *Proceedings of SIGGRAPH '05 Posters*, 2005.

[9] Tonchidot Corporation, http://www.tonchidot.com/

[10] Yosuke Horiuchi, Tomoo Inoue, Ken-ichi Okada, "Virtual theatrical space linked with a physical miniature stage for multiple users' easy image share", Transactions of the Virtual Reality Society of Japan, Vol.16, No.4, 2011.

[11] Geyer, F., Pfeil, U., Budzinski, J., Hochtl, A., Reiterer, H., AffinityTable, "A hybrid surface for supporting affinity diagramming", , In *Proceeding of the 13th IFIP TC 13 Int'l Conf. on Human-Computer Interaction*, Springer-Verlag, 2011,pp.477-484,

[12] Clifton Forlines, Alan Esenther, Chia Shen, Daniel Wigdor, and Kathy Ryall, "Multi-user, multi-display interaction with a single-user, single-display geospatial application", In *Proceedings of UIST '06*, ACM, Montreux, 2006, pp.273-276.

[13] Andy Wu, Derek Reilly, Anthony Tang, and Ali Mazalek, "Tangible navigation andobject manipulation in virtual environments", In *Proceedings of TEI '11*, ACM, Funchal , 2011, pp.37-44.

[14]PrimeSence , http://www.primesense.com/

[15] OpenNI, http://openni.org/

[16] OpenNI, http://openni.org/about

[17]NITE, http://www.primesense.com/nite

[18]SensorKinect, https://github.com/avin2/SensorKinect