InfrActables: Multi-User Tracking System for Interactive Surfaces

Christoph Ganser[1] Adrian Steinemann Andreas Kunz[1]

ICVR – Innovation Center Virtual Reality
Swiss Federal Institute of Technology
Zurich

ICVR – Innovation Center Virtual Reality
Swiss Federal Institute of Technology
Zurich

ICVR – Innovation Center Virtual Reality
Swiss Federal Institute of Technology
Zurich

ABSTRACT

We present InfrActables, a wireless technology for human computer interaction devices. It allows multiple users to interact simultaneously on back-projection displays. It recognizes each device's position, orientation, and identification, but also enables the tools to communicate their states to the application that the user interacts with. This makes it possible to build complex interaction devices for direct manipulation with buttons, sliders, and other input capabilities. The technology was developed for a computer-supported environment, allowing multiple users to interact simultaneously on a surface using multiple styli and other user input devices. The principle of operation is not limited to two-dimensional surfaces, three-dimensional user input devices can also profit from its advantages.

CR Categories and Subject Descriptors: I.3.1 [Computer Graphics]: Hardware Architecture: Input Devices. 1.3.6 [Computer Graphics]: Methodology and Techniques: Device Independence, Ergonomics, Interaction Techniques.
Additional Keywords: Tracking, User Input Devices, Single Display Groupware, Stylus-based User Input Device

1 INTRODUCTION

Computer supported collaborative work is gaining more and more attention from industry and research. This not only because of the increasing need for remote collaboration between teams at different locations, but also because today's development goals are getting more and more complex, making interdisciplinary team work necessary. Communication between the team members gets even more complicated because of the technical terminology used in different disciplines. Whereas a group of individual, distributed users can benefit from a rich range of conferencing products on the market, only a few tools are available for electronic meetings of distributed and co-located teams, allowing simultaneous interaction.

Tools supporting visual communication by sharing graphical content and sketching can be a great support for co-located collaboration and communication. Thus, it is not astonishing that today's interactive whiteboards are widely used in industry and education. The products that are available on the market are designed for vertical interactive surfaces and allow one single user to interact at a time. This is suitable for presentation applications. But research in the field of Single Display Groupware (SDG) has shown, that communication in discussion and solution finding [1] can be improved by allowing multiple users to interact with a shared whiteboard at the same time.

Previous research [2] has shown that these non-hierarchical meetings can be supported better by a horizontal interactive surface like a table than by a vertical display. Most of the tracking technologies used for vertical displays turn out to be unsuitable for horizontal surfaces, as a user will place elbows and/or other things on the table.

2 REQUIREMENTS

The following requirements were set up for the interactive table in our computer supported environment [3]. Four people should be supported to interact simultaneously on a horizontal back-projection display. No cables should hinder the user's interaction on the display. The input devices should be identifiable to the system. This will allow the mapping of a device to attributes like color, line width, or to users, e. g. "Peter made this stroke".

The interaction devices' detection has to be robust, since users will put all kinds of objects onto the interaction surface like elbows, mobile phones, coffee cups, and laptops.

In order to allow smooth sketching, the system should support at least a refresh rate of 10 Hz. An analysis of the scribble writing developed for the Palm PDA indicated a minimum refresh rate of 12 Hz. Following [4], 10 Hz is the bottom line. However, a higher update rate is preferable.

The delay between the user's motion and the system's graphical response should be less than 120 milliseconds. We measured these response times in existing commercial single user systems.

For the styli, we defined three different device states. "Hovering" allows controlling and moving the pointer. "Drawing" is used for sketching, and "Right Button" is used to bring up a context menu.

The extensibility of the system is another requirement, which allows the tracking technology to be used in many applications with various user interface concepts.

3 RELATED WORK

Many input technologies for interactive surfaces exist on the market and in research. Most of them are specialized in presentation applications on an electronic whiteboard. We present only the few technologies that are related to track multiple input devices and that can optionally identify the input devices' states.

LiveBoard [5] and its software suit 'Tivoli' [6] are a pioneering work on electronic whiteboards. The Tivoli paper envisions simultaneous multi-user interaction and remote collaboration.

Chen and Davis [7] used laser pointers and thus allow multiple users to interact on a high resolution tiled display. In Oh and Sturzlinger [8], a system is described that enables three users to interact simultaneously with laser pointers on a front projection display. The devices are cable-bound. In order to identify the devices, they are turned off and on again every third frame.

Wacom Intuos2 graphic tablets [9] offer a feature called "Dual Track". It allows up to two input devices like stylus or mouse to be tracked. The tablets are available up to a size of 616 x 446 mm.

[1] (ganser | kunz)@inspire.ethz.ch

Proceedings of the IEEE Virtual Reality Conference (VR'06)
1087-8270/06 $20.00 © 2006 IEEE

DiamondTouch [4] is an outstanding multi-user input technology allowing the user to interact with his body (Finger, Hand, Arm) on a horizontal or vertical front projection display. The detection of the user's finger is realized by capacitive sensing. A high-frequency signal is sent from the interaction table through the user's body into the receiver, which is integrated in the chair. The interactive surface measures 42 inches in the diagonal. In addition to the user's body, a cable-bound pen connected to a separate receiver is available to the user as an additional pointing device.

Six-DOF tracking systems offer additional degrees for pitch, yaw, and elevation of multiple tracked devices. Besides many advantages, they tend to be very expensive.

A motion capturing system [10] detects up to 120 tracked targets using a similar technique to our approach. A microcontroller is used to transmit a unique code from the tracking target to the multiple optical detection units. The system costs more than 50'000$ and features neither device states nor a synchronization between the devices and the detection unit. [11] is closely related to [10], implementing an HDLC protocol with a data rate of 7/8 bit per second from an IR diode over a standard camera. A cable-bound system for finger tracking can be found in [12]. The active LEDs are time sequentially addressed. States are not present as the system was developed for finger tracking.

4 INFRACTABLES

Out of the requirements, we decided to implement the following functions into InfrActables: Simultaneous position and orientation tracking of multiple interaction tools, identification of these tools using an ID, and the transmission of the devices' states to the application. Even though these tasks have to be solved wirelessly, they still allow many approaches in principle to the overall solution.

The usage of RF technology was especially tempting, as it would facilitate the transmission of almost unlimited device state information with the ID of the emitting device.

RF technology like RF transceivers (15.00 $ per input device) for the states, and computer vision for the tracking would lead to the problem of association, as both streams need to identify themselves with the system. For this reason, we decided to go for a pure optical system.

Shape recognition of interaction devices would restrict the construction of the stylus. A changeable device shape would be needed to transmit the devices' states. These considerations brought us to a very general solution, in which the devices communicate in a time and in a location multiplexed way using infrared diodes and a camera. This means that every device has at least one diode to transmit a binary code. The codes from the devices appear in the camera's image and have to be analyzed to generate the events for the user interface components.

Affordable IEEE 1394 cameras (around 1200 $) feature a frame rate between 74 Hz [13] and 200 Hz [14]. Compared to base technologies like sound and RF, this is rather a low sampling rate, defining the bottleneck for the InfrActables. This led us to the design decision that a synchronized system, in which all parts are in the same state, would bring a performance advantage. According to the Nyquist sampling theorem, an unsynchronized system would need a frame rate (sampling rate) twice as high as the communication rate.

4.1 Principle of Operation

A bit code is transmitted from the interaction devices to the detection system. This code must contain the ID and the state of the input device. The device's location is detected from the position, in which the code appears in the camera's image. Thus, a code consisting completely of zeros can not be used, as it would never light up the IR-LED and stay invisible to the camera.

The bit code is transmitted by the IR-LED mounted on the interaction device through the back projection display to a camera underneath the interaction surface. An optical IR-filter blocks all light in the visible spectrum. This simplifies the detection in the captured images, as no interference with the projected image occurs. The computer now processes the captured images and transfers the corresponding events to our event propagation system, which routes them to the application. The application can now react on the user's input and draw a sketched line into the frame buffer, for instance, giving the user the visual feedback.

4.2 Synchronization

In order to ensure the synchronization between all components, a custom-made electronic device (SyncBox) was built. Its main functionality is communicating with the processing unit what bit should be acquired from the input devices. The SyncBox must then signal the bit position to the input devices and trigger the camera. The communication with the users' input devices is accomplished using a 36 kHz modulated infrared signal. Figure 1 shows the set-up in principle.

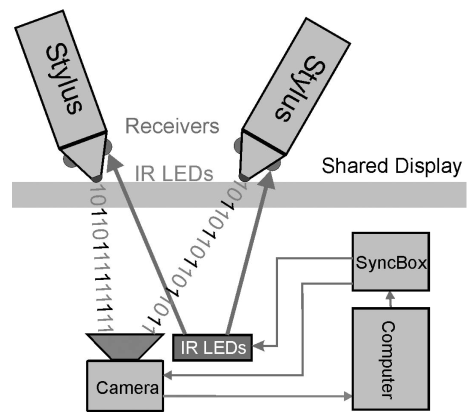

Figure 1. Principle of operation

The overall system clock is determined by the processing unit. If the OS suddenly lags for some milliseconds, the whole tracking system will wait. This principle has also shown itself to be useful for debugging our tracking software, allowing a step by step operation of the system. The processing unit flags the SyncBox over the parallel port what bit position is expected in the acquired frame. The SyncBox uses the length of the modulated IR flash to signal the bit position to the device, and triggers the camera.

The InfrActables' mode of operation can be described in principle as a multi-channel simplex communication between the interaction devices and the application code. The processing unit acts similar to a reading master on an I^2C bus and generates the clock signal.

4.3 Bit Code

For our first test of the InfrActables in the computer supported environment mentioned earlier, we used a code with a length of five bits (3 ID, 2 states). This allows us to use seven interaction devices with four states (Table 1). Using a camera with a frame rate of 60 Hz gives a refresh rate of 12 Hz (frame rate/bit code length) to the input devices.

In other applications, more IDs and fewer states could be preferable. On the other hand, a single user input device with many states (rotary switch, slider) needs only one ID but many bits for the states. InfrActables turns out to be a flexible tool box allowing many different user interaction devices to be created.

ID	Interaction Device
111	Stylus1
110	Stylus2
…	…
001	Device7

State	State of Stylus
11	Drawing
10	Hover
01	Popup-Menu

Table 1. Bit code for ID and device state

The bit code consisting of ID and state does not conform to the requirements known from serial data transmission standards like V.24, in which additional bits are used to signal the beginning and the end of the transmission, and a parity bit helps to detect errors. As it is unknown, when the input devices will appear or disappear in the frames acquired by the camera (e. g. the user suddenly lifts the pen), start and stop bits are required. If a new blinking IR-LED is detected, the start bit can be avoided by waiting for a second period before the ID is determined. As the devices support the hovering state, the additional latency occurs only once at a non-critical moment. In case the user removes the stylus from the detection area, the end of the bit code is forced to be zero. Although the state 00 is not used, a corrupted ID causes the wrong virtual pointer to jump back and forth to the spot at which the other device was removed. This makes the use of a stop bit favorable. Parity bits have not been used so far, as the system runs stable enough for our application.

As the frame rate of the camera is the bottleneck of our tracking system, the update rate for a set-up with 5 data bits plus one stop bit is limited to 12.1 Hz for a 73 Hz camera and to 33.3 Hz for a 200 Hz camera, if all bits have to be individually detected. Since this is relatively slow for sketching and writing, a smarter update algorithm was developed. It normally queries only the function bits, while the ID bits are only analyzed if a new LED appears in the frame. This increased the update rate to 24.3 Hz and to 66.67 Hz, respectively. Another optimization can be applied by using every detected bit to trigger a "moved" event. State 11 will cause a "moved" event to be sent in every frame.

Classless Inter-Domain Routing is a well known concept that could be used for our bit code. Its intention was to break up the strict partition of the network addresses. Concerning the InfrActables, this would mean that an individual length for the ID and the state could be used for every device. To extract the ID or the state from a received code, a bitwise AND operation must be applied to the ID mask or the inverted ID mask as a second operand, see Table 2. The variable division of the bit code into ID and state is sensible for applications using input devices which feature few and many states together. But this concept also affects the optimized update scheme presented in the previous paragraph, as all bits of the device with the longest state variable must be queried. This concept has not been implemented so far.

Code	ID Mask	ID (Code & ID Mask)	State (Code & ~ID Mask)
1 0110	1 0000	**1** 0000	0 0110
010 11	111 00	**010** 00	000 **11**
011 01	111 00	**011** 00	000 **01**
001 01	111 00	**001** 00	000 **01**

Table 2. Flexible ID and state resolution

The previous paragraphs showed some issues related to the use of the bit code. We can now deduce what signals the SyncBox must be able to communicate to the input devices:

A straight forward solution would be to signal each bit position using a modulated IR pulse with a unique length. This would limit the extensibility of our system. A more economic approach sends out signals, which indicate the interaction device how the "pointer" must be manipulated to determine the bit being sent to the camera, e.g.: Goto ID 0, Goto state 0, Goto parity 0, Shift to next bit, Stop bit.

4.4 IDs and LED Topology

As the bit code allows several applications for different user input devices, the same turns out to be true for the use of LEDs with the same or different IDs mounted on a device.

A device with one LED and one ID is the trivial case. It can be tracked in only two dimensions. Adding an additional LED makes it possible to also detect the orientation. However, the initial orientation may be wrong by 180 degrees. A third LED is used to clarify the orientation. In addition, device states must not be limited to the transmitted binary code. It is also possible to assign multiple IDs logically to a device. As an example: using a device with two LEDs with different IDs would reveal the position, orientation and add a high resolution state variable described by the distance between the LEDs. With this strategy, much more sophisticated devices can be generated.

4.5 Hardware

The SyncBox is driven by an ATmega16 microcontroller. It generates the required 36 kHz synchronization signal for the input devices. This signal is sent to an array of infrared diodes. The interaction devices have an infrared receiver demodulator (SFH-5110-36), connected to the external interrupt of the microcontroller inside the device. The 36 kHz modulation inhibits any interference between the transmitting signal, the bit code (the devices' answers), and the synchronization signal. All interaction devices are built on the same hardware platform based on an ATtiny26 microcontroller evaluating the state of the connected keys, reacting on the received synchronization signals, and controlling the flashing of the IR-LED. The IR-LEDs in the InfrActables emit infrared light between 740 and 850 nm in order to be visible in the spectrum of a normal camera. For the receivers, the flash of the SyncBox must be at 950 nm. A step-up DC-DC converter (MAX1674) allows us to use 1.5V to 5V batteries. This flexibility is needed to minimize the physical size of the InfrActables.

4.6 Recognition Software

The recognition software uses three abstraction layers to detect the input device's position, orientation, state, and identification (see Figure 2).

The first layer detects the spots in the image (LEDCenter), resulting in a list containing the detected spots of the IR-LEDs. The segmentation algorithm has been optimized under the premise that the captured images will contain mostly black pixels and the shapes to be segmented are only convex. With a pointer to a 32 bit unsigned integer iterated over the image, four pixels can be compared to zero in one step. Each pixel of a non-zero quadruple is then compared to a threshold value. As only convex shapes are detected, the algorithm only needs to keep track of the previous scan line. During the segmentation, the center of the spots is calculated, using a center of gravity calculation. Small spots below a certain number of pixels will be ignored.

The second layer (LEDPointer) assigns the recognized LEDCenters to objects representing the physical LEDs. This assignment is based on motion estimation and on a simplified best matching algorithm. Unmatched LEDCenters create new LEDPoints. LEDPoints, which are not matched to any LEDCenter, are deleted at the end of the system's period.

The third layer (LEDDriver) contains the logic for interpreting the state and for evaluating the position and orientation of a device from various LEDPoints (see LED topology). It has the function of a driver for each type of user input device like stylus or for any other user device with more DOFs. The drivers generate tracking events, which are passed to the client code via an event handler.

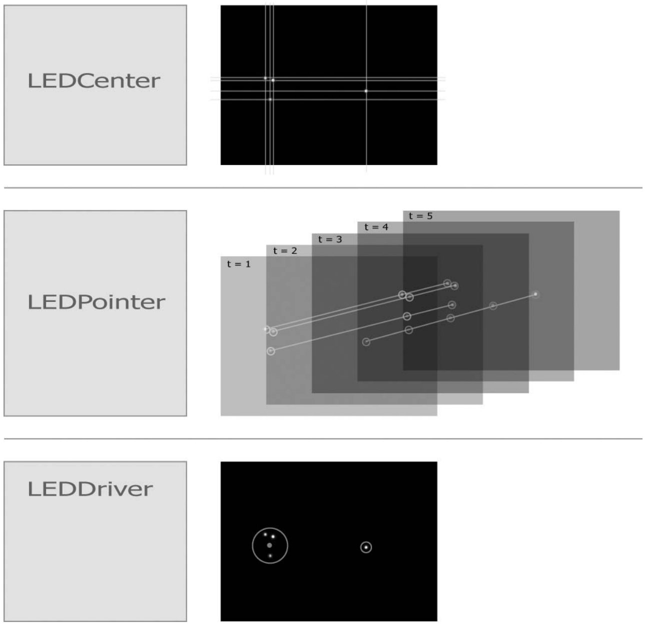

Figure 2. Recognition software layers

5 CONCLUSION

We presented InfrActables, a tracking technology that allows multiple users to interact simultaneously on a back projected display, using pens (styli) and other user input devices.

The technology can be seen as a toolbox, useful to create a large set of user interfaces. The key parameters for the application of InfrActables are:

- The update rate may go up to 200 Hz if only one device without any state is used (bits: ID=0, states=0, stop bit=1). Using CameraLink cameras instead of Firewire would allow even higher rates at higher costs.
- The number of user input devices, for example 64 devices without any states (bits: ID=6, states=0, stop=1)
- The number of states needed by the application. For example, 256 states can be supported for one single device (bits: ID=0, states=8, stop bit=1)

Developers have to balance these parameters according to the requirements of the desired interaction and their set-ups.

6 OUTLOOK

As the technology is developed now, we will intensify our future research towards the creation of novel user interfaces and interaction techniques with InfrActables.

Instead of sending out simple synchronization signals we would like to establish a more sophisticated communication from the application to the InfrActables. This will allow the SDG to give a visual or a haptic feedback to the user through the input device.

We also want InfrActables to enable a seamless multi-display, multi-user interaction. Especially for collocated groups, the combination of horizontal displays with multiple vertical displays seems promising to us.

7 ACKNOWLEDGMENT

This work has been carried out within the togETHer (ETH World) and CRiON (KTI) projects. We would like to thank these organizations for their support. We also would like to thank Thomas Kennel, Samira Andersson, Ramon Hofer, Thomas Kaufmann, and the employees of the ETH workshops for building the set-up, the styli, and the other interaction tools. We also would like to thank Roland Kehl and his colleagues from the Computer Vision Group for their help concerning the calibration.

REFERENCES

[1] Jason Stewart, Benjamin B. Bederson and Allison Druin. Single display groupware: a model for co-present collaboration. in Proceedings of the SIGCHI conference on Human factors in computing systems. 1999. Pittsburgh, Pennsylvania, United States: ACM Press.

[2] Norbert A. Streitz, Jörg Geissler, Torsten Holmer, Shin'ichi Konomi, Christian Müller-Tomfelde, Wolfgang Reischl, Petra Rexroth, Peter Seitz and Ralf Steinmetz, *i-LAND: an interactive landscape for creativity and innovation*, in *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*. 1999, ACM Press: Pittsburgh, Pennsylvania, United States. p. 120-127.

[3] C. Ganser, T. Kennel, N. Birkeland and A. Kunz. Computer-Supported Environment for Creativity Processes in Globally Distribued Teams. in International Conference on Engineering Design. 2005. Melbourne, Australia: Design Society.

[4] Paul Dietz and Darren Leigh. *DiamondTouch: A Multi-User Touch Technology*. in *CHI*. 2001. Orlando, Florida: ACM.

[5] Scott Elrod, Richard Bruce, Rich Gold, David Goldberg, Frank Halasz, William Janssen, David Lee, Kim McCall, Elin Pedersen, Ken Pier, John Tang and Brent Welch. *Liveboard: A Large Interactive Display Supporting Group Meetings, Presentations and Remote Collaboration*. in *CHI*. 1992. Monterey, California, USA: ACM.

[6] Elin Rønby Pedersen, Kim McCall, Thomas P. Moran and Frank G. Halasz. Tivoli: an electronic whiteboard for informal workgroup meetings. in Proceedings of the SIGCHI conference on Human factors in computing systems. 1993. Amsterdam, The Netherlands.

[7] Xing Chen and James Davis, Lumipoint: multi-user laser-based interaction on large tiled displays. Displays, 2002. **23**: p. 205-2122.

[8] J.-Y. Oh and W. Stuerzlinger, *Laser Pointers as Collaborative Pointing Devices*. Graphics Interface 2002, 2002: p. 141-149.

[9] Wacom Intuos2 graphic tablet: http://www.wacom.com/.

[10] PhaseSpace, http://www.phasespace.com/.

[11] Darnell J. Moore, Roy Want, Beverly L. Harrison, Anuj Gujar and Ken Fishkin, Implementing phicons: combining computer vision with infrared technology for interactive physical icons, in Proceedings of the 12th annual ACM symposium on User interface software and technology. 1999, ACM Press: Asheville, North Carolina, United States. p. 67-68.

[12] AR-Tracking, http://www.ar-tracking.de/.

[13] AVT, Marlin F-033B; http://www.alliedvisiontec.com/.

[14] Point Grey, Dragon Fly Express; http://www.ptgrey.com/.