Gestaction3D: a platform for studying displacements and deformations of 3D objects using hands.

Diane Lingrand[1,3], Philippe Renevier[3] , Anne-Marie Pinna-Déry[1,3], Xavier Cremaschi[1], Stevens Lion[1], Jean-Guilhem Rouel[1], David Jeanne[2], Philippe Cuisinaud[1] and Julien Soula[1*]

[1] Polytech' Nice - Sophia Antipolis, CS dept
[2] Univ. Nice - Sophia Antipolis, Ergonomy dept
[3] Rainbow Team - I3S Univ. Nice-Sophia Antipolis - CNRS
B.P. 145 - F 06903 Sophia Antipolis Cedex - FRANCE
Diane.Lingrand@unice.fr
*: now at CSTB Sophia Antipolis

Abstract. We present a low-cost hand-based device coupled with a 3D motion recovery engine and 3D visualization. This platform aims at studying ergonomic 3D interactions in order to manipulate and deform 3D models by interacting with hands on 3D meshes.
Deformations are done using different modes of interaction that we will detail in the paper. Finger extremities are attached to vertices, edges or facets. Switching from one mode to another or changing the point of view is done using gestures. The determination of the more adequate gestures is part of the work.

1 Motivations

In the early 80's, mice and graphic screens led to a revolution in computer interfaces and quickly became the by far most common 2D devices. Today, a growing number of 3D displays appears on the market: active or passive 3D glasses, 3D Head Mounted Displays and now 3D LCD monitors. However, 3D devices such as pointers, 3D mice and gloves are not widely used, due to the high cost, a lack of applications and most importantly the low Ergonomy making their usage difficult.

Our motivations come from two main applications. The first one is about Computer Graphics and 3D builders such as the well known Blender, 3DS Max or Maya. Building 3D worlds and 3D objects is commonly done using 2D devices (such as mouse and keyboard) and visualized on a 2D monitor. Designers of 3D worlds and 3D objects use 3 orthogonal planes in order to imagine the 3D world. Some 3D operations are made difficult and counter-intuitive by the limitations induced. Our 3D device enables intuitive manipulations in 3D space by capturing the user's 3D motion. To provide visual feedback to the user in this context, 3D vision is also mandatory.

The second application comes from the medical field where the images are often in 3D. Segmenting anatomical structures are done using different segmentation methods that may be automatic in some particular cases but often need user intervention for algorithm initialization, during convergence for local minimum avoidance or at the end, for refinement. This is the case when the physician needs very precise result, for example in brain anatomy labeling for spectroscopy. The human interaction with the model is actually tedious and physiologists are not as familiar with 3D projections as computer graphists: there is a need for an intuitive and 3D interface.

All the children are able to simply create and modify 3D models using their fingers on plasticine. Our idea is based on using the hands, and, specially the fingers extremities in order to deform 3D meshes. However, how many hands and fingers to use and how to use them needs to be more precisely studied. Thus, Gestaction3D aims at building a platform for 3D interactions testing. The system is based on Computer Vision reconstruction and built with very cheap components in order to easily modify it and produce several prototypes at low cost.

2 Related work

The dexterity of human hands allows us to execute precise tasks at high speed. If necessary, the hands can easily be trained. Hand motion alone or with instruments have been employed with the aid of Computer Vision in HCI since several years [13, 11, 1, 3, 2, 17, 10, 5, 9].

Some systems use detection in 2D such as in the Magic Table [1] where colored tokens are detected on a table in order to interact both with real and virtual planar objects. Some others use 3D detection such as the GestureVR [13] or the VisionWand [3]. GestureVR [13] allows spatial interactions using 2 cameras with the recognition of 3 hand gestures. Two fingers and their orientations are detected. It permits also 3D drawing, robot arm manipulation and 3D scene navigation. The VisionWand [3] is a passive wand with colored extremities that is used for interactions with 2D objects and menus selection. The system allows the recognition of 9 different gestures.

The motivations of the different works are manipulation of objects [1, 3, 14], pointing to objects or menus [11], gesture recognition for gesture command [15] or American Sign Language detection[6] .

Moeslund and colleagues [11] have developed a system for pointing task using computer vision and a magnetic tracker mounted on the stereo glasses.

Smith and colleagues [14] have explicited constraints that allow the manipulation of 3D objects using 2D devices. They apply their system for objects manipulation in a room (chairs, tables, ...).

Computer Vision based approaches using free hands for gesture recognition are still in a stage of research [10, 5] even if some results are promising [17]. Limiting the known motions to a small set of gestures is mandatory both for the user (more gestures imply more learning) and the recognition system.

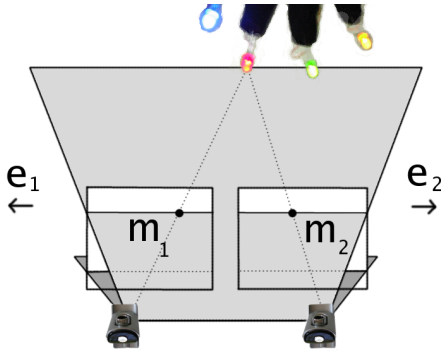

Fig. 1. Standard geometry of cameras.

In this work, we need to study how to interact with 3D objects in order to deform them according to different modes of interaction we will detail later in the paper. We also need to move our point of observation during the interaction. We want to be able to use our system both in an immersive room (dark) and in an usual office or in classroom. We want to use a passive system and decided to use colored and comfortable gloves. We focus also to increase the space of interaction that is really too much limited in systems such as the P5 gloves from *Essential Reality*.

3 Platform description

Gloves are hand made from thin silk black gloves with 5 different colors located at each finger, switches and batteries. The stereo acquisition is done using two webcams Philips Fun II aligned in order to use the simplification of epipolar geometry in the case of standard geometry (see figure 1). Thus, the depth is easily recovered: depth of fingers is inversely proportional to the 2D disparity between the 2 projections [8].

It is well known in Computer Vision that the extraction of points of interest and matching between two views is a difficult problem and that mistakes at this step may lead to the failure of the entire 3D reconstruction process. In our case, colored LEDs are easily segmented and 2D localized using the HSV color system in order to allow different lighting conditions. The colors avoid the difficult matching points step [18]. Calibration is done in order to exploit the whole field of view with respect to the amplitude of displacements in the scene. Then, depth is directly recovered from 2D disparity.

The platform is composed by a server and a client. The server is responsible of the acquisition of the 3D positions of the fingers and of the gesture recognition (see figures 2 and 3). We will detail this part in the next paragraph (4). The client is responsible for the rendering process and for applying the displacement and

Fig. 2. User 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.

deformations to the scene. Communication between the acquisition computer and the rendering machine is done with the VRPN library [16] in order to make its integration with Virtual Reality Systems possible.

Among the different models of deformable objects [12], we chose the CGAL [7] implementation which manages useful operations such as mesh refinement. The 3D scene is then managed using OpenSG and displayed using stereo rendering to enable visual feedback to the user. On a desktop computer, the system uses VRex stereo glasses on a CRT monitor to enable visual feedback to the user. In an immersive room, the local stereo rendering is used.

4 Objects modeling and deforming

As a feedback to the user, detected positions of fingers extremities are rendered using small colored spheres (see figure 3 bottom right). At the very beginning of our study, we observed that it is mandatory to have a 3D display in order to know if fingers are in front of or back to the objects we want to deform. We also observed that smoothing the mesh enables an interaction closer to the plasticine.

In order to interact with a 3D mesh, several fingers extremities are followed. A vertex is attached to a finger when the finger is detected in the 3D neighborhood of the vertex. It is detached when the finger moves rapidly away from the vertex. When a vertex is attached to a finger, it moves according to 3D displacements of the finger. Depending on user's ability to move separately different fingers, different fingers can simultaneously be attached to different vertices, permitting to deform a 3D object as it were in deformable plasticine.

We developed different modes of interaction:

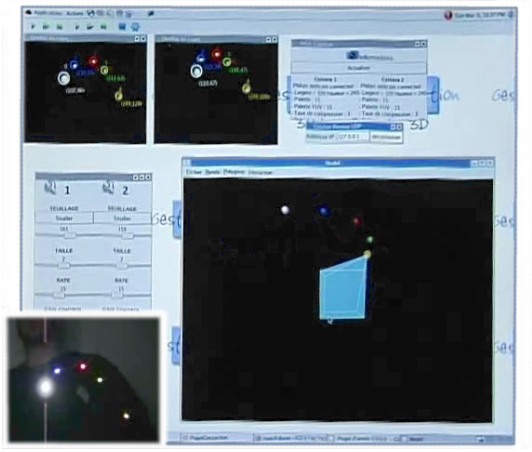

Fig. 3. Focus on the application. On top left, the two acquired images with the detection of fingers extremities. On bottom left, a view of the user. On center right, the fingers in 3D (balls centered on the 3D reconstructed points). The yellow finger is attached to a vertex and moving it backward.

vertex mode: the vertex is moving according to the corresponding finger motion. Several vertices can move simultaneously.

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.

edge deletion: edges that are selected are deleted.

extrusion: an edge is selected by its two vertices and it is extruded using vertices displacements.

facet division: when 2 vertices are selected, the corresponding facet is divided into two parts, adding a new edge.

The different modes can simply be activated using the keyboard. However, it is probably not satisfying to switch from natural displacements to keyboard, even if a second user is assigned to the keyboard and can react to vocal command. We wanted to evaluate this point using different gestures to switch from a mode to another. Actually, these are simple gestures, mostly planar and should be improved later, for example using more sophisticated gesture recognition such as [17, 4] (learning of the gesture structure).

5 Gestures interviews

A first user-centered design was performed in order to establish gestures for the scene management according to the "3D interactions in a virtual environment" taxonomy of Bowman [2]:

camera displacements: translation and rotation relative to the camera center (user point of view);
single object displacements: translation and rotation relative to the object gravity center;
object selection: single selection or multiple selection;
gesture on/off: enabling or disabling the gesture recognition in order to ensure that the user really intends to issue a gestural command.

The final users are students in computer science. The design method is classic in the CHI field. First, we asked to the users, thanks to a web form, some information especially about: (i) their customs, (ii) how they could accept a new input device for 3D interactions and (iii) whether they thought using one or two hands during 3D interactions, etc. The form was also opened to non-student people, their answers consolidating students' ones. Once answers analyzed, we interviewed ten potential users (3 women and 7 men) in order to find the most intuitive gestures for the scene management.

Our final users are essentially (80%) male and young. They are familiar with 3D virtual environments (3D games, 3D conception). They are also aware of the difficulties of interacting in 3D environment with the 2D mouse and the keyboard (this requires to know many keyboard shortcuts to be efficient).

Interviews were made in front of slides video-projected on screen. Slides sequentially illustrated the initial state of the 3D scene and the final state of the scene after each action (camera displacements, single object displacement and object selection). At the end of the interview, we discussed the gestures with the user. Each interview was filmed.

After the interview analysis, we met four interviewed students (other could not come) in a participative design session.

The result of this process is a first collection of gesture (shown on the movie). In many situations, users prefer a two-handed interaction. For some actions, like selecting (figure 4) or taking an object, there is a consensus on "intuitive" gesture. But, for the other actions, like orienting the camera (figure 5), there are many gesture propositions. Consequently we decide to follow the consensus, when existed, and otherwise to impose one or two gestures for a command. One recurrent remark is about the non-dominant hand and a third modality. In front of a small screen, users prefer to use the keyboard rather than the second hand, but in front of a large screen (as it was the case during interviews), functionalities are expected to be attached to the second hand. A suggestion is also to use a carpet (or any device on the floor) for moving in the virtual scene, and also to try vocal recognition to limit the number of gestures.

6 Conclusion and perspectives

The platform described in this paper allows users to build and deform 3D objects using a low-cost passive glove. A first study has been done using users interviews in order to determine gestures associated to several commands: camera and objects displacements and selection. We now need to evaluate these gestures on the

Fig. 4. Selection gesture: users agree for the same gesture !

Fig. 5. Gesture for rotation: different users mean different gestures !

prototype and also to evaluate the most appropriate commands for interaction modes switching.

This platform enables us to make further studies on gestures recognition with more elaborate models such as those described in [10, 5]. We are now ready for further studies improving the interactions with 3D worlds.

7 Acknowledgments

Authors would like to thank all people how have encouraged this work at the engineering school Polytech'Nice-Sophia and at the CSTB Sophia Antipolis where this platform has been tested.

References

1. F. Bérard. The Magic Table: Computer-Vision Based Augmentation of a Whiteboard for Creative Meetings. In *IEEE Workshop on Projector-Camera Systems, in conj. with ICCV (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 Symposium on User Interface Software and Technology (UIST)*, pages 173–182, Vancouver (Canada), 2003.
4. J. J. Corso, G. Ye, and G. D. Hager. Analysis of Multi-Modal Gestures with a Coherent Probabilistic Graphical Model. *Virtual Reality (VR)*, 9(1):93, Dec. 2005.
5. K. G. Derpanis. A Review of Vision-Based Hand Gestures. internal report, Centre for Vision Research, York University (Canada), 2004.
6. K. G. Derpanis, R. P. Wildes, and J. K. Tsotsos. Hand Gesture Recognition within a Linguistics-Based Framework. In T. Pajdla and J. Matas, editors, *European Conference on Computer Vision*, volume LNCS 3021, pages 282–296, Prague (Czech Republic), May 2004. Springer.
7. 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.
8. O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
9. A. Jaimes and N. Sebe. Multimodal Human Computer Interaction: A Survey. In *IEEE International Workshop on Human Computer Interaction in conjunction with ICCV*, Beijing (China), Oct. 2005.
10. T. B. Moeslund and L. Nørgaard. A Brief Overview of Hand Gestures used in Wearable Human Computer Interfaces. Technical Report CVMT 03-02, Laboratory of Computer Vision and Media Technology, Aalborg (Denmark), 2003.
11. T. B. Moeslund, M. Störring, and E. Granum. A Natural Interface to a Virtual Environment through Computer Vision-estimated Pointing Gestures. In *Int. Workshop on Gesture and Sign Language based Human-Computer Interaction (GW 2001)*, pages 59–63, London (UK), Apr. 2001.
12. J. Montagnat and H. Delingette. A review of deformable surfaces: topology, geometry and deformation. *Image and Vision Comput.*, 19(14):1023–1040, Dec. 2001.
13. J. Segen and S. Kumar. Gesture VR: Vision-Based 3D Hand Interface for Spatial Interaction. In *International Conference on Multimedia (ACM Multimedia)*, pages 455–464, Bristol (UK), Sept. 1998. ACM Press.
14. G. Smith, T. Salzman, and W. Stürzlinger. 3D Scene Manipulation with 2D Devices and Constraints. In *Graphics Interface*, pages 135–142, Ottawa (Ontario, Canada), June 2001.
15. 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.
16. R. M. Taylor, T. Hudson, A. Seeger, H. Weber, J. Juliano, and A. Helser. VRPN: A Device-Independent, Network-Transparent VR Peripheral System. In *ACM Symposium on Virtual Reality Software & Technology (VRST)*, Banff (Canada), Nov. 2001. ACM, SIGGRAPH, and SIG-CHI, ACM Press.
17. G. Ye, J. J. Corso, and G. D. Hager. Gesture Recognition Using 3D Appearance and Motion Features. In *IEEE Workshop on Real-Time Vision for Human-Computer Interaction (in conj. with CVPR)*, Washington, DC (USA), June 2004.
18. Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78(1-2):87–119, 1994. Appeared in October 1995, also INRIA Research Report No.2273, May 1994.