Towards better interactions with a 3D deformable model with your hands

D. Lingrand[1]* S. Meunier[1] Ph. Renevier[1] A.-M. Pinna-Déry[1] M. Riveill[1] J. Soula[2]

[1]Rainbow - I3S Univ. Nice - Sophia Antipolis / CNRS [2] CSTB Salle Immersive Le Corbusier
Sophia Antipolis - France

ABSTRACT

In this work, we present a low-cost and modular prototype dedicated to study 3D interactions in order to easily deform a 3D models using hands. This is done using a simple colored glove coupled with a 3D motion recovery engine and 3D visualization.

We present different interaction schemes that we want to study using real interactions rather than a wizard of oz. This work is a first step towards different questions on 3D interactions we want to experimentaly address.

Keywords: gestures, 3D interaction.

Index Terms: H5.2 [Information Systems]: INFORMATION INTERFACES AND PRESENTATION—User Interfaces;

Figure 1: LEFT: User wearing stereo glasses in front of the system . On the left computer, the VRPN server and the two webcam widgets with the colored LEDs detected. On the right computer, the VRPN client with the 3D rendering. RIGHT: Focus on the glove extremities

1 MOTIVATIONS

There is a lack of a user-friendly and intuitive interaction device and technique for interacting with 3D objects. For instance, segmenting 3D data or designing 3D models remain cumbersome tasks involving heavy user interaction through usual 2D devices.

Mice are widely used for 2D interactions but there is no consensus on a single 3D device, merely because of the high cost, a lack of applications and most importantly the low ergonomic making their usage difficult. As pointed out in [2], the wide variety of 3D applications makes it better to have several specialized devices well adapted to some group of application than a single one.

Taking into account the fact that modeling and deforming 3D forms using plasticine seems to be natural to everybody and that deforming a 3D surface on a computer using mice and keyboard is not so obvious, we decided to use the hands for such interaction.

In [7], we interviewed with users familiar to 3D virtual environments and aware of the difficulties of interaction in such environments with 2D mouse and keyboard. The interviews were made in front of slides projected on a screen and the users were asked to interact with hands in order to do several actions such as objects selection, displacements and deformations. No constraint on hands motion was imposed. For some actions such as object selection, the results were coherent and it was possible to propose intuitive patterns of interaction. In some other cases, no consensus could be extracted from the users feedback, partly due to the unconstrained framework but also due to the difficulty of simulating such interactions with a wizard of oz. Thus, we decided to build a real platform in order to test different schemes of interaction and to compare them on real actions. This work constitutes a first step and is based on the separation of commands from 3D displacements. It is built with very cheap components in order to easily modify it and produce several prototypes at low cost. The hardware components may be improved when we are fully satisfied with the prototype developed.

*e-mail: lingrand@i3s.unice.fr

2 RELATED WORK

Due to their dexterity and easy training, hand motion alone or with instruments have been employed with the aid of Computer Vision in HCI since several years [9, 8, 1, 3, 2, 12, 6].

Computer Vision based approaches using free hands for gesture recognition are still in a stage of research [8] even if some results are promising [12]. The use of these technique must be either in controlled environment (lighting ...) either limiting the number of recognized gestures. As for example, in the GestureVR [9], three different hand gestures are used, combined with orientation of 2 fingers.

Thus, numerous works use intermediate objects manipulated with the hands, more easily detected by vision algorithms. The Magic Table [1] uses the detection of colored tokens on a white table, the VisionWand [3] is a passive wand with colored extremities in order to facilitate the visual detection.

The motivations of the different works are manipulation of objects [1, 3], pointing to objects or menus [8], gesture recognition for gesture command [10]. In the different approaches for interacting with a 3D world, there are mainly two different kind of actions: real spatial interactions and command interactions, such as mode selection or menu selection which are mostly planar.

In this work, we decided to split the actions into different modes of interactions and to switch from one mode to another using few very simple gestures, mostly in a plane parallel to the focal plane of the cameras. The set of gestures is of limited size, which is mandatory both for the user (more gestures imply more learning) and the recognition system. We will now present the implemented platform.

3 PLATFORM DESCRIPTION

In order to make the integration of our device with Virtual Reality Systems possible, we have chosen the VRPN library [11] and architected the system as a server/client (see figure 1). The server is responsible for the acquisition of the 3D positions of the fingers and for sending them regularly to the client. The client is responsible for applying the displacement and deformations to the scene and for the rendering process.

3.1 The 3D device

Gloves are hand made from thin black gloves with different colored LEDs located at each finger. Colored LEDs are easily segmented and localized in 2D using the HSV color system. This color representation provides the color value independently from its intensity and saturation, providing more robustness against varying lighting conditions. As shown on right of figure 1, several LED for each finger allow to avoid occlusions when the finger is not exactly in front of the cameras.

The stereo acquisition is done using two webcams aligned in order to use the simplification of epipolar geometry in the case of standard geometry. Thus, the depth is easily recovered: depth of fingers is inversely proportional to the 2D disparity between the 2 projections [5]. The precision obtained on 3D coordinates is sufficient to be able to deform 3D surface with short training. As a feedback to the user, detected positions of fingers extremities are rendered using small colored spheres in the 3D scene. The 3D scene is displayed using stereo rendering to enable visual feedback to the user, and the geometry is managed using the CGAL [4] library.

3.2 Modes of interaction

We developed different modes of interaction for objects deforming and moving. In the **vertex mode**, a vertex is attached to a finger when the finger is detected in his 3D neighborhood. It is detached when the finger moves rapidly away from the vertex. When attached, the vertex is moving according to the corresponding finger motion. Depending on user's ability to move separately different fingers, several vertices can move simultaneously. In the **facet mode**, one finger is attached to a facet which translates along its normal according to the translation component on this axis of the finger displacement. Edges that are selected are deleted in the **edge deletion mode** and geometry is thus updated avoiding holes in the surface. In the **extrusion mode** a facet is extruded along its normal. In the **facet division mode**, when 2 vertices are selected, the corresponding facet is divided into two parts, adding a new edge.

The different modes of interaction can simply be activated using gestures made from the combinations of four discretized directions: up, down, left and right. In the following table, we present a possible set of Gestures for activating different modes, choosing gestures closed to mnemonic when possible ("V" for Vertex mode, ...).

vertex mode		vertex deletion	
facet mode		extrusion	
edge deletion		facet division	
orientation		translation	

The gesture capture mode is enabled when the thumb and the major fingers are close together and become disabled when these fingers are spread. In this current work, the dominant hand is both responsible of 3D interactions and gestures commands. This could easily be changed if needed in this architecture.

4 Conclusion and perspectives

The platform described in this paper is a first prototype that allows users to build and deform 3D objects using 3D finger displacements in different interaction modes selected by gestures, both in office or immersive room environments. This has been built in order to be easily modified according to different experiments we want to perform. We plan to make several ergonomic studies on different configurations with users manipulating a real application instead of a wizard of oz.

The questions we want to address are the following:

- Is the separation of commands from 3D interactions a good one (planar gestures for commands and 3D interactions for geometric features displacements) ?
- Would it be better to use another modality than gestures for the commands (for example vocal command) ?
- Is it better to perform all manipulations with one hand or is it better to assign the commands to the non-dominant hand and the 3D interactions to the dominant hand ?
- Would it be better to deform a 3D mesh using the dominant hand while navigating in the scene (position, orientation, scale) using the non-dominant hand ?
- Are 3 fingers enough or too much for the 3D interactions ?
- Do we need more or less geometrical operations ?
- Would it be more useful to have higher level operation enabling the user to ignore the meshes but to concentrate on the surface ? This is actually what we are doing when playing with plasticine. Higher level operations concern the merging or the splitting of surfaces.
- Do we need individual displacements of geometric primitives or is it more useful to influence also the neighbors which corresponds to add elastic forces between vertices ?

The problem we address concerns the way of navigating and deforming a surface and the selection of the current mode of interaction, knowing that these actions could be done one by one or at the same time. The level of details of deformations (vertex level or macroscopic level) is also an interesting question.

All these questions need several studies that we plan to do in the following years. Then, the final system will not necessarily use exactly the same technology (cameras, LEDs, ...) and may use more robust (and expensive) components.

Acknowledgements

Authors would like to thank all people how have encouraged this work at Polytech'Nice-Sophia and at the CSTB Sophia Antipolis.

References

[1] F. Bérard. The Magic Table: Computer-Vision Based Augmentation of a Whiteboard for Creative Meetings. In *IEEE PROCAM*, 2003.

[2] D. A. Bowman, E. Kruijff, J. J. LaVIOLA, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2004.

[3] X. Cao and R. Balakrishnan. VisionWand: Interaction Techniques for Large Displays Using a Passive Wand Tracked in 3D. In ACM UIST, pp 173–182, Vancouver, Canada, 2003. ACM Press.

[4] A. Fabri, G.-J. Giezeman, L. Kettner, S. Schirra, and S. Schönherr. On the Design of CGAL a Computational Geometry Algorithms Library. *Software - Practice and Experience*, 11(30):1167–1202, 2000.

[5] O. Faugeras. *3D Computer Vision*. MIT Press, 1993.

[6] A. Jaimes and N. Sebe. Multimodal Human Computer Interaction: A Survey. In IEEE W. on HCI (in conj. ICCV), China, Oct. 2005.

[7] D. Lingrand, P. Renevier, A.-M. Pinna-Déry, *etal* Gestaction3D: a platform for studying displacements and deformation of 3D objects using hands. In CADUI, pp 105 – 114, Romania, 2006.

[8] T. B. Moeslund, M. Störring, and E. Granum. A Natural Interface to a Virtual Environment through Computer Vision-estimated Pointing Gestures. In *W. on Gest. and Sign Lang. based HCI*, pp 59–63, UK, 2001.

[9] J. Segen and S. Kumar. Gesture VR: Vision-Based 3D Hand Interface for Spatial Interaction. In *ACM Int.Conf. on Multimedia* , pp 455–464, Bristol, UK, Sept. 1998.

[10] T. Starner, B. Leibe, D. Minnen, T. Westyn, A. Hurst, and J. Weeks. The perceptive workbench: Computer-vision-based gesture tracking, object tracking, and 3D reconstruction for augmented desks. *Machine Graphics and Vision (MGV)*, 14:59–71, 2003.

[11] R. M. Taylor *etal* VRPN: A Device-Independent, Network-Transparent VR Peripheral System. In ACM UIST, Canada, 2001.

[12] G. Ye, J. J. Corso, and G. D. Hager. Gesture Recognition Using 3D Appearance and Motion Features. In IEEE W. on Real-Time Vision for HCI (in conj. with CVPR), Washington DC, USA, June 2004.