Interactive Perceptualization

Ken Hinckley and George Robertson
Microsoft Research
Tel: (206)-703-9065
{kenh, ggr}@microsoft.com

Abstract

Video input (camera-based monitoring of the user and his or her actions) offers an extremely promising medium for a rich vocabulary of user interaction. We see a clear role for video as an input channel to advance the use of highly interactive tools and visualizations. But by itself, video input cannot solve all of the problems of the user interface. Video input needs to be carefully coupled with other means of providing enhanced manipulation as well as with salient feedback channels so that the user can act and perceive in a tightly coupled loop.

Introduction

People do not learn by being pummeled with information, but rather they learn by doing, by actively investigating the relationships and structure of information [3]. As interface designers, our goal should be *interactive perceptualization* [2]: allow users to see, hear, and feel what happens to the information in response to their actions.

There are two dimensions of interactive perceptualization which the interface should seek to maximize:

- *Action*: the computer's ability to recognize and/or track our full range of physical actions and use these as elements of the human-computer dialogue; and

- *Perception*: the extent to which the computer can help users to see, feel, or hear the results of their actions.

On the action dimension, there is a mismatch between the control structures supported by computer interfaces and our underlying human capabilities. In the traditional graphical user interface, all of the human's motor capabilities are reduced to pressing a few buttons on the keyboard with the nonpreferred hand and moving a single point (the cursor) with the preferred hand.

On the perception dimension, the computer presents a world of hidden state that one often cannot see, hear, or feel. This information should be tangible to the user and should be manipulable in response to the user's actions to afford a high degree of action-perception coupling.

Video

Video input offers one approach to improve things along the *action* dimension. There are many degrees-of-freedom of body motion which are potentially useful for user interfaces, but it is often inconvenient to directly measure these motions with an input device.

For example, head motion parallax is one of the most important spatial depth cues, but 3D computer interfaces do not commonly support such cues, and when they do, cumbersome head-tracking gear or a head-mounted display is required. A monitor-mounted camera could provide an elegant head motion input device by tracking the head position. There is a user interface problem, however, involving states of the interaction (whether the user's head is being tracked or not), how to perform transitions between these states, and how to provide the user with effective feedback of such transitions.

By itself, video can only provide part of the desired state information: it can provide a continuous stream of the head position and it can provide and out-of-range signal (when the camera sees no head to track). But as an interaction technique, we would really like head-coupled motion parallax to provide a third state where the user can freeze the image on the screen at a desired viewpoint [5][6]. This might be necessary to see small details or read text in the scene, or to perform precise selection of objects in the scene using a mouse. We are currently exploring several alternatives to augment video-based head tracking with discrete state transitions. Some possible solutions include using a footpedal, selecting a special command with the mouse, or making a hand gesture which enables / disables head tracking.

Input as Output

Another approach to enhancing the possibilities for action lies at the input device level: replacing or augmenting the mouse and keyboard with more expressive device(s) that capture a wider range of human manual skills, such as the ability to use both hands to help perform a compound task, or the ability of our hands to precisely control and sense pressure. As we see it, the research question is not "Which is better: video or input devices?" Rather, the question is how to best integrate these modalities to provide tight action/perception coupling.

Unlike video, physical input devices also serve as passive haptic output devices: the user can feel his or her hand touching, moving, or releasing the input device. We say that the devices are "passive" because the computer can't directly control or alter these aspects of the interaction. But interface designers can and should take advantage of these properties of devices to introduce physical contact and muscular tension into the dialog. These haptic qualities can be used to phrase low-level tasks together which emphasize the higher-level cognitive chunks that the interface designer wishes to establish as a mental model [1].

We believe this approach complements video quite nicely. Video can provide rich information about the position of the user's body, whether or not the user is sitting in front of the computer, or whether or not there is more than one person in the room, to name a few. On the other hand, physical contact with input devices can provide an effective input/feedback coupling to delimit physical gestures: the contact and physical motion can be detected by the system, and the user senses these same properties as haptic feedback. Physical devices can also embody affordances which indicate their intended use [5].

From the Screen to the Environment

Despite rapid progress in computer hardware, many users are still stuck with a 14 or 15 inch "desktop" display which is barely larger than a single sheet of paper. In terms of the user interface, screen real estate is perhaps the most valuable (and enduring) resource to consider. Unfortunately the clutter of title bars, menu bars, button bars, task bars, and scroll bars leaves relatively little of the display area available for the user's actual work. Controls and feedback mechanisms which are tied to visual displays are furthermore demanding of our visual attention; we believe it would be better to remain visually focused on the information which one is authoring or exploring, rather than the ancillary widgets that surround it.

We believe that a key theme for the next generation of user interfaces will be moving controls, affordances, and background monitoring information off of the screen and into the user's physical environment. Our example of video-based head tracking provides an instance of this: there is no GUI control on the screen for adjusting motion parallax; the user's own body *is the control* for this capability.

By adding some physical input devices, the interface can potentially achieve the seemingly contradictory goals of removing visual affordances from the display surface while still giving the user obvious indications of what actions are possible. For example, imagine a physical tablet which has dedicated regions that mimic the iconify, maximize, and close window controls for desktop windows. The actual windows on the screen would no longer need to display these controls, since the controls are always accessible on the physical tablet.

Another simple example would be to use a trackball, operated by the nonpreferred hand, to scroll through documents or web pages. With a dedicated transducer for scrolling, there is no longer a real need for *graphical* scroll bars. This screen real estate could be freed up to display the actual document.

We believe that audio feedback is also a key part of this action/perception design balance. For example, one argument against removing graphical scroll bars from the screen is that they do provide some information: they indicate how long the document is and they indicate the user's position within that document. This information does not have to be delivered visually, however; approximately the same information could be provided through interactive audio feedback (for example, large documents could sound "heavier" than small documents [4]).

Conclusion

We have outlined a vision of the user interface which strives for interactive perceptualization, a carefully balanced and closely tied coupling between the user's actions and the user's perceptions. Perceptual user interfaces offer opportunities to improve the ability of computing systems to respond to user actions, but the success of PUI's depends upon corresponding support and enhancements on the user perception side of the balance. We have identified exploration and integration of video-based input with input device approaches, haptic issues, and interactive audio feedback as avenues for pursuing these research goals.

References

1. Buxton, W., "Chunking and Phrasing and the Design of Human-computer Dialogues," Information Processing '86, Proc. of the IFIP 10th World Computer Congress, ed. by H. J. Kugler. Amsterdam: North Holland Publishers, 1986, pp. 475-480.
2. Buxton, W., "Human skills in interface design," in *Interacting With Virtual Environments*, ed. by L. MacDonald and J. Vince. New York: Wiley, 1994, pp. 1-12.
3. Card, S. K., Robertson, G. G., Mackinlay, J. D., "The Information Visualizer, an Information Workspace," Proc. ACM CHI'91 Conference on Human Factors in Computing Systems, 1991, pp. 181-187.
4. Gaver, W. W., "The SonicFinder: An interface that uses auditory icons," Human-Computer Interaction, 4 (1), 1989, pp. 67-94.
5. Hinckley, K., Pausch, R., Goble, J., Kassell, N., "A Survey of Design Issues in Spatial Input," ACM UIST'94 Symp. on User Interface Software & Technology, 1994, pp. 213-222.
6. Hix, D., Templeman, J., Jacob, R., "Pre-Screen Projection," ACM CHI'95 Conference on Human Factors in Computing Systems, 1995, pp. 226-233.