METAPHORS OF 3D INTERACTION IN IMMERSED MEDIUMS BASED ON GESTURES RECOGNITION (THE "JEDI[1] METAPHOR").

Andrei Manitiu, Jean Pierre Jessel, René Caubet
manitiu@irit.fr
IRIT - Equipe Synthèse d'Image
Université "Paul Sabatier" (Toulouse III)
118, rte. de Narbonne, 31000 Toulouse, France
tel. : (0)5 61 55 63 13
fax : (0)5 61 55 62 58

Abstract

The interaction in virtual worlds implies the use of specific input/output devices (using exteroception and proprioception). In the case of total immersion, the user will not see the input devices any more. Therefore he must "feel them". The dataglove is the most used input device to handle objects and to interact with widgets. We will propose an interaction method based on gesture recognition applied to 3D navigation.

Keywords

3D Interaction, 3D vision, gesture recognition, immersed workspaces, navigation, simulation, glove, manipulation.

INTRODUCTION

The interaction in virtual worlds is a vast research field, concerning stereoscopic vision, object handling, navigation in virtual worlds etc. These directions have been exploited for a long time and many solutions have been offered. A global view over previous work shows little cohesion of the solutions. Generally, each paper focuses on only one or two aspects of the interaction and ignores the others. Then, we can notice a strong orientation of each paper towards specific problems involved in the 3D interaction. We will make a short review of the main methods of classification used so far. A first classification clearly distinguishes between the interaction in the total immersion worlds (use of the HMD) and in the non-immersed workspaces (stereoscopic glasses). This difference appears primarily on the nature of the physical devices used for the interaction. In the case of the non-immersed workspaces, a large number of input devices were developed (Spacemouse, Spaceball, Polhemus, 3D mouse and joysticks etc). The vision device is generally a Head Mounted Display, for the immersed vision. For input devices we usually find Polhemus tracking devices dissimulated under different aspects, gloves and video cameras for gesture recognition.

This paper will consider only the case of the immersed media and will particularly study the metaphors of navigation used and their possibilities of being generalized for satisfying other aspects of the interaction. We will examine some already proposed solutions and then we will present our model which uses gesture recognition to carry out navigation in total immersion as well as many other tasks.

Research overview

A problem that preoccupied a lot the researchers is the navigation in virtual universes [Colin90]. The technical and visual solution that everyone seems to agree on, is the metaphor of the "flight": in virtual reality "no one walks, everyone flies" [Hayes93]. Its simplicity of implementation and use fills in the lack

[1] Jedi - Star Wars characters able to manipulate distant objects using their hands (and the force)

of realism that characterizes this metaphor. An extension of this metaphor automates some parameters of movement [Macki90]. For example, during a displacement towards a specific destination, the movement speed can be automatically computed according to the distance to the target.

The solutions for navigation control in immersed media are generally split in the use of the HMD equipped with orientation / position trackers and the use of the glove to handle the viewpoint.

The use of the HMD has been proposed more than 30 years ago [Suthe68]. Using it for navigation control is more recent and raises some particular questions. The first one refers to the fact that in everyday life we use the head movement only for significant changes of the viewpoint and not for navigation. When in front of their desktop, people use only their eyes to focus on various objects, whereas the HMD with retina motion control are still at experimental stage. Moreover, for the forward / backward movement of the viewpoint, this solution is not well suited, thus the solutions proposed often use an auxiliary device. Turner [Turne91] uses physical models to control the movement of a virtual camera. Several other systems such as the "Zashiki-Warashi system" [Yoshi94] or the "ALIAS 3D modeling package" [Alias93], use a second device to control navigation. Furthermore, it has been proved that nausea effects may appear if the computing power of a system using a HMD-like device does not allow real time navigation (in other words if there is a delay between the movement of the head and the response of the viewpoint). Another undesirable effect also occurs[Oman91] when displacement in the universe does not correspond to the speeds/accelerations of the head.

A solution which mitigates these disadvantages was brought up by the model of miniature worlds [Stoak95] consisting in adding its miniature copy to a virtual world, the whole in an immersed medium. This metaphor introduces a second dynamic viewpoint on the universe. The process is a 3D extension of the 2D dynamic map model which usually can be seen in games. The user is equipped with a HMD and a tracker integrated in a ball, which represents the miniature world that can be manipulated by the hand. The main viewpoint will not be affected by ball movements. This one is especially used to offer an overall picture of the universe. The disadvantage of this metaphor is that the user has his hand occupied with a specific tool for observation and that he must change this tool to handle the objects in the universe, while remaining completely immersed.

Generally, research went on the use of a second device only when a very precise navigation is needed. Hinckley [Hinck94] presents a model of data visualization from a cerebral scanner, by using two objects : a miniature head and a cutting plan.

Gesture recognition uses mainly two methods of perceiving the gesture: the dataglove which directly returns the angles between various knuckles [Kessl95], and the video camera, which acquires hand movement [Mulde98]. If the use of the glove implies wearing a glove which can encumber the movements, the returned data is more precise than the analysis arisen from image recognition in the case of a camera. Thus, the glove considerably reduces the computing time. This is very important in an application whose goal is not gesture recognition itself, but uses gestures in a more broad context that often requires some complex computations. Moreover, one camera often asks for the use of a special contrasted glove and requires particular lighting conditions [Mulde98], leading to a reduced space of movement.

The use of gesture recognition for navigation has been already considered. "Visual Computing Laboratory" [Katke98] uses a camera to interpret gestures in order to remotely specify the basic movements of a robot. As for T. Baudel and Al. [Baudel95] , they use a glove in order to control navigation in a graphic presentation. The main interest of this application lies in the possibility of controlling one or more unspecified interfaces (tape recorder, video tape recorder, computer) "in the dark", as well as in immersed workspaces. Nevertheless, we can see a certain tiredness of the hand which operates the glove after a longer period of use.

We find the same advantages in the use of the glove for the interaction in immersed worlds. Compared to the physical objects surrounding it, the user is "in the dark", it cannot know the exact position of the peripherals which surround him.

PROPOSAL FOR A NAVIGATION MODEL USING GESTURE RECOGNITION: THE "JEDI METAPHOR"

It is important to recall that our main goal is to find an interaction metaphor coherently adapted to immersed virtual worlds, for which the only way of visualization is the HMD. In this context, the "coherent" term indicates the use of a single input device for all the tasks of interaction implied by such an approach (navigation, distant selection and manipulation etc.).

Why the glove - problem presentation

Many papers tackled the subject of the use of the glove for various separate tasks. Jaques and Al. [Jaque94] use the gestures to simulate a manufacture whereas Rubio and Al. [Rubio95] use the glove to produce a platform of virtual sculpture. In order that the navigation solution that we propose to be usable, it is necessary that the mechanism used may adapt itself to perform several tasks in the same time. We have seen that there are different ways to handle navigation, such as the HMD integrated position tracker and we have also talked about the problems that its use leads to. This is why we conceived a navigation way as intuitive as possible, while respecting the standards of coherence. In this case, it seems natural to choose the same device to handle objects and to control navigation in the virtual universe.

The second question that arises consists in finding a way to give the user the possibility to indicate to the system the single action he wish to complete at a certain moment (e.g. navigation). The selection method used by most of the applications are the toolbars and the menus. The experience shows that it is very difficult to interact with 3D widgets in absence of some force feedback devices, in spite of a good stereoscopic vision. Let us note that this is only a technical limitation, the evolution of the peripherals making now affordable force feedback joysticks, wheels and mouse. While not being our main goal, we try to find a metaphor which can be easily extended to the use of the gloves with force feedback.

The use of menus being thus too heavy, we have adopted the gesture recognition to make a choice in a pallet of actions which we do not restrict. The gesture is the oldest mean of communication that the humans know - it uses the sense of proprioception, which was poorly adapted to the data-processing use. The proprioception means the a priori knowledge by the human of the positions of various parts of his body. This is why the gestures do not make use of any other of our senses and can consequently be used without any difficulty in a completely immersed universe.

Notion and classification of gestures

In our approach, the gesture recognition is a way, not a purpose. This is why our recognition engine is relatively simple, based on proximity techniques. The hand model that we adopted comprises 17 articulations, located as in the `figure 1` : 15 for the articulations of the phalanges and two for the handle. A gesture is detected if a subset of the angles matches, with little variation, a matrix of given angles (static angles). We will call these predetermined gesture. Not all the angles are necessarily taken into account, thus the user can handle some degrees of freedom, by modifying the angles not used in the gesture (dynamic angles).

We considered two types of predetermined gestures: simple gestures and composite gestures (gestures made up of several gestures). Generally, a simple gesture is characterized by a sum of angles between the various articulations of the hand. The simple gesture provokes an event as soon as this gesture is detected. A composite gesture consists of some (in general of two) simple gestures. In this case, an event is only triggered if the sequence of simple gestures composing the gesture is detected.

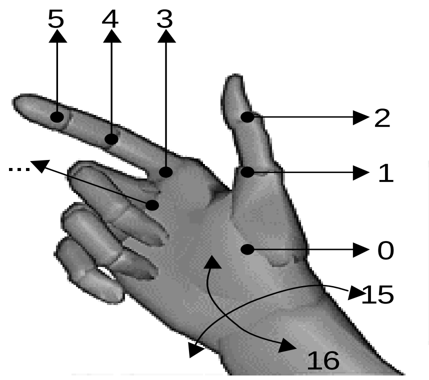
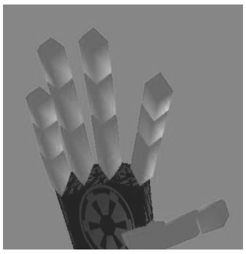


Figure 1. glove joints

The second classification that we used classifies the gestures in static and dynamic gestures. A static gesture is one for which all the angles are static or a gesture starting an action which does not take account of dynamic angles. The dynamic one, once detected, associates the matrix of the dynamic angles to the event, by making possible further manipulations (for controlling a process) while the event is still active.



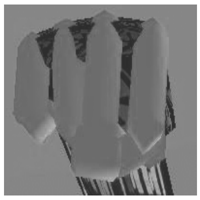


Figure 2.: example of complex gesture formed of two simple gestures.

In our example (`figure 2`), a complete fist opening followed by a fist closing constitutes our composed gesture which starts an event. All the intermediate positions of the fingers are not taken into consideration. Nevertheless, the continuation of the two gestures must take place within a delay indicated in the gesture definition matrix. In this example, each of the simple gestures uses the articulations 0 - 14. In other terms, the position of the handle is ignored.

The delay indicates the maximum time (in 1/10 sec.) allowed for the succession of the two simple

0	Index of the first joint that will determine the gesture (I_{first}) - first static angle
14	Index of the last joint that will determine the gesture (I_{last}) - last static angle
0.019610	Angle of the first joint of the first simple gesture
-0.022740	Angle of the second joint of the first simple gesture
….	
0.004782	Angle of the last joint of the first simple gesture
0.4	Tolerence (T_o)
-1.294260	Angle of the first joint of the second simple gesture
-0.443430	Angle of the second joint of the second simple gesture
…	…
-1.519800	Angle of the last joint of the second simple gesture
25	Delay

Matrix definition for the dynamic gesture

gestures. Once the first simple gesture is detected, after this delay, the gesture will be lost, unless the second simple gesture is detected, in which case an event is generated. The tolerance (T_o) indicates the precision of detection of the simple gesture. It represents the maximum variation between the angles of the gesture definition matrix (A_{def}) and the angles of the finger (A_{finger}), from which a gesture is recognized. This relation can be written in two ways, according to the interpretation we want to give. Thus, a gesture is recognized if

1. for each I between I_{first} and I_{last}
   abs(A_{def} - A_{finger}) < = To or .

2 $$\sum_{i=I_{first}}^{I_{end}} \mathrm{abs}\left(A^{i}{}_{def} - A^{i}{}_{finger}\right) \le T_o$$

The choice of the formula is related to the way we want to interpret the gesture. The first formula takes into account each angle individually, by introducing a local precision inverse proportional to the tolerance. The second one describes the gesture like a unit, the precision being considerably reduced. If we replace the sum by the product, we keep the overall concept, but with a much better precision. The tolerance indicated in the matrix must be relative to the used formula.

After some use and handling tests, we found that the first approach is more suited for our goals, because we use the gesture especially as a way of decision. Nevertheless, by varying the tolerance, we can obtain more or less precise gestures, according to the needs. For example, for a static simple gesture we use a rather small tolerance, the gesture itself must be very precise, because it triggers an event as soon as detected. For a composed gesture we will choose either a great tolerance and a small delay, or a small tolerance and large delay. The gesture we took as example starts a critical action: the application shutting down. Thus, we used a relatively small tolerance for a rather average delay.

On the other hand, a dynamic gesture generally implies a larger flexibility of handling, because it must be detected during the entire action - a more significant tolerance will thus be chosen. We used this type of gesture to control navigation in the virtual world. As the gesture recognition usually employs very elaborate mathematical ways (neuronal networks, statistical functions etc), the absence of a genuine mathematical apparatus in our model can be seen like a weakness. This is not the case, because, as we mentioned, our goal is not to produce a gesture recognition engine, but to use it to control an application. So far, the data processing for a virtual world requires a great computational power, thus a reduction of the number of operations needed for auxiliary handling is welcomed. This is why we strongly simplified the process of recognition, to obtain results which makes this process very useful for interactions in complex virtual worlds.

Navigation

In order to validate our work, we applied this metaphor to a precise task, witch is the navigation in a completely immersed workspace. As navigation is a basic task in the process of interaction, it must be carried out without effort but with precision and flexibility. At the same time, switching-over to navigation mode must be done instantaneously and the system should not leave this mode without explicit intervention of the user.

Two methods were likely to satisfy these needs. The first consists in creating a gesture to enter in navigation mode and a second to leave it. Between the two gestures, the user can use the glove to control the navigation. The second method implies the use of one gesture only, navigation being done as long as this gesture is detected.

A more accurate analysis of the first model reveals the "critical points" or the "points of conflict". A critical point is an action which is likely to be interpreted in two different ways by the system. Let us take the example of ending the navigation mode: when the user is about to make the gesture of output, the movements which lead the hand into the position to make the gesture will be interpreted like a navigation, even if they were used to make a gesture. This is why this method, certainly easier to understand and to use, request a deeper analysis of the gesture which uses other more advanced techniques of recognition (prediction, analyzes of the entire movement, etc).

After several tests, we adopted the second model, for reasons of ease of use and for its resemblance to a handling way we are accustomed in everyday life. This model corresponds to what we called "the metaphor of the tap": when someone wants to open a tap he serves the handle (simple gesture), turns the handle (dynamic gesture) and reopens the handle (end of the gesture). We will use this metaphor for our navigation, by associating it to a dynamic simple gesture. This gesture breaks up into two parts: the static angles which determine the gesture and the free angles which constitute the freedom degrees of handling. We evaluated a dozen gestures (with various matrices of definition of the static and dynamic angles) from the point of view of flexibility, precision and robustness of handling. Among these gestures, we retained two, particularly intuitive. The two gestures we selected use for static angles the angles between articulations 6-14 (see figure 1). The gesture will thus be recognized as soon as the user closes the auricular, the annular and the major (`figure 3`). In this case the free angles will be those of the inch, of indexes and of the

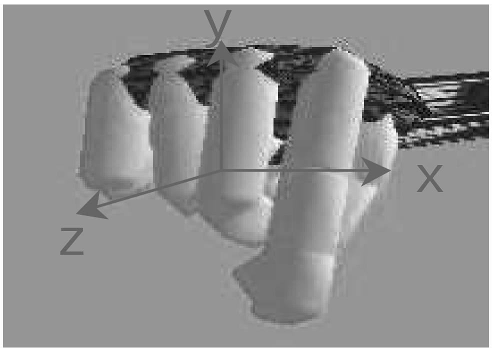


Figure 3. : navigation gesture

handle. Among them, only three will be used to handle the three freedom degrees of the movement (note: even if the index and the inch comprise three degree of freedom each one, we make no distinction between them, the voluntary movement of only one phalange being very difficult to control - more precisely, the degree of freedom considered is the sum of the three angles between the phalanges).

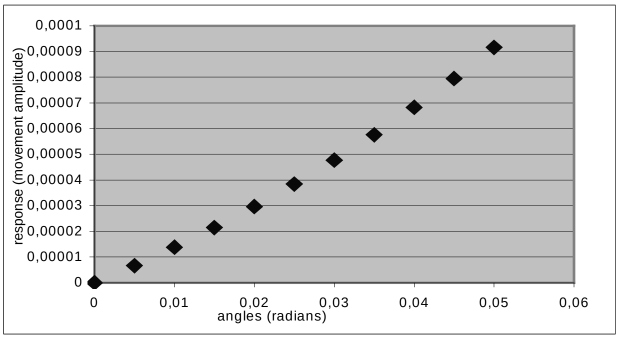
Here are the tables of association for the two gestures:

	First gesture	**Second gesture**
handle (15)		Y
handle (16)	Z	X
inch	Y	
indexes	X	Z

Ergonomic considerations excepted, the difference between the two gestures is purely technical - this is why we will detail only one of them.

We saw that the use of the dynamic gestures is very intuitive. Moreover, the glove offers a very high degree of handling accuracy, because its sensors (optical fiber) provide a very good spatial resolution compared to other type of sensors such as magnetic trackers. This good resolution was a problem during our tests, because we have noticed that the smallest movements of a phalange caused an answer and thus an action, whereas the human carries out permanently some small muscular movements of the finger, not controlled by the brain. For the navigation system to be usable, the output angles should initially be filtered in order to eliminate these involuntary movements. Then, our studies showed that unconsciously, the user does not associate the action with the movement itself, but with the speed/acceleration of this movement. This is why we had to adapt our model to this perception, by computing a response curve of the glove which corresponds to the users' behavior. This curve breaks up into two parts: the first part is supposed to eliminate the involuntary movements whereas the second causes a natural response compared to the movement.

Excepting the multiplying factors, the curves are the same for the three degrees of freedom. For small angles (lower than 0,06 rad), we used an exponential curve which seems to better filter the small movements.



Y=EXP(13*X)/10000-0,0001;

For more significant angles, which imply a voluntary movement, experiments show that a parabolic curve of second degree is more suitable. These curves are an indication of the principle we used for handling navigation with the glove. The two curves are continuous in the critical point of 0.05 rad. which we chose as limit of the involuntary movement. Traced accordingly to experimental response data, these functions seem to describe a natural movement and answer, which makes the task of navigation very intuitive.

The results we obtained can be applied to any handling by the means of dynamic gestures, navigation being an example of it.

Visual feedback

Any interaction supposes a bi-directional dialogue between the user and the system. We saw that the input is represented by gestures. The system must answer the user that it has recognized his gesture and that it is ready to carry out the task requested or to pass in another operating mode.

We applied this model in order to ask the system to change from visualization to navigation mode using the glove. When the gesture of navigation is detected, two indicators appear (see figure 4). The evolution of the two indicators is related to the navigation in the universe : the road advances or moves back at a speed proportional to the rate of advance and the target moves on the grid according to the rotational movement amplitude of the viewpoint. This metaphor provides an additional feedback and helps the user to control his gesture better.

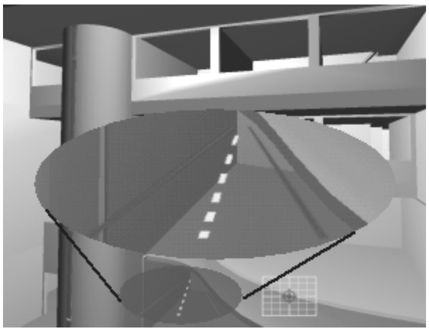


Figure 4 : navigation feedback

When the user is ready to carry out the gesture to leave the application (gesture made up of two simple gestures), the system visually informs him when it detects the beginning of the gesture. If the first simple gesture was carried out accidentally, the user must simply pay attention not to confirm the output by carrying out the second simple gesture. After a time of approximately two seconds, the first gesture is lost and the warning disappears.

During the tests carried out on a dozen of people, we noticed that the control of the gestures is not very easy. Under these circumstances, our research showed that a visual answer clearly improves comprehension of the system by creating a link between the haptic, proprioceptive, and visual user's senses. We have also noticed that for any type of action, it is necessary to find a visual metaphor which reinforces the perception of the associated action and which facilitates the control of this type of interaction.

Conclusion and future works

Performance measurements obtained so far by our system from the point of view of speed, precision and comfort of navigation show the power of the "Jedi metaphor". Moreover, these give us good reasons to believe that such a metaphor can be widen to be coherently used for a multitude of handling tasks. Our future work will focus on two aspects: the extension of the concept of dynamic gesture and composed gesture, as well as the application of the metaphor for other operations.

The gesture recognition engine could be improved to take into account gestures containing movements of the fingers (e.g.: the forward movement of the index could push a button). Moreover, the gestures which we examined exploit only the movements of the hand. By adding a Polhemus sensor, we will be able to take into account the arm, its movement and its orientation. Though very useful for applications such as navigation, it should not be forgotten that the precision of these sensors is significantly reduced - it will thus be necessary to find more elaborate filtering functions to increase this precision. The management of the arm movement could also enrich the composed gestures by adding the "gestures in motion". Once developed, this metaphor will help us to conceive a total immersed interface using gestures. Thus, we will seek metaphors for the 3D menus, pointers, buttons etc. Naturally, the use of a glove with force feedback will reinforce the realism of the interaction and will open new possibilities.

The metaphor of the Jedi that we proposed is a typical result of synergy between several fields of research. It is well known that the process of interaction does not rely only on a single discipline, but concerns at the same time sociologists, data processing specialists, ergonomicists, psychologists etc. Nevertheless, work on the interfaces is carried out often only in one or two fields at the same time. In our research we associated the 3D vision with the gesture recognition and the guiding principles of navigation, which gave us as a result a metaphor which is likely to be widened to satisfy several handling.

REFERENCES

[Alias93] Alias Research, Inc. "*ALIAS Reference Guide*." Toronto, Ontario, Canada., 1994

[Baudel95] T. Baudel and M. B.Lafon "*CHARADE: Remote Control of Objects using Free-Hand Gestures*" L.R.I. - CNRS URA 410

[Colin90] Colin Ware and Steven Osborne "*Exploration and virtual camera control in virtual three dimensional environements*" In Proceedings of 1990 Symposium on Interactive 3D Graphics, pp 175-183

[Hayes93] Nancy Hayes "*Nobody walks in VR - they all fly*" IEEE Computer Graphics and applications, pp 85, May 1993

[Hinck94] Ken Hinckley, Randy Pauusch, John C. Goble and Neal F. Kassel "*Passive real-world interface props for neurosurgical visualisation*", In Proceedings of CHI'94, pp. 452-458. ACM SIGCHI, April 1994.

[Jaque94] Mark W. S. Jaques & David J. Harrison

"*Using gestures to interface with a Virtual Manufacturing Package*" - Interface to real & virtual worlds - InforMatique Montpellier 94 pp.231-239

[Kathe98] Arun Katkere "*Gesture recognition*" research at Visual Computing Laboratory - 1996-1998

[Kessl95] Kessler, G. Drew, Larry F. Hodges and Neff Walker. "*Evaluation of the CyberGlove(TM) as a Whole Hand Input Device*." - HCI - December, 1995.

[Macki90] Jock D. Mackinlay, Stuart K. Card, George G. Robertson "*Rapid controlled movement through a virtual 3D workspace*" In Proceedings of SIGGRAPH'90 pp 171-176

[Mulde98] Axel Mulder "*Human movement tracking technology: resources*" - School of Kinesiology, Simon Fraser University, July 1994 (last additions: 8 May 1998)

[Oman91] C. M. Oman "*Sensory conflict in motion sickness : an observer theory approach*", in Stephen R Ellis, editor, "Pictorial Communication in Real and Virtual Environments", pp. 326 - 376. 1991

[Rubio95] F. Rubio, P. Torguet, R. Caubet, "*Atelier de sculpture virtuelle multi-utilisateurs*", actes de Interaction Homme Machine 95 Toulouse, Octobre 95

[Stoak95]Richard Stoakley, Matthew J. Conway, Randy Pausch : "*Virtual Reality on a WIM: Interactive Worlds in Miniature*", In Proceedings of CHI'95 pp 265-272. ACM SIGCHI, March 1995

[Suthe68] I. E. Sutherlant "*Head Mounted three dimensional display*" In Proceedings of the Fall Joint Computer Conference, 33, pp. 757-764, 1968

[Turne91] Russel Turner, Francis Balaguer, Enrico Gobbetti and Daniel Thalman. "*Physically-based interactive camera motion using 3D input devices*", In Proceedings of Computer Graphics International '91, pp 135-145

[Yoshi94] T. Yoshimura, Y. Nakamura and M. Sugiura "*3D direct manipulation interface : Development of the zashiki-warashi system*", 18(2) pp. 201-207, 1994