A Camera-Based Touch Interface for Pervasive Displays

Gerald D. Morrison

SMART Technologies Inc., 1207 – 11 Ave. SW, Suite 300,
Calgary, AB, Canada T3C 0M5
GeraldM@smarttech.com

Abstract. Human-computer interaction using large-format pervasive displays is an active area of research that focuses on how humans can better work with displayed information that may be all around them. In order for this to happen, there must be an enabling technology that creates the interface between users the underlying systems driving the displays. Touch capabilities in a large-format display would be advantageous as a large and pervasive display area is informationally dense and touch provides a natural, life-size interface to that information. Since not all displays are large-format but interactivity may still be desirable, it is optimal if the touch enabling technology can scale to various sizes. This paper describes a new enabling technology in the form of a camera-based input device which uses smart cameras to analyze a scene directly in front of a computer display whether that display is wall sized or desktop sized. The system determines where a user has touched the display, and then treats that information as a mouse click, thereby controlling the computer. The technology is ideally suited to pervasive display systems where scalability or unique installations are desired.

1 Introduction

Research and applications in man-machine interfaces or devices often use touch as an enabler. The ability to touch a computer display and manipulate displayed objects or control applications is a natural extension of everyday human experience where touch is a critical sense. As a result, many types of touch systems have been developed to address this need from point-of-sale systems, to information kiosks, to electronic whiteboards and conferencing. However, there are unique considerations for the different applications. For example, when using a desktop display for touch input as well as pen input, the system must recognize finger for touch but must also have some ability to handle palm reject when being used with a pen. As displays get larger, multi-user collaboration is important and the ability to recognize two or more simultaneous inputs is desirable. Wall-size touch displays are very interesting for users as they allow them to work with computer application in sizes that are comparable with the size of the human body. Human-body sized interfaces naturally facilitate interac-

tivity with the user and collaboration among several users and this paper will describe two such research installations.
Although many touch technology systems are suitable for small display types, not all commercially available touch screen technologies are suitable for very large format displays. Machine vision, however, is quite suitable and can scale from desktop display formats to room-sized displays. Vision systems are attractive because they are more versatile than other input devices since they add capabilities similar to human vision. They can also be made inexpensively as more functionality is integrated into a single chip.
The novel technology presented in this paper uses a vision system comprising of smart CMOS cameras in the corners of a display to look along the surface of the display and determine the location of an object in front of, or in contact with, the display. Typically, the object of interest is the user's finger. The images collected by the cameras are processed in such a way as to recognize various attributes of the object(s), such as location relative to the display in three dimensional space (for one example). This information is then used in feedback to the computer generating the display, enabling touch control of the application.
There can be as few as two cameras in a system (for desktop displays) or there can be as many as are required to observe the display (a wall-sized system for example, needs multiple pairs of cameras to observe the entire surface). The vision architecture and some research applications will be described as well as some thoughts on future directions.

2 Methodology

The recent ability to embed significant processing power onto the same printed circuit board as a camera sensor, as opposed to the traditional approach of using a frame grabber and discrete computing system, has led to tighter integration of the camera components and a smaller size for commercial deployment. In this section, the hardware is briefly described. More detail can be found in [1].

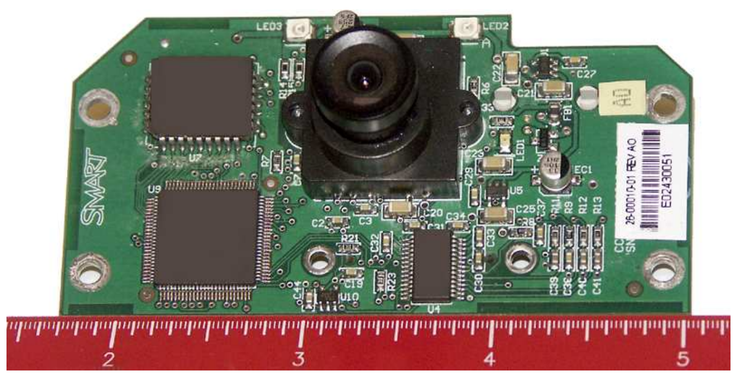

Fig. 1. A custom designed DSP-based Smart Camera showing size in inches.

The novel technology presented in this paper uses smart CMOS cameras (see Fig. 1) in the corners of a display to look along the surface of the display and determine the location of an object in front of, or in contact with, the display.

Typically, the object of interest is the user's finger. The images collected by the cameras are processed in such a way as to recognize various attributes of the object(s), such as location relative to the display in three-dimensional space (as one example). This information is then used in feedback to the computer generating the display so that touch control of the application can be enabled. The diagram of the electronic system architecture is shown in Fig. 2. Note that there can be many cameras in a system but for single rectangular systems, the number of cameras is typically limited to four (one in each corner). Other systems can function with as few as two cameras or as many as are required to observe the display (a wall size system needs multiple pairs of cameras to observe the entire surface – see Fig. 3 for example).

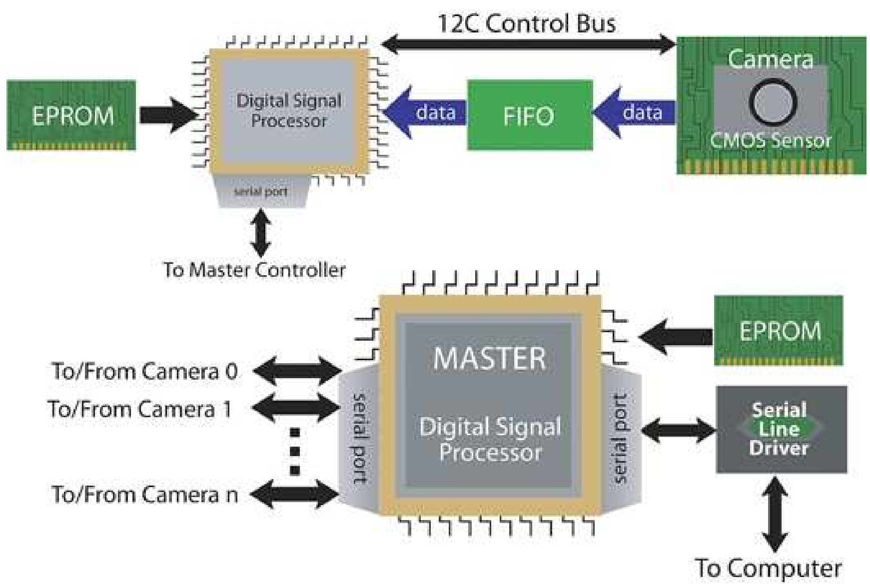

Fig. 2. The hardware architecture used for the camera-based touch enabler.

3 Implementation Results

This section describes a few interesting research projects that were built using the technology described in the previous section. Figure 3 shows a large wall display that was built for studies being conducted a Georgia Tech. The system uses 12 cameras to provide a very robust touch system and to cover the 17-foot span while still providing excellent functionality (writing, click and drag, double-click, etc.). In the figure, three projected images are seamed together to make one large homogeneous display. Conversely, in the Georgia Tech research, the displayed information is constructed from a six-projector array that overlaps so that the user shadow is significantly reduced [2].

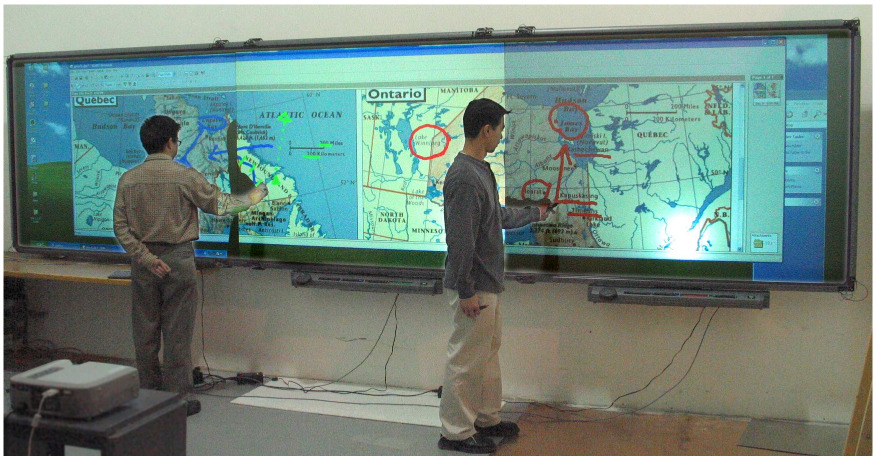

Fig 3 A 17-foot-wide wall display that is multi-user touch sensitive. System pictured uses three projectors to form one image.

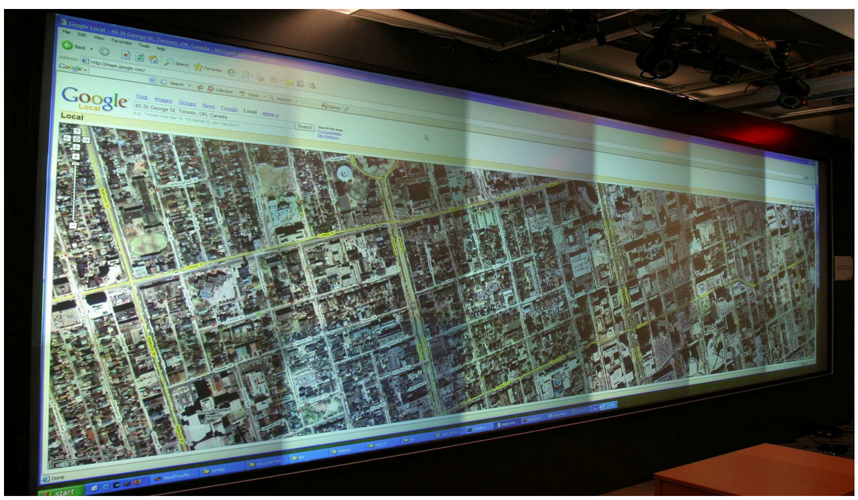

Fig 4 An 18 x 6 foot touch enabled high resolution wall display using 18 projectors.

Large format high resolution displays are becoming increasingly important in scientific and educational institutions, and projector arrays are a common way of constructing them [3]. The compelling visuals, information density and high resolution make them ideal for a variety of applications. Until now the display could only be manipulated through traditional input methods and so some applications, like those that use hand-writing for example, were awkward and ineffective. By adding touch capability, new applications and uses are possible. In fact, it is now possible to replace a traditional classroom blackboard with a fully electronic, fully interactive, and more informationally dense wall-size equivalent. Figure 4 shows an example of this type of system that was constructed for the University of Toronto. It uses 18 projectors and the display surface measures 18 feet wide by 6 feet tall. In this system, touch is enabled by cameras mounted along the top periphery of the display.

Machine vision can be used to create a touch interface of arbitrary size due to the fact that the resolution of the system is optically based. This means that for any projected display, there is a proportional correspondence between the display pixels and the camera pixels. Therefore, if the resolution of the cameras system allows a touch resolution of say, one display pixel, that touch resolution will not change no matter how small or how large the display pixel is. In other words, whether an SXGA display occupies a 15″LCD or is displayed on a 50′ wall, the capabilities of the machine vision system remain the same as long as the camera resolution is sufficient for pointer identification. This optical scaling ability is an important feature when making large format displays touch sensitive, but it also allows for creation small display touch systems. Figures 3 and 4 show two types of wall-sized touch displays. These are presently the largest touch interactive displays available

Fig 5**. A 17" touch display with finger and pen input.**

Figure 5 shows a 17" LCD system that is suitable for desktop applications or point-of-sale terminals. This system is unique in the it works with both passive input (a finger) and active input (a tethered IR pen). The touch enabling infrastructure consists of 2 cameras and an opposing mirror. Image analysis is performed on the superposition of direct and reflected images, thereby providing a superb palm-reject feature. An automatic mode switch is located in the pen docking recess to provide a highly intuitive and seamless switch from pen mode to finger mode.

Figure 6 shows an 84” camera-based touch sensitive display that can be used for presentation, training or any type of computer application control. The upper corners of the display are not covered thereby revealing the location of the cameras.

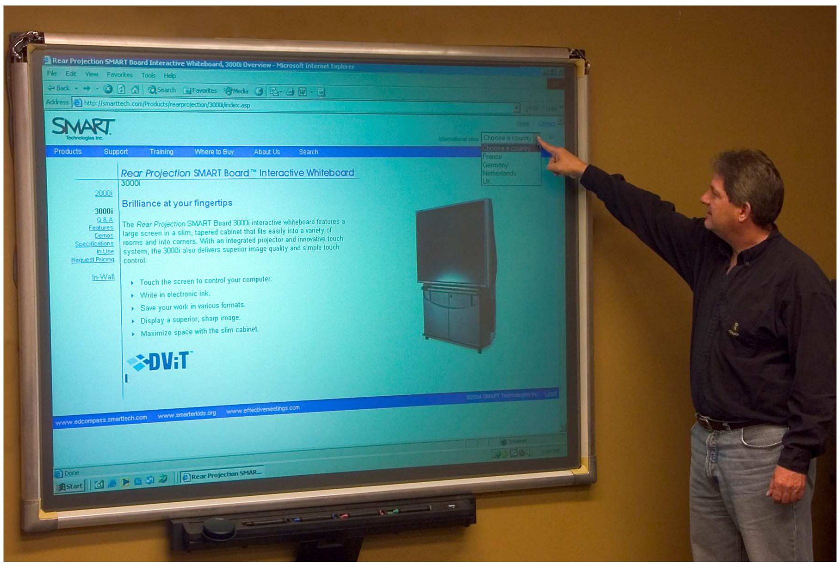

Fig 6: An 84” diagonal touch display suitable for training and presentations.

With all of these systems it is possible to control the computer by touching (clicking) the icons, dragging information, writing text, or anything else that can be accomplished with a mouse. The camera-based touch enabling technology is unique as it allows for operation from small-size to wall-size displays.

4 Future Directions

Now that camera-based touch technology as been demonstrated, much can be done to exploit its vision capabilities. One exciting area is that of on-contact gesture control. This is different from typical hand waving gestures in that the gestures are formed by actions that are only in contact with the surface. As an illustration, the right-mouse-click function is very useful in a variety of applications. However, touch systems that provide on a single touch point cannot accommodate this easily as the touch contact is always interpreted as a left-click action. However, with a vision system, it is possible to provide an intuitive right click action by using the multiple touch capabilities of the technology. This can be done, for example, by touching the display with the right

index finger and then subsequently touching the display with the right middle finger (while holding the index finger in place). This multiple simultaneous touch action is an unambiguous gesture that will always be interpreted correctly. Another possibility is two finger scrolling for scrolling through documents that can be implemented by simply touching the display with two fingers and then dragging in the direction in the desired scroll. Again, this requires a multiple simultaneous contact that is unambiguous from a single contact.
Artificial vision is a voluminous area of research and even cognition-like abilities have been demonstrated. With the application of machine vision to touch enabled displays, we now have the capability to augment human computer interaction within pervasive display systems in ways that go beyond the traditional touch-as-mouse activities.

References

1. Morrison, G.:, A CMOS Camera-based Man-Machine Input Device for Large-Format Interactive Displays. IEEE Computer Graphics and Applications, 25(4). p. 52-57
2. Summet, J., Flagg, M., Rehg, J., Corso, G., Abowd, G.: Increasing the Usability of Virtual Rear Projection, International Workshop on Projector-Camera Systems, Nice, France, October 2003
3. Bezerianos, A., Balakrishan, R.: View and space management on large displays, IEEE Computer Graphics and Applications, 25(4). p. 34-43