A two-handed 3D interaction technique for controlling an ambient environment

Muhammad Tahir, Eric Lecolinet
ENST, 46 rue Barrault, 75634 Paris
cedex 13, France
tahir@enst.fr, elc@enst.fr

ABSTRACT

Controlling a rich multimedia environment (e.g. TV, DVD, VCR etc.) requires users to manipulate multiple remote controls. Users must control these numerous devices that are unaware of each other, independently and in a heterogeneous way. The principles of ambient intelligence may offer an efficient solution to such type of problem in order to facilitate the interaction along with the control of such environments.

We propose a two-handed 3D interaction technique that is based on gestural interaction. This interaction is performed by using ARToolkit markers. This solution has been preferred to bare hand recognition for efficiency reasons.

This technique allows the progressive learning of the gestures by getting feedback from the visual representation that is displayed on a PC screen or a TV set. This idea is inspired from marking and control menus. It allows a hybrid interaction, as it defines less number of gestures for the expert mode and it uses the indications from the visual representation for novice mode. We have considered different types of interactions and designs that have been implemented in a preliminary prototype. We now plan to perform experiments to improve the interaction designs.

1. INTRODUCTION

Ambient intelligence is the study of environments that recognise the presence of users, their movements, and respond to their actions in a natural way. The key characteristics of an ambient environment are intelligence, awareness, interaction and ubiquity. Gesture recognition can for instance be used to interpret the bodily movements of the users. In an intelligent room the appliances can thus be handled by hand motion [2] or by the optical tracking of fingers position [5][6]. However gesture recognition is a difficult problem because of the inaccuracy of recognition algorithms. The high level interpretation of gestures is an even more difficult task. Some authors have proposed gesture recognition toolkits for user interfaces as an attempt to solve this problem [8]. The use of tangible objects, and the knowledge of context awareness constitute another way of solving the interaction problems in ambient intelligent systems [3].

An ambient user may use cameras alongwith some physical objects to interact with the environment. Such an environment lies in the category of Augmented Reality (AR). AR deals with the combination of real world and computer generated virtual world. It is related to the live video imagery e.g. motion tracking, marker recognition etc [13], which is processed by a video device and augmented by the addition of computer generated graphics. A user of an ambient environment needs feedback and control over the operations performed. User interfaces provide useful feedback in the form of results of operations and gives an opportunity to refine his actions. When the user interacts with the interfaces, he selects an option or generates some marks to execute an operation. Marking menu allows the user to select the desired item from a menu of items either in menu mode or in mark mode [11]. During the selection operation the user keeps its attention focused on the menu and gets control of the operation [12]. Similarly, fluid composition of operations is handled in [14] which provides crossing as an interaction technique for handling pen-based applications.

Keeping in mind the study of these articles and their results, we concentrated on these ideas: defining efficient hand gestures for controlling an environment and using some intermediate tool (i.e. tangible objects) as a help to achieve the result of those gestures. As a result we propose a new two-handed 3D interaction technique that let users interact with both hands. Currently, we have developed a prototype that uses ARToolkit to detect different patterns on the markers. These markers are detected by a web camera and are used for the selection and navigation of different operations. Currently, two sides of the right hand marker and one side of the left hand marker are used. This solution has been preferred to bare hand recognition for efficiency reasons because the later involves getting the value of fingers position, movement and color information. This requires a lot of calculation each time when hand moves.

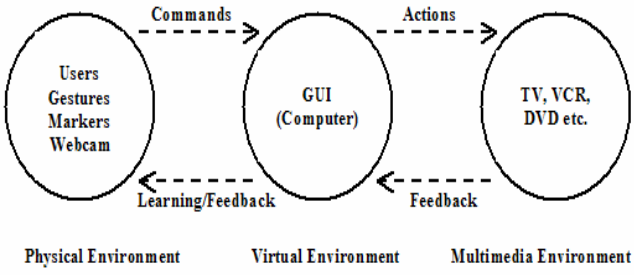

Figure 1. General System Overview.

We have identified several categories of typical operations, which are either discrete or continuous such as boolean operations, selection in a list or a tree, continuous tuning etc. Our goal is to develop a set of 3D interaction styles that will make it possible to control such operations in a simple way. Our aim is to develop a set

of interaction styles that would be the equivalent of 2D widgets or controls in the 3D world.

2. TYPES OF OPERATIONS

We have divided the different operations into two main categories: discrete and continuous operations. Discrete operations are performed by changing the states of operations like changing a value from 0 to 1 (Boolean change). These operations are performed just by moving and presenting the OK pattern to camera in the specified direction. For example turning ON the volume state of the TV to mute i.e. Mute ON.

To perform an operation, user needs to start from somewhere in the GUI. This point is named as center point. The center point is the point from where the operation starts. It is defined as the circle in the middle of the GUI. There is no need to define a center point for discrete operations. Positioning the hand in the proper region and presenting OK marker in front of the webcam selects the operation. It removes the unnecessary movement of hand and thus improves the navigation, selection and manipulation.

Continuous operations are the operations performed by continuously changing its state, for example increasing and decreasing the volume level of DVD Player. This resembles to the situation of changing the value on the slide bar with the mouse. These continuous operations include increasing/decreasing the volume, navigating through the all TV channel list and through the favourite TV channel list.

While changing the value of the continuous operations a center point is required, which is then used as the starting point of the operation. For example to change the volume, we move our right hand marker in the upward direction so that it moves in the proper area of the graphical interface, then turning the square marker towards OK pattern executes the required operation. Similarly moving the hand in the opposite direction, selects the inverse value of that operation.

Table1: Continuous and Discrete operations (TV)

Discrete	Continuous
Device ON	Vol +
Device OFF	Vol -
Next Channel	All Channels list navigation
Previous Channel	Favorite channels list navigation
Mute ON	
Mute OFF	

3. MARKERS ANATOMY & BEHAVIORS

Markers are physical objects that contain patterns on it (figure 2 & 3). For example in figure 2, white square with NL and black square around it forms one pattern. Similarly in figure 3, the marker contains two patterns OK and NR. We use markers having a square shape and a maximum of four useable sides by wearing it over any hand. These markers can be rotated in four directions, and each marker may manipulate four different operations. These markers contain different patterns (images) on them (see figure 2 & figure 3).

There are three types of patterns used: NR & NL patterns, used for the navigation of right and left hands respectively, and an OK pattern for right hand. These patterns are detected by the camera and manipulated by the ARToolkit. Currently, to control the TV operations we are using the two sides of right hand marker and one side of the left hand marker. We use the right and front side of the right square marker to interact with the graphical representation (GUI). The OK pattern is used to finally "select" an operation (i.e. for the confirmation of the operation) while the front side contains the NR pattern that provides the navigation through the GUI.

The GUI represents different TV designs in different ways so that a user can interact with them. For example one GUI represents the TV channels with respect to category of channels and other with numbers.

We use the front portion (NL) of the left square marker for the navigation of interface e.g. moving around the left side of the GUI such as in the list of three hundred channels (one hundred channels per visualization) traversing back the previous hundred channels. This supports the right hand marker for the selection of lengthy channel list of 300. It is more time consuming if we use the right hand to display, navigate and select the channels. We use only the left hand in collaboration with the right hand to navigate and display the next channel list. In this way both hands can be used in parallel. The navigation is performed when the user moves and changes the environment (such as changing the value of slider bar) with the NR or NL markers and selection is performed when that value is finally chosen by the OK marker.

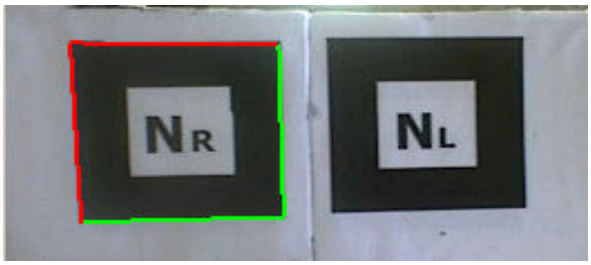

Figure 2. Two square markers (right & left hands).

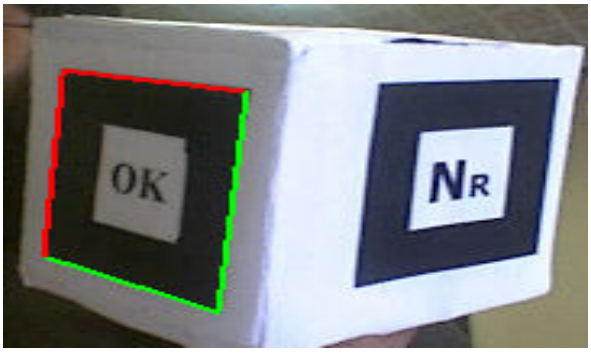

Figure 3. Right hand square marker with both patterns.

4. INTERACTION WITH THE GUI

We have applied our technique to new menu designs for the selection of TV channels in different ways e.g. by entering the number of channel, by choosing the channel from the list and from the different categories These designs have their own visual representations.

The camera connected to the computer detects the both square markers as soon as they appear in front of it and General Control Menu appears. General Control Menu displays different devices, which can then be selected one by one by using the right hand marker movement.

Right hand and left hand markers are defined by the black circle and light grey circle respectively on the GUI. Both circles change their positions when the markers on both hands move.

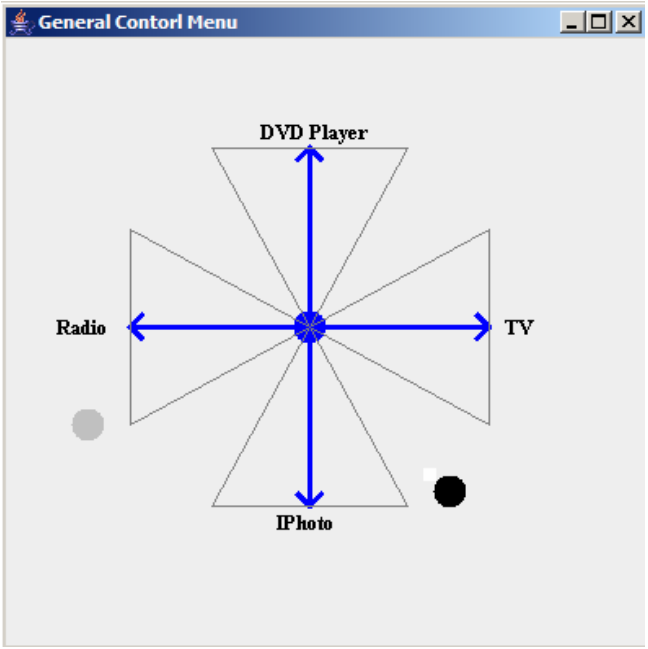

Figure 4. General Control Menu.

The small triangles in grey colour around arrows represent the specified selection region of the operations. When the right hand (black circle) is in any one of the four triangles, turning OK side of the marker selects the required device e.g. turning on TV (in this case, TV option is selected and it shows the specific TV GUI for example as shown in the figures 5)

TV Control Menu design 1 represents the different operations e.g. Channel list, Next Channel, Mute ON/OFF, Volume + /-, and a navigation operations. Navigation operations (next list, prev list) are used to control the TV channel list with both hands. When we move the left hand in the next list square (region), the second channel list on the right side appears ranging channels from the continuation of previous channels. Similarly with the right hand, previous hundred channels can be displayed.

In the second design menu (figure 5), TV channels are listed by their category. These include music, sports, cinema, documentary etc. Left hand marker is used to select the category and right hand to select the required channel of category. Design 4 (figure 6), represents the alphabetical prototype presentation of TV channels. Left hand selects the alphabets category and right hand further selects the channels starting with that alphabet. This design provides a sorted list of channels. For example when the left hand is at A, all channels starting with A are shown at the right side of the figure (AA, AB, AC, AD, AE, AG, AG, AH, AI, AJ). When the right hand side is positioned at AA, it produces the list of all channels starting with AA in above two rows (AAA to AAF and AA1 to AA3). This provides more specific and deeper navigation of TV channels in the hierarchy of the tree.

Design 5 (figure 7) provides a help to user to enter the channel number. Now the center point for the start of operation has changed specially for this design. User starts with the right hand from the start box, selects each number from each list of columns and ends at the end box. As soon as the black circle reaches in the end box the required channel is selected. Another variation of this design is to fix left hand in the start box one time, and then changing the channels with right hand only. When the black circle is in the start box or in the end box, it provides feedback by changing the color of box. When the channel is selected it also displays the channel number on the GUI so that the user is sure of its selection. Although this design provides five hundred channels selection, it may be extended upto 1000 channels.

In the same figure channel number 091 is selected. Here are two important points: firstly, the channel number selected *last* is the chosen one. For example to select 9 in the second column, black circle crossed the 5,6,7 and 8. Similarly it has crossed the 5, 4, 3, and 2 to select 1 from the third column. In the same way to select number 4, the user needs not to take care of its hand position and he may come from 0, 1, 2, 3 or from 9, 8, 7, 6, 5 in the second or third column. This provides free hand movement for operation selection and thus reduces tension and fatigue in the hand. Secondly, the user may modify its selection any time. For example if the user has selected the channel number 145 but not yet finished the operation, may come back and change the channel number to e.g. 142 or 125 etc. This provides flexibility in navigation and selection of an operation.

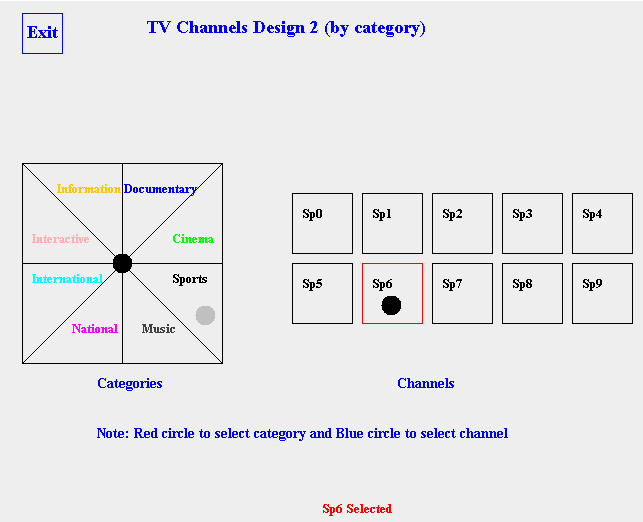

Figure 5. Design 2, TV channels selection (categories).

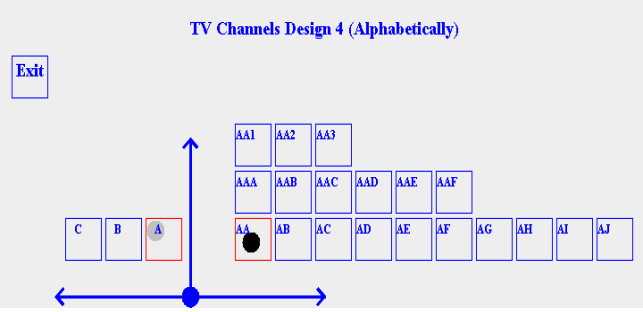

Figure 6. Design 4, TV channels (alphabetically).

Design 5 (figure 7) has different variations in which the columns are represented in curved styles. These curves provide selection of certain numbers of first column in less time. These also provide more space between columns at the upper and lower regions providing the benefit of easy hand movement and selection.

For current implementation, Java 2D is used for the graphical representation of the user interfaces and ARToolkit for the manipulation of markers. The physical equipments used are paper markers, a Philips webcam and a computer (Compaq Presario running Windows XP) with Java and C compilers.

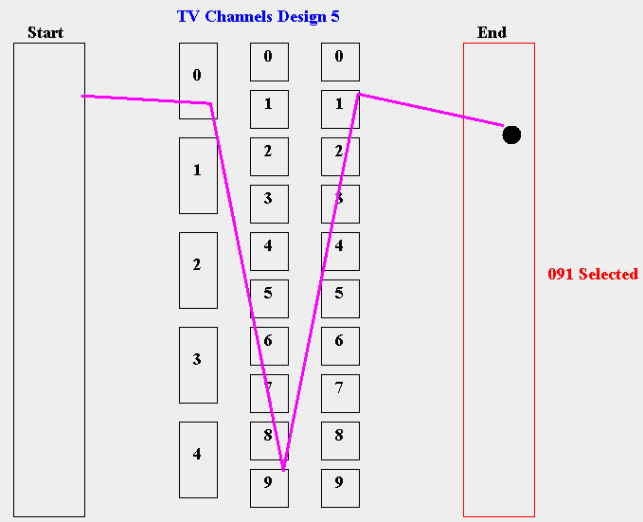

Figure7. Design 5, TV channels (channel numbers entry).

5. DISCUSSIONS

Our technique arises various discussion points:

To select an operation, a user positions the hand in the operation region and then uses the OK marker. Is not it better to use OK marker by putting it in the operation area at once?

Using only the OK marker may create a problem at the initialisation of GUI. At that time OK marker may already exist in the region of an operation or a little movement of a hand may put the marker in the unspecified region, which we don't want to select, so it may select an unwanted operation unexpectedly.

During the test of the designs, it has been observed that the 3D movement of the hand in the vertical (upward, downward) direction produces more fatigue than moving it in horizontal (left, right) and diagonal (45°) direction. Therefore the frequently used commands are placed in the left, right and diagonal lines (regions) and others are used in upward and downward direction. This improves the learning speed of user operations that ultimately reduces the time to manipulate an operation.

Also we have used the Next Channel operation in the upward diagonal line and Mute ON/OFF in the downward diagonal line to maintain consistency for these discrete operations. Mute ON/OFF operation is less frequently used than vol+/- operation. Because vol+/- operation is a continuous operation and it resembles the natural movement of human hand in upward and downward direction (for increasing and decreasing the volume) respectively, so it has been positioned at this place.

Some operations do not require the positioning of the marker at the center point of GUI, so it may be necessary that that operation area is visible all the time (e.g. Mute ON/OFF) because user need to position the hand at the proper area and then performs the operation.

These designs also provide feedback about the imprecise control of the hands. For example in the designs 2, 4 & 5 user need to be careful in the selection of channels because sometimes a little movement of hand may change the selection position of the target operation.

6. CONCLUSION AND FUTURE WORK

We have presented a new interaction technique for an ambient environment specifically for TV control. It provides a way to select a specific TV channel from a long list of TV channels efficiently. The other TV operations can be performed without the need of remote control. It provides an alternative approach to handle multimedia devices. Future work will include usability tests, further implementations and will extend this work for the other household appliances to handle more generic situations. We will also work on the improvement of markers, gestures and the design of interaction styles. Finally, this research work is not restricted to use only ARToolkit, but to use other technologies such as RFID, Phidgets etc.

7. REFERENCES

[1] Alluse Y, Baugette magique et Vision GPU, Internship report, ENST, Paris.

[2] Irie K, Wakamoura N., and Umeda K. Construction of an Intelligent Room Based on Gesture Recognition: Operation of Electric Appliances with Hand Gestures. In Proceedings of IEEE 2004.

[3] Loenen Evert J. Van, On the role of Graspable objects in AmI paradigm, Philips Research Lab, The Netherlands.

[4] Lenman Sören, Bretzner L, Thuresson B, Computer Vision based hand gesture interfaces for Human-Computer interaction, KTH Sweeden.

[5] Duric Z, Wechsler H, Yven J, Adaptive and Smart Interface for VCR Remote Control using Hand Gestures, IEEE 2002.

[6] Marten J.B., Kok A.J.F., Experiencing 3D Interactions in Virtual Reality (VR) and Augmented Reality (AR), EUSAI, The Netherlands.

[7] Acampora G, Loia V, Nappi M, Ricciardi S, Human-Based Models For Smart Devices in Ambient Intelligence, IEEE, ISIE 2005.

[8] Rubine Dean, Specifying gestures by example, Computer Graphics, Volume 25, Number 4, July 1991.

[9] Julia L, Faure C, Pattern recognition and beautification for a pen based interface, IEEE 1995.

[10] Zhao S, Balakrishnan R, Simple vs Compound Mark Hierarchical Marking Menus, UIST 04, Santa Fe, New Mexico, USA.

[11] Kurtenbach, G., The design and evaluation of Marking Menus, PhD Thesis, 1993, University of Toronto.

[12] Pook S, Lecolinet E, Vaysseix G, and Barillot E. ControlMenu: Execution and Control in a Single Interactor. In Extended Abstracts of CHI2000, pages 263–264, 2000.

[13] Kato H., BillingHurst M, Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System, IWAR 1999.

[14] Apitz G., Guimbretière F., CrossY: A Crossing-Based Drawing Application, UIST 04.