NAVIGATION IN VIRTUAL REALITY USING MICROSOFT KINECT

Florian Kammergruber, Andreas Ebner, Prof. Willibald A. Günthner
Technische Universitaet Muenchen, Institute for Materials Handling, Material Flow, Logistic, Germany

ABSTRACT: *Currently interaction in a virtual reality system is performed by the help of a tracking system, or with some special input devices. Some examples of these well-known devices are motion platforms, haptic interaction devices with force feedback, data gloves, and navigation devices like a flystick or a game controller. Not all of them can be used in the construction industry regarding their sturdiness and ease of use. There are continuously new developments, especially in the area of mobile input devices. These are driven by the game console market, which satisfies the requirements of the users for wireless input devices and for innovative approaches which integrate the player interactively with his body movements in the game. The latest of such devices was developed by Microsoft for the Xbox game console: "Kinect". This device processes video and infrared information of a camera system to enable markerless tracking, which is based on certain geometries and reference images of the user. The recorded three-dimensional body movements are used to create a skeletal model of the user and his entire body can be tracked.*

The use of these properties allows an officially released SDK from Microsoft. During the development of new interaction concepts for highly mobile virtual reality systems, the Kinect has been integrated into a virtual reality system as an input device for navigation. With a mobile Virtual Reality system in a building container, the system can be used for planning purposes and navigation through the construction site. Predefined gestures and movements of the user enable a totally free and intuitive navigation and control of the system in the immersive environment. The implemented approach at the Technischen Universitaet Muenchen is characterized by extremely low-cost hardware, and coded as a plug-in for the RTT DeltaGen software.

KEYWORDS: *navigation, interaction, Microsoft Kinect, gesture tracking*

1. 3D NAVIGATION IN VIRTUAL REALITY

An essential property of Virtual Reality (VR) is its interactivity, which allows the user to interact with the virtual world. Numerous input devices that serve as a man-machine interface within the VR system were developed for it. Possible characteristics of the interaction are navigation in the virtual world, tampering with objects and feedback via a force feedback. The requirements for the interaction are speed, purposefulness and especially intuitive handling, which may not have a recognizable delay between input and output for real time suitability and has a high positional accuracy in space. Using suitable and intuitive operating metaphors, the conducted actions and gestures of the input devices are converted into movements and actions in the virtual world. This is necessary due to the abstraction of the model character, as the 3D model is only displayed but does not exist in reality. The operating metaphor therefore is based on the logic of how human body movements can be used for a control to steer something which does not exist in reality.

For most interactions in VR or input devices, a continually tracking or a spatial determination is necessary (Ong and Nee, 2004). Spatial movements in the real world are transferred to the digital and virtual world with the help of the calculated coordinates. According to the specific application and purpose of usage, input devices can be used without a tracking system.

1.1 Optical Tracking Systems

Due to their basic physical principle tracking systems, they can be divided into acoustic, electromagnetic, inertial, mechanical and optical tracking as well as hybrid systems. In the last few years, optical tracking systems have been the most prevalent. An optical tracking system operates with one or more digital cameras to record objects that will be followed in the tracking area. The tracking area can be extended scalable by using additional cameras. Video or infrared cameras are used for orientation within the tracking area using reference geometries like marker, targets, shapes or edges (Wagner, 2005).

The optical tracking using infrared cameras is especially popular and excels in its high precision and sturdiness by the exact calculation of position even in the case of rapid movements. A "flashing" infrared light sent out by the cameras is reflected by the special markers. Using the known constellations and dimensioning of the single targets, the position of the target recorded out of different perspectives can be calculated using triangulation. These infrared systems with cameras, software and special markers are very expensive to purchase. The reference geometries can be of passive or active form whereby passive targets are recognized by the cameras from their pattern or their solid geometries. In contrast, active targets send out their own signals by being equipped with LEDs, which send out infrared light (see Fig. 1).

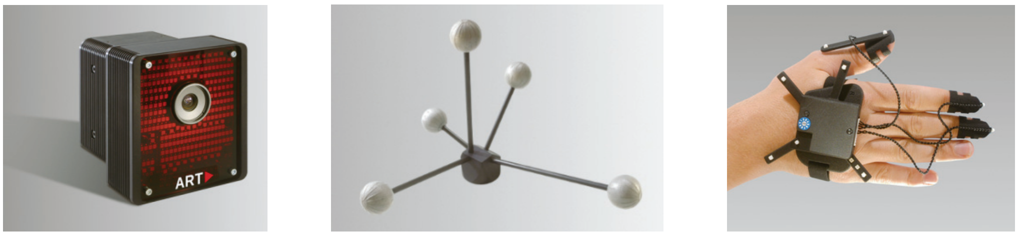


Fig. 1: Infrared tracking camera as well as a passive and active target (ART, 2012)

Aside from the infrared systems, optical image processing systems are increasingly used in which a video camera (in the simplest case a webcam) evaluates the picture of the surroundings with image-processing software. This technology is most frequently used in the area of augmented reality. In this case, the system orients itself from the targets, which are mostly made up of 2D matrix codes and have a black/white pattern that can easily be processed. Such video systems cost less than infrared systems, as the camera and the software are significantly cheaper. In return, the tracking is less sturdy and depends a lot on the lighting conditions as well as the angle between the camera and the 2D matrix codes.

1.2 Input Devices

For the complete depiction, an overview is given of common - and in some cases - special input devices as a man-machine interface to VR systems, which are especially intended for navigation in the virtual world. In the area of mobile input devices in VR, there are constantly new developments or prototypes designed for specific applications (Bowmann, 2005). On the other hand, there is the market for gaming consoles which creates an increasing variety of mobile interaction devices. This is based on the market demands of the users for wireless interactive devices and on the demands of the providers who, by using innovative approaches, try to interactively integrate the gamer into the game.

1.2.1 Motion platforms

To depict the natural walking movement of a person in VR, different technical concepts were developed to enable a movement close to reality. The problem is that the user can move unrestrictedly in a virtual and three-dimensional world, but is limited by the size of the room in the real world. Therefore, the developments aim to find a solution where the user moves on the spot, while the floor moves under him. The technical solutions range from the "Virtusphere", a pivoted hollow sphere (Medina et al., 2008), the "CyperCarpet" a turning table set up with many spheres and a treadmill in front of a power wall (Wulz, 2008). Fig. 2 shows the mentioned solutions.

Fig. 2: Motion platforms Virtusphere, CyperCarpet and treadmill

These input devices are technically very sophisticated, but inflexible in their assembly position and very expensive to build.

1.2.2 Navigation devices

Aside from the motion platforms, there are navigation devices that allow navigation in six degrees of freedom. One of these navigation devices is the game controller (gamepad) from the area of gaming consoles and computers (Riva et al., 2007). It has an ergonomic design and is used by gamers to move within the game world and carry out various actions.

Fig. 3: “RumblePad 2” made by the Logitech Co.

This wireless device has a range of up to 10 meters. As seen in Fig. 3, there are two analogue steering elements (control sticks) for command input as well as an 8-way direction block and ten freely programmable buttons.

A further input device usable for VR is the Wii controller distributed by the Nintendo Co. for their gaming console, which was already modified for a use in the area of VR (Lee, 2008). The ergonomically-shaped device can be controlled simply by pressing a few buttons and is connected by Bluetooth. As a new and popular feature, the Wii controller has three integrated acceleration sensors which recognize the position of the controller in the user’s hand and its acceleration.

Fig. 4: Nintendo Wii controller with Nunchuck extension

The device, which can be obtained for almost $30, can be used individually or with further accessories as the Nunchuck, for example. Fig. 4 shows the Wii controller and the separately available Nunchuck extension, which has a control stick and more buttons.

Aside from the Wii Controller, the Nintendo Co. has also developed a “balance board” for the Wii gaming console for interaction with the feet (see Fig. 5). It is equipped with four corner pressure sensors that record the applied forces and thus the exertion of force by the feet or the body (center of gravity). The balance board is battery-powered and connected by Bluetooth.

Fig. 5: Nintendo Wii balance board

The balance board was already successfully applied in another VR project of Haan et al. (2008) and also by Hilsendeger et al. (2009) as an affordable interactive medium. They used the possibility of calculating the body center of gravity to navigate dynamically in the VR using the user’s weight relocation.

However, input devices - which imitate a pointer and can thus portray a beam metaphor - are the industry's most-used solution for interacting and navigating in VR. For these devices, the term flystick has been commercially established, as it is very suitable for the flying through virtual scenes. Fig. 6 shows two different flysticks that are controlled like a pistol and can be bought for approx. $1,000.

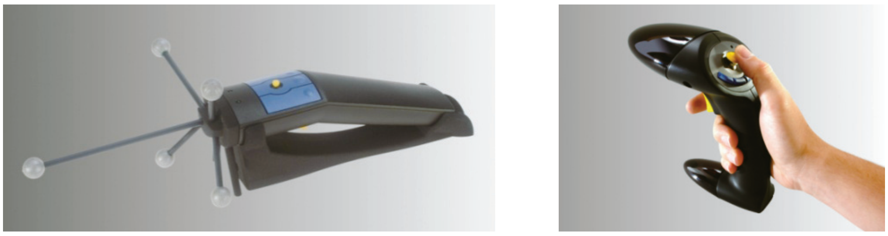


Fig. 6: Flystick models made by the A.R.T. Co. (ART, 2012)

Usually, the flysticks are connected with the VR control computer by wireless and allow navigation in all six degrees of freedom. With their beam gesture, they are very suitable to operate virtual menus in three-dimensional space (Bormann 1994). For navigation and interaction, the flysticks must be equipped with passive or active tracking targets so their position can be determined by optic tracking systems. Additionally, the flysticks are equipped with several control buttons where defined actions can be controlled.

1.3 Summary

In tracking procedures, it is shown that the optical infrared systems dominate the current VR equipment. Acoustic, electromagnetic, inertial and mechanical ones are only rarely used for commercial purposes. The purchase and usage of professional optical infrared tracking systems must still be assessed for approx. $15,000 (ART tracking). This high purchase price is an obstacle for many companies using VR systems. Furthermore, the pursuit of objects and people demands equipping them with suitable tracking targets.

One can see that for VR input devices, the solutions and products shown are also partially very well engineered and can be used in the industry without a problem. These devices can, however, be very expensive and unwieldy, especially if they are motion platforms or haptic input devices with force feedback. This paper describes the development and adaptation of new intuitive operating metaphors and new interactive concepts that can be implemented with the help of Microsoft Kinect. In the foreground is the very mobile usage of the input device as well as its very low purchase price. The device is able to unite the tracking system and the input device into one single system unit. The input has to be done as intuitively as possible and adjusted to the needs of navigation in VR.

2. MICROSOFT KINECT

The interactive device Kinect (see Fig. 7) developed by the Microsoft Co. for the Xbox gaming console was launched into the market in November of 2010. Initially, it was intended as an extension of the gaming console to create a more interactive and natural gaming experience. Using gestures and movements, the user is able to interact with the game world. Suitable games were developed for the specific use of this type of game navigation.

Fig. 7: Microsoft Kinect camera

However, it did not take a long for the first computer experts to dare to modify the camera. The Kinect was connected to a PC and drivers for it were developed by them. Originally, this was not officially supported by Microsoft. The first attempts of the software developers ranged from a three-dimensional recording of rooms and objects, to a multitouch interface where photographs could be rotated and enlarged with two hands. Non-experienced developers who used this technology that had been developed for years obtained fantastic results from Kinect within a few days. Early in 2012, the first Kinect camera intended for use with a PC was released. It featured an improvement in the close range of the tracking: Kinect for Windows 7. An official SDK was released for this. The new Kinect was used for the functions and interactions described in this paper.

The Kinect is an input device for the detection of movements. The camera system enables the recording of three-dimensional body movements of a person using the collaboration of a 3D camera which has optical and infrared sensors and a stereophonic microphone. It combines both video and infrared information. From the recording of the RGB camera and the depth information, the software calculates a skeleton model of the user with the corresponding nodal points. Arms, legs, head and torso of the skeleton model move with only slight delay synchronous to the user. Even in the case of changing users with different body sizes and measurements, the camera recognizes their contours without problems. Astonishingly, neither calibration for a user nor special markers on the body are necessary. The Kinect combines both the functioning of a tracking system and the properties of an input device in a system environment. It enables the control and interaction with a PC through a Natural User Interface (NUI) using gestures. The skeletal information has spatial coordinates of joints in a human body. The x,y,z coordinates of joints are referenced to the Kinect hardware. The following joints are utilized for the recognition of gestures in our project: head, left elbow, right elbow, hip center, hip left, hip right, shoulder center, shoulder left, shoulder right, left wrist, right wrist, left hand and right hand. Fig. 8 shows all the joints and the RGB picture.

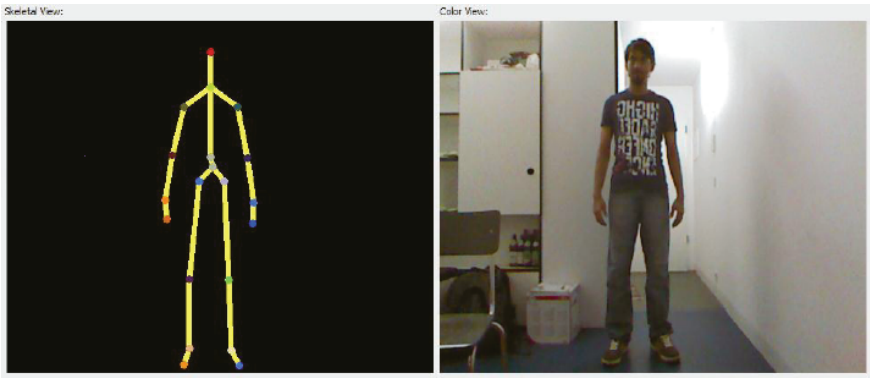


Fig. 8: Skeletal View by the Kinect camera und RGB picture by the camera

With the skeleton model supplied by Kinect with its joint points, navigation in VR can be enabled using defined gestures. For this, it is necessary to define the single gestures for the different movements and determine them according to the joint points. In this paper, Version 10.0 of the RTT Deltagen VR Visualization software is used. As a research partner of RTT, it is possible to obtain an SDK for RTT Deltagen, which enables it to access the position and orientation of the camera and manipulate it. With the SDK by RTT a plug-in was developed that reverts to the data of the Kinect camera and implements a corresponding translation and rotation of the observer camera using the different gestures. Thus, the position of the observer in the digital scene can be controlled and navigation in VR using gestures is enabled. The preceding papers allowed a navigation using different methods, all of them demanding an input device on the user. For the methods developed so far, expensive equipment is necessary. In this project, Kinect is suggested as an inexpensive tool for control and navigating. The advantages of NUI are: low learning effort for using and controlling, learning by doing, and securing a direct interaction.

2.1 Definition of Gestures and Implementation

A small software was developed to obtain and gather the data from the Kinect sensor and to detect the gestures. With the help of information on joints, many gestures are recognized. The coordinates are obtained and drawn in an application. The red circles represent the right hand and the green circles the left hand. The blue circles have head, spine and hip joints. The speed is shown in the scroll bar and a text box. The detected gesture is shown in a text box. A provision is made to adjust the angle of the Kinect camera as well. The following Fig 9 shows the described software window in the idle position.

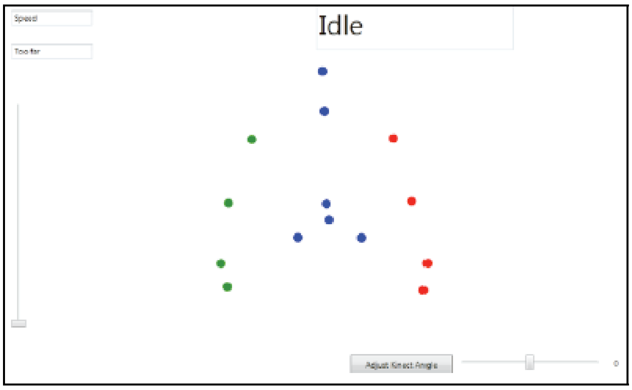
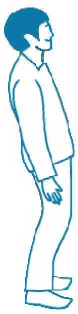
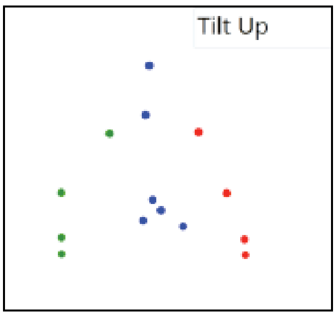
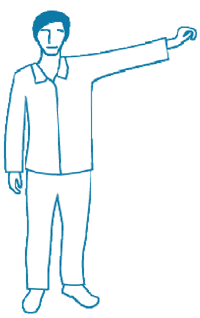
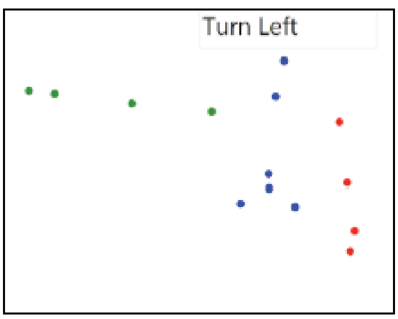
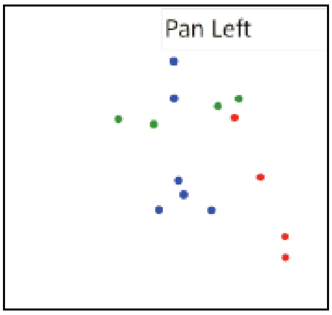
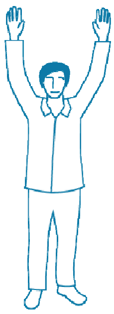
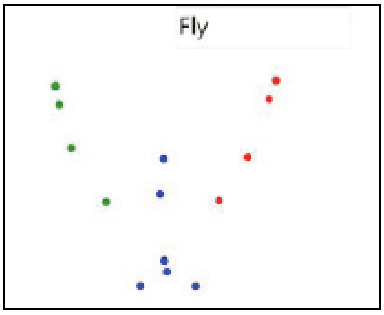
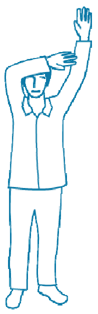
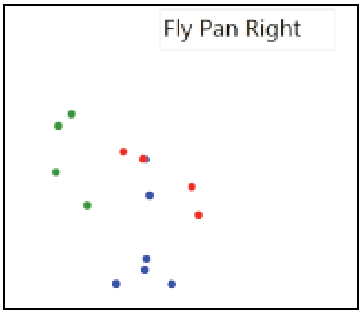


Fig. 9: Software window depicting the idle state

In the next step, it was necessary to determine solid borders or areas for the recorded joint points. From these, the software can recognize which gestures are made by the user at the current point in time. The first column of Table 1 shows a few of the defined and combined gestures that the user can make in front of the VR system in the recording area of the Kinect camera. Of course, the corresponding opposites of the gestures shown such as turn right, tilt down, pan right, etc also exist. Column 2 shows the recorded picture of the developed software and the recognition of the gesture by the software. Column 3 describes the characteristic properties of the gestures, so that they can be recognized as well as the resulting navigation movement.

Table 1: Gestures with Software Recognition and Characteristics

Gesture	Software Recognition	Characteristic
		Bend backwards with hands down → Tilt Up
		Raise the left hand above shoulder sideways with the right hand down → Turn Left

		Move the left hand across body with right hand down → Pan Left
		Raise both arms → Fly, move forward
		Raise both arms and move right hand across head → Fly and Pan Right

The source code necessary for development is displayed exemplarily in Figure 10. In it, the single rules for the software to recognize and differentiate the gestures are stored. The x,y,z coordinates of the different joints are taken as a basis and the gesture is defined according to their relative position to one another. If, for example, both wrists are above the hip and the head, the gesture “Fly” is recognized. If, furthermore, the x coordinate of the elbow is smaller than that of the shoulder joint, the gesture “fly pan right” is recognized. In this case, the origin of the coordinate systems lies on the middle hip joint point.

```
if (handright.Position.Y > 0 && handleft.Position.Y < 0)//Left hand down & right hand moving
{
    if (elbowright.Position.Y > shoulderright.Position.Y)//right hand raised sideways
        txtinstructions.Text = "Turn Right";
    if (elbowright.Position.X < shoulderright.Position.X)// right hand across spine
        txtinstructions.Text = "Pan Right";
}
if (handleft.Position.Y > 0 && handright.Position.Y < 0)
{
    if (elbowleft.Position.Y > shoulderleft.Position.Y)
        txtinstructions.Text = "Turn Left";
    if (elbowleft.Position.X > shoulderleft.Position.X)
        txtinstructions.Text = "Pan Left";
}
if (handleft.Position.Y > 0 && handright.Position.Y > 0)//Both hands above waist
{
    if (handleft.Position.Y > head.Position.Y && handright.Position.Y > head.Position.Y)
        // Both hands above head
    {
        txtinstructions.Text = "Fly";
        if (head.Position.Z < hipcenter.Position.Z)// Bend Forward
            txtinstructions.Text = "Fly Tilt Down";
        if (head.Position.Z > hipcenter.Position.Z + .2) //Bend Backward
            txtinstructions.Text = "Fly Tilt Up";
        if (elbowright.Position.X < shoulderright.Position.X) //Right hand across head
            txtinstructions.Text = "Fly Pan Right";
        if (elbowleft.Position.X > shoulderleft.Position.X) //Left hand across head
            txtinstructions.Text = "Fly Pan Left";
    }
```

Fig. 10: Pseudo code explaining the function determining the gestures with comments

In investigations, these gestures have been proven to be resilient and intuitive and can be used for the navigation in six degrees of freedom in the VR. For that, the developed gestures are sorted out from the software and recognized as described. Then they are transferred to the RTT plug-in, which implements these gestures into a translation and/or rotation of the observing camera.

2.2 Possibilities and Limitations

One of the greatest obstacles for the application of the Kinect camera was the SDK of the Kinect, which was written in C#. However, the programming language C++ was necessary for an integration and application in the RTT SDK. Therefore, it was necessary to rewrite the Kinect SDK into C++. Unfortunately, the support by Microsoft and the templates for the SDK in Kinect in C++ were very rudimentary.

Basically, you can conclude that the application of Kinect in VR offers new cheap possibilities in man-machine communication. Each user has the possibility of independently defining new or different gestures and using them for the interaction. The approaches analyzed in this paper only sufficed for the suitability of Kinect as a navigation interface. A direct manipulation of objects in the virtual world and an interaction with them was not implemented. For this operation, new operating metaphors would be necessary, as the Kinect camera cannot interpret and evaluate the hand's grabbing motions, for example. The problem of the somewhat coarse recording and tracking of the user indicates that a head tracking is currently not possible. Thus, one cannot deliver a user-specific perspective of the standard view in VR systems. Further developments concerning accuracy of resolution of the Kinect must be awaited. Moreover, the condition of light and the surroundings also present a challenge. The room cannot be too dark, as no clean tracking of the user via the camera can be obtained. Often this is exactly the case in VR systems and could be a problem area in the application of the camera in CAVE facilities.

Aside from recording limitations, ergonomic aspects also play a role in evaluating the Kinect. Human beings are not made for moving a long time with outstretched arms, as they become heavy and start hurting. This especially happens during longer navigation and application sessions. For short VR sessions, this problem is not a limitation. However, this problem also occurs in existing interactive devices. Regarding ergonometrics, it can also be determined that the Kinect as such does not supply haptic feedback to the user, neither for the activation of a function (e.g. pressing of a button) nor for possible collisions within the virtual world.

3. USE IN THE CONSTRUCTION INDUSTRY

In construction planning, it is necessary to create the possibility of analyzing selected conditions in a spatial model. In the area of VR visualization, there are many previous research papers that show how construction sites can be depicted in VR (Doulis et al., 2007) for detecting collisions between static elements of the construction site and observing the current level of planning and progressed. Generally, VR is thus a useful instrument to create an improved vision and a common view with reference to future intentions.

Due to external influences at construction sites, there are often unexpected delays that, in turn, change the construction process. The detailed plans at the beginning of the construction thus become obsolete and another plan must be created at once and all construction site workers must be informed about the newly planned processes. With the help of VR, possible planning mistakes can be easily recognized and all participants can be given a clear understanding of the planned alterations. Current planning can be visualized graphically and all people participating actively at the construction site can discuss the corresponding problems together. For use in construction sites, it is necessary for VR, to employ highly mobile VR systems. So construction site workers accept the new technology, a simple intuitive and especially resilient solution for the interaction with the system is necessary. The navigation via the Kinect camera offers these properties. Fig. 11 shows, as an example, how a mobile VR-system could be integrated into a construction site container. A large visualization screen enables a detailed and comprehensive observation of the construction site. The size of the screen also permits mutual observation by several participants affected by the planning. The Kinect camera set above the visualization enables a free and intuitive navigation in the immersive environment. Using the explained gestures, the user can navigate freely through the VR visualization and emphasize the observation of selected details. The navigation can be done by any user who stands in the center of the observation area of the Kinect camera.

Fig. 11: Construction trailer with integrated mobile VR-System and Kinect navigation

Using the research approach explained here, a prototypical implementation of control by gestures for mobile VR systems was created that also enables a sensible usage of VR in the construction industry at the construction site. Construction managers and planners can react to altered fringe conditions during construction and better assess the consequences of their actions. Due to the transparency obtained by VR, planning during and before implementation can be improved and costly mistakes in the construction process avoided. The intuitive control by hand gestures lowers the inhibition level of users concerning the usage of the system and simplifies the interaction with the VR compared with current interactive devices. Furthermore, the Kinect camera significantly lowers the investment costs for a mobile VR system as no expensive infrared tracking systems is necessary. The investment for the hardware is thus reduced to a visualization display (flat screen), a powerful graphics computer and the Kinect camera. This system and the developed interaction techniques to visualize accompanying construction should be implemented and tested in a pilot construction site at the Technische Universitaet Muenchen to evaluate the potential of the new concept.

4. ACKNOWLEDGMENTS

The author gratefully acknowledges the support of the TUM Graduate School's Faculty Graduate Center Mechanical Engineering at the Technical University of Munich.

This research project was performed on behalf of the Federal Logistics Association (BVL) and funded by the Federal Ministry of Economics and Technology via the Federation of Industrial Cooperative Research Associations "Otto von Guericke" (AIF) by a resolution of the German Bundestag.

5. REFERENCES

ART - Advanced Realtime Tracking GmbH (2012): Company website, Retrieved January 27, 2012 http://www.ar-tracking.de/.

Bormann S. (1994): Virtuelle Realität – Genese und Evaluation, Bonn: Addison-Wesley.

Bowman D.A., Kruijff E., LaViola J., Poupyrev I.: 3D User Interfaces. Boston: Addison-Wesley, 2005.

Doulis M., Vogel M., Pflüger J., Rietmann M., Raps M. (2007): 4Dive - A 4D Interface for the Visualization of Construction Processes in a Virtual Environment. Proc. of the Conference on Construction Applications of Virtual Reality, 2007.

Haan de G., Griffith E., Post F. (2008): Using the Wii Balance Board as a low-cost VR interaction device, In: Proceedings of the 2008 ACM symposium on Virtual reality software and technology, Bordeaux, 289-290.

Hilsendeger A., Brandauer S., Tolksdorf J., Fröhlich C. (2009): Navigation in Virtual Reality with the Wii Balance Board, In: Virtuelle und Erweiterte Realität, 6. GI-Workshop VR/AR, Aachen: Shaker, 269-280.

Lee J. (2008): Hacking the Nintendo Wii Remote, In: IEEE Pervasive Computing, Vol. 7, 39-45.

Medina E., Fruland R., Weghorst S. (2008): Virtusphere: Walking in a Human Size VR “Hamster Ball”, In: Proceedings of the Human Factors and Ergonomics, Vol. 52, 2102–2106.

Ong S. K. and Nee A. Y. C. (2004): Virtual and Augmented Reality Applications in Manufacturing, London: Springer.

Riva G., Mantovani F., Capideville C., Preziosa A., Morganti F., Villani D., Gaggioli A., Botella C., Alcaniz M. (2007): Affective Interactions Using Virtual Reality: The Link between Presence and Emotions, In: CyberPsychology & Behavior, Vol.10, Nr. 1, 45-56.

Wagner M. (2005): Tracking with multiple sensors, Dissertation, Munich: Technische Universitaet Muenchen.

Wulz J. (2008): Menschintegrierte Simulation in der Logistik mit Hilfe der Virtuellen Realität, Dissertation, Munich: Technische Universitaet Muenchen.