RC 21883 (98425) 3 November 2000 **Computer Science**

IBM Research Report

Portable Whiteboard System with Vision Input

Ferdinand Hendriks, Xiping Wang, Belle Tseng, Zon-yin Shae
IBM Research Division
Thomas J. Watson Research Center
P. O. Box 704
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - T. J. Watson - Tokyo - Zurich

Portable whiteboard system with vision input

Ferdinand Hendriks, Xiping Wang, Belle Tseng, Zon-yin Shae

IBM Research Division, Thomas J.Watson Research Center,
30 Sawmill River Road, Hawthorne, NY 10532

Keywords: Whiteboard, stylus, pen, vision, collaboration, collinear optics

ABSTRACT

We present a novel whiteboard system that uses one or more active input devices. The system is especially suitable for situations in which several users provide input on a large writing surface which also serves as a projection surface. That surface can have general orientation. We have implemented a system that uses an inexpensive, infrared-emitting stylus and an off-the-shelf videoconferencing camera fitted with an IR transmitting/visible blocking filter to capture handwritten strokes. The system uses a calibration method that uses projective mapping thus allowing off-axis camera placement. We also propose a self-calibrating system in which the capture device shares its optical system with the projector. In a system with multiple local users the question of "who wrote what" becomes relevant. Therefore, we address the question of attaching identity to stroke information.

1. INTRODUCTION

In the recent past, a number of inexpensive whiteboard systems have appeared[1,2] that address the ease and cost-of- use issues in electronic whiteboards. They use a combination of ultrasound and infrared techniques to implement digitizer and pen identification functions. Both systems use capture hardware that must be affixed to the writing surface. We base our whiteboard system on the belief that in the future the use of compact and inexpensive cameras and projectors will increase significantly. If this scenario holds true, then such cameras can perform the function of a 2D digitizer with the advantage that the board is not wired to the computer. We may see the gradual replacement of small displays with bright digital projection displays that can project onto many kinds of surfaces. Extrapolating further, we may envision projectors that incorporate the camera function using collinear, shared optics. Large screens facilitate and encourage meetings with multiple users in front of a single large screen and they promote a sense of being immersed in the application. With multiple participants the question "Who wrote what" arises, creating the need to associate written strokes with the person who wrote them. This creates the problem of how to transmit this identity to the whiteboard system. We suggest several approaches ranging from pen personalization to writer recognition We report our experiences with the novel input technique and suggest a framework for the stroke identification system. While our system is essentially a front projection version of Xerox Parc's Liveboard[3], we feel that our system has lowered the threshold for users because it is portable and because of the huge cost reductions that have taken place in projection displays and video capture cameras.
Furthermore, our system can be deployed on walls as well as on desks.

2. SYSTEM ARCHITECTURE

Several workplace field studies have shown the importance of physical whiteboards which are a locale for discussion and collaboration. However, physical whiteboards are only visible locally in one place. Their information is not easily available for remote sites. Especially, strokes written on the whiteboard with ink can not be erased or edited by remote users. Hence, real-time multi-user physical whiteboard collaboration is limited locally to people in the same room. With the development of our light pen technology, system architecture of a networked collaboration system is shown in Figure 1. It consists of two virtual, physical whiteboards connected via the Internet. The virtual strokes generated by electronic light pens are captured by a low cost camera. Low cost projectors can then be used to display the strokes on the whiteboards. Consequently, we can virtualize the physical whiteboard, and enable the remote collaboration for discussion, design, and editing in a networked environment.

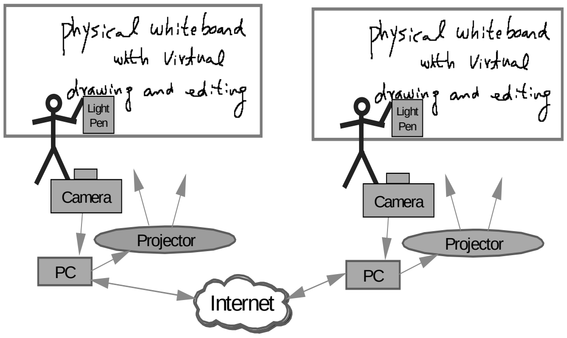


Figure 1 System Architecture

3. IMPLEMENTATION

3.1 System client

A high-level view of our system client is shown in Figure 2. The whiteboard's requirements are that is must be suitable for projection and suitable as a writing substrate. We found that existing whiteboards are quite suitable. Our system also captures audio. We use a wireless microphone which communicates with the system's audio capturing system. Audio is compressed according to the G.723 standard. Positioning of the camera and projector are in principle arbitrary. However we reiterate that it is preferable to have a ceiling mounted projector and camera to avoid occlusion of the lightpen by the user.

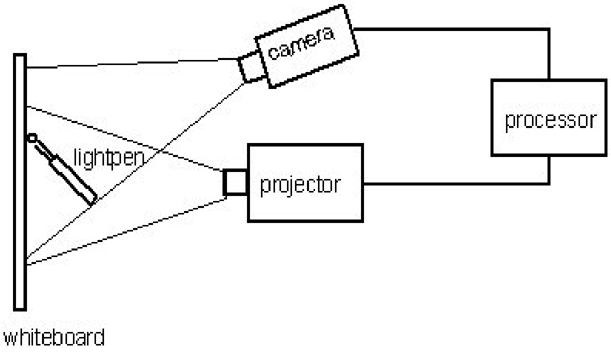


Figure 2 System client running a front

3.2 Input device

We now turn to the details of the input device. In principle, a vision system is capable of tracking the motion of the tip of a pen over a surface. However, it is not easy for such a system to establish whether the pen is in contact with the surface. To make both tasks less error-prone and robust, we chose a writing instrument that features an IR-emitting (at a wavelength of 940 nm) light at its nib, aimed back at the camera. To avoid blocking the LED, it is best to ceiling mount the camera.

Choice of an LED with wide radiation pattern makes the system less sensitive to the attitude of the pen. A microswitch senses contact with the writing surface. Clearly, the pen must be in view of the camera and the light emitter must be as close as possible to the pen contact point. While at first this seems a serious drawback, we found that it encourages good habits of the stylus user: if the emitter is visible to the camera it is also visible to the audience (if any) . Other pen designs were evaluated, including those that rely on IR reflection off the surface. The layout of the pen is shown in Figure 3.

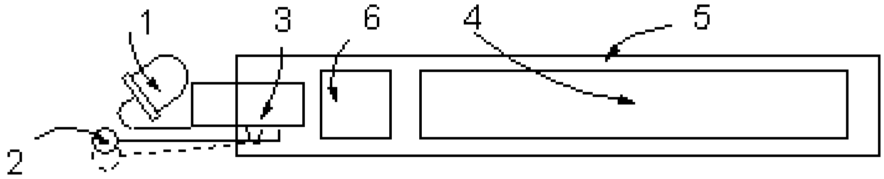


Figure 3 Active light pen with backward-looking light source.

The IR signal of the pen may be of many types. In a computer lab environment, DC IR suffices to capture the pen's location. For enhanced robustness, carrier-modulation techniques such as those used in remote control units are suitable. The modulation can be used to envcode user id's. Attaching the userid to the pen has the advantage that it does not have to be discovered by the system. The downside is that the user's identity is then easily confused when users exchange pens without expressing their identity to the pen, for example through the use of switches, magnetic reader and similar techniques. A further complication of a modulated IR system is that it requires some form of synchronization with the capture system. Mizui[4] offers an interesting solution to the synchronization problem. Unfortunately, the low-cost cameras we prefer to use typically do not offer synchronization options.

3.3 Spot finding and tracking

The current position of the light pen is acquired by means of computer vision. That is, a video camera is used to capture pictures of the light pen, and coordinates of the pen are then obtained by processing the video signal from the camera. As a general requirement, the processing must be fast enough to track movement of the light pen in real time. Fortunately, this can be simply achieved by a thresholding process since the light pen always produces a bright spot in a video frame. Figure 4 shows the position of the light pen in the current video frame and its movement to the adjacent video frame.

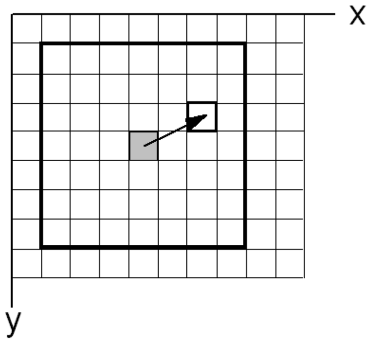


Figure 4 Spot tracking principle. If the current spot is the gray location, a search is done in a sufficiently large square neighborhood (here shown as 7 x 7). The arrow indicates the motion vector from the current spot to the newly found spot belonging to the next captured video frame.

The initial position of the light pen is found by full-searching the first video frame for the brightest pixel that exceeds a pre-determind threshold $T = (1-a)M$, where a is usually chosen to be .2, see Figure 5. Positions of the light pen in consecutive video frames are extracted using a local search, limited to a square neighborhood of the current position, to speed up the spot finding. This is possible because adjacent video frames are highly correlated when using a high capture rate. In practice, however, the brightest pixel in a video frame sometimes does not represent the accurate position of the light pen due to ambient light.. For example, a LED typically generates a light spot that has a Gaussian-like distribution in density as shown in Figure 5, if the LED is aimed directly at the camera. If the light pen is not exactly facing the camera, then the density distribution shape will change. Therefore, to improve the accuracy of the spot finding algorithm, we search for a group of brightest pixels instead of a single brightest pixel and then find the final current position of the light pen by averaging the coordinates of the pixels in the group defined as those pixels with intensity greater than the threshold T.

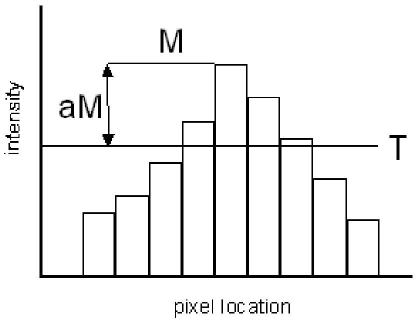


Figure 5 To find the current pen position we average the position of those pixels that exceed a given threshold intensity.

3.4 Stroke smoothing

Despite the precautions taken in acquiring and tracking the light spot, the spots captured by the above vision algorithm usually do not produce a sufficiently smooth stroke. To overcome this problem, a simple low pass filtering algorithm is applied to the captured coordinates of the light pen. The basic concept of filtering can be described by the following formula

$$p(x_i, y_i) = \frac{1}{N} \sum_{j=0}^{N-1} a_i \; q(x_{i-j}, y_{i-j}), \qquad (1.1)$$

where $p(x_i, y_i)$ denotes the filtered current position, a_i are waiting factors, $q(x_{i-j}, y_{i-j})$ the previous position and N is the number of taps of the filter. Usually, N=2 is sufficient.

3.5 Stroke formation

We define a stroke as a set of connected dots from the above vision capture system. Since the light pen only has two states, on and off, the starting point and ending point of a stroke cannot be directly obtained from the light pen. It therefore requires additional image processing to determine start and end of the light pen device. Starting of the light pen always corresponds to a transition from off to on, while ending of the light pen always to a transition from on to off. Based on this observation, the start point and end point of a stroke can be found by detecting lighting transition of the light pen.

In our experimental system, an off-the-shelf color videoconferencing camera (Videolabs Flexcam) and an IBM Intellistation M Pro with dual 400 Mhz Pentium processors are used to capture the location of the light pen. We block the visible enviromental lights with an IR pass/visible block filter (Schott Glass RG-830) RG-830) and inspect only the red component of a video frame. The capture card is an Osprey 100 which is capable of producing a 640 x 480 video frame resolution at a rate of 30 frames per second. With our above image processing algorithm, we can successfully capture approaching the maximum capture rate. In our experiment, we found that the threshold can be set to less than 50% of saturation. And a 2nd order low pass filter produces a satisfactory result. It should be pointed out that a threshoulding based spot finding algorithm, in general, is sensitive to the enviromental lighting. However, with IR optical filtering we successfully eliminated the effects of environmental lighting and we observed that our vision algorithm is insensitive to the threshold.

3.6 Calibration

The details of the calibration are as follows. When the projector and camera are separate entities, an intial calibration is required which requires only a few seconds of user participation. Essentially, the user is prompted to click on a number of fiducial points with an active, IR light-emitting pen. The system then estimates the projective transformation which causes the projected light output to most closely follow the pen motion. A minimum of four points is required; we currently use 9 points in the projection plane[1]. The relationship among the physical whiteboard and the camera's capture plane is as shown in Figure 6.

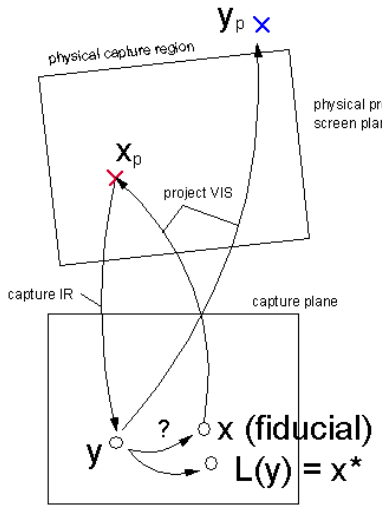


Figure 6 Illustration of the calibration procedure. Fiducial points $\mathbf{x}$ are given in the camera capture plane. The projector projects these at location $\mathbf{x_p}$ in the physical whiteboard plane. The capture system images $\mathbf{x_p}$ at location $\mathbf{y}$ in the capture plane. We seek the projective map $\mathbf{L}$ such that $\mathbf{L}(\mathbf{y}) = \mathbf{x}^*$ and $\left\|\mathbf{x} - \mathbf{x}^*\right\|$ is minimum.

Referring to Figure 6, the relationship between the i-th observed point $\mathbf{y}_i$ and the i-th fiducial point $\mathbf{x}_i$ is

$$\mathbf{L}(\mathbf{y}_i) = \mathbf{x}_i, \qquad i = 1, 2, \ldots N \tag{1.2}$$

where

$$\mathbf{L}(\mathbf{y}) = \frac{\mathbf{Ay} + \mathbf{b}}{\mathbf{c}^{\mathbf{T}}\mathbf{y} + 1} \tag{1.3}$$

and

$$\mathbf{A} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}, \tag{1.4}$$

and

$$\mathbf{b} = \begin{Bmatrix} b_1 \\ b_2 \end{Bmatrix}, \qquad \mathbf{c} = \begin{Bmatrix} c_1 \\ c_2 \end{Bmatrix}. \tag{1.5}$$

If we define the projective map using the vector

$$\mathbf{z} = (m_{11}, m_{22}, m_{21}, m_{22}, b_1, b_2, c_1, c_2)^T \tag{1.6}$$

then the calibration equations are

$$\mathbf{M}(\mathbf{z}) = \mathbf{n} \tag{1.7}$$

where

$$\mathbf{M} = \sum_{i=1}^{N} \begin{pmatrix} y_1^i & y_2^i & 0 & 0 & 1 & 0 & -x_1^i y_1^i & -x_1^i y_2^i \\ 0 & 0 & y_1^i & y_2^i & 0 & 1 & -x_2^i y_1^i & -x_2^i y_2^i \end{pmatrix} \tag{1.8}$$

and

$$\mathbf{n} = \sum_{i=1}^{N} \begin{Bmatrix} x_1^i \\ x_2^i \end{Bmatrix} \tag{1.9}$$

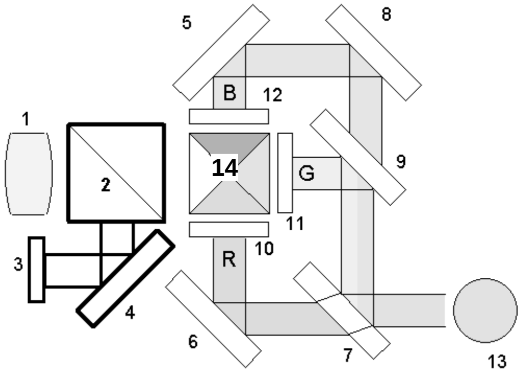
Although (1.3) is nonlinear, (1.7) is linear and is conveniently solved using LU-decomposition. The least squares solution to (1.7) is

$$\mathbf{z}^* = \left(\mathbf{M}^{\mathbf{T}}\mathbf{M}\right)^{-1} \mathbf{M}^{\mathbf{T}}\mathbf{n} \tag{1.10}$$

The system maintains its calibration as long as the screen camera and projector - and their settings - do not change.

4. TRACKING PROJECTOR

It occurred to us that the primary source of calibration errors is the fact that the camera and/or the projector may not at all lie along the normal of the projection surface and that the projector and camera optics may be vastly different. This led us to the idea of combining the camera and projector such that they share a single set of refractive elements. That would also take care of lens aberrations such as radial distortion that cannot be compensated by the projective mapping (1.3). The internal elements of such a camera/projector are shown in Figure 7. Most of the components shown constitute a digital projector using LCD shutters. The new elements – shown in bold lines allow the projector to look at its own projected image, or, as is done here, just at the IR band. One of the challenges of designing a combined camera/projector is keeping the distance from the objective to the LCD shutters short to avoid an excessively long projection distance.



5. USERID

When using a large whiteboard it is more likely that more than one user contributes to the whiteboard data. In some settings it may be desirable to know “Who wrote this stroke.”
We indicate three principles on which stroke user identification can be based.

1. Biometrics
2. Pen personalization
3. Personal Area Network (PAN)

The first method has the advantage of requiring no action from the user during use of the lightpen. On the other hand a training phase is required which may be lengthy and tedious. If the stroke coordinates contain timestamps the identity of the user may also be inferred from the dynamics of the handwriting.
The first principle has the advantage of requiring no action from the writer. Unfortunately, this method vastly increases the amount of information per stroke, has significant latency and misclassifies a few percent of the writer's strokes. The second method, pen personalization, requires registration: an action that associates a pen with a user. For example, the user may have to input a number into the pen or click on his name in the graphical user interface. The pen then transmits this code to the capture system. One can personalize the pen in several ways. One of these involves carrier modulation of each IR light. Another uses a unique spatial light pattern that can be template-matched by the vision system. We mentioned the added complexity of carrier modulation. Phase locking of the pen and capture system is not strictly needed, see for example Mizui[4].
The third method, PAN, also requires no conscious user effort once each user is on the PAN. The mere act of touching the pen would establish the user/pen link required by the whiteboard system.

Finally, if there is only one user at the whiteboard the userid defaults to the "console userid." This is the case in Microsoft's NetMeeting and Lotus' Sametime collaboration products.

6. CONCLUSION

We have implemented a portable whiteboard system using front projection and a vision-based stroke capture system. The system uses an off-the-shelf PC video camera and PC projection system. We have applied the system to recording concurrent audio and stroke information. In order to alleviate the sensitivity of the calibration to any perturbation of the optical parameters of the capture and projection optics, we proposed a modification to digital projectors. The modification places the capture sensor inside the projector such that it shares the projection optics. Finally, we outlined solutions to the question "who wrote what?" in the context of stroke based user input. Attaching identification information to stroke data is of benefit in online and offline processing, such as retrieval of stroke based data.

ACKNOWLEDGEMENTS

We gratefully acknowledge valuable discussions with Juerg von Kaenel and Stu Feldman on issues regarding collaborative whiteboarding and Fuad Doany for help with the optical design of tracking projectors.

REFERENCES

1. Mimio: htttp://www.mimio.com
2. e-Beam: http://www.e-beam.com
3. S. Elrod, R. Bruce, R. Gold, D. Goldberg, F. Halasz, W. Janssen, D. Lee, K. McCall, E. Pedersen, K. Pier, J. Tang, B. Welch," Liveboard: A Large Interactive Display Supporting Group Meetings, Presentations, and Remote Collaboration," *Conference proceedings on Human factors in computing systems* , 1992, Pages 599 – 607.
4. S. Mizui, US Patent 5,901,236, "Method and Apparatus for Measuring Position by Image Processing," May 4, 1999.