The World through the Computer: Computer Augmented Interaction with Real World Environments

Jun Rekimoto and *Katashi Nagao*
Sony Computer Science Laboratory Inc.
Takanawa Muse Building,
3-14-13, Higashi-gotanda, Shinagawa-ku,
Tokyo 141 Japan
{rekimoto,nagao}@csl.sony.co.jp
http://www.csl.sony.co.jp/person/{rekimoto,nagao}.html

ABSTRACT
Current user interface techniques such as WIMP or the desk-top metaphor do not support real world tasks, because the focus of these user interfaces is only on human–computer interactions, not on human–real world interactions. In this paper, we propose a method of building computer augmented environments using a situation-aware portable device. This device, called *NaviCam*, has the ability to recognize the user's situation by detecting color-code IDs in real world environments. It displays situation sensitive information by superimposing messages on its video see-through screen. Combination of ID-awareness and portable video-see-through display solves several problems with current ubiquitous computers systems and augmented reality systems.

KEYWORDS: user-Interface software and technology, computer augmented environments, palmtop computers, ubiquitous computing, augmented reality, barcode

INTRODUCTION
Computers are becoming increasingly portable and ubiquitous, as recent progress in hardware technology has produced computers that are small enough to carry easily or even to wear. However, these computers, often referred to as PDAs (Personal Digital Assistant) or palmtops, are not suitable for traditional user-interface techniques such as the desk-top metaphor or the WIMP (window, icon, mouse, and a pointing device) interface. The fundamental limitations of GUIs can be summarized as follows:

Explicit operations GUIs can reduce the cognitive overload of computer operations, but do not reduce the volume of operations themselves. This is an upcoming problem for portable computers. As users integrate their computers into their daily lives, they tend to pay less attention to them. Instead, they prefer interacting with each other, and with objects in the real world. The user's focus of interest is not the human–computer interactions, but the human–real world interactions. People will not wish to be bothered by tedious computer operations while they are doing a real world task. Consequently, the reduction of the amount of computer manipulation will become an issue rather than simply how to make existing manipulations easier and more understandable.

Unaware of the real world situations Portability implies that computers will be used in a variety of situations in the real world. Thus, dynamical change of functionalities will be required for mobile computers. Traditional GUIs are not designed for such a dynamic environment. Although some context sensitive interaction is available on GUIs, such as *context sensitive help*, GUIs cannot deal with real world contexts. GUIs assume an environment composed of desk-top computers and users at a desk, where the real world situation is less important.

Gaps between the computer world and the real world Objects within a database, which is a computer generated world, can be easily related, but it is hard to make relations among real world objects, or between a real object and a computer based object. Consider a system that maintains a document database. Users of this system can store and retrieve documents. However, once a document has been printed out, the system can no longer maintain such an output. It is up to the user to relate these outputs to objects still maintained in the computer. This is at the user's cost. We thus need computers that can understand real world events, in addition to events within the computer.

Recently, a research field called *computer augmented environments* has been emerged to address these problems [18]. In this paper, we propose a method to build a computer augmented environment using a portable device that has an ability to recognize a user's situation in the real world. A user can see the world through this device with computer augmented information regarding that situation. We call this interaction style *Augmented Interaction*, because this device enhances the ability of the user to interact with the real world environment.

This paper is organized as follows. In the next section, we briefly introduce the idea of proposed interaction style. The

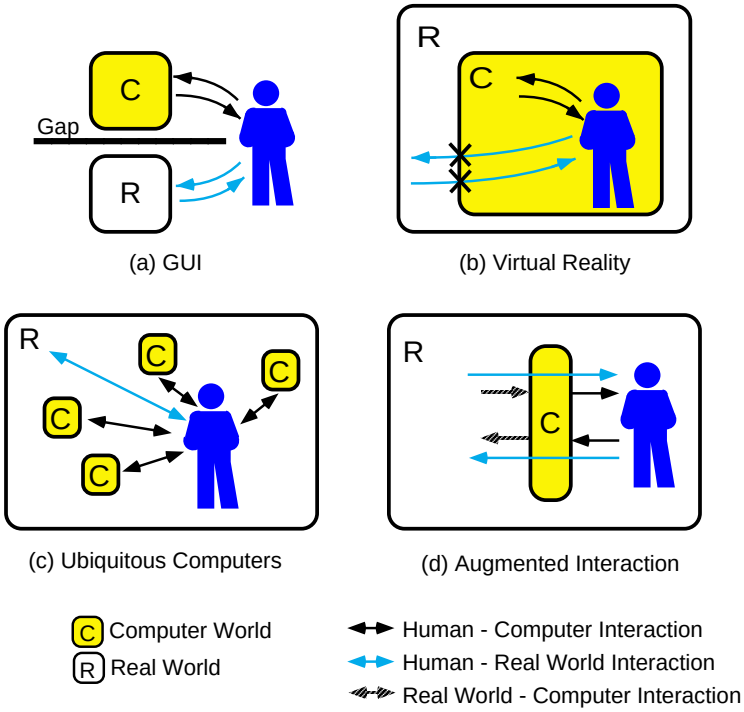

Figure 1: A comparison of HCI styles

following three sections present the NaviCam system, its applications, and its implementation issues. Comparison to other work and our future plans are also discussed in the RELATED WORK section and the FUTURE DIRECTIONS section, respectively.

SITUATION AWARENESS AND AUGMENTED INTERACTION

Augmented Interaction is a style of human-computer interaction that aims to reduce computer manipulations by using environmental information as implicit input. With this style, the user will be able to interact with a real world augmented by the computer's synthetic information. The user's situation will be automatically recognized by using a range of recognition methods, that will allow the computer to assist the user without having to be directly instructed to do so. The user's focus will thus not be on the computer, but on the real world. The computer's role is to assist and enhance interactions between humans and the real world. Many recognition methods can be used with this concept. Time, location, and object recognition using computer vision are possible examples. Also, we can make the real world more understandable to computers, by putting some marks or IDs (bar-codes, for example) on the environment.

Figure 1 shows a comparison of HCI styles involving human–computer interaction and human–real world interaction.

(a) In a desk-top computer (with a GUI as its interaction style), interaction between the user and the computer is isolated from the interaction between the user and the real world. There is a gap between the two interactions. Some researchers are trying to bridge this gap by merging a real desk-top with a desk-top in the computer [12, 17]. (b) In a virtual reality system, the computer surrounds the user completely and interaction between the user and the real world vanishes. (c) In the ubiquitous computers environment, the user interacts with the real world but can also interact with computers embodied in the real world. (d) Augmented Interaction supports the user's interaction with the real world, using computer augmented information. The main difference between (c) and (d) is the number of computers. The comparison of these two approaches will be discussed later in the RELATED WORK section.

NAVICAM

As an initial attempt to realize the idea of Augmented Interaction, we are currently developing a prototype system called *NaviCam* (NAVIgation CAMera). NaviCam is a portable computer with a small video camera to detect real-world situations. This system allows the user to view the real world together with context sensitive information generated by the computer.

NaviCam has two hardware configurations. One is a palmtop computer with a small CCD camera, and the other is a head-up display with a head-mounted camera (Figure 2). Both configurations use the same software. The palmtop configuration extends the idea of position sensitive PDAs proposed by Fitzmaurice [9]. The head-up configuration is a kind of *video see-through HMD* [2], but it does not shield the user's real sight. Both configurations allow the user to interact directly with the real world and also to view the computer augmented view of the real world.

The system uses color-codes to recognize real world situations. The color-code is a sequence of color stripes (red or blue) printed on paper that encodes an ID of a real world

Figure 2: Palmtop configuration and head-up configuration

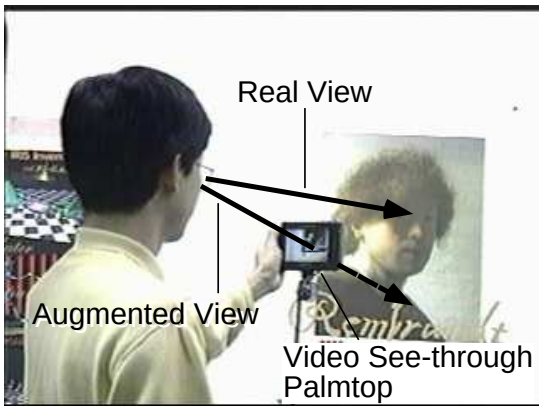

Figure 3: The magnifying glass metaphor

object. For example, the color-code on the door of the office identifies the owner of the office. By detecting a specific color-code, NaviCam can recognize where the user is located in the real world, and what kind of object the user is looking at. Figure 5 shows the information flow of this system. First, the system recognizes a color-code through the camera. Image processing is performed using software at a rate of 10 frames per second. Next, NaviCam generates a message based on that real world situation. Currently, this is done simply by retrieving the database record matching the color-coded ID. Finally, the system superimposes a message on the captured video image.

Using a CCD camera and an LCD display, the palmtop NaviCam presents the view at which the user is looking as if it is a transparent board. We coined the term *magnifying glass metaphor* to describe this configuration (Figure 3). While a real magnifying glass optically enlarges the real world, our system enlarges it in terms of *information*. Just as with a real magnifying glasses, it is easy to move NaviCam around in the environment, to move it toward an object, and to compare the real image and the information-enhanced image.

APPLICATIONS

We are currently investigating the potential of augmented interaction using NaviCam. There follows some experimental applications that we have identified.

Augmented Museum

Figure 4: NaviCam generates information about Rembrandt

Figure 4 shows a sample snapshot of a NaviCam display. The system detects the ID of a picture, and generates a description of it. Suppose that a user with a NaviCam is in a museum and looking at a picture. NaviCam identifies which picture the user is looking at and displays relevant information on the screen. This approach has advantages over putting an explanation card beside a picture. Since NaviCam is a computer, it can generate personalized information depending on the user's age, knowledge level, or preferred language. Contents of explanation cards in today's museums are often too basic for experts, or too difficult for children or overseas visitors. NaviCam overcomes this problem by displaying information appropriate for the owner.

Active Paper Calendar

Figure 6 shows another usage of NaviCam. By viewing a calender through NaviCam, you can see your own personal

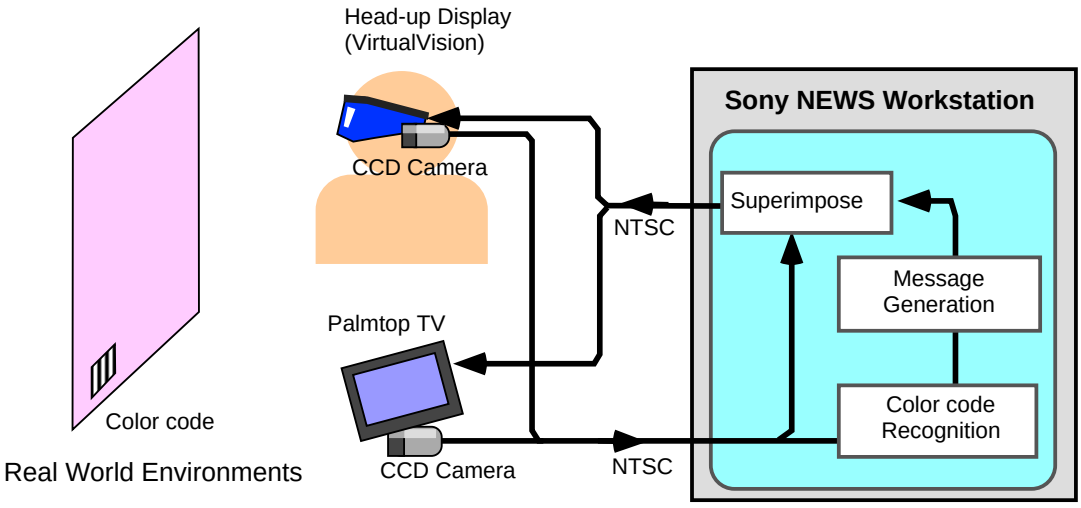

Figure 5: The system architecture of NaviCam

Figure 6: Viewing a paper calender through NaviCam

Figure 7: A pseudo-active office door greets a visitor

schedule on it. This is another example of getting situation specific and personalized information while walking around in real world environments. NaviCam can also display information shared among multiple users. For example, you could put your electronic annotation or voice notes on a (real) bulletin board via NaviCam. This annotation can then be read by other NaviCam equipped colleagues.

Active Door

The third example is a NaviCam version of the active door (Figure 7). This office door can tell a visitor where the occupier of the office is currently, and when he/she will come back. The system also allows the office occupier to leave a video message to be displayed on arrival by a visitor (through the visitor's NaviCam screen). There is no need to embed any computer in the door itself. The door only has a color-code ID on it. It is, in fact, a passive-door that can behave as an active-door.

NaviCam as a collaboration tool

In the above three examples, NaviCam users are individually assisted by a computer. NaviCam can also function as a collaboration tool. In this case, a NaviCam user (an operator) is supported by another user (an instructor) looking at the same

Figure 8: NaviCam can be used as a collaboration tool

screen image from probably a remote location. Unlike other video collaboration tools, the relationship between the two users is not symmetric, but asymmetric. Figure 8 shows an example of collaborative task (video console operation). The instructor is demonstrating which button should be pressed by using a mouse cursor and a circle drawn on the screen. The instructor augments the operator's skill using NaviCam.

Ubiquitous Talker: situated conversation with NaviCam

We are also developing an extended version of NaviCam that allows the user to operate the system with voice commands, called *Ubiquitous Talker*. Ubiquitous Talker is composed of the original NaviCam and a speech dialogue subsystem (Figure 11). The speech subsystem has speech recognition and voice synthesis capabilities. The NaviCam subsystem sends the detected color code ID to the speech subsystem. The speech subsystem generates a response (either voice or text) based on these IDs and spoken commands from the user. The two subsystems communicate with each other through Unix sockets.

An experimental application developed using Ubiquitous Talker is called the *augmented library*. In this scenario, Ubiquitous Talker acts as a personalized library catalogue. The user carries the NaviCam unit around the library and the system assists the user to find a book, or answers questions about the books in the library (Figure 9).

Figure 9: Ubiquitous Talker being used as a library guide

Ubiquitous Talker would also be an important application in the AI research area. Recognizing dialogue contexts remains one of the most difficult areas in natural language understanding. Real-world awareness allows a solution to this problem. For example, the system can respond to a question such as "Where is the book entitled Multimedia Applications?" by answering "It is on the bookshelf *behind* you.", because the system is aware of which bookshelf the user is looking at. It is almost impossible to generate such a response without using real world information. The system also allows a user to use deictic expressions such as "*this* book", because the situation can resolve ambiguity. This feature is similar to multimodal interfaces such as Bolt's *Put-That-There* system [4]. The unique point in our approach is to use real world situations, other than commands from the user, as a new modality in the human–computer interaction.

For a more detailed discussion of Ubiquitous Talker's natural language processing, please refer to our companion paper [13].

IMPLEMENTATION DETAILS

At this stage, the wearable part of the NaviCam system is connected to a workstation by two NTSC cables and the actual processing is done by the workstation. The workstation component is an X-Window client program written in C. What appears on the palmtop TV is actually an X-window displaying a video image. Video images are transmitted from the video capturing board by using DMA (direct memory access), processed in the system, and sent to the X-Window through the shared-memory transport extension to X.

The following are some of the software implementation issues.

Color code detection

The system seeks out color codes on incoming video images. The image processing is done by software. No special hardware is required apart from video capturing.

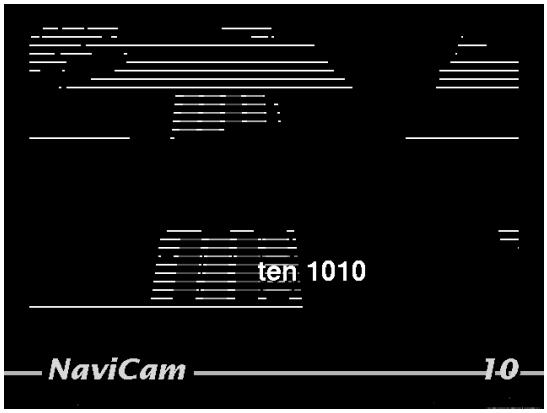

Figure 10: Detecting a color code: a snapshot of what the system is really seeing

The color-code detection algorithm is summarized as follows. First, the system samples some scan lines from the video image (Figure 10). To locate any red and blue bands, each pixel in the scan line is filtered by a color detecting function based on its Y (brightness), R (red) and B (blue) values. Any color bands detected become candidates for a color code. We use the following equations to extract red and blue pixels:

$$\begin{cases} C_1 Y + C_2 < Y - 3R < C_3 Y + C_4 \\ C_5 Y + C_6 < Y - 3B < C_7 Y + C_8 \end{cases} \quad (1)$$

where $Y = R + G + B$, and $C_1, \ldots, C_8$ are constant values. These constants are calculated from sampled pixel values of color-bar images under various lighting conditions. A pixel that satisfies equation 1 is taken as a red pixel. To detect blue pixel, another set of constants $(C'_1, \ldots, C'_8)$ is used.

Next, the system selects the most appropriate candidate as the detected color code. Final selection is based on checks for consistency of distance between the color bands. The detected code is then used to generate information on the screen.

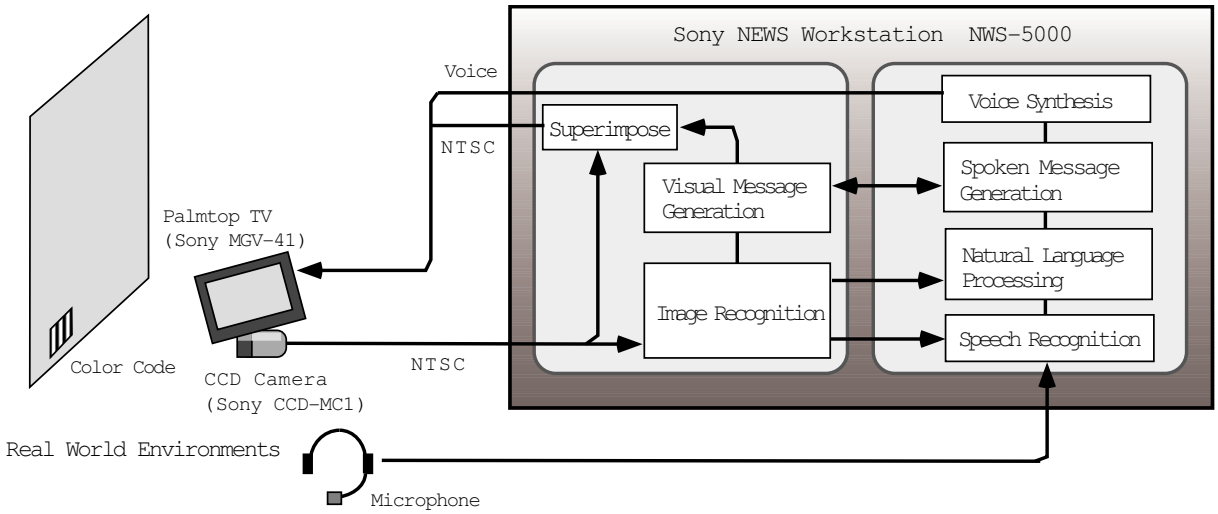

Figure 11: The architecture of Ubiquitous Talker

Using above algorithm, the system can recognize 4-bit color-code IDs (3cm $\times$ 5cm in size) at a distance of 30cm – 50cm using the consumer-based small CCD camera (Sony CCD-MC1). IDs are placed in various environments (e.g., offices, libraries, video studios) so the lighting condition also changes depends on the place and the time. Even under such conditions, the color-detecting algorithm was quite robust and stable. This is because equation 1 compensates an effect on pixel values when lighting condition changes.

Superimposing information on a video image

The system superimposes a generated message on the existing video image. This image processing is also achieved using software. We could also use chromakey hardware, but the performance of the software based superimposition is satisfactory for our purposes, even though it cannot achieve a video-frame rate. The message appears near the detected color code on the screen, to emphasize the relation between cause and effect.

We use a 4-inch LCD screen and pixel resolution is 640 $\times$ 480. The system can display any graphic elements and characters as the X-Window does. However, it was very hard, if not impossible, to read small fonts through this LCD screen. Currently, we use 24-dot or 32-dot font to increase readability. The system also displays a semi-transparent rectangle as a background of a text item. It retains readability even when the background video image (real scene) is complicated.

Database registration

For the first three applications explained in the APPLICATIONS section, the system first recognizes IDs in the real world environment, then determines what kind of information should be displayed. Thus, the database supporting the NaviCam is essential to the generation of adequate information. The current implementation of the system adopts very simplified approach to this. The system contains a group of command script files with IDs. On receipt of a valid ID, the system invokes a script having the same ID. The invoked script generates a string that appears on the screen. This mechanism works well enough, especially at the prototype stage. However, we obviously need to enhance this element, before realizing more complicated and practical applications.

RELATED WORK

In this section, we discuss our Augmented Interaction approach in relation to other related approaches.

Ubiquitous computers

Augmented Interaction has similarities to Sakamura's *highly functionally distributed system* (HFDS) concept [14], his TRON house project, and *ubiquitous computers* proposed by Weiser [16]. These approaches all aim to create a computer augmented *real* environment rather than building a *virtual* environment in a computer. The main difference between ubiquitous computing and Augmented Interaction is in the approach. Augmented Interaction tries to achieve its goal by introducing a portable or wearable computer that uses real world situations as implicit commands. Ubiquitous computing realizes the same goal by spreading a large number of computers around the environment.

These two approaches are complementary and can support each other. We believe that in future, human existence will be enhanced by a mixture of the two; ubiquitous computers embodied everywhere, and a portable computer acting as an intimate assistant.

One problem with using ubiquitous computers is reliability. In a ubiquitous computers world, each computer has a different functionality and requires different software. It is essential that they collaborate with each other. However, if our everyday life is filled with a massive number of computers, we must anticipate that some of them will not work correctly, because of hardware or software troubles, or simply because of their dead batteries. It can be very difficult to detect such problem among so many computers and then fix them. Another problem is cost. Although the price of computers is getting down rapidly, it is still costly to embed a computer in every document in an office, for example.

In contrast to ubiquitous computers, NaviCam's situation aware

approach is a low cost and potentially more reliable alternative to embedding a computer everywhere. Suppose that every page in a book had a unique ID (e.g. bar-code). When the user opens a page, the ID of that page is detected by the computer, and the system can supply specific information relating to that page. If the user has some comments or ideas while reading that page, they can simply read them out. The system will record the voice information tagged with the page ID for later retrieval. This scenario is almost equivalent to having a computer in every page of a book but with very little cost. ID-awareness is better than ubiquitous computers from the viewpoint of reliability, because it does not require batteries, does not consume energy, and does not break down.

Another advantage of an ID-awareness approach is the possibility of incorporating existing ID systems. Today, barcode systems are in use everywhere. Many products have barcodes for POS use, while many libraries use a barcode system to manage their books. If NaviCam can detect such commonly used IDs, we should be able to take advantage of computer augmented environments long before embodied computers are commonplace.

Augmented Reality

Augmented reality (AR) is a variant of virtual reality that uses see-through head mounted displays to overlay computer generated images on the user's real sight [15, 8, 6, 2, 7, 5].

AR systems currently developed use only locational information to generate images. This is because the research focus of AR is currently on implementing correct registration of 3D images on a real scene [1, 3]. However, by incorporating other external factors such as real world IDs, the usefulness of AR should be much more improved.

We have built NaviCam in both head-up and palmtop configurations. The head-up configuration is quite similar to other AR systems, though currently NaviCam does not utilize locational information. We thus have experience of both head-up and palmtop type of augmented reality systems and have learned some of the advantages and disadvantages of both.

The major disadvantage of a palmtop configuration is that it always requires one hand to hold the device. Head-up NaviCam allows for hands-free operation. Palmtop NaviCam is thus not suitable for some applications requiring two handed operation (e.g. surgery). On the other hand, putting on head-up gear is, of course, rather cumbersome and under some circumstances might be socially unacceptable. This situation will not change until head-up gear becomes as small and light as bifocal spectacles are today.

For the ID detection purpose, head-up NaviCam is also somewhat impractical because it forces the user to place their head very close to the object. Since hand mobility is much quicker and easier than head mobility, palmtop NaviCam appears more suitable for browsing through a real world environment.

Another potential advantage of the palmtop configuration is that it still allows traditional interaction techniques through its screen. For example, you could to annotate the real world with letters or graphics directly on the NaviCam screen with your finger or a pen. You could also operate NaviCam by touching a menu on the screen. This is quite plausible because most existing palmtop computers have a touch-sensitive, pen-aware LCD screen. On the other hand, a head-up configuration would require other interaction techniques with which users would be unfamiliar.

Returning to the magnifying glass analogy, we can identify uses for head-up magnifying glasses for some special purposes (e.g. watch repair). The head-up configuration therefore has advantages in some areas, however, even in these fields hand-held magnifying lenses are still dominant and most prefer them.

Chameleon - a spatially aware palmtop

Fitzmaurice's *Chameleon* [9] is a spatially-aware palmtop computer. Using locational information, Chameleon allows a user to navigate through a virtual 3D space by changing the location and orientation of the palmtop in his hand. Locational information is also used to display context sensitive information in the real world. For example, by moving Chameleon toward a specific area on a wall map, information regarding that area appears on the screen. Using locational information to detect the user's circumstances, although a very good idea, has some limitations. First, location is not always enough to identify situations. When real world objects (e.g. books) move, the system can no longer keep up. Secondly, detecting the palmtop's own position is a difficult problem. The Polhemus sensor used with Chameleon has a very limited sensing range (typically 1-2 meters) and is sensitive to interference from other magnetic devices. Relying on this technology limits the user's activity to very restricted areas.

FUTURE DIRECTIONS

Situation Sensing Technologies

We are currently just using a color-code system and a CCD camera to read the code, to investigate the potential of augmented interaction. This very basic color-code system is, however, unrealistic for large scale applications, because the number of detectable IDs is quite limited. We plan to attach a line-sensor to NaviCam and use a real barcode system. This would make the system more practical.

Situation sensing methods are not limited to barcode systems. We should be able to apply a wide range of techniques to enhance the usefulness of the system.

Several, so-called next generation barcode systems have already been developed. Among them, the most appealing technology for our purposes would seem to be the *Supertag* technology invented by CSIR in South Africa [11]. Supertag is a wireless electronic label system that uses a battery less passive IC chip as an ID tag. The ID sensor is comprised of a radio frequency transmitter and a receiver. It scans hundreds of nearby tags simultaneously without contact. Such wireless ID technologies should greatly improve the usefulness of augmented interaction.

For location-detection, we could employ the global positioning system (GPS) which is already in wide use as a key-component of car navigation systems. The personal handy phone system (PHS) is another possibility. PHS is a micro-cellular wireless telephone system which will come into oper-

ation in Japan in the summer of 1995. By sensing which cell the user is in, the system can know where the user is located.

A more long-range vision would be to incorporate various kinds of vision techniques into the system. For example, if a user tapped a finger on an object appearing on the display, the system would try to detect what the user is pointing to by applying pattern matching techniques.

Obviously, combining several information sources (such as location, real world IDs, time, and vision) should increase the reliability and accuracy of situation detection, although the inherent problems are not trivial. This will also be another future direction for our research.

Inferring the user's intention from the situation

Recognized situations are still only a clue to user's intentions. Even when the system knows where the user is in and at which object the user is looking, it is not a trivial problem to infer what the user wants to know. This issue is closely related to the design of agent-based user interfaces. How do we design an agent that behaves as we would want? This is a very large open-question and we do not have immediate answer to this. It may be possible to employ various kinds of intelligent user interface technologies such as those discussed in [10].

CONCLUSION

In this paper, we proposed a simple but effective method to realize computer augmented environments. The proposed augmented interaction style focuses on human–real world interaction and not just human–computer interaction. It is designed for the highly portable and personal computers of the future, and concentrates on reducing the complexity of computer operation by accepting real world situations as implicit input. We also reported on our prototype system called NaviCam, which is an ID-aware palmtop system, and described some applications to show the possibilities of the proposed interaction style.

ACKNOWLEDGMENTS

We would like to thank Mario Tokoro for supporting our work. We would also like to thank Satoshi Matsuoka, Shigemitsu Oozahata and members of Sony CSL for their encouragement and helpful discussions. Special thanks also go to Yasuaki Honda for Figure 7 and assisting video production, and to Tatsuo Nagamatsu for information on the video capturing hardware.

REFERENCES

1. Ronald Azuma and Gary Bishop. Improving static and dynamic registration in an optical see-through HMD. In *Proceedings of SIGGRAPH '94*, pp. 197–204, July 1994.

2. Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *Computer Graphics*, Vol. 26, No. 2, pp. 203–210, 1992.

3. Michael Bajura and Ulrich Neumann. Dynamic registration correction in augmented-reality systems. In *Virtual Reality Annual International Symposium (VRAIS) '95*, pp. 189–196, 1995.

4. R. A. Bolt. Put-That-There: voice and gesture at the graphics interface. *ACM SIGGRAPH Comput. Graph.*, Vol. 14, No. 3, pp. 262–270, 1980.

5. Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of UIST'93, ACM Symposium on User Interface Software and Technology*, pp. 145–155, November 1993.

6. Steven Feiner, Blair MacIntyre, and Doree Seligmann. Annotating the real world with knowledge-based graphics on a see-through head-mounted display. In *Proceedings of Graphics Interface '92*, pp. 78–85, May 1992.

7. Steven Feiner, Blair MacIntyre, and Doree Seligmann. Knowledge-based augmented reality. *Communication of the ACM*, Vol. 36, No. 7, pp. 52–62, August 1993.

8. Steven Feiner and A. Shamash. Hybrid user interfaces: Breeding virtually bigger interfaces for physicall smaller computers. In *Proceedings of UIST'91, ACM Symposium on User Interface Software and Technology*, pp. 9–17, November 1991.

9. George W. Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Communication of the ACM*, Vol. 36, No. 7, pp. 38–49, July 1993.

10. Wayne D. Gray, William E. Hefley, and Dianne Murray, editors. *Proceedings of the 1993 International Workshop on Intelligent User Interfaces*. ACM press, 1993.

11. Peter Hawkes. Supertag - reading multiple devices in a field using a packet data communications protocol. In *CardTech/SecurTech '95*, April 1995.

12. Hiroshi Ishii. TeamWorkStation: towards a seamless shared workspace. In *Proceedings of CSCW '90*, pp. 13–26, 1992.

13. Katashi Nagao and Jun Rekimoto. Ubiquitous Talker: Spoken language interaction with real world objects. In *Proc. of IJCAI-95*, 1995.

14. Ken Sakamura. The objectives of the TRON project. In *TRON Project 1987: Open-Architecture Computer Systems*, pp. 3–16, Tokyo, Japan, 1987.

15. Ivan Sutherland. A head-mounted three dimensional display. In *Proceedings of FJCC 1968*, pp. 757–764, 1968.

16. Mark Weiser. The computer for the twenty-first century. *Scientific American*, 1991.

17. Pierre Wellner. Interacting with paper on the DigitalDesk. *Communication of the ACM*, Vol. 36, No. 7, pp. 87–96, August 1993.

18. Pierre Wellner, Wendy Mackay, and Rich Gold. Computer augmented environments: Back to the real world. *Communication of the ACM*, Vol. 36, No. 7, August 1993.